REGULATORY TOXICOLOGY AND PHARMACOLOGY 22, 283–291 (1995)

An Empirical Examination of Factors Influencing Prediction of Carcinogenic Hazard across Species

GEORGE M. GRAY,*[,1] PING LI, ILYA SHLYAKHTER, AND RICHARD WILSON†

*Center for Risk Analysis, Harvard School of Public Health, 718 Huntington Avenue, Boston, Massachusetts 02115; and †Department of Physics, Harvard University, Jefferson Laboratories, Cambridge, Massachusetts 02138

Received September 1, 1995

This study was stimulated by a recent U.S. Environmental Protection Agency (EPA, 1994) statement in draft environmental carcinogen risk assessment guidelines: "Several kinds of observations from animal studies can contribute to the judgment whether animal responses indicate a significant carcinogenic hazard to humans." We have investigated each of these kinds of observation using the cancer bioassay data system database. We obtained concordances from rat to mouse (and vice versa) for various subgroups of chemicals as follows: chemicals that induced tumors at multiple sites, chemicals that induce cancer in both sexes, chemicals that display reduced latency, and chemicals increasing the rates of rare tumors. The concordances are much higher for these chemical subgroups than the chemical groups that induce tumor at a single site, in only one sex, or without reduced latency, respectively. Thus, our findings support some of the EPA's suggested factors.

INTRODUCTION

Long-term rodent bioassays are a primary source of information about potential human carcinogenic risk. However, there is considerable scientific uncertainty in generalizing the results of these tests from rodents to humans and in extrapolating the dose response from the high doses used in the animal bioassay to the lower doses encountered by humans in the workplace or environment. Scientists and risk assessors have been searching for factors which influence the ability of animal tests to predict human hazard. The U.S. Environmental Protection Agency (EPA), in draft revisions to its carcinogen risk assessment guidelines (EPA, 1994), identified several observations from animal bioassays which the agency believes increase the likelihood that a rodent carcinogen could also be a human carcinogen. According to EPA a compound would be considered more likely to be a human carcinogenic hazard if it (i) causes an increase in uncommon tumor type(s); (ii) causes increased tumor rates at multiple sites; (iii) induces tumors by more than one route of administration; (iv) causes tumors in multiple species, strains, or in both sexes of one test species; (v) is associated with reduced latency in tumor formation; (vi) leads to metastatic tumors; (vii) causes an unusual (presumably large) magnitude of tumor response; (viii) increases the proportion of malignant tumors in animals; or (ix) is associated with a clear dose-related increase in tumor rate. Many of these factors have been proposed as increasing human carcinogenic hazard before (e.g., EPA, 1986; IARC, 1987; Huff *et al.*, 1991) although often with little or no supporting evidence (e.g., Byrd, 1988). The identification of these factors has been based on biological intuition and some experimental evidence. Our goal is to quantitatively evaluate these factors in a large database.

Ideally, to identify important factors which may influence the predictive value of rodent carcinogenicity tests for humans we would study the response of humans and laboratory rodents to known human carcinogens. However, due to the rather limited number of known human carcinogens which have been tested in standard animal bioassays, few conclusions can be drawn. On the other hand, there are a large number of chemical carcinogenicity tests that have been conducted on the two common laboratory species, rats and mice. Concordance of carcinogenic response across these phylogenetically similar species should probably be considered an upper bound on the concordance that could be expected between rodents and humans.

Several investigators have examined the prediction of carcinogenic response across species including Crouch and Wilson (1979), Purchase (1980), Haseman and Huff (1987), Gold *et al.* (1989, 1992), Byrd *et al.* (1990), and Piegorsch *et al.* (1992). All find relatively high concordance of carcinogenic response between rodent species. Some have investigated the predictive ability of particular subsets of chemicals. For example,

[1] To whom correspondence should be addressed.

0273-2300/95 $12.00

Gold *et al.* (1989) found that cross-species prediction was higher for mutagens than nonmutagens and for substances which are toxic at low doses. Piegorsch *et al.* (1992) demonstrated that cross-species concordance is higher for chemicals with a high carcinogenic potency. Because carcinogenic potency is correlated (at least in the standard bioassay) with acute toxicity (Parodi *et al.*, 1982; Zeise *et al.*, 1984), this is equivalent to an increase in concordance with low maximum tolerated dose (MTD) as found by Gold *et al.* (1989).

This study uses the cancer bioassay data system (CBDS), a computerized database of long-term animal bioassays conducted by the National Cancer Institute and National Toxicology Program, to evaluate several of the factors listed previously to see which increase the concordance of carcinogenic response between rats and mice. Specifically, we examine whether chemicals which cause increases in rare tumors, which induce tumors at multiple sites, which cause increased tumor rates in both sexes of a test species, which reduce latency of neoplasia, or which cause a clear dose-related trend in tumor incidence are better cross-species predictors of carcinogenicity. We hope that this empirical analysis will allow identification of factors which influence prediction of carcinogenic response in general. We realize that knowledge of biological and pharmacokinetic factors, which may strongly influence carcinogenic response in different species, would be very useful. However, mechanistic information of sufficient detail is available on only a handful of the chemicals being studied. In addition, study of the general attributes of chemicals will be useful in the case of newly tested chemicals with few mechanistic data yet developed. We also hope that this type of analysis will be useful as a hypothesis-generating tool.

We examine the positive predictive value (PPV) and negative predictive value (NPV) across species for chemicals with these attributes. We begin with the initial hypothesis that, in fact, there are no differences between chemicals that influence response in two different species. We are, therefore, testing the EPA statements as alternative hypotheses. There is one strong assumption which we make in using these data. We assume that the standard 2-year bioassay, with exposures near the maximum tolerated dose, is appropriate for evaluating carcinogenicity. It must be clear that we are only evaluating the question of prediction of high-dose animal carcinogenic response. The study is relevant to the validity of extrapolation to low-dose human risk only if further assumptions are made. An additional assumption is also made—that the group of chemicals tested by NTP from the universe of chemicals is representative of the particular chemical for which we may wish to make a prediction. It is, in fact, unlikely that tested chemicals are a random selection because testing has focused on chemicals thought likely to be carcinogenic or with widespread human exposure.

