Estimates of the Proportion of Chemicals That Were Carcinogenic or Anticarcinogenic in Bioassays Conducted by the National Toxicology Program

Kenny S. Crump,1 Daniel Krewski,2 and Cynthia Van Ladingham1

1The K.S. Crump Group, Inc., Ruston, LA 71270 USA; 2Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada

Estimates were made of the proportion of chemicals that were carcinogenic, anticarcinogenic, or either in 397 long-term bioassays conducted by the National Toxicology Program (NTP). The estimates were obtained from the global pattern of p-values obtained from statistical tests applied to individual experiments. These tests accounted for multiple comparisons using a randomization procedure and were found to be at the correct level of significance. Representative estimates of the proportion of carcinogens [with 90% confidence intervals (CI)] compared to the NTP estimates were as follows: male mice, 0.32 (CI, 0.19–0.44), NTP = 0.29; female mice, 0.28 (CI, 0.15–0.41), NTP = 0.34; male rats, 0.35 (CI, 0.23–0.47), NTP = 0.36; female rats, 0.34 (CI, 0.21–0.46), NTP = 0.28; all sexes and species, 0.59 (CI, 0.49–0.69), NTP = 0.51. Representative estimates of the proportion of anticarcinogens were as follows: male mice, 0.34; female mice, 0.27; male rats, 0.40; female rats, 0.44; all sexes and species, 0.66. Thus, there was as much or more evidence in this study for anticarcinogenesis as carcinogenesis. Even though the estimators used were negatively biased, it was estimated that 85% of the chemicals were either carcinogenic or anticarcinogenic at some site in some sex–species group. This suggests that most chemicals given at high enough doses will cause some sort of perturbation in tumor rates. Key words: carcinogens, carcinogenicity, meta-analysis, NTP bioassays, POLY3 test. Environ Health Perspect 107:83–88 (1999). [Online 16 December 1998] http://ehpnet1.niehs.nih.gov/docs/1999/107p83-88crump/abstract.html

The National Cancer Institute and, more recently, the National Toxicology Program (NTP) have been routinely testing chemicals for carcinogenic potential in long-term rodent bioassays for about 25 years (1–5). Most of the chemicals have been tested in separate groups of animals composed of males and females of two species (generally rats and mice). These bioassays were initially designed to screen for potential human carcinogens, and it was anticipated that further studies would be conducted on the chemicals identified as carcinogenic. Because NTP bioassays involve relatively few animals (generally 50 animals are tested at each of 2 or 3 dose levels, in addition to 50 control animals, in each of the 4 sex–species groups), relatively high doses were used to minimize the possibility that a potential human carcinogen would not be identified (3,4). Because of financial and time constraints, followup studies have often not been conducted, and results from NTP bioassays have been used extensively in the regulation of chemicals—not only to identify potential carcinogenic hazards to humans, but also to determine dose–response relationships. More than 400 chemicals have been tested to date by the NTP, about one-half of which have been identified as carcinogenic in at least one sex–species group (1–5). Some researchers have expressed concern at finding such a high percentage of carcinogens and believe that many of these carcinogenic responses are generalized indirect reactions to high-dose toxicity, which consequently may not be relevant to humans exposed to much lower levels (5). On the other hand, it has been pointed out that most chemicals were selected for testing by the NTP because they were suspected carcinogens and are not necessarily representative of all chemicals in commercial use (3,6,7).

In the current study, we developed independent estimates of the proportions of chemicals that were carcinogenic in each sex–species group, as well as the proportion of chemicals that were carcinogenic in any group, among chemicals tested by the NTP, and compared these estimates to the proportions identified by the NTP. We also estimated the proportions of chemicals that were anticarcinogenic (i.e., caused dose-related decreases in tumor incidence in one or more tumor categories) and the proportions that were either carcinogenic or anticarcinogenic. The NTP did not routinely evaluate chemicals for anticarcinogenesis. These estimates are based on a substantially different methodology (8–11) than that used by the NTP to identify carcinogens. Unlike the NTP approach, the estimation procedure used here did not involve determination of whether an individual chemical was carcinogenic. Rather, we examined the overall distribution of p-values obtained from statistical tests applied to each of the NTP bioassays. If there were no carcinogens present, these p-values would theoretically be uniformly distributed. The proportion of carcinogens is estimated from the pattern of departure from the uniform distribution. The approach does not rely on a single cutoff for p-values (e.g., p = 0.05) and accounts appropriately for the proportion of p-values in any range (e.g., an excess of p-values in the range 0.1–0.2, or a deficit in the range 0.8–1.0).

