

Defining hormesis

EJ Calabrese* and LA Baldwin

Environmental Health Sciences, Morrill I, N344, University of Massachusetts, Amherst, Massachusetts 01003, USA

Much confusion surrounds the concept of hormesis and what its biological meaning represents. This paper provides a definition of hormesis that addresses its historical foundations, quantitative features, and underlying evolutionary and toxicologically based mechanistic strategies. Hormesis should be considered an adaptive response characterized by biphasic dose responses of generally similar quantitative features with respect to amplitude and range of the stimulatory response that are either directly induced or the result of compensatory biological processes following an initial disruption in homeostasis. Given the limited magnitude of the stimula-

tory response (i.e., usually 30–60% greater than controls at maximum), heightened study design and replication requirements are often necessary to ensure reliable judgments on causality. Even though hormesis is considered an adaptive response, the issue of beneficial/harmful effects should not be part of the definition of hormesis, but reserved to a subsequent evaluation of the biological and ecological context of the response. *Human & Experimental Toxicology* (2002) 21, 91–97.

Key words: adaptive response; biphasic; hormesis; overcompensation; risk assessment; U-shaped

Introduction

The phenomenon of hormesis is becoming more broadly discussed in the biomedical literature, especially in toxicology and radiation biology/health physics as well as in the general scientific and lay literature. What characterizes much of this literature is the lack of a generally agreed upon definition of hormesis with respect to conceptual understanding, quantitative features, mechanistic framework, and biological significance. A plethora of terms has been applied to similar descriptive dose–response phenomena such as beneficial effects of low doses, intermediate disturbance hypothesis, subsidy–stress gradient, U-shaped, J-shaped, biphasic, stimulatory–inhibitory, facilitation–inhibition, reverse, bidirectional, dual, bell-shaped, compensatory and paradoxical dose responses as well as a string of biological ‘laws’ including those of Hebb,¹ Yerkes–Dobson^{2,3} and Arndt–Schulz.⁴ Such terminological diversity for similar-appearing descriptive dose–response phenomena reflects, at least in part, significant professional/academic isolation and lack of conceptual integration across scientific disciplines. This lack of consistency impedes progress to design and test hypotheses related to this phenomenon and to differentiate and generalize complexities of biological responses to low-dose exposures.

The current paper offers a definition of hormesis that is based on a comprehensive assessment of the historical literature relevant to the concept of hormesis in the chemical and radiation domains from the late nineteenth to the middle of the twentieth century^{4–8} and an assessment of several thousand articles with evidence of hormetic effects based on quantitative evaluation criteria.^{9–11}

Decoupling beneficial effects from the definition of hormesis

The concept of a beneficial effect within the context of a dose–response study is difficult to determine due to considerable biological complexity and the fact that beneficial effects are often seen with reference to a specific and relative setting. What may be beneficial for the individual due to low-dose exposures may be harmful for a population. Longevity may be enhanced at low doses but at the expense of fecundity or the reverse. What may be beneficial may be different when assessing the effects of the treatment on the host or the attacking organism. A cancer chemotherapeutic drug may be effective at high doses due to inhibitory effects on cell proliferation, but harmful to the patient at lower doses where it may stimulate cell proliferation and therefore tumor growth. In this case, the low dose may be assumed to be harmful to the patient while enhancing the tumor.¹² In a similar situation, a high dose of antibiotic may be bactericidal, thereby permitting the

*Correspondence: Edward J. Calabrese, Environmental Health Sciences, Morrill I, N344, University of Massachusetts, Amherst, Massachusetts 01003, USA.
E-mail: edwardc@schoolph.umass.edu

patient to survive.⁴ However, at lower doses the treatment may enhance the survival of the bacteria to the detriment of the patient. Certain cardiac glycosides may enhance cell proliferation of prostatic smooth muscle at low doses while having an inhibitory effect at higher doses. Yet, low-dose exposure may enhance the likelihood of functional impeding of urine flow in males.¹³ In this case, the hormetic-stimulatory response at low dose would not be beneficial for the patient. These examples illustrate that the definitional characterization of hormetic dose responses as a beneficial effect at low doses is often complex, situation specific, sometimes overly simplistic, encouraging of ideologically based support or criticism of the hormesis concept and therefore not generally useful. This does not mean that beneficial or harmful characterizations should not be made. Such judgments need to be made, but at a subsequent and more advanced stage of analysis.

