Hormesis: Scientific Foundations

Edward J. Calabrese, Ph.D.

Department of Public Health
Environmental Health Sciences
Morrill I, N344
University of Massachusetts
Amherst, MA 01003
Phone: 413-545-3164
Fax: 413-545-4692
E-mail: edwardc@schoolph.umass.edu
Hormesis: Scientific Foundations

1. Introduction

I first became introduced to the concept of hormesis, but not the term, when I was in my third year of college. I was a student in a plant physiology class; one of our experiments was to conduct a dose response experiment in which a plant growth inhibitor was applied to the soil in which peppermint plants were grown. This laboratory exercise was simply designed to demonstrate the concept of a dose-response relationship in which plant growth would be progressively reduced at higher doses. One day toward the end of the semester the professor indicated that the peppermint plants were not only not showing reduced growth but were actually being stimulated. This was not supposed to happen; it was simply a laboratory exercise; he felt that we may have used the wrong chemical, perhaps mislabeled the plants, or perhaps it was possible we might have discovered something new, but this was unlikely. He invited students to explore this with him after the semester was over.

He encouraged me to repeat the study following his original directions and to determine if our group had actually followed them. As it turned out, in the subsequent experiments we were able to observe the high dose inhibition but also the stimulation if the dose was sufficiently lowered. Based on these data, it appeared that our group may have made a dilutional error when we made up our stock solutions, thereby leading to stimulation at the lower concentrations.

This experiment was replicated repeatedly with similar findings, up to eight times, at which point we seemed convinced that the effects were real. Similar experiments were then conducted in hydroponics rather than in soil in order to avoid the soil-chemical
interaction. Again, a low dose stimulation and a high dose inhibition was observed; we replicated these findings numerous times in separate experiments; the findings were published in the European journal, Physiologica Plantarum (Calabrese and Howe, 1976). While this was my first exposure to the concept of hormesis, I was still unaware of the term, never used it in this paper, and simply referred to the response as a low dose stimulation and a high dose inhibition.

I went on to become a toxicologist, focusing on high dose experiments and mechanisms of toxicity. Then many years later, I received a brochure for a conference on Radiation Hormesis in 1985. The phenomenon seemed, with its J-shaped curve, conceptually similar to the plant response so many years ago. I became interested and was invited to the meeting if I would summarize the literature on chemical hormesis. This activity refocused my attention on the concept of hormesis, something that had long be absent from my thinking and research activities. Yet over the past 20 years this concept has come to dominant my professional life.

While this conference got me reacquainted with the concept of hormesis, I was not intellectually inspired to pursue it. However, in 1989, Leonard Sagan and Sheldon Wolff had a debate in the journal Science on the topic of radiation hormesis (Sagan, 1989; Wolff, 1989). I read it carefully and felt that these articles simply repeated the debate of the conference of four years before. So I called Leonard Sagan and suggested that the field needed to better understand this concept and that notable progress had not been made in the past four years. We agreed to work together to create an organization called BELLE, which stands for the Biological Effects of Low Level Exposure, (www.belleonline.com), that would address the scientific foundations of hormesis.
It was the creation of BELLE, with its continuing focus on the topic of hormesis that finally got me to redirect much of my professional activities to the study and understanding of the nature of the dose response in the low dose zone and the concept of hormesis. Yet, strong progress was not made until 1996. It was at this time that I received a telephone call from the Texas Institute for Advancement of Chemical Technologies (TIACT). They wanted to fund a study on hormesis, determining for once and for all, so to speak, whether hormesis was real or not. When I visited the advisory committee I was told that they wanted to determine whether hormesis occurred in humans focusing on several agents in great detail. I told them that this was not the ideal way to approach the problem. I indicated that if hormesis were real, it should be evolutionarily based and seen across all forms of life from microbes, to plants to animals. I argued that the assessment should be broad, seeking generalization, not narrowly focused on several detailed case studies in humans. My approach was rejected and funding turned down. However, about two months later they reversed their decision and funded my more general approach. It was the initial funding from TIACT and then from U.S. governmental agencies, principally the U.S. Air Force, which has allowed me to pursue the hormesis question. Now let us explore concept of hormesis itself.
2. Definition of Hormesis

Hormesis is a dose response phenomenon characterized by a low dose stimulation and a high dose inhibition (Figure 1). The term hormesis (meaning to excite) was created in 1943 by Southam and Ehrlich in studies dealing with the effects of red cedar extracts on fungal metabolism. In the early decades of the 20\textsuperscript{th} century the hormesis concept was embodied in the Arndt-Schulz Law and Hueppe’s Rule, terms which have gradually faded from use. The shape of this dose-response curve may be either an inverted U- or a J-shape depending on the endpoint measured. In cases of endpoints such as growth, longevity, fecundity and memory retention, the response would be an inverted U-shape. In the case of disease incidence (e.g., tumor incidence) it would be a J-shape.

