

A critique of the use of hormesis in risk assessment

Kirk T Kitchin^{*1} and J Wanzer Drane²

¹*Environmental Carcinogenesis Division, National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, NC, USA;*

²*Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA*

There are severe problems and limitations with the use of hormesis as the principal dose-response default assumption in risk assessment. These problems and limitations include: (a) unknown prevalence of hormetic dose-response curves; (b) random chance occurrence of hormesis and the shortage of data on the repeatability of hormesis; (c) unknown degree of generalizability of hormesis; (d) there are dose-response curves that are not hormetic, therefore hormesis cannot be universally generalized; (e) problems of *post hoc* rather than *a priori* hypothesis testing; (f) a possible large problem of 'false positive' hormetic data sets which have not been extensively replicated; (g) the 'mechanism of hormesis' is not understood at a rigorous scientific level; (h) in some cases hormesis may merely be the overall sum of many

different mechanisms and many different dose-response curves – some beneficial and some toxic. For all of these reasons, hormesis should not now be used as the principal dose-response default assumption in risk assessment. At this point, it appears that hormesis is a long way away from common scientific acceptance and wide utility in biomedicine and use as the principal default assumption in a risk assessment process charged with ensuring public health protection. *Human & Experimental Toxicology* (2005) 24, 249–253

Key words: default assumption; dose-response; hormesis; risk assessment

A special 2001 issue of *Critical Reviews in Toxicology* was guest edited by Drs Edward Calabrese and Linda Baldwin.¹ This issue contains about 340 pages of articles on hormesis, most written by Calabrese, and includes constructive criticism and commentaries by other individuals including Drs Wayne B Jonas, Kenny Crump and Arthur Upton.^{2–4} In this paper quotations from these three people together with those of Lave, Kitchin, and Christiani and Zhou will be used in our critique of the use of hormesis in risk assessment.^{5–7}

In this paper, our intent is to answer the question 'At what point, if any, could or should

hormesis be employed as the principal dose-response default assumption in risk assessment?'

A short answer would be: hormesis should not now be used as the principal dose-response default assumption in risk assessment. As noted by Lave,⁵ 'Thus, the task of demonstrating that hormesis is true is challenging, difficult and time consuming. We are therefore unlikely to see hormesis play an important role in regulation for many years.' We agree with Lave and think that at this point, hormesis is a long way away from common scientific acceptance, utility in biomedicine and use in risk assessment to ensure public health protection.

Statements about the future possible use of hormesis in risk assessment are very speculative, as are predictions of political elections. Some of the weaknesses of hormesis as a dose-response theory and/or basis for risk assessment are given below. Many of these weaknesses are well illustrated by quotations from prior articles on hormesis, which are numbered in presentation order and grouped by the contributing author. These quotations have been annotated in indented format by the authors (Kitchin and Drane)

Disclaimer: This paper has been reviewed in accordance with the policy of the National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, and approved for publication. We thank Steve Little and Chris Gordon for reviewing this manuscript as part of EPA clearance procedures. Approval does not signify that the contents necessarily reflect the views and policies of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

*Correspondence: KT Kitchin, MD-143-06, Environmental Carcinogenesis Division National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA
E-mail: kitchin.kirk@epa.gov

1. 'Most crows are black. If you see a white crow it is surprising and proves that all crows are not black. It tells you little else however.'²

Thus dose-response curves that appear to be, or actually are, hormetic do exist. This 'fact' does nothing to establish the prevalence, generalizability, utility or mechanism of hormetic dose-response curves.

2. 'Literature reviews, no matter how well done, cannot be better than the data contained in the original studies themselves.'²

The original dose-response studies cited as evidence of hormesis were not designed to prove, or disprove, the existence of hormetic dose-response curves. This makes literature reviews and interpretations based on such a literature review a *post hoc* matter and not an *a priori* test of a scientific hypothesis.

3. 'Publication is no guarantee of quality.'²

The scientific standards of repeatability by other laboratories, understanding at a rigorous scientific level and generalizability to the status of a scientific theory or law are not met by mere publication in one or more journal articles.

4. 'In the former one would want to assure proper dose verification, randomization of samples, blindness of outcome measures, proper statistical analysis, and full reporting of all data.'²

All five of these factors contribute to a quality dose-response study. Too often one or more of them are lacking in studies interpreted to be hormetic. For example, trying to do a current modern day statistical analysis of the 1951 Moskwa and Ber seedling growth study⁸ without the original data and using the results in a *post hoc* manner to argue for hormesis or for the use of hormesis in risk assessment is extremely problematic. Obtaining the needed future research funding to pay for such high quality *a priori* dose-response studies would be a difficult task.