Unravelling historical confusions: is the hormetic stimulation the result of a direct stimulation or an overcompensation response?

In the early-to-mid-decades of the twentieth century, a significant issue in the area of radiation-induced biological effects emerged as to whether reported stimulatory responses due to low doses of radiation were the result of a direct stimulation (i.e., often referred to as a biopositive effect) or an overcompensatory response following injury. The Arndt-Schulz law, which was based on the research of Hugo Schulz in the 1880s, assumed that a direct stimulatory response accounts for the low-dose stimulatory phase of the dose response. This was viewed by leading experts in radiation biology/health such as Shields Warren during the 1940s–60s as incompatible with substantial experimental data indicating that stimulation caused by low-dose radiation exposure occurred only as a result of an overcompensation reparative response to an initial disruption in homeostasis (see Calabrese and Baldwin⁷). Thus, the Arndt-Schulz law was essentially discounted by mainstream radiation health researchers. The lack of resolution of this issue has continued to the present time and provides a principal basis for current confusion over the concept of hormesis.

Defining hormesis

Hormesis is an adaptive response characterized by biphasic dose responses of generally similar quan-

titative features with respect to amplitude and range of the stimulatory response that are either directly induced (i.e., direct stimulation hormesis [DSH]) or the result of compensatory biological processes following an initial disruption in homeostasis (i.e., overcompensation stimulation hormesis [OCSH]).

Overcompensation stimulation hormesis

Overcompensation hormesis is an adaptive response to low levels of stress or damage resulting in enhanced fitness for some physiological systems for finite periods and, under specific defined circumstances such as colony growth, indefinitely. It results from a modest overcompensation to a disruption in homeostasis. The key conceptual features of OCSH are the disruption of homeostasis, the modest overcompensation, the reestablishment of homeostasis and the adaptive nature of the process. Figure 1 depicts the general form of the OCSH dose–response relationship including the temporal sequence of the dose response.¹⁴

The ‘disruption of homeostasis’ phrase establishes the toxicological nature of hormesis distinguishing it from the concept of essentiality of nutrients and DSH (Figure 1). Disruption of homeostasis, within the context of hormesis, is not restricted to gross toxicological damage whereby macromolecular changes predominate but should be more broadly seen as comprising a continuum from a general stress response, as evidenced by alterations in glucocorticoid levels, to those changes that include limited macromolecular damage. The ‘modest overcompensation’ feature of the process leading to the expression of hormesis is essential because it functionally links hormetic responses to homeostasis, a universal biological concept, providing the theoretical foundation for the broad generalizability of hormetic phenomena.

This modest overcompensation response suggests a highly regulated, optimization process providing additional adaptive equity as a type of biological insurance policy that remains after the costs of tissue repair have been satisfied. This concept implies a continuous responding to compensatory regulatory messages until the homeostatic condition is reestablished. Efficiency in reestablishing homeostasis demands that resources be appropriately allocated. Compensation responses should be quantitatively linked to the extent of damage incurred; that is, the repair response would correspond to the extent of the damage, with sufficient, but not excessive, biological resources allocated to ensure that the repair function is completed.

Hormesis represents the advantage gained by the individual from resources initially and principally

