TOXICOLOGICAL SCIENCES **62,** 330–338 (2001)

The Frequency of U-Shaped Dose Responses in the Toxicological Literature

Edward J. Calabrese[1] and Linda A. Baldwin

Department of Environmental Health Sciences, School of Public Health and Health Sciences, University of Massachusetts, Amherst, Massachusetts 01003

Received November 17, 2000; accepted April 17, 2001

Hormesis has been defined as a dose-response relationship in which there is a stimulatory response at low doses, but an inhibitory response at high doses, resulting in a U- or inverted U-shaped dose response. To assess the proportion of studies satisfying criteria for evidence of hormesis, a database was created from published toxicological literature using rigorous *a priori* entry and evaluative criteria. One percent (195 out of 20,285) of the published articles contained 668 dose-response relationships that met the entry criteria. Subsequent application of evaluative criteria revealed that 245 (37% of 668) dose-response relationships from 86 articles (0.4% of 20,285) satisfied requirements for evidence of hormesis. Quantitative evaluation of false-positive and false-negative responses indicated that the data were not very susceptible to such influences. A complementary analysis of all dose responses assessed by hypothesis testing or distributional analyses, where the units of comparison were treatment doses below the NOAEL, revealed that of 1089 doses below the NOAEL, 213 (19.5%) satisfied statistical significance or distributional data evaluative criteria for hormesis, 869 (80%) did not differ from the control, and 7 (0.6%) displayed evidence of false-positive values. The 32.5-fold (19.5% vs 0.6%) greater occurrence of hormetic responses than a response of similar magnitude in the opposite (negative) direction strongly supports the nonrandom nature of hormetic responses. This study, which provides the first documentation of a data-derived frequency of hormetic responses in the toxicologically oriented literature, indicates that when the study design satisfies *a priori* criteria (i.e., a well-defined NOAEL, ≥ 2 doses below the NOAEL, and the end point measured has the capacity to display either stimulatory or inhibitory responses), hormesis is frequently encountered and is broadly represented according to agent, model, and end point. These findings have broad-based implications for study design, risk assessment methods, and the establishment of optimal drug doses and suggest important evolutionarily adaptive strategies for dose-response relationships.

Key Words: **hormesis; compensatory responses; overcompensation; U-shaped; J-shaped; dose response; low doses; risk assessment; extrapolation.**

The occurrence of hormesis in the toxicological sciences has a long and controversial history (Calabrese and Baldwin, 2000a,b,c,d,e). Evidence supporting the existence of hormesis is substantial, with numerous reproducible examples suggesting potential broad generalizability (Calabrese *et al*., 1999). However, little information exists concerning the frequency of hormesis within the toxicological literature; that is, how often one would expect to observe hormesis given appropriate study design parameters. Two databases were previously created from the published literature to quantify aspects of hormetic responses in toxicological studies. In the case of Davis and Svendsgaard (1994), an attempt was made to estimate the incidence of hormetic responses based on the frequency of deviation from control responses independent of study design, NOAEL (no observed adverse effect level), and statistical significance. The second database (Calabrese and Baldwin, 1997a,b) focused on describing the quantitative features of the hormetic dose response and issues relating to generalizability rather than frequency in the toxicological literature.

Taking into consideration the limitations of the previous databases and incorporating suggestions by Crump (2001), a new database was created to assess the proportion of studies in the toxicological literature satisfying criteria for evidence of hormesis consistent with the definition of Stebbing (1998). Rigorous *a priori* entry criteria were established based on study design characteristics to identify data sets with the potential to detect a hormetic effect. Data sets meeting these criteria, independent of outcome, were entered into the database. Subsequent application of *a priori* evaluative criteria identified those dose-response relationships satisfying requirements for evidence of hormesis.

METHODS

Journal selection. Because a broad range of experimental models, end points, and agents, including mixtures, was desired, two environmentally-oriented toxicological journals (*Environmental Pollution,* 1970–1998; *The Bulletin of Environmental Contamination and Toxicology,* 1966–1998) and one pharmacologically oriented toxicological journal (*Life Sciences*, 1962–1998) were selected. Use of these journals ensured broad coverage of the toxicological literature without truncated end-point selection associated with more specialized journals. This was viewed as a desirable and necessary journal selection strategy at this stage of project development, as it would offer

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

[1] To whom correspondence should be addressed. Fax: (413) 545-4692. E-mail: edwardc@schoolph.umass.edu.

greater opportunity to address issues of generalizability. Furthermore, inclusion of approximately 30 years of articles from each journal ensured the opportunity to incorporate independent peer review over prolonged periods, studies reflecting changes in toxicological funding priorities (thereby enhancing the range of chemicals, end points, and hypotheses assessed), improvements in study design, analysis, and technical developments as the field evolved, and assessment of historical trends if needed.

