Available online at www.sciencedirect.com
SCIENCE DIRECT®
Regulatory Toxicology and Pharmacology 38 (2003) 52–57
Regulatory Toxicology and Pharmacology
www.elsevier.com/locate/yrtph

Implications of the lack of accuracy of the lifetime rodent bioassay for predicting human carcinogenicity

Fanny K. Ennever and Lester B. Lave*

Graduate School of Industrial Administration, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Received 30 December 2002

Abstract

The NTP lifetime rodent bioassay (LRB) is the "gold standard" for predicting human carcinogenicity. Unfortunately, little attempt has been made to validate it against human carcinogenicity. Here we show that the extremely limited data available do not support either of the two common interpretations of LRB results. If a risk-avoidance interpretation is used where any positive result in a sex/species combination is considered positive, 9 of the 10 known human carcinogens tested are positive, but an implausible 22% of all chemicals are positive. If a less risk averse interpretation is used where only chemicals positive in both rats and mice are considered positive, only 3 of the 6 known human carcinogens tested are positive. In either interpretation, some known human carcinogens are not positive in the LRB, potentially allowing widespread human exposure to misidentified chemicals. Improving the predictive accuracy of the LRB and other tests for human carcinogenicity requires that test results be validated against the known human carcinogenicity of chemicals. This will require redirecting available resources from screening chemicals to validating carcinogenicity tests as well as a substantial investment in epidemiology to identify more known human carcinogens and presumed human non-carcinogens.

1. Introduction

Cancer is a dread disease; one in four Americans will develop cancer in his or her lifetime. Lifestyle changes, particularly quitting smoking, have the greatest potential to reduce premature deaths from cancer. However, prevention efforts have also concentrated on identifying chemicals that can cause cancer, even though chemical exposures probably contribute at most a few percent of the total cancer occurrence (Gold et al., 2002). Exposure to synthetic chemicals is uncontrollable, involuntary, inequitable, unfamiliar, not observable, and scientifically controversial, all factors known to increase public outrage (Slovic, 1999).

If society cared only about detectable cancer risks, prevention would focus on the 87 agents, viruses, mixtures or circumstances identified as known human carcinogens by the International Agency for Research on Cancer (IARC, 2002). However, epidemiology is limited in its ability to detect effects, and can only find disease resulting from past, usually poorly measured, exposures (Huff, 1999). Society wants to reduce exposures both to existing and new carcinogens, even when exposures are too low to be detected by epidemiology.

The experimental approach to detecting carcinogens has been to expose animals, mainly rats and mice, to a chemical to determine whether tumors develop (Rall, 2000; Roe, 1998). The past two decades have seen vigorous debate over the utility of the tests (Davies et al., 2000; Gold et al., 2002; Gori, 2001; Haseman et al., 2001; Johnson, 2002; Tomatis, 2002, and references cited therein). The result has been a bifurcation in perception. Toxicologists are generally skeptical about the biological and statistical relevance of the rodent bioassay to human cancer risk (Davies and Monroe, 1995; Gold et al., 2002; Gori, 2001; Haseman, 2000; Miller and Davis, 2001) "In the face of these shortcomings, many experts believe the scientific value of the 2-year bioassay is highly limited—barely worth the investments in personnel, animals, money, and time"

* Corresponding author. Fax: 1-412-268-7357.
E-mail address: lave@andrew.cmu.edu (L.B. Lave).

0273-2300/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0273-2300(03)00068-0

(Schmidt, 2002) At best, the LRB should be used out of prudence for lack of a better alternative (Huff, 1999; Rall, 2000; Ruden, 2002; Tomatis, 2002; VanDoren, 1996).

