PB88-127139

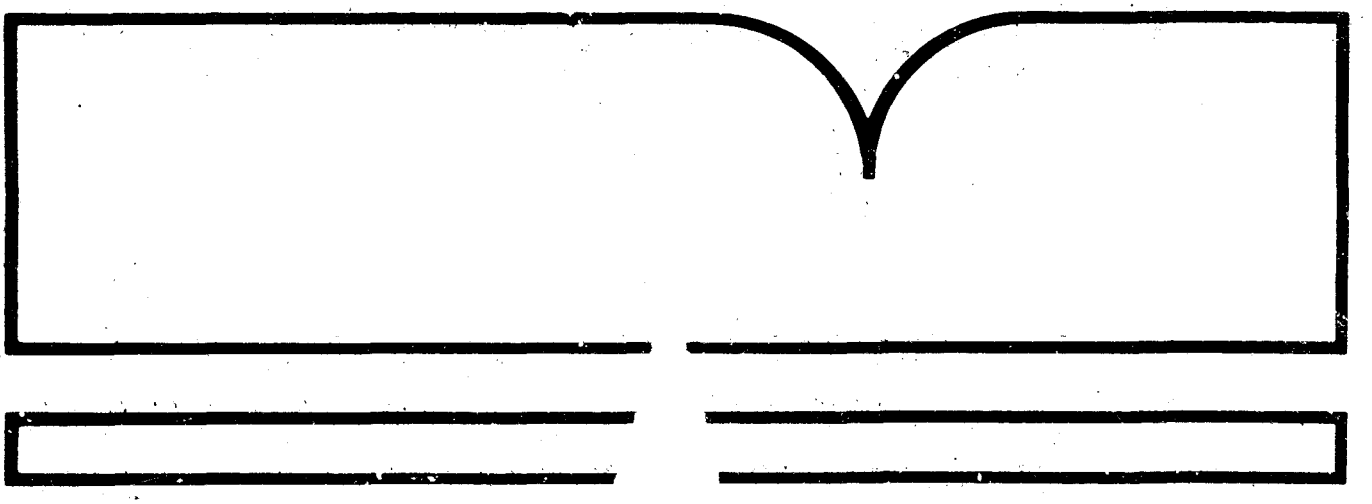
Investigation of Cancer Risk Assessment Methods Volume 3. Analyses

Clement Associates, Inc., Ruston, LA

Prepared for

Environmental Protection Agency, Washington, DC

Sep 87



TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/6-87/007d	2.
3. PB88-127139	
4. TITLE AND SUBTITLE Investigation of Cancer Risk Assessment Methods: Volume 3. Analyses	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. DESCRIPTOP	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT Distribute to public	19. SECURITY CLASS (This Report) Unclassified	21. NO. OF PAGES 196
	20 SECURITY CLASS (This page) Unclassified	22. PRICE

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

ILLUSTRATIONS

TABLES

TABLES

Section 1

METHODOLOGY

INTRODUCTION

One goal of this project is to examine various methods for analyzing bioassay data to determine which methods produce results that correlate well with the results obtained from epidemiological data and to characterize the uncertainties involved. For this to be possible, reasonable, alternative methods of analysis need to be defined. Recall that in the introductory section (in Volume 1 of this report) were listed the components of risk assessment and several approaches for each component; that list is reproduced in Table 1-1. Consider Figure 1-1, which depicts the process of risk assessment based on bioassay data: for several experiments in each of a few species, particular carcinogenic responses yield estimates of RRDs that are combined in some way to yield the final estimate. The components listed in Table 1-1 correspond to the different levels in the tree shown in Figure 1-1 and the approaches specify how to handle the corresponding level. The basic method for defining analysis methods has been to select different combinations of the approaches, as is described in this section.

Also in this section is a description of the methods used to compare the bioassay-based results to the epidemiologically derived estimates. A nonparametric generalized rank test is used to evaluate the correlation between the two sets of estimates. When specific point estimates from the bioassay analyses are employed as predictors, their performance is compared on the basis of the fit of a straight line with slope of one to the data. Three approaches (loss functions) used to fit the line are described.

ANALYSIS OF BIOASSAY DATA

For each chemical being analyzed, the procedure described here is followed to derive the RRDs of interest. For each carcinogenic response coded from a study testing the chemical of interest, the multistage model that best describes the response rates from all dose groups is fit to the dose-response data. The multistage model has the form

$$P(d) = 1 - \exp(-(q_0 + q_1 d + \ldots + q_k d^k)), \qquad (1\text{-}1)$$

where P(d) is the probability of cancer when exposed to average daily dose d; $q_0, q_1, \ldots, q_k \geq 0$ and k is equal to one less than the number of dose groups. The model is fit by an updated version of GLOBAL82 (1) that gives maximum likelihood, lower bound, and upper bound estimates for the dose D such that

$$\frac{P(D) - P(0)}{1 - P(0)} = 0.25,$$

i.e. D is the dose corresponding to an extra risk of one in four. This dose will be called a risk related dose (RRD) corresponding to a risk of one in four. Similar definitions of doses corresponding to the particular levels of risk can be found in the literature. Sawyer *et al.* (2), for example, discuss "TD50", the daily dose required to halve the probability of remaining tumorless.

Actually, the model is fit to each combination of dose and response values that might arise by combination of the approaches listed in Table 1-1. In particular, the components that affect the fitting of the model are numbered 4, 5, and 6 in that table; 20 combinations of the approaches to these components are possible. Hence, as many as 20

models have been fit to each response. (Many responses have been fit by only 10 models since the data needed to analyze only the effective number of animals, approach b to component 6, are not generally available in the published literature.) The triple of estimates composed of the 95% lower bound, the MLE, and the 95% upper bound for the RRD corresponding to an extra risk of 0.25, which is labeled (D_L, D_M, D_U), is available for each of the models fit to each response.

Definition of Analysis Methods

Each analysis method specifies which species of animal to consider, what criteria the experiments on those species must satisfy, which responses within those experiments to consider, and which of the 20 model results to use. (Throughout this report, "experiment" denotes the data from all dose groups in a single bioassay of one species and one sex of test animal, except when results for two sexes are reported together and cannot be separated.) In every case, the first step is to assign one triple to each experiment, selecting the triple from the responses that are eligible for that method (components 7 and 8). The triple that is selected is the one that has the smallest D_L, lower limit on RRD. This procedure is adopted because we are interested in the evidence for carcinogenicity and the manipulation of that evidence. The eligible response with the smallest D_L is the one that is consistent with the highest carcinogenic potential and may, therefore, provide the best evidence of carcinogenicity from the experiment under consideration.

Given this method of assignment of RRD triples to experiments, and given the approaches listed in Table 1-1, 88320 possible analysis methods could be defined. Thirty-eight analyses were run as the first set (Table 1-2). The thirty-eight analysis methods fall into five categories that are defined by the manner in which the data from individual experiments are combined to yield the twelve values that are

of interest in this investigation. Those twelve values are the minimum, the first quartile, the median, and the third quartile of the lower bounds, MLEs and upper bounds. The five categories are described below.

No Averaging. The first category of analyses includes those that treat each species, each study within species, and each sex within study separately (approach a for component 9, a for 10, and d for 11; see Table 1-1). Let Y^k_i be the ith lower bound for RRD in species k, and assume that the Y^k_i values are ordered with Y^k_1 the smallest and $Y^k_{n(k)}$ the largest; n(k) is the number of experiments in species k. Define the species-specific quartile values as follows:

$$Y^k_{1Q} = \begin{cases} Y^k_1 \text{ for } n(k) \leq 4 \\ Y^k_{\lfloor n(k)/4 \rfloor} \text{ for } n(k) \geq 5 \end{cases} \qquad (1\text{-}2)$$

$$Y^k_{2Q} = \begin{cases} Y^k_1 \text{ for } n(k) \leq 2 \\ Y^k_{\lfloor (n(k)+1)/2 \rfloor} \text{ for } n(k) \geq 3 \end{cases}$$

$$Y^k_{3Q} = \begin{cases} Y^k_{n(k)} \text{ for } n(k) \leq 4 \\ Y^k_{n(k)-\lfloor (n(k)/4 \rfloor} \text{ for } n(k) \geq 5 \text{ .} \end{cases}$$

Then the minimum and quartile values of the lower bounds for the analysis are defined by

$$Y_{min} = \min_k (Y^k_1) \qquad (1\text{-}3)$$

$$Y_{1Q} = \min_k (Y^k_{1Q})$$

$$Y_{2Q} = \text{median}_k (Y^k_{2Q})$$

$$Y_{3Q} = \max_k (Y^k_{3Q}).$$

The minimum and maximum over species are adequate to define the first and third quartiles, respectively, because rarely are there more than

four species tested for any chemical. The MLE and upper bound values are defined in exactly the same manner. Analyses 0 through 8c, 11c, 11d, and 25 are in this category.

Averaging Over Sex. The second category of analyses is represented by analysis 9, where the results from different sexes tested in the same study are combined. A study may include as many as two experiments. Two experiments are considered to be from the same study only if they were carried out in the same laboratory by the same experimentors, the same moiety of chemical was used, the same strain of animal was tested, the numbers of animals initially on test were nearly identical, and the study protocols were nearly identical. If that is the case, this analysis methods calls for harmonically averaging the values from the two experiments, lower bound with lower bound, MLE with MLE, and upper bound with upper bound. The weights for the average are equal to the initial numbers of animals on test in each experiment. After the averaging, one triple is associated with each study. These can be ordered, the species-specific quartiles defined, and the minimum and quartile values for the analysis defined in exactly the same manner as the first category.