Screening protocol. All articles were initially screened in ascending chronological order beginning with volume 1, number 1 of each journal through 1998, with the exception of *Life Sciences*. Due to the increasingly large number of articles published per year in this journal (by the end of 1979 approximately 6000 articles had been screened with an annual publication rate increasing to over 600 articles), a decision was made to limit additional screening to 6 years, approximately equally spaced over the remaining 19 years of publication (1982, 1985, 1988, 1992, 1995, 1998). During the initial screening, exclusion and entry criteria described below were applied to all dose-response relationships reported in tabular or graphical form in each article. Dose-response relationships meeting the entry criteria were later examined with evaluative criteria described below for satisfying or not satisfying evidence of hormesis. The initial screening and the subsequent application of evaluative criteria were performed by the two authors; the results of the application of evaluative criteria were examined a second time by one of the authors.

Exclusion criteria. Only studies with experimental data were considered. Review articles, abstracts, non-English language articles, epidemiologic studies, and field studies were excluded. Studies lacking any of the following conditions were excluded: (1) a concurrent control; (2) the capacity to achieve responses greater than (or less than, depending on end point) the control response (e.g., studies where the end point was survival and the control response was 100% or where the end point was tumor incidence and the control response was zero); (3) at least two doses below the NOAEL; and (4) at least one dose showing *a priori* criteria-based inhibition.[2]

NOAEL designation. The NOAEL designation represents a unique dose that can be satisfied by only one dose. In the hormesis database this dose is satisfied by definitional determinants such that this dose represents the highest dose not differing from the control and having defined decrements at immediately higher doses. Any dose lower than this designated NOAEL that displays a response below that of the control would be interpreted as displaying either variability or error. As a result of this definition of NOAEL and applying it consistently throughout the database, possible subjective reinterpretation and designation of the NOAEL dose was prevented. The implications of this scheme were to allow for the inclusion of negative variability/error in the dose-response relationship below a designated NOAEL to permit false-positive estimation. If this approach had not been followed, some dose-response relationships could have been eliminated from satisfying entry criteria, ultimately resulting in a higher proportion of studies satisfying the evaluative criteria.

Residual bias may occur as a result of the NOAEL designation used in this assessment. Some doses that are characterized as NOAELs may in fact display evidence of low/modest toxic responses. However, if the decrement does not achieve a certain designated level (e.g., statistical significance, percent decrement), a determination could be made for that dose being the NOAEL. Thus, it is possible to inappropriately designate a bona fide LOAEL (lowest observed adverse effect level) as a NOAEL. This concern is widely recognized in regulatory toxicology and is one of the reasons why the NOAEL has been broadly criticized with respect to its no adverse effect designation. This possible limitation has led to proposals for application of statistical procedures, such as the benchmark dose (BMD), to estimate the NOAEL. If a NOAEL is actually a LOAEL in the current hormesis database, this would have implications for detection of hormesis at lower doses in the dose response spectrum. In fact, it could limit the potential detection to possibly one dose under certain study design scenarios. Again, even this one dose may still actually represent a type of LOAEL, if in fact it too had low residual deficits. This suggests that for dose responses in the present hormesis database where the NOAEL reflects a dose with a slight/modest toxic response, a false-negative potential for hormesis estimation may exist.

A decision was made in the development of the criteria to include as NOAELs for evaluative purposes doses that could satisfy evaluative criteria for evidence of hormesis. Although it is possible that one could have eliminated NOAELs within an evaluative designation, this approach was rejected, since the NOAEL, when it exceeds the control value, could be considered as being in the hormetic zone. This is because the designation of the NOAEL is not a perfect representation of the zero equivalent point (i.e., the highest dose with a response equal to the control response), but could err on either side of the control for real biological effect purposes. For this reason, it was decided that it would be unfair to bias a determination against a hormetic perspective. It should be noted that it was argued above that mischaracterization of a LOAEL with a NOAEL could lead to false-negative representation. However, allowing a NOAEL to be positively identified as a hormetic response is not a misrepresentation.

Entry criteria. The entry criteria were designed to ensure consistency with the U (or inverted U) shape of the hormetic dose-response relationship. That is, all studies needed to have sufficient evidence to demonstrate the occurrence of high-dose inhibition based on statistical and/or quantitative criteria, a NOAEL, and doses below the NOAEL that were to be evaluated for the potential of a low-dose stimulatory response based on statistical and/or quantitative criteria. Studies satisfying these general criteria were placed into one of three entry criteria tiers (T1, T2, T3) presented in Table 1: T1 includes dose-response relationships subjected to hypothesis testing; T2 was designed to identify dose-response relationships lacking hypothesis testing but reporting standard deviation (SD) or standard error of the mean (SEM) information, thereby providing information on the distribution of the data. T3 was designed to identify dose-response relationships defined only by data points reflecting mean/median values with no reference to variation.

Evaluative criteria. All dose-response relationships meeting the entry criteria were then subjected to evaluative criteria for evidence of hormesis (Table 1). An outcome satisfying criteria for evidence of hormesis is considered indicative of a dose-response relationship demonstrating stimulation at low doses and inhibition at higher doses. (See Fig. 1 for examples of data sets satisfying evaluative criteria.)

