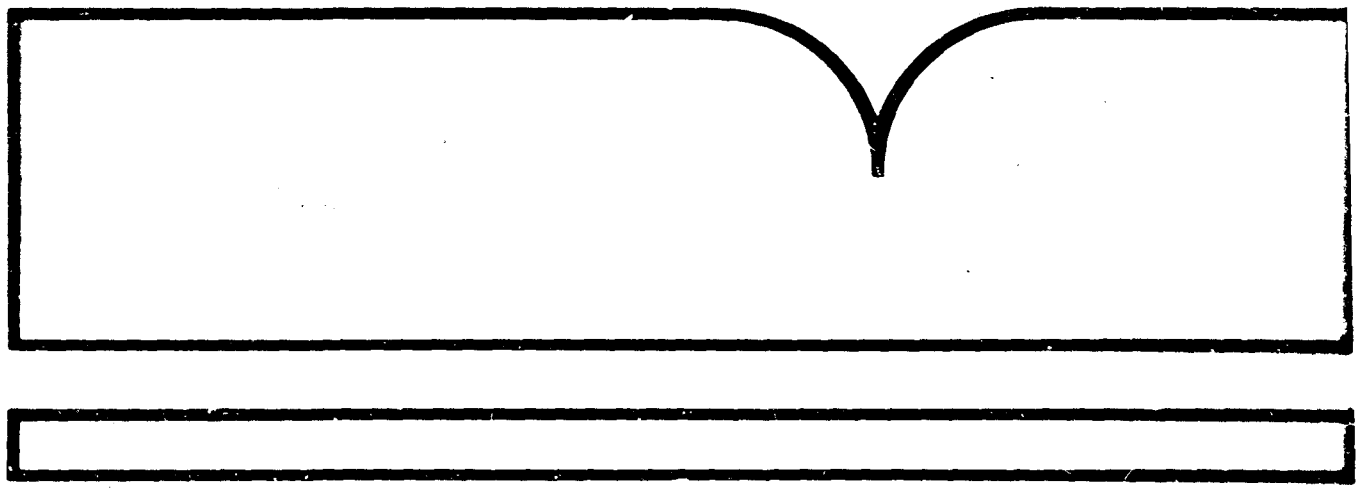
EPA 600-6-87-007a

PB88-127105

INVESTIGATION OF CANCER RISK ASSESSMENT METHODS: SUMMARY

Clement Associates, Incorporated
Ruston, LA

Sep 87



TECHNICAL REPORT DATA *(Please read Instructions on the reverse before completing)*	PB88-127105
1. REPORT NO. EPA/600/6-87/007a	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Investigation of Cancer Risk Assessment Methods: Summary	5. REPORT DATE September 1987 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Bruce C. Allen, Annette M. Shipp, Kenny S. Crump, Bryan Kilian, Mary Lee Hogg, Joe Tudor, Barbara Keller	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clement Associates, Inc. 1201 Gaines Street Ruston, LA 71270	10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO. 68-01-6807
12. SPONSORING AGENCY NAME AND ADDRESS Office of Health and Environmental Assessment Carcinogen Assessment Group (RD-689) U.S. Environmental Protection Agency Washington, DC 20460	13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA/600/21
15. SUPPLEMENTARY NOTES EPA Project Officer: Chao Chen, Carcinogen Assessment Group Office of Health and Environmental Assessment, Washington, DC (382-5719)	

16. ABSTRACT The major focus of this study is upon making quantitative comparisons of carcinogenic potency in animals and humans for 23 chemicals for which suitable animal and human data exists. These comparisons are based upon estimates of risk related doses (RRDs) obtained from both animal and human data. An RRD represents the average daily dose per body weight of a chemical that would result in an extra cancer risk of 25%. Animal data on these and 21 other chemicals of interest to the EPA and the DOD are coded into an animal data base that permits evaluation by computer of many risk assessment approaches.

This report is the result of a two-year study to examine the assumptions, other than those involving low dose extrapolation, used in quantitative cancer risk assessment. The study was funded by the Department of Defense [through an inter-agency transfer of funds to the Environmental Protection Agency (EPA)], the EPA, the Electric Power Research Institute and, in its latter stages, by the Risk Science Institute.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
Distribute to public	Unclassified	69
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

DISCLAIMER

This document has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. The information in this document has been funded by the U.S. Environmental Protection Agency, the Department of Defense (through Interagency Agreement Number RW97075101), the Electric Power Research Institute, and the Risk Science Institute.

CONTENTS

ILLUSTRATIONS

TABLES

BRIEF SUMMARY OF MAJOR CONCLUSIONS

The major focus of this study is upon making quantitative comparisons of carcinogenic potency in animals and humans for 23 chemicals for which suitable animals and human data exists. These comparisons are based upon estimates of "RRDs" obtained from both animal and human data. An RRD represents the average daily dose per body weight of a chemical that would result in an extra cancer risk of 25%. Animal data on these and 21 other chemicals of interest to the EPA and the DOD are coded into an animal data base that permits evaluation by computer of many risk assessment approaches.

The major findings of this study are as follows:

1. Animal and human RRDs are strongly correlated. The knowledge that this correlation exists between animal and human carcinogenicity data should strengthen the scientific basis for cancer risk assessment and cause increased confidence to be placed in estimates of human cancer risk made from animal data.

2. In the majority of cases considered, analysis methods for bioassay data that utilize lower statistical confidence limits as predictors yield better predictions of human results than do the same methods using maximum likelihood estimates.

3. Analysis methods for animal data that utilize median lower bound RRDs determined from the ensemble of data for a chemical generally yield better predictions of human results than analyses that utilize minimum RRDs calculated from all the studies available.

4. Use of the "mg intake/kg body weight/day" (body weight) method for animal-to-human extrapolation generally causes RRDs estimated from animal and human data to correspond more closely than the other methods evaluated, including the "mg intake/m^2 surface area/day" (surface area) method.

5. The risk assessment approach for animal data that was intended to mimic that used by the EPA underestimates the RRDs (equivalent to overestimating human risk) obtained from the human data in this study by about an order of magnitude, on average. However, it should be understood that the risk assessment approaches implemented in this study are computer automated and do not always utilize the same data or provide the same result as the EPA approach.

6. Reasonable risk analysis methods can be defined for the chemicals in this study that reduce the residual loss (roughly the average multiplicative factor by which the RRD predictors obtained from the animal data are inconsistent with the ranges of human RRDs consistent with the human data) to 1.7. This is not the same as saying that the predictors are accurate to within a factor of 1.7, because the estimated ranges of human RRDs that are consistent with the human data cover an order of magnitude or more for most chemicals.

7. It has been possible to identify a set of analysis methods using the median lower bound estimates that are most appropriate for extrapolating risk from animals to humans, given the current state of knowledge and data analysis. It is possible to use the information and results presented in this investigation to calculate ranges of risk estimates that are consistent with the data and also incorporate many uncertainties associated with the extrapolation procedure.

8. The many components of risk assessment are interrelated and evaluation of risk assessment methods should focus on the complete risk assessment process rather than on individual components.

9. The data base and methods used in this study can provide a useful basis for evaluating various risk assessment methods.

This study only compared human and animal results for a relatively high risk level. It did not examine the uncertainty inherent in the low dose extrapolation process.

INTRODUCTION

This report is the result of a two year study to examine the assumptions, other than those involving low dose extrapolation, used in quantitative cancer risk assessment. The study was funded by the Department of Defense [through an interagency transfer of funds to the Environmental Protection Agency (EPA)], the EPA, the Electric Power Research Institute and, in its latter stages, by the Risk Science Institute. The objectives of the study are as follows:

1. To identify and express quantitatively uncertainties that are involved in the process of risk estimation, excluding the uncertainties in the low dose extrapolation model;
2. To examine the impact of the different assumptions that are made in risk estimation;
3. To compare results calculated from human and animal data, including the identification of the assumptions that produce the best correlation of risk estimates between humans and animals;
4. To develop guidelines for presenting a range of risk estimates based on different but scientifically acceptable assumptions or

assumptions that have considerable backing in the scientific community.

These objectives are pursued using empirical methods in which carcinogenicity data for 44 chemicals are analyzed systematically in a variety of ways. Particular attention is placed on those 23 chemicals for which there exist data from both animal and human studies suitable for making quantitative comparisons.

Table 1 contains a list of components of a quantitative risk assessment based upon animal data. Each component requires a decision on the part of the risk assessor for which there is no unique "correct" choice. Also listed in Table 1 are various possible approaches to each component. The choices that a risk assessor makes for these components affect the resulting estimates of risk. The choices for these components therefore are related to the uncertainty in assessment of risk from animal data.

Objective 2 is pursued by making different risk estimates for the 44 chemicals in the study by systematically varying the approaches to the components listed in Table 1. Examination of the distributions of the changes in the estimates associated with different approaches to the various components permits the examination of the impact of the various approaches (assumptions). These distributions also relate to the uncertainties in the process of risk estimation, so this work also applies to Objective 1.

A major part of the study involves making comparisons between risk estimates derived from animal data and those derived from human data for those 23 chemicals for which suitable data are found to exist for both animals and humans. This work addresses the question of whether correlations exist between animal and human data and therefore is of

fundamental importance to the scientific validity of quantitative risk assessment. The practice of making quantitative estimates of human risk from animal data is based upon the hypothesis (heretofore essentially untested) that such correlations do in fact exist. If quantitative correlations can be shown to exist, then these correlations can provide a stronger scientific basis for risk assessment. Further, evaluation of the correlations and determination of those approaches to the components listed in Table 1 that produce the best correlations can suggest better risk assessment methods and assist in evaluating and presenting the uncertainty in risk estimates derived using those methods, in accordance with Objectives 3 and 4.

DATA BASE

At the beginning of the project EPA provided a list of 40 chemicals for inclusion in the project that are of interest to the agency. This list was supplemented by adding additional chemicals for which suitable quantitative data are available from both animal and human studies and deleting a few chemicals from the original list for which suitable animal bioassay data could not be located, which brought the total number of chemicals studied to 44 (Table 2). The first step in the project was to collect the relevant carcinogenicity data from the literature on each of these chemicals. Initially the data collection included information on pharmacokinetics, metabolism, and mutagenicity in addition to that on carcinogenicity, and data on these topics was collected for several chemicals. However, collection of this information was discontinued early in the project due to resource limitations.

Data Matrix: An intensive search was made for animal or human carcinogenicity studies on these chemicals. Sources searched include our company's files, computerized data bases (Medline, Chemical Exposure, Biosis, Embase, and NTIS), publications of governmental and other

official organizations (IARC monographs, EPA health assessments and similar documents, and NCI and NTP technical reports) and a carcinogenicity data base complied by Gold *et al.* (1). The relevant articles were obtained and summary information extracted from them was coded into a computerized data base called the Data Matrix.

The Data Matrix includes information on species, sex, route of exposure, length of exposure, length of observation, whether a positive carcinogenic response was observed and whether a data set is suitable for quantitative risk estimation. Data sets on animal studies that satisfy this latter condition are coded into a more detailed data base called the Animal Data Base. A list of the chemicals included in the study, the number of carcinogenicity data sets summarized in the Data Matrix, and the number of those in various categories that are coded into the Animal Data Base is given in Table 2. As can be seen from this table, a total of 1233 data sets (a data set is generally composed of all the dose response data from a given sex and species of animals exposed via a reasonably common protocol in a study) from 736 studies (a study generally consists of all of the data in a single primary reference) are summarized in the Data Matrix.

Animal Data Base: All of the bioassays that are considered to be at least minimally acceptable for quantitative risk estimation are coded into the computerized Animal Data Base. The criteria that a data set needs to satisfy for inclusion are as follows:

- the test species is a non-human mammalian species;
- the protocol includes matched controls, preferably vehicle (or sham inhalation) treated animals;
- dosing is consistent within a dose group, with dosages and dosing pattern clearly stated;

- a single route of exposure is employed (early in the project it was decided not to continue to code experiments that exposed the animals by skin painting or subcutaneous injection; therefore the Data Base is not complete with respect to these routes of exposure);
- the test compound is administered alone or in an acceptable vehicle, without pretreatment or concurrent treatment of any kind;
- tumor incidence is reported as number of tumor-bearing animals as opposed to number of tumors.

Table 2 provides a summary of the data included in the Animal Data Base for each of the 44 chemicals. For these chemicals, 631 data sets are included in the Data Base.

The data are coded into the data base in sufficient detail to permit a wide range of analyses to be applied to the data, including analyses that evaluate the approaches listed in Table 1. Included in the data base is the following information, whenever available: species, strain, and sex; weight data; food intake data; detailed exposure protocol including route and time pattern of exposure; initial number of animals per dose group; numbers of animals per dose group having various tumor responses (see below) and number per dose group examined for each tumor response; time until first development of each tumor type coded and number in each dose group alive at this time. Within a single data set, the following tumor responses are coded, whenever possible:

- those that occur significantly more often in any dosed group compared to the control group;
- the tumor type most nearly significant, in cases in which none are significant;
- the combination of all significantly increased tumors;

- the combination of all significantly increased malignant tumors;
- all tumors[1];
- all malignant tumors[1];
- the tumor considered to be the response of interest in humans (if known).