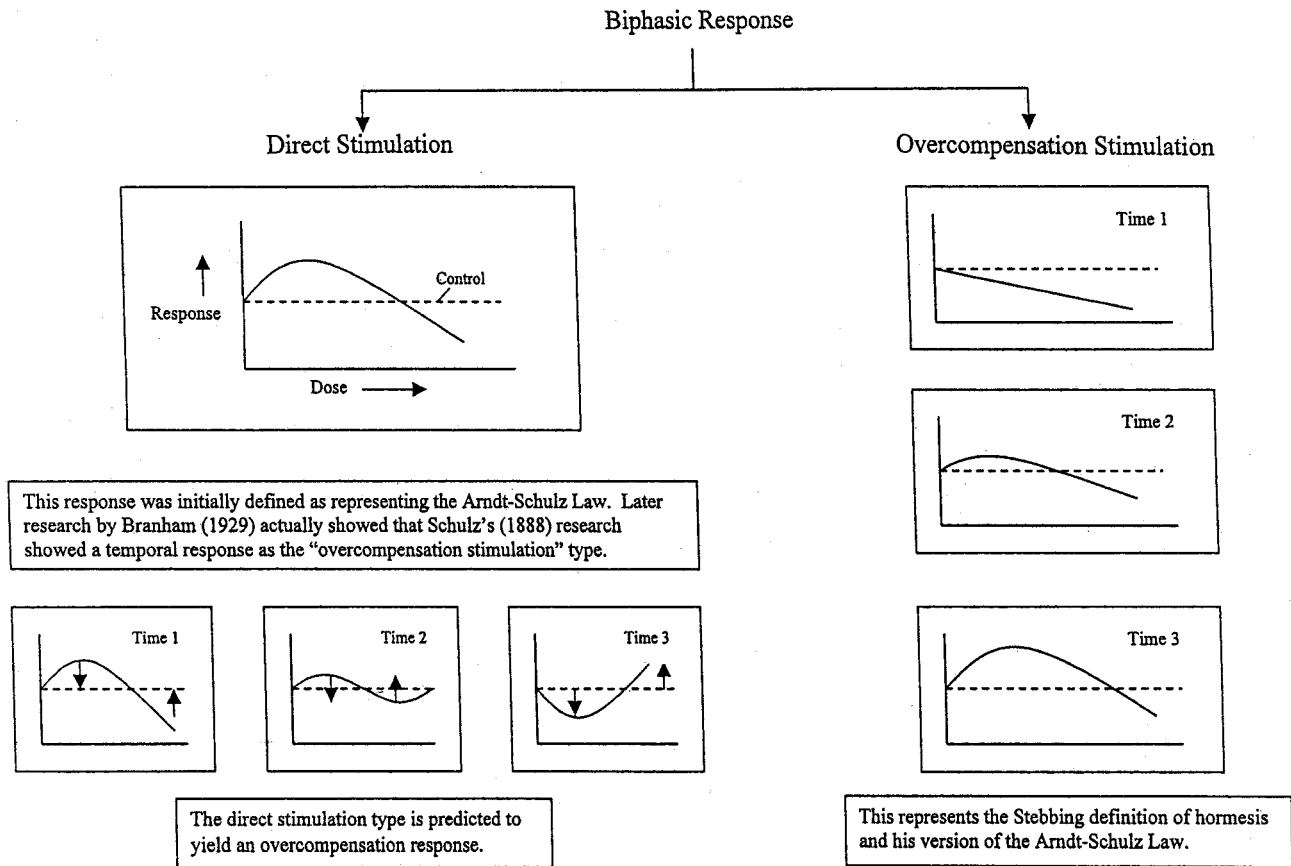


Figure 1 Comparison of direct stimulation and overcompensation stimulation hormesis (modified from Calabrese and Baldwin¹⁴)

allocated for repair activities but modestly in excess of that needed to repair the immediate damage. This process could also readapt the organism against damage from a subsequent and more massive exposure within a limited time period. Therefore, the limited overcompensation response may satisfy two functions: the assurance that the repair was adequately accomplished in a timely fashion and protection against subsequent and possibly more massive results. The value of this latter function is commonly assessed in chemical and radiation toxicological studies of the adaptive response. In this case a low dose (e.g., X-rays, many heavy metals, organic solvents such as carbon tetrachloride, endogenous compounds such as β -amyloid peptide) administered prior to a higher and more threatening dose of the same agent, often reduces the toxic potential of the subsequent massive exposure. Furthermore, if no subsequent toxic exposure occurs, the overestimated application of resources to the initial damage (i.e., the overcompensation response) may be employed for other useful functions (e.g., reducing background stressor damage, providing additional vegetative growth, etc.). This is, in fact, what is typically measured in studies assessing hormesis.