Averaging Over Study. The third category entails combining studies within species (Analysis 10). Note that different experiments falling under the same study are not averaged, so each study may contain more than one triple of estimates. Once again, let Y^k_i be the ith ordered lower bound from an experiment testing species k (the same procedure is followed for MLEs and upper bounds). Species-specific minimum values are obtained by harmonically averaging the smallest Y^k_i values from each study. Species-specific quartiles are obtained by randomly sampling a single Y^k_i value from each study and then harmonically averaging the values selected. A total of 100 samples is taken for each species so that when the averages are ordered, the 25th, 50th, and 75th estimate

the first, second, and third quartiles, respectively. The weight attached to each study for the harmonic averages is the total number of animals initially on test from all experiments under that study. The minimum value associated with the analysis is the smallest of the species-specific minimums and the quartiles associated with the analysis are defined from the species-specific quartiles as shown in Eq. 1-3.

Averaging Over Species. Examples of the fourth category of analyses are provided by Analyses 11a and 11b, in which results from different species are averaged. Once again, species-specific results are averaged harmonically; in this case an unweighted average is used. To obtain the minimum average lower bound, one selects the smallest lower bound found among experiments in each species, then these species minimums are averaged. The lower bound quartile values associated with the analysis are estimated by random sampling: 100 times, a lower bound is randomly selected from each species and the average computed. When ordered, the 25th, 50th, and 75th average represent the first, second, and third quartile, respectively. The same procedure is followed for MLEs and upper bounds.

Averaging Over Sex, Study, and Species. The final category of analyses includes Analyses 12 through 24d. In this category, results are sequentially averaged over experiment within study, over study within species, and finally, over species. Note that at each step, one triple (averaged) is associated with each study, species, and analysis, respectively. Consequently, only one average lower bound, one average MLE, and one average upper bound is associated with such an analysis. As a result, the minimum and all quartiles of the lower bounds are the same, i.e. the one overall average of lower bounds. The same is true of the MLEs and upper bounds. All averaging is accomplished by a harmonic average with weights equal to the initial number of animals on test when averaging over experiment within study, the total number of animals from

all experiments contained within a study when averaging over study within species, and the constant 1 when averaging over species.

Description of Approaches to Components

The previous discussion has described the manner in which the different analysis methods are defined. In so doing, it has described the approaches to several of the components of risk assessment. In particular, it has been shown how the approaches to components 4, 5, and 6 combine to yield 20 possible sets of input for the multistage model for each carcinogenic response. In addition, the approaches to components 9, 10, and 11 combine to define the five categories of analyses presented above. Some of the approaches to these and other components may require further explanation. This is provided below.

Length of Observation. For component 1, the length of observation criterion, when any experiment is considered (approach a) a correction factor of the form

$$(T_e/T)^3$$

is multiplied by the RRD estimates to adjust the latter for experiments of length T_e. T is the length of the standard experiment for the animal on test (Table 1-3). If T_e is greater than T, no correction is applied. When using approach b, i.e. when including only those experiments for which T_e is 90% or more of T, no correction factor is applied to any experiment.

Units of Human-Animal Equivalence. The approaches to component 4 specify the units assumed to yield equivalence between humans and animals with respect to carcinogenic potency of the test chemical. When

possible, experiment-specific, indeed dose-group-specific, body weights and food intakes have been used to convert among the units. Standard values (Table 1-3) have been used when necessary.

Calculation of Average Dose. Component 5 relates to the calculation of average doses for each dose group. Either all dosing is considered (approach a), in which case the average is calculated over the entire experiment, or only dosing over the first 80% of the experiment is considered (approach b), in which case average daily dose is based on that time period as well. The 80% figure is predicated on the assumption that exposures during the last 20% of the life of an animal are unlikely to affect cancer incidence due to the latency period. Crump and Howe (4) observe that "chemically induced tumors are apt to have a latent period of about 1/5 of the life span of the species." (The same assumption is used when specifying approach b to component 2, i.e., when restricting experiments to those with "long" dosing of 80% or more.) As an example of the difference between the approaches to component 5, consider a 100-week study with a dose group receiving 1 mg/kg/day for 90 weeks. In approach a, average daily dose is calculated as

$$(90 \cdot 1 + 10 \cdot 0)/100 = 0.9 \text{ mg/kg}.$$

For approach b, the average daily dose is

$$(80 \cdot 1)/80 = 1 \text{ mg/kg}.$$

Tumor Type to Use. As discussed in Volume 2 of this report, special codes have been designated for two types of response, all tumors and the combination of all significantly increased tumors. The codes used, in general, are based on the International Classification of Diseases for

Oncology (5). The topology-morphology code applied for all tumors is 1000-8000; for the combination of significantly increased tumors it is 1000-7000. These are the responses included in approaches b and a, respectively, to component B.

Definition of Standard Methods

Analysis O resembles the procedure employed by EPA's Carcinogenic Assessment Group in many respects: $mg/m^2/day$ are the units assumed to yield human and animal equivalence; species, studies, and experiments are not combined so that the minimum lower bound comes from the most sensitive species and sex and from the experiment yielding the smallest RRD lower bound (largest upper bound on risk); and route of exposure is limited to the more common routes, inhalation, gavage, and oral, unless humans are exposed by some other route. Of course, no automatic procedure can exactly duplicate the decision-laden process of risk assessment. Nevertheless Analysis O is one reasonable procedure and, more importantly for this project, is the one that serves as a template for defining other analyses.

Another standard method has been defined. It is called Analysis 30 and has been used as a template to define an additional set of twenty analysis methods (Table 1-4). Analysis 30 differs from Analysis O in that mg/kg/day rather than $mg/m^2/day$ are the units assumed to yield equivalence between humans and animals for extrapolation of RRD estimates. Moreover, the route of administration of the test chemicals is not limited to any particular route; injection and instillation studies are included along with gavage, oral, or inhalation experiments. The eighteen methods that are single-component variations of Analysis 30, *i.e.* Analyses 31 through 50 (Analyses 39 and 40 were not performed), are not duplicated in Analyses 1 through 25, except for

Analysis 31 which is the same as Analysis 3b and Analysis 38 which is the same as 4a.

This alternative standard and its single-component variants are not used in the majority of the analyses performed for this project. That set of methods was defined only after the bulk of the analysis was completed. Its purpose is to provide information on the uncertainty associated with single components of the risk assessment process. It is used to investigate component-specific uncertainty or variability in the manner described later in this section. However, information on the predictiveness of the estimates from the supplemental analyses is considered when the best method(s) are identified.

Table 1-5 gives a verbal description of the 38 initial analyses and the 19 supplemental analyses.

Sieve

In addition to criteria restricting the type of experiments that are used in some analyses, another procedure has been set up to select subsets of the data for analysis. This procedure is called a sieve and operates as follows. Criteria are defined that rank experiments in terms of preference for analysis. Say a rank of 1 is preferred over 2, 2 over 3, etc. The experiments and responses that are used in any specific analysis are those that have the lowest rank; if there are any rank 1 data sets those and only those are used, if no rank 1 data sets are available all the rank 2 data sets are used, etc. This procedure is an attempt to use the best data that are available but yet to do something when the best type of data is unavailable.

The sieve may have more than one level. That is, a selection from among the experiments may be made on the basis of one criterion and then the

selected bioassays may be subjected to further screening on the basis of another criterion. In each case, the best, the lowest rank, data sets survive the screening and become available for analysis. If one criterion is based on properties of the carcinogenic responses (e.g., rank 1 is given to responses that show a significant relationship to dosing) as opposed to another criterion that is based on features of the entire experiment (e.g., rank 1 is given to experiments that have at least 50 dosed animals) the former screening is applied first. In that way, the greatest amount of data passes from one level of the sieve to the other; individual responses, not entire experiments, are eliminated at the first stage. Also, for the example given above, a significant response (*i.e.* evidence of carcinogenicity) is to be preferred, no matter how many animals are tested. This is not guaranteed to happen if the first screening is based on number of dosed animals.

The sieve technique is designed to work with any of the analyses defined in terms of the components of risk assessment as described above. The sieve is applied only after any inclusion criteria specific to an analysis. For example, in Analysis 1, only experiments that lasted at least 90% of the standard experiment length are included; the sieve is applied after that selection is made. Note also that the selections that define the analyses are unlike the selection procedure for the sieve. The analysis-defining selections do not rank studies. If there are not experiments that lasted at least 90% of the standard experiment length for a chemical, that chemical is not included in Analysis 1. The sieve technique does rank experiments so that the best can be used; a chemical cannot be excluded from an analysis because of the action of the sieve.

The sets of analyses that employ a sieve use one or both of two screens. The first examines each response to see if a significant dose-related effect on response rates is evident. Priority (*i.e.*, the lowest rank)

is given to those responses that do show such a significant relationship. The second screen is based on a combination of two features of each experiment, the length of observation and the number of dosed animals.