Using a similar methodology, Crump et al. (10) estimated that there were appreciably more liver carcinogens (166) among 390 NTP bioassays than were identified by the NTP (108). Other studies (12–14) have examined the frequency of carcinogenic or anticarcinogenic responses in the NTP database. However, these investigations focused on specific levels of statistical significance (e.g., 0.05 or 0.01). We believe that the methodology used in the current paper, which exploits the entire distribution of p-values, may allow for a more definitive assessment of carcinogenesis and anticarcinogenesis in NTP bioassays.

Methods

This study was based on data from 397 long-term carcinogenicity bioassays conducted by the NTP (10) and stored in either the Carcinogenesis Bioassay Data Base (CBD; 320 studies), or its successor, the Toxicity Data Management System (TDMS; 77 studies). Most of these studies involved mice and rats, and our analysis was restricted to these two species. Generally, males and females of a species were tested in separate experiments, most of which involved two or three dose levels, in addition to a control group. The highest dose used in each experiment was an estimate of the maximum tolerated dose. Most of the experiments used 50 animals per dose group, although control groups in a few of the earlier studies contained as few as 10 animals. Experiments that the NTP considered inadequate for determining carcinogenicity were eliminated, which left a total of 211 chemical–species combinations, of which 103 were considered adequate, 17 were marginal, and 81 were inadequate.

Address correspondence to K.S. Crump, The K.S. Crump Group, Inc., 602 E. Georgia Avenue, Ruston, LA 71270 USA. The careful review of this work by Joe Haeman is gratefully acknowledged. This research was supported by Health Canada contract no. 5167. Received 17 April 1998; accepted 8 September 1998.
Table 1. Numbers of studies with various NTP classifications and percent significant (p<0.05) in the present analysis.

| NTP classifications | Positive (%) | Equivocal (%) | Negative (%) | Concordance* (%) |
|---------------------|--------------|---------------|--------------|------------------|
| Male mice           | 108 (90.7)   | 24 (20.2)     | 24 (20.2)    | 84.7             |
| Female mice         | 129 (87.6)   | 24 (20.2)     | 22 (17.6)    | 86.9             |
| Male rats           | 130 (76.9)   | 42 (28.6)     | 19 (12.5)    | 80.2             |
| Female rats         | 103 (83.5)   | 38 (26.3)     | 22 (14.4)    | 83.9             |
| All sexes and species | 201 (83.1)  | 60 (21.7)     | 136 (7.4)    | 85.6             |

*Percentage of experiments classified as positive by NTP with POLY3 p-value <0.05, plus percentage of experiments classified as equivocal or negative by the NTP with p-value >0.05.

Categories of tumors that were subjected to statistical analysis were defined similarly to the categories routinely analyzed by the NTP and were reviewed by Joseph Haseman of the NTP for consistency with NTP practice. There was a total of 93 possible tumor categories for CBDS studies and 79 for TDMS studies. The test statistic, T, for an effect in a sex-species experiment was the largest of the POLY3 test statistics, after application of a continuity correction, derived from any tumor category. To ensure that the test based on T had the proper false positive rate, its p-value was determined using a randomization procedure (18,19). In this procedure animals from a given sex-species experiment were randomly reassigned to treatment groups a total of 400 times, and the POLY3 was reapplied to each set of randomized data. To control for potential differences in life span among different treatment groups, in these random reassignments animals were stratified into four groups according to elapsed time from beginning of study until death, and the total number of animals in each treatment-time category was kept fixed at its value in the original data. The p-value for the test statistic, T, was defined as the proportion of times (out of a total of 400) for which the largest POLY3 test statistic from the randomly created data set equaled or exceeded T. The p-value for a test for an effect in any sex-species experiment was defined as 1 - (1 - Pmin)k, where Pmin was the minimum p-value from all sex-species experiments, and k was the number of such experiments in a study (in the majority of studies, k = 4).

