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