While hormesis has often been defined as a beneficial effect at low doses, a decision on whether a hormetic dose-response exists or not should precede a subsequent and independent decision on whether the response is potentially beneficial, harmful or of unknown health implications. A decoupling of the definition of hormesis from whether or not the stimulatory response is beneficial will help enhance the likelihood that the hormetic concept will be evaluated on its scientific merits rather than in a political or other context.

3. Quantitative Features of the Hormetic Dose Response

An essential feature of the hormetic dose response is that it is a normal component of the traditional dose response. That is, at doses above the no observed adverse effect level (NO(A)EL) the hormetic dose response and the traditional sigmoidal dose response are the same. At below NO(A)EL doses the threshold dose response model predicts that
there is no treatment-related difference between the control group response and the below NO(A)EL treatment group responses. In the threshold model, the only “difference” between the control value and the responses of doses at and below the NO(A)EL is understood to be random variation. In contrast, the inverted U- or J-shaped hormeric models predict a modest treatment-related response will occur immediately below the NOAEL, a response that is reproducible, frequently statistically significant and often explained by specific toxicological/molecular mechanisms.

3.1. Magnitude of Stimulatory Response

The hormetic dose response is best evaluated within the context of a dose-time response since the low dose stimulatory response often represents a modest overcompensation response following an initial disruption (i.e., toxicity) in homeostasis (Figure 2).

The overcompensation response results from biological compensatory processes that allocate resources slightly in excess of that needed to ensure a return to homeostasis. This “extra” allocation of resources (i.e., adaptive response) leads to the hormetic stimulation, accounting for why hormesis is optimally seen within a dose-time effect evaluation. Such dose-time-response features may also explain why the magnitude of the hormetic stimulation is limited to percentile rather than fold increases over control values. This framework also provides the experimental basis for why hormesis is often quite challenging to prove and reproduce. This is because the maximum stimulatory response is usually only approximately 30-60% greater than controls, often being difficult to distinguish from normal variation. Consequently, assessing the viability of an hormetic hypothesis typically requires: (1) the need for strong study designs with
multiple doses properly spaced below the NO(A)EL range, (2) larger numbers of subjects especially at lower doses to enhance improved statistical power, (3) a time component in the study design and (4) the greater than normal need to establish the reproducibility of the findings. In addition, mechanistic explanations have often been demanded to ensure biological/toxicological understandings and predictive applications.

Even though the time component is crucial in assessing hormesis, most investigations do not simultaneously include many doses along with multiple time points. Most relevant papers contain either one (i.e., multiple doses) or the other (i.e., multiple time points). However, the hormesis database has approximately one thousand examples of studies demonstrating hormesis using the study-designs of multiple doses and time points (Calabrese and Blain, 2005). In the absence of multiple time points the hormetric response could be easily missed and underestimated.

A number of studies reveal a low dose stimulation and a high dose inhibition without an apparent compensatory response. In such cases the stimulation occurs as a result of a direct stimulatory response. This is also an example of hormesis but one that most likely acts via a different mechanism than the overcompensation process.

Regardless of the biological model, the modest stimulatory response is the most significant feature of the hormetric dose-response, regardless of whether the response is a modest overcompensation to a disruption in homeostasis or of a direct stimulatory nature. However, there is little biological understanding of the factors that constrain the magnitude of the stimulatory response across such a broad biological diversity of models and endpoints. For example, the magnitude of the hormetric stimulation has not been
explained by the proportion of receptors occupied for an active stimulatory response pathway (Roy and Rai, 2004; Dong et al., 1991).

The survival value of a modest overcompensation stimulatory response may be evaluated within the context of its overall damage repair function. Based on the hormetic dose response, at low levels of injury more resources are allocated than required to repair damage and re-establish homeostasis. In contrast, compensatory responses are often inadequate for full repair following more massive damage. The modest “over” allocation of resources that result in the hormetic stimulatory response may be viewed as an inefficiency in which more resources were appropriated than were needed to simply repair damage. However, the preponderance of the evidence suggests that this interpretation is not correct; the induced hormetic response represents an altered phenotype with enhanced efficiency that not only repairs the induced damage but one that confers an additive spectrum of adaptations such as higher tolerance to subsequent challenge (as is the case with auto or heteroprotection phenomena in toxicology and as preconditioning as used in the biomedical/clinical literature) as well as one that leads to a reduction in background injury or both. Further, these changes often affect complex parameters such as growth, longevity and various types of behavioral performance. Hormetic responses, therefore not only enhance adaptive capacity, reduce and repair damage but also influence a broader spectrum of endpoints which have biological significance. The new phenotype achieves this adaptive state within a context of optimized and reallocation of resources when the biological system is slightly stressed. The broad generalizability of the modest nature of overcompensation response suggests a high level of regulatory control.
3.2. Width of Stimulatory Response

