Misconception 6—Cancer risks to humans can be assessed by standard high-dose animal cancer tests

Approximately half of all chemicals that have been tested in standard animal cancer tests, whether natural or synthetic, are rodent carcinogens (table 4; Gold & al. 1989a; Gold & al. 1999; Gold & al. 1997a). Why do so many test positive? A reasonable explanation is that the design of these experiments produces effects that would not occur at lower doses. In standard cancer tests, rodents are given chronic, near-toxic doses, the maximum tolerated dose (MTD). The rationale for this experimental design was based on a consensus in the 1970s that chemicals with carcinogenic potential would be rare and, therefore, experiments had to be designed to maximize the chance of finding an effect. Since the costs of conducting these tests are high—currently $2 million to $4 million per chemical (US National Toxicology Program 1998)—a limited number of animals would be put on test (50 in each of three groups: the controls, a group receiving a high dose, and a group receiving half the high dose). Because of the small number of animals on test, the studies lack statistical power and, therefore, the doses were set as high as the animals would tolerate while living long enough to get cancer, since cancer is a disease of old age. Evidence is accumulating that cell division

Table 4: Proportion of chemicals evaluated as carcinogenic

Chemicals tested in both rats and mice [(a)]	
Chemicals in Carcinogenic Potency Database (CPDB)	350/590 (59%)
Naturally occurring chemicals in the CPDB	79/139 (57%)
Synthetic chemicals in the CPDB	271/451 (60%)
Chemicals tested in rats and/or mice [(a)]	
Chemicals in the CPDB	702/1348 (52%)
Natural pesticides in the CPDB	37/72 (51%)
Mold toxins in the CPDB	14/23 (61%)
Chemicals in roasted coffee in the CPDB	21/30 (70%)
Commercial pesticides	79/194 (41%)
Innes negative chemicals retested [a]	17/34 (50%)
Physician's Desk Reference (PDR): drugs with reported cancer tests [(b)]	117/241 (49%)
FDA database of drug submissions [(c)]	125/282 (44%)

Sources: (a) ***Carcinogenic Potency Database*** (http://potency.berkeley.edu; Gold & al. 1999; Gold & Zeiger 1997); (b) Davies & Monro 1995; (c) Contrera & al. 1997.
Note: 140 drugs are in the databases of both the Food and Drug Administration (FDA) and the *Physician's Desk Reference* (***PDR***).

caused by the high dose itself, rather than the chemical per se, is increasing the carcinogenic effects and, therefore, the positivity rate. High doses can cause chronic wounding of tissues, cell death, and consequent chronic cell division of neighboring cells. This is a risk factor for cancer (Ames & al. 1996) because, each time a cell divides, the probability increases that a mutation will occur, thereby increasing the risk for cancer.

At the low levels to which humans are usually exposed, such increased cell division does not occur. The process of mutagenesis and carcinogenesis is complicated because many factors are involved: e.g. DNA lesions, DNA repair, cell division, clonal instability, apoptosis (cell suicide in response to DNA damage), and p53 (a cell cycle control gene that is mutated in half of human tumors) (Christensen & al. 1999; Hill & al. 1999). The normal endogenous level of ***oxidative DNA lesions*** in cells is appreciable (Helbock & al. 1998). In addition, tissues injured by high doses of chemicals have an inflammatory immune response involving activation of white cells in response to cell death (Adachi & al. 1995; Czaja & al. 1994; Gunawardhana & al. 1993; Laskin & Pendino 1995; Laskin & al. 1988; Roberts & Kimber 1999; Wei & al. 1993a; Wei & al. 1993b). Activated white cells release mutagenic oxidants (including peroxynitrite, hypochlorite, and H_2O_2). Therefore, the very low levels of chemicals to which humans are exposed through water pollution or synthetic pesticide residues may pose no, or only minimal, cancer risks because these effects do not occur at low doses.