The general public, however, tends to consider the animal tests to be highly reliable (Slovic et al., 1997). A recent formulation acknowledges that the rodent bioassay does give many false positives (chemicals that cause cancer in rodents but not so in humans). However, it asserts that these false positives may be identified by consideration of mode of action (Ettlin and Prentice, 2002) and that the high sensitivity at least gives confidence in a negative result (Barlow et al., 2002; Rall, 2000). A few false positives have been identified by probable mode of action (α-2-μglobulin) but most modes of action are unknown. False negatives also occur (Johnson, 2001; Lave et al., 1988), which identify carcinogenic chemicals as benign, leading to human exposure and cancers. The only way to eliminate false negatives would be to classify all chemicals as carcinogenic, depriving society of many valuable pharmaceuticals and other chemicals. In deciding how to interpret a positive or negative outcome from a LRB, the decision maker needs to know the accuracy of the test (Johnson, 2001).

2. Calculations of the accuracy of the rodent bioassay

Defenders of the rodent bioassay, including advocacy groups, offer two principal justifications: (1) all known human carcinogens cause cancer in rodents and (2) only 5–10% of all chemicals are rodent carcinogens (CCHE, 2002; Rall, 2000; Tennant et al., 2001). Unfortunately, these two justifications interpret "cause cancer in rodents" differently. A LRB produces myriad data with different interpretations (Ashby, 2001; Byrd, 1988; Crump et al., 1999; Haseman et al., 1996; Johnson, 2000; Tennant et al., 2001). Most discussions either consider the result positive (i.e., capable of causing cancer in humans) if any rodent group tested is a positive or insist that only trans-species carcinogens are positive (Davies and Monroe, 1995; Fung et al., 1995; Johnson, 2002; Kodell et al., 1999). In the NTP data, 22% of chemicals are positive if any positive is a positive, and 6.8% are positive if only trans-species chemicals are positive (Fung et al., 1995). Not all chemicals were tested in both rats and mice, which lowers the positive rate. Whatever its limitations, this dataset is representative of how chemicals are actually tested.

Johnson (2001) lists only 10 known human carcinogens among the hundreds of chemicals tested by the NTP. Only three of the six tested in both species caused cancer in both; one caused cancer in neither (see Table 1).

Table 1
Data from Johnson (2001) on the results of the NTP LRB for the 10 known human carcinogens tested

Known human carcinogen	Carcinogenicity in rats	Carcinogenicity in mice
Thiotepa	+	+
Benzene	+	+
Benzidine and dyes	+	+
1,3-Butadiene	Not tested	+
Ethylene oxide	Not tested	+
8-Methoxypsoralin	+	Not tested
Nickel compounds	+	–
Asbestos	+	Not tested
Talc	+	–
Aspirin/phenacetin/ caffeine (APC)	–	–

Many known human carcinogens have been tested in non-NTP lifetime rodent bioassays. (Davies et al., 2000; IARC, 2002). Of those tested adequately, all were positive in at least one test, with the possible exception of arsenic (Tennant et al., 2001). This result says more about the persistence of toxicologists than about the ability of a standard protocol to predict human carcinogenicity. Bioassays using strains other than those used by the NTP give discordant results, i.e., positives in one system are frequently negative in another (Ettlin and Prentice, 2002; Fung et al., 1995; Johnson, 1999). Thus, switching protocols might change which known human carcinogens are false negatives, but, short of identifying all chemicals as positive, none is able to eliminate all false negatives.

Validating the LRB has not been important to the NTP, since only 10 of the hundreds of chemicals tested were known human carcinogens. Assessing the accuracy of the NTP LRB on the basis of 10 chemicals is problematic, but these are the best data available. Assuming a positive in any sex/species rodent group is a positive in humans, 9 of the 10 chemicals were positive, a sensitivity of 90%. This interpretation still leads to 10% false negatives but more troubling is classifying 22% of the chemicals tested as positive (Fung et al., 1995). Requiring a trans-species response to classify a chemical as a human carcinogen, 3 of the 6 chemicals are positive, a sensitivity of 50%, leading to 50% false negatives. Of all chemicals tested, 6.8% are positive in both species (Fung et al., 1995).