Significance of a relationship between dose and response rates is assessed by use of the Fisher's exact test (6) and the Cochran-Armitage trend test (7). If the response rate in any treated group is significantly greater than that in the control group at the 0.05 level as determined by Fisher's exact test or if the trend of response rates is significant at the 0.05 level as determined by the Cochran-Armitage trend test, then the response is considered significant and is given lower rank. Otherwise, the higher rank is assigned. This screening is called the significance screening.

The ranking scheme based on experimental protocol, *i.e.* on length of experiment and number of treated animals, is depicted in Table 1-6. Note that this is just one of infinitely many ranking schemes possible. Of the two features (experiment length and number of dosed animals) slightly more weight, in terms of the perceived quality of the study, has been given to length of observation. This part of the sieve is labeled the quality screen.

Given the two screens described above, four sets of analyses have been defined, one with no screenings, one with the significance screen alone, one with the quality screen alone, and one with both screens. As described earlier, when both screens are used, the significance screen applied to individual responses operates before the quality screen. Of course, the entire sieve procedure comes into play only after the application of the exclusion criteria that define each analysis method.

COMPARISON WITH EPIDEMIOLOGICAL RESULTS

Once the bioassay data have been analyzed by the many methods defined, one wishes to use the results to compare and evaluate those methods. The techniques that have been selected to do this use the RRD estimates obtained from the epidemiological data as the basis for comparison. A method of bioassay analysis that yields estimates "close to" the epidemiologically derived estimates is judged to be satisfactory. The following describes the techniques for determining how close a set of bioassay-based estimates is to the human-based estimates.

Correlation Analysis

In the analysis of the epidemiological data, we 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_{HU} and RRD_{HL}, respectively. The interval $[RRD_{HL}, RRD_{HU}]$ represents the range of estimates that are in some sense consistent with the epidemiological data because of data uncertainty or statistical variability.

Because of the many bioassays for any given chemical and because of statistical variability within each bioassay, the bioassay analysis results may also be reasonably characterized by a range of RRD estimates The interval selected in the correlation analysis to represent that range is defined by the median RRDs; it extends from the median of the lower bound estimates to the median of the upper bound estimates. That choice of interval considers statistical variability in the sense that both lower and upper statistical confidence limits are used in its definition. Moreover, the use of median values avoids some difficulties with outliers and behaves properly in an asymptotic sense. Should anomalous results appear in some bioassays, estimators of the

appropriate range such as that from the minimum lower bound to the maximum upper bound are adversely affected. Such "minimum-to-maximum" ranges are highly sensitive to outliers and once outliers appear, those estimators are not self-correcting. As more bioassays of a particular chemical are performed (and so as the chance of outliers increases) the minimum lower bound and maximum upper bound estimates can only get more extreme. Median values, on the other hand, should behave more properly in the sense of discounting truly anomolous results and converging to the "true" value. Let $[L_{2Q}, U_{2Q}]$ be the interval from the median (second quartile) of the lower bound RRD estimates to the median of the upper bound RRD estimates obtained for any given chemical.

We are interested in the correlation between the epidemiologically based estimates and the bioassay-based estimates. That is, we wish to know if chemicals with larger estimated human RRDs also tend to have larger estimated animal RRDs. In the absence of known or suspected distributions for the RRD estimates, nonparametric tests of correlation are appropriate. The standard nonparametric measures of correlation (notably Spearman's rho) use the ranks of point estimates without consideration of variabilities. When the variability or uncertainty about point estimates is not the same for each observation, as is the case with the data in the present analysis, such a method may be inappropriate. Ng (8) has proposed a concept of generalized ranks that does consider variabilities as reflected in the intervals surrounding a point estimate. That concept is used to rank the human intervals, (RRD_{HL}, RRD_{HU}), and, separately, to rank the animal intervals, (L_{2Q}, U_{2Q}), and to determine the degree of correlation between the two sets of intervals.

A partial ordering for the intervals is defined as follows (the definition is given in terms of the animal intervals; exactly equivalent definitions hold for the human intervals). Let interval i,

corresponding to chemical i, be labelled $(L_{2Qi}, U_{2Qi}) = I_i$. Then I_i is less than I_j if $L_{2Qi} < L_{2Qj}$ and $U_{2Qi} < U_{2Qj}$ (if $U_{2Qi} = U_{2Qj} = \infty$, then I_i is less than I_j if $L_{2Qi} < L_{2Qj}$). I_i is greater than I_j if I_j is less than I_i. Otherwise, I_i and I_j cannot be ordered (we will say they are "tied").

A ranking of the intervals can be defined on the basis of the partial ordering. Let n_i be the number of intervals less than interval I_i and let m_i be the number of intervals tied with I_i. Define the rank of I_i, R_i, to be

$$R_i = n_i + m_i/2 + 1.$$

We will use R_i to denote the rank of the ith chemical when based on the animal intervals and S_i to denote the rank of the ith chemical when based on the human intervals. Ng (8) has shown that the ranks so defined have desirable properties including the fact that the sum of the ranks, ΣR_i or ΣS_i, is $N\cdot(N+1)/2$ (*i.e.* the sum of the ordinary ranks of N numbers) and that these generalized ranks reduce to ordinary ranks if the partial ordering is also a total ordering. The R_i and S_i values are used here to estimate correlations.

By analogy to Spearman's rho, a correlation coefficient, ρ, is defined as follows:

$$\rho = \frac{(\sum_i R_i S_i) - 1/4(N(N+1)^2}{[\sum_i (R_i-\bar{R})^2 \cdot \sum_i (S_i-\bar{S})^2]^{1/2}} \qquad (1\text{-}4)$$

Note that $\bar{R} = \bar{S} = (N+1)/2$. The statistic ρ behaves appropriately for a measure of correlation ($-1 \leq \rho \leq 1$; larger positive (negative) values indicate more positive (negative) correlation; etc.).

The significance of ρ is assessed by simulation. The R_i's are held fixed at their observed values while the S_i's are permuted over the set of observed S_i's. That is, each permutation sets $p(i)$, $i = 1,2,\ldots,N$, from the set $(1,2,\ldots,N)$ such that $p(i) \neq p(j)$ for $i \neq j$. Let $S_i' = S_{p(i)}$. The correlation coefficient is evaluated for each permutation (in the numerator of equation 1-4, R_i is paired with S_i'). If among a total of M permutations, K of them yield coefficients at least as large as the observed ρ, then the probability of observing a coefficient as large as or larger than ρ under the null hypothesis of no correlation is estimated to be K/M. The null hypothesis is rejected in favor of the alternative, $\rho > 0$, for small values of K/M. In the present analysis, 10,000 permutations were created (M = 10,000).

Prediction Analyses

The correlation analysis just discussed concentrates on intervals of RRD estimates to determine whether or not the human and animal estimates generally behave in the same way (*i.e.*, RRDs for chemical i are lower when estimated from the epidemiology when they are lower when based on the bioassay). If that correlation analysis is positive, then it is reasonable to go on to ask if particular points obtained via bioassay analysis are good predictors of the results that are obtained directly from epidemiological investigation. At this stage also, one can examine the magnitude of errors, *i.e.* the uncertainty that results from the use of any predictor. The following is a description of the methods employed to compare and evaluate various predictors.

Unlike the correlation analysis, the prediction analysis selects a single point from the bioassay analysis results as the estimate of RRD for each chemical. Each of the 38 analyses described in Table 1-2 could supply any number of predictors. The four that have been investigated are the minimum of the lower bound estimates, L_M, the median of the

lower bound estimates, L_{2Q}, and the minimum and median of the maximum likelihood estimates, MLE_M and MLE_{2Q}, respectively. These values are available for each chemical analyzed by each of the thirty-eight methods.

The behavior and properties of the predictors are assessed, again, by comparison with the epidemiologically derived estimates. Those human estimates are <u>not</u> distilled to a single point. Instead the best estimate, RRD_H, and/or the interval from RRD_{HL} to RRD_{HU} form the basis for evaluating the predictors. In particular, a straight line with slope of 1 is fit to the base ten logarithmic transform of predictor values and the logarithmic transform of the human estimates. That is, the bioassay-based estimate of the human RRD corresponding to a risk of one in four, H_A, is given by

$$\log_{10}(H_A) = \log_{10}(P_i) + C,$$

where P_i is the predictor from the analysis (either L_M, L_{2Q}, MLE_M, or MLE_{2Q}) for chemical i and C is the y-intercept to be estimated. This relationship implies that

$$H_A = P_i \cdot 10^C,$$

i.e. that a linear relationship exists between the untransformed estimates. Consequently, the potency of chemicals as determined from bioassay data relative to the estimates of human potency is not related to their absolute potency, which seems reasonable.

Suppose that $A_i = \log_{10}(P_i) + C$, where P_i is one of the predictors for chemical i from the bioassay data as described above, for any given value of C. The y-intercept, C^*, is determined (*i.e.* the line is fit to the data) by minimizing the sum of the losses for each chemical

associated with the predicted RRD, A_i, and the estimates derived from the epidemiological data. Clearly, a loss function must be defined in order to carry out this procedure.