METHODS

Database

The carcinogenesis bioassays performed under NCI/NTP program are the primary source of data for our analysis. The data are contained in the CBDS and include results up to December 1987 (Arnold, 1988), at which point the CBDS format was abandoned by NTP. The extent of the database in this analysis is slightly different from that in Byrd *et al.* (1990). The database consists of bioassays on 18 strains of mice, hamsters, and rats and seven routes of administration (details in Byrd *et al.*, 1990) and includes a total of 587 combinations of substances. Most of these bioassays have data for both male and female animals.

The computation methods were developed by Dr. Edmund Crouch and are described in Byrd *et al.* (1990). We will reiterate a few key factors. First, CBDS experiments had to meet certain criteria before being analyzed. Each experiment consisted of groups of animals treated by a chemical substance at different dose. Each combination of chemical, species, strain, sex, and route of administration was considered a separate experiment. The following three criteria were set in our study:

1. Each experiment had to have at least one control group, and each group was required to have data available on some animals.
2. Relative magnitudes of the first dose administered to the different dosed groups were taken as representative (i.e., doses changed during the bioassay, the beginning dose was used).
3. Only long-term experiments (the longest lifetime recorded for any animal exceeded 70 weeks) were included.

Because we are primarily interested in the interspecies concordance between mouse and rat, only those substances tested on both mouse and rat are analyzed. All the mouse strains and rat strains were used. Furthermore, only experiments conducted using food, water, or gavage as routes of exposure were used in the analysis.

Second, the statistical significance within each experiment was calculated. For each dosed group, a one-sided Fisher's exact significance value was obtained by comparing the tumor incidence with the control group. The numbers of tumors in each group were those at the end of the experiment. The numbers of animals at risk were taken to be either the initial numbers of animals in a group ("unadjusted" analysis) or the number of animals when the first animal was found to have a tumor of the classification under analysis ("adjusted" analysis). In addition, a mortality-adjusted Cochran–Armitage dose trend test (Gart *et al.*, 1986) was performed.

For each experiment, the response was classified as positive, negative, or weak. A response was considered

TABLE 1
Summary of Responses in the CBDS Database[a]

	No. of chemicals	%
Chemicals tested on both mouse and rat	317	
Chemicals positive in either mouse or rat	268	85
Chemicals positive in both species	155	49
Chemicals positive in single site	156	49
Chemicals positive in two sites	60	19
Chemicals positive in three or more sites	67	21
Chemicals positive in single sex of one species	118	37
Chemicals positive in both sexes of one species	150	47

[a] Weak evidence counted as positive.

positive if the Fisher's exact test $P < 0.025$ for any one dosed group of $P < 0.05$ for two dosed groups. A weak response was defined if the criteria for positive were not met but a dose-trend test with $P < 0.025$ was observed. Otherwise, the response was negative. In our investigations the weak responses can be assigned to either the positive or the negative response category. The weak was included as positive throughout this analysis. A summary of the responses in the database is provided in Table 1.

The same tumor classification system used by Byrd *et al.* (1990) was implemented in this work. There are a total of 102 classifications which include both tumor types and tumor sites (shown in Table 2). Among the 102 classifications, 66 were classified as malignant primary neoplasms and 36 as benign primary neoplasms. This classification excludes metastatic lesions or unassignable lesions as to primary or metastatic.

A total of 317 chemicals tested on both mouse and rat satisfy all of the requirements. Further statistic details of these chemicals are shown in Table 1. Within the 317 substances that met our criteria, 109 are mutagens and 208 are nonmutagens. Data on mutagenicity was provided by Zeiger (1994).

Data Analysis

To investigate the factors identified by EPA as potentially increasing the likelihood of a "serious" carcinogenic effect, we investigated interspecies and intraspecies concordances grouping chemicals in the CBDS to correspond to the different factors. In the context of this paper, concordance refers to the agreement of tumor induction at any site between different species/strain/route/sex combinations in response to the same chemical substance. Higher concordance or positive predictive value indicates a better chance of observing a response in one group when it is already found in another group. The obvious notion is that a good interspecies concordance in animals may imply a good concordance between animals and humans (Crouch and Wilson, 1979).

Tumor responses were analyzed on the basis of 2 × 2 concordance tables. A sample concordance table is shown in Table 3. The columns and rows represent the same or different groups of tumor site/type classifications in one or another combination of species, strains, routes, and sexes. The numbers in the cells designated A–D represent numbers of tested chemical substances that had positive or negative response for an animal group (species or sex). Perfect concordance would be defined as B = C = 0, meaning that a tumor found in the row selection always predicts a tumor in the column selection and not finding a tumor in the row selection guarantees not finding a tumor in the column selection. Throughout our analysis, a chemical went to the positive cells for a species when it caused a tumor in at least one sex and in at least one site.

A concordance plot (Byrd *et al.*, 1991) was used to illustrate both PPV and NPV between two groups of tumor site/type classifications. The PPV and NPV are defined as PPV = A/(A + B) and NPV = D/(C + D). Confidence limits for PPV and NPV were obtained by treating each value as binomial samples from an infinite population of substances. The reported error bars represent a confidence interval corresponding to a 68%, one-sided binomial distribution (approximately 1 SD).

Definitions

It was required that we interpret many of the EPA factors to conduct the analysis. Within the CBDS the following definitions were used for the EPA factors examined (EPA statement first, our definition following). The exact EPA statement is as follows: "Several kinds of observations from animal studies can contribute to the judgment whether animal responses indicate a significant carcinogenic hazard to humans. . . . The following observations add significance:"

Uncommon tumor types. Chemicals which cause an increase in a type of tumor with a historical background rate lower than 2% in control animals. Common tumors are defined as those with a background rate higher than 10% for a particular species/strain/sex. About 80% of the 102 classifications were considered rare tumors.

Tumors at multiple sites. Chemicals causing increase in tumors in two sites or three or more sites in at least one group (sex, species) of treated animals were examined. These are compared to chemicals which induce tumors in a single site.

Tumors by more than one route of administration. The CBDS has too few chemicals with multiple tests via different exposure routes to examine this factor.