Early in the study individual animal pathologies were coded whenever possible, which would make possible time-to-tumor analyses. However, this work was discontinued due to limited resources after such data had been coded for about about five chemicals.

Selection of Chemicals for Animal-Human Comparisons: For a chemical to be included in the analyses comparing results in animals and humans, data had to be available from both human and animal studies that would support the quantitative comparisons conducted and for which reasonably strong positive evidence of carcinogenicity exists in either the animal or the human data. A list of the chemicals satisfying these requirements and which are therefore included in the comparative analyses is presented in Table 3. Thirteen industrial chemicals are included in this list, seven drugs, a food contaminant (aflatoxin), a food additive (saccharin), and tobacco smoke.

It is neither necessary nor sufficient that a chemical be unequivocally carcinogenic in humans in order to be included. Thus, a chemical such as saccharin, which has been associated with cancer only in laboratory rodents, is included while bis(chloromethyl) ether is not included, even though sufficient evidence apparently exists to establish that bis(chloromethyl) ether is carcinogenic in humans (2). The reasons such

[1] Interstitial cell tumors of the testes in male F344 rats, mammary gland benign tumors in female Sprague-Dawley rats, malignant lymphomas in AKR and AKR/J mice, and mammary tumors in MTV+ mice are not included in these groups. These tumors have a very high background rate of occurrence in the indicated species, which would tend to obscure dose-related effects at other sites.

chemicals are not included generally relate to limitations regarding the data on human exposures. Of the 23 chemicals or chemical groups that IARC considered in 1982 to have "sufficient" evidence of human carcinogenicity, 11 are included in this study. Twelve other chemicals are included; three are considered to provide "limited" evidence, eight to provide "inadequate" evidence in support of human carcinogenic effects, and cigarette smoke has not been formally evaluated by IARC.

It was considered important that the study not be limited to chemicals whose carcinogenicity in humans has been firmly established. One of the ultimate goals of the study is to compare the predictions of carcinogenic potency of chemicals derived from animal data with the corresponding potency in humans. If such comparisons are restricted to confirmed human carcinogens, the ability of the animal data to predict human results might be overestimated. The same would be true if the study is restricted to confirmed animal carcinogens. Although a similar study by the National Academy of Sciences was restricted to confirmed human carcinogens, the authors recognized the potential for bias in this approach (3).

A thorough search was conducted for useful epidemiological data on the chemicals selected. Individual researchers were queried regarding unpublished data that would be helpful in our analyses, possible updates of their work and, particularly, additional information on exposure.

ANALYSIS OF EPIDEMIOLOGICAL DATA

Calculation of Risk Related Doses (RRDs): The epidemiological data on the 23 chemicals in Table 3 vary greatly in format and quality. Three distinct types of studies are represented: prospective cohort studies (including clinical trials), case-control studies, and (in the case of

aflatoxin) a cross-sectional comparison of cancer rates and levels of exposure in different populations. Even within one of these categories, the individual studies differ considerably with respect to such factors as duration of exposure, latency, and methods for reporting results. Because of the wide variations in data from the epidemiological studies, a systematic, standardized method of recording the human data (like that developed for the bioassay data base) is not considered feasible. Instead, the epidemiologic data for each chemical is considered as a whole and risk estimates are developed using general guidelines whose purpose is to insure that, to the extent possible, the methodology 1) can be employed with a minimal amount of data, 2) makes best use of the data, and 3) ensures that risk estimates made from data of differing types and quality are comparable.

The majority of epidemiological studies considered are prospective studies. The minimum amount of information required for an analysis of a prospective study consists of a single group with known cumulative dose (expressed in ppm-years, for example) and observed and expected numbers of cancers. Additional information on observed and expected responses categorized by exposure group is accommodated whenever available and may provide better estimates of carcinogenic potency. Using the linear dose response model for relative risk of $RR = 1 + \beta d$, where d is cumulative dose, the potency parameter β is estimated by fitting this model to the epidemiologic data by the method of maximum likelihood. Comparable linear dose response approaches are applied to case control and cross-sectional epidemiological studies.

The parameter β is used in conjunction with a life table analysis that employs U.S. sex- and age-specific mortality rates for the cancer in question to estimate the "extra risk" of death by cancer from a specified human exposure pattern. Extra risk is defined as $(P - P_0)/(1 - P_0)$, where P is the lifetime probability of death from the cancer under

consideration in the presence of the postulated exposure and P_0 is the background lifetime probability in the absence of exposure. Extra risk may be interpreted as the probability of death from the cancer under consideration, given that without the exposure death would have been due to some other cause.

A constant daily exposure for 45 years beginning at age 20 is used as the reference human exposure pattern for the calculation of human risk. This pattern is taken as a compromise between the exposure patterns found in most of the epidemiological studies (which are of occupationally exposed cohorts for the most part), and constant lifetime exposure beginning early in life that is typical of animal bioassays. The endpoint estimated is the daily dose rate in mg/kg/day under this exposure pattern that will produce an extra risk of 0.25. This daily dose rate is called a "risk related dose" (RRD). Since the extra risk measured in most of the epidemiological studies is less than 0.25, estimation of RRDs will generally require extrapolation beyond the dose ranges of the epidemiological data. On the other hand, an extra risk of 0.25 can generally be measured directly in standard animal bioassays; consequently, use of 0.25 as a reference risk should make the analyses of the animal data robust with respect to the dose response model selected. The choice of a reference risk of 0.25 therefore represents a compromise designed to minimize the extrapolation required beyond the dose and response ranges in the animal and human studies.

Exposures in the epidemiologically studied cohorts are frequently the source of considerable uncertainty in the analyses. For example, exposures in occupational cohorts are often measured infrequently and those measurements that are made are sometimes of uncertain relevance to exposures of specific workers. It is considered to be important to quantify this uncertainty, although such quantification is difficult. The approach adopted is to estimate uncertainty factors that represent

our impression of the uncertainty of the dose estimates for any given study. These factors are applied to estimate upper and lower bounds for the exposures in the epidemiological studies. To promote uniformity in determining these factors, fairly specific guidelines for their calculations were adopted *a priori* and followed consistently for each chemical. A single investigator (B.A.) developed the bounds for each chemical and for each study. As additional studies were analyzed, the uncertainty bounds derived earlier were reviewed and occasionally revised. To minimize the possibility of unintentional bias, all of the analyses of the epidemiological data were performed independently of the analyses of the animal data.

The upper and lower bounds on exposures in the epidemiological cohorts are applied, along with statistical confidence limit procedures, to estimate upper and lower bounds for β. These bounds are then translated into upper and lower bounds for the RRD. The analysis of each epidemiological study therefore produced a best estimate RRD and corresponding lower and upper bounds, RRD_L and RRD_U, that reflect both the statistical uncertainty in the observed cancer responses in the epidemiological studies and the uncertainty in the exposure levels.

In many cases, more than one triple (RRD_L, RRD, RRD_U) for a chemical is available from the epidemiologic literature, either because of more than one study or more than one carcinogenic response analyzed. Rather than combining results for different responses or from different studies, a single triple is selected to represent the potency of a given chemical. The triple that is selected is one that corresponds best with the consensus of opinion about the carcinogenic effect of the chemical determined from all the literature reviewed. However, the results from a study or particular response in a study are not used if the dose-response model provided a poor fit to the data or if the study is deemed to be markedly inferior to other studies providing RRD estimates. In

the case of vinyl chloride, for example, a liver cancer response is chosen since angiosarcoma of the liver is considered to be undeniably linked to vinyl chloride exposure whereas respiratory cancer, another endpoint analyzed, is not so clearly linked. Another example is provided by isoniazid. Overall, the literature on isoniazid does not conclusively demonstrate its carcinogenicity in humans let alone indicate any particular site of action. Hence, the response selected is all malignant neoplasms, and, moreover, the triple chosen is one that has an infinite upper bound (consistent with no carcinogenic effect). Figure 1 displays the endpoints used to calculate human RRDs for each chemical.

ANALYSIS OF ANIMAL DATA

Two approaches for comparing the results of bioassay analyses to the estimates derived directly from the epidemiology were considered. First, correlation analyses were used to determine if the human carcinogenicity data are correlated at all in a quantitative sense with the animal data. These analyses involve the triples of RRDs derived from the human data and corresponding triples (RRD_{AL}, RRD_{A}, RRD_{AU}) obtained from the animal data. If the correlation analysis is positive, then it is reasonable to ask if particular RRD estimates obtained from animal data are good predictors of the results obtained directly from epidemiological studies. At this stage one can also examine the magnitude of errors, *i.e.* the uncertainty that results from the use of any predictor. Both correlation and prediction analyses require RRDs from animal data that are similar to those obtained from the epidemiological data.

Calculation of RRDs from Animal Data. For each carcinogenic response coded from a study testing the chemical of interest, a multistage model is fit to the dose-response data (4). The model is fit by an updated

When averaging is carried out at every level - over sex, study, and species (Analyses 12 through 24d) the averaging serves to define a unique triple for each chemical. For the remaining analyses, the collection of RRDs must be further condensed to obtain a unique triple for each chemical.

For analyses in which no averaging is conducted (Analyses 0 - 8c and 25), two predictors from the lower bounds on RRDs are selected: one, L_M, by taking the minimum of the lower bound RRDs, and the other, L_{2Q}, by taking the second quartile (median) of the lower bound RRDs, first within a species, and then taking the median of the species-specific medians. This approach to computing medians is similar to the method of averaging described above, and is designed to insure that different species contribute equally to the RRDs. The maximum likelihood RRDs and upper bound RRDs are similarly combined and consequently two different types of triples are produced: (L_M, MLE_M, U_M) and (L_{2Q}, MLE_{2Q}, U_{2Q}). For analyses in which only partial averaging is conducted (Analyses 9 -11b), the approach taken can be roughly described as the same as that just described for the case of no averaging, except applied to those RRDs remaining after the appropriate averaging process is complete. Thus two sets of triples from the animal data are produced for all analyses except those for which averaging is carried out at every level (Analyses 12 - 24d)

Data Sieve. In an effort to make the Animal Data Base as complete as possible, all data satisfying the minimal criteria listed earlier are included. This results in there being data of highly variable quality in the data base. In the analysis methods discussed thus far, no account is taken of the quality of the data; data from poorer studies (*e.g.* those using very few animals or observing the animals only for a short period of time) are treated the same as data from studies of

higher quality. To address this problem, a data sieve was designed such that, when applied, only higher quality data are used in an analysis.

The sieve is composed of two screens that can operate either separately or in tandem. The first, the significance screen, examines each data set for a statistically significant ($p < 0.05$) increase in responses at any treatment group over that in the control group by Fisher's exact test, or for a statistically significant dose-response trend by the Cochran-Armitage test. If at least one of the data sets for a chemical eligible for an analysis satisfies this condition, all data sets for that chemical not satisfying the condition are deleted from the analysis. If no data sets for a chemical satisfy the condition, then none of the data sets for that chemical are deleted on the basis of the significance screen.

The second screen, called the quality screen, screens on the basis of the length of observation and the number of dosed animals. Each data set is assigned a rank according to the scheme depicted in Table 6. All data sets assigned a rank that is higher than the lowest rank of any data set otherwise eligible for an analysis are excluded from the analysis.

The sieve is applied to the data sets that would otherwise be eligible for a particular analysis. When both screens are employed, the significance screen is applied first. The sieve is designed to select the best data sets pertaining to a chemical among those eligible for a particular analysis, but not to be the basis for the exclusion of any chemical from an analysis. Note in this regard that there is no way that use of either screen can cause all of the data for a chemical to be eliminated from an analysis.

INVESTIGATION OF COMPONENT-SPECIFIC UNCERTAINTY

The importance of individual components and choices for those components (listed in Table 1) to risk assessment are investigated by constructing histograms of the ratios RRD_x/RRD_{30} of RRDs obtained from animal data for the various chemicals, where RRD_{30} represents an RRD obtained from Analysis 30, and RRD_x represents an RRD obtained from an analysis that differs from 30 with respect to an approach to a single risk assessment component. Specifically, x is allowed to range over Analyses 31 to 50, as each of these differ from Analysis 30 only in the approach to a single component. Since human data are not required for this investigation, data for all 44 chemicals represented in the data base (Table 2) are utilized. Only median lower bound predictors (L_{20}s) are considered.

Table 7 summarizes results of this analysis by presenting modes and dispersion factors for the histograms. The dispersion factor is the average factor by which the ratios differ from the mode. A mode close to 1.0 indicates that the single approach that differs from that used in Analysis 30 makes little difference, on average, in the RRD obtained. A large dispersion factor indicates that the effect of the approach under consideration is highly chemical-specific.

The dispersion factors for Analyses 31 - 34, which differ from Analysis 30 only in the dose measure assumed for animal-human equivalence, are all relatively small. This indicates that changing this dose measure has about the same effect as multiplying RRDs by a fixed constant.

