J. ENVIRONMENTAL SYSTEMS, Vol. 9(2), 1979-80

AN EXAMPLE OF DECISION-MAKING ON ENVIRONMENTAL CARCINOGENS: THE DELANEY CLAUSE

WILLIAM P. DARBY, PH. D.
Department of Technology and Human Affairs
Washington University

ABSTRACT

This analysis focuses on the decision-making process of regulating environmental carcinogens as carried out under the Delaney Clause of the Federal Food, Drug, and Cosmetics Act. The study concludes that regulation of environmental carcinogens in this way is unavoidably based upon a relative assessment of risks and benefits, even though the Delaney Clause is commonly believed to preclude such an assessment. The analysis further concludes that by waiting for evidence acceptable to the scientific community before acting, the decision-maker may unknowingly be operating under a relative assessment of risks and benefits not in accord with his or her own perception.

INTRODUCTION

The current controversy involving action by the Food and Drug Administration in regulating the use of saccharin as a food additive provides a useful example of regulating environmental carcinogens under the "Delaney Clause" of the Federal Food, Drug and Cosmetics Act, which states:

> Provided, that no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal, or if it is found, after tests which are appropriate for the evaluation of the safety of food additives, to induce cancer in man or animal . . . [1].

The Delaney Clause represents an "all-or-nothing" approach to regulating environmental carcinogens. That is, it allows no regulatory action (for example, establishing maximum safe tolerance levels) other than a ban, and presumably does not allow a risk/benefit analysis as the basis of, or even as input to, this phase of regulatory decision-making.

doi: 10.2190/3RTA-83T0-GUFQ-HTME
http://baywood.com

The objective of this paper is to show that action taken on environmental carcinogens under the Delaney Clause (and, in fact, any public policy action predicated on "purely-scientific" evidence) ultimately and unavoidably includes an assessment of risks and benefits. It is not the objective of this paper to treat the epidemiological questions of defining a dose-response function or extrapolating animal data to humans. The focus is on the role of scientific experimentation in decision-making on environmental carcinogens under the Delaney Clause.

CLASSICAL STATISTICAL INTERPRETATION OF SCIENTIFIC INVESTIGATIONS

The results of any scientific investigation can only be properly interpreted using the principles of statistical inference. Despite the common practice of calling the results of tests on chemicals for carcinogenicity and other effects (toxicity, mutagenicity, teratogenicity, tumorigenicity, etc.) either "positive" or "negative," the sense of certainty conveyed by these terms is false. There is always some chance that the conclusion drawn from the experimental results is incorrect.

Among the scientific community, the results of investigations are usually interpreted using the techniques of classical (as opposed to Bayesian) statistics. Investigations of the carcinogenicity of a chemical substance usually compare a group of subjects (laboratory animals, for example) treated with the chemical substance to a control group of similar subjects not treated with the chemical. The comparison is carried out in terms of the frequencies with which carcinogenesis is observed in the two groups.

The initial (null) hypothesis is that the frequencies of carcinogenic effects are the same in both the treatment group and the control group. The data from the investigation are interpreted to determine if the null hypothesis can be rejected in favor of an alternative hypothesis: that the frequency of carcinogenesis among the treatment group is greater than the frequency of carcinogenesis among the control group. Interpreting the experimental data results in either accepting the null hypothesis or rejecting it in favor of the alternative hypothesis.

In performing the statistical interpretation of the experimental data, one of two errors can result:

1. the null hypothesis is rejected, concluding an increase in carcinogenesis as the result of treatment with the chemical substance, when in reality no increase results (Type I statistical error); or
2. the null hypothesis is accepted, concluding no increase in carcinogenesis as the result of treatment with the chemical substance, when in reality an increase does result (Type II statistical error).

EXPERIMENTAL DESIGN

As shown in the appendix, the results of experiments intended to show chemical carcinogenesis are usually displayed in the form of contingency tables similar to the example shown in Table 1.

