Available online at www.sciencedirect.com
SCIENCE DIRECT®
Regulatory Toxicology and Pharmacology 41 (2005) 128–133
Regulatory Toxicology and Pharmacology
www.elsevier.com/locate/yrtph

# Are tumor incidence rates from chronic bioassays telling us what we need to know about carcinogens?

David W. Gaylor*

*Gaylor and Associates, LLC, Eureka Springs, AR 72631, USA*

Received 22 September 2004
Available online 19 December 2004

**Abstract**

Chronic bioassays for over 500 chemicals have been conducted under the auspices of the National Cancer Institute and/or the National Toxicology Program (NTP) to screen chemicals for carcinogenicity, providing a wealth of information about bioassays. Typically, chemicals are administered for two years to both sexes in each of one strain of rats and mice generally at the maximum tolerated dose (MTD), MTD/2, MTD/4 (in recent years), as well as unexposed control animals. In an attempt to ascertain the sensitivity of this bioassay to detect animal carcinogens tested at the MTD for the current experimental design, the false negative rate (failure to detect increased tumor rates) was investigated. This was accomplished by examining the tumor incidences from over 150 NTP bioassays and estimating the probability that a statistically significant ($P \leqslant 0.01$) dose–response trend would be obtained at one or more tissue sites in either sex of rats or mice if 200, rather than 50, animals were used per dose group. This provides an estimate of the proportion of chemicals that were not declared high-dose animal carcinogens due to the limited sample size of 50 animals per species–sex–dose group. In this series of chemicals tested, 97/156 (62%) were identified by the NTP to show some or clear evidence of carcinogenicity. With an increase of the number of animals per dose group from 50 to 200, it is estimated that 92% of these chemicals would show statistically significant ($P \leqslant 0.01$) dose–response trends at one or more tissue sites in either sex of rats or mice. Many of these chemicals are not genotoxic. Some chemicals had no structural alerts for carcinogenicity, but were tested because of potentially high human exposure. This analysis suggests that almost all of the chemicals selected would produce a statistically significant increase in tumor incidence at the MTD with larger sample sizes. Hence, this MTD bioassay screen is not distinguishing between true carcinogens and non-carcinogens. Rather, the screen is simply failing to detect the weaker carcinogens at the MTD. More than 30% of chemicals tested failed to detect statistically significant dose–response trends for tumors because of inadequate sample sizes of 50 animals per dose. Presumably, little or no action would have been taken to regulate exposures to these chemicals as potential carcinogens due to lack of a positive test result. This analysis does not suggest that most chemicals are carcinogenic at human exposure levels nor does it suggest that more than 50 animals should be tested per dose group. With an MTD that may produce a difference (up to 10%) in weight gain between treated and control animals, there quite possibly is cytotoxicity at the MTD. Increased carcinogenicity would be expected from increased opportunities for mutagenic activity during regenerative cell replication to compensate for cytotoxicity. Since it appears that almost all chemicals tested adequately at the MTD will demonstrate carcinogenicity, it is tempting to surmise that this is due in large part to one or more nearly universal modes of action, such as, regenerative cell replication at the MTD rather than due to some unique carcinogenic property of a chemical. That is, the current bioassay possibly is just primarily a screen for the more potent cytotoxins at the MTD, rather than a screen specifically for carcinogenicity. This issue should be examined and suggests that cytotoxicity and cell proliferation should be considered in setting the MTD, particularly for non-genotoxic (non-DNA reactive) chemicals. From a public health view, it is prudent to assume that most chemicals could demonstrate increased tumor incidence rates at the MTD in rodents. The current standard NTP bioassay provides sufficient data to estimate a benchmark dose associated with a specified low tumor incidence to be used as a point-of-departure for cancer risk assessments. The question that should be investigated by a bioassay is not whether a chemical is a carcinogen at the MTD, but what is the relationship between dose and cytotoxicity and/or other modes of action that could produce an excess of tumors?

* Fax: +1 479 253 1092.
*E-mail address:* DavidGaylor@earthlink.net.