Three different forms of loss function are considered. The first and simplest, called DISTANCE^2, defines the loss associated with the prediction for chemical i to be proportional to the square of the distance from the predicted value to the interval defined by the lower and upper endpoints of the epidemiologically derived RRDs. Though this formulation of loss is straightforward, it does have some disadvantages. First, it does not consider the best estimates of RRDs obtained from the epidemiological analysis, the RRD_Hs. Moreover, it cannot be applied when the animal predictors can have infinite values. Since MLE_M and MLE_{2Q} can be infinite, but L_M and L_{2Q} cannot, DISTANCE^2 can be used to evaluate only the latter two predictors. This same difficulty with infinite values arises when the loss function utilizes the RRD_K estimates, which may indeed be infinite. Because of these limitations of DISTANCE^2 and because we wish to consider possibly infinite estimates (particularly in RRD_H since we made a point of including in these analyses chemicals that may not be carcinogenic as determined by epidemiological investigation, *i.e.* for which RRD_H is infinite) two other loss functions have been developed. These are called CAUCHY and TANH. All three forms of loss function are described in detail below.

<u>DISTANCE^2 Loss Function</u>. Loss associated with chemical i is defined solely in terms of the interval ($RRD_{HL,i}$, $RRD_{HU,i}$). That loss is given by

$$l_{1,i} = \begin{cases} 0 & \text{if } \log_{10}(RRD_{HL,i}) \le A_i \le \log_{10}(RRD_{HU,i}), \\ d^2 & \text{if } A_i < \log_{10}(RRD_{HL,i}), \\ k \cdot d^2 & \text{if } A_i > \log_{10}(RRD_{HU,i}). \end{cases}$$

Here d is the absolute distance between A_i and the closest of $\log_{10}(RRD_{HL,i})$ and $\log_{10}(RRD_{HU,i})$. The constant k is the asymmetry parameter.

If, as appears reasonable, it is worse to overpredict RRDs, then $k > 1$ can be used to reflect the belief about the degree of asymmetry. Nevertheless, this approach to fitting the line is essentially equivalent to determining the line that is closest to the intervals defined by the lower and upper endpoints of the human estimates. The total loss associated with a particular analysis is the unweighted sum of the losses associated with each chemical in the analysis.

A simple extension of the reasoning presented in the discussion of the loss function l_1 allows definition of a fitting algorithm for results expressed as intervals in both the horizontal and vertical directions. Such results are obtained in the correlation analyses. The extended l_1 routine has been run with the same intervals used to determine correlations. Such a procedure allows us to identify individual chemicals whose intervals of RRD estimates are far from the fitted line and, therefore, may be thought of as outliers and may, in fact, detract from the correlation.

CAUCHY and TANH Loss Functions. Suppose $H_i = \log_{10}(RRD_{H,i})$, the logarithmic transform of the best estimate from the epidemiological analysis for chemical i. Then, recalling that $A_i = \log_{10}(P_i) + C$, we wish to find C that minimizes

$$\sum_i l(A_i, H_i) \cdot W_i \qquad (1\text{-}6)$$

where $l(-,-)$ is a nonnegative loss function, W_i is the weight attached to the loss for chemical i, and the sum runs over all chemicals in the analysis. Considering that A_i or H_i may be infinite, these are properties that we considered it desirable for l to have:

P1: $l(A,H) = 0$ if $A = H$

P2a: $l(\infty, H) < \infty$ and $l(A,\infty) < \infty$

P2b: $l(\infty, \infty) = 0$

P3: $l(A_1, \infty) > l(A_2, \infty)$ if $A_1 < A_2$

P4: $l(A, H) < l(A, \infty)$ for $A, H < \infty$

P5: $l(A_1, H) > l(A_2, H)$ if $A_1 > A_2 > H$ or $A_1 < A_2 < H$

P6: $l(A_1, H) \leq l(A_2, H)$ for $H - A_1 = A_2 - H > 0$.

These properties can be interpreted as stating that loss is minimized only when the prediction matches the observation (P1); that loss is finite even for infinite RRD estimates (P2a) and that predictions of infinite RRDs are good (*i.e.*, have zero loss) when the observed RRDs are infinite (P2b); that the loss is less if the prediction is larger when the observed RRD is infinite (P3); that the loss is greater when the observed RRD is infinite than when the observation is finite if the prediction is finite (P4); that the farther away from the observed RRD one goes in one direction the greater is the loss (P5); and that the loss from underestimating an observed RRD by a certain amount is no greater than the loss from overestimating by that amount (P6). The last property allows one to choose an asymmetric loss function if one wants to reflect the belief that it is worse to overestimate RRDs than to underestimate them.

Unfortunately, it is easy to show that the properties P1 through P6 are mutually inconsistent. Two approaches have been taken to get a set of consistent properties to motivate the choice of loss function. The first approach is to drop property 3. This is the only property that prevents the loss function from being expressed as a function of the distance (A-H). Consequently, we defined l_2 as follows:

$$l_2(A,H) = 1 - (1/(1+f(\sigma(A-H))\cdot(A-H)^2)) \qquad (1\text{-}7)$$

where $\sigma(-)$ is the sign function and $f(-)$ is some positive function allowing the introduction of asymmetry. It is clear that l_2 satisfies P1, P2a, P2b, and P4 - P6. Moreover, once we are given a set of H_i's and P_i's, then we can approximate P3 but retain the other properties. To do so, infinity must be approximated by some large number. That number must be chosen large enough so that P4 remains true (for the given set of H_i's and reasonable lines, *i.e.* values of C) but not so large that the inequality described in P3 gets washed out (*i.e.*, appears only after many decimal places). For this project infinity is estimated by 10^{15}, *i.e.* on the $\log_{10}$ scale by 15. It can be noted in passing that a symmetric version of $l_2(A,H)$ behaves like squared-error loss when A and H are close together. This formulation of loss is called the CAUCHY loss function since it resembles the distribution function of the CAUCHY probability distribution.

As an alternative approach, we can retain P3 but weaken P4 slightly to get

P4': $l(A, H_1) < l(A, \infty)$ for $A < H_1 < \infty$.

The set of properties obtained by replacing P4 with P4' precludes the use of a loss function expressed as a simple function of distance, (A-H), alone. The next best (easiest) approach would be to try

$$l(A,H) = | g(A) - g(H) |.$$

However, since P2a implies that g(H) goes to a limit as $H \to \infty$ and P1 implies that g(H) is monotone, then g(H) must "flatten out" at some point, *i.e.* for H large enough. Hence it will be possible to find A_1, H, and A_2 such that $l(A_1,H) > l(A_2, H)$ even though $H - A_1 = A_2 - H > 0$ (in violation of P6).

Consequently, we are led to try a slightly different loss:

$$l(A, H) = \begin{cases} g(H) - g(A), & A \leq H \\ m[g(H) - g(2H-A)], & A > H, \end{cases}$$

where g is some monotone increasing function and $m \geq 1$. When this is the case, $l(A_1, H) = g(H) - g(A_1)$ and $l(A_2, H) = m[g(H) - g(2H - A_2)] = m[g(H) - g(A_1)]$ when $H-A_1 = A_2-H > 0$, and so P6 is not violated.

If we adopt this form of the loss function, then both $\lim_{x\to\infty} g(x)$ and $\lim_{x\to-\infty} g(x)$ must be finite (by P2a). A monotone increasing function that has all these nice properties is

$$g(x) = \tanh(c_1 x) = (e^{c_1 x} - e^{-c_1 x})/(e^{c_1 x} + e^{-c_1 x}).$$

The constant $c_1 > 0$ can be thought of as a scaling factor. The resulting loss function (called TANH) is

$$l_3(A,H) = \begin{cases} \tanh(c_1 H) - \tanh(c_1 A), & A \leq H \\ m[\tanh(c_1 H) - \tanh(c_1(2H-A))], & A > H. \end{cases} \qquad (1\text{-}8)$$

The factor m is chosen to reflect the desired degree of asymmetry. For this investigation, asymmetry considerations have been examined by setting m equal to 1.5, 2, 5, 10, 50, and 100. Larger values of m reflect stronger beliefs about the undesirability of overestimating RRDs. Small c_1 shrinks everything (except infinity) toward zero where tanh is more nearly linear, so that loss when an infinite value is observed or predicted is exaggerated compared to the loss when both observed and predicted are finite. Small enough c_1 may also make P4 true for any given set of observations and reasonable values of C. A value of 0.1 has been assigned to c_1 throughout these analyses.

Given the alternatives l_2 and l_3, we can calculate the loss contributed by any given chemical. What remains is to specify the weights, W_i, that allow accumulation of the individual losses into an overall loss value as shown in Eq. (1-6). It seems clear that less weight should be given to chemicals whose human RRD estimates are less certain, i.e. to those whose intervals surrounding H_i are longer. Once again, the problem of infinite values exists, in this case infinite interval lengths. Consequently one should consider positive, monotone decreasing functions that go to a positive limit as the representation for the W_i's.

Let $D_i = \log_{10}(RRD_{HU,i}) - \log_{10}(RRD_{HL,i})$. We wish to have

$$W_i = h(D_i),$$

where $\lim_{x \to \infty} h(x) = r > 0$. The function selected is

$$h(x) = \coth^2(x) = ((e^x + e^{-x})/(e^x - e^{-x}))^2.$$

Note that $\lim_{x \to \infty} \coth^2(x) = 1$. Also consider the following. If we were doing ordinary weighted least squares, we would want the weights proportional to the inverses of the variances. In our case, we have quasi-variance represented by the intervals from RRD_{HL} to RRD_{HU}. In the ordinary situation, the intervals would be proportional to the standard deviations and so weights could be formulated in terms of the inverses of the interval lengths squared. Note that coth(x) behaves like 1/x for x close to zero, so that $\coth^2(x)$ would behave like $1/x^2$. That is, if we choose to use $\coth^2(D_i)$, we have a function that mimics ordinary least squares for small D_i but that allows us to consider infinite-valued D_i.