The dose width of the hormeric stimulatory response is a critical dose-response feature with toxicological, public health, medical and risk assessment implications. The width of the hormeric stimulatory response has been typically 20-fold or less of the zero equivalent point (ZEP) (i.e., dose where the response crosses the control value) dose and often less than 10-fold. The hormeric response is first observed as the dose decreases from the ZEP/NO(A)EL and may occur over a 5 – 10-fold range below the ZEP, returning toward control values at lower concentrations. The hormeric stimulatory response immediately below the threshold is expected since the compensatory range would be continuous with the above NOAEL response zone. The modest width of the stimulatory response probably results from the homogeneous nature of biological models used for study and experimental conditions which are constant across treatment groups. Nonetheless, there are many examples where the dose width of the stimulatory zone exceeds 100-fold immediately below the ZEP and even greater than 1,000-fold. While mechanistic explanation(s) accounting for the stimulatory zone width variation has yet to be adequately addressed, it is likely that much of this variation may be explained by population-heterogeneity. While the experimental emphasis on homogeneous biological systems may often lead to an hormeric response, a population response may be far less clear, if it represents a composite of heterogeneous subgroups.

3.3. Dose Range and Hormesis

While hormeric effects are frequently reported in the $10^{-5}$ to $10^{-9}$ M region (Calabrese, 2005a, 2005b; Calabrese and Baldwin, 2003, 2001), opioids commonly induce biphasic responses at concentrations as low as $10^{-12}$ to $10^{-15}$ M (Calabrese, 2001).
A more extreme case is reported by Roy and Rai (2004) on the effects of cAMP on immune function (phagocytosis) revealing a clear hormetic effect at least as low as $10^{-18}$ M. At this concentration there was about one cAMP molecule per 3,300 cells, suggesting the existence of chemically mediated bystander effects. In contrast to the above example, hormetic-like biphasic dose responses in A431 cells by EGF occur when approximately 50% of the high affinity receptors are occupied, amounting to about 3000 EGF molecules per cell. Despite such differences in agonist concentrations and biological systems between the two cell types, the quantitative features of the dose response relationships were remarkably similar, suggesting that such highly diversified biological systems use different mechanism-based biological tactics to achieve the similar strategic goal of constraining the hormetic features of the dose response within defined limits.

4. Frequency

The frequency of hormetic dose responses in the toxicological literature has been addressed by Calabrese and Baldwin (2001, 2003) using *a priori* entry and evaluative criteria. Such criteria were applied to all articles published in three journals (i.e., Environmental Pollution, Bulletin of Environmental Contamination and Toxicology, and Life Sciences) from their inception to the present. These journals were chosen because they publish articles on a wide range of toxicology models, endpoints and agents from environmental to pharmacologically oriented toxicology. This assessment provided an estimation of the frequency of hormetic effects but also addressed in a more qualitative manner the issue of generalizability. In general, the entry criteria required, at minimum, the presence of a LOAEL, NOAEL and several doses below the NOAEL. The evaluative
criteria required evidence of a low dose (below NOAEL) stimulatory response based on statistically significance or alternative criteria equaling or exceeding the statistical criteria in rigor.

Of the nearly 21,000 articles assessed, 98% did not satisfy the entry criteria principally because the experiments lacked an adequate number of properly distributed doses. Of the remaining 2% satisfying the entry criteria, approximately 40% satisfied the evaluative criteria. These findings represent the first objective estimate of the frequency of hormetic dose responses in the toxicological literature. The findings were particularly significant because the figure of 40% is high, especially within the context of the rigorous criteria employed. These findings suggest that hormetic dose response relationships are a central feature in the toxicological sciences.

The hormesis-frequency database was also employed to provide an evaluation of which dose response model was more frequent in the toxicological literature, the threshold or the hormetic model (Calabrese and Baldwin, 2003). Almost 1,800 doses below the NOAEL from 664 dose response relationships were evaluated. While the threshold model predicts a 1:1 ratio of responses “greater than” to “less than” the control response (i.e., random variation), a 2.5 to 1 ratio was observed. These highly statistically significant findings indicate that the threshold model failed to provide adequate prediction in the low dose zone while the hormetic model was very effective. These findings challenged the belief in the primacy of the threshold model in toxicology and offered considerable support for the hormetic dose response model. More recent investigations with approximately 57,000 dose responses with 2,200 chemicals from the U.S. National Cancer Institute (US NCI) all with five concentrations and a concurrent
control have confirmed the basic conclusion noted above that the hormetic dose response occurs approximately 2.5 times more frequently than the threshold model in fair head-to-head competition (Calabrese et al., 2006).

5. Generalizability of Hormesis

For hormesis to be a broadly generalizable biological concept, it should be evolutionarily based being observed across the spectrum of plant, microbial and animal models, reported in all organ systems and for numerous endpoints, and induced by essentially all classes of chemicals and physical stressor agents. Even though this is a high standard for generalizability, even in the biological domain, hormesis model predictions perform well in each of these domains (Calabrese and Blain, 2005).