0273-2300/$ - see front matter 
doi:10.1016/j.yrtph.2004.11.001

*Keywords:* Maximum tolerated dose; Cytotoxicity; Regenerative cell replication; Benchmark dose; Cancer risk assessment; Bioassay

## 1. Introduction

One purpose of US National Toxicology Program (NTP) chronic bioassay testing program is to evaluate chemicals for carcinogenic activity. Chemicals are generally administered for two years following weaning in both sexes of rats and mice. Usually 50 animals per group are tested at the maximum tolerated dose (MTD), MTD/2, MTD/4 (in recent years), and unexposed control animals. This test was devised to have reasonable power to detect moderate increases in tumor incidence at high doses without using an inordinately large number of test animals. Over 500 chemicals have now been tested under the auspices of the NTP, providing a wealth of information on chronic bioassays.

Because of the large number of tissue sites examined, some with background tumor incidence rates high enough to produce spurious higher rates in dosed animals due to inherent chance experimental variation, there is concern about the false positive rate for chemicals tested in the standard NTP bioassay (Fears et al., 1977). Haseman (1983) indicated that the NTP tended to use a statistical significance level of $P \leqslant 0.05$ for rare tumors (historical background rates of less than 1–2%) and required $P \leqslant 0.01$ for more common tumors. Using these rules, Haseman (1983) calculated the false positive rate for carcinogenicity to be about 8% for the NTP bioassay. This appears to be a reasonable level.

Haseman (1983) acknowledges that although his paper deals primarily with false positives, examination of the statistical properties of the standard bioassay would be incomplete without consideration of the false negative rate. Haseman (1984) provides a table that shows the false negative rate to be 30% for a chemical that induces a rare tumor in 13% of animals at the high dose. For a chemical that doubles the tumor incidence for a common tumor from 30% in the controls to 60% at the high dose, the false negative rate is also 30%. The false negative rate would be even higher for chemical carcinogens producing fewer tumors. Fears et al. (1977) show that 5-fold increases in rare tumors are not likely to be detected. Crump et al. (1996) indicate that there are more liver carcinogens than are detected by the NTP bioassay. In an overview of NTP results, Huff and Haseman (1991) state that: "unfortunately, little is known about the false negative rate, which is of more importance to public health." Haseman and Elwell (1996) state that it is difficult to assess false negative rates because of the limited sensitivity of the NTP bioassay to detect subtle carcinogenic effects.

The primary purpose of this paper is to estimate the statistical false negative rate for carcinogenicity based on a retrospective analysis of tumor incidence rates observed in a series of standard NTP 2-year chronic rodent bioassays and an evaluation of the effect of using a sample size of 50 animals per group at the MTD. From the wealth of information provided by the NTP tests, the following questions can be addressed: (1) should the sample size be changed, (2) should the MTD be re-defined, and (3) are chronic bioassays similar to the NTP test that are being conducted around the world by government agencies, private industry, and academic institutions telling us what we need to know about chemical carcinogenicity? Several authors have suggested that testing chemicals at high doses produces cytotoxicity that results in regenerative cell replication that provides additional opportunities for cancerous mutagenic events to occur (e.g., Ames and Gold, 1990a,b; Clayson et al., 1989; Cohen and Ellwein, 1990; Conolly and Lutz, 2004; Counts and Goodman, 1995; Lutz, 1998; Schulte-Hermann et al., 1991).

## 2. Methods

A false negative decision about the carcinogenicity of a chemical occurs when the bioassay fails to produce a statistically significant increased tumor incidence when in fact the chemical truly causes an increase in the tumor incidence at the doses tested. This is a statistical limitation resulting from the number of animals (generally 50) used per species–sex–dose group.