The modes associated with all of the other analyses fall in the interval [0.8, 1.25]. This suggests that the change from Analysis 30 encompassed in these analyses do not affect the RRD calculations much, on average.

Analyses 45 - 47 each differ from Analysis 30 only in the manner in which results from different studies are combined, and each is associated with a relatively small dispersion factor. This indicates that the manner used to combine data is relatively unimportant; all approaches considered give roughly comparable results.

The remaining analyses differ from Analysis 30 with respect to components that relate to length of study (Analysis 37), length of dosing (Analysis 38), exposure route (Analyses 37 and 38), tumor type to use (Analyses 41 - 44), and species to use (Analyses 49 - 50). These analyses are associated with larger dispersion factors, suggesting that there is greater uncertainty associated with these risk assessment components. This suggests that further research related to these components could reduce the overall uncertainty in risk assessment.

METHODS FOR COMPARISON OF ANIMAL AND HUMAN RESULTS

Correlation Analysis. This analysis is intended to determine whether RRDs derived from the animal data (the animal results) are correlated with those derived from the human data (the human results). The analysis of the individual epidemiological studies on each chemical have produced a "best" estimate of the RRD corresponding to a one-in-four risk, RRD_H, and upper and lower bounds on that dose, RRD_{HL} and RRD_{HU}, respectively. The interval $[RRD_{HL}, RRD_{HU}]$ represents the range of RRDs that are in some sense consistent with the epidemiological data, taking into account data uncertainty and statistical variability. A similar interval is required from the animal data to compare with that derived from the human data. The interval selected for this comparison is $[L_{2Q}, U_{2Q}]$, the medians of the lower and upper bounds on the RRDs estimated from the animal data.

A statistical test was conducted for each of the selected methods of bioassay analysis to determine if the RRDs estimated from animal data were significantly correlated with those estimated from human data. Specifically, the test determined whether the intervals defined by the upper and lower bounds for the human RRDs were significantly correlated with the corresponding intervals calculated from the animal data. A generalization of Spearman's rho statistic (5) was used that applies to intervals rather than individual points. In this statistic, the interval for one chemical was considered to rank higher than that for a second chemical if both the lower and upper bounds of the first interval were larger than the respective bounds for the second interval. The statistical significance of a particular analysis was evaluated by randomly reassigning the human intervals to chemicals while keeping the animal intervals assigned to the correct chemicals (a permutation test). The p-value of the statistical test represents the probability that, given the animal and human intervals calculated, a correlation as large or larger than that observed could have occurred by a random assignment of these intervals to chemicals.

Prediction Analysis. If the correlation analysis just discussed finds a positive correlation between the animal and human RRDs, it is reasonable to determine which particular estimates derived from the animal data best predict the results obtained directly from the epidemiological data, and to determine how well these estimates predict the animal results. The prediction analysis therefore selects a single estimator from the bioassay results as the estimate of RRD for each chemical. Four types of estimates are investigated: the minimum and median of the lower bound estimates (L_M and L_{50}) and the minimum and median of the maximum likelihood estimates (MLE_M and MLE_{50}).

are not distilled to a single point, but rather the
, and the interval [RRD_{HL}, RRD_{HU}] are used to
tors. In these evaluations, a straight line with
the base ten logarithmic transform of predictor
e human RRDs. Plots of these fits are produced with
lotted vertically and the predictors derived from the
ted on the horizontal axis. The unit slope insures that
hip estimated on the basis of the logarithmic transformed
linear relationship on the basis of the untransformed data.
lationship is equivalent to assuming that RRDs estimated from
ta are a constant multiple of the RRDs estimated from human

e fitting was accomplished by minimizing a loss function calcu-
on the basis of the animal and human RRDs, the straight line, and
s function. Three types of loss functions are considered. The
lest, called $DISTANCE^2$, is the squared vertical distance (on the log
le) from the interval [RRD_{HL}, RRD_{HU}] plotted on the vertical axis to the prediction line. If the prediction line passes through this interval the loss is taken to be zero. This loss function has two potential drawbacks: 1) it makes use only of the endpoints of the interval and does not take into account the best estimate, RRD_H; 2) it cannot be applied when the predictor RRDs can be infinite, as is the case when MLE_M and MLE_{2Q} are used as the predictors. Because of these drawbacks, and to evaluate how robust our conclusions are to our choice of loss function, two additional loss functions, CAUCHY and TANH are defined.

RESULTS OF ANIMAL AND HUMAN COMPARISONS

CORRELATION ANALYSES

Table 8 contains the correlation coefficients and their associated p-values corresponding to each of the initial 38 methods of analyzing the bioassay data studied. Figures 2 through 8 contain graphs of selected analyses. This summary reports only results from analyses that applied the data sieve described earlier. Use of the sieve gave a higher correlation in 28 of the 38 analyses and in each of the 10 exceptions the reduction in the correlation was marginal.

The results in Table 8 provide a strong indication of a positive correlation between the animal and human RRD estimates. Thirty-five of the 38 analyses had a p-value less than 0.05, indicating a statistically significant positive correlation between the animal and human RRDs. Fifteen of the analyses had a p-value of 0.0001 or smaller, including the Base Case analysis which attempts to mimic the analysis method used by the USEPA. Not only are the correlation coefficients statistically significantly positive, but they are sizable in an absolute sense as well. Twenty-six of the analyses yield a correlation coefficient larger than 0.7.

Given these results, it is highly unlikely that these correlations are due to chance. It is also highly unlikely that they are due to bias in the methods employed. Unlike the earlier study by the National Academy of Sciences (6), this study was not limited to chemicals that were unequivocally carcinogenic in both animals and humans; thus this potential source of bias was avoided. All animal analyses were conducted using a computer program that avoided chemical-specific decisions by an investigator that might perhaps unconsciously be biased towards improving the correlations. Although the analyses of the human

data did require judgements involving individual chemicals, these judgements were made blind, without knowledge of the outcome of the animal analyses. Thus, by any reasonable standard, the animal RRDs are substantially correlated with the human RRDs. This correlation is very important because it demonstrates that it is scientifically feasible to estimate human risk from animal data.

Discussed below are highlights of the correlation analysis results as they relate to specific individual or groups of analyses.

Analyses that Average Over Sex, Study, and Species (Analyses 12-24d). Analyses that average response at all levels generally did not perform as well as comparable analyses that did not average. Analyses that do not average at every level utilize the median of the individual animal RRDs. This result suggests that median RRDs from animal data correlate better with human data than average RRDs. However, the differences between the correlations in analyses that average and comparable analyses that utilize median RRDs is small in many cases.

Analyses that Use Data From Longer Studies or That Dose for Longer Periods (Analyses 1, 2, 13). These analyses generally perform more poorly than comparable analyses that are not so limited (Analyses 0,12). This result is somewhat surprising. It suggests that the timing of the dose is of secondary importance to the amount of the dose, at least when dose is averaged over the length of the experiment as it is in this study.

Analyses that Use the Same Exposure Route or Tumor Response as the Human Data (Analyses 3a, 8c, 25). Analysis 3a that uses the same exposure route as humans and 8c that involves a tumor response that is seen in humans both provide somewhat poorer results than Analysis 0 that does not make these restriction. On the other hand, Analysis 25 that uses

both the same route and response as in humans has a somewhat larger correlation than Analysis 0. These mixed results suggest that, given the uncertainties in the present study with respect to the human RRDs, it does not appear necessary to base a risk assessment on a lesion known to result in humans from exposure to the chemical in question. Similarly, it does not appear to be essential to limit animal data to experiments employing the same route of exposure as humans experience.

Analyses Based on Only Malignant Tumors (Analyses 7, 14). These analyses provide essentially the same correlations as their counterparts (Analyses 0 and 12) that use both benign and malignant tumors, despite the fact that the human results are for malignant tumors exclusively. This suggests that there is no clearcut choice between use of malignant tumors only and use of both benign and malignant in risk assessment and that reasonable risk assessment methods could be based upon either approach.

Analyses Restricted to Specific Species (Analyses 11b, 11c, 11d). Analysis 11b that averages results from mice and rats provides essentially the same correlation as Analysis 11a that averages results from all species. This may be a reflection that the vast majority of the data in the Animal Data Base is from either mouse or rat studies (cf. Table 2). RRDs from rat studies (Analyses 11c), mouse studies (Analyses 11d), and both mouse and rat studies (11b) give nearly identical results.

Choice of Dose Units (Analyses 4a, 4b, 4c, 4d, 24a, 24b, 24c, 24d). Selection of dose units for assumed animal-human equivalence has very little effect upon the correlations; this is expected because relatively few studies in the Animal Data Base include study-specific data on body weight, food consumption, and other variables that affect calculation of the dose measure. However, this choice can have a major effect upon the

actual extrapolated human estimates derived from animal data. This important issue will be explored in connection with the prediction analyses in the next section.

Identification of Analyses Yielding Higher Correlations. Analysis 3b (Figure 3) yields the highest correlation, $\rho = 0.90$. Interestingly, this analysis is the least restrictive of all, being the only one that involves instillation, injection, and implantation studies as well as the more standard gavage, inhalation, and oral studies. This analysis was the only one that included chlorambucil, chromium, and melphalan, since data from experiments using the standard routes of exposure were not available for these chemicals. The correlation analysis was repeated for Analysis 3b with these three chemicals omitted to determine if the high correlation is related to the addition of these chemicals to the analysis. The resulting correlation was 0.88, which is very close to the original value, $\rho = 0.90$, and is still notably better than the correlation obtained from any other analysis.

Aside from Analysis 3b, no other analysis stands out from the others. The next highest correlation is 0.81 (Analysis 25) and another 16 analyses yield correlations between 0.76 and 0.81. The higher correlation obtained from Analysis 3b which employs routes of exposure not normally used for risk assessment suggests that inclusion of these routes may allow improved estimates for some human carcinogens that, for some reason, are not easily shown to be carcinogenic in animals via routes through which humans are normally exposed. Further investigation of this issue may be warranted.

PREDICTION ANALYSES

In the prediction analyses a single RRD estimated from the animal data is used to predict the RRDs obtained from the human data. The fidelity of the prediction is measured by three loss functions: DISTANCE^2, CAUCHY, AND TANH. Thus, whereas the correlation analyses consider only whether higher ranked animal RRDs are associated with higher ranked human RRDs, the prediction analyses examines the ability of the animal bioassays to predict human risk. It also includes an examination of the magnitude of the errors resulting in prediction of human RRDs from animal RRDs.

As in the correlation analysis, the use of the sieve to screen the data appears to be appropriate and useful. This is particularly true when predictors other than the lower bound median, L_{2Q}, are used. While application of the sieve increased average loss for some analysis methods when L_{2Q} was the predictor used, this can probably be largely attributed to confounding associated with use of the sieve and to random factors. It is concluded that definition and application of some data screening procedure that eliminates from consideration experiments of lesser quality should accompany assessments of risk that depend on animal data.

Evaluation of Animal to Human Conversion Methods. Heretofore, animal-to-human extrapolation has generally been conducted by assuming that equal doses will produce the same lifetime risks in animals and humans when both animal and human doses are measured in the same particular units. The dose units studied in this report (mg/kg body weight/day, mg/m^2 surface area/day, ppm in air or water, and mg/kg body weight/lifetime) have all been applied in the past. Because of differences between animals and humans in body weights, life spans, etc., use of different units will produce different estimates of human

risk. There is limited scientific support for use of any particular dose units (7). However, results from the present study can be used to empirically evaluate these different conversion approaches. Specifically the "conversion factor" 10^c, where c is the y-intercept from the best fitting line on the log-log plots of human and animal RRDs, is an estimate of the amount the RRDs obtained from the animal data would have to be multiplied by in order to agree, on average, with the RRDs obtained from the human data. A conversion factor larger than 1 indicates that the RRDs obtained from animal data tend to underestimate those obtained from human data and vice-versa.

Table 9 contains these conversion factors for two loss functions (CAUCHY AND TANH) and for three different sets of analyses chosen such that the analyses within a set differ only with respect to the dose units assumed to yield equivalence between animals and humans. These sets are (0,4a,4b,4c,4d), (12,24a,24b,24c,24d), and (31,30,32,33,34). This table indicates that use of the mg/kg/lifetime dose measure leads to overestimation of the human risk, for all analysis methods considered, by estimated factors ranging from 10 to 150. Similarly, use of mg/m^2 surface area/day also leads to overestimation of risk, by factors ranging from 1.6 to 12. This is significant because this is the dose measure generally used by EPA to estimate human risk. Actually, the extent of overestimation by EPA may be greater that indicated in this table (cf. Table 10); EPA's analysis method generally uses additional conservative assumptions (such as taking the animal data indicative of the highest risk rather that using medians or averaging over studies) not applied in the analysis methods listed in Table 9. (However it should be kept in mind that none of the analyses methods studied will faithfully reproduce EPA's risk assessment results.)