5. 'Investigate models where hormesis does not occur to find out why.'²

If hormesis is a generalizable and unifying hypothesis, then it should occur either nearly 100% or 100% of the time. Thus we should ask the questions 'Why is hormesis not observed in every single dose-response study?' Not all dose-response studies result in hormetic dose-response curves. Using the data compiled by Calabrese and Baldwin,¹ 99.6% of the 20285 journal articles examined did *not* show positive evidence of hormesis. Why should a regulatory agency do

risk assessments based on a dose response theory (hormesis) which is not observed in many or all scientific experiments?

6. 'Although there are many convincing examples of hormesis, the overall prevalence of hormesis is an open question.'³

Some of the more convincing examples of hormesis may occur in the areas of dose-response curves with essential nutrients and some pharmaceuticals which are toxic at greater than therapeutic concentrations. However convincing some hormetic examples are, the prevalence of hormesis among chemical exposures in general is not known.

7. 'No matter how the Calabrese *et al.* database is evaluated, it is difficult to see how it can be used to estimate the prevalence of hormetic responses in general. As noted earlier, even if the number of the 1000+ studies demonstrating hormesis was known with certainty, this would only provide the numerator for a percentage. It is by no means clear what number should be used for the denominator.'³

Prevalence is a ratio of the number of 'cases' in a population divided by the total number of subjects, items or events in the population under investigation. An estimate of the prevalence is obviously the number of cases in the sample divided by the total sample size. Thus, Calabrese *et al.* have tried to estimate the prevalence of hormesis and have given us their best estimate of 86 hormetic journal articles in 20285 total journal articles examined (0.4% prevalence). Crump refers to 'the 1000+ studies demonstrating hormesis' as standing alone without adequate knowledge of the number of those studies reviewed that did not show evidence of hormesis. We know less about the number of dose-response curves that did not show evidence of hormesis or how representative the three journals studied (*Environmental Pollution*, *Bulletin of Environmental Contamination and Toxicology* and *Life Sciences*) are of the total biomedical literature.

8. 'The attempts at estimating the prevalence of hormesis reviewed herein did not adequately control for false positives...'³

In a prevalence study, cases showing evidence of hormesis must be validated to be hormetic instead of proceeding on the interpretation and belief that hormesis is present. Without validation, some of those impressions of hormesis will certainly represent false positives. A curve that appears to show hormesis does not mean hormesis is present. That is, at a very fundamental level there needs to be a definition or test of hormesis that would allow investigators to validate the presence of suspected hormesis regard-

less of the investigators' interpretive views. We do not have that definition which, from a mathematical and biophysical point of view, is a set of axioms representing an abstraction of the hormetic process. Those axioms could then be used to create conjectures, which when proved, become theorems. Those theorems, when subjected repeatedly to experimental evidence, either validate our presumed knowledge of the hormetic process or refute our current working model of dose-response relationships. Another way of begging for a definition based on fundamental scientifically valid axioms is the question, 'Will the experiment that showed 'hormesis' in the original study give the same results upon repetitions of the study?' This has been rarely demonstrated. If it is demonstrated even once, does that mean the mechanisms of hormesis are well understood? The answer currently is 'no.'

9. 'If the data set was the most hormetic looking out of 100 examined, then to conduct a statistical test for hormesis at the standard 0.05 level one should use $P=0.0005$ (the solution to $1 - (1 - P)\exp 100 = 0.05$) rather than $P = 0.05$.'³

If one wishes an experiment-wide $\alpha = 0.05$ when testing 100 data sets for positive evidence of hormesis, the probability that one or more data sets will be truly positive is 0.05 only if α for each of the 100 independent data sets is tested at the 0.0005128 level of probability. Conversely, if α is set equal to 0.05, as is often done in *post hoc* analysis, then testing 100 data sets for hormetic evidence will show positive results 5 ± 2.1794 times out of the hundred trials (mean \pm standard deviation). Thus, finding a single statistically significant positive hormetic dose-response curve in such a situation is hardly substantial positive evidence for hormesis.

10. 'Calabrese and Baldwin¹ selected only data sets that *a priori* appeared to be hormetic, so that there is no way to control for the false positive rate, or to generalize the conclusions from their database. Also, they analysed their database using an *ad hoc* scoring system that is difficult to interpret and does not control the false positive rate.'³

Remarks on false positives are under quotation no. 8 above. The absence of non-cases (in which hormesis is not observed) reduces their investigations to purely descriptive or observational studies, as there is no referent or control group (in the normal epidemiological sense). Even if the scoring system appears to be a good one, there is no way of determining how investigations with nonhormetic outcomes would score. Relative to normal models of the scientific method, investigations of hormesis are largely missing

three very important components of the scientific method (steps 2, 3 and 4), which looks somewhat like:

1. Observation \rightarrow 2. Induction \rightarrow 3. Abstraction
 \rightarrow 4. Deduction \rightarrow 1. Observation \rightarrow 2. Induction
 \rightarrow (etc.)