Tumors in multiple species, strains, or in both sexes. Chemicals which cause tumors in both species in the bioassay have perfect predictive value in our analysis

TABLE 2
Tumor Classification

1. Skin, breast, papilloma, +
2. Respiratory, oral + papilloma, +
3. GI papilloma, +
4. Urinary, reprod. papillma, +
5. Skin, breast adenoma, +
6. Respiratory, oral adenoma, +
7. Liver adenoma, +
8. GI adenoma
9. Urinary, reprod. adenoma
10. Pituitary adenoma
11. Endocrine + adenome, +
12. Skin, urinary + adenomas
13. Reprod., endocrine + adenomas
14. Tubular cell adenoma, +
15. Follicular, clear cell adenomas
16. Cortical adenoma
17. Skin, breast, liver, and cystadenomas
18. GI, urinary, reprod. cystadenomas
19. Endocrine cystadenomas
20. Acinar cell adenoma
21. Keratoacanthoma, +
22. Tubular adenoma, +
23. Interstitial cell tumor
24. Pheochromocytoma, +
25. Skin, breast fibroma
26. Blood, bone fibroma
27. Fibroma, other sites
28. Lipoma, +
29. Leiomyoma, +
30. Endometrial stromal polyp, +
31. Fibroadenoma, +
32. Hemangioma, +
33. Osteoma, +
34. Hamartoma, +
35. Ganglioneuroma, +
36. Chromophobe adenoma
37. Skin, breast carcinoma
38. Blood, bone carcinoma
39. Lung carcinoma
40. Oral, GI carcinoma
41. Urinary carcinoma
42. Reproductive carcinoma
43. Pituitary carcinoma
44. Endocrine carcinoma
45. Brain carcinoma
46. Skin, breast papillary carcinoma
47. Lung papillary carcinoma
48. GI, urinary papillary carcinoma
49. Uterus, ovary papillary carcinoma
50. Thyroid papillary carcinoma
51. Skin + squamous carcinoma
52. Lung squamous carcinoma
53. Oral, GI squamous carcinoma
54. Urinary, reprod. squamous carcinoma
55. Skin, GI basal cell carcinoma
56. Urinary transitional cell carcinoma
57. Skin, breast adenocarcinoma
58. Lung adenocarcinoma
59. Oral, GI adenocarcinoma
60. Urinary, reprod. adenocarcinoma
61. Endocrine, brain adenocarcinoma
62. Islet cell carcinoma
63. Bile duct carcinoma
64. Hepatocellular carcinoma
65. Alveolar, broncheolar carcinoma
66. Chromaphobe carcinoma
67. Tubular cell adenocarcinoma
68. Thyroid follicular cell carcinoma
69. Cortical carcinoma
70. Clear cell carcinoma
71. Adnexal, sebaceous, + carcinoma
72. Thymoma
73. Granulous cell carcinoma
74. Interstitial cell carcinoma
75. Pheochromocytoma, +
76. Skin sarcoma, +
77. Other sites sarcoma, +
78. Blood, bone sarcoma, +
79. Liposarcoma
80. Leiomyosarcoma, +
81. Endometrial stromal sarcoma, +
82. Carcinosarcoma, +
83. Mesothelioma, osteosarcoma, +
84. Teratoma, +
85. Hemangiosarcoma, +
86. Granular cell tumor, +
87. Glioma
88. Oligodendroglioma, +
89. Astrocytoma
90. Olfactory neuroblastoma, +
91. Neurofibrosarcoma
92. Lymphoma
93. Lymphocytic lymphoma
94. Histiocytic lymphoma
95. Mixed lymphoma
96. Malignant reticulosis
97. Leukemia
98. Myelomonocytic leukemia
99. Lymphocytic leukemia
100. Plasmacytic leukemia, +
101. Granulocytic leukemia
102. Monocytic leukemia

(by definition) and so this question was not examined. We do evaluate whether chemicals which increase tumors of both sexes of one species are better predictors of carcinogenicity in the other species than are those increasing tumors in only one sex.

Progression of lesions from preneoplastic to benign to malignant. This factor is not examined. It will be the focus of a future analysis.

Reduced latency of neoplastic lesions. These are assumed to be chemicals which cause the appearance of tumors in treated animals sooner than in controls. We examined predictive value for chemicals causing any tumor in treated animals 4, 13, or 26 weeks earlier than controls. We also investigated the difference between chemicals causing reduced latency in both dose groups or only one.

Metastases. This factor is not evaluated.

Unusual magnitude of tumor response. We defined this factor to mean chemicals which cause a very strong positive dose-related increase in tumors in treated animals. We compare chemicals with trend tests significant at the $P < 0.05$, $P < 0.025$, or $P < 0.01$ levels in one species to the standard definition of positive in the other. Use of significance level as a measure of magnitude of response is reasonable because of the standard experimental design in NTP bioassays, but might present problems in other situations.

Proportion of malignant tumors. Our tumor type grouping only allows this distinction to be made for a few tumor types, therefore this factor is not evaluated.

Dose-related increases. Here we distinguished between statistically significant responses only by pair-

TABLE 3
Sample Concordance Table

	Column		
Row	+	−	
+	A	B	A + B
−	C	D	C + D
	A + C	B + D	A + B + C + D

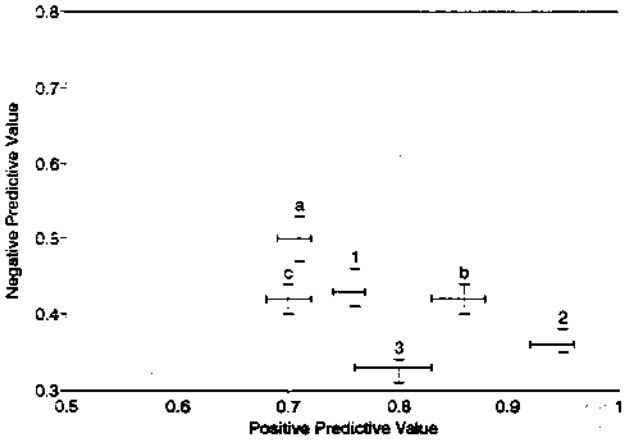

FIG. 1. Concordance of carcinogenic response between mice and rats for any tumor type (1, mouse to rat; a, rat to mouse), from rare tumors to any tumor (2, mouse to rat; b, rat to mouse), and from common tumors to any tumor (3, mouse to rat; c, rat to mouse).