For each analysis method and for each predictor, lines have been fit to the results using both loss functions l_2 and l_3. In both cases

$W_i = \coth^2(D_i)$ is the weighting scheme employed. For those predictors that are guaranteed to be finite (L_M and L_{20}) the loss defined by l_1 (distance to the interval) with W_i's set equal to 1 have also provided estimates of the best fitting lines. Average loss for any analysis and for any loss function is the total loss (weighted sum of the chemical-specific losses) divided by the sum of the weights.

Uncertainty

Two types of uncertainty are investigated in this project, what we have called residual uncertainty and component-specific uncertainty. The methods for quantifying these uncertainties are described below.

Residual Uncertainty. The lines fit to the data using any of the loss functions described above will not eliminate all uncertainty. That is, there will remain differences between the values predicted on the basis of the best-fitting line and the observed epidemiologically derived estimates. The DISTANCE^2 loss function is used to quantify those differences.

Let A_i be the prediction (in this case from the DISTANCE^2-fitted line) for chemical i in any particular analysis. Let G_i be defined as follows:

$$G_i = 1 \text{ if } \log_{10}(RRD_{HL,i}) \leq A_i \leq \log_{10}(RRD_{HU,i}),$$
$$10^{A_i}/RRD_{HU,i} \text{ if } A_i > \log_{10}(RRD_{HU,i})$$
$$RRD_{HL,i}/10^{A_i} \text{ if } A_i < \log_{10}(RRD_{HL,i}).$$

Then the average of the G_i's over all the chemicals in the analysis yields a result called the residual uncertainty factor. It is the average factor by which the predicted values must be multiplied or divided in order to yield predictions that lie within the intervals

defined by the $RRD_{HL,i}$'s and the $RRD_{HU,i}$'s.

Alternatively, one can examine separately those chemicals for which the human RRDs are overestimated by the best-fitting line $[A_i > \log_{10}(RRD_{HU,i})]$ and those for which they are underestimated $[A_i < \log_{10}(RRD_{HL,i})]$. These are the two groups of chemicals for which the best-fitting line does not intersect the interval of human RRD estimates. Factors analogous to the uncertainty factor defined above, but pertinent to one or the other of these two groups, are defined in a slightly different manner.

For those chemicals whose human intervals lie completely below the line $[A_i > \log_{10}(RRD_{HU,i})]$, recall that their contribution to the symmetric DISTANCE^2 loss is of the form

$$d^2 = (A_i - \log_{10}(RRD_{HU,i}))^2.$$

Let D^2 be the average value of these d^2 values, *i.e.* D^2 is an average squared distance on the log scale. Then an average factor by which the natural-scale predictions (10^{A_i}) must be multiplied in order to get predictions within the human intervals is 10^{-D}, where D is the positive square root of D^2.

Similar results hold for those chemicals with intervals that lie above the line. In that case, the contribution to the DISTANCE^2 loss is

$$d^2 = (\log_{10}(RRD_{HL,i}) - A_i)^2.$$

Let E^2 be the average of these squared distances. Then 10^E is the average factor by which the natural-scale predictions must be multiplied to give values with the human intervals. E is the positive square root of E^2. Both 10^{-D} and 10^E are presented.

Component-Specific Uncertainty. A histogram approach has been used to investigate the uncertainty associated with any one component of risk assessment. Only the supplemental analyses (Analyses 30 through 50) are used in this investigation. However, all chemicals with relevant animal bioassay data can be used since epidemiological data is not required.

For any given predictor (e.g., the median lower bound RRD estimate) each analysis results in a single result for each chemical, P_i. Let us denote the dependence of the results on the analysis by letting $P_{X,i}$ be the result for chemical i in Analysis X. Component-specific uncertainty addresses the issue of how the $P_{X,i}$ values change with the analyses, X. The investigation of this uncertainty is limited to analyses that differ from a standard analysis (Analysis 30) in only one component. Analyses 31 through 50 are such single-component variants of Analysis 30. The ratios $R_{X,i} = P_{X,i}/P_{30,i}$, where $X = 31,32,\ldots,50$, are the raw data for this component-specific uncertainty investigation.

For each Analysis, X, for X between 31 and 50, inclusive, there is a corresponding histogram of the ratios, $R_{X,i}$. The cut points of the histogram are 0, 0.01, 0.02, 0.05, 0.10, 0.20, 0.50, 0.80, 1.25, 2.0, 5.0, 10.0, 20.0, 50.0, 100.0, and ∞. Each ratio is located in one of the subintervals defined by these cut points. Its location indicates how the results for the corresponding chemical change when the component associated with the analysis (the one that differs from the standard choice, that in Analysis 30) is changed.

For each histogram, the mode is determined. Moreover, a dispersion factor (≥ 1) is defined that indicates how spread out the ratios are. Let I be the subinterval containing the mode of the distribution and let G_I be the geometric mean of the endpoints of interval I. For example, if I = [0.8, 1.25], then $G_I = ((0.08)\cdot(1.25))^{1/2} = 1$. Generally, let J

be any subinterval with geometric mean G_J. For the intervals on the ends of the histograms, [0, 0.01] and [100, ∞], the geometric means are determined from the ratios that fall within them. If, for instance, two ratios are greater than 100, say 400 and 1000, then the geometric mean for the interval [100, ∞] in the histogram in question is $((400)\cdot(1000)^{1/2} = 632$. This procedure is followed since the entries in the intervals on the ends of the histograms may vary over many orders of magnitude, unlike the entries in any other interval. It does not appear reasonable or consistent to fix means in these cases.

The dispersion factor for any histogram is defined as

$$\frac{\Sigma_J(N_J \cdot G_J/G_I)}{\Sigma_J N_J}$$

where N_J is the number of ratios (chemicals) in interval J and the sums run over all intervals.

The dispersion factor indicates the average factor by which the ratios differ from the mode. A dispersion factor of 1 corresponds to a histogram with all the ratios in one subinterval. Since the mode can be construed as an indication of the direction and magnitude of the change in RRD estimates when the approach to a single component is changed, the dispersion factor indicates how consistent that change in estimates is. It is the average factor by which the RRD estimates from the altered (nonstandard) analysis must be multiplied or divided to yield ratios in the interval of the mode. Since the altered analyses differ from the standard in only one component and a histogram corresponds to one altered analysis, a dispersion factor is associated with one component and indicates how well-determined are the changes that result from a change in approach to that component.

REFERENCES

1. 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.

2. Sawyer, C., Peto, R., Bernstein, L., and Pike, M. C. (1984). Calculation of carcinogenic potency from long-term animal carcinogenesis experiments. Biometrics 40:27-40.

3. Freireich, E., Gehan, E., Rall, D., Schmidt, L., and Skipper, H. (1966). Quantitative comparison of toxicity of anti-cancer agents in mouse, rat, hamster, dog, monkey, and man. Cancer Chemotherapy Reports 50:219-244.

4. Crump, K. and Howe, R. (1980). A Small Sample Study of Some Multivariate and Dose Response Permutation Tests for Use with Teratogenesis or Carcinogenesis Data. Prepared for the Food and Drug Administration under contract to Ebon Research Systems, 34 pages.

5. International Classification of Diseases for Oncology (1976). World Health Organization. Geneva, Switzerland.

6. Bickel, P. J. and Doksum, K. A. (1977). Mathematical Statistics. Holden-Day, Inc., San Francisco.

7. Armitage, P. (1955). Tests for linear trends in proportions and frequencies. Biometrics 11:375-386.

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

Table 1-1

APPROACHES TO RISK ASSESSMENT COMPONENTS

1. Length of the 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 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.

Table 1-1 (continued)

APPROACHES TO RISK ASSESSMENT COMPONENTS

7. Malignancy status to consider

   a. Consider malignant tumors only.
   b. Consider both benign and malignant tumors.

8. Particular tumor type to use

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

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 Initial Standard (Analysis 0).