Using the estimate of the dose–response trend obtained from past NTP studies for each specified tumor type/tissue site in males or females of rats or mice and the standard error of the trend, it is possible to estimate the approximate probability (power) of detecting a statistically significant trend as a function of the sample size. For example, the Cochran–Armitage dose–response trend test (Armitage, 1971) is of the form

$$Z = \text{slope}/(\text{standard error of the slope}),$$

where $Z$ is a standard normal deviate with a standard deviation of one. Suppose the test for a dose–response trend produced a one-sided $P$ value of 0.09, i.e., $Z = 1.34$, with 50 animals per dose group. Based strictly on the statistical test, this would be considered lack of evidence for carcinogenicity, i.e., categorized as non-carcinogenic. However, if more animals had been used per dose group, the standard error of the slope would have been

reduced by a factor proportional to the square root of the sample size. If 200 animals had been used per dose group, the standard error of the dose–response slope would have been reduced by a factor of two giving an expected value of $Z = (2 \times 1.34) = 2.68$, which indicates a statistically significant trend with $P < 0.01$. Now, the chemical likely would be categorized as carcinogenic, not because its carcinogenic potential changed, but because the sample size changed. Because of inherent experimental variation, the value of $Z$ would not necessarily double with a quadrupling of the sample size. For the above example, the probability is 63.7% that the approximately normal trend test with an expected value of $Z = 2.68$ exceeds the $P \leqslant 0.01$ critical value of $Z = 2.33$.

Hence, for each existing statistical trend test from the published bioassay results it is possible to estimate the probability that a statistically significant trend would be detected with a larger sample size. Assuming independence of tests, which certainly is the case for dose–response trends occurring in different species and/or sexes for a chemical, the power $\{1 - (\text{false negative rate})\}$ of statistical tests to detect one or more dose–response trends can be estimated as a function of sample size for each chemical tested. For example, suppose for a chemical there are two tissue sites that could produce a statistically significant dose–response trend where the estimated probability of detecting a statistically significant trend is 0.8 for one tissue site in rat males and 0.6 for one tissue site in female mice with 200 animals per dose group and relatively small probabilities of detecting a significant trend at other sites. The probability of not detecting a significant trend for both sites is $(1 - 0.8) \times (1 - 0.6) = 0.08$. The probability of detecting a statistically significant trend for at least one site for the chemical is at least $\{1 - (1 - 0.8)(1 - 0.6)\} = 0.92$ with 200 animals per group. Thus, based on the results from testing a chemical with 50 animals per group, the probability of at least one tissue site with a statistically dose–response trend can be estimated for a chemical with a specified larger sample size, e.g., 200 animals per group. Haseman and Lockhart (1993) indicate relatively few correlations between chemically related site-specific carcinogenic effects.

Similar to NTP procedures, testicular tumors in male rats are not considered. In the analysis in this paper, preneoplastic effects were not used to provide additional support for carcinogenicity. Obviously, it is not possible to know what decisions would have been reached on a case-by-case basis by NTP reviews. With 200 animals per group, even rare tumors with a spontaneous 1% tumor incidence rate could occur with a frequency adequate to occasionally lead to false positive test results. To protect against a high false positive rate with 200 animals per group, statistical significance of $P \leqslant 0.01$ was required for all tissue sites in the following analyses.

Early NTP Technical Reports did not contain enough detail to estimate the power of statistical tests. The power analysis was conducted in this paper for NTP chemicals published in Technical Reports No. 250–450. Only chemicals tested in both sexes of rats and mice were examined resulting in 156 cases. Two trend tests were routinely conducted by the NTP, one assuming all tumors in dead and moribund animals were fatal and the other assuming all tumors were non-fatal (incidental). If there were no indication of the more appropriate test, to remain conservative the least significant test was used here.

## 3. Results

In the series of NTP Technical Reports No. 250–450, a total of 156 chemicals were tested in both sexes of rats and mice. In these bioassays, generally 50 animals were used per dose group. Among these chemicals, the NTP determined that 97/156 (62%) were positive (some or clear evidence of carcinogenicity). This may reflect, in part, the ability of the chemical selection committee to choose mostly carcinogens for testing based on genotoxicity and/or structure activity relationships. However, many chemicals are selected for testing because of potentially high human exposures without any suspected carcinogenic structural alerts. Many chemicals without suspected potential for carcinogenicity were animal carcinogens at high doses.