Interpreting the results shown in Table 1 involves recognizing that there is some spontaneous, or background, carcinogenesis rate which is independent of treatment with the chemical substance, and which governs the frequency of carcinogenesis among the control group. Determining whether the carcinogenesis rate among the treatment group is greater than the rate among the control group hinges upon choosing an upper limit for the number of cancers which could be observed among the treatment group and still be reasonably attributed to the background rate and its normal random variations. If the observed number of cancerous animals in the treatment group does not exceed the chosen value, then the null hypothesis is accepted.

Choosing the upper limit by the principles of experimental design requires a trade-off of the acceptabilities of committing each type of error (incorrectly accepting or incorrectly rejecting the null hypothesis), and also a judgment of how sensitive the test ought to be to true differences between the carcinogenesis rate after treatment with the chemical substance and the background rate, if differences do exist. As discussed in the appendix, these differences are stated in terms of the odds ratio, where an odds ratio equal to one indicates that the true treatment and background carcinogenesis rates are identical. An odds ratio greater than one indicates that the true treatment carcinogenesis rate is greater than the true background carcinogenesis rate.

The investigator may choose a very high cutoff value (the number of cancerous animals in the treatment group for which the null hypothesis will still be accepted). This is equivalent to choosing a very low probability value for incorrectly accepting the null hypothesis (significance level). The higher the cutoff value for which the null hypothesis will still be accepted, the higher the true value of the odds ratio needs to be to maintain the same probability of erroneously accepting the null hypothesis, given specified numbers of test animals in each test group and a specified total (of both groups) of cancerous animals. Similarly, the higher the cutoff value for which the null hypothesis will still be accepted and the closer the odds ratio is to unity, the higher the probability that the statistical test will not be sensitive enough to detect the small differences between the true treatment and background carcinogenesis rates. In this case, the chances of incorrectly accepting the null hypothesis are high.

Typically, the designs of scientific investigations evolve from economic constraints on such criteria as numbers of laboratory animals used, and from convention, which sets a value of 0.05 as the maximum acceptable probability of incorrectly rejecting the null hypothesis. Rarely are tradeoffs involving the

Table 1. Example Results of Carcinogenicity Test[a]

	Number of animals		
	With Tumors	*Without Tumors*	*Totals*
Treatment Group	A	B	45
Control Group	C	D	42
Totals	12	75	87

Null Hypothesis: True Odds Ratio = 1; Alternate Hypothesis: True Odds Ratio > 1.

Criterion for Rejecting Null Hypothesis: A>	*Probability of Incorrectly Rejecting Null Hypothesis*	*Probability of Incorrectly Accepting Null Hypothesis for True Odds Ratio =*			
		1.5	*5*	*10*	*20*
0	1.000	0.[b]	0.	0.	0.
1	0.999	0.0001	0.	0.	0.
2	0.991	0.001	0.	0.	0.
3	0.955	0.009	0.	0.	0.
4	0.856	0.042	0.0001	0.	0.
5	0.670	0.134	0.001	0.	0.
6	0.429	0.313	0.009	0.0004	0.
7	0.211	0.556	0.043	0.004	0.
8	0.075	0.783	0.151	0.025	0.003
9	0.018	0.927	0.377	0.116	0.025
10	0.003	0.985	0.683	0.362	0.142
11	0.0002	0.999	0.920	0.743	0.507

[a] The marginal totals used in this example are the results of the 1977 Canadian study of saccharin, for second-generation (F_1) male rats [2, 3]. The treatment group was fed a standard laboratory ration with saccharin added to comprise 5.0 per cent of the diet. All twelve animals with tumors (4 benign; 8 malignant) were found in the treatment group. This was the most sensitive group in the study to the carcinogenic effect of saccharin [4].
[b] 0. means less than 0.00005.

relative consequences of the two types of error, used by the investigator in determining the maximum acceptable probability of incorrectly accepting the null hypothesis.

In the public policy arena, however, the decision-maker may be particularly concerned with criteria overlooked by the investigator. Not only might trade-offs made by the decision-maker differ markedly from trade-offs made by the investigator, but neither might realize what (or even, that) trade-offs are incorporated into the interpretation of the results provided to the decision-maker by the investigator, and upon which the decision-maker bases his action.