Table 9 indicates that the dose measure mg/kg/day provides more nearly unbiased estimates of human risk when the most appropriate analysis

method as determined in the prediction analysis (i.e. method 30) is used. Interestingly, this measure also generally provided about the smallest loss among the five dose measures, although the differences in loss were small, as expected.

There is no obvious *a priori* reason why any particular dose measure is the "correct" one to use for animal-to-human conversions. Results from the present study can be used empirically to determine appropriate conversion methods. Specifically, multiplication of the animal RRD by the conversion factor, 10^c, provides an estimate of the human RRD in which the bias due to systematic differences in animal and human risk estimates found in this study have been eliminated. With this approach, the dose units can be selected on the basis of those that, along with other facets of an analysis, produced the best correlations between animals and humans (or smallest losses). Application of the correction factor 10^c eliminates the bias associated with any method by correcting for any overestimation or underestimation produced, on average, by that method.

Predictors. Of the four types of predictors investigated (L_M, L_{2Q}, MLE_M, MLE_{2Q}), the lower bound median is clearly superior to the others. This is the indicated by all three loss functions used. Consider the twenty analyses 0-11d (cf. Table 4). With DISTANCE^2 loss, L_{2Q} gave a smaller loss than L_M in every case (MLE_M and MLE_{2Q} are not considered with this loss function); with TANH loss, L_{2Q} gave a smaller loss than the other three types of predictors in 18 analyses; with CAUCHY loss, L_{2Q} gave a smaller loss than the other three types of predictors in 15 analyses.

The superiority of L_{2Q} over the predictors based on maximum likelihood estimates may be related to the fact that small changes in the bioassay data can result in sizable changes in MLE estimates of RRDs. This

suggests that the large-sample theoretical properties of MLEs (such as consistency and asymptotic efficiency) are not operative to any practical extent in this situation, given the usual sample sizes encountered in bioassays. The lack of stability of the MLEs is even more of a problem when extrapolating to low dose or low risk. Regulatory agencies have in the past relied more heavily on lower bound RRDs than on maximum likelihood estimates, mainly in the interest of being protective of human health. This study provides additional support for that policy since the lower bound median is, in fact, a better predictor of human risk estimates than are the MLE predictors (in the sense of providing smaller loss).

Comparison of Analysis Methods. Given that the superiority of L_{20} over the other predictors has been established, it is desirable to identify which analysis methods based upon this predictor provide the best estimates. This task is complicated by the fact that three different loss functions have been defined, and these do not agree completely with respect to the analysis yielding smallest loss. Moreover, it seems unlikely that there would exist a single "best" method. Consequently, we have identified a small set of analysis methods that perform relatively well with respect to all three loss functions.

Several such analysis methods, along with others that are of general interest are listed in Table 10. All of the results in this table are from applying the L_{20} estimator, except in the one case noted on the table. The "incremental normalized loss" presented in this table is a summary loss measure synthesized from all three loss functions. For each loss function separately, it is possible to determine for a particular analysis the amount of additional loss over the minimum contributed by that analysis. The sum of these additional losses over the three loss functions defines the total incremental normalized loss. The "conversion factors" listed in Table 10 are the average factors,

10^C, by which RRDs obtained from the animal data would have to be multiplied by in order to agree, on average, with the RRDs obtained from the human data; these factors were discussed in an earlier section. The last column in Table 10 contains values of the residual error, which represents the average distance on a log-log plot from the interval defined by the human RRDs to the line that fits best, given the animal RRD predictors and the intervals determined by the human RRDs. This residual error represents roughly the average multiplicative error in estimating the human RRDs from the animal data that is not explainable by the uncertainty in the human RRDs (this uncertainty being expressed by the intervals $[RRD_L, RRD_U]$ estimated from the human data). The residual error is in essence an additional expression of loss.

The Base Analysis (Analysis 0) employing the minimal lower bound estimator, L_M (second row of Table 10) has both the largest normalized loss and the largest residual error. Moreover, RRDs derived from this analysis underestimate the human RRDs on average by a factor of 12. By all standards, this method is the poorest of those listed. This method is also perhaps most like that presently employed by EPA. Modification of this method by using the median lower bound estimator, L_{50}, rather than L_M, as represented in the first row of Table 10, provides an improvement in terms of normalized loss, residual error, and requiring a smaller conversion factor. These results illustrate further the finding discussed earlier that analysis methods that use median lower bound RRDs as estimators provide smaller losses than analysis methods that use minimum estimates.

Use of malignant tumors only, rat data only, or mice data only (Analyses 7, 11c, and 11d, respectively) did not provide clear improvements over estimates that included data on nonmalignant tumors and data from different species.

Analyses 30, 31, 43, 45, and 47 are presented as a group of analyses that generally perform well. All of these analyses use the mg/kg/day method of extrapolating from animals to humans (except 31, which utilizes the $mg/m^2/day$ method), and all include routes of exposure (instillation, injection, and implantation) not normally used in quantitative risk assessment. Analyses 30, 45, and 47 differ only in the way RRDs are combined and give fairly comparable results; Analysis 45 which averages RRDs from different sexes in the same study, might be considered to perform the best overall, as it has both the smallest normalized loss and residual uncertainty. This analysis also had the largest correlation (0.91) of those in Table 10. Analysis 43 employs a different carcinogenic endpoint than the others, namely total tumor-bearing animals. Although this analysis has a small normalized loss, its residual uncertainty factor is 40% larger than any from Analyses 30, 45 and 47.

Options for Presenting a Range of Risk Estimates. Guidelines are provided for presenting a range of risk estimates for a risk assessment based on Analyses 30, 31, 43, 45, and 47. Three options are considered. The first entails selecting, *a priori*, one method from the recommended set. The results of that method, including the uncertainty quantified by the residual uncertainty factor, are taken as the representative range of risk estimates. The second option uses all the methods. The range it produces includes any value that could be obtained from any one or more of the methods, and so can be considered to give the maximum range consistent with the recommended set. Although the third option also considers all methods in the recommended set, it summarizes the results by the smallest range of estimates that is consistent with the predictions of all the analyses. As with the first option, the last two incorporate the residual uncertainty factors to define the ranges of estimates.

DISCUSSION

GENERAL CONSIDERATIONS

The animal data base and the methods used in this study provide a useful basis for evaluating quantitative risk assessment. Their use in the present context has demonstrated the strong positive correlation between the animal and human risk estimates and hence relevance of animal carcinogenicity experiments to human risk estimation. Moreover, it has been possible to identify methods of analysis of the bioassay data, including the choice of the median lower bound predictor, that satisfactorily predict risk-related doses in humans. Application of these methods has led to suggested guidelines concerning the prediction of human risks and the presentation of ranges of estimates incorporating the relevant uncertainties.

There are, however, certain features of this investigation that should be borne in mind when evaluating the results of this study. These are summarized below.

- A risk level of 0.25 is used throughout.
- The bioassay data is rather crude in several respects. We have already referred to the data deficiencies and their impact on the ability to perform some analyses.
- The epidemiological data is of variable quality. Some degree of subjectivity is inherent in the estimates of uncertainty associated with the epidemiological RRDs.
- Different forms (complexes) of some chemicals were grouped together.
- Other approaches to the components could be defined and investigated.

- The three loss functions employed in the prediction analysis lack an underlying statistical development and so have been used merely to rank the analysis methods.
- Many other analysis methods could be investigated.

DIRECTIONS FOR FUTURE RESEARCH

In the course of the previous discussion, several proposed extensions of this project have been mentioned. Several fall under the heading of sensitivity analyses of the results already obtained. These include investigation of the robustness of the results to reasonable alternative choices for the epidemiological estimates; examination of other means to analyze bioassay data, including time-to-tumor analyses; and investigation of the effect of using lower levels of risk, say 10^{-6}, which are of direct regulatory concern. A detailed statistical development of the loss functions used here (or a general development for certain classes of loss functions) might be of general interest. The data that is available from this project could provide an interesting and pertinent example to which that development could apply.

Also discussed in connection with component-specific uncertainty are efforts directed at reducing or explaining that uncertainty. The greatest uncertainties are related to the components specifying how to handle experiments of different lengths of dosing, routes of exposure, or test species and specifying the carcinogenic responses to use. Many aspects of these components and their uncertainties can be addressed in an investigation of pharmacokinetics. The data base contains detailed data on the timing and intensity of exposure for each bioassay, so a pharmacokinetic study, which requires such information, is entirely feasible with the currently collected data. Two specific proposals are discussed here.

Risk estimates incorporating pharmacokinetic data could be used to determine appropriate surrogate doses. It is sometimes assumed that a given dose measured as average concentration of the active metabolite at the target tissue will produce the same risk in animals and humans. However, given the many differences between animals and humans (size, life span, and metabolic rates, to mention a few), it is not clear which, if any, surrogate dose is the most appropriate. This issue is similar to that of choice of the most appropriate surrogate dose measure for animal to human extrapolation (*e.g.* mg/kg/day versus mg/m^2/day) considered in this study and can be studied in a similar manner. Risk estimates using pharmacokinetic data could be used to determine empirically the most appropriate surrogate dose. Even though the range of RRDs consistent with the human data generally cover a range of an order of magnitude or greater, the potential surrogate doses cover an even wider range. Just as the present study indicates that certain dose measures appear to predict human results well in conjunction with appropriate choices for other risk assessment components, a study using pharmacokinetic data should allow similar conclusions regarding the surrogate dose. A preliminary investigation indicates that possibly 16 of the 23 chemicals with suitable human data used in this study might also have data that would support a risk assessment that incorporates pharmacokinetic data.

A second potentially useful investigation incorporating pharmacokinetic data involves using the data in the data base on different routes of exposure to study the best means of extrapolating from route to route in animal studies. Risk assessment methods, including the ones examined in this study, often assume a given dose rate involves the same risk, regardless of route. This clearly is a gross oversimplification. The animal data collected for this study contains numerous examples of carcinogenicity studies on the same chemical and animal species, but for

which exposure is through different routes. Those studies could be used to determine how pharmacokinetic data could best be applied to perform route-to-route extrapolation. Since human data would not be essential in these investigations, our total data base that encompasses 44 chemicals could be used.

The question of different chemical classes and the consistency that may be apparent within any of the classes is deserving of further study. It would be reasonable to couple this work with pharmacokinetic methods. In the present data base, several classes are represented. However, the number within any particular class is somewhat limited. An expanded data base may be necessary for a thorough investigation.

In fact, one desirable goal in and of itself, but one that would enhance the prospects for successful completion of these other proposals, is the maintenance and updating of the bioassay data base. All aspects of this, including accumulation of more data sets for the chemicals already included and addition of more substances, may be necessary. Some revamping of the data coding format may also make future analyses easier and more accurate. Especially for pharmacokinetic studies, for instance, dose patterns could be recorded on a daily rather than weekly basis.

As a counterpart to the bioassay data base enhancement, updating and augmenting the epidemiological data is essential. Since the epidemiological data (in particular, data on exposure) is the single most limiting factor preventing use of human data, any hope of increasing the size of the sample of chemicals useful in estimating conversion factors and residual uncertainty must be based on an effort to acquire such data. For those chemicals already analyzed, more specific exposure data would reduce the uncertainty bounds surrounding epidemiological RRD estimates and refine our estimates. As is the case

with the bioassay data, much of the limitation or uncertainty is solely a matter of inadequate reporting of data.

It should be noted in passing that the methods and portions of the computer programs developed and applied in this project may be useful in other contexts. Of particular interest is a study of other types of health effects, *e.g.* reproductive effects. The investigation of these issues could include determinations of uncertainty as well as identification of the most appropriate methods. Other projects, including investigation of other types of extrapolations, *e.g.* from one temporal dosing pattern to another or from rats to mice, could also be facilitated by use of the data base, methods, and programs developed in the present work.

Finally, one would like to investigate cancer risk assessment methods appropriate when data available to a particular assessment are limited. We have mentioned this problem in connection with component-specific uncertainty (*i.e.* noting that confounding like that affecting those uncertainty calculations will often be present in any given risk analysis setting) and in connection with the set of recommended bioassay analysis methods. In the latter instance, it was pointed out that each analysis in the recommended set, save for Analysis 17, is capable of being applied to any data base but that data limitations due to incomplete data presentation may entail that Analyses 20 and 43 are not possible. The remaining analyses (30, 31, 45, and 47) can be performed no matter what the data set contains, but they may be seriously affected by the extent and nature of the contents.

Consequently, the following investigation is proposed as a means of studying the effects of the limitations on the data for any chemical of interest and of determining how best to extrapolate risks to humans. Pick the data in the data base that most nearly matches the data for the

chemical in question. The matching may be based on species, routes of exposure, and quality of the data. Moreover, one may wish to restrict attention to chemicals that are in the same class of the substance of interest. Suppose, for example, a volatile organic chemical is under investigation and that the only data available are from rat inhalation studies. Then, the proposed procedure would first select rat inhalation bioassays conducted using appropriate chemicals (*i.e.*, perhaps limited to volatile organics). The components of risk assessment not fixed by the selection could be varied and the method that works best with the selected data would be the basis for extrapolating to humans risks due to the chemical in question. Since we also have a recommended set consisting of methods that appear to perform well for the data and chemicals considered as a whole, the risks estimated on that basis (*i.e.* using the recommended set) would be available for comparison. These estimates reveal what would happen if other species, other routes, and other chemicals are included. The relationship between the estimates obtained by the two approaches would suggest a general type of uncertainty attributable to use of a limited data base (in this example, rat inhalation studies). A pilot study could investigate the feasibility of such a chemical-specific approach to risk assessment.