Where no hypothesis testing was performed, a difference of at least two SD or two SEM between the control and treatment group was considered indicative of potential statistical significance. Although our intent was to standardize all such data to conform to a similar distribution (i.e., SEM), this was not possible because of considerable variability in the nature and specificity of the information provided (e.g., out of a total of 196 dose-response relationships in category T2, 66 distributions were reported as SEM, 60 were reported as SD, and 70 were not identified). Consequently, we used the distribution provided in

[2] For purposes of this study, the NOAEL was defined as the highest dose with a response not statistically significantly different with respect to adverse responses from the control in studies where hypothesis testing was performed; in studies lacking hypothesis testing and in studies where hypothesis testing was performed but statistical significance was not observed with respect to adverse effects, the NOAEL was defined as the highest dose with a response $\geq$ 90% of the control for inverted U-shaped dose-response relationships or as the highest dose with a response $\leq$ 110% of the control for U- or J-shaped dose-response relationships. Inhibition was defined as occurring when: (1) the response for at least one dose higher than the NOAEL was statistically significantly different from the control in studies where hypothesis testing was performed; (2) the response for at least one dose higher than the NOAEL showed no 2$\times$ SD/SEM overlap with the control response in studies where only data distribution was reported; or (3) in the absence of statistical significance or nonoverlapping distributions, the response for at least two doses higher than the NOAEL was $<$ 90% of the control for inverted-U shaped dose-response relationships or $>$ 110% of the control for U- or J-shaped dose-response relationships.

TABLE 1
Summary of *a Priori* Entry and Evaluative Criteria with Descriptions of Outcomes Satisfying Evidence of Hormesis

Category	T1	T2	T3
Entry criteria			
Hypothesis testing	Yes	No	No
Data distribution reported	Not relevant	Yes	No
Minimum no. doses below NOAEL[a]	2	2	2
Minimum no. doses above NOAEL[b]	1 dose with a statistically significant response or 2 doses with responses $<$ 90% of control	1 dose with a response showing no 2 $\times$ SD/SEM overlap with control or 2 doses with responses $<$ 90% of control	2 doses with responses $<$ 90% of control
Evaluative criteria[c]			
Outcome satisfying evidence of hormesis[d]			
Including and/or below the NOAEL responses	At least 1 stimulatory dose with a statistically significant response or at least 3 doses with responses $\geq$ 110% of control	At least 1 stimulatory dose with a response showing no 2 $\times$ SD/SEM overlap with control or at least 3 doses with responses $\geq$ 110% of control	At least 3 doses with responses $\geq$ 110% of control

Note. T1, dose-response relationships subjected to hypothesis testing; T2, designed to identify dose-response relationships lacking hypothesis testing but reporting SD or SEM, thus providing data distribution information; T3, designed to identify dose-response relationships defined only by data points reflecting mean/median values with no reference to variation.

[a]NOAEL, no observed adverse effect level. For the purposes of this study the NOAEL was defined as the highest dose with a response not statistically significantly different from the control with respect to adverse effects in studies where hypothesis testing was performed; in studies lacking hypothesis testing and in studies where hypothesis testing was performed but statistical significance was not observed with respect to adverse effects, the NOAEL was defined as the highest dose with a response $\geq$ 90% of the control for inverted U-shaped dose-response relationships or as the highest dose with a response $\leq$ 110% of the control for U- or J-shaped dose-response relationships.

[b]For the purposes of this study inhibition must be demonstrated as follows: at least one dose higher than the NOAEL with a response statistically significantly different from the control in studies where hypothesis testing was performed; at least one dose higher than the NOAEL with a response showing no 2 times SD/SEM overlap with the control in studies where data distribution was reported; or at least 2 doses higher than the NOAEL with responses $<$ 90% of the control.

[c]Please note that these descriptions apply to inverted U-shaped dose-response relationships; in the case of J- (or U-) shaped dose-response relationships the evaluative response criterion value including and/or below the NOAEL is $\leq$ 90% of control and the evaluative response criterion above the NOAEL is $>$ 110% of control.

[d]Consistent with the U- (or inverted U-) shape of the hormetic dose-response relationship.

the paper, recognizing that the comparison between studies would lack the intended uniform comparability.

In cases where data were graphically represented, on some occasions error bars were depicted for treatment data points, but not for the control. In those cases the dose responses were considered indicative of potential statistical significance if the error bars (SD/SEM $\times$ 2) of the treatment did not cross the control value.

In order to avoid exclusion of potentially relevant data below the NOAEL and to enhance the rigor of evaluative criteria, dose-response relationships with at least three doses with responses $\geq$ 110% of control (or with responses $\leq$ 90% for J- or U-shaped curves), i.e., alternative quantitative criteria were considered satisfying evidence of low-dose stimulation in the absence of statistical significance or potential statistical significance as determined by data distribution.

In order to avoid exclusion of potentially relevant data due to absence of a statistically significant or potentially statistically significant inhibitory response at high doses, dose-response relationships with at least two doses with responses $<$ 90% of control (or $>$ 110% for J- or U-shaped curves) were considered satisfying evidence of inhibition in the absence of statistical significance or potential statistical significance as determined by data distribution.