MINIMUM RISK

The total risk (expected loss) involved in interpreting a set of experimental results at a specific cutoff value can be calculated if the corresponding values of a, the maximum acceptable probability of incorrectly rejecting the null hypothesis (Type I statistical error) and of b, the accompanying probability of incorrectly accepting the null hypothesis (Type II statistical error) have been determined. The value of b is a function of the true odds ratio, which is theoretically unknowable. However, for this analysis, it will be assumed that the true odds ratio can be reasonably estimated.

Combining these probability values with judgments of the relative undesirability ("cost") of committing a Type I error and a Type II error, respectively C_a and C_b, the total risk, R_T, can be calculated as the expected loss due to inferential error:

$$R_T = C_a a + C_b b \tag{1}$$

For example, assume that the results of a particular experiment are interpreted at a cutoff value which corresponds to a = 0.01, and the required sensitivity of the test to true values of the odds ratio results in b = 0.05. Also, assume that the relative undesirability of committing each type of error can be represented by distributing 1,000 units between C_a and C_b. If an individual chooses $C_a = 900$ and $C_b = 100$, the total risk (R_T) is 14 units. However, if the choice is $C_a = 100$, and $C_b = 900$, the total risk is 46 units. At these values for a and b, the minimum risk will be asymptotic to 10 at $C_a >> C_b$ and the maximum risk will be asymptotic to 50 at $C_a << C_b$. Figure 1 illustrates this behavior, and shows that, assuming the individual intuitively wants to minimize Equation 1, accepting a = 0.01 and b = 0.05 as maximum probability values for interpreting the experimental results incorrectly, means that the individual estimates $C_a/C_b \geqslant 100$. Under the assumption that an individual behaves to minimize total risk as expressed in Equation 1, whenever he or she chooses to interpret the results of a scientific investigation at a particular cutoff value, that action determines the values of a, b, and most importantly, C_a/C_b. (While the latter two values depend on the unknowable true value of the odds ratio, there is a true value for the odds ratio, and if it can be reasonably estimated, values for b and C_a/C_b can also be reasonably estimated.) Thus, whenever the results of an investigation are subjected to statistical interpretation, a judgment of the relative costs of erroneous conclusions is inevitable.

COSTS OF ERRONEOUS CONCLUSIONS: THE INVESTIGATOR AND THE DECISION-MAKER

The value of C_a/C_b may have little meaning for the investigator. He or she typically does not recognize costs of erroneous conclusions. On the other hand, the decision-maker is keenly aware of the meaning of C_a/C_b.

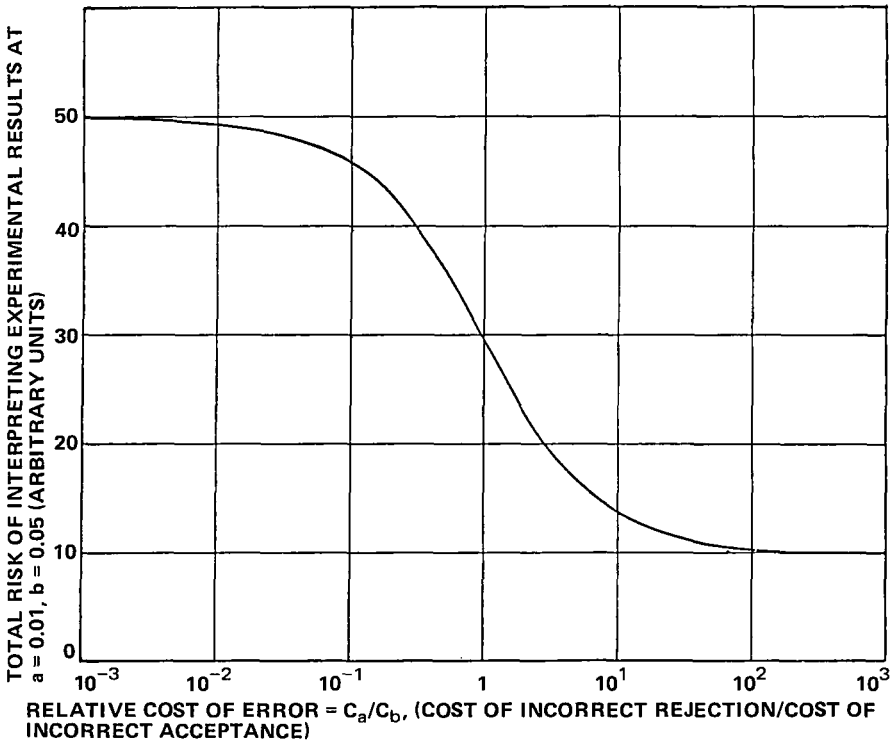

Figure 1. Typical total risk curve for interpreting experimental results.