The estimate used for the proportion of positive experiments was

\[ K(a) = \frac{[F(a) - a]}{(1 - a)}, \]

where F(a) was the empirical distribution function of p-values (i.e., F(a) was the proportion of experiments in which the p-value was ≤a). If the statistical test is operating at the correct level of significance (e.g., if its p-value is uniformly distributed whenever treatment does not affect tumor rates) then the estimator, K(a), is negatively biased for every value of a, and the bias of K(a) decreases, and its variance increases, as a approaches 1 (8,11).

A simulation was conducted to determine the properties of the estimator K(a) in the null case (i.e., the situation in which none of the chemicals studied by the NTP were either carcinogenic or anticarcinogenic). In this exercise, a set of null data was fabricated for each NTP experiment by randomly reassigning animals in the manner described earlier. Then our estimation procedure was applied to the set of fabricated data rather than to the actual data.

367 experiments involving male mice, 373 of female mice, 366 of male rats, and 367 of female rats, which collectively tested a total of 397 chemicals (more precisely, chemical-exposure route categories, as a few chemicals were tested by different routes). A designation of "inadequate" usually resulted from either 1) a determination that the experimental animals did not survive long enough to provide an adequate test of carcinogenicity or 2) a determination that the test animals in the high-exposure group could have withstood a higher exposure.

We developed estimates of the proportion of carcinogenic chemicals in each of the four sex-species groups and of the proportion that were carcinogenic in any one of the four groups. We developed corresponding estimates for the proportion of chemicals that were anticarcinogenic or either carcinogenic or anticarcinogenic. A chemical could be classified as both carcinogenic and anticarcinogenic if it caused a dose-related increase in tumor incidence at one site and a dose-related decrease at a different site.

The estimation procedure used p-values from statistical hypothesis tests applied to individual sex-species experiments. These p-values were derived from application of the POLY3 test for a treatment-related trend in tumor incidence (15) to data from all of the dose groups on each of a number of individual tumor categories that were present in a given sex-species experiment. After evaluating a number of statistical tests, we selected the POLY3 test because significance levels from this test appeared to be closest to the nominal values (10,16). Although the POLY3 test had not been used historically by the NTP, it has recently become the NTP’s preferred method for analyzing animal carcinogenic data (17).
To reduce the effect of random variation, this procedure was repeated 10 times for each experiment, and the estimator \( K(a) \) calculated from the combined data. To reduce the amount of computer time required, \( p \)-values were based on 100 random assignments, rather than 400.

**Results**

Originally, the NTP classified individual tumor categories as P (positive for carcinogenicity), E (equivocal), N (negative), or I (inadequate). Later the classification scheme was changed to CE (clear evidence), SE (some evidence), EE (equivocal evidence), NE (no evidence), and IS (inadequate study). To have a common classification system for all NTP studies, we designated P, CE, and SE as positive (P), E and EE as equivocal (E), and N and NE as negative (N). Data from experiments with a classification of I or IS were not used. An experiment was then classified by the highest classification of any tumor site (i.e., if at least one tumor category was classified as P, then the experiment received that classification; otherwise, if at least one tumor category was classified as E, then the experiment received that classification, otherwise the experiment was classified as N). The same procedure was used to categorize a study based on the experiment-wise classifications.

Table 1 compares the carcinogenicity classifications obtained by the NTP (modified as explained in the previous paragraph) to those obtained from our analysis using a conventional significance level of 0.05. Experiments classified by the NTP as positive had \( p \)-values <0.05 in our analysis in 91% of male mice experiments, 88% of female mice experiments, 77% of male rat studies, and 84% of female rat experiments. Correspondence was better in experiments classified as negative by the NTP, as only 4.4–8.2% of these had \( p \)-values <0.05. Between 26 and 33% of the studies classified as equivocal by the NTP had \( p \)-values <0.05. In overall study classifications, a \( p \)-value <0.05 was obtained in 84% of the studies classified as positive by the NTP, 7.4% of those classified as negative, and 22% of those classified as equivocal.