A convincing case can also be made for the hormetic dose response being of fundamental biological importance and having potentially broad applications for toxicology, pharmacology and clinical medicine. The hormetic response is adaptive in nature, being selected to enhance performance in a broad spectrum of functional domains, thereby promoting the survival interests of the species. Since the hormetic dose-response has been incorporated into the basic biology of theoretically all organisms, such relationships are commonly found in functional and regulatory processes governing dozens of essential receptor systems as shown in the table (Table 1) that control the broad range of critical processes underlying most organismal activities.

As noted above, several hormetic databases have been created that deal with environmental and biomedical agents. These databases provide large numbers of examples derived from a broad array of biological models and experimental frameworks.
A series of figures have been selected to illustrate the generalizability of hormesis and the general consistency of these dose response functions (Figure 3). These are but a few of the many thousands of similar examples in the peer-reviewed literature demonstrating the occurrence of hormesis.

6. Reproducibility

The most critical factor in the evaluation of hormetic dose response relationships is the reproducibility of findings. Reproducibility is typically incorporated into the routine evaluation of experimental data by investigators as well as being addressed during the peer-review process. However, the challenges of reproducibility are heightened due to the modest size of the stimulatory responses.

In order to better assess the level of confidence investigators have in the reliability of their reported hormetic low dose stimulatory responses, major reviews on hormesis and immune responses and in tumor cell lines have included direct quotes on how these researchers considered the specific responses that illustrate hormetic biphasic responses (Calabrese, 2005, 2005a). Of the nearly 550 graphically illustrated hormetic dose responses in these reviews, over 90% acknowledged the biphasic nature of the dose response, with a large number devoting considerable focused assessment/discussion on the reproducible nature of the findings, possible mechanisms that may account for these responses as well as their clinical implications. Such detailed consideration of the reliability of the observed hormetic responses as provided by the investigators provide an insight to the degree of confidence that exists with respect to reproducibility of hormetic responses by the authors.
6.1. Hormetic Mechanisms

Since hormetic effects have been reported in a broad array of biological models, for numerous organs and endpoints and chemical/physical stressors, no single mechanism is likely to account for these phenomena. Only limited attempts have been made to assess underlying mechanisms of hormetic dose responses in toxicology. Most toxicological researchers who have reported the hormetic U- or J-shaped dose response have not attempted to experimentally define the mechanisms by which such shifts in the dose response occur. In contrast, the field of pharmacology, in which hormetic effects are also commonly seen, has developed the theoretical foundations and practical means to deconstruct and then reassemble such dose-responses usually with the assistance of synthetic agonists and antagonists of receptors which mediate the broad spectrum of hormetic biphasic effects. While there are numerous ways in which biphasic responses may occur, a basic framework is via the involvement of high and low affinity receptor subsets. That is, a single agonist with differential binding (i.e., high and low receptor affinities) affecting two opposite acting receptors will induce hormetic-like biphasic dose responses in numerous biological systems as has been shown for dozens of receptor systems as given in the Table 1.

7. Making Hormesis the Default Assumption in Risk Assessment

The key potential environmental application of the hormesis concept is its adoption as the default model used in risk assessment. As default models for risk assessment, the EPA employs the threshold model for non-carcinogens and LNT modeling for carcinogens in the absence of overwhelming evidence to the contrary.
Recently, it has been proposed that hormesis could be adopted as the default model in risk assessment based on objective criteria such as generalizability of the model, frequency in the toxicological literature, capacity to quantify false positives and negatives, relevance to endpoints of public health interest, capacity for validation, and capacity to harmonize both non-carcinogen and carcinogen risk assessment (Calabrese, 2005b).

Carcinogen risk assessment has been problematic because of the costs of compliance with estimated de minimus risks and the inability to assess the validity of such risk estimates. In the late 1970’s the U.S. government undertook an assessment of the LNT model using the carcinogen acetylaminofluorene in a “mega-mouse” study in which 24,000 animals were used. A special expert committee (seven members, four consultants to the members and one overall advisor) of the U.S. Society of Toxicology in 1981 provided a detailed assessment of the dose-time-relationships of this study (SOT, 1981). As shown in the figure (Figure 4) their evaluation demonstrated an hormetic-like, J-shaped dose-response for bladder cancer as noted in their statement that this “study provides more than evidence of a “threshold”. It provides statistically significant evidence that low doses of a carcinogen are beneficial”. A similar mega-mouse type experiment was conducted in the 1970’s in the U.S. with respect to ionizing radiation employing more than 15,000 mice in a single experiment and over 30,000 in a collective series of studies. As seen in the next figure (Figure 5) once again an horometic dose-response relationship was demonstrated for total tumors for males and females.

Hormetic dose responses have been reported for a wide range of other carcinogens and this has been reflected across initiation, promotion and progression
endpoints. While observations of an hormetic effect have often been addressed retrospectively, more recently Japanese investigators have developed prospective studies with liver carcinogens using the National Toxicology Program (NTP) standard animal model, the Fischer 344 rat. Such investigations have often demonstrated J-shaped dose responses using hepatic foci and hepatic tumors as endpoints along with substantial mechanistically oriented findings. Of particular note is that the DDT findings illustrate a high dose enhanced foci formation but a clear J-shape response for this endpoint in the low dose zone, consistent with the hormetic hypothesis (Figure 6).