The modest extra resources to assure reestablishment of the homeostatic condition have been broadly adopted by many species. Despite this common adaptive strategy, various biological systems may have evolved different specific approaches to achieve the compensatory response, depending on the significance of the function needing restoration, the availability of resources, as well as the extent to which biological redundancy occurs in the affected systems. This is analogous to the case with other adaptive strategies such as enzyme-mediated xenobiotic detoxification/excretion processes where probably all species follow the general norm of converting lipophilic substances to more hydrophilic metabolites but may use different specific chemical substrate strategies (e.g., glucose versus sulfate, glycine versus glutamate) to achieve this hydrophilic metabolite detoxification/excretory goal. Thus, the process of natural selection of hormetic strategies within the diverse range of biological species is likely to follow a generally similar broad goal with specific strategies tailored to the unique ecological niche features of the species. Within an evolutionary paradigm of diversity linked to a common framework, the nature of the hormetic dose-response curve across

species is quantitatively consistent suggesting a high degree of genetically based conservation.

Direct stimulation hormesis

Examples of hormetic dose responses exist in which detailed temporal features were included without the observation of an overcompensating response. Such findings indicate that hormetic responses can occur via direct (biopositive) mechanisms. However, as suggested above, lack of temporal features in most studies precludes such differentiation.

DSH displays similar quantitative features as OCSH with respect to amplitude and dose range of the stimulatory response. This suggests that it is also tightly regulated with major resource constraints. The endpoints that are assessed represent functions that maintain normal multisystem responsiveness and homeostasis. The physiological systems and endpoints measured with DSH are often those reported in experiments in which OCSH is observed. This suggests that OCSH and DSH may be mediated via similar regulatory systems and, therefore, are bounded by similar resource and system plasticity constraints, accounting for their common quantitative features. However, the initial action that generates the DSH is not a response to a disruption in homeostasis but an adaptive response that operates within normal maintenance functions that allow for metabolic excursions within the twofold range of background. It would use fewer resources as compared to OCSH since there is no obvious damage to repair and disruption to overcome. Nonetheless, it represents a type of steady-state adaptive response that reflects normal, modulatory physiological dynamics.

Qualitative/quantitative features of hormesis

Qualitative features

Hormetic responses are characterized as biphasic dose–response relationships exhibiting a low-dose stimulation and a high-dose inhibition. That is, both the stimulatory and inhibitory dimensions of the hormetic phenomenon must be present to satisfy the qualitative definition of hormesis. This is necessary in order to establish the hormetic response within the traditional toxicological dose–response continuum. Dose–response relationships exhibiting stimulation at low doses but where the inhibitory response is not demonstrated either because the response at higher doses does not diminish below control values or because the upper end of the dose–response spectrum was not assessed do not satisfy this definition.

Whether the hormetic response displays a U- or an inverted U-shaped dose response is a function of the end-point measured. For example, an inverted U-shaped dose response would be observed when the endpoints were longevity or growth; a U-shaped dose response would be seen when the endpoints were disease incidence such as cancer or heart disease. Consequently, hormesis is a general term for biphasic dose–response relationships of a U- or inverted U-shaped nature.

Quantitative features

Further confusing the understanding of the term hormesis is that the historical use of this term did not define nor imply specific quantitative and temporal features of the dose–response relationship. Based on an investigation of several thousand published studies offering qualitative consistency with the hormetic dose–response relationship, Calabrese and Baldwin^{9,10} noted that such effects could be quantitatively characterized by a maximum stimulatory response that generally did not exceed twofold of the control with most maximum responses only 30–60% greater than controls. The width of the stimulatory response was typically (i.e., 90% of 2609 examples) in the 5- to 100-fold dosage range, immediately below the toxicity threshold; reliable exceptions to the 5- to 100-fold stimulatory dose range exist in which ranges $\geq 10^3$ -fold of dose have been reported.

The fact that the stimulatory zone can be so broad suggests that multiple mechanisms are involved. Furthermore, while evidence exists that overcompensation to a disruption in homeostasis may extend over a 100-¹⁵ to at least a 300-fold¹⁶ dose range, the direct stimulatory response may have the capacity to affect stimulatory responses over a range considerably larger than observed with the overcompensation-based phenomenon based on preliminary assessments of selected pharmacological dose–response systems.¹⁴ While this remains essentially an unexplored area, further subclassification according to the range of stimulation may be necessary, but as of yet there is insufficient biological understanding to guide on how to proceed.