Assessment of false-positive responses. An indication of the frequency of false-positive responses (i.e., to what extent the positive findings could be accounted for by chance or random variation) was obtained by assessing the responses of treatment doses below the NOAEL and comparing the proportion of negative findings to positive findings. This is based on the assumption that if chance or random variation was responsible for the positive findings (i.e., a hormesis designation) then the number of negative responses should approximate the number of positive responses. It should be noted that although the NOAEL dose was included when assessing dose-response relationships with the evaluative criteria (Table 1), only treatment doses below the NOAEL were evaluated for false-positive responses. The NOAEL by definition cannot display an adverse (or negative) response, and its inclusion in the assessment of false-positives would therefore bias the outcome. By excluding the NOAEL values in this assessment, bias favoring false-positive estimation was minimized. This approach therefore provides a rate of false-positive/negative estimates that could be applied to the total rather than deriving the absolute number by direct estimation.

When the evaluative criteria were based on the response of single doses, the proportion of false-positive findings was derived by dividing the total number of doses below the NOAEL showing significant or potentially significant negative responses by the total number of significant or potentially significant responses of a positive and negative nature for both the hypothesis testing (T1) and distributional data (T2) categories. A similar procedure was employed to estimate false-positive findings when alternative quantitative criteria were used.

Assessment of false-negative responses. An indication of the frequency of false-negative responses was obtained by assessing the proportion of dose-response relationships satisfying the alternative quantitative evaluative criteria to the total number of dose-response relationships not satisfying evaluative criteria in the hypothesis testing category T1. This procedure was also applied

Entry criteria satisfied:

- **Concurrent control**
- **NOAEL established (see Exclusion Criteria description)**
- **Bona fide inhibition (see Exclusion Criteria description)**
- **At least 2 doses below the NOAEL**

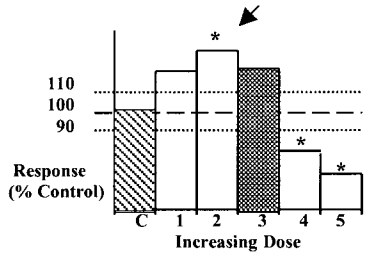
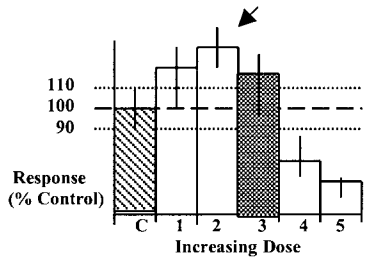
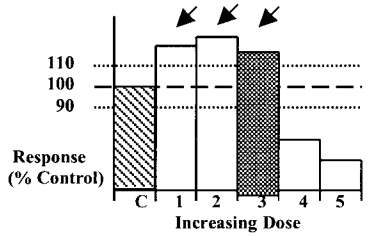
Evaluative criteria And category	Evidence of hormesis
Statistical Significance **(at least 1 stimulatory response at and/or below NOAEL that is statistically significant)** **T1**	
Data Distribution **(no 2 X SD/SEM overlap of at least 1 treatment response at and/or below NOAEL with control response)** **T2**	
Alternative Quantitative **(at least 3 doses at and/or below NOAEL with responses ≥ 110% of control response)** **T1, T2, T3**	

FIG. 1. Hypothetical data sets comprising a control (C, hatched lines) and five treatment groups (doses) satisfying entry criteria; data sets showing evidence of hormesis for each evaluative approach (statistical significance, data distribution, and alternative quantitative) and category. See Table 1 for category descriptions. NOAEL dose is indicated by cross-hatching; arrows indicate response(s) satisfying evaluative criteria for hormesis; *statistical significance at $p \leq 0.05$. Error bars represent the mean ± 2 SD. Please note that these descriptions apply to inverted U-shaped dose-response relationships; in the case of J- (or U-) shaped dose-response relationships, the evaluative response criterion value including and/or below the NOAEL is ≤ 90% of control.

to dose-response relationships that failed to satisfy evaluative criteria for the distributional data category T2.

RESULTS

Frequency of Hormetic Effects

Table 2 presents the results of application of the entry criteria organized by journal and year of publication. Of the 20,285 articles screened, 195 articles (1%) contained 668 dose-response relationships meeting the entry criteria. The number of articles screened was equally divided between the environmentally oriented journals (51.5%; 10,462 articles published in *Environmental Pollution* and *The Bulletin of Environmental Contamination and Toxicology*) and the more pharmacologically oriented journal (48.4%; 9823 articles published in *Life Sciences*). Approximately 1% of the articles in each journal

TABLE 2
Summary of Results of Application of *a Priori* Entry Criteria for Articles, Organized by Journal and Publication Year