While non-carcinogen risk assessment has been much less controversial, especially in comparison to carcinogen risk assessment, hormeric effects occur independently of endpoint, affecting the broad spectrum of non-carcinogen endpoints. The quantitative features of hormetic responses for carcinogens and non-carcinogens are similar; in theory, therefore, the hormetic concept could provide a toxicologically based means for harmonization of risk assessment procedures for both carcinogens and non-carcinogens.

8. Why Was Hormesis Rejected/Marginalized by the Toxicology Community?

Even though there is substantial evidence to support the hormetic dose response model, why did the toxicology community reject and/or marginalize it, as seen through the activities of major professional societies, textbook content, regulatory activities, educational programs and massive research initiatives. This is important for the field of toxicology to resolve since the central pillar of this discipline is the dose-response relationship. What factors, therefore, led the field of toxicology to make a critical error
on a “core” belief. Furthermore, how could the toxicology community continue to reinforce and propagate this error in its teaching and related regulatory activities? The historical reasons for why the hormesis hypothesis failed to thrive throughout the 20th century and to the present are complex and not entirely scientific. There are several key features that have been dominant. They include:

1). Association with Homeopathy: The hormesis concept had an early and prolonged association with homeopathy, which has been in a long conflict with what is typically termed “traditional” medicine. This association was based on how Schulz (1887, 1888) interpreted and generalized the findings concerning the effects of chemical disinfectants on yeast metabolism. In these experiments, low doses of the disinfectants stimulated metabolism in yeast while high doses were inhibitory. He interpreted the low dose stimulation as an adaptive response. Homeopathic treatments are designed to evoke symptoms at low doses that are believed to be curative or adaptive in nature; these low dose-induced symptoms are intended to be similar to those symptoms seen in diseased patients. This sequence constitutes the “like cures like” concept, that is, the Simile Principle of homeopathy. Schulz believed that the homeopathic treatment induced a specific or tailored type of adaptive capacity to the specific symptoms of the patient which would assist in resisting a pathogen infection rather than in direct killing of pathogens. Thus, for Schulz his observations of yeast responses to low doses of chemical disinfectants provided the underlying scientific foundations by which homeopathy was likely to work. Despite the efforts of Schulz and his supporters from the late 1880’s to the early decades of the 20th century, in the conflict between homeopathy and traditional medicine, homeopathy clearly was the loser, becoming greatly diminished by the early
decades of the 20th century. Another “victim” of this conflict was the concept of hormesis which became widely seen as collaterally damaged goods (e.g., guilt by association). In his highly-regarded text, Principles of Bacteriology, Hueppe in 1896 strongly affirmed that the hypothesis (i.e., hormesis) of Schulz was a reproducible phenomenon; he asserted that its close association with homeopathy should not affect its assessment by the scientific and medical communities, a suggestion that was not widely followed.

The profession of toxicology has its roots in the field of pharmacology which itself is part of the core of traditional medicine. It is not surprising, therefore, that medically oriented and pharmacologically trained toxicologists [i.e., essentially the entire first generation of U.S. toxicologists that created and grew the Society of Toxicology (SOT)] became part of the biomedical and regulatory establishments and simply maintained the practice of distancing their field and its activities from homeopathy, including the shunning of the concept of hormesis.

2). **Toxicology: A High Dose/Few Dose Discipline**: Toxicology developed as a high dose-few doses testing discipline, with a focus on finding safe levels of exposure to industrial chemicals, pesticides and drugs. So wedded to this perspective it took toxicology nearly 70 years to break its historical bondage to the LD50, which had its derivation by Trevan in 1927. Despite this liberation from the LD50 concept, toxicology still remains deeply tied to its high dose-few doses historical foundations as seen in national governmental hazard assessment research programs.

3). **Hormesis: Hard to Prove**: Hormesis is difficult to prove with adequate scientific certainty because of its modest stimulatory response that requires extraordinary allocation
of resources and time to establish. Researchers must also demonstrate that the alleged hormetic response is not simply background variation (e.g., noise) but is a reproducible phenomenon. Since toxicology has been a high dose discipline using few doses it is not hard to see how the “field” of toxicology could easily assume that modest stimulatory (i.e., hormetic) responses in below threshold zones were simply normal background variation. With this focus on high doses, hormesis became easily ignored, trivialized or directly disputed and became, for the toxicologist, the proverbial “road less traveled”.