To assess the overall accuracy of the NTP LRB, we must also know its specificity (percent of true negatives among known human non-carcinogens). Specificity cannot be estimated directly, because epidemiologic studies can only give an upper limit on potency, not prove a lack of carcinogenic effect (Huff, 1999). However, prevalence puts a lower limit on the specificity, because the sum of false positives and true positives is equal to the prevalence: The percentage of false positives must be less than or equal to the prevalence. If the 133

Table 2
The accuracy of the LRB for predicting human carcinogens assuming any positive rodent bioassay result is positive (sensitivity = 90%, prevalence = 22%, see text) for a 10% prevalence of human carcinogens

	Humans −	Humans +	Total
Rodents −	77	1	78
Rodents +	13	9	22
Total	90	10	

chemicals Fung et al. (1995) used to estimate prevalence contained no real human carcinogens, the *minimum* value for the specificity is one minus the prevalence; and any real human carcinogens it contained would increase the specificity. Thus, if the prevalence is 22% (any positive is positive), the specificity is 78% or higher, and if the prevalence is 6.8% (only trans-species are positive), the specificity is 93% or higher. The sensitivity and specificity of rats towards mice (and mice towards rat) carcinogenicity are all 70–75% (Johnson, 2001; Lin et al., 1995). The responses between rodents and humans are unlikely to be more similar than the responses between rats and mice (Lave et al., 1988). However, we will show that even with these possibly inflated sensitivity and specificity estimates, the LRB has limited value to regulators.

Without testing, each chemical's probability of being a carcinogen equals the percentage of carcinogens in the group. The information value of a test is the increased likelihood that a chemical is a carcinogen, given the test outcomes. Example calculations for the LRB are given in Tables 2 and 3 and overall results are shown in Figs. 1 and 2. If any positive in a rodent group is considered positive (Fig. 1), a negative result in the LRB reduces the initial probability of carcinogenicity by a factor of 8 (e.g., 10/100 before testing reduces to 1/78); a positive result in the LRB increases the initial probability by a factor of 4 (e.g., 10/100 before testing increases to 9/22). For requiring both species to be positive to be considered positive (Fig. 2), a negative result in the LRB reduces the initial probability by a factor of 2 (e.g., 5/100 before testing reduces to 2.5/93); a positive result in the LRB increases the initial probability by a factor of 7 (e.g., 5/100 before testing increases to 2.5/6.8).

Table 3
The accuracy of the LRB for predicting human carcinogens considering only trans-species results positive (sensitivity = 50%, prevalence = 6.8%, see text) for a 5% prevalence of human carcinogens

	Humans −	Humans +	Total
Rodents −	90.7	2.5	93.2
Rodents +	4.3	2.5	6.8
Total	95	5	

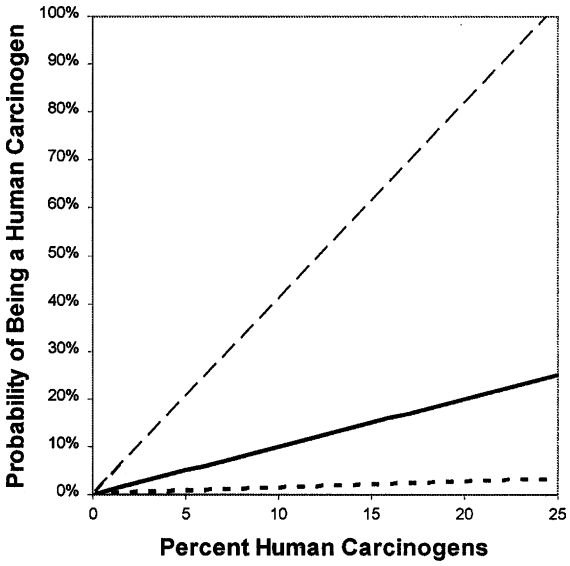

Fig. 1. The probabilities that a chemical is a human carcinogen as a function of the percent human carcinogens among the set of chemicals, given for positive LRB results (– –) and negative LRB results (- - - -), compared to the before-testing probability (—). Scenario is any positive rodent bioassay result is positive, sensitivity = 90%, prevalence = 22% (see text).