Environmental Pollution				*Bulletin of Environmental Contamination and Toxicology*				*Life Sciences*			
Year	Published	Entered	No. d-r	Year	Published	Entered	No. d-r	Year	Published	Entered	No. d-r
								1962	127	0	0
								1963	162	0	0
								1964	217	0	0
								1965	333	1	1
				1966	38	1	4	1966	299	1	1
				1967	46	0	0	1967	350	0	0
				1968	45	0	0	1968	349	0	0
				1969	45	0	0	1969	344	1	1
1970	13	0	0	1970	103	0	0	1970	170	2	3
1971	23	0	0	1971	100	1	4	1971	168	1	2
1972	28	0	0	1972	130	0	0	1972	131	2	4
1973	52	2	17	1973	136	1	2	1973	247	1	1
1974	54	2	6	1974	234	1	1	1974	444	6	7
1975	58	0	0	1975	259	4	9	1975	459	0	0
1976	53	2	6	1976	246	3	11	1976	438	2	4
1977	88	0	0	1977	235	0	0	1977	494	5	8
1978	80	1	1	1978	247	2	10	1978	643	9	26
1979	83	0	0	1979	428	3	5	1979	593	4	9
1980	83	0	0	1980	312	1	3				
1981	86	0	0	1981	270	3	11				
1982	83	2	9	1982	248	2	6	1982	741	5	9
1983	65	2	3	1983	213	0	0				
1984	109	0	0	1984	210	1	2				
1985	107	3	50	1985	254	2	11	1985	623	11	27
1986	51	3	14	1986	283	1	2				
1987	138	2	66	1987	322	0	0				
1988	180	4	11	1988	260	3	5	1988	610	8	13
1989	159	2	3	1989	271	6	12				
1990	161	4	25	1990	276	3	17				
1991	115	2	8	1991	283	1	1				
1992	133	0	0	1992	277	5	27	1992	596	13	33
1993	150	1	2	1993	271	0	0				
1994	152	1	11	1994	277	3	7				
1995	168	0	0	1995	267	6	26	1995	658	12	20
1996	162	1	4	1996	289	4	14				
1997	127	1	1	1997	279	4	14				
1998	297	3	14	1998	250	1	2	1998	627	11	42
Total	3058	38	251	Total	7404	62	206	Total	9823	95	211

Note. Environmental Pollution was divided into two series during the years 1980–1986 (Series A, Ecological and Biological; Series B, Chemical and Physical). Series A was selected for screening. *Life Sciences* was divided into two parts during the years 1970–1973 (Part 1, Physiology and Pharmacology; Part 2, Biochemistry, General, and Molecular Biology). Part 1 was selected for screening. Due to the increasingly large number of articles published per year in this journal additional screening was limited to 6 years (1982, 1985, 1988, 1992, 1995, 1998); d-r, dose-response relationships meeting *a priori* entry criteria.

contained dose-response relationships meeting the entry criteria (*Environmental Pollution*, 1.2% $^{38}/_{3058}$; *The Bulletin of Environmental Contamination and Toxicology*, 0.8% $^{62}/_{7404}$; *Life Sciences*, 1.0% $^{95}/_{9823}$). The number of dose-response relationships meeting the entry criteria was approximately equally divided among the three journals (*Environmental Pollution*, 37.5% $^{251}/_{668}$; *The Bulletin of Environmental Contamination and Toxicology*, 30.8% $^{206}/_{668}$; *Life Sciences*, 31.5% $^{211}/_{668}$).

Figure 2 presents the results of application of the evaluative criteria to the 668 dose-response relationships satisfying the entry criteria organized by category. Two hundred forty-five (245) dose-response relationships (36.7% of 668) from 86 articles (0.4% of 20,285) satisfied the requirements for evidence of hormesis. Eighteen articles containing 118 dose-response relationships were in *Environmental Pollution*; 28 articles containing 68 dose-response relationships were in *The Bulletin of Environmental Contamination and Toxicology*; and 40 articles containing 59 dose-response relationships were in *Life Sciences*.

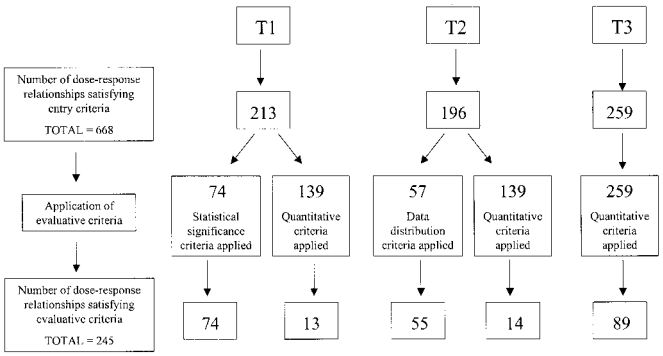


FIG. 2. Summary of application of the evaluative criteria to the 668 dose-response relationships satisfying the entry criteria, organized by category. See Table 1 for category descriptions.

Table 3 presents the results of the 245 dose-response relationships satisfying evaluative criteria for evidence of hormesis organized by experimental model, end point, and agent, including mixtures. A total of 73 different agents and mixtures from a broad range of chemical classes is represented.

Assessment of False-Positive Responses

Tables 4 and 5 present the results of the assessment of false-positive responses for each entry/evaluative criteria category. The collective findings indicate that the false-positive rate from the various categories was approximately 4%. When the false-positive/negative values were totaled, it yielded a net 1.4% false-positive estimate. This would reduce the 36.7% hormesis frequency to 35.3% (Table 6). These findings indicate that the methodology was not very susceptible to false-positive/negative error.