In the case of action on food additives under the Delaney Clause, if the scientific investigator incorrectly rejects the null hypothesis, concluding that the additive causes cancer when in reality it does not, and the decision-maker consequently bans the chemical, C_a is the cost of banning a non-carcinogenic chemical. In addition to the transaction costs of the ban, C_a also includes the benefits of the chemical which users must forego and the possible additional cost of using substitutes, if they are available. Similarly, C_b, the cost of incorrectly accepting the null hypothesis, is interpreted by the decision-maker as the cost of allowing a carcinogenic chemical to remain in use as a food additive. This cost includes the increased risk of cancer for the population of users. Thus, when the results of a scientific investigation are viewed in their role in the decision-making process under the Delaney Clause, the value for C_a/C_b, which is inevitably determined, is really a relative assessment of the benefits and risks of the chemical under test.

THE DELANEY CLAUSE: RISKS AND BENEFITS

The widely-accepted notion that the Delaney Clause does not allow a risk/benefit analysis is false. A risk/benefit assessment is inevitably included in every

action under the Delaney Clause, because the notion that there exists a purely scientific judgment is also false. If the decision-maker accepts the interpretation of the experimental results provided by the investigator, the decision-maker implicitly accepts the value of C_a/C_b, determined by the investigator's method of interpretation, as a relative risk/benefit assessment of the chemical under test. Not only might this assessment be very different from the decision-maker's own perception of risks and benefits, but neither the decision-maker nor the investigator might realize that a specific value (C_a/C_b) has been accepted as a risk/benefit assessment, and that the entire decision-making process hinges on it.

REGULATING ON THE BASIS OF SCIENTIFIC EXPERIMENTATION

For a given set of experimental results (numbers of test animals in the treatment group and in the control group, and total number of cancerous animals in both groups), and a specific value for the true odds ratio, the lower the probability of incorrectly rejecting the null hypothesis (banning a non-carcinogenic chemical) an individual is willing to accept, the higher the probability of incorrectly accepting the null hypothesis (not banning a carcinogenic chemical) he or she must be willing to accept.

For example, in Table 1, assume that the scientific community is not willing to reject the null hypothesis until greater than nine of the twelve cancerous animals are found in the treatment group. This cutoff value corresponds to a significance level of 0.018. If the true value of the odds ratio is estimated to be 20, the probability of incorrectly accepting the null hypothesis is 0.025. Under these conditions, the total risk, calculated using Equation 1, approaches the asymptotic minimum when C_a/C_b exceeds 100. Thus, if the decision-maker waits for evidence acceptable to the scientific community before banning the chemical, his judgment is based on a risk/benefit assessment that mistakenly banning a non-carcinogenic substance is at least 100 times as undesirable as mistakenly not banning a carcinogenic substance.

Similarly, if the decision-maker estimates that that mistakenly not banning a carcinogenic substance is 100 times as undesirable as mistakenly banning a non-carcinogenic substance, then the null hypothesis must be rejected if greater than seven of the twelve cancerous animals are found in the treatment group. This results in a significance level of 0.211 and (estimating the true value of the odds ratio at 20) a probability of incorrectly accepting the null hypothesis equal to 0.0002. These probability values minimize Equation 1 over the set of all possible choices of the cutoff value for interpreting the experimental results, if $C_a/C_b = 0.01$. It is highly unlikely that experimental results which only achieve a significance level of 0.211 would be accepted by the scientific community as evidence of a carcinogenic effect.