Figure 1 shows the cumulative distributions of \( p \)-values obtained from the simulation of the null case in which treatment has no effect on tumor rates. Each of the sex- and species-specific distributions appear to be a good approximation to the uniform distribution (represented by the line \( y = x \)), a condition that is required for the theoretical properties of our estimation procedure (described in the Methods section). The greatest departure from the uniform distribution occurred in the case of female mice, where the cumulative distribution function (CDF) graph lay slightly below the line \( y = x \) for values of \( p \) between 0.5 and 0.9. If this tendency occurred in the real data, it would result in a slight underestimate of the proportion of carcinogens. The simulated cumulative distribution for all sexes and species is more irregular and has a “sawtooth” appearance for small values of \( p \). This effect is apparently due to the limited number of reassignments used to calculate a \( p \)-value. Because only 100 reassignments were used in this simulation study, the smallest possible nonzero sex- and species-specific \( p \)-value was 0.01. However, the smallest \( p \)-value for all sexes and species was larger, and in most studies (those studies involving mice and rats of each sex for a total of four experiments) was

\[
1 - (1 - 0.01)^4 = 0.0394.
\]

This “sawtooth effect” would be less of a factor in the analyses of the actual data, where 400 reassignments were conducted. However, even with 100 reassignments, the cumulative distribution for all sexes and species appears to closely approximate the line \( y = x \) for larger \( p \)-values (e.g., larger than 0.6), which is the range of \( p \)-values of principal interest in our estimation procedure.

Figure 2 displays the cumulative distributions of \( p \)-values obtained for the actual data. In contrast to the simulated distributions obtained for the null situation, these distributions show significant departures above the line \( y = x \), indicative of a carcinogenic effect in many experiments. The largest such departure occurs, as expected, in the graph for all sexes and species.

Figure 3 shows values of the estimator \( K(a) \) of the proportion of carcinogens, with corresponding 90% confidence intervals, as a function of the parameter \( a \). The horizontal line in these graphs represents the proportion of carcinogens based on the NTP classifications (Table 1). Recall that \( K(a) \) is negatively biased for all values of \( a \), and this bias decreases with increasing \( a \). However, the variance of \( K(a) \) increases with increasing \( a \), which accounts for the fact that confidence intervals in Figure 3 become progressively wider as \( a \) increases. These properties suggest emphasizing estimates associated with larger values of \( a \), but not so large that the confidence intervals become excessively wide. The estimates for male and female mice and male rats agree closely with the proportion of carcinogens obtained by the NTP over a range of values of \( a \). Although our estimates are above the NTP proportion over an extended range of \( a \) values for female rats and for all sexes and species, the 90% confidence intervals on our estimates generally incorporate the proportion of carcinogens identified by the NTP.
Figures 4 and 5 show, respectively, estimates of the proportions of anticarcinogens and the proportions of chemicals that were either carcinogenic or anticarcinogenic. Comparing Figures 3 and 4, except for female mice, there is evidence of an equal or greater proportion of anticarcinogens than carcinogens. For all sexes and species, estimates of the proportions of carcinogens plus estimates of the proportion of anticarcinogens exceeds 1 over a large range of values of $a$. This suggests that a number of chemicals were both carcinogenic and anticarcinogenic—causing dose-related increases in tumors in at least one site in one sex–species group and dose-related decreases at another site. Table 2 displays estimates of the proportion of chemicals that were carcinogenic, anticarcinogenic, or either based on the specific value $a = 0.75$. This value was selected as representative based on the trade-off of having $a$ large in order to reduce bias, but not so large that the variance was excessively large. Table 2 also shows the proportions of carcinogens identified by the NTP. Our estimates of the proportion of carcinogens in the different sex–species groups were similar, ranging from 0.28 in female mice to 0.35 in male rats. These estimates are also fairly similar to the NTP proportions. Our estimate of the proportion of chemicals that were carcinogenic in any sex–species group was 0.59, which was larger than the NTP proportion of 0.51. However, the NTP value was contained in our 90% confidence interval of 0.49–0.69.