Analyses of the limited data on dose-response in bioassays are consistent with the idea that cell division from cell-killing and cell replacement is important. Among rodent bioassays with two doses and a control group, about half the sites evaluated as target sites are statistically significant at the MTD but not at half the MTD ($p < 0.05$). Ad libitum feeding in the standard bioassay can also contribute to

the high positivity rate (Hart & al. 1995a). In mice fed a restricted number of calories, cell division rates are markedly lower in several tissues than in mice fed ad libitum (Lok & al. 1990). Linearity of response to increasing dosage seems less likely than has been assumed because of the ***inducibility*** of the numerous defense enzymes that deal with exogenous chemicals as groups (e.g. oxidants, electrophiles) and thus protect us against the natural world of mutagens as well as the small amounts of synthetic chemicals to which we are exposed (Ames & al. 1990b; Calabrese & Baldwin 2001; Luckey 1999; Munday & Munday 1999; Trosko 1998).

Risk assessment requires additional biological data

More than a decade ago, we argued that risk assessment for humans requires data on the mechanism of carcinogenesis for each chemical (Ames & Gold 1990; Ames & al. 1987). Historically, standard practice in regulatory risk assessment for chemicals that induce tumors in high-dose rodent bioassays has been to extrapolate risk to low dose in humans by multiplying rodent potency by human exposure, i.e. by assuming linearity in the dose response. Without data on the mechanism of carcinogenesis, however, the true human risk of cancer at low dose is highly uncertain and could be zero (Ames & Gold 1990; Clayson & Iverson 1996; Gold & al. 1992; Goodman 1994). Adequate risk assessment from animal cancer tests requires more information for a chemical, about pharmacokinetics, mechanism of action, apoptosis, cell division, induction of defense and repair systems, and differences among species. Several mechanisms have now been identified that indicate that carcinogenic effects at the high doses of rodent tests would not be relevant to the low doses of most human exposures (e.g. saccharin, BHA, chloroform, *d*-limonene). Under the new *Guidelines for Cancer Risk Assessment* from the US Environmental Protection Agency (EPA), these mechanisms are to be considered in

evaluating the dose-response, method of risk assessment, and relevance to humans; the default linear extrapolation has been replaced by this more scientific approach (US Environmental Protection Agency 1999).

Examples of such biologically based mechanisms include cell proliferation following cytotoxic effects at high doses of saccharin, only in the male rat urothelium; the cytotoxicity results from formation of a precipitate in rat urine, which is a species-specific response. For several chemicals, studies show an association between cell division in the rodent liver and cancer (e.g. chloroform, oxazepam, 2,4-diaminotoluene) (Ames & Gold 1990; Ames & al. 1993a; Butterworth & Bogdanffy 1999; Cohen 1998; Cunningham & al. 1994a; Cunningham & al. 1991; Cunningham & al. 1994b; Heddle 1998). Some chemicals (e.g. *d*-limonene, induce kidney tumors in male rats by a mechanism that is not relevant to humans: accumulation of a male rat-specific protein (α_{2u}-globulin) resulting in toxicity to the kidney, sustained cell proliferation, and kidney tumors. Humans do not synthesize α_{2u}-globulin or any protein that can function like it (Swenberg & Lehman-McKeeman 1999) and, therefore, the carcinogenic effect in male rats is not predictive of a cancer hazard to humans. Some chemicals induce thyroid follicular-cell tumors at high doses by a metabolic inactivation of the thyroid hormones T_3 and T_4, which results in increased levels of thyroid-stimulating hormone levels, sustained proliferation of cells in the thyroid, and tumor formation (McClain 1990). Humans are less sensitive to this secondary, threshold mechanism than rats (McClain 1994; US Environmental Protection Agency 1998a).

The US EPA's evaluation of chloroform provides an example of the new emphasis on incorporating more biological information into evaluations of cancer test results and risk assessment. The EPA concluded that chloroform-induced tumors were secondary to toxic effects that occur at high dose. Therefore, the EPA relied on a nonlinear dose-

response approach with a margin of exposure to estimate cancer risk for humans. They concluded that

> chloroform is likely to be carcinogenic to humans by all routes of exposure under high-exposure conditions that lead to cytotoxicity and regenerative hyperplasia in susceptible tissues. Chloroform is not likely to be carcinogenic to humans by any route of exposure under exposure conditions that do not cause cytotoxicity and cell regeneration. (US Environmental Protection Agency 2002)

Is selection bias causing the high positivity rate?