REFERENCES

1. Gold, L., Sawyer, C., Magaw, R., Backman, G., de Veciana, M., Levinson, R., Hoooper, N., Havender, W., Bernstein, L., Peto, R., Pike, M., and Ames, B. (1984). A carcinogenic potency database of the standardized results of animal bioassays. Environmental Health Perspectives 58:9-319.

2. International Agency for Research on Cancer (1982). Chemicals, industrial processes and industries associated with cancer in humans. IARC Monographs on the Carcinogenic Risk of Chemicals to Humans Supplement 4.

3. National Academy of Sciences Executive Committee (1975). Contemporary Pest Control Practices and Prospects. Pest Control: An Assessment of Present and Alternative Technologies, Vol. 1.

4. Howe, R. and Crump, K. (1982). GLOBAL 82: A Computer Program to Extrapolate Quantal Animal Toxicity data to Low Doses. Prepared for the Office of Carcinogen Standards, OSHA, U. S. Department of Labor, Contract 41USC252C3.

5. Ng, T. (1985). A Generalized Ranking on Partially Ordered Sets and Its Applications to Multivariate Extensions to Some Nonparametric Tests. Unpublished report.

6. National Academy of Sciences Executive Committee (1975). Contemporary Pest Control Practices and Prospects. Pest Control: An Assessment of Present and Alternative Technologies, Vol. 1.

7. Crump, K., Silvers, A., Ricci, P., and Wyzga, R. (1985). Interspecies comparison for carcinogenic potency to humans. Principles of Health Risk Assessment. Ricci, P. (ed.). Prentice Hall. pp. 321-372.

8. U. S. Environmtneal Protection Agency. Addendum to the Health Assessment Document for Dichloromethane (Methylene Chloride). EPA-600/8-82-004FF.

Table 1

APPROACHES TO RISK ASSESSMENT COMPONENTS

1. Length of experiment
   a. Use data from any experiment but correct for short observation periods.
   b. Use data from experiments which last no less than 90% of the standard experiment length of the test animal.

2. Length of dosing
   a. Use data from any experiment, regardless of exposure duration.
   b. Use data from experiments that expose animals to the test chemical no less than 80% of the standard experiment length.

3. Route of exposure
   a. Use data from experiments for which route of exposure is most similar to that encountered by humans.
   b. Use data from any experiment, regardless of route of exposure.
   c. Use data from experiments that exposed animals by gavage, inhalation, any oral route, or by the route most similar to that encountered by humans.

4. Units of dose assumed to give human-animal equivalence
   a. mg/kg body wt/day.
   b. ppm in diet.
   c. ppm in air.
   d. mg/kg body wt/lifetime.
   e. mg/m^2 surface area/day.

5. Calculation of average dose
   a. Doses expressed as average dose up to termination of experiment.
   b. Doses expressed as average dose over the first 80% of the experiment.

6. Animals to use in analysis
   a. Use all animals examined for the particular tumor type.
   b. Use animals surviving just prior to discovery of the first tumor of the type chosen.

7. Malignancy status to consider
   a. Consider malignant tumors only.
   b. Consider both benign and malignant tumors.

8. Tumor type to use
   a. Use combination of tumor types with significant dose-response.
   b. Use total tumor-bearing animals.
   c. Use response that occurs in humans.
   d. Use any individual response.

Table 1 (continued)

APPROACHES TO RISK ASSESSMENT COMPONENTS

9. Combining data from males and females
   a. Use data from each sex within a study separately.
   b. Average the results of different sexes within a study.

10. Combining data from different studies
   a. Consider every study within a species separately.
   b. Average the results of different studies within a species.

11. Combining data from different species
   a. Average results from all available species.
   b. Average results from mice and rats.
   c. Use data from a single, preselected species.
   d. Use all species separately.

NOTE: Underlines indicate approach used in base analysis (Analysis 0).

Table 2

SUMMARY OF ANIMAL DATA BY CHEMICAL

Chemical	No. Reviewed		Number of Data Sets Coded in Animal Database												
			Oral			Gavage			Inhalation			Other			
	Studies[a]	Data Sets[a]	R[b]	M[b]	O[b]	R	M	O	R	M	O	R	M	O	Total
Acrylonitrile	10	19	9	0	0	4	0	0	6	0	0	0	0	0	19
Aflatoxin	62	86	23	0	3	3	1	0	0	1	0	0	2	0	33
Allyl Chloride	2	6	2	2	0	0	0	0	0	0	0	0	0	0	4
4-Aminobiphenyl	8	8	0	2	0	0	0	0	0	0	0	0	0	0	2
Arsenic	16	33	7	2	0	0	0	0	0	0	0	5	2	1	17
Asbestos	39	84	1	0	8	0	0	0	11	0	9	17	0	18	64
Benzene	13	26	0	0	0	6	2	0	2	4	0	0	1	0	15
Benzidine	8	10	0	0	0	1	0	0	0	0	0	1	1	0	3
Benzo[a]pyrene	42	51	0	0	0	1	2	0	0	0	0	3	2	6	14
Cadmium	26	30	3	0	0	1	1	0	1	0	0	1	1	0	8
Carbon Tetrachloride	8	21	0	0	0	2	4	0	0	0	0	0	0	0	6
Chlorambucil	3	6	0	0	0	0	0	0	0	0	0	0	1	0	1
Chlordane	3	8	2	6	0	0	0	0	0	0	0	0	0	0	8
Chloroform	12	31	1	1	2	3	13	0	0	0	0	0	0	0	20
Chromium	12	16	1	0	0	0	0	0	0	0	0	6	1	0	8
Cigarette Smoke	37	41	0	3	0	0	0	0	3	0	1	1	0	1	9
3,3-Dichlorobenzidene	6	8	4	0	1	0	0	0	0	0	0	0	0	0	5
1,2-Dichloroethane	5	14	0	0	0	2	2	0	2	2	0	0	0	0	8
Dichloromethane	6	14	2	2	0	0	0	0	6	2	0	0	0	0	12
Diethylstilbestrol	61	81	0	2	0	0	0	0	0	0	0	2	9	3	16
Diphenylhydrazine	2	2	0	0	0	0	0	0	0	0	0	1	0	0	1
Epichlorohydrin	7	7	1	0	0	0	0	0	2	0	0	0	0	0	3
Estrogen	24	34	1	0	0	0	0	0	0	0	0	0	0	1	2
Ethylene Dibromide	7	19	0	0	0	2	2	0	4	2	0	0	0	0	10
Ethylene Oxide	10	15	0	0	0	1	0	0	9	0	0	0	0	0	10
Formaldehyde	9	15	0	0	0	0	0	0	5	4	1	0	0	0	10

Table 2 (continued)

SUMMARY OF ANIMAL DATA BY CHEMICAL

Chemical	No. Reviewed		Number of Data Sets Coded in Animal Database												
			Oral			Gavage			Inhalation			Other			
	Studies[a]	Data Sets[a]	R[b]	M[b]	O[b]	R	M	O	R	M	O	R	M	O	Total
Hexachlorobenzene	4	7	2	2	2	0	0	0	0	0	0	0	0	0	6
Hydrazine	15	31	0	7	0	4	8	0	0	0	0	0	0	0	19
Isoniazid	25	66	4	17	4	0	11	0	0	0	0	0	4	0	40
Lead	22	33	9	2	2	0	0	0	0	0	0	0	1	1	15
Melphalan	4	7	0	0	0	0	0	0	0	0	0	0	1	0	1
Methotrexate	9	16	0	2	2	0	0	0	0	0	0	2	0	0	6
Mustard Gas	2	4	0	0	0	0	0	0	0	0	0	0	4	0	4
2-Naphthylamine	23	37	1	1	9	1	4	1	0	0	0	0	4	0	21
Nickel	37	77	0	0	1	0	0	0	5	0	0	20	1	1	28
Nitrilotriacetic Acid	7	18	9	6	0	0	0	0	0	0	0	0	0	0	15
Phenacetin	13	21	5	6	0	1	0	0	0	0	0	0	0	0	12
Polychlorinated Biphenyls	9	12	3	3	0	0	0	0	0	0	0	0	0	0	6
Reserpine	2	6	2	2	0	0	0	0	0	0	0	0	0	0	4
Saccharin	19	27	14	1	1	0	0	0	0	0	0	0	2	0	18
2,3,7,8-Tetrachloro-dibenzo-p Dioxin	11	19	3	0	0	2	3	0	0	0	0	0	2	0	10
Tetrachloroethylene	5	14	0	0	0	2	2	0	2	2	0	0	0	0	8
Toxaphene	1	4	2	2	0	0	0	0	0	0	0	0	0	0	4
Trichloroethylene	39	34	0	0	0	4	8	0	3	3	2	0	2	0	22
2,4,6-Trichlorophenol	1	4	2	2	0	0	0	0	0	0	0	0	0	0	4
Vinyl Chloride	35	65	4	0	0	4	0	0	23	20	1	4	0	0	56
Vinylidene Chloride	17	46	2	0	0	3	2	0	10	15	0	0	0	0	32
TOTAL	736	1233	119	73	35	47	65	1	94	55	14	63	41	32	631

[a]A study is generally comprised of all information contained in a single primary reference on a simple chemical. A data set generally comprises all of the dose response data from a given sex and species to animals via a common protocol in a study.

[b]R = rat; M = mouse; O = other species.

Table 3

CHEMICALS FOR WHICH MINIMAL HUMAN AND ANIMAL DATA EXIST FOR QUANTIFYING CARCINOGENIC POTENCY

Chemical	Use[a]	Evidence for Carcinogenicity (IARC classification scheme) In Humans	In Animals
Aflatoxin (AF)	F	Limited	Sufficient
Arsenic (AS)	IC	Sufficient	Inadequate
Asbestos (AB)	IC	Sufficient	Sufficient
Benzene (BN)	IC	Sufficient	Limited
Benzidine (BZ)	IC	Sufficient	Sufficient
Cadmium (CD)	IC	Limited	Sufficient
Chlorambucil (CB)	D	Sufficient	Sufficient
Chromium (CR)	IC	Sufficient	Sufficient
Cigarette smoke (CS)[b]		- -	- -
Diethylstilbestrol (DS)	D	Sufficient	Sufficient
Epichlorohydrin (EC)	IC	Inadequate	Sufficient
Estrogens (ES) (conjugated)	D	Sufficient	Inadequate
Ethylene oxide (EO)	IC	Inadequate	Limited
Isoniazid (IS) (isonicotinic acid hydrazide)	D	Inadequate	Limited
Melphalan (ML)	D	Sufficient	Sufficient
Methylene chloride (MC)	IC	Inadequate	Sufficient[c]
Nickel (NC)	IC	Limited	Sufficient
Phenacetin (PH) (analgesics containing phenacetin)	D	Sufficient	Limited
Polychlorinated biphenyls (PC)	IC	Inadequate	Sufficient
Reserpine (RS)	D	Inadequate	Limited
Saccharin (SC)	F	Inadequate	Limited
Trichloroethylene (TC)	IC	Inadequate	Limited
Vinyl Chloride (VC)	IC	Sufficient	Sufficient

[a]IC = industrial chemical; D = drug; F = food additive or contaminant.
[b]Not considered in IARC monographs, although tobacco smoke is acknowledged by IARC as a known human carcinogen.
[c]Although classified as "Inadequate" by IARC (2), results of studies completed since IARC evaluation indicate that the evidence for the carcinogenicity of methylene chloride in animals is now "Sufficient" (8).

Table 4

DESCRIPTIONS OF INITIAL ANALYSES

Analysis	Template[a]	Differences[b]
0	Base Analysis	[described in Table 1]
1	0	limited to experiments of long observation
2	0	limited to experiments of long dosing
3a	0	route like human route only
3b	0	any route
4a	0	mg/kg/day
4b	0	ppm diet
4c	0	ppm air
4d	0	mg/kg/lifetime
5	0	doses averaged over first 80% of experiment
6	0	early deaths eliminated
7	0	malignant responses only
8a	0	combination of significant responses only
8b	0	total tumor-bearing animals only
8c	0	response that human get only
9	0	results averaged over sex within study
10	0	results averaged over study within species
11a	0	results averaged over all species
11b	0	results averaged over rats and mice only
11c	0	rat data only
11d	0	mouse data only
12	0	results averaged over sex, study, and species
13	12	limited to experiments of long dosing and observation
14	12	malignant responses only
15	14	limited to experiments of long dosing and observation
16	12	combination of significant responses only
17	16	limited to experiments of long dosing and observation
18	12	combination of malignant significant responses only
19	18	limited to experiments of long dosing and observation
20	12	total tumor-bearing animals only
21	20	limited to experiments of long dosing and observation
22	12	total malignancy-bearing animals only
23	22	limited to experiments of long dosing and observation

Table 4 (continued)

DESCRIPTIONS OF INITIAL ANALYSES

Analysis	Template[a]	Differences[b]
24a	12	mg/kg/day
24b	12	ppm diet
24c	12	ppm air
24d	12	mg/kg/lifetime
25	0	route and response that humans get only

[a]The template is the analysis which most closely resembles a given analysis.