4). Lack of Scientific Leadership: There was a lack of scientific leadership advocating an objective consideration of hormesis during the 1920’s-1940’s; this was a critical time period since this was when the field of toxicology was consolidating concepts and principles and becoming responsive to national obligations to provide guidance to agencies such as the FDA, USPHS and others. At the same time the concept of hormesis (i.e., the Arndt-Schulz Law as it was often called during the 1920’s and 1930’s) had a number of powerful, persuasive and committed scientific opponents, especially Professor Alfred J. Clark, a leading pharmacologist at the University of Edinburgh. Clark was quite prestigious, based on his work on receptor theory and quantitative aspects of pharmacology; he was also the author of the seminal text entitled “Handbook of Pharmacology” in 1937, a book that influenced the education of pharmacologists and toxicologists well into the 1970’s. Of particular importance was the fact that this major text included substantial refutation of the Arndt-Schulz Law, to which there was no published response, Schulz having died in 1932. Given Clark’s prestige and the influence of his textbook, routine educational practices hostile to the hormesis concept
became institutionalized and a central core of pharmacological/toxicological education and training from the 1930’s onward.

Beyond his direct opposition to the Arndt-Schulz Law, Clark had close professional associations with similarly prestigious and highly influential scientists and biostatisticians whose contributions shaped the course of quantitative biology including pharmacology, toxicology and related disciplines for decades. These included J.H. Gaddum, a quantitative pharmacologist also with an influential pharmacology text, amongst whose numerous scientific honors included being formally recognized for his critical research contributions to the area of chemical nerve transmission during the Nobel acceptance speech by Dale in 1936, recognition for the basic formulation applying the law of mass-action to the different types of inhibition and co-creator of probit analysis. The second colleague of Clark, C.I. Bliss, a biostatistican and co-creator of the probit analysis, prodigiously published the applications of dose response models to numerous biological disciplines; this ensured a consistent quantitative modeling of dose-response data to pharmacology, toxicology, food science, entomology, microbiology and other disciplines with a decidedly environmental focus. The integration of the work of Gaddum and Bliss within the scientific framework of Clark provided strong and continuing visible leadership to the fields of pharmacology and toxicology. With this type of arrangement there were no apparent efforts by those individuals, to model low dose responses that were consistent with the inverted U- or J-shaped dose response of hormesis, in fact, quite the contrary.

5). Statistical Modeling Procedures Precluded Possible Consideration of Hormesis: From the 1930’s onward, probit analysis became widely used across numerous biology
disciplines dealing with all or none responses such as with dosage-mortality curves. While probit analysis typically yields a straight line dose response it soon became recognized that there were numerous exceptions to this assumption at low doses. Such divergences from the straight line, which resembled a concave curve at lower concentrations, suggested a threshold concentration. To estimate a threshold dose the famous biostatistician, R.A. Fisher, developed the maximum likelihood estimation process, that is a series of successive approximations which converge to yield the most stable value above zero. In addition to the estimation of threshold doses, Bliss provided a method to estimate a probit value for both 0% and 100% responses, even though probits are values that > 0% and are < 100%.

These statistical manipulations strongly influenced modeling responses at low doses. Most significant is that they resulted in constraining low dose responses to be above that of the control value even if the responses were less than control values. These actions reinforced the belief in the sigmoidal nature of the dose response, denying even the possibility of hormetic responses. In this way the field of toxicology with its early quantitative modeling methods prevented the concept of hormesis from being considered; this also had the effect of reinforcing the belief that the hormetic effect was not a real biological phenomenon but simply background variation. In fact, this is precisely what Bryan and Shimkin (1943) did in their modeling of benzpyrene enhanced tumor incidence, despite below control values for several responses in the low dose zone. The constraining of tumor responses in the modeling process to be above control values was adopted by EPA in the late 1970’s, thereby denying the possibility of hormetic responses.
9. Integrating the Causes of Why Hormesis was Rejected and its Rebirth

Having lost the battle of textbook content, professional societal influence and governmental directives, the concept of hormesis went into an intellectual retreat for decades but still remained alive. In fact, hormetic-like biphasic effects continued to be published in a wide range of biological/toxicological journals with a bewildering array of descriptive terms such as U-shaped, J-shaped, inverted U, biphasic, hormesis, opposite effect, paradoxical, dual effect, subsidy-gradient, bi-directional, preconditioning response, ambiguous response, Yerkes-Dodson Law, Arndt-Schulz Law, Hueppe’s Rule along with others. Given this publication record over the next 40 years it is not unexpected that the concept of hormesis had a number of scientific proponents. However, the concept of hormesis never achieved significant and lasting penetration within the so-called mainstream of the scientific/toxicological communities despite numerous quality publications in respected and highly visible journals. However, the incentive to reconsider the dose-response paradigm ironically emerged as a response to EPA regulation of carcinogens which assumes linearity at low dose and whose application led to prohibitively expensive compliance/remediation programs.