Limited insight exists concerning why the stimulatory range of hormetic dose–response relationships can vary widely based on research in pharmacology and experimental psychology. In the field of pharmacology, the administration of parathyroid hormone to pancreatic islets cells *in vitro* affects a highly reproducible hormetic dose–response relationship concerning the release of insulin. However, if the level of calcium in the medium is changed it alters the nature of this hormetic dose response by changing the stimulatory range from approximately 8- to

100-fold but not the amplitude of the response.¹⁷ To our knowledge, this represents the first pharmacological/toxicological example of an experimental modulation of the stimulatory range of the hormetic dose–response relationship.

The range of the stimulatory response has been readily assessed in the field of experimental psychology where more complicated study designs have been routinely used. For example, it is common that the effects of different levels of stress on various types of performance are evaluated. However, experimenters often incorporate a second variable — tasks of different complexity — to be solved. In these experiments, the hormetic dose response is typically seen to have similar amplitude across the different levels of stress but the range of the stimulatory response is much more restricted under conditions of greater complexity. These types of hormetic-like dose–response relationships have been referred to as Hebb’s law (i.e., when there is a single level of complexity) and the Yerkes-Dodson law (i.e., when there are multiple levels of complexity).

These two models (i.e., release of insulin from pancreatic islet cells and those representing examples of the Yerkes-Dodson law), by which the range of the stimulatory response may be modified, have significant implications not only for the design of toxicological investigations but also for understanding the underlying mechanisms that account for the range of stimulatory responses in the low-dose zone.

Nomenclature

The term hormesis was selected to represent the biphasic dose–response phenomenon described here because of its widespread use in the fields of radiation biology/health physics and ecological and human toxicology. In addition to numerous articles, two books have been published written on the topic.^{18,19} Even though there has been a lack of precision/agreement over the meaning of hormesis it has been reasonably focused on and consistent with the currently proposed interpretation and quantitative characteristics. Other terms like U-shaped, J-shaped, biphasic, stimulatory–inhibitory, dual, and bidirectional are valuable but too general. The terms intermediate disturbance hypothesis and subsidy–stress gradient are more specific, and probably are examples of OCSH but need to be better demonstrated and assessed. While we believe that hormesis is a highly predictable process, its characterization as a ‘law’ is excessive and unnecessary. Thus, we believe that the term hormesis warrants the primary focus for common use in this area.

In Figure 2, a hormetic nomenclature is proposed that both recognizes the descriptive similarity of hormetic-like biphasic dose–response relationships, as well as general features of differentiation based on quantitative aspects of dose responses and temporal responsiveness. The three basic features involve hormetic-like responses in the presence or absence of temporal data. If appropriate temporal data are