Since the results of high-dose rodent tests are routinely used to identify a chemical as a possible cancer hazard to humans, it is important that we try to understand how representative the 50% positivity rate might be of all untested chemicals. If half of all chemicals (both natural and synthetic) to which humans are exposed would be positive if tested, then the utility of a rodent bioassay to identify a chemical as a "potential human carcinogen" is questionable. To determine the true proportion of rodent carcinogens among chemicals would require a comparison of a random group of synthetic chemicals to a random group of natural chemicals. Such an analysis has not been done.

A counter argument to the idea that the 50% positivity rate is due to the effects of administering high doses is that so many chemicals are positive because they were selected for testing on the grounds that they were expected to be carcinogenic. We have discussed that this is a likely bias since cancer testing is both expensive and time consuming, making it prudent to test suspicious compounds (Gold & al. 1998); however, chemicals are selected for cancer-testing for many reasons other than suspicion, including the extent of human exposure, level of production and occupational exposure,

and scientific questions about carcinogenesis. Moreover, if the main basis for selection was that chemicals were suspected carcinogens, then one should select mutagens (80% are carcinogens compared to 49% of nonmutagens); yet, 55% of the chemicals tested are nonmutagens (Gold & al. 1998). The idea that chemicals are selected for testing because they are likely to be carcinogenic, rests on an assumption that researchers have adequate knowledge about how to predict carcinogenicity and that there is consensus about the criteria; that is, the idea that bias in the positivity rate is due to selection requires that there is shared, adequate knowledge of what is likely to be carcinogenic.

However, while some chemical classes are more often carcinogenic in rodent bioassays than others—e.g. nitroso compounds, aromatic amines, nitroaromatics, and chlorinated compounds—several results suggest that predictive knowledge is highly imperfect, even now after decades of testing results on which to base predictions have become available. For example, a prospective prediction exercise was conducted by several experts in 1990 in advance of the 2-year bioassays by the United States National Toxicology Program (NTP). There was wide disagreement among the experts as to which chemicals would be carcinogenic when tested; accuracy varied, thus indicating that predictive knowledge is uncertain (Omenn & al. 1995). One predictive analysis for a randomly selected group of chemicals has been conducted using a computerized method based on chemical structure; among 140 randomly selected chemicals, 65 (46%) were predicted to be carcinogenic if tested in standard bioassays (Rosenkranz & Klopman 1990). Another argument against the hypothesis of selection bias is the high positivity rate for drugs (table 4), because drug development tends to select chemicals that are not mutagens or expected carcinogens.

A study by Innes & al. (1969) has frequently been cited (Ames & al. 1987, Letters) as evidence that the positivity

rate is low, because only 9% of 119 chemicals tested (primarily pesticides) were positive. However, the Innes tests were only in mice, had only 18 animals per group, and were terminated at 18 months. This protocol lacked the power of modern experiments, in which both rats and mice are tested, with 50 animals per group for 24 months. When 34 chemicals for which Innes obtained negative results were retested in other strains of mice or in rats, using more adequate protocols including higher doses and longer experiment length, 17 of the 34 formerly negative chemicals tested positive (table 4) (Cohen 1995; Cohen & Lawson 1995; Gold & al. 1999; Gold & al. 1997a).

Thus, it seems likely that a high proportion of all chemicals, whether synthetic or natural, might be "carcinogens" if run through the standard rodent bioassay at the MTD. For nonmutagens, carcinogenicity would be primarily due to the effects of high doses; for mutagens, it would result from a synergistic effect between cell division at high doses and DNA damage (Ames & Gold 1990; Ames & al. 1993a; Butterworth & al. 1995). Without additional data on the mechanism of carcinogenesis for each chemical, the interpretation of a positive result in a rodent bioassay is highly uncertain. The carcinogenic effects may be limited to the high dose tested.

Problems in extrapolating carcinogenicity between species

The use of bioassay results in risk assessment requires a qualitative species extrapolation from rats or mice to humans. The accuracy of this extrapolation is generally unverifiable, since data on humans are limited. Ultimately one wants to know whether the large number (many hundreds) of chemicals that have been shown to be carcinogenic in experimental animals would also be carcinogenic in humans. This question cannot be answered by reversing the

question—that is, by asking whether the small number of chemicals that are carcinogenic to humans are also carcinogenic in rodent bioassays—because, even if most human carcinogens were carcinogenic to experimental animals, the converse does not necessarily follow, as can be demonstrated by a simple probabilistic argument (Freedman & Zeisel 1988).