Because of inherent experimental variation not all of the chemicals deemed carcinogenic in the standard 50 animal per group bioassay would necessarily produce a statistically significant dose–response trend at the $P \leqslant 0.01$ level if 200 animals per dose group had been used. The estimated expected number of these 97 chemicals that would produce statistically significant $P \leqslant 0.01$ dose–response trends at one or more tissue sites is at least 95.1. Similarly, of the 23 chemicals tested with 50 animals per dose group that the NTP indicated provided only equivocal evidence of carcinogenicity, it was estimated that at least 22.5 of these chemicals would produce statistically significant ($P \leqslant 0.01$) dose–response trends with 200 animals per dose group at one or more tissue sites. It is not surprising that increasing the sample size would increase the evidence for carcinogenicity from equivocal to positive for most of these chemicals. Of the 36/156 (23%) chemicals considered by the NTP as not showing any evidence of carcinogenicity, with 200 animals per dose group it is estimated that at least 26.0 of these chemicals would produce statistically significant ($P \leqslant 0.01$) dose–response trends at one or more tissue sites. The above results are summarized in Table 1. Overall, it is estimated that at least 92% of all chemicals tested at the MTD would produce a statistically significant ($P \leqslant 0.01$) dose–response trend at one or more tissue sites.

Table 1
Estimated number of positive chemicals expected to produce statistically significant ($P \leqslant 0.01$) dose–response trends at one or more tissue sites with 200 animals per dose group

| Original NTP carcinogenic category with 50 animals per dose group | Estimated number of chemicals based on 200 animals per dose group | | Total |
|---|---|---|---|
| | Negative | Positive[a] | |
| Negative | 10.0 | 26.0 | 36 (23%) |
| Equivocal | 0.5 | 22.5 | 23 (15%) |
| Positive (some or clear evidence) | 1.9 | 95.1 | 97 (62%) |
| Total | 12.4 (8%) | 143.6 (92%) | 156 |

[a] Expected number likely to be at least this large.

## 4. Example

To illustrate how the expected numbers of chemicals listed in Table 1 for statistically significant ($P \leqslant 0.01$) dose–response trends with 200 animals were compiled across chemicals, one case (Rotenone) is illustrated. The results of the standard bioassay are presented in the NTP Technical Report Series No. 320, Toxicology and Carcinogenesis Studies of Rotenone in F344/N Rats and B6C3F$_1$ Mice; January, 1988. The evidence for carcinogenicity of rotenone was considered equivocal by the NTP. For follicular cell adenoma or carcinoma of the thyroid gland in male rats a dose–response trend was reported with a non-significant $P = 0.139$. This corresponds to a standard normal $Z$-score of 1.085. A quadrupling of the sample size to 200 animals per dose group would be expected to increase the $Z$-score by a factor of $\sqrt{4} = 2$ to 2.17. The probability that a $Z$-score with an expected value of 2.17 exceeds 2.33 required for statistical significance of $P \leqslant 0.01$ is 0.436. Similarly, in male rats for parathyroid adenoma the dose–response trend test with 50 animals per dose group had a $P = 0.119$ or $Z = 1.18$. With 200 animals per dose group a $Z$-score of $(2 \times 1.18) = 2.36$ would be expected with a probability of 0.512 for achieving statistical significance ($P \leqslant 0.01$) with a $Z$-score exceeding 2.33. In female rats, sarcoma, fibrosarcoma, or myxosarcoma of subcutaneous tissue produced a dose–response trend test with $P = 0.121$ corresponding to a $Z = 1.17$. With 200 animals per dose group the expected $Z = 2.34$ has a probability of 0.504 of producing a statistically significant ($P \leqslant 0.01$) dose–response trend. No other tissue sites in either sex of rats or mice showed marginal evidence of a dose–response trend. Thus, for rotenone three sites showed equivocal evidence of carcinogenicity, but none approached statistical significance ($P \leqslant 0.05$) required for rare tumors with 50 animals per dose group. The probability that at least one of these three tumor types would produce a statistically significant ($P \leqslant 0.01$) dose–response trend is $[1 - (1 - 0.436)(1 - 0.512)(1 - 0.504)] = 0.863$. That is, there is at least an 86.3% chance that rotenone would have produced at least one statistically significant ($P \leqslant 0.01$) dose–response trend with 200 animals per dose group. The probability could be slightly higher since there were other tissue sites that could produce a dose–response trend, but were not included because of their relatively low probabilities. Thus, rotenone contributes 0.863 to the expected number of positive chemicals with 200 animals per dose group that were classified as equivocal in the original test with 50 animals per dose group. A similar analysis was conducted for each of the 156 chemicals to compile the results listed in Table 1.