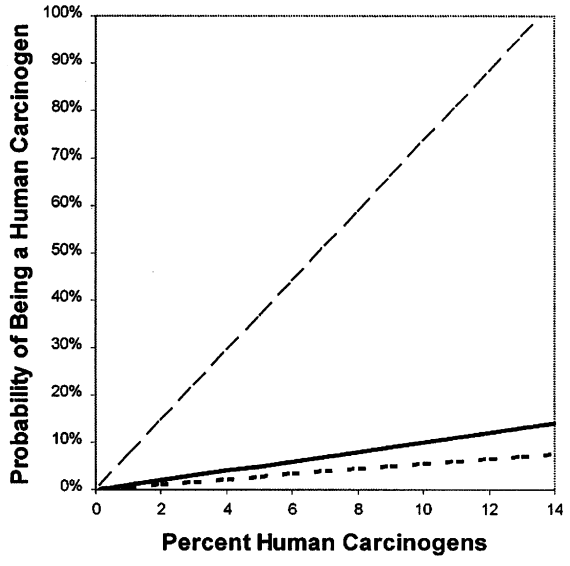

Fig. 2. Same as Fig. 1 except scenario is only trans-species results are positive, sensitivity = 50%, prevalence = 6.8% (see text).

3. Discussion

Tables 2 and 3 show a disturbingly large proportion of incorrect predictions. If we accept any positive as a positive, the sensitivity appears to be high (90%), but 10% of the human carcinogens are still false negatives. This assumption implies that at least 22% of chemicals tested will be positive in the LRB, an implausibly high number, as well as one that identifies too many positive chemicals for effective priority setting. For example,

40% of marketed drugs and food additives are rodent carcinogens (Davies and Monroe, 1995; Johnson, 2002). If we define only trans-species results as positive, then we miss half of known human carcinogens.

The current tests would be better at predicting human carcinogenicity if toxicologists had focused on that goal. They would have calculated the predictivity of each test and examined how best to interpret test results. The key to developing better tests is validating assays for human carcinogenicity. The first step is a rigorous examination of the sensitivity and specificity of each available test for human carcinogenicity. Judging some tests to be inferior because they cannot predict rodent carcinogenicity is a serious mistake (Zeiger, 1998). But the human carcinogenicity or non-carcinogenicity of too few chemicals is known (many of the substances listed in IARC Group A are human viruses, radioisotopes, mixtures, or exposure circumstances not amenable to testing). The second step is rigorous epidemiologic studies to add to the list of known human carcinogens that are available to be tested in assays. Currently, the majority of epidemiologic studies for carcinogens are being performed on substances already known to be human carcinogens (Karstadt, 1998). More challenging will be to establish criteria for probable human non-carcinogens, to assess the specificity of assays directly. The 16 used in the ILSI-HESI Alternative to Carcinogenicity Testing (van der Laan and Spindler, 2002) is a good start. We must identify chemicals to which humans have been exposed at high dosage over several decades with no indication of increased cancer incidence. The NTP must determine criteria for the presumption that a chemical is not a human carcinogen, e.g., the estimated upper bound of potency gives less than a 10^{-6} lifetime risk at a reasonable maximum exposure. The goal is to develop a validation set with at least 50 human carcinogens and 50 human non-carcinogens (or carcinogens of very low potency).