The implications of the above example are clear. In the former case, which fits the model of the Delaney Clause well, when a decision-maker waits until

evidence acceptable to the scientific community is available before regulating a suspected carcinogen, the regulation is based upon an assessment that banning a non-carcinogenic substance is at least 100 times as undesirable as not banning a carcinogenic substance. In the latter case, if the decision-maker estimates that not banning a carcinogenic substance is 100 times as undesirable as banning a non-carcinogenic substance, then he or she must be willing to regulate, based upon evidence of carcinogenic properties which would not normally be acceptable to the scientific community. In both cases, a risk/benefit assessment is an unavoidable part of the decision-making process.

SUMMARY AND CONCLUSIONS

It has been shown that a risk/benefit assessment is part of every public policy action which is based upon the interpretation of the results of a scientific investigation. The risk/benefit assessment comes in the form of a judgment about the relative undesirability of performing a statistical analysis of the experimental data which either incorrectly rejects or incorrectly accepts the null hypothesis.

The assessment may be included implicitly (and perhaps, unknowingly), when the decision-maker acts only upon experimental evidence which is acceptable to the scientific community. In this case, the decision-maker accepts the assessment which is contained in criteria established by the scientific community for judging experimental results. Neither the experimenter nor the decision-maker may be aware of the particular assessment, which underlies the so-called "purely-scientific" decision. On the other hand, the assessment may be included explicitly by the public policy decision-maker in deciding if the experimental results reported by the experimenter provide sufficient evidence upon which to act, whether or not the evidence is acceptable to the scientific community.

Examination of Table 1 shows that for a given set of experimental results (numbers of test animals in the treatment group and in the control group, and total number of cancerous animals in both groups) and for specific values for the true odds ratio, the lower the acceptable probability of incorrectly rejecting the null hypothesis, the higher the allowable probability of incorrectly accepting the null hypothesis must be. The inverse is also true. If it is assumed that by only rejecting the null hypothesis when a reasonably small probability of incorrect action is achieved, the scientific community assesses the desirability of incorrect acceptance of the null hypothesis to be greater than incorrect rejection, (i.e., $C_a/C_b > 1$), the Delaney Clause is seen in a different light. A public policy decision-maker predisposed to protect human health might be expected to assess incorrect rejection of the null hypothesis more desirable than incorrect acceptance (i.e., $C_a/C_b < 1$). If the null hypothesis is incorrectly rejected, a non-carcinogenic substance will be banned. If the null hypothesis is incorrectly accepted, a carcinogenic substance will not be banned. Barring deleterious

effects of substitutes, the former incorrect action does not adversely affect human health; the latter incorrect action does.

Consequently, the Delaney Clause, often thought to be a strong safeguard of human health, and to preclude a judgment based on a risk/benefit assessment, may be a rather weak mechanism which always includes a risk/benefit assessment. Given an explicit risk/benefit assessment in which the risks far outweigh the benefits, a decision-maker would probably ban a substance based on evidence of carcinogenesis which would not be significant enough to be accepted by the scientific community, and which might not trigger the Delaney Clause mechanism.

APPENDIX

Statistical Methodology

Determining whether a greater tumorigenesis rate exists for laboratory animals treated with a test substance (treatment group) than for laboratory animals not treated with a test substance (control group) can be formulated in terms of the classical statistical technique of testing for independence in a 2 X 2 (fourfold) contingency table. Thus, if the treatment group (S) and the control group ($\bar{S}$), which consist respectively of E and F animals, are categorized with respect to the presence (T) or absence ($\bar{T}$) of tumors, the laboratory results can be expressed as:

	Tumorigenesis		
Treatment	T	$\bar{T}$	*Totals*
S	A	B	E
$\bar{S}$	C	D	F
Totals	G	H	

The numbers of animals with tumors in the two sample groups (treatment and control) are independent variables with binomial distributions characterized by parameters $p_1 = P(T/S)$ and $p_2 = P(T/\bar{S})$. The case of independence is, therefore, that $p_1 = p_2$.

Several types of statistical tests exist to test hypotheses concerning p_1 and p_2 [5, 6]. The approach chosen here is Fisher's exact test, which does not require large sample sizes and minimum cell frequencies [5, 7-9]. It does, however, require that the marginal totals are fixed. That is, totals of T, $\bar{T}$, S, $\bar{S}$, are G, H, E, and F, respectively. The Fisher exact test evaluates the probabilities of tables which have E, F, G, and H as the marginal totals under various hypotheses concerning p_1 and p_2.