## 5. Discussion

In an attempt to ascertain the false negative rate (failure to detect carcinogenesis) for the current NTP experimental protocol, based on the tumor incidence data compiled by the NTP for 156 chemicals tested in both rats and mice with 50 animals per dose group, the probability was estimated of obtaining one or more statistically significant ($P \leqslant 0.01$) dose–response trends by increasing the number of animals to 200 per dose group for these 156 chemicals. Sixty-two percent of these chemicals were declared carcinogens by the NTP with 50 animals per dose group. With 200 animals per dose group, 92% of these chemicals are predicted to produce a statistically significant ($P \leqslant 0.01$) dose–response trend at one or more tissue sites. This provides an estimate that at least 30% of the chemicals tested with the standard protocol failed to detect carcinogenesis due to the limited number (50) of animals used per dose group. Clearly, an even higher proportion of chemicals would produce statistically significant dose–response trends when tested near the MTD as the sample size increases. This supports the contention of Ashby and Purchase (1993) that most chemicals would be carcinogenic if typical bioassays were conducted in enough strains of animals.

This does not imply that the NTP miscalculated the carcinogenicity of these chemicals under the conditions of the test protocol. Rather, this analysis demonstrates the lack of the power of the standard bioassay with 50 animals per dose group to detect carcinogens at the MTD. It is disturbing that more than 30% of chemicals tested with the NTP protocol fail to detect carcinogenicity. High false negative rates can be expected in similar chronic bioassays conducted around the world by government agencies, private industry, and academic institutions. Unfortunately, concern for carcinogenicity would be minimal or not be considered in risk assessments of non-genotoxic carcinogens that fail to unequivocally produce carcinogenicity in bioassays. Presumably, concern for genotoxic chemicals might remain even if carcinogenicity were not demonstrated in a bioassay.

If tested in an adequate number of animals, it appears that almost all chemicals would demonstrate carcinogenicity at high doses. An artificial category of carcinogenic or non-carcinogenic for chemicals that is based on an inadequate sample size of 50 animals per dose group is misleading. Since the standard bioassay cannot adequately distinguish between carcinogens and non-carcinogens tested at the MTD, it would appear prudent from a public health standpoint to assume that all chemicals may be carcinogenic at the MTD in animals. Since it appears that almost all chemicals are carcinogenic at high doses in rodents, this negates the value and need for a chemical screening test at the MTD. It is not being recommended here that sample sizes should be increased to verify the apparent outcome that almost all chemicals are carcinogenic at the MTD. The important question is not: is a chemical a carcinogen? Rather, the question that needs to be answered is: at what dose does a chemical cause cancer? The important issue to be addressed by a bioassay is not a statistical test for carcinogenicity. Rather, the important issue to be addressed with bioassay data is estimation of the relationship between dose and tumor incidence or precursors to cancer.

Presumably, for genotoxic chemicals that are DNA-reactive and depending on the pharmacokinetics, extremely low doses may cause some cancer. Since it appears that almost all chemicals cause cancer at the MTD in rodents, including non-genotoxic chemicals, one hypothesis is that at high doses cytotoxicity results in regenerative cell replication that provides additional opportunities for cancerous mutagenic events to occur (Ames and Gold, 1990a,b; Clayson et al., 1989; Cohen and Ellwein, 1990; Conolly and Lutz, 2004; Counts and Goodman, 1995; Lutz, 1998; Schulte-Hermann et al., 1991). Should the MTD be based on cytotoxicity? Should the bioassay be designed to examine the relationship between dose and cytotoxicity? These are questions that should be addressed that are beyond the scope of this paper.