Evidence about interspecies extrapolation can, however, be obtained by investigating whether chemicals that are carcinogenic in rats are also carcinogenic in mice, and visa versa. If mice and rats are similar with respect to carcinogenesis, this provides some evidence in favor of interspecies extrapolations; conversely, if mice and rats are different, this casts doubt on the validity of extrapolations from mice to humans.

One measure of interspecies agreement is concordance, the percentage of chemicals that are classified the same way as to carcinogenicity in mice and rats (i.e. results are concordant if a chemical is a carcinogen in either both species or in neither, and results are discordant if a chemical is a carcinogen in one species but not in the other). Observed concordance in bioassays is about 75% (Gold & al. 1997a; Gold & al. 1998), which may seem low since the experimental conditions are identical and the species are similar. The observed concordance is just an estimate based on limited data. We have shown by simulations for 300 ***NCI / NTP*** bioassays of chemicals tested in both rats and mice (which have an observed concordance of 75%), that an observed concordance of 75% can arise if the true concordance is anything between 20% and 100% (Freedman & al. 1996; Lin & al. 1995) and, indeed, observed concordance can seriously overestimate true concordance. Thus, it seems unlikely that true concordance between rats and mice can be estimated with any reasonable degree of confidence from bioassay data.

Problems in using results of animal cancer tests for regulatory risk assessment

We have discussed the problems in deriving valid human risk assessments from the limited data from animal cancer tests (Bernstein & al. 1985; Gold & al. 1998). Standard practice in regulatory risk assessment for a given rodent carcinogen has been to extrapolate from the high doses of rodent bioassays to the low doses of most human exposures by multiplying ***carcinogenic potency*** in rodents by human exposure. Strikingly, however, due to the relatively narrow range of doses in 2-year rodent bioassays and the limited range of statistically significant tumor incidence rates, the various measures of potency obtained from 2-year bioassays, such as the EPA's q_1^* value, the TD_{50}, and the lower confidence limit on the TD_{10} (LTD_{10}) are constrained to a relatively narrow range of values about the MTD, in the absence of 100% tumor incidence at the target site, which rarely occurs (Bernstein & al. 1985; Freedman & al. 1993; Gaylor & Gold 1995; Gaylor & Gold 1998; Gold & al. 1997a). For example, the dose usually estimated by regulatory agencies to give one cancer in a million can be approximated simply by using the MTD as a surrogate for carcinogenic potency. Gaylor and Gold (1995) have shown that the "virtually safe dose" (VSD) can be approximated by the MTD/740,000 for rodent carcinogens tested in the bioassay program of the NCI/NTP. The MTD/740,000 was within a factor of 10 of the VSD for 96% of carcinogens. This is similar to the finding that in near-replicate experiments of the same chemical, potency estimates vary by a factor of 4 around a median value (Gaylor & al. 1993; Gold & al. 1989b; Gold & al. 1987b).

Using the benchmark dose approach proposed in the EPA carcinogen guidelines, risk estimation is similarly constrained by bioassay design. A simple, quick, and relatively precise determination of the LTD_{10} can be obtained by the maximum tolerated dose (MTD) divided by 7 (Gaylor &

Gold 1998). Both linear extrapolation and the use of safety or uncertainty factors proportionately reduce a tumor dose in a similar manner. The difference in the regulatory "safe dose," if any, for the two approaches depends on the magnitude of uncertainty factors selected. Using the benchmark dose approach of the proposed carcinogen risk assessment guidelines, the dose estimated from the LTD_{10} divided, for example, by a 1000-fold uncertainty factor is similar to the dose of an estimated risk of less than 10^{-4} using a linear model. This dose is 100 times higher than the VSD corresponding to an estimated risk of less than 10^{-6}. Thus, whether the procedure involves a benchmark dose or a linearized model, cancer risk estimation is constrained by the bioassay design.