[b]The differences listed are the ways in which the analysis in question differs from its template. For Analyses 0, no "differences" are defined. The approaches to this analysis are indicated in Table 1.

Table 5

DESCRIPTIONS OF SUPPLEMENTAL ANALYSES

Analysis	Template[a]	Differences[b]
30	0	mg/kg/day; any exposure route
31	30	$mg/m^2/day$
32	30	ppm diet
33	30	ppm air
34	30	mg/kg/lifetime
35	30	limited to experiments of long observation
36	30	limited to experiments of long dosing
37	30	route like humans only
38	30	oral, gavage, inhalation, or route like humans
41	30	malignant responses only
42	30	combination of significant responses only
43	30	total tumor-bearing animals only
44	30	response that humans get only
45	30	results averaged over sex within study
46	30	results averaged over study within species
47	30	results averaged over all species
48	30	results averged over rats and mice only
49	30	rat data only
50	30	mouse data only

[a]The template is the analysis which a given analysis most closely resembles.

[b]The differences listed are the ways in which the analysis in question differs from its template.

Table 6

RANKS BASED ON LENGTH OF EXPERIMENT
AND NUMBER OF TREATED ANIMALS

Length of Experiment[a]	Number of Dosed Animals		
	50+	15-49	< 15
≥ 75%	1	2	5
50-75%	3	4	7
< 50%	6	8	9

[a]These values are expressed as percentages of the standard experiment length of the test species.

Table 7

COMPONENT-SPECIFIC UNCERTAINTY: MODES AND DISPERSION FACTORS FOR RATIOS OF RRDS[a], BY SUPPLEMENTAL ANALYSIS[b]

Analysis	Number of Chemicals	Mode of Histogram	Dispersion Factor[c]
31	44	.05 - .1	2.3
32	44	.2 - .5	1.7
33	44	.2 - .5	1.8
34	44	.02 - .05	1.3
35	40	.8 - 1.25	28.5
36	34	.8 - 1.25	86.0
37	24	.8 - 1.25	5.3
38	40	.8 - 1.25	33.7
41	39	.8 - 1.25	290.6
42	29	.8 - 1.25	75.6
43	31	.8 - 1.25	39.6
44	37	.8 - 1.25	54.1
45	44	.8 - 1.25	1.2
46	44	.8 - 1 25	1.7
47	44	.8 - 1.25	2.2
48	43	.8 - 1.25	23.2
49	39	.8 - 1.25	39.6
50	36	.8 - 1.25	335.6

[a]The ratios are of the chemical-specific RRD estimates from the indicated analysis to those of Analysis 30 (cf. Table 5).

[b]The analyses were performed with the L_{20} predictor and using the full sieve.

[c]The dispersion factor is the average factor by which the ratios differ from the mode.

Table 8

CORRELATION COEFFICIENTS AND ASSOCIATED p-VALUES, BY ANALYSIS METHOD[a]

Analysis	Number of Chemicals	ρ	p-value
0	20	.78	.0001
1	18	.68	.0015
2	19	.49	.0153
3a	17	.73	.0007
3b	23	.90	<.0001
4a	20	.78	.0001
4b	20	.76	.0001
4c	20	.78	<.0001
4d	20	.78	<.0001
5	20	.79	<.0001
6	6	.79	.0342
7	19	.76	.0001
8a	13	.56	.0214
8b	17	.66	.0022
8c	18	.76	.0001
9	20	.76	.0003
10	20	.77	.0002
11a	20	.76	<.0001
11b	20	.76	<.0001
11c	19	.79	<.0001
11d	13	.76	.0023
12	20	.75	<.0001
13	18	.43	.0416
14	19	.71	.0005
15	18	.46	.0316
16	13	.49	.0436
17	11	.58	.0301
18	10	.73	.0090
19	9	.79	.0058
20	17	.63	.0043
21	13	.38	.1023
22	15	.35	.1036
23	13	.18	.2821
24a	20	.75	.0001
24b	20	.74	.0001
24c	20	.74	.0001
24d	20	.75	<.0001
25	16	.81	.0002

[a]A sieve to screen the data has been used.

Table 9

CONVERSION FACTORS[a] CORRESPONDING TO VARIOUS DOSE UNITS, BY METHOD OF ANALYSIS[b]

Units	Analysis Method	
$mg/m^2/day$	Restricted routes, unaveraged (0)	1.58 - 2.07
	Restricted routes, averaged (12)	3.47 - 5.61
	Unrestricted routes, unaveraged (31)	8.45 - 12.02
mg/kg/day	Restricted routes, unaveraged (4a)	0.28 - 0.40
	Restricted routes, averaged (24a)	0.43 - 0.61
	Unrestricted routes, unaveraged (30)	1.08 - 1.70
ppm diet	Restricted routes, unaveraged (4b)	0.59 - 1.17
	Restricted routes, averaged (24b)	1.77 - 2.95
	Unrestricted routes, unaveraged (32)	4.52 - 5.94
ppm air	Restricted routes, unaveraged (4c)	0.83 - 1.06
	Restricted routes, averaged (24c)	1.82 - 2.96
	Unrestricted routes, unaveraged (33)	1.89 - 6.61
mg/kg/life	Restricted routes, unaveraged (4d)	10.40 - 16.67
	Restricted routes, averaged (24d)	19.63 - 23.12
	Unrestricted routes, unaveraged (34)	72.95 - 79.62

[a]The multiplicative factor by which bioassay-based RRDs overestimate, on average, RRDs obtained from human data.

[b]The range given is that suggested by the CAUCHY and TANH loss functions; all results based upon median lower bound (L_{2Q}) estimator.

Table 10

COMPARISON OF RESULTS FOR SELECTED ANALYSES[a]

Analysis	Number of Chemicals	Correlation Coefficient	Total Incremental Normalized Loss[b]	Bias-Correcting Conversion Factors[c]	Residual Uncertainty Factor[d]
0	20	0.78	1.15	1.6 - 2.1	5.3
0[e]	20	0.78	1.71	12 - 12	16.2
7	19	0.76	1.40	1.6 - 3.6	5.4
11c	19	0.77	0.62	0.81 - 1.9	4.5
11d	13	0.76	1.01	3.7 - 4.3	3.1
17	11	0.58	0.27	2.8 - 2.8	4.2
20	17	0.67	0.62	0.69 - 0.78	7.1
30	23	0.91	0.39	1.1 - 1.7	2.0
31	23	0.90	0.53	8.5 - 12	2.0
43	17	0.74	0.28	0.18 - 0.29	2.8
45	23	0.91	0.27	1.2 - 1.7	1.7
47	23	0.89	0.28	1 - 1.7	1.8

[a]The results correspond to the member of the pair (with sieve, without sieve) that gives best results. For Analyses 11c, 20, and 43 this is without the sieve; for other analyses this is with the sieve. The median lower bound predictor, L_{2Q}, is used in all analyses except for the exception noted.

[b]This value is not the same as that in Table 2-8 because the inclusion of the supplemental analyses reduced the minimum average loss for two of the three loss functions and increased the maximum loss for all three of the functions.

[c]These values are the factors, 10^c, based on the y-intercepts from the CAUCHY and TANH loss functions (cf. Tables 2-13 and 2-17) and represent the average ratio of human RRDs to animal RRDs.

[d]Residual uncertainty is from Table 2-21 or 2-22. It is the factor computed for all chemicals and represents the average factor by which a prediction must be multiplied or divided in order to eliminate uncertainty not due to uncertainty in the human estimates.

[e]Using minimal lower bound estimator L_M.

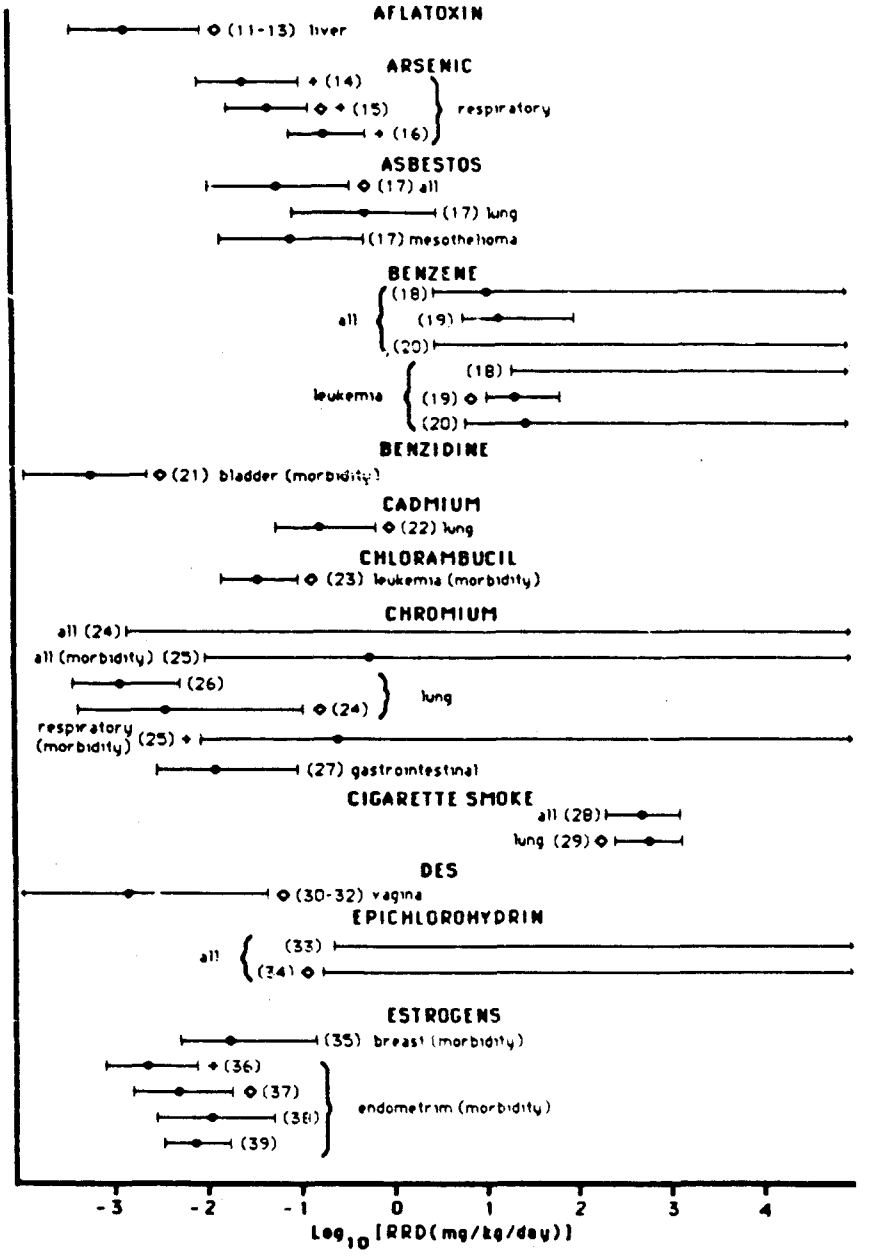


Figure 1. Best estimates and upper and lower bounds for RRDs from each human study.

◇ Marks the data selected to represent the chemical when comparisons with bioassay-based estimates are made.

+ Marks poor fit of linear dose-response model to data.