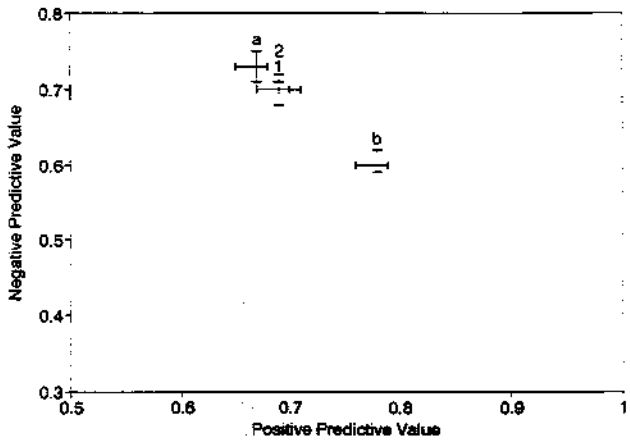

FIG. 2. Concordance of carcinogenic response within a species. Male mouse to female mouse, 1; female mouse to male mouse, 2. Male rat to female rat, a; female rat to male rat, b.

wise comparison versus those with clear dose-related increases in tumor rate. Chemicals which show a $P \leq 0.05$ Cochran–Armitage test for dose response are compared to those which are only significant at $P \leq 0.05$ by Fisher exact test for at least one dose group but do not have a significant dose trend.

RESULTS

We find, similar to other investigators discussed previously, that a positive response in mice is associated with a positive response in rats 76% of the time. The positive predictive value from rats to mice is approximately 71% (shown in Fig. 1). This can be taken as a general predictive rate across species in the high-dose chronic rodent bioassay. It has been noted that chance alone would lead to an approximately 43% PPV from rats to mice and 48% from mice to rats (Gold *et al.*, 1989). Positive predictive value from male mice to female mice, and vice versa, is essentially identical (69%). However, a difference is observed for rats depending on whether the prediction is from response in male rats to female rats (67%) or from female rats to males (78%) (Fig. 2).

First we examine whether chemicals increasing rates of rare tumors are more concordant across species than all chemicals or than those increasing common tumors. We find that chemicals increasing rates of rare tumors (sometimes along with other tumor sites) have better predictive value across species than those increasing only common tumors. Figure 1 also shows a plot of positive and negative predictive values for chemicals increasing rare tumors and those increasing common tumors. Those increasing rare tumors have a statistically significantly greater predictive value than all chemicals (95 vs 76% for mice to rats; 86 vs 71% for rats to mice) or those increasing common tumors (95 vs 80% for mice to rats; 86 vs 70% for rats to mice). This confirms an unpublished conclusion by Byrd *et al.* (1991).

We next investigated whether there is a difference in positive predictive value of carcinogenic response across species for chemicals which cause an increase in tumor rate at one, two, or three or more sites in the predicting species. Figure 3 shows that chemicals which cause an increase in tumors at multiple sites (either two or three or more) have much higher mouse to rat positive predictive value than chemicals causing an increase at a single site (96–98 vs 57%). The same is true for the rat to mouse prediction (85–90 vs 56%). Interestingly, there did not seem to be much difference in predictive value between chemicals causing increased tumor rates in two organs and those causing increases in three or more. It appears that for either

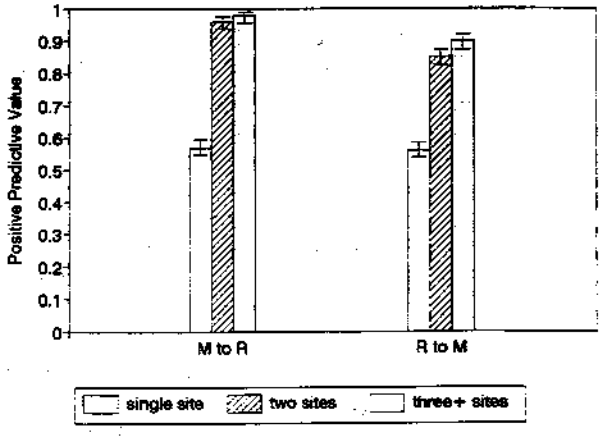

FIG. 3. Prediction of carcinogenic response across species for chemicals inducing tumors in one, two, and three or more sites.

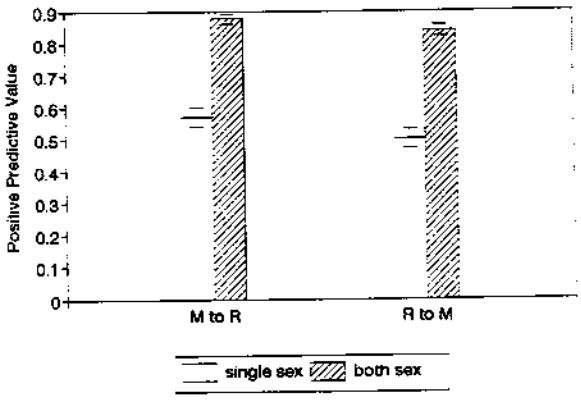

FIG. 4. Prediction of carcinogenic response across species for chemicals inducing tumors in just one or both sexes of one test species.

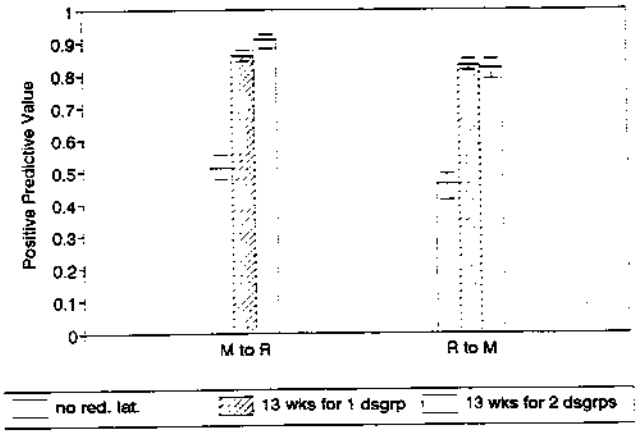

FIG. 5. Influence of reduced latency of tumor development on cross species prediction of carcinogenicity including difference between reduced latency in one dose group or two.

comparison, if a chemical causes multiple tumors in one species it is virtually certain to increase the tumor rate in the other species as well. Interestingly, for the group of chemicals inducing tumors at two or more sites the correlation between potency and cross-species prediction (Gold *et al.*, 1989; Piegorsch *et al.*, 1992) was not observed (data not shown).