The principal feature of this challenge to low dose linearity in the early 1980’s, however, was not the claim that hormesis necessarily existed and was generalizable; rather, it was that linearity at low doses was most likely an incorrect model and that most dose-responses probably acted via a threshold model. In practical terms it is not possible to readily distinguish between the “linear at low dose” model and the threshold model due to limited data that are provided in the vast proportion of toxicological studies. In fact, the stronger potential challenge to the linear at low dose model lies establishing that
the hormetic dose response exists, is broadly generalizable and relevant to all types of toxins including carcinogens. If this were established, it could provide a scientific framework to challenge the biological plausibility of linearity at low dose modeling and make a credible case for the existence of thresholds for carcinogens. Thus, a major challenge of the past 20 years has been whether there is adequate evidence in the toxicological literature to support the scientific foundations of the hormetic hypothesis. Since this search has yielded copious examples of reliable hormetic dose responses in the toxicological literature (Calabrese and Blain, 2005) how could the large number of objective and active toxicologists over the past several decades continue to have either missed or dismissed it?

The most important factor leading to this failure to recognize hormesis has been that hormesis has simply not been seen or seen often, or convincingly by most toxicologists. This is most likely due to the use of study designs geared to estimate LOAELs and NOAELs which have too few doses and with no explicit intention to assess below NO(A)EL doses. This failed testing perspective has been strongly influenced by agencies such as EPA, whose regulatory agenda has been overly fixated on short-term goals such as NOAEL derivation rather than the scientific understanding of the nature of the dose response. The control of the intellectual agenda in the field of toxicology and risk assessment by regulatory agencies, along with their capacity to dominate and direct research funding, subtly, but powerfully, prevented hormesis from being seriously considered by the research community. In addition, we have observed that a substantial proportion of rodent NOAELs display a low degree of toxicity even though the response is not statistically significantly different than control values (Calabrese and Baldwin,
2003a). The next lowest dose (i.e., dose below the NOAEL) also frequently displayed what we referred to as “residual” toxicity, although less than that seen at the NOAEL.

If one combines high dose testing, too few doses, and residual toxicity at and below the NOAEL, the field of toxicology was simply not able to assess the hormesis hypothesis. Linked to these technical issues, decades of historical bias, regulatory assumptions providing a basis for the broad acceptance of threshold and linear models, mainstream statistical models that denied a possible hormetic interpretation and a national funding program to reinforce such perspectives, it is not hard to determine why the hormesis hypothesis became and continues to be engrained in marginalization.

10. Discussion

There is now convincing evidence that hormesis is real, reproducible, and, in properly design studies, very common, in fact, more common than any other dose response model. Equally important is the fact that hormesis is known to occur in large numbers of biological models, for hundreds of different endpoints, and with many chemically diverse agents. By all reasonable measures and means of judgment, it is highly generalizable and a basic feature of life processes, a component of evolutionary biology. As such, the concept of hormesis pervades not only environmental toxicology but all biological disciplines that deal with dose-response relationships and adaptive responses. Yet, hormesis remains a paradigm lost, the victim of its own difficulty to be proven, and a history where political disputes between traditional medicine and homeopathy led to its constantly reinforced marginalization over the entire 20th century. The causes of this marginalization are historical, cultural, political and scientific and
remain deeply embedded within society. They also remain difficult to overcome despite the sustained body of evidence mounted on its behalf. Structural impediments to its success are found in the lack of widespread agreement on a definition of hormesis, no common terminology across scientific disciplines, dominating government regulatory authorities that appear more suspicious than curious of its potential scientific foundations and societal implications, constraining statistical evaluative methods that permit only above control values to be modeled, national legislation establishing environmental criteria that often excludes a consideration of hormesis, risk communication strategy procedures that deny the possibility that it exists and government-mediated hazard assessments that are too limited to address its existence. Yet, despite what appears to be continuing overwhelming challenges and highly stacked odds against its recognition and acceptance, hormesis has made substantial gains in scientific interest and acceptability over the past decade based on its presence in recent editions of leading textbooks, articles published in peer-reviewed papers, sessions at major national professional societal conferences, invited seminars at leading institutions, incorporation into academic education programs, governmental funding, and the decision of leading national and international advisory groups to formally include it into their investigative agenda and numerous articles written in the popular press.

Despite this progress, most of the articles that publish bona fide evidence supporting hormesis do not use the term, often employing general terms such as U-shaped, biphasic and stimulatory responses. This lack of the term hormesis may be related to the fact that investigators who see hormetic-like biphasic responses have not yet become sufficiently acquainted with the hormesis concept and its literature or are
resistant to use it for historical reasons cited here. There has also been a lack of recognition by the general biomedical community that the widespread occurrence of biphasic dose response relationships of similar quantitative features may be examples of hormesis. Despite the many thousands of such examples in the biomedical literature, the biomedical community has been strikingly silent on whether this common dose-response phenomenon reflects a general biological principle. This lack of integrative thinking is indicative of the academic isolation of modern science where the important achievements of specialization are rewarded to the frequent exclusion of integration.

If widely accepted, the hormetic concept has the potential to affect significant changes in the biomedical sciences as toxicological training and education, hazard assessment as reflected in animal model selection, endpoints measured, study design, type of statistical analyses used, and risk assessment practices.

11. Conclusions

Perspective #1

The threshold and linear dose response models fail to make accurate predictions in the below threshold zone.