Our estimates of the proportion of anticarcinogens in sex–species groups ranged from 0.27 to 0.44 (Table 2). Except for female mice, these estimates were larger than our corresponding estimates of the proportion of carcinogens. We estimated that 0.66 of the chemicals were anticarcinogenic in at least one sex–species group. Our estimates of the proportion of chemicals that were either carcinogenic or anticarcinogenic in sex–species groups ranged from 0.57 to 0.67. We estimated that 85% (95% confidence interval (CI), 78–91%) of chemicals were either carcinogenic or anticarcinogenic in at least one sex–species group.

**Discussion**

Using methods similar to those employed here, Crump et al. (10) estimated a considerably larger proportion of liver carcinogens in NTP studies (0.43; 90% CI, 0.35–0.51) than was identified by the NTP (0.28). We had expected to find similar evidence in the present study for an excess of chemicals carcinogenic at any site, but we did not. In retrospect, we believe this difference is due mainly to the manner in which the NTP evaluated studies. The NTP tended to evaluate the data for each tumor category individually in order to determine whether there was a treatment-related increase in tumors at that site. A chemical was labeled a carcinogen if it was deemed to be carcinogenic at any site. It is clearly more likely that a chemical will be labeled a carcinogen at some site using this procedure than at a particular site (e.g., liver). Expressed differently, the NTP did not use a formal method for correcting for multiple comparisons and consequently may have been prone to accept weaker statistical evidence for an effect at any tumor site than for an effect at a particular site. In contrast, our analysis did adjust appropriately for multiple comparisons, as evidenced by the uniform distribution of $p$-values in the

---

**Figure 3.** Estimates of the proportion of carcinogens in NTP bioassays. Thin lines indicate 90% confidence intervals. Black horizontal lines indicate NTP estimate.

**Figure 4.** Estimates of the proportion of anticarcinogens in NTP bioassays. Thin lines indicate 90% confidence intervals.
null situation (Fig. 1). We recommend that the NTP consider formally adjusting for multiple comparisons in their statistical analysis, just as we have done. Such an approach would make determinations of statistical significance more easily interpreted and could lead to more equitable classifications of carcinogenic responses across different sex–species groups.

Another factor that may have contributed to our finding a larger proportion of liver carcinogens than the NTP stemmed from the high frequency of dose-related increases among liver tumors and the occurrence of some of these tumors through mechanisms that were of questionable relevance to humans. To compensate, the NTP may have used more stringent criteria for evaluating liver tumors than tumors at other sites.

Although we did not find evidence for larger proportions of carcinogens than were identified by the NTP, there were differences between our findings and those of the NTP for individual chemicals (Table 1). Based on using a conventional significance level of 0.05 in our analysis, the overall concordance between our analysis and the NTP ranged from 0.91 for male mice to 0.85 for male rats. We were not able to improve this degree of concordance appreciably by using different significance levels for our analysis. There are a number of potential reasons for these differences. Whereas we used a single statistical test for trend in our analysis (the POLY3 test), the NTP used a variety of statistical tests, including pairwise comparisons. The NTP pooled control groups from several studies in their statistical evaluations of some of the earlier studies. We did not have complete records for identifying these pooled controls and consequently used matched controls in our analyses of some of these studies. The NTP sometimes took into account historical control data, but we did not. In some instances, the NTP took into account data that were not available in CBDS or TDMS. For example, in some equivocal situations, the NTP collected additional kidney pathology that was not added to CBDS or TDMS. In at least one study, an entire group of dosed animals that was critical to the NTP evaluation was not included in CBDS or TDMS. Whereas we evaluated sex–species groups in isolation, the NTP sometimes labeled an insignificant trend in a particular sex as clear evidence of carcinogenicity if the evidence at the same site in the opposite sex was unequivocal. Male rats had a high background rate of testicular tumors, and dose-related trends in these tumors generally were not taken into account by the NTP. (However, removing chemicals from our analysis for which the strongest effect was seen in these tumors did not appreciably improve the concordance between our analysis and that of the NTP.) It should be kept in mind that it was not the purpose of the present work to critique NTP classifications of individual chemicals, but rather to make independent estimates of the proportion of chemicals that were carcinogenic, anticarcinogenic, or either based on the information contained in the CBDS or TDMS databases. The data and procedures we used provide a statistically valid method for deriving such estimates.