We are under no illusion that epidemiological studies will be easy, inexpensive, or even conclusive in most cases (Gori, 2001; Ward et al., 2003). However, the need for validation is recognized for new assays such as endocrine disruption, transgenic assays, and reproductive toxicity (Ashby, 2001; Reicke and Stahlmann, 2000; U.S. EPA, 1999; van der Laan and Spindler, 2002). Our analysis shows an equally urgent need for validation of the LRB for cancer, as well as other tests that may have been discarded for lack of agreement with the LRB (Zeiger, 1998). This can proceed in parallel with the epidemiologic studies, starting with Group A carcinogens, many of which have not been tested in the NTP LRB or other tests, and continuing with newly identified human carcinogens and presumed non-carcinogens. Without validation on human carcinogens, there is no scientific data supporting the superiority of new carcinogenicity tests, such as transgenic rodents.

Collecting new data for validation is preferable to massaging existing data. Genotoxicity, pharmacology, site and mode of action have been proposed as alternatives to considering only trans-species carcinogens a human threat (Davies et al., 2000; Ettlin and Prentice, 2002; van der Laan and Spindler, 2002). Applying these proposals to the 10 known human carcinogens (Johnson, 2001) and the subset of 133 chemicals used to develop the prevalence information (Fung et al., 1995) would yield different estimates for sensitivity and prevalence, but cannot escape the inevitable trade-off between sensitivity and specificity. Expanding the dataset for validation will cover more types of human carcinogens and non-carcinogens, making modifications to the design and interpretation of all assays more likely to be applicable to unknown chemicals.

4. Conclusions

What should society do until there is a carcinogenicity test with demonstrated high sensitivity and specificity? Companies and regulators already recognize the deficiencies of the LRB and have found other answers. For example, many identified animal carcinogens continue to be used in drugs, pesticides and food additives (Davies and Monroe, 1995; Johnson, 2002; Rakitsky et al., 2000). Interestingly, the vast majority of these animal carcinogens were non-genotoxic, indicating that lack of genotoxicity is being used to cast substantial doubt on the relevance of rodent carcinogenicity in the commercialization of new chemicals (Kowalski, 2001; MacGregor et al., 2000). Quantitative exposure limits can be calculated for chemicals known to be genotoxic and assumed, but not known, to be carcinogenic (Fiori and Meyerhoff, 2002). Until a satisfactory test is found, chemicals could be regulated as carcinogens if they are genotoxic, or based on their non-cancer toxicity if they are not genotoxic.

The fundamental difficulty with screening chemicals for carcinogenicity has been the lack of feedback, or incorrect feedback. The failure to validate the NTP LRB against human carcinogens means that no one knows the accuracy of this test. Accuracy was assumed to be high by some toxicologists without data. Many toxicologists were skeptical but had no better alternative. The problem was compounded by then discarding theories and other tests because they failed to predict rodent carcinogenicity.

The small amount of available data suggests that the NTP LRB produces many false positives and false negatives. The social cost of three decades of reliance on the NTP LRB is hundreds of potentially valuable chemicals that were discarded because they are rodent carcinogens as well as human exposure to perhaps tens of chemicals that were not positive in the LRB, but are

human carcinogens. Toxicology must focus on the goal, human carcinogenicity, not surrogates. When we have a set of chemicals with known human carcinogenicity, toxicology will be able, for the first time, to focus on the right goal, instead of pouring hundreds of millions of dollars into tests whose accuracy was unknown and later shown to be low.

Acknowledgments

We thank the Heinz Endowments for support and Daniel Byrd, Herbert Rosenkranz, Gilbert Omenn, and two anonymous reviewers for comments.

References

Ashby, J., 2001. Expectations for transgenic rodent cancer bioassay models. Toxicol. Pathol. 29 (Suppl. 1), 177–182.

Barlow, S.M., Greig, J.B., Bridge, J.W., Carere, A., Carpy, A.J.M., Galli, C.L., Kleiner, J., Knedsen, I., Koeten, H.B.W.M., Levy, L.S., Madsen, C., Mayen, S., Narbonne, J.-F., Pfannkuchi, F., Prodanchuk, M.G., Smith, M.R., Steinberg, P., 2002. Hazard identification by methods of animal-based toxicology. Food Chem. Toxicol. 40 (2–3), 145–191.