Perspective #2

The threshold and linear dose response models have been significantly out-competed by the hormetic dose response model in multiple, independent comparisons.

Perspective #3

There is little toxicological justification for the continued use of the threshold and linear dose response models to estimate below threshold responses or in the low dose zone.
Perspective #4

Give Perspectives 1-3, there is no scientific basis to use the threshold or linear dose response models in risk assessment practices. This has significant implications for current standards based on the threshold and linear models and future risk assessment practices.

Perspective #5

HORMESIS: A concept with much supportive experimental evidence that is reproducible.

Perspective #6

HORMESIS: Based on Perspective #5 it should be considered as a real concept in the biological sciences.

Perspective #7

HORMESIS IS GENERALIZABLE
- Across biological models
- Across endpoints measured
- Across chemical class/physical agents

Perspective #8

Based on Perspective #7, HORMESIS is evolutionarily based, with broad potential implications.

Perspective #9

HORMESIS: Very common in toxicological/pharmacological literature, making it a central concept.

Perspective #10

HORMESIS: A normal component of the traditional dose response, being graphically contiguous with the NO(A)EL.
Perspective #11

HORMESIS: Readily definable quantitative features, that are broadly generalizable, making it reasonably predictable.

Perspective #12

HORMESIS: Far more common than the threshold and linear dose response models in fair head-to-head comparisons; this would make the hormetic model the most dominant in toxicology.

Perspective #13

HORMESIS: No single specific hormetric mechanism; there appears to be a common biological strategy underlying such phenomena.

Perspective #14

HORMESIS: Important implications for toxicology, risk assessment, risk communication, cost-benefit assessments, clinical medicine, drug development and numerous other areas.

Perspective #15

HORMESIS: Should become the object of formal evaluation by leading advisory bodies such as the National Academy of Sciences.
12. References


Figure 1. Schematic hormetic dose responses (i.e., inverted U- and J-shaped)
Figure 2. Overcompensation stimulation (hormesis) within a dose-time-response relationship. Response (R) on the vertical axis, dose (D) on the horizontal axis.

At Time 1 there is a dose dependent decrease, consistent with a toxic Response.

At Time 2 there is the start of a compensatory response, as is evident by the low stimulatory response within the low dose range.

At Time 3 the compensatory response achieves its maximum increase over controls at the low dose. At higher doses a complete compensatory response is not achieved. As time progresses the low dose stimulatory response may expected to return to the control value.
Figure 3. Selected examples of hormetic dose responses published in the peer-reviewed literature

A) Source: Jefferson and Aguirre, 1980

![Graph showing hormetic dose response of methanol on fruit fly longevity.](image)

B) Source: Lee et al., 1985

![Graph showing effect of sodium arsenate on Syrian hamster cell line LSH/ss LAK survival.](image)
Figure 3. Continued

C) Source: Marushige and Marushige, 1998

![Graph C: Effect of pervanadate on the cell rounding in trigerminal neurinoma cells](image1)

D) Source: Mead and Pentreath, 1998

![Graph D: Effect on proliferation of 1321N1 cells, a human astrocytoma cell line, using the neutral red assay](image2)
Figure 3. Continued

E) Source: Parkhurst et al., 1981

F) Source: Parmelee et al., 1993
Figure 3. Continued

G) Source: Sandifer and Hopkin, 1997

![Graph showing the effect of lead and copper on springtail survival.]

H) Source: Varlinskaya et al., 2001

![Graph showing the effects of acute ethanol on overall social activity of adolescent rats.]

Effects of acute ethanol on overall social activity of adolescent rats tested on postnatal day 30.
Figure 3. Continued

I) Source: Vieira et al., 2000

J) Source: Cicero and Badger, 1977
Figure 4. Bladder tumor incidence adjusted for time in ED01 megamouse study (SOT, 1981).
Figure 5. Radiation megamouse tumor study based on data from Ullrich, and Storer, 1979; Ullrich et al., 1976.
Figure 6. Effect of DDT on number of GST P-positive foci in F344 rat livers. Note: as the dose decreases the J-shaped dose-response becomes evident (Source: Sukata et al., 2002).
Table 1. A partial listing of receptor systems displaying biphasic dose-response relationships.

<table>
<thead>
<tr>
<th>Receptor systems displaying biphasic dose-response relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenosine</td>
</tr>
<tr>
<td>Adrenoceptor</td>
</tr>
<tr>
<td>Bradykinin</td>
</tr>
<tr>
<td>CCK</td>
</tr>
<tr>
<td>Corticosterone</td>
</tr>
<tr>
<td>Dopamine</td>
</tr>
<tr>
<td>Endothelin</td>
</tr>
<tr>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>Estrogen</td>
</tr>
<tr>
<td>5-HT</td>
</tr>
<tr>
<td>Human chorionic gonadotrophin</td>
</tr>
<tr>
<td>Muscarinic</td>
</tr>
</tbody>
</table>