In a similar study that estimated the number of liver carcinogens in the NTP database (10), the presence of anticarcinogenic effects caused the CDFs of p-values to fall below the graph y = x for values of p near 1.0. This effect was a source of additional negative bias in the (one-point) estimator, K(a). To minimize this effect, Crump et al. (10) used the "two-point estimator,"

\[ K(a, b) = \frac{[bF(a) - aF(b)]}{[b - a]}, \]

where a and b are two judiciously chosen points between 0 and 1.0. This estimator reduces to the one-point estimator, K(a), used in the present study when b = 1.0. This refinement was not needed in the present study because, except possibly for female mice, the CDFs do not exhibit an effect of anticarcinogenesis (Fig. 2). The reason for this was that in the present study the test statistic was the maximum of individual test statistics computed from each individual tumor type. The maximum test statistic over several tumor sites will be less affected by anticarcinogenic activity at one or a few sites than a test statistic for a single site will be affected by anticarcinogenesis at that site. Thus, in this respect, estimation of the proportion of experiments that showed carcinogenic activity at any site was simpler than estimating the corresponding proportion for a particular site.

In the previous study of liver carcinogens (10), p-value CDFs simulated for the null case exhibited irregularities due to the discreteness of the data. This was particularly a

![Figure 5. Estimates of the proportion of chemicals that were either carcinogenic or anticarcinogenic in NTP bioassays. Thin lines indicate 90% confidence intervals.](image)

### Table 2. Representative estimates* (with 90% confidence intervals) of the proportion of chemicals that were carcinogenic, anticarcinogenic, or either

| Carcinogenic | Anticarcinogenic | Either |
|--------------|-----------------|--------|
| Male mice    | 0.32 (0.19–0.44)| 0.29   |
| Female mice  | 0.28 (0.15–0.41)| 0.35   |
| Male rats    | 0.35 (0.23–0.47)| 0.36   |
| Female rats  | 0.34 (0.21–0.46)| 0.28   |
| All sexes and species | 0.35 (0.24–0.49)| 0.51   |

*Based on the estimator K(a), with a = 0.75.

Environmental Health Perspectives • Volume 107, Number 1, January 1999
problem for female rats, which exhibited the smallest spontaneous rates of liver tumors. Such irregularities are not apparent in the simulated null distributions in the present study (Fig. 1). This difference is due to the fact that in the present study the test statistic was defined as the maximum value obtained from analysis of a number of individual tumor categories. The average number of tumor categories analyzed ranged from 12 in male mouse CBDS studies to 30 in male rat TDMS studies. Defining the test statistic as the maximum number of test statistics gave a denser range of possible $p$-values, which resulted in a test statistic that was more uniformly distributed in the null case, than defining the test statistic in terms of only a single tumor category.

The present study found evidence for a larger proportion of anticarcinogens than carcinogens among NTP chemicals. An analysis that used a conventional significance level of 0.05 to detect effects would not have found this because the proportion of chemicals having a $p$-value < 0.05 for anticarcinogenesis was smaller than the corresponding proportion for carcinogenesis in every group except female rats. Nevertheless, the overall distribution of $p$-values provided evidence for a larger proportion of anticarcinogens. We estimated that 66% of NTP chemicals were anticarcinogenic in at least one site in at least one sex–species group. Finding a larger proportion of anticarcinogens than carcinogens was surprising to us because chemicals were selected for study by the NTP on the basis of suspected carcinogenicity and also because anticarcinogenesis should be inherently more difficult to detect than carcinogenesis in NTP bioassays. More than 90% of the background incidences of the tumor categories used in our analysis were less than 0.05. It would be difficult to detect dose-related decreases in tumor responses in a tumor category with a background rate as small as 0.05.