Clearly, assuming that most chemicals are carcinogenic at the MTD does not imply that most chemicals cause cancer at much lower human exposure levels. Various indices of carcinogenic potency have been based upon the estimated low-dose tumor incidence slopes or using linear extrapolation from higher experimental doses estimated to produce a specified excess tumor incidence (e.g., $TD_{50}$, $TD_{25}$, $TD_{10}$, or $TD_{01}$). For ranking potential carcinogenic hazards, Ames et al. (1987) developed the HERP index, which is the ratio of the expected human exposure level relative to the $TD_{50}$ (dose that induces cancer in 50% of otherwise tumor-free animals in a standard lifetime). This appears to be useful for establishing levels of concern and setting priorities for further investigations and research. A relatively quick estimate of the $TD_{50}$ can be obtained from the MTD established in a 90-day study (Gaylor and Gold, 1998). This can provide a basis for a preliminary cancer risk assessment of a chemical that has not been studied in a 2-year bioassay. Ashby and Tennant (1991) accurately predicted carcinogenicity for about 80% of 301 NTP studies based on structural alerts and the *Salmonella* test.

There appears little need to test chemicals in 2-year bioassays to establish carcinogenicity at or near the MTD. However, bioassays are needed to provide a dose–response relationship between dose and carcinogenic events to perform cancer risk assessments. Current EPA (1999) carcinogen risk assessment guidelines suggest a benchmark dose with a low excess tumor rate (1–10%) as a point-of-departure for cancer risk assessment. As a default value, the lower 95% confidence limit on an extra tumor incidence of 10% ($BMDL_{10}$) is used as a point-of-departure. For genotoxic chemicals, linear extrapolation from the $BMDL_{10}$ to zero generally is used for low-dose cancer risk assessments. For chemicals with established nonlinear dose–response curves in the low dose range, the $BMDL_{10}$ is divided by uncertainty factors to arrive at a reference dose presumed to have negligible, if any, carcinogenic activity. This process requires tumor incidence dose–response data to estimate the $BMDL_{10}$. Tumor incidence dose–response relationships can be established from the data provided by the standard NTP bioassays. This is a necessary and valuable contribution of the NTP bioassays. The current NTP bioassays with three doses (generally the MTD, MTD/2, and MTD/4) plus controls with 50 animals per dose group in both sexes of rats and/or mice provide sufficient data for calculating benchmark doses. In the absence of even a weakly significant dose–response curve an estimate of the $BMDL_{10}$ may not be obtainable. In such cases, the MTD could serve as a conservative point-of-departure for cancer risk assessments. For non-DNA reactive chemicals, perhaps the MTD should be based upon cytotoxicity or some other effect depending on the mode of carcinogenic action. Also, pharmacokinetics need to be considered in establishing benchmark doses.

## 6. Conclusions

The current NTP bioassay in rats and mice with doses at or near the MTD with 50 animals per dose group is not adequate to distinguish between rodent carcinogens and non-carcinogens. Over 92% of chemicals tested would be expected to produce statistically significant ($P \leqslant 0.01$) dose–response trends for tumor incidence at one or more tissue sites if 200 animals per dose group had been used. Many of these chemicals are not genotoxic. It is surmised that cytotoxicity at high doses near the MTD could result in regenerative cell replication that provides an increased opportunity for carcinogenic mutations to occur (Ames and Gold, 1990a,b; Clayson et al., 1989; Cohen and Ellwein, 1990; Conolly and Lutz,

2004; Counts and Goodman, 1995; Lutz, 1998; Schulte-Hermann et al., 1991). Presumably, this mode of action would not be present, or at most negligible, at low chemical doses. Hence, the current bioassay possibly is primarily a screen for the more potent cytotoxins at the MTD in rodents, rather than a screen specifically for carcinogenicity. This issue should be examined and suggests that cytotoxicity and cell proliferation should be considered in setting the MTD, particularly for non-genotoxic (non-DNA reactive) chemicals.

Since most chemicals appear to be carcinogenic in rodents at the MTD with adequate sample sizes, there appears to be little need to screen chemicals in 2-year chronic bioassays for carcinogenicity at the MTD. However, bioassays are needed to provide a dose–response relationship between dose and carcinogenic effects for use in cancer risk assessments. Current EPA (1999) carcinogen risk assessment guidelines utilize a benchmark dose associated with a low excess tumor incidence rate (1–10%) as a point-of-departure for cancer risk assessment. The standard NTP bioassay with three doses (generally the MTD, MTD/2, and MTD/4) plus controls with 50 animals per dose group in both sexes of rats and/or mice provides sufficient data for calculating benchmark doses. For this critical information, it is not recommended that the number of doses or the number of animals per dose should be changed for the NTP bioassay.