Hypotheses concerning p_1 and p_2 can be stated based upon several different measures, including the difference $(p_1 - p_2)$ and the ratio (p_1/p_2). Because of ease of formulation (Dunnett, 1977) the measure used in this analysis is the odds ratio, α, which can be expressed as

$$\alpha = \frac{p_1(1-p_2)}{p_2(1-p_1)} \tag{1}$$

Since interest is typically in the number of animals in the treatment group with tumors (A), the probability of any particular value, a, under the assumptions of fixed marginal totals and an odds ratio α_0 is

$$P(A = a/E,F,G,\alpha_0) = \frac{\binom{E}{a}\binom{F}{G-a}\alpha_0^a}{\sum_i \binom{E}{i}\binom{F}{G-a}\alpha_0^i} \tag{2}$$

where $\binom{x}{y}$ is the binomial coefficient, that is:

$$\binom{x}{y} = \frac{x!}{y!\,(x-y)!} \tag{3}$$

and the summation over i is from max(0,G-F) to min(E,G). Typically, these values are 0 and G.

The null hypothesis, that the two tumorigenesis rates p_1 and p_2 are equivalent, is stated, using the odds ratio, as $\alpha = 1$. The alternate hypothesis of interest is $p_1 > p_2$. In terms of the odds ratio, this is equivalent to $\alpha > 1$. Interpretation of the experimental results is based upon choosing a value of a as the maximum value for which the null hypothesis will still be accepted. That is, a is chosen such that the probability of values of $A > a$ occurring when the null hypothesis is true (significance level, probability of incorrectly rejecting the null hypothesis, probability of Type I error) is sufficiently small.

Note that this process is somewhat different from the common practice of choosing a significance level for interpreting the results. This is necessary because of the discrete nature of Fisher's exact test. There are a finite number of possible values of A, given the condition of fixed marginal totals. Each value of A corresponds to a specific contingency table, with a certain probability of occurrence for a particular value of the odds ratio. Consider for example, the case with the marginal totals:

	T	$\tilde{T}$	Totals
S	A	B	5
$\tilde{S}$	C	D	5
Totals	4	6	10

Only five possible tables exist, with the associated probabilities (calculated using Equation 2, with $\alpha = 1$):

a	*b*	*c*	*d*	*P(A = a/5,5,4,1)*	*P(A≤a/5,5,4,1)*	*P(A>a/5,5,4,1)*
0	5	4	1	0.0238	0.0238	0.9762
1	4	3	2	0.2381	0.2619	0.7381
2	3	2	3	0.4762	0.7381	0.2619
3	2	1	4	0.2381	0.9762	0.0238
4	1	0	5	0.0238	1.0000	0.0000

If a significance level of, say, 0.05 were chosen, then certainly A = 4 would be sufficient to reject the null hypothesis that $\alpha = 1$. However if the rejection criterion were $A > 3$, then the significance level is really 0.0238 (<0.05). If the rejection criterion were $A > 2$, then the significance level is really 0.2619

(> 0.05). There is no straightforward way to interpret this example at a significance level of 0.05. Although modifications do exist to permit this, this analysis will not include them, for simplicity [10, 11].

The probability of rejecting the null hypothesis for alternative values of the odds ratio (power of the test) can also be easily calculated using Equation 2. For the preceding example, the powers of the test for several different values of the odds ratio are shown below:

Rejection Criterion $A >$	*Significance Level*	*Power of the test*						
		$\alpha = 0.5$	$\alpha = 1.0$[a]	$\alpha = 2.0$	$\alpha = 4.0$	$\alpha = 6.0$	$\alpha = 8.0$	$\alpha = 10.0$
0	0.976	0.919	0.976	0.995	0.999	1.000	1.000	1.000
1	0.738	0.513	0.738	0.893	0.967	0.986	0.992	0.995
2	0.262	0.107	0.262	0.487	0.713	0.816	0.871	0.905
3	0.024	0.005	0.024	0.081	0.203	0.306	0.387	0.452

[a]By definition, the power of the test with $\alpha = 1.0$ is the significance level.