To test the assumption that chemicals that cause an increase in tumors of both sexes of one species are more likely than single-sex carcinogens to increase tumors in the other species, we calculated the positive predictive value from mice to rats and rats to mice for chemicals which cause tumors in one sex only and those causing tumors in both sexes (Fig. 4). From mice to rats the PPV for single-sex carcinogens is 57% and for chemicals causing tumors in both male and female mice the PPV is 88%. The results are equally impressive for the rat to mouse comparison. The single-sex PPV is 50%, while the PPV for chemicals causing tumors in both sexes is 84%. This also confirms the unpublished conclusion of Byrd *et al.* (1991).

Evaluation of chemicals which cause tumors to appear more rapidly in treated than in control animals found similar results whether reduced latency was defined as the appearance of tumors in treated animals 4, 13, or 26 weeks earlier than controls. Figure 5 compares the PPV for chemicals which caused tumors earlier in treated animals 13 weeks or more than controls in either one or both dose groups to chemicals which did not reduce tumor latency but were still positive. Clearly, reduced latency is associated with much higher PPV across species. For mouse to rat comparison PPV jumped from 51% with no reduced latency to 87% with reduced latency in one dose group. Similarly, the rat to mouse PPV went from 46% with no reduced latency to 85% with 13 weeks or more reduced latency in one dose group. There is a small caveat here; this definition of reduced latency, which is convenient for our analysis, is coupled with the definition of increased carcinogenic response. However, this coupling is unimportant because of the next test noted below.

The one factor which does not seem to influence cross-species prediction of carcinogenic response is the magnitude of the increase in tumors. Figure 6 shows that the PPV from mice to rats is essentially the same whether a chemical causes an increase in tumors with a dose response significant at $P < 0.05$, $P < 0.025$, or $P < 0.01$. The same is true for prediction from rats to mice. From mouse to rat the PPVs are 77, 78, and 81%, respectively. A slight increase is also seen for rat to mouse predictions at 71, 73, and 76%. However, all the PPV differences are within the uncertainty of PPV evaluation.

To investigate whether the presence of a clear dose–response relationship increased the cross-species prediction of carcinogenic response, we looked at predictive values for those chemicals causing tumors with

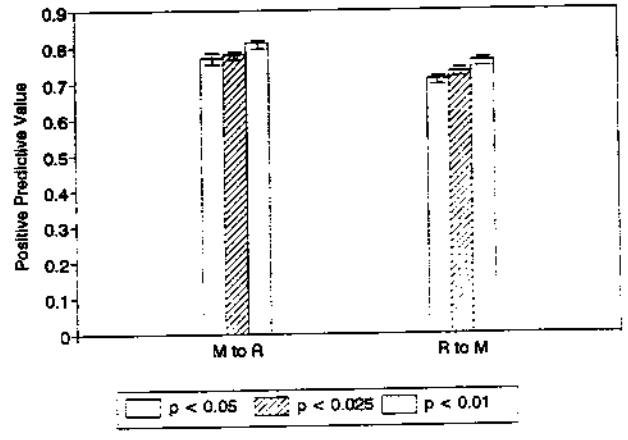

FIG. 6. Cross-species prediction of cancer response for chemicals demonstrating dose–response trends with different levels of significance.

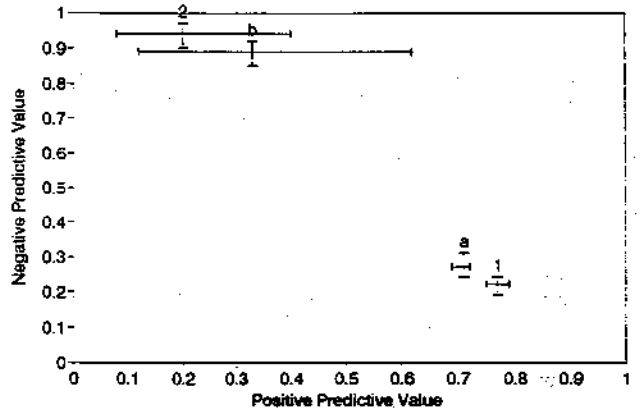

FIG. 7. Difference in cross-species prediction of carcinogenicity between chemicals demonstrating statistically significant dose–response relationships (1, mouse to rat; a, rat to mouse) and those with only significant pairwise comparisons to controls (2, mouse to rat; b, rat to mouse).

a statistically significant dose–trend test compared to those which caused a statistically significant Fisher exact test with at least one dose group but did not have a significant trend. The group of chemicals that cause increased tumor rates by the Fisher exact test only is quite small, increasing the size of the error bars, but comparison with those causing dose-related increases finds a much higher cross-species PPV for those chemicals with significant dose trends (Fig. 7). It is interesting to note the increase in negative predictive value in the group of chemicals which are positive only by the Fisher exact test.

Finally, although it is not an explicit criterion in the EPA proposed guidelines, it is generally believed that there are important differences in carcinogenic potential between mutagenic and nonmutagenic chemicals (Ashby and Tennant 1988, 1991; Butterworth, 1989). In this analysis mutagenic chemicals clearly have better PPVs than all chemicals and nonmutagenic chemicals have worse PPVs (Fig. 8). From mice to rats the PPV for mutagens was 84% and for nonmutagens was 71%. The PPV from rats to mice was 77% for mutagens and 67% for nonmutagens.

DISCUSSION

Our investigation of factors which EPA believes "can contribute to the judgment whether animal responses indicate a significant carcinogenic hazard to humans" finds that most do indeed increase cross-species prediction of carcinogenicity. As might be expected, the biological intuition, informed by years of experience, which underlies the list of important factors is a good guide. Many of the factors we have investigated have long been believed to increase the confidence of cross-species generalization of carcinogenic hazard. For example, in addition to EPA, the Food and Drug Administration has maintained for many years that chemicals which increase the rates of rare tumors in animal bioassays are to be considered potentially greater human hazards than those which increase only common tumors. Again, the empirical basis for this claim, prior to this study, is unclear.