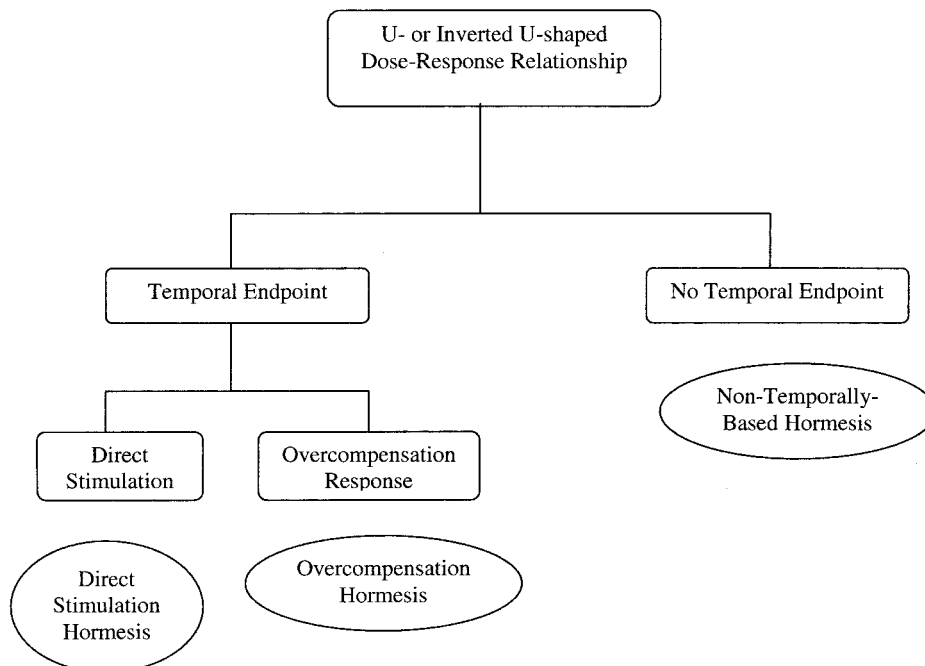


Figure 2 Schematic of hormetic nomenclature

available then it may be possible to differentiate between directly stimulatory or overcompensation stimulatory hormetic responses. A further common level of differentiation may be applied to the three general dose–response classifications based on quantitative features of the dose response concerning the magnitude and range of the stimulatory response. Under the assumption that hormesis represents a modest overcompensation to a disruption in homeostasis, there is a biologically based expectation that any overcompensation response would be limited; such responsiveness would assure that homeostasis would be efficiently re-established.¹¹ Based on such observations it is judged that maximum stimulatory responses greater than three- to fourfold are likely to represent different phenomena than hormesis.

Differentiating the definition of hormesis from the proof of hormesis

The principal problem with the above definition of hormesis is that of determining if the definition has been satisfied, especially the low-dose stimulation. The recognition that hormetic effects in the stimulatory range are likely not to exceed 30–60% of the control places heightened experimental requirements on claims that the dose response was a real stimulation not accounted for by normal variation. While there is no absolute guidance to be offered in this area, demands to derive a causality conclusion require consideration of the strength and appropriateness of the study design, adequacy of statistical power and reproducibility of findings. The demands on factors impacting decisions on proof become even more difficult when the temporal parameter is included because of the multidose, multitime period study design considerations. This is a significant contributory factor leading to the more limited number of studies that adequately document both the dose and temporal features of the hormetic phenomenon.

The challenge of proof also requires the use of a biological model and endpoint selection that can be assessed within the context of a hormetic dose–response relationship. That is, the endpoint must have the potential to display a biphasic dose–response relationship and temporal responsiveness. Animal models with disease incidence essentially negligible would be unable to assess the occurrence of possible biphasic responses. This is a serious experimental issue since some commonly used cancer and teratogenicity bioassay models have been selected in part because of a low background disease incidence. Likewise, if the initial comparison data

were normalized to 100% and these values could not be increased it would not be possible to estimate stimulatory responses, but only decreased responses. Thus, while it is essential to have a clear definition of hormesis it is also important that investigators interested in studying this phenomenon be properly guided with respect to model and end-point selection, temporal considerations and study design/statistical power and replication concerns.

Common strategy, but no single hormetic mechanism

The common features of hormetic dose–response relationships that are extremely widespread across the biological and toxicological sciences suggest a common regulatory strategy for biological resource allocation as well as plasticity of regulatory processes within the context of an evolutionary framework. Thus, even though the definition of hormesis is of a descriptive nature, its generalizability indicates the occurrence of basic biological regulatory processes and strategies.

The issue of whether there is a hormetic mechanism may be evaluated within the above framework. Current evidence suggests that the key feature of hormetic dose responses is that resource allocation must be carefully controlled and regulated via physiological set points linked to molecular switching mechanisms. This framework provides the basis by which direct stimulatory or overcompensation stimulatory effects are regulated and display similar hormetic-like biphasic quantitative dose–response relationships. Within this context, there is no expectation that a single hormetic mechanism would have evolved and be broadly applicable. While hormetic responses would be expected to occur in most tissues, precisely how such biological responses occur would be biologically framed within the unique endogenous and exogenous environments of each biological subsystem. While it is clear that hormetic dose–response relationships display limited amplitude variation, the range of the stimulatory response may be very broad. Such recognition is critical to understand for hazard assessment, risk assessment and therapeutic purposes.