Assessment of False-Negative Responses

There were 139 dose-response relationships in T1 that satisfied entry but failed to satisfy evaluative criteria for hormesis. Thirteen of these 139 dose-response relationships satisfied the alternative quantitative evaluative criteria. This value provides an estimate of the false-negative rate for hormesis of 9.3% within the hypothesis testing criteria (category T1, Fig. 2). A similar procedure applied to the distributional data revealed a false-negative rate of 10.1% (category T2, Fig. 2). A direct comparison of dose-response relationships satisfying evaluative criteria for both hypothesis testing and alternative quantitative criteria revealed that such dose responses were approximately twice as likely to satisfy the evaluative criteria for hypothesis testing than for alternative criteria (i.e., of the 75 dose-response relationships satisfying hypothesis testing criteria, 38 also satisfied alternative criteria). That is, it is twice as difficult to have three doses at and/or below the NOAEL with responses $\geq$ 110% of the control response as to have one of these doses with a responses statistically significantly greater than the control. These findings not only strongly support the use of the methodology to estimate false-negative rates, but also indicate that the actual false-negative rates are likely to be higher than estimated. These results also suggest that the findings provided in the alternative criteria for hormetic estimates are considerably more rigorous than the hypothesis testing and distributional methods.

DISCUSSION

The findings indicate that in studies satisfying entry criteria, 36.7% satisfied the evaluative criteria for a hormetic response. Although the above assessment indicates that the study findings cannot be accounted for by false-positive responses or by

TABLE 3

Summary of the Dose-Response Relationships Satisfying *a Priori* Entry and Evaluative Criteria for Evidence of Hormesis Organized by Experimental Model, Endpoint, and Agent, Including Mixtures

	No. of dose-response relationships		
	Satisfying entry criteria	Satisfying evaluative criteria	%
Experimental model			
Plant	309	138	45 (138/309)
Vertebrate	266	83	31 (83/266)
Invertebrate	48	9	19 (9/48)
Microbe	43	15	35 (15/43)
Protozoan	2	0	0 (0/2)
Total	668	245	
Endpoint analyzed			
Metabolic	231	97	42 (97/231)
Growth	183	72	39 (72/183)
Reproductive	124	43	35 (43/124)
Molecular	64	11	17 (11/64)
Behavioral	36	17	47 (17/36)
Physiologic	17	3	18 (3/17)
Survival	13	2	15 (2/13)
Total	668	245	
Agent and Mixtures			
Effluents	116	67	58 (67/116)
Pesticides	114	30	26 (30/114)
Metals	85	30	35 (30/85)
Petroleum products/ constituents	24	15	62 (15/24)
Alcohol production wastes	19	16	84 (16/19)
Polycholorinated biphenyls	17	11	65 (11/17)
Solvents	10	6	60 (6/10)
Miscellaneous	283	70	25 (70/283)
Total	668	245	

Note. % = no. satisfying evaluative criteria/ no. satisfying entry criteria. Metabolic endpoints include enzyme activities, photosynthesis rate, respiration rate, protein synthesis, etc. Physiologic endpoints include muscle contraction/ relaxation, blood pressure, heart rate, etc. Pesticides include insecticides, herbicides, fungicides and ectoparasiticides. Miscellaneous includes a variety of agents and mixtures (e.g., pharmaceutical products, receptor agonists and antagonists, detergents, etc.).

TABLE 4
Assessment of False Positive Responses in Cases Where *a Priori* Evaluative Criteria Were Based on the Response of Single Doses below the NOAEL

Criteria			Number treatment doses below NOAEL			
Entry	Evaluative	Category	Total	Positive hormesis evidence	Chance positive evidence	False positive rate
Hypothesis testing	Statistical significance	T1	551	129	1	1/130 = 0.008 (0.8%)
Data distribution	Data distribution	T2	538	84	6	6/90 = 0.067 (6.7%)
			1089	213	7	7/220 = 0.032 (3.2%)

Note. Positive hormesis evidence, treatment doses below NOAEL with statistically significant or potentially significant stimulatory responses. Chance positive evidence, total number of treatment doses below NOAEL with statistically significant or potentially significant inhibitory responses. False positive rate, false positive/positive + false positive.

random variation, there are fundamental limitations in the current study methodology that are likely to yield a tendency for false-negative conclusions (values lower than actual hormesis estimates). The false-negative rate was nearly three times greater than the false-positive rate (i.e., 9.7% vs 3.5%). The false-negative criteria were established as being twice as rigorous as the false-positive estimation procedure. Finally, while false-positive evaluation was able to be applied to all possible instances for positive responses, this was not the case for the 170 negative dose responses in the alternative quantitative criteria for which no validation procedure is available. Given these three factors, it is likely that the 36.7% estimate of hormetic dose-response frequency is conservative and is likely somewhat higher.

In addition, the study did not take temporal factors into consideration. Numerous investigations exist that demonstrate stimulatory responses occur only following a disruption in homeostasis, that is, after an initial decrement in response (Stebbing, 1998; Calabrese, 2001). If responses were not taken at multiple times during the experiment, possible stimulatory responses could be missed, leading to false-negative conclusions.

It is interesting to note that of 1089 treatment doses below the NOAEL using hypothesis testing and distributional data entry criteria (Table 4), 213 (19.5%) of the treatment doses were determined to satisfy hormesis evaluative criteria. Only seven treatment doses (7/1089 = 0.6%) were significantly below the control. This suggests that hormetic responses in these categories occurred approximately 30-fold (19.5%/0.6% = 32.5) more frequently than a response of similar magnitude in the opposite (negative) direction. This finding, which employs the treatment doses below the NOAEL as the unit of comparison, provides striking support for the position that hormetic effects cannot be attributed to chance.