In the present study we find that there are indeed characteristics of chemicals, or of the tumors that they induce, which increase the cross-species PPV for a carcinogenic response. These include induction of uncommon tumor type(s), tumors at multiple sites, and tumors in both sexes of one test species. In addition, we find that chemicals causing a tumorigenic response characterized by reduced latency in tumor formation or causing a clear dose-related increase in tumor rate, rather than simply an increase by pairwise comparison but no dose response, have increased positive cross-species predictive power.

Interestingly, the overlap of chemicals between these factors is quite high. For example, 102 of 112 chemicals which induce tumors at multiple sites were found to induce tumors in both sexes and 103 of that 112 had a significant dose trend. This leads to the possibility that the factors investigated are not independent descriptions of a group of chemical carcinogens. For instance, we find, as noted by Ashby and Tennant (1991), that mutagenic chemicals are more likely than nonmutagens to be multiple-site carcinogens. In the same way, mutagens are also associated with rare tumors much more than nonmutagens. Identification of the correlation between the many factors investigated in this study could yield one (or more) attribute that increases cross-species PPV. If a single (or perhaps two or three) attribute of a chemical could be generally agreed to signal likely cross-species prediction of carcinogenic response, then attention could be focused on understanding the cross-species generalizability of animal carcino-

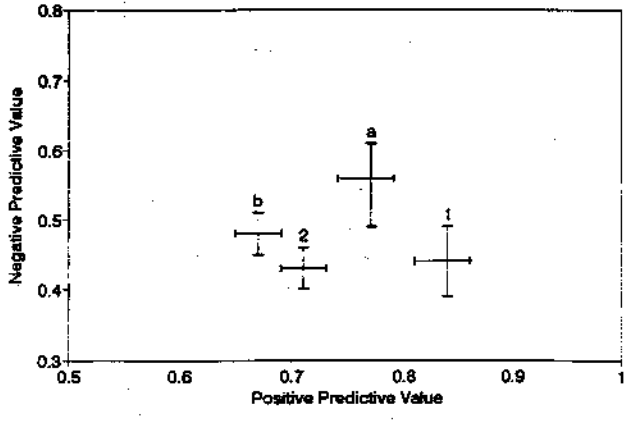

FIG. 8. Cross-species concordance of carcinogenic response for mutagenic (1, mouse to rat; a, rat to mouse) and nonmutagenic chemicals (2, mouse to rat; b, rat to mouse).

gens lacking the attribute(s). This could also have important consequences for carcinogen classification.

We had expected that there would be an increase in cross-species prediction for chemicals which cause a large tumor response, characterized by a dose–response trend positive at $P < 0.01$ compared to those with the standard $P < 0.05$ response. This we expected because it is far less likely that a chemical which displays this strength of response would be judged carcinogenic for spurious experimental reasons. However, our analysis shows no difference between the two groups.

Several aspects must be considered in the use of these factors to interpret possible human carcinogenicity. First, we only analyze cross-species concordance of high-dose, lifetime exposure. It is possible that some of these factors would not influence generalization across species under different exposure conditions. Second, we are examining average results for many chemicals without specific information on potential mechanisms of action. In cases in which mechanistic information is available it may provide additional evidence for the cross-species generalizability (or lack thereof) of carcinogenic response. Our goal is to understand in general the response of chemicals in different species with the hope of informing inferences about the large number of chemicals which may lack mechanistic data (Weisburger, 1990). We have also relied on a strictly statistical measure to classify carcinogens as positive or negative. It is clear that within the NTP program there are tumor responses that are statistically significant but considered not biologically significant, just as there are nonstatistically significant responses sometimes considered biologically significant. Our choice was to remove as much subjective influence as possible given that the many different chemicals had been reviewed by different peer review panels over the years. We believe that it is unlikely that our method of defining carcinogens leads to a significantly different result than would another method.

In conclusion, this investigation finds support for many of the factors stated as influencing cross-species prediction of carcinogenic response by the EPA (1994). All of the factors tested, with the exception of increased magnitude of response, significantly improved the positive predictive value of carcinogenic response between rodents in the NTP bioassay program. Higher concordance between rodent species might indicate higher concordance between human and rodent species. Future work will investigate the cross-species prediction of a protective effect against carcinogenicity. Many of the chemicals tested in the NTP bioassay program cause a statistically significant decrease in tumor rate in some sites. We will investigate whether this effect is concordant across species and whether there are factors which influence the prediction of protection.

ACKNOWLEDGMENTS

This work was conducted under EPA Cooperative Agreement 822917-01-0. Dr. Edmund Crouch wrote the original computer programs and gave us invaluable help in using them. We benefited greatly from discussions with Drs. Rita Schoeny, John S. Evans, and Lorenz Rhomberg.

REFERENCES

Ashby, J., and Tennant, R. W. (1988). Chemical structure, salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. *Mutat. Res.* **204,** 17–115.

Ashby, J., and 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.

Butterworth, B. E. (1989). Nongenotoxic carcinogens in the regulatory environment. *Regul. Toxicol. Pharmacol.* **9,** 244–256.

Byrd, D. (1988). New statistical approaches to the quantitative interpretation of toxicology data. *J. Am. College Toxicol.* **7,** 559–563.

Byrd, D., Crouch, E. A. C., and Wilson, R. (1990). Do mouse liver tumors predict rat tumors: A study of concordance between tumors induced in different sites in rats and mice. In *Mouse Liver Carcinogenesis: Mechanism and Species Comparisons,* p. 19–41. A. R. Liss, New York.

Byrd, D., Crouch, E. A. C., and Wilson, R. (1991). Organ specificity of rodent tumor induction: A test of statistical and graphical methods by examination of tumor induction from ingestion of selected substances. In *Risk Analysis* (C. Zervos, Ed.). Plenum Press, New York. [This was the written report of a poster presentation at the meeting of the Society for Risk Analysis, Washington, DC, November 1988, where 30 different plots were presented.]

Crouch, E. A. C. (1983). Uncertainties in interspecies extrapolations of carcinogenicity. *Environ. Health Perspect.* **50,** 321–327.

Crouch, E. A. C., and Wilson, R. (1979). Interspecies comparison of carcinogenic potency. *J. Toxicol. Environ. Health* **5,** 1095–1118.