Conclusion

This paper argues that hormesis is an adaptive response with distinguishing dose–response characteristics that is induced by either direct acting or overcompensation-induced stimulatory processes at low doses. In biological terms, hormesis represents

an organismal strategy for optimal resource allocation that ensures homeostasis is maintained. This strategy dictates the quantitative features of the dose–response relationship that typifies hormesis including the modest amplitude of the stimulatory response, the range of the stimulatory response and the relationship of peak stimulatory zone to the onset of toxicity regardless of mechanism by

which the low-dose stimulation originated (i.e., direct acting stimulation versus overcompensation stimulation following initial toxicity). Since numerous mechanisms have evolved to achieve this resource allocation regulatory goal, no single hormetic mechanism is expected, but a common evolutionary-based homeostasis maintenance regulatory strategy is evident.

References

1. Hebb DO. Drives and the C.N.S. (conceptual nervous system). *Psychol Rev* 1955; **62**: 243.
2. Broadhurst PL. Emotionality and the Yerkes-Dodson law. *J Exp Psychol* 1957; **54**: 345–52.
3. Teigen KH. Yerkes-Dodson: a law for all seasons. *Theory Psychol* 1994; **4**: 525–47.
4. Calabrese EJ, Baldwin LA. Chemical hormesis: its historical foundations as a biological hypothesis. *Toxicol Pathol* 1999; **27**: 195–216. (Reprinted in *Human Exp Toxicol* 2000; **19**: 31–40).
5. Calabrese EJ, Baldwin LA. Can the concept of hormesis be generalized to carcinogenesis. *Reg Toxicol Pharmacol* 1999; **28**: 230–41.
6. Calabrese EJ, Baldwin LA. Radiation hormesis: its historical foundations as a biological hypothesis. *Hum Exp Toxicol* 2000; **19**: 41–75.
7. Calabrese EJ, Baldwin LA. Radiation hormesis: the demise of a legitimate hypothesis. *Hum Exp Toxicol* 2000; **19**: 76–85.
8. Calabrese EJ, Baldwin LA. Tales of two similar hypotheses: the rise and fall of chemical and radiation hormesis. *Hum Exp Toxicol* 2000; **19**: 85–97.
9. Calabrese EJ, Baldwin LA. The dose determines the stimulation (and poison): Development of a chemical hormesis database. *Int J Toxicol* 1997; **16**: 545–59.
10. Calabrese EJ, Baldwin LA. A quantitatively-based methodology for the evaluation of chemical hormesis. *Hum Ecol Risk Assess* 1997; **3**: 545–54.
11. Calabrese EJ, Baldwin LA. The frequency of U-shaped dose–responses in the toxicological literature. *Toxicol Sci* 2001; **62**: 330–38.
12. Foekens JA, Sieuwerts AM, Stuurman-Smeets EMJ, Dorssers LCJ, Berns EMJJ, Klijn JGM. Pleiotropic actions of suramin on the proliferation of human breast cancer cells *in vitro*. *Int J Cancer* 1992; **51**: 439–44.
13. Chueh S-C, Guh J-H, Chen J, Lai M-K, Teng C-M. Dual effects of ouabain on the regulation of proliferation and apoptosis in human prostatic smooth muscle cells. *J Urol* 2001; **166**: 347–53.
14. Calabrese EJ, Baldwin LA. U-shaped dose–responses in biology, toxicology, and public health. *Annu Rev Pub Health* 2001; **22**: 15–33.
15. Garceau D, Yamaguchi N, Goyer R. Hepatic adrenoceptors involved in the glycogenolytic response to exogenous (–)-norepinephrine in the dog liver *in vivo*. *Life Sci* 1985; **37**: 1963–70.
16. Smith EC. Effects of ultra-violet radiation and temperature on *Fusarium*: II. Stimulation. *Bull Torrey Bot Club* 1935; **62**: 151–64.
17. Fadda GZ, Akmal M, Lipson LG, Massry SG. Direct effect of parathyroid hormone on insulin secretion from pancreatic islets. *Am J Physiol* 1990; **258**: E975–E84.
18. Luckey TD. *Hormesis with ionizing radiation*. Boca Raton, FL: CRC Press, 1980.
19. Luckey TD. *Radiation hormesis*. Boca Raton, FL: CRC Press, 1991.