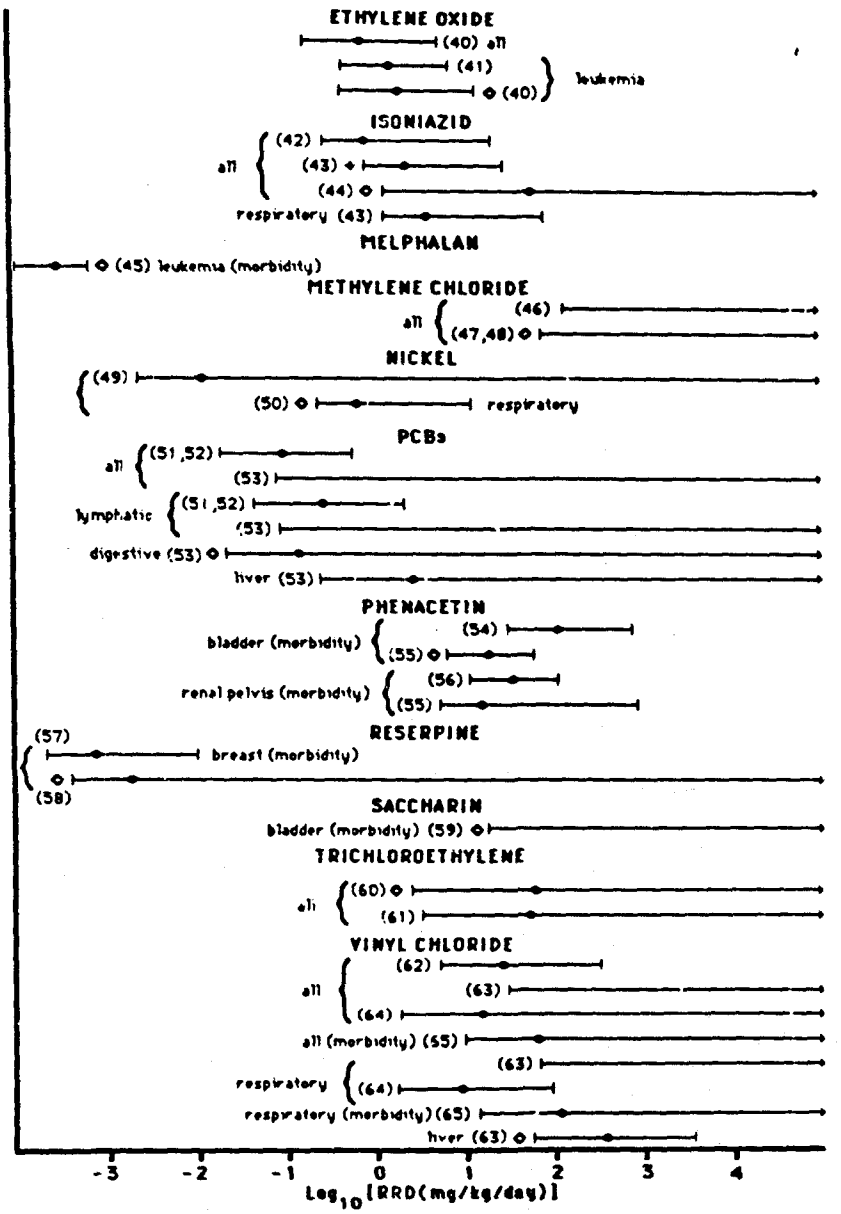


Figure 1. Best estimates and upper and lower bounds for RRDs from each human study.

◇ Marks the data selected to represent the chemical when comparisons with bioassay-based estimates are made.

+ Marks poor fit of linear dose-response model to data.