Gart, J. J., Krewski, D., Lee, P. N., Tarone, R. E., and Wahrendorf, J. (1986). *Statistical Methods in Cancer Research: Volume III—The Design and Analysis of Long-Term Animal Experiments,* IARC Scientific Publication 79. IARC, Lyon, France.

Gold, L. S., Bernstein, L., Magaw, R., and Slone, T. H. (1989). Interspecies extrapolation in carcinogenesis: Prediction between rats and mice. *Environ. Health Perspect.* **81,** 211–219.

Gold, L. S., Manley, N. B., and Ames, B. N. (1992). Extrapolation of carcinogenicity between species: Qualitative and quantitative factors. *Risk Anal.* **12,** 579–587.

Haseman, J. K., and Huff, J. E. (1987). Species correlation in long-term carcinogenicity studies. *Cancer Lett.* **37,** 125–132.

Haseman, J. K., Huff, J. E., Zeiger, E., and McConnel, E. E. (1987). Comparative results of 327 chemical carcinogenicity studies. *Environ. Health Perspect.* **74,** 229–235.

Huff, J., Haseman, J., and Rall, D. (1991). Scientific concepts, value, and significance of chemical carcinogenesis studies. *Annu. Rev. Pharmacol. Toxicol.* **31,** 621–652.

International Agency for Research on Cancer (IARC) (1987). Overall evaluations of carcinogenicity: An updating of IARC monographs volumes 1 to 42. In *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Supplement 7.* World Health Organization, Lyon, France.

Parodi, S., Taningher, M., Baero, P., and Santi, L. (1982). Quantitative correlations amongst alkaline DNA fragmentation, DNA covalent binding, mutagenicity in the Ames test and carcinogenicity for 21 compounds. *Mutat. Res.* **93,** 1–24.

Piegorsch, W. W., Carr, G. J., Portier, C. J., and Hoel, D. G. (1992). Concordance of carcinogenic response between rodent species: Potency dependence and potential underestimation. *Risk Anal.* **12,** 115–121.

Purchase, I. F. H. (1980). Inter-species comparisons of carcinogenicity. *Br. J. Cancer* **41,** 454–468.

United States Environmental Protection Agency (EPA) (1986). *Guidelines for Carcinogen Risk Assessment,* 51FR33994. EPA, Washington, DC.

United States Environmental Protection Agency (EPA) (1994). *Draft Revisions to the Guidelines for Carcinogen Risk Assessment,* EPA/600/BP-92/003. Office of Research and Development, EPA, Washington, DC.

Weisburger, E. K. (1990). Mechanistic considerations in chemical carcinogenesis. *Regul. Toxicol. Pharmacol.* **12,** 41–52.

Zeise, L., Crouch, E. A. C., and Wilson, R. (1984). Use of acute toxicity to estimate carcinogenic risk. *Risk Anal.* **4,** 187–199.

TABLE 1
Summary of Responses in the CBDS Database[a]

	No. of chemicals	%
Chemicals tested on both mouse and rat	317	
Chemicals positive in either mouse or rat	268	85
Chemicals positive in both species	155	49
Chemicals positive in single site	156	49
Chemicals positive in two sites	60	19
Chemicals positive in three or more sites	67	21
Chemicals positive in single sex of one species	118	37
Chemicals positive in both sexes of one species	150	47

[a] Weak evidence counted as positive.

positive if the Fisher's exact test $P < 0.025$ for any one dosed group of $P < 0.05$ for two dosed groups. A weak response was defined if the criteria for positive were not met but a dose-trend test with $P < 0.025$ was observed. Otherwise, the response was negative. In our investigations the weak responses can be assigned to either the positive or the negative response category. The weak was included as positive throughout this analysis. A summary of the responses in the database is provided in Table 1.

The same tumor classification system used by Byrd *et al.* (1990) was implemented in this work. There are a total of 102 classifications which include both tumor types and tumor sites (shown in Table 2). Among the 102 classifications, 66 were classified as malignant primary neoplasms and 36 as benign primary neoplasms. This classification excludes metastatic lesions or unassignable lesions as to primary or metastatic.

A total of 317 chemicals tested on both mouse and rat satisfy all of the requirements. Further statistic details of these chemicals are shown in Table 1. Within the 317 substances that met our criteria, 109 are mutagens and 208 are nonmutagens. Data on mutagenicity was provided by Zeiger (1994).

Data Analysis

To investigate the factors identified by EPA as potentially increasing the likelihood of a "serious" carcinogenic effect, we investigated interspecies and intraspecies concordances grouping chemicals in the CBDS to correspond to the different factors. In the context of this paper, concordance refers to the agreement of tumor induction at any site between different species/strain/route/sex combinations in response to the same chemical substance. Higher concordance or positive predictive value indicates a better chance of observing a response in one group when it is already found in another group. The obvious notion is that a good interspecies concordance in animals may imply a good concordance between animals and humans (Crouch and Wilson, 1979).

Tumor responses were analyzed on the basis of 2 × 2 concordance tables. A sample concordance table is shown in Table 3. The columns and rows represent the same or different groups of tumor site/type classifications in one or another combination of species, strains, routes, and sexes. The numbers in the cells designated A–D represent numbers of tested chemical substances that had positive or negative response for an animal group (species or sex). Perfect concordance would be defined as B = C = 0, meaning that a tumor found in the row selection always predicts a tumor in the column selection and not finding a tumor in the row selection guarantees not finding a tumor in the column selection. Throughout our analysis, a chemical went to the positive cells for a species when it caused a tumor in at least one sex and in at least one site.

A concordance plot (Byrd *et al.*, 1991) was used to illustrate both PPV and NPV between two groups of tumor site/type classifications. The PPV and NPV are defined as PPV = A/(A + B) and NPV = D/(C + D). Confidence limits for PPV and NPV were obtained by treating each value as binomial samples from an infinite population of substances. The reported error bars represent a confidence interval corresponding to a 68%, one-sided binomial distribution (approximately 1 SD).