Haseman and Johnson (13) concluded that much of the anticarcinogenesis seen in NTP bioassays was indirectly caused by a dose-related reduction in weight gain. This mechanism may not be operative except at very high doses. However, this uncertainty is not limited to this mechanism or even anticarcinogenesis, as there is generally uncertainty regarding whether an effect seen in a high-dose NTP bioassay, either carcinogenic or anticarcinogenic, will occur at lower doses.

This study estimated that 85% of the chemicals studied by the NTP were either carcinogenic or anticarcinogenic at some site in some sex–species group of rodents. It should be kept in mind that the estimator used to obtain this estimate is inherently negatively biased. This suggests that most chemicals, when given at sufficiently high doses, may cause perturbations that affect tumor responses, causing increases at some sites and decreases at others. These effects will undoubtedly be smaller at lower doses, but identification of doses at which they do not occur will be highly problematic.

**References and Notes**

1. Huff J, Cirvillo J, Haseman J, Bucher J. Chemicals associated with site-specific neoplasia in 1394 long-term carcinogenesis experiments in laboratory rodents. Environ Health Perspect 52:247–270 (1991).
2. Haseman JK, Huff JE, Zeiger E, McConnell EE. Comparative results of 327 chemical carcinogenicity studies. Environ Health Perspect 74:229–235 (1987).
3. Fung VA, Barrett JC, Huff J. The carcinogenesis bioassay in perspective: application in identifying human cancer hazards. Environ Health Perspect 103:880–883 (1995).

4. National Research Council. Issues in Risk Assessment. Washington, DC:National Academy Press, 1993.
5. Ames BN, Gold LS. Chemical carcinogenesis: too many rodent carcinogens. Proc Natl Acad Sci USA 87:772–776 (1990).
6. Fung VA, Huff JE, Weisburger EW, Hoel DG. Predictive strategies for selecting 379 NCI/NTP chemicals evaluated for carcinogenic potential: scientific and public health impact. Fundam Appl Toxicol 20:413–436 (1993).
7. Huff JE. Issues and controversies surrounding qualitative strategies for identifying and forecasting cancer causing agents in the human environment. Pharmacol Toxicol 72(suppl 1):12–27 (1993).
8. Bicks M, Bleuer S, Krewski D. Estimation of the proportion of positives in a sequence of screening experiments. Can J Stat 24:1–16 (1996).
9. Bicks M, Krewski D. Statistical issues in the analysis of the long-term carcinogenicity bioassay in small rodents: an empirical evaluation of statistical decision rules. Fundam Appl Toxicol 12:202–221 (1989).
10. Crump K, Krewski D, Wang Y. Estimates of the number of liver carcinogens in bioassays conducted by the National Toxicology Program. Risk Anal 18:299–308 (1998).
11. Crump K, Krewski D. Estimation of the number of studies with positive trends when studies with negative trends are present. Can J Stat (1998).
12. Haseman J. A re-examination of false-positive rates for carcinogenicity bioassays. Fundam Appl Toxicol 3:334–339 (1983).
13. Haseman J, Johnson F. Analysis of National Toxicology Program rodent bioassay data for anticarcinogenic effects. Mutat Res 350:131–141 (1995).
14. Haseman JK, Ewell M. Evaluation of false positive and false negative outcomes in NTP long-term rodent carcinogenicity studies. Risk Anal 16:813–820 (1996).
15. Bailar A, Portier C. Effects of treatment-induced mortality and tumor-induced mortality on tests for cancer in small samples. Biometrics 44:417–431 (1988).
16. Dinse G. A comparison of tumor incidence analyses applicable in single-sacrifice animal experiments. Stat Med 13:899–706 (1994).
17. NTP. Toxicology and Carcinogenesis Studies of Diethanolamine (CAS No. 111-42-2) in F344/N Rats and B6C3F1 Mice (Dermal Studies)—Draft Technical Report 478. Research Triangle Park, NC:National Toxicology Program, 1996.
18. Farrar D, Crump K. Exact statistical tests for any carcinogenic effect in animal bioassays. Fundam Appl Toxicol 11:592–663 (1988).
19. Farrar D, Crump K. Exact statistical tests for any carcinogenic effect in animal bioassays. II. Age-adjusted tests. Fundam Appl Toxicol 15:710–721 (1990).