Table 1-2

APPROACHES USED FOR INITIAL THIRTY-EIGHT ANALYSES

Analysis	Component										
	1	2	3	4	5	6	7	8	9	10	11
0	a[a]	a	c	e	a	a	b	d	a	a	d
1	b*	a	c	e	a	a	b	d	a	a	d
2	a	b*	c	e	a	a	b	d	a	a	d
3a	a	a	a*	e	a	a	b	d	a	a	d
3b	a	a	b*	e	a	a	b	d	a	a	d
4a	a	a	c	a*	a	a	b	d	a	a	d
4b	a	a	c	b*	a	a	b	d	a	a	d
4c	a	a	c	c*	a	a	b	d	a	a	d
4d	a	a	c	d*	a	a	b	d	a	a	d
5	a	a	c	e	b*	a	b	d	a	a	d
6	a	a	c	e	a	b*	b	d	a	a	d
7	a	a	c	e	a	a	a*	d	a	a	d
8a	a	a	c	e	a	a	b	a*	a	a	d
8b	a	a	c	e	a	a	b	b*	a	a	d
8c	a	a	c	e	a	a	b	c*	a	a	d
9	a	a	c	e	a	a	b	d	b*	a	d
10	a	a	c	e	a	a	b	d	a	b*	d
11a	a	a	c	e	a	a	b	d	a	a	a*
11b	a	a	c	e	a	a	b	d	a	a	b*
11c	a	a	c	e	a	a	b	d	a	a	c*[b]
11d	a	a	c	e	a	a	b	d	a	a	c*[b]
12	a	a	c	e	a	a	b	d	b*	b*	a*
13	b*	b*	c	e	a	a	b	d	b*	b*	a*
14	a	a	c	e	a	a	a*	d	b*	b*	a*
15	b*	b*	c	e	a	a	a*	d	b*	b*	a*
16	a	a	c	e	a	a	b	a*	b*	b*	a*
17	b*	b*	c	e	a	a	b	a*	b*	b*	a*
18	a	a	c	e	a	a	a*	a*	b*	b*	a*
19	b*	b*	c	e	a	a	a*	a*	b*	b*	a*
20	a	a	c	e	a	a	b	b*	b*	b*	a*
21	b*	b*	c	e	a	a	b	b*	b*	b*	a*
22	a	a	c	e	a	a	a*	b*	b*	b*	a*
23	b*	b*	c	e	a	a	a*	b*	b*	b*	a*
24a	a	a	c	a*	a	a	b	d	b*	b*	a*
24b	a	a	c	b*	a	a	b	d	b*	b*	a*
24c	a	a	c	c*	a	a	b	d	b*	b*	a*
24d	a	a	c	d*	a	a	b	d	b*	b*	a*
25	a	a	a*	e	a	a	b	c*	a	a	d

[a]The letters in this table correspond to the labeling of approaches in Table 1-1.

[b]Analyses 11c and 11d differ in that the single species considered in 11c is rats and for 11d it is mice.

*An asterisk marks those approaches that differ from those in Analysis 0, the first standard.

Table 1-3

STANDARD VALUES USED IN ANALYSIS OF ANIMAL BIOASSAY DATA

Animal	Surface Area Coefficient[a]	Body Weight (kg)	Food Consumption (mg/day)	Inhalation Rate (m^3/day)	Drinking Water Rate (mg/day)	Experiment Length (weeks)
Dog	10.1	12.7	508000	1.5	350000	312
Guinea pig	9.5	0.43	12900	0.074	145000	104
Hamster	9.0	0.12	9600	0.037	30000	104
Monkey	11.8	3.5	140000	1.4	450000	364
Mouse	9.0	0.03	3900	0.05	6000	104
Rabbit	10.0	1.15	33900	1.6	300000	156
Rat	9.0	0.35	17500	0.26	35000	104

[a]Surface area in m^2 is calculated as $KW^{2/3}/100$ where W is weight in kilograms and K is the surface area coefficient (2).

Table 1-4

APPROACHES USED FOR SUPPLEMENTAL ANALYSES

Analysis	Component										
	1	2	3	4	5	6	7	8	9	10	11
30	a[a]	a	b	a	a	a	b	d	a	a	d
31	a	a	b	e*	a	a	b	d	a	a	d
32	a	a	b	b*	a	a	b	d	a	a	d
33	a	a	b	c*	a	a	b	d	a	a	d
34	a	a	b	d*	a	a	b	d	a	a	d
35	b*	a	b	a	a	a	b	d	a	a	d
36	a	b*	b	a	a	a	b	d	a	a	d
37	a	a	a*	a	a	a	b	d	a	a	d
38	a	a	c*	a	a	a	b	d	a	a	d
39	a	a	b	a	b*	a	b	d	a	a	d
40	a	a	b	a	a	b*	b	d	a	a	d
41	a	a	b	a	a	a	a*	d	a	a	d
42	a	a	b	a	a	a	b	a*	a	a	d
43	a	a	b	a	a	a	b	b*	a	a	d
44	a	a	b	a	a	a	b	c*	a	a	d
45	a	a	b	a	a	a	b	d	b*	a	d
46	a	a	b	c	a	a	b	d	a	b*	d
47	a	a	b	a	a	a	b	d	a	a	a*
48	a	a	b	a	a	a	b	d	a	a	b*
49	a	a	b	a	a	a	b	d	a	a	c*[b]
50	a	a	b	a	a	a	b	d	a	a	c*[b]

[a]The letters in this table correspond to the labeling of approaches in Table 1-1.

[b]Analyses 49 and 50 differ in that the single species considered in 49 is rats and in 50 it is mice.

*An asterisk marks those approaches that differ from those in Analysis 0, the first standard.

Table 1-5

DESCRIPTIONS OF ALL ANALYSES

Analysis	Template[a]	Differences[b]
0	Initial Standard	$mg/m^2/day$, no averaging of results; oral, gavage, inhalation or route like humans
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 anomals 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 1-5 (continued)

DESCRIPTIONS OF ALL 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
30	Alternative standard	mg/kg/day; no averaging; any 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 most closely resembles and on which is based the analysis in question. Analyses 0 and 30 are the two standards; they have no template but rather are the bases for defining the other analyses.

[b]The differences listed are the ways in which the analysis in question differs from its template. For Analyses 0 and 30, for which there are no templates, no "differences" are defined. In these two cases the approaches for several prominent components are listed.

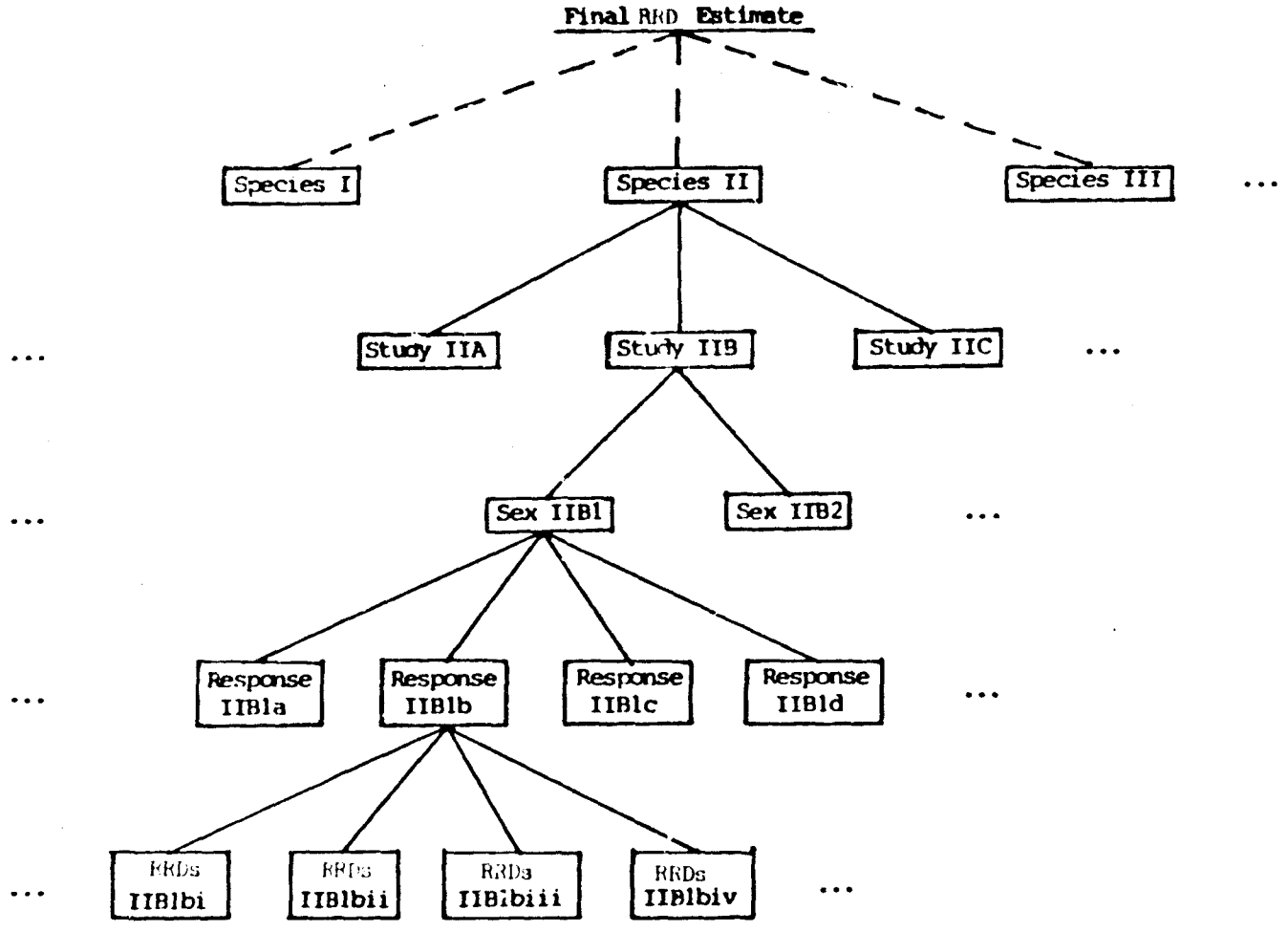
Table 1-6

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 1-3 lists the standard experiment lengths.

Figure 1-1


Final RRD Estimate
Species I
Species II
Species III
...
Study IIA
Study IIB
Study IIC
...
Sex IIB1
Sex IIB2
...
Response IIB1a
Response IIB1b
Response IIB1c
Response IIB1d
...
RRDs IIB1bi
RRDs IIB1bii
RRDs IIB1biii
RRDs IIB1biv
...