Figure 2

Correlation of Animal and Human RRDs - Analysis 0

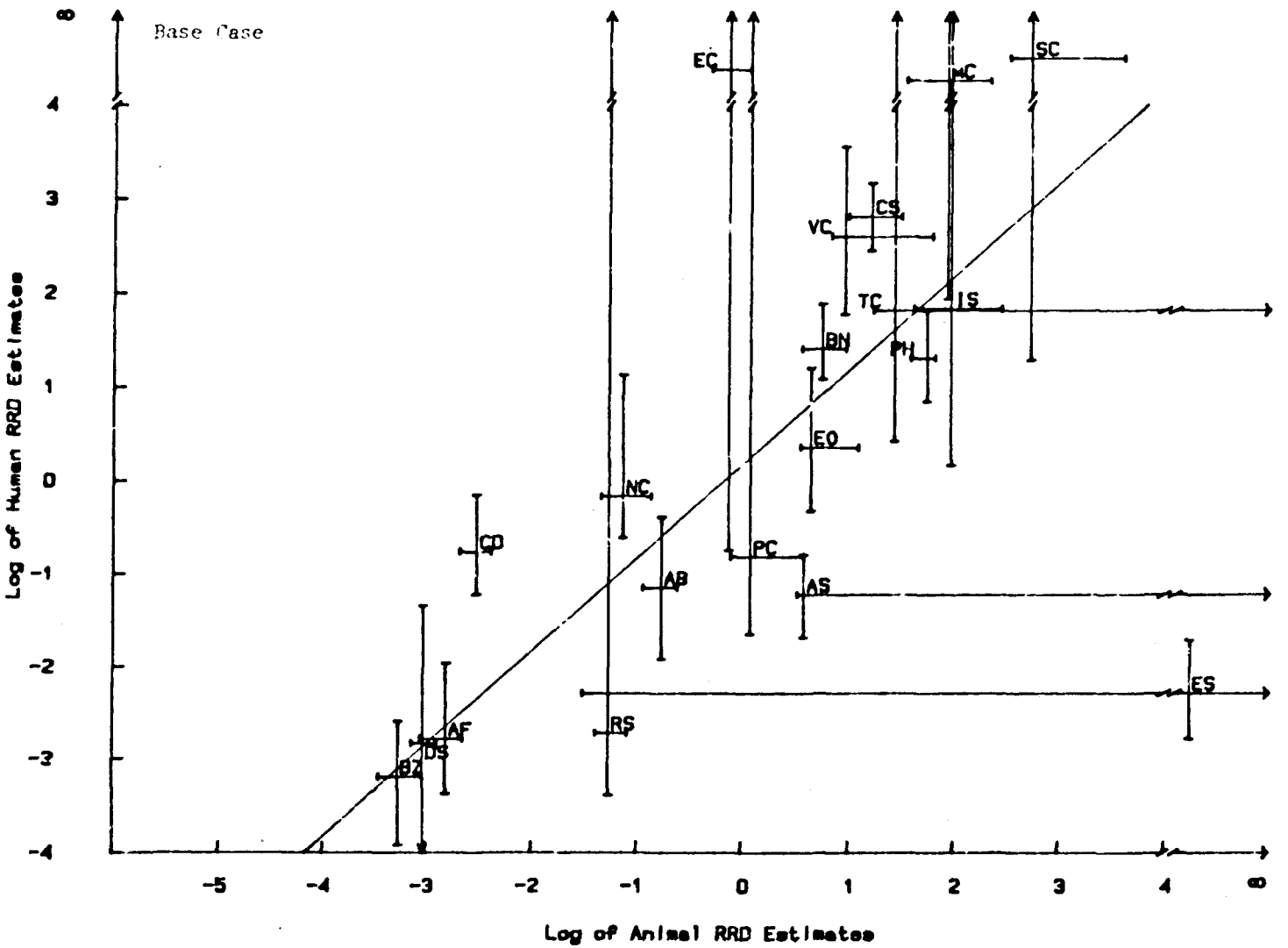


Figure 3

Correlation of Animal and Human RRDs - Analysis 3b

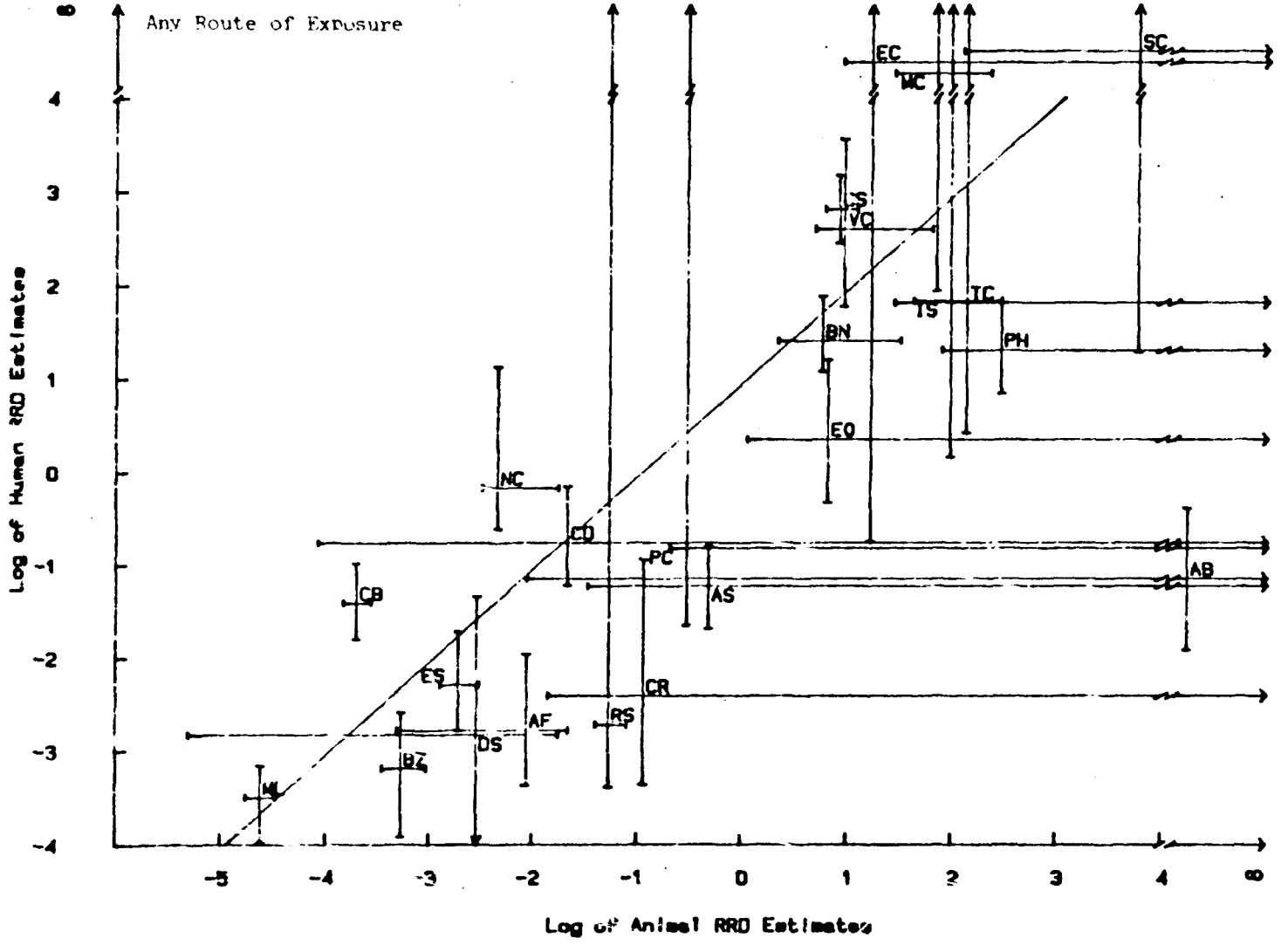


Figure 4

Correlation of Animal and Human RRDs - Analysis 7

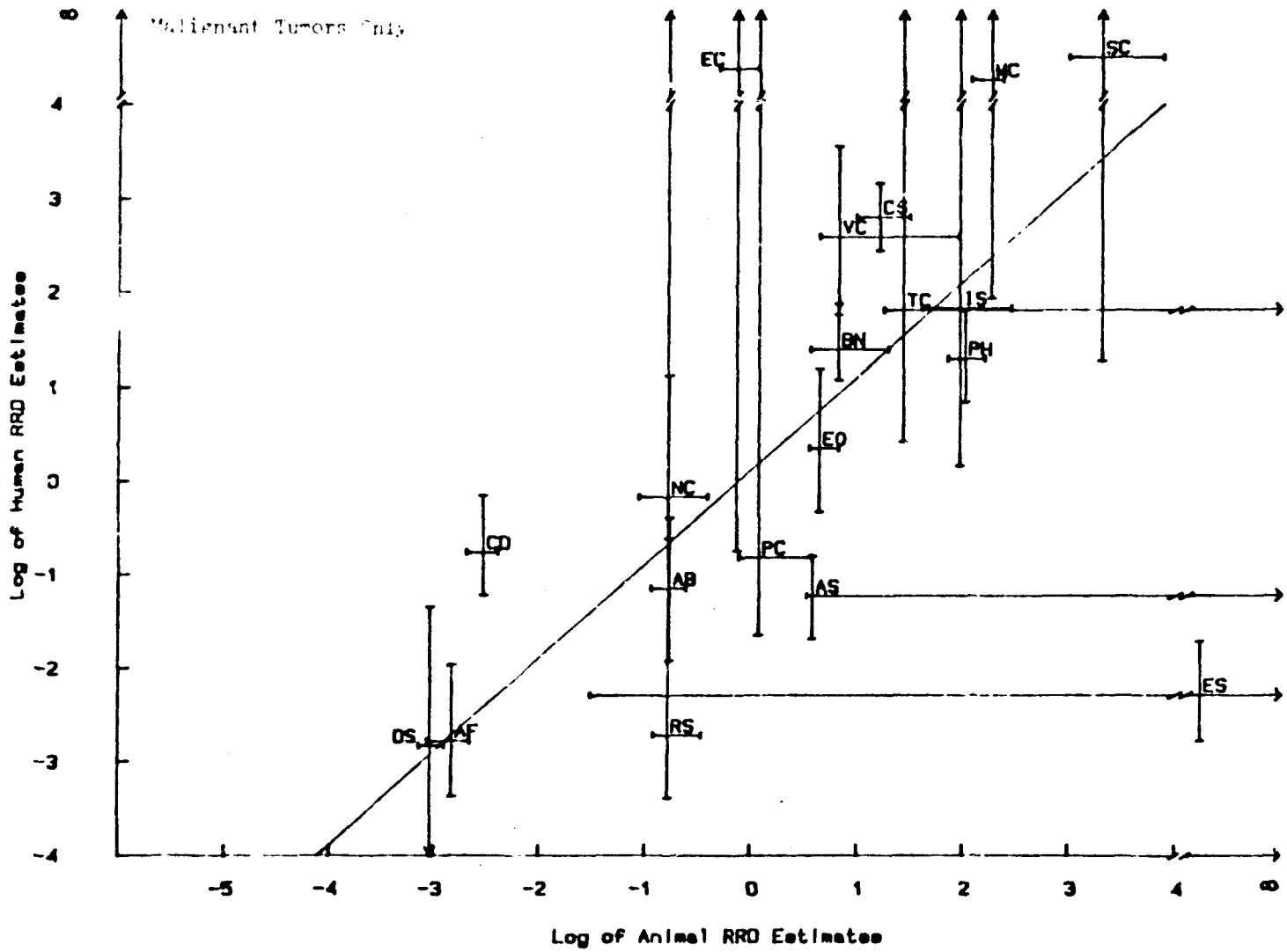


Figure 5

Correlation of Animal and Human RRDs - Analysis 11c

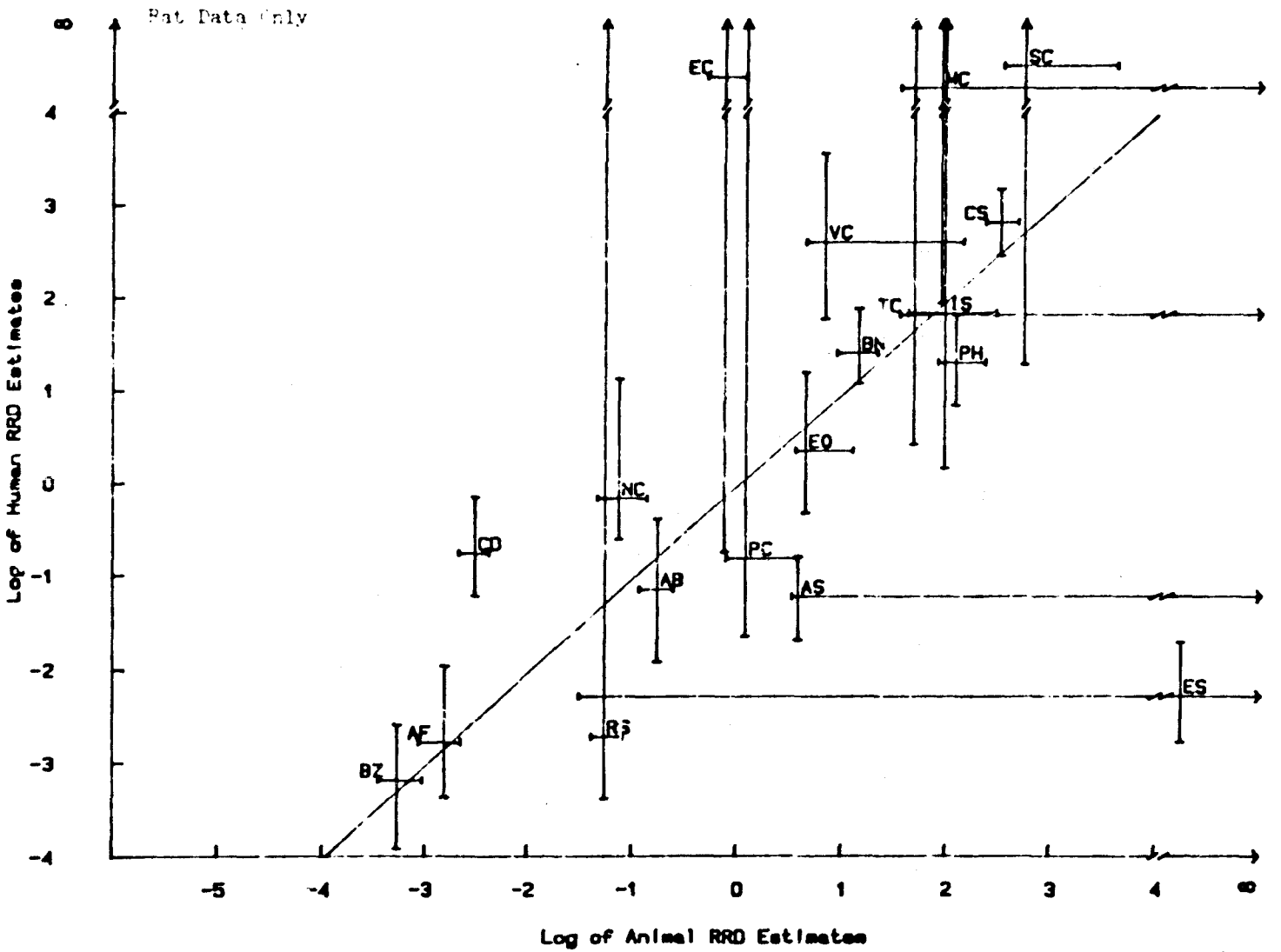


Figure 6

Correlation of Animal and Human RRDs - Analysis 11d

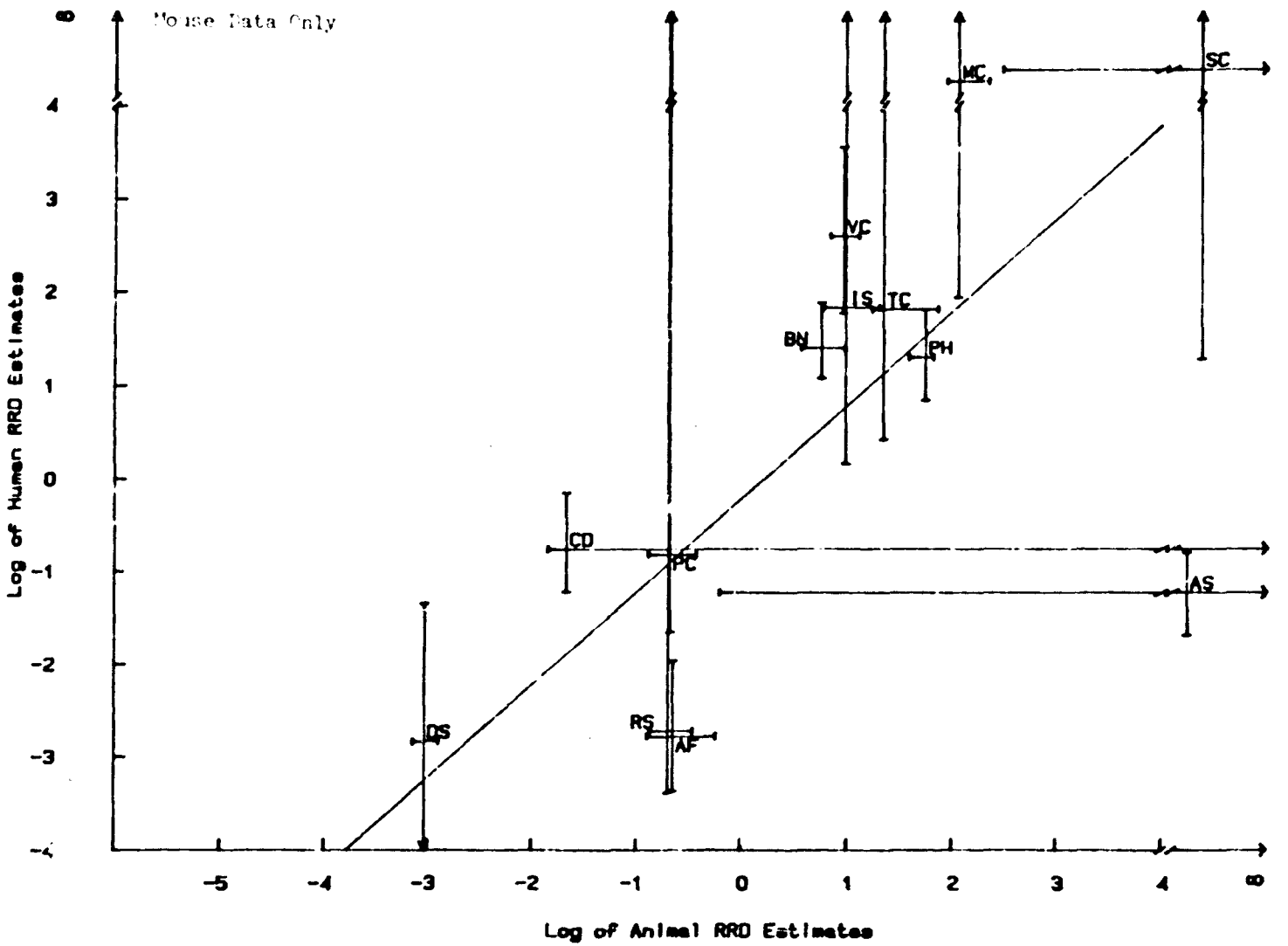


Figure 7

Correlation of Animal and Human RRDs - Analysis 12

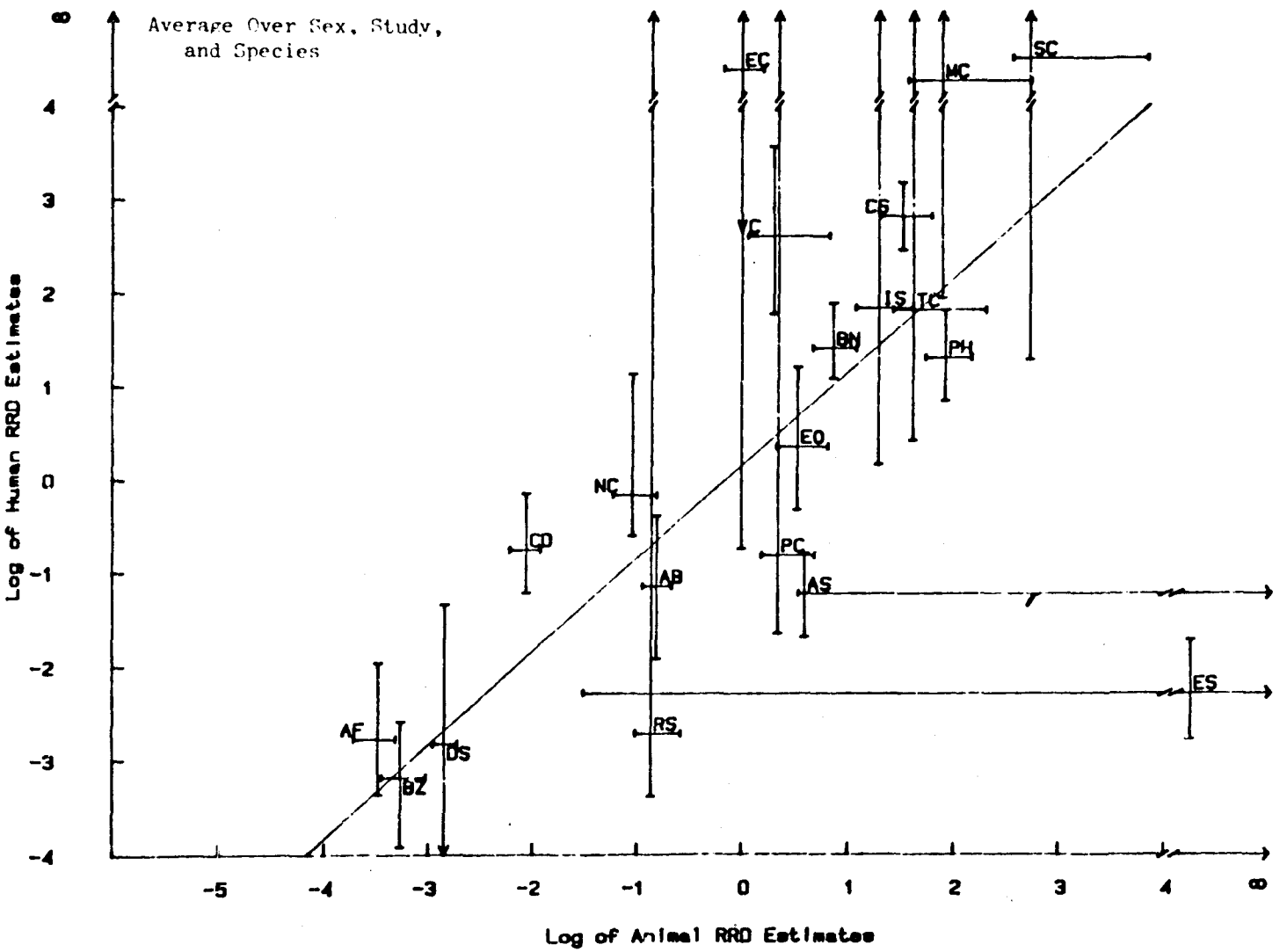


Figure 8

Correlation of Animal and Human RRDs - Analysis 25