The probability value complementary to the power (1-power) is the probability of accepting the null hypothesis for alternative values of the odds ratio. For values of the odds ratio other than 1, it is the probability of incorrectly accepting the null hypothesis (probability of a Type II error). These probabilities are also shown below for various rejection criteria.

Rejection Criterion $A >$	*Significance Level*	*Probability of incorrectly accepting null hypothesis*						
		$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 2.0$	$\alpha = 4.0$	$\alpha = 6.0$	$\alpha = 8.0$	$\alpha = 10.0$
0	0.976	0.081	0.010	0.005	0.001	0.000	0.000	0.000
1	0.738	0.487	0.160	0.107	0.033	0.014	0.008	0.005
2	0.262	0.893	0.611	0.513	0.287	0.184	0.129	0.095
3	0.024	0.995	0.949	0.919	0.796	0.694	0.613	0.548

The above table indicates that once the rejection criterion is established for a particular set of fixed marginal totals, then both the probabilities of incorrectly rejecting the null hypothesis (significance level) and of incorrectly accepting the null hypothesis are also established. The probability of incorrectly accepting the null hypothesis is ultimately determined by the true value of the odds ratio. For example, using the preceding table, if it is decided to reject the null hypothesis (that $\alpha = 1.0$) if the observed value of A exceeds 2, then the probability of incorrectly rejecting the null hypothesis is 0.262. This value is much higher than the value of 0.05 typically used in ordinary scientific research. Even at this large a probability of incorrectly rejecting the null hypothesis, the probability of incorrectly accepting the null hypothesis is also quite large. Achieving a probability of incorrect acceptance of the null hypothesis no greater than the probability of incorrect rejection of the null hypothesis can only be achieved if the true value of the odds ratio is greater than 4.

If the criterion for rejecting the null hypothesis is set at observed values of A which exceed 3 (i.e., A = 4), then the probability of incorrectly rejecting the null hypothesis is 0.024. This value is more in keeping with values normally encountered in interpreting the results of scientific research. However, the accompanying probabilities of incorrectly accepting the null hypothesis are extremely high. Values of incorrect acceptance of the null hypothesis exceed 0.024 unless the true value of the odds ratio is above 375.

REFERENCES

1. U.S. Congress House Committee on Appropriations, Agriculture-Environmental and Consumer Protection Appropriations for 1975, Hearings, 93rd Congress, 2nd Session, Part 8, Food and Drug Administration, Study of the Delaney Clause and Other Anticancer Clauses, U.S. Government Printing Office, 1974.
2. Office of Technology Assessment, U.S. Congress, Cancer Testing Technology and Saccharin, U.S. Government Printing Office, 1977.
3. M. D. Reuber, Preliminary Review of the Carcinogenicity Studies on Saccharin, typescript, September 12, 1977.
4. Saccharin and Its Salts-Proposed Rule and Hearing, *Federal Register, 42:*73, p. 20000, April 15, 1977.
5. J. J. Gart, The Comparison of Proportions: A Review of Significance Tests, Confidence Intervals and Adjustments for Stratification, *Review of the International Statistical Institute, 39,* pp. 148-169, 1971.
6. C. W. Dunnett and M. Gent, Significance Testing to Establish Equivalence Between Treatments, With Special Reference to Data in the Form of 2 X 2 Tables, *Biometrics, 33,* pp. 593-602, December, 1977.
7. R. A. Fisher, The Logic of Inductive Inference, *Journal of the Royal Statistical Society, Series A, 98,* pp. 93-94, 1935.
8. E. L. Lehmann, *Testing Statistical Hypotheses,* John Wiley and Sons, Inc., New York, New York, pp. 140-146, 1959.
9. Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice,* The MIT Press, Cambridge, Massachusetts, pp. 364-366, 1975.
10. S. Siegel, *Nonparametric Statistics for the Behavioral Sciences,* McGraw-Hill Book Company, New York, New York, pp. 95-104, 1956.
11. K. D. Tocher, Extension of the Neyman-Pearson Theory of Tests to Discontinuous Variates, *Biometrika, 37,* pp. 130-144, 1950.

Direct reprint requests to:

William P. Darby, Ph.D.
Department of Technology and Human Affairs
Washington University
St. Louis, MO 63130