Byrd, D.M., 1988. New statistical approaches to the qualitative interpretation of toxicology data. J. Am. Col. Toxicol. 7, 559–563.

Center for Children's Health and the Environment (CCHE), 2002. Available from: http://www.childenvironment.org.

Crump, K.S., Krewski, D., Van Landingham, C., 1999. Estimates of the proportions of carcinogens and anticarcinogens in bioassays conducted by the US National Toxicology Program. Application of a new meta-analytic approach. Ann. N. Y. Acad. Sci. 895, 232–244.

Davies, T.S., Monroe, A., 1995. Marketed human pharmaceuticals reported to be tumorigenic in rodents. J. Am. Col. Toxicol. 14 (2), 90–107.

Davies, T.S., Lynch, B.S., Monro, A.M., Munro, I.C., Nestmann, E.R., 2000. Rodent carcinogenicity tests need be no longer than 18 months: an analysis based on 210 chemicals in the IARC monographs. Food Chem. Toxicol. 38 (2–3), 219–235.

Ettlin, R.A., Prentice, D.E., 2002. Unexpected tumour findings in lifetime rodent bioassay studies—what to do. Toxicol. Lett. 128 (1–3), 17–33.

Fiori, J.M., Meyerhoff, R.D., 2002. Extending the threshold of regulation concept: de minimis limits for carcinogens and mutagens. Regul. Toxicol. Pharmacol. 35, 209–216.

Fung, V.A., Barrett, J.C., Huff, J., 1995. The carcinogenesis bioassay in perspective: application in identifying human cancer hazards. Environ. Health Perspect. 103 (7–8), 680–683.

Gold, L.S., Ames, B.N., Slone, H., 2002. Misconceptions about the causes of cancer. In: Paustenbach, D. (Ed.), Human and Environmental Risk Assessment: Theory and Practice. Wiley, New York, pp. 1415–1460.

Gori, G.B., 2001. The costly illusion of regulating unknowable risks. Regul. Toxicol. Pharmacol. 34 (3).

International Agency for Research on Cancer (IARC), 2002. Available from: http://monographs.iarc.fr/nomoeval/grlist.html.

Haseman, J., Melnick, R., Tomatis, L., Huff, J., 2001. Carcinogenesis bioassays: study duration and biological relevance. Food Chem. Toxicol. 39, 739–744.

Haseman, J.K., 2000. Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. Drug Metabol. Rev. 32 (2).

Haseman, J.K., Elwell, M.R., 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 (6), 813–820.

Huff, J., 1999. Long-term chemical carcinogenesis bioassays predict human cancer hazards. Issues, controversies, and uncertainties. Ann. N. Y. Acad. Sci. 895, 56–79.

Johnson, F.M., 1999. Carcinogenic chemical-response "fingerprint" for male F344 rats exposed to a series of 195 chemicals: implications for predicting carcinogens with transgenic models. Human Ecol. Risk Assess. 5 (2), 427–443.

Johnson, F.M., 2000. The "rodent carcinogen" dilemma: formidable challenge for the technologies of the new millennium. Ann. N. Y. Acad. Sci. 919, 288–299.

Johnson, F.M., 2001. Response to Tennant et al.: attempts to Replace the NTP Rodent Bioassay With Transgenic Alternatives Are Unlikely to Succeed. Environ. Mols. Mutagen. 37, 89–92.

Johnson, F.M., 2002. How many food additives are rodent carcinogens. Environ. Mol. Mutagen. 39 (1), 69–80.

Karstadt, M., 1998. Availability of epidemiologic data for chemicals known to cause cancer in animals: an update. Am. J. Indus. Med. 34 (5), 519–525.

Kodell, R.L., Chen, J.J., Jackson, C.D., Gaylor, D.W., 1999. Using short-term tests to predict carcinogenic activity in the long-term bioassay. Human Ecol. Risk Assess. 5 (2), 427–443.