The data further revealed that the general occurrence for hormetic dose responses was widely incorporating of biological model, end points, and chemical classes. These findings represent the first attempt to assess the frequency of hormetic responses within the context of a biological/toxicological model based on study design, dose response, and statistical features. The results are particularly noteworthy, as they directly challenge the long-held view that hormetic responses should be seen as statistical exceptions, paradoxical findings, or otherwise unexpected events.

TABLE 5
Assessment of False Positive Responses in Cases Where *a Priori* Evaluative Criteria Were Based on the Response of Multiple Doses below the NOAEL

Criteria			Number dose responses with ≥ 3 doses below NOAEL			
Entry	Evaluative	Category	Total	Positive hormesis evidence	Chance positive evidence	False positive rate
Hypothesis testing	Alternative quantitative	T1	49	7	2	2/9 = 0.222 (22.2%)
Data distribution	Alternative quantitative	T2	68	10	1	1/11 = 0.09 (9%)
Alternative quantitative	Alternative quantitative	T3	104	40	0	0 (0%)
			221	75	3	3/78 = 0.038 (3.8%)

Note. The values of 110% and 90% refer to inverted U-shaped dose-response relationships; the values would be reversed in the case of U- (or J-) shaped dose-response relationships. Positive hormesis evidence, total number of dose-responses with at least 3 doses below the NOAEL with a response ≥ 110%. Chance positive evidence, total number of dose-responses with at least 3 doses below the NOAEL with a response ≤ 90%. False positive rate, false positive/positive + false positive.

TABLE 6
Adjustment for Potential False Positive Responses of the Number of Dose-Response Relationships Satisfying Evidence of Hormesis

Criteria						
Entry	Evaluative	Categories	Unadjusted positives/total	Est. false positive potential	No. false positive	Adjusted positives/total
Hypothesis testing	Statistical significance	T1	74/213	0.8%	1 (0.8%×74)	73 (74–1)
Hypothesis testing	Alternative quantitative	T1	13/139[a]	22.2%	3 (22.2%×13)	10 (13–3)
Data distribution	Data distribution	T2	55/196	6.7%	4 (6.7%×55)	51 (55–4)
Data distribution	Alternative quantitative	T2	14/139[a]	9.0%	1 (9.0%×14)	13 (14–1)
Alternative quantitative	Alternative quantitative	T3	89/259	0%	0	89 (89–0)
			245/668 =36.7%			236/668 = 35.3%

Note. The unadjusted positives/total value (i.e., 245/668 = 36.7%) includes correction for potential false negatives (i.e., positive responses with alternative quantitative criteria in categories T1 and T2).

[a]This is based on the assumption that these categories adjust for false negative responses by employing alternative quantitative evaluative criteria in cases where the data do not satisfy statistical significance (T1) or potential statistical significance as indicated by data distribution (T2). The estimated potential false negative responses are 9.3% (13/139) for dose-response relationships satisfying hypothesis testing entry criteria and 10.1% (14/139) for dose-response relationships satisfying distributional data entry criteria.

Although the above findings suggest that hormetic responses are quite common if assessed with the appropriate study design criteria, only 1% of the more than 20,000 published articles contained data meeting the study design criteria for entry into the database. This emphasizes the fact that very few published studies have the potential for detecting hormetic responses in the low-dose region of dose-response relationships. In fact, the criteria used in the present study ignored temporal features. If adequate temporal features were required, the proportion of studies satisfying entry criteria would have been far less than the 1%. Yet, if hormetic effects are to be adequately characterized, multiple appropriately spaced doses need to be assessed over multiple periods. The dual combination of multiple doses and periods places extraordinary demands on the investigation and are generally ignored, at least in part, thereby affecting the opportunity to assess hormetic effects. Thus, it is not surprising that hormetic effects have been considered exceptions or paradoxical responses, as our findings indicate that only 1 of 100 studies has the appropriate dosage design needed to assess this hypothesis.

Although there are multiple reasons why entry criteria were not satisfied, the most likely reason is that the hormetic evaluation has high study design criteria requirements, especially with respect to the number of doses and doses below the NOAEL. At a minimum, four doses plus a concurrent control are required, with two of the four doses being below the NOAEL. Historically, there has been a strong emphasis on high-dose evaluation, as these responses are often more definitive and publishable for defining the NOAEL. These factors minimize the proportion of experiments that emphasize below-NOAEL responses. Likewise, there has been the long-standing belief that responses below the NOAEL are most likely due to normal variation and not reproducible treatment effects. It is this very central assumption of modern experimental and regulatory toxicology that the present findings challenge. Yet, it is this historically controlling assumption that has strongly influenced past toxicological study designs and contributes to the observation that 99% of studies do not satisfy the entry criteria for hormetic responses.