Section 2

RESULTS

This section describes the results of the evaluation of the animal bioassay data and of its comparison with the epidemiologically derived risk estimates. The evaluation is logically divided into two steps. The first is a correlation analysis, the goal of which is to determine whether or not the estimates of risk-related doses obtained from analysis of the bioassay data (the animal estimates) are correlated with the estimates obtained from epidemiology (the human estimates). If no correlation is found, then it may not be appropriate to attempt to estimate human risk from animal data. If, on the other hand, a positive correlation does exist, then it seems reasonable to assume that the animal models are relevant to human risk estimation and to proceed to the second step, that of identification of useful predictors. The goal of that process is to determine which particular point estimates calculated from the bioassay data can be satisfactorily employed as predictors of the human RRDs, and to evaluate the variability (the remaining uncertainty) associated with the identified predictors.

The correlation analysis reveals that there is, indeed, a significant positive correlation between the human estimates and most of the animal estimates. Those analysis methods that demonstrate the best correlations provide viable alternatives for the choice of the predictors. The detailed results of the correlation and prediction analyses are described below.

CORRELATION ANALYSIS

Table 2-1 presents the correlation coefficient estimates (and their associated p-values) corresponding to each method of analyzing the bioassay data. The four columns of that table represent the four data sieves we have defined. Graphs of animal RRDs vs. human RRDs for many of the analysis methods are contained in Figures 2-1 to 2-34. (Abbreviations for all the chemicals considered in this project are

given in Table 2-2).

Generally speaking, the results in Table 2-1 show a strongly positive relationship between animal and human RRDs. The number of analyses resulting in correlation coefficients greater than 0.6 ranges from 26 to 29 out of 38, depending on the sieve used. When the full sieve is used, 26 analyses yield results with $\rho > 0.7$. In no instance did a negative value for ρ obtain. Out of the 38 p-values associated with the analyses employing the full sieve, 28 were less than 0.01, and 35 less than 0.05.

Given these results, it is inconceivable that these correlations are due to chance. It is also highly unlikely that they are due to bias in the methods employed. The coding of the animal data into the computerized data base was made by individuals who were unaware of the results for the RRD estimates for the human data. The calculations for the animal RRDs were made using uniform approaches implemented by an impartial computer program. Although the calculations of the human RRDs were made individually and required judgements on the part of the analyst, they also were made blindly without knowledge of animal RRDs for any of the chemicals.

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 both possible and scientifically appropriate to estimate human risk from animal data. The range of finite, best RRD estimates from human data represented by these 23 chemicals spans roughly six orders of magnitude (from $10^{-3.5}$ mg/kg/day for melphalan to $10^{2.6}$ mg/kg/day for saccharin). Human and animal RRD estimates are fairly consistent over this range considering the crudeness of much of the underlying data (see, for example, Figure 2-12).

These analyses are considered in greater detail below. Individual analysis methods will be studied; methods that yield the best correlations are identified and discussed. Similarly, methods that yield the poorest correlations will be discussed. We begin with an evaluation of the sieve.

Evaluation of Sieve

The purpose of the sieve is to select only the better data for analysis whenever data of varying quality are available, while at the same time not excluding any chemicals from analysis on the basis of the sieve. The sieve consists of two parts: a quality screen that discriminates among data sets on the basis of number of animals tested and length of observation, and a significance screen that selects only data sets in which a statistically significant response was found, whenever such data sets are available for a chemical (cf. Section 1). The idea behind the sieve is that use of better data should improve the observed correlations between the human and animal results. This desired result is in fact achieved, since ρ is higher when the full sieve is used in 28 out of the 38 analyses (Table 2-1). The effect of the sieve can be observed by comparing, for example, the graphs in Figures 2-1 through 2-4 or in Figures 2-9 through 2-12. The data appear to be more closely grouped about the best-fitting line when the screenings are applied (especially when the full sieve is applied, Figures 2-4 and 2-12) than when they are not applied (Figures 2-1 and 2-9). This improvement in correlations when better data are used is further evidence that the observed correlations between the animal and human RRDs are real.

Aside from Analyses 3b, 8a and 11a, almost all of the benefit obtained from applying a sieve is seen when the significance screen is applied. That screen limits attention to the carcinogenic responses that are significantly dose-related when such responses are available. The quality screen, which focuses on the number of animals tested and the length of the experiment, does not appear to provide much of an improvement over and above the significance screen for most analysis methods (compare the third and fourth columns of Table 2-1). Such a result is expected for analyses, like number 1, that already restrict attention to a subset of the experiments, in this case the "long" experiments. It is somewhat surprising in other cases, especially since the majority of the carcinogenic responses in the data base are there because of their apparent dose-related action. It is possible that the criterion limiting attention to studies utilizing gavage, inhalation, or oral routes of exposure (or the route most similar to human exposure), a

criterion underlying every analysis except 3b, also had the side effect of eliminating many of the studies that would receive lower ranks in the quality screening. This is conceivable if, for example, other methods of dosing involved fewer animals (possibly because these routes are more difficult to administer) or if other routes tend to involve bolus doses that might cause early deaths (carcinogenic or non-carcinogenic) and consequently shorten the duration of observation. It is possible that experiments employing "nonstandard" routes of dosing were designed to investigate special questions and so may not have been overly concerned with number of animals or length of observation.

Also in relation to the action of the sieves, those analyses that average RRD values at each stage (over sex within study, over study within species, and finally across species; Analyses 12 through 24d) are relatively impervious to the application of any screenings. The correlation coefficients within any of the rows in Table 2-1 corresponding to those analyses are very similar, no matter which sieve is applied (cf. Figures 2-28 through 2-31). It seems likely that the averaging that occurs in these analyses acts in much the same way as the screens are intended to work; much as a sieve acts to eliminate outliers, so averaging works to pull outliers toward the "middle" of the results. This effect is enhanced by the use of harmonic averaging which severely limits the influence of infinite values. Since RRDs are bounded below but not above, infinite-valued estimates are obvious candidates for outliers. Similarly, the action of both the quality screen and the significance screen would tend to eliminate infinite-valued estimates since experiments that are too short or employ too few animals would tend to find no carcinogenicity of a chemical (*i.e.*, give infinite RRD estimates) and responses not significantly related to dose generally also produce infinite-valued RRDs.

Analyses other than 12 through 24d employ averaging, but not at every level. Analysis 9 averages only across sex within study; Analysis 10 only across study within species; Analysis 11a only across species; and Analysis 11b only across the species rats and mice. Since the experiments employing species other than rats and mice do not appear to be as "clean" as those using rats and mice (compare the first columns in the rows corresponding to Analyses 11a and 11b and note the sizable

increase in ρ when a quality screen is applied to Analysis 11a), let us concentrate on Analyses 9, 10, and 11b (cf. Figures 2-18 through 2-25). In those cases, one notes a similar but slightly lessened independence from the sieve. Especially when averaging across sex or across study, an effect similar to that of the screenings may already be in operation.

In this connection note that Analysis 12, which averages at all levels and that uses the same data as Analysis 0, results in larger correlation coefficients than does Analysis 0 when the significance screen is not used (columns 1 and 2 of Table 2-1). Correlation coefficients associated with Analysis 12 are somewhat smaller than those associated with Analysis 0 when the significance screen is employed (third and fourth columns) and, moreover, the application of the significance screen to Analysis 0 produces larger coefficients than the nonscreened Analysis 12. This suggests that application of an appropriate sieve may be a better approach than merely averaging at all levels. Since use of a sieve appears to improve most analyses, and since use of the full sieve is about as good or better for most analyses than use of either screen by itself, the remaining discussion will emphasize analyses that employ the full sieve.

Analyses that Use Combination of All Significant Individual Responses

Two of the endpoints defined and included in the bioassay data base whenever possible are the combination of all individual carcinogenic responses that are significantly dose-related and the combination of all such responses that are malignant. Analyses that use the first of these, combination of significant responses, i.e. Analyses 8a, 16, and 17, provide relatively poor correlations ($p < 0.6$) no matter which data screening procedure is implemented (cf. Figures 2-14 and 2-32). The p-values associated with these correlations range between 0.02 and 0.05 which, given the number of comparisons performed, might reasonably be considered only marginally significant. The response, combination of significant responses, could not be defined for every study or every chemical; only 13 of the 23 chemicals had one or more experiments presenting data that allow calculation of this endpoint. It is the case that the experiments that do provide the necessary information in this

regard are generally more complete and better studies, notably the NTP bioassays. This is indicated by the fact that rank 1 studies (those observing over 50 dosed animals for at least 75% of the standard observation period) are available for 12 of the 13 chemicals included in Analyses 8a and 16. That being the case, it is less likely that the relatively poor correlation is due to use of data of poorer quality.

Interestingly, the analyses using the combination of malignant statistically dose-related responses, Analyses 18 and 19, provide very good (and in some cases, the largest) correlation coefficients, ranging from 0.73 to 0.79. However, the difficulty of defining the response is even more severe with this endpoint than with the previous one. No more than 10 chemicals included studies with the necessary information. Note that the p-values associated with these chemicals range between 0.003 and 0.009; such p-values are associated with ρ's on the order of 0.61 when more chemicals are included (cf. Analysis 2, no screens). So, while use of this endpoint may well be appropriate, more data would have to be made available before any stronger conclusion would be warranted.