Kowalski, L.A., 2001. In vitro carcinogenicity testing: present and future perspectives in pharmaceutical development. Curr. Opin. Drug Discov. Develop. 4 (1), 29–35.

Lave, L.B., Ennever, F.K., Rosenkranz, H.S., Omenn, G.S., 1988. Information value of the rodent bioassay. Nature 336 (6200).

Lin, T., Gold, L.S., Freedman, D., 1995. Carcinogenicity tests and interspecies concordance. Stat. Sci. 10 (4), 337–353.

MacGregor, J.T., Casciano, D., Muller, L., 2000. Strategies and testing methods for identifying mutagenic risks. Mutat. Res. 455 (1–2), 3–20.

Miller, R.T., Davis, B.J., 2001. Summary of panel discussion: past, present and future of toxicologic pathologists and the contributions to hazard identification and molecular mechanisms of action. Toxicol. Pathol. 29 (1), 156–157.

Rakitsky, V.N., Koblyakov, V.A., Turusov, V.S., 2000. Nongenotoxic (epigenetic) carcinogens: pesticides as an example. A critical review. Teratogen. Carcin. Mut. 20 (4), 229–240.

Rall, D.P., 2000. Laboratory animal tests and human cancer. Drug Metabol. Rev. 32 (2), 119–128.

Reicke, K., Stahlmann, R., 2000. Test systems to identify reproductive toxicants. Andrologia 32 (4–5), 209–218.

Roe, F.J., 1998. A brief history of the use of laboratory animals for the prediction of carcinogenic risk for man with a note on needs for the future. Exp. Toxicol. Pathol. 50 (4–6), 271–276.

Ruden, C., 2002. The use of mechanistic data and the handling of scientific uncertainty in carcinogen risk assessments. The trichloroethylene example. Regul. Toxicol. Pharmacol. 35 (1), 80–94.

Schmidt, C.W., 2002. Assessing assays. Environ. Health Perspect. 110 (5), A248–A251.

Slovic, P., 1999. Emotion, sex, politics, and science: surveying the risk-assessment battlefield. Risk Anal. 19 (4), 689–701.

Slovic, P., Malmfors, T., Mertz, C.K., Neil, N., Purchase, I.F., 1997. Evaluating chemical risks: results of a survey of the British Toxicology Society. Human Exp. Toxicol. 16 (6), 289–304.

Tennant, R., Haseman, J., Stoll, R.E., 2001. Transgenic assays and the identification of carcinogens. Environ. Mol. Mutagen. 37 (1), 86–92.

Tomatis, L., 2002. The IARC monographs program: changing attitudes towards public health. Int. J. Occup. Environ. Health 8 (2), 144–152.

United States Environmental Protection Agency, 1999. Endocrine Disruptor Screening Program Web Site. Available from: http://www.epa.gov/scipoly/oscpendo/index.htm.

van der Laan, J.W., Spindler, P., 2002. The in vivo rodent test systems for assessment of carcinogenic potential. Regul. Toxicol. Pharmacol. 35 (1), 122–125.

VanDoren, P.M., 1996. The effects of exposure to "synthetic" chemicals on human health: a review. Risk Anal. 16 (3), 367–376.

Ward, E.M., Schulte, P.A., Bayard, S., Blair, A., Brandt-Rauf, P., Butler, M.A., Dankovic, D., Hubbs, A.F., Jones, C., Karstadt, M., Kedderis, G.L., Melnick, R., Redlich, C.A., Rothman, N., Savage, R.E., Sprinker, M., Toraason, M., Weston, A., 2003. Priorities for development of research methods in occupational cancer. Environ. Health Perspect. 111 (1), 1–12.

Zeiger, E., 1998. Identification of rodent carcinogens and noncarcinogens using genetic toxicity tests: premises, promises, and performance. Regul. Toxicol. Pharmacol. 28 (2), 85–95.