The selection of the three journals noted in the Methods section was designed to achieve a broad representation of biological models, agents tested, and end points assessed. Although this approach was generally successful in achieving these goals, there were several important omissions or underrepresentations in certain categories. For example, the number of microbiological models was minimally represented; likewise, studies involving various types of radiation were also minimal. Nonetheless, areas such as microbiological responses and effects of radioactivity have been extensively documented and are represented in the earlier and separate database developed by Calabrese and Baldwin (1997a,b). Such underrepresentation in the present study is believed to be a result of the journal selection rather than a biological restriction of the hormetic response.

The findings presented here add to and strengthen the earlier reports on the potential widespread generalizability of hormesis. They provide a useful complement to the Calabrese and Baldwin (1997a,b) hormesis database, which includes several thousand examples of dose-response relationships satisfying quantitative criteria for assessing hormesis, as well as study replication and mechanistic findings that account for the biphasic nature of the dose and temporal responses of the hormetic phenomenon. Although numerous examples of apparent hormetic responses exist independent of chemical, biological model, and end point, the previous database cannot address the issue of frequency of occurrence of hormetic responses. The current study addresses this limitation and suggests that hormetic responses are commonly encountered if the study design is appropriate.

The present findings have important implications for the

design, conduct and interpretation of toxicological investigations as well as the potential to alter current concepts of NOAEL and challenge findings of risk assessment modeling activities commonly used for regulatory practices that assume linearity in low-dose areas. More specifically, for hormetic effects to be properly assessed, it is important that consideration be given to animal model and end-point selection. For example, an assessment of end points such as mutagenicity, carcinogenicity, and teratogenicity within a hormetic framework cannot be made using models with zero or negligible background/control incidence. It is also important to establish in a reliable manner the NOAEL for end points of interest and to include multiple and carefully spaced doses below the NOAEL. Furthermore, it may be necessary to include a temporal component within the study design if the hormetic mechanism represents an overcompensation response (Hart and Frame, 1996; Morré, 2000; Stebbing, 1998). The above suggestions are not trivial recommendations, as they require the commitment of substantial additional resources. Nonetheless, these features are necessary to more properly determine the nature of the dose response in the low-dose zone.

These findings address fundamental aspects of the nature of the dose response in the low-dose zone and suggest the need to incorporate U-shaped features in future modeling aspects of biological responses. Although the current investigation has focused on toxicologically derived data, sufficient data exist within the original hormesis database to indicate that this phenomenon is operational and similarly significant across the broad spectrum of biological, pharmacological, and other biomedical disciplines.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Kenny Crump (ICF Clements) for his guidance and technical assistance. This effort was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant F49620-98-1-0091. The U.S. Government is authorized to reproduce and distribute for Governmental purposes notwithstanding any copyright notation thereon.

REFERENCES

Calabrese, E. J. (in press). Overcompensation stimulation: A mechanism for hormetic effects. *Crit. Rev. Toxicol*.

Calabrese, E. J., and Baldwin, L. A. (1997a). A quantitatively-based methodology for the evaluation of chemical hormesis. *Hum. Ecol. Risk Assess*. **3,** 545–554.

Calabrese, E. J., and Baldwin, L. A. (1997b). The dose determines the stimulation (and poison): Development of a chemical hormesis database. *Int. J. Toxicol*. **16,** 545–559.

Calabrese, E. J., and Baldwin, L. A. (2000a). Chemical hormesis: Its historical foundations as a biological hypothesis. *Hum. Exp. Toxicol*. **19,** 2–31.

Calabrese, E. J., and Baldwin, L. A. (2000b). Radiation hormesis: Its historical foundations as a biological hypothesis. *Hum. Exp. Toxicol*. **19,** 41–75.

Calabrese, E. J., and Baldwin, L. A. (2000c). Radiation hormesis: The demise of a legitimate hypothesis. *Hum. Exp. Toxicol*. **19,** 76–84.

Calabrese, E. J., and Baldwin, L. A. (2000d). Tales of two similar hypotheses: The rise and fall of chemical and radiation hormesis. *Hum. Exp. Toxicol*. **19,** 85–97.

Calabrese, E. J., and Baldwin, L. A. (2000e). The marginalization of hormesis. *Hum. Exp. Toxicol*. **19,** 32–40.

Calabrese, E. J., Baldwin, L. A., and Holland, C. D. (1999). Hormesis: A highly generalizable and reproducible phenomenon with important implications for risk assessment. *Risk Anal.* **19,** 261–281.

Crump, K. S. (2001). The regulatory implications of hormesis: Is hormesis a universal phenomenon? *Crit. Rev. Toxicol.*, in press.

Davis, J. M., and Svendsgaard, D. J. (1994). Nonmonotonic dose-response relationships in toxicological studies. In *Biological Effects of Low Level Exposures: Dose-Response Relationships* (E. J. Calabrese, Ed.), pp. 67–85. CRC Press, Boca Raton.

Hart, R. W., and Frame, L. T. (1996). Toxicological defense mechanisms and how they may affect the nature of dose-response relationships. *BELLE Newsl.* **5,** 1–16.

Morré, D. J. (2000). Chemical hormesis in cell growth: a molecular target at the cell surface. *J. Appl. Toxicol.* **20,** 157–163.

Stebbing, A. R. D. (1998). A theory for growth hormesis. *Mutat. Res.* **403,** 249–258.