(and the cycle continues as long as the working scientific models can be improved and refined). So far evidence for hormesis has largely been advanced in respect to step 1 (observation), but evidence for hormesis is inadequate in steps 2, 3 and 4 (induction, abstraction and deduction).

11. 'There is no basis in a "superior" science such as physics, chemistry, biochemistry, endocrinology or pharmacology that explains what hormesis is at the level of atoms, molecules and/or cellular macromolecules.'⁶

The present theory of hormesis lacks the details required to build a sophisticated multicomponent dose-response model. Words and biological concepts like evolutionary pressure, modest overcompensatory reaction, stimulated immune system, antimutagenic biosystem and adaptive response do not provide the needed level of scientific sophistication for quantitative risk assessment. Greater scientific detail is required to build a strong mathematical and scientific theory useful to extrapolate beyond available experimental data. Experimental data needed to justify particular hormetic mechanisms often do not exist. Other opposing dose-response theories (receptor-ligand, threshold, one hit, multihit etc.) do much better at meeting this high standard.

12. 'A problem that hormesis has in being more scientifically accepted is (a) proving that only one mechanism accounts for both the "beneficial" and "toxic" parts of the biphasic dose-response curve and (b) giving substantial evidence against the interpretation that "hormesis" is the sum of many different mechanisms which add up to either "beneficial" or "toxic" in two different parts of the dose-response curve. Some examples of hormesis may consist of an initial beneficial dose region where several mechanisms are operating (just for the sake of argument let us say 3 mechanisms) and the overall sum of these 3 mechanisms is "beneficial" to the organism. At higher, toxic, doses, many more mechanisms are operating (just for the sake of argument let us say 8 mechanisms) and the sum of all these 8 mechanisms puts the organism in the "toxic" part of the biphasic dose-response curve.'⁶

There are many examples of hormetic dose-response curves that may be multicomponent in nature. Also there are several good compilations of important biological defence mechanisms which may contribute to experimentally observed hormesis. Several examples are given below:

- a) In the 1999 article by Pollycove and Feinendegen,⁹ at least nine major defence mechanisms are mentioned – reduced glutathione, superoxide dismutase, catalase, peroxidase (antioxidant prevention), the many different enzymes of the repair of DNA damage and removal of persistent DNA alterations by apoptosis, differentiation, necrosis and the immune system. Considering the known biological complexity of DNA repair enzymes, apoptosis, differentiation, necrosis and immune surveillance, it is likely there are at least 100 biological defence mechanisms. Why then should we view an observed hormetic dose-response relationship as anything more than a sum of many different effects and processes?
- b) Teeguarden et al. list nine defence mechanisms that can lead to apparent hormesis (attenuation of uptake processes, increased excretion, reduced bioactivation, increased detoxification, altered disposition, competition for receptor, cell cycle kinetics (DNA repair, cell proliferation, apoptosis), receptor upregulation and/or down regulation and immune response).¹⁰
- c) Recently, a mechanism and modelling based argument for multiple, rather than single, component processes contributing to hormesis has been developed by Conolly and Lutz.¹¹ Their examples include i) antagonistic action of two adenosine receptor subtypes, ii) homo- and heteroligand dimers of androgen receptor complexes, iii) induced DNA repair by a treatment chemical and iv) cell division delay caused by a treatment chemical.

13. 'Even if hormesis is biologically true, its assessment is limited due to the difficulties of study design, biological markers selection, statistical power considerations, model and end-point selection and risk model approaches.'⁷

First of all being true (in the sense of one or more positive examples) does not mean being universally true. Being universal does not mean being measurable. But let us suppose for the sake of argument that hormesis is biologically valid, universal and measurable. These three properties simply set the stage for well defined *a priori* experiments to measure the presence or absence of hormesis and test the hormetic hypothesis over different regions of the experimental dose range (the *x* axis). Without a mathematical or statistical model, however, the experiment will want for statistical power to reject the hypothesis of nonhormetic responses

in favour of the interpretation of a positive hormetic response.