Analyses That Utilize Malignant Neoplasms Only

Analysis 7 is identical to Analysis 0 except that the former analysis utilizes animal data on malignant neoplasms only, where the latter analysis permits data on benign neoplasms to be used as well; Analyses 12 and 14 have a similar relationship. Analyses 7 and 14, utilizing data on malignant neoplasms, yield results that are quite similar to those obtained from Analyses 0 and 12, respectively, analyses that used data on both benign and malignant neoplasms. The graphs for Analyses 0 and 7 (Figures 2-4 and 2-14) are very similar, the major difference being that data for benzidine are utilized in Analysis 0 but not Analysis 7. It is important to note that inclusion of both benign and malignant tumors does not degrade the correlations (in fact, it improves them somewhat) despite the fact that the human results are for malignant tumors exclusively.

epidemiologically derived estimates even when no screening is used. This may reflect an underlying difference in the overall quality of rat and mouse experiments.

One of the few analyses that derives some benefit from the quality screening, over and above that obtained by significance screening, is 11a. The same is not true for Analysis 11b. This suggests that the improvement obtained by screening the quality of the data in Analysis 11a is derived primarily from elimination of experiments in species other than rats or mice that were too short or that tested too few animals. Indeed, the correlation coefficients for Analyses 11a and 11b are nearly identical when the quality screen is applied (columns 2 and 4 of Table 2-1; cf. Figures 2-23 and 2-25). With that screening, either of these two analyses compares favorably with the standard analysis (Analysis 0) and are similar to the results obtained from rats alone or mice alone. This is perhaps not surprising given the preponderance of rat and mouse experiments in the data base and the previously noted similarity of rat-alone and mouse-alone correlation coefficients when the data is appropriately screened.

Choice of Dose Units

Analyses 0, 4a, 4b, 4c, and 4d show correlations obtained when five different dose units are used for extrapolating risk from animals to humans [namely, $mg/m^2/day$ (Analysis 0), mg/kg/day (Analysis 4a), ppm in diet (Analysis 4b), ppm in air (Analysis 4c), mg/kg/lifetime (Analysis 4d)]. The choice of dose units makes little difference in the correlations. A similar phenomenon is observed for Analyses 12, 24a, 24b, 24c, and 24d. These analyses average over sex, study, and species and, again, differ only in the choice of dose units used in animal-to-human extrapolation. The similarity of the correlation coefficients is expected because few data sets included study-specific data (primarily body weights and food consumption variables) that affect the calculation of the dose measure. Chemicals for which such data are not available would tend to maintain their same relative positions in all of these analyses and so no change of ranks would be apparent. The merits of the various dose units are considered further in the evaluation of the

prediction analyses.

Identification of Analyses Yielding Higher Correlations

Analysis 3b yields the highest correlation; when the full sieve is applied, $\rho = 0.90$ (cf. Figure 2-12). This analysis method also yields correlation coefficients that are among the best when less than the full sieve is used. Interestingly, Analysis 3b is the least restrictive of the methods examined. Whereas all other analyses are restricted to experiments that expose animals by gavage, inhalation, oral, or the route of exposure that humans encounter, 3b also allows instillation, injection, and implantation experiments. These additional routes are often not considered in quantitative risk assessment.

This discussion of Analysis 3b provides an opportunity to consider the effect that changes in the data have on the correlation coefficients. Some changes in data are the result of changing the criteria used to pick experiments and carcinogenic responses for particular analysis methods. In this case, allowing all routes of exposure adds three chemicals that only have studies that expose animals by "nonstandard" means, chlorambucil, chromium, and melphalan. Moreover, RRD estimates for certain other chemicals change dramatically when all routes are allowed. Arsenic is a prime example; note the change in location of the animal lower bound for this chemical (compare Figure 2-4 to Figure 2-12). The animal lower bound RRD for arsenic in Analysis 3b is derived from an experiment in which exposure was accomplished via intratracheal instillation and for which a dose-related increase in lung tumors was found. Note that the animal upper limit RRD is infinite whether or not all routes of exposure are included in the analysis, a fact consistent with the commonly held view that arsenic has not been shown conclusively to be carcinogenic in animals.

The correlation coefficient, ρ, is derived from the ranks determined by the relative positions of the intervals. In Analysis 0, the human rank of arsenic is 6 and its animal rank is 15, a major discrepancy. In Analysis 3b, with the addition of the three chemicals, arsenic's human rank increases to 9 but its animal rank, due to the reduction in the lower bound discussed above, rises to only 16.5. So, while the values

of the RRDs can and do change substantially, the ranks based on the RRDs may be relatively insensitive to those changes.

Nevertheless, ranks can be greatly altered. Comparing the same two Analyses, 0 and 3b, one notes a great change in the lower and upper bounds associated with estrogen. This results in a change of rank, from 11.5 in Analysis 0 to 6 in Analysis 3b, despite the addition of three more chemicals in the latter analysis. Estrogen has only two bioassays, one an implantation study and the other a feeding study. The feeding study which was the only animal estrogen study utilized in Analysis 0, failed to find any significantly dose-related responses (hence the infinite MLE for estrogen in Analysis 0); however, the implantation study entered Analysis 3b and included an increased incidence of kidney tumors. So, while we are interested in changes in the underlying base of data and their effect on risk estimates, one must be aware that other changes may be confounded with the changes in which we are interested. Thus, with estrogen, including other routes (the implantation study) actually eliminates the feeding study (because of the action of the sieve), an unforeseen change in the underlying data. Another manifestation of this is the fact that certain chemicals are eliminated from some analyses because they lack the data to support those methods.

One can attempt to minimize such confounding data dependency. It is of interest, for example, whether the high correlation obtained in Analysis 3b is due to the addition of the three chemicals mentioned earlier, chlorambucil, chromium, and melphalan. The human and animal ranks of chromium and melphalan are well matched, but those for chlorambucil are 7.5 and 3, respectively, showing moderate discrepancy. If these three chemicals are not included in Analysis 3b, so that the only differences between Analyses 0 and 3b are due to inclusion of additional routes of exposure, then $\rho = 0.88$ when the full sieve is applied. This is very close to the original ρ, 0.90, and is still notably better than the correlation obtained from any other analysis.

Nevertheless, it is not possible to conclude unequivocally from these analyses that the improved correlation due to inclusion of additional routes of exposure will hold generally and is not simply a feature of the particular data available for analysis. A substantial part of the

improved correlation is due to a data-dependent change in one chemical (estrogen). This may, however, be an indication that inclusion of additional 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.

Aside from Analysis 3b, no other analysis stands out as being superior to the others based on the correlation analysis. The largest of the remaining correlation coefficients is 0.81 (Analysis 25, full sieve) and another 16 of the 38 analyses performed with the full sieve yield ρ's between 0.76 and 0.81. These and perhaps many of the other analyses provide ample indication that animal-based estimates of RRDs are applicable to estimation of human RRDs and can be considered viable alternative procedures for use in human risk assessment.

PREDICTION ANALYSIS

The three loss functions described in the discussion of methodology (Section 1 of this volume) have been used to determine the lines of best fit for 304 sets of predictions (the four predictors, L_M, L_{2Q}, MLE_M, and MLE_{2Q}, obtained from the initial 38 analyses performed with and without the sieve; see Tables 2-3, 2-4, and 2-5). In the case of the DISTANCE^2 loss function (Table 2-3) there are only 152 (= 304/2) analyses. This is because this loss cannot be applied to the predictors, MLE_M and MLE_{2Q}, that can assume infinite values.

The prediction investigation has been limited to the analyses performed without any data screening or with the full sieve, i.e. both the significance and quality screens. The other sieves examined in the correlation analyses, those with either the significance screen or the quality screen alone, are not employed here. The quality screen alone did not appear to provide substantial improvement over no sieve in the correlation analysis. The significance screen accounted for the majority of the benefit obtained by screening the data, but adding the quality screen did not undercut the benefit seen and, for certain analyses (including one of the most promising ones, Analysis 3b)

provided notable additional improvement in the correlation. Addition of the quality screen ensures that possibly questionable experiments of short duration or employing few test animals do not compromise the results obtained from the lower ranked studies (the "better" studies). This may be particularly important when extreme values, such as the minimum lower bound or minimum MLE, are used as predictors.

Sieve

Examination of Tables 2-3 through 2-5 reveals that use of the sieve generally does improve the predictive ability of the bioassay results. No matter which loss function is used, the predictive power of those analyses employing the minimum values (L_M or MLE_M) is improved (in the sense of yielding smaller average loss) 74% to 82% of the time when the sieve is applied. This effect is seen slightly more often among the analyses that do not average over sex, study, and species (Analyses 0 through 11d and 25) than among those that do average over sex, study, and species (Analyses 12 through 24d). In the first set, improvement with application of the sieve is noted 77% to 91% of the time, depending on the loss function chosen. Improvement occurs in 56% to 75% of the analyses in the second set. As was noted in the correlation analysis, the effect of averaging in some sense mimics the desired effect of a sieve by eliminating outliers.

One would expect the effect of the sieve to be less pronounced when median values are used as predictors. While this is true for the median lower bound predictor, L_{2Q} (the sieve improves no more than 50% of the analyses), the sieve reduces loss using MLE_{2Q} in almost 80% of the analyses. The benefit obtained by using medians is