## References


Ames, B.N., Gold, L.S., 1990a. Too many rodent carcinogens: mitogenesis increases mutagenesis. Science 249, 970–971.

Ames, B.N., Gold, L.S., 1990b. Chemical carcinogenesis: too many rodent carcinogens. Proc. Natl. Acad. Sci. 87, 7772–7776.

Ames, B.N., Magaw, R., Gold, L.S., 1987. Ranking possible carcinogenic hazards. Science 236, 271–280.

Armitage, P., 1971. Statistical Methods in Medical Research, second ed. John Wiley and Sons, New York, pp. 363–365.

Ashby, J., Purchase, I.F.H., 1993. Will all chemicals be carcinogens to rodents when adequately evaluated? Mutagenesis 8, 489–495.

Ashby, J., Tennant, R.W., 1991. Definitive relationships among chemical structure, carcinogenicity, and mutagenicity for 301 chemicals tested by the U.S. NTP. Mutat. Res. 257, 229–306.

Clayson, D.B., Nera, E.A., Lok, E., 1989. The potential for the use of cell proliferation studies in carcinogen risk assessment. Regul. Toxicol. Pharmacol. 9, 284–295.

Cohen, S.M., Ellwein, L.B., 1990. Cell proliferation in carcinogenesis. Science 249, 1007–1011.

Conolly, R.B., Lutz, W.K., 2004. Nonmonotonic dose–response relationships: mechanistic basis, kinetic modeling, and implications for risk assessment. Toxicol. Sci. 77, 151–157.

Counts, J.L., Goodman, J.I., 1995. Principles underlying dose selection for, and extrapolation from, the carcinogen bioassay: dose influences mechanism. Regul. Toxicol. Pharmacol. 21, 418–421.

Crump, K., Haseman, J., Krewski, D., Wang, Y., 1996. Estimates of the number of liver carcinogens in bioassays conducted by the National Toxicology Program. The Toxicologist 30: No. 1, Part 2 (Abstract No. 1041).

Fears, T.R., Tarone, R.E., Chu, K.C., 1977. False positive and false negative rates for carcinogenicity screens. Cancer Res. 37, 1941–1945.

Gaylor, D.W., Gold, L.S., 1998. Regulatory cancer risk assessment based on a quick estimate of a benchmark dose derived from the maximum tolerated dose. Regul. Toxicol. Pharmacol. 28, 222–225.

Haseman, J.K., 1983. A re-examination of false-positive rates for carcinogenesis studies. Fund. Appl. Toxicol. 3, 334–339.

Haseman, J.K., 1984. Statistical issues in the design, analysis, and interpretation of animal carcinogenicity studies. Environ. Health Perspect. 58, 385–392.

Haseman, J.K., Elwell, M.R., 1996. Evaluation of false positive and false negative outcomes in NTP long-term rodent carcinogenicity studies. Risk Anal. 16, 813–820.

Haseman, J.K., Lockhart, A.M., 1993. Correlations between chemically related site-specific carcinogenic effects in long-term studies in rats and mice. Environ. Health Perspect. 101, 50–54.

Huff, J., Haseman, J., 1991. Long-term chemical carcinogenesis experiments for identifying potential human cancer hazards: collective database of the National Cancer Institute and National Toxicology Program (1976–1991). Environ. Health Perspect. 96, 23–31.

Lutz, W.K., 1998. Dose–response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear–sublinear curves, practical thresholds, and J-shapes. Mutat. Res. 405, 117–124.

Schulte-Hermann, R., Bursch, W., Parzefall, W., 1991. Mitogenesis and programmed cell death as determinants of carcinogenicity of nongenotoxic compounds. In: Butterworth, B.E., Slaga, T.J., Farland, W., McClain, M. (Eds.), Chemically Induced Cell Proliferation: Implications for Risk Assessment. Wiley-Liss, New York, pp. 237–244.

US Environmental Protection Agency. 1999. Guidelines for Carcinogen Risk Assessment. NCEA-F-0644, Risk Assessment Forum. Washington, DC.