In the papers reviewed by the Calabrese group,¹ there are many examples of measurable responses that appear to be hormetic. There are sufficient numbers of defence mechanisms that ought to be able to lead to hormetic dose response functions under some circumstances. But at present, there is no experimental evidence that hormesis is universal. Furthermore, mathematical abstraction from experimental observations (such as exists in the area of receptor-ligand theory, for example) is not present. It is possible that in the future several plausible theories of hormesis might be developed. Usually, that is the case before a scientific theory can be more fully developed and accepted. Finally, Calabrese and Baldwin do not provide a statement as to why the positive evidence for hormesis is absent.¹ For those experiments where there appears to be evidence of hormesis, there is no valid way of ascertaining the validity of the assertion of the presence of hormesis.

14. 'In light of the foregoing findings, national and international study groups generally have concluded that given appropriate adjustments for the dose, dose rate, and quality of radiation, the weight of evidence supports the use of the linear-nonthreshold dose-response model for radiation protection purposes in assessing the risks of mutations, chromosome aberrations and certain types of cancer in populations exposed to low-level ionizing radiation.'⁴

Study groups concerned with the protection of population from the risks of chemical exposures have generally used linear or threshold types (with safety or uncertainty factors) of risk assessment procedures. To date, hormetic approaches to radiation or chemical risk assessment have not been widely accepted in scientific circles or employed by government regulatory agencies which are charged with protecting public health.

Conclusion

There are severe problems and limitations with the use of hormesis as the principal dose-response default assumption in risk assessment. These problems and limitations include:

- a) unknown prevalence of hormetic dose-response curves (items 6–10, 13);^{1,3}
- b) random chance occurrence of hormesis and the shortage of data on the repeatability of hormesis (items 2, 3, 8, 10);
- c) unknown generalizability of hormesis (in part because the prevalence of hormetic dose-response curves is unknown) (items 1–10, 12, 13);^{1–3}

- d) severely limited generalizability of hormesis because numerous examples of nonhormetic dose-response are known (items 1 and 5);^{1,2}
- e) the argument for using hormesis as the principal default assumption in risk assessment is based on *post hoc* rather than *a priori* testing of the hypothesis of hormesis (items 2, 3 and 7);³
- f) with retrospective *post hoc* searching of the scientific literature for hormetic dose-response curves, 'false positive' data sets may be a very significant problem (items 6–8, 10);^{1,3,7}
- g) the mechanism of hormesis is not understood at levels of scientific detail similar to the detailed alternative theories of dose-response such as Michaelis–Menten, receptor-ligand binding, single and multiple hit models or the Moolgavkar–Knudson carcinogenesis model (item 11);^{6,7,12}
- h) some observed cases of hormesis may merely be the overall sum of many different dose-response curves – some beneficial and some toxic (item 11).⁶

References

- 1 Calabrese EJ, Baldwin LA. Hormesis: a generalizable and unifying hypothesis. *Crit Rev Toxicol* 2001; **31**: 353–424.
- 2 Jonas WB. A critique of 'The scientific foundations of hormesis'. *Crit Rev Toxicol* 2001; **31**: 625–29.
- 3 Crump K. Evaluating the evidence for hormesis: a statistical perspective. *Crit Rev Toxicol* 2001; **31**: 669–79.
- 4 Upton AC. Radiation hormesis: data and interpretations. *Crit Rev Toxicol* 2001; **31**: 681–95.
- 5 Lave LB. Hormesis: policy implications. *J Appl Toxicol* 2000; **20**: 141–45.
- 6 Kitchin KT. Defining, explaining and understanding hormesis. *Hum Exp Toxicol* 2002; **21**: 105–106.
- 7 Christiani DC, Zhou W. Hormesis: the new approach in risk assessment? *Hum Exp Toxicol* 2002; **21**: 399–400.
- 8 Moskwa Z, Ber A. The influence of organic solvents on the growth of plants. *Experientia* 1951; **7**: 136–37.
- 9 Pollycove M, Feinendegen LE. Molecular biology, epidemiology, and the demise of the linear no-threshold (LNT) hypothesis. *Comptes Rendus Acad Sci Paris Life Sci* 1999; **322**: 197–204.
- 10 Teeguarden JG, Dragan Y, Pitot HC. Hazard assessment of chemical carcinogens: the impact of hormesis. 2000; **20**: 113–20.
- 11 Conolly RB, Lutz WK. Non-monotonic dose-response relationships: mechanistic basis, kinetic modeling, and implications for risk assessment *Toxicol Sci Rev* 2004; **77**: 151–57.
- 12 Ruffolo RR Jr. Review: the important concepts of receptor theory. *J Auton Pharmacol* 1982; **2**: 277–95.