Definitions

It was required that we interpret many of the EPA factors to conduct the analysis. Within the CBDS the following definitions were used for the EPA factors examined (EPA statement first, our definition following). The exact EPA statement is as follows: "Several kinds of observations from animal studies can contribute to the judgment whether animal responses indicate a significant carcinogenic hazard to humans. . . . The following observations add significance:"

Uncommon tumor types. Chemicals which cause an increase in a type of tumor with a historical background rate lower than 2% in control animals. Common tumors are defined as those with a background rate higher than 10% for a particular species/strain/sex. About 80% of the 102 classifications were considered rare tumors.

Tumors at multiple sites. Chemicals causing increase in tumors in two sites or three or more sites in at least one group (sex, species) of treated animals were examined. These are compared to chemicals which induce tumors in a single site.

Tumors by more than one route of administration. The CBDS has too few chemicals with multiple tests via different exposure routes to examine this factor.

Tumors in multiple species, strains, or in both sexes. Chemicals which cause tumors in both species in the bioassay have perfect predictive value in our analysis

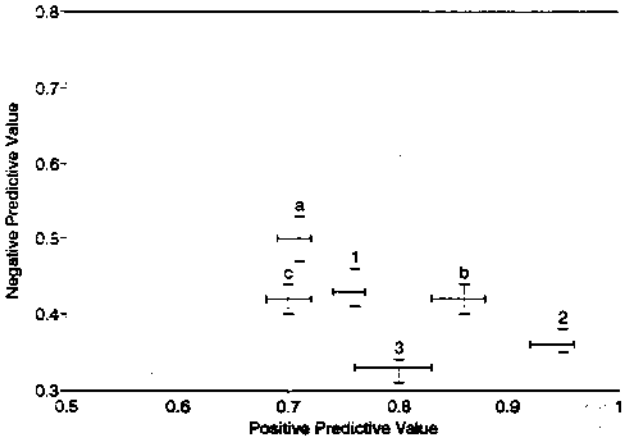

FIG. 1. Concordance of carcinogenic response between mice and rats for any tumor type (1, mouse to rat; a, rat to mouse), from rare tumors to any tumor (2, mouse to rat; b, rat to mouse), and from common tumors to any tumor (3, mouse to rat; c, rat to mouse).

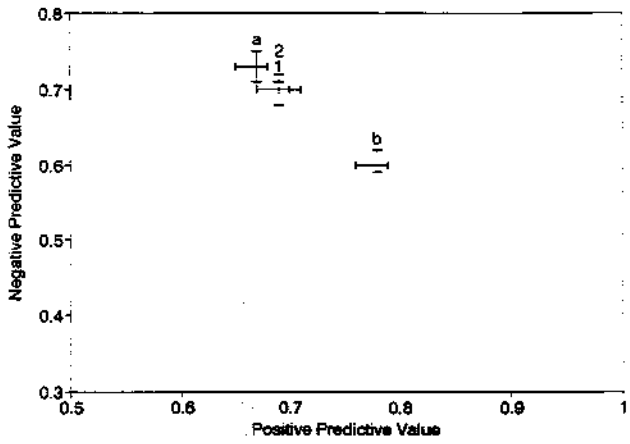

FIG. 2. Concordance of carcinogenic response within a species. Male mouse to female mouse, 1; female mouse to male mouse, 2. Male rat to female rat, a; female rat to male rat, b.

wise comparison versus those with clear dose-related increases in tumor rate. Chemicals which show a $P \leq 0.05$ Cochran–Armitage test for dose response are compared to those which are only significant at $P \leq 0.05$ by Fisher exact test for at least one dose group but do not have a significant dose trend.

RESULTS

We find, similar to other investigators discussed previously, that a positive response in mice is associated with a positive response in rats 76% of the time. The positive predictive value from rats to mice is approximately 71% (shown in Fig. 1). This can be taken as a general predictive rate across species in the high-dose chronic rodent bioassay. It has been noted that chance alone would lead to an approximately 43% PPV from rats to mice and 48% from mice to rats (Gold *et al.*, 1989). Positive predictive value from male mice to female mice, and vice versa, is essentially identical (69%). However, a difference is observed for rats depending on whether the prediction is from response in male rats to female rats (67%) or from female rats to males (78%) (Fig. 2).

First we examine whether chemicals increasing rates of rare tumors are more concordant across species than all chemicals or than those increasing common tumors. We find that chemicals increasing rates of rare tumors (sometimes along with other tumor sites) have better predictive value across species than those increasing only common tumors. Figure 1 also shows a plot of positive and negative predictive values for chemicals increasing rare tumors and those increasing common tumors. Those increasing rare tumors have a statistically significantly greater predictive value than all chemicals (95 vs 76% for mice to rats; 86 vs 71% for rats to mice) or those increasing common tumors (95 vs 80% for mice to rats; 86 vs 70% for rats to mice). This confirms an unpublished conclusion by Byrd *et al.* (1991).

We next investigated whether there is a difference in positive predictive value of carcinogenic response across species for chemicals which cause an increase in tumor rate at one, two, or three or more sites in the predicting species. Figure 3 shows that chemicals which cause an increase in tumors at multiple sites (either two or three or more) have much higher mouse to rat positive predictive value than chemicals causing an increase at a single site (96–98 vs 57%). The same is true for the rat to mouse prediction (85–90 vs 56%). Interestingly, there did not seem to be much difference in predictive value between chemicals causing increased tumor rates in two organs and those causing increases in three or more. It appears that for either

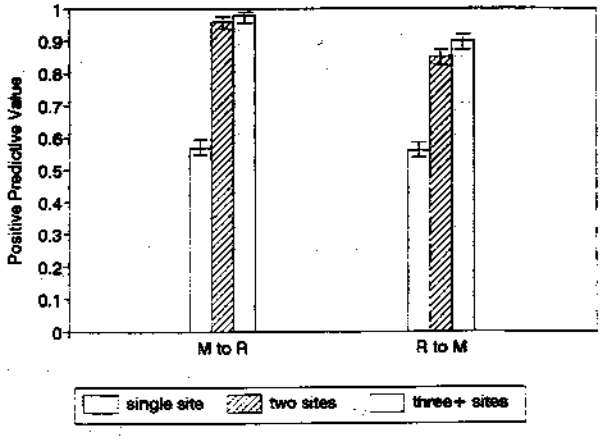

FIG. 3. Prediction of carcinogenic response across species for chemicals inducing tumors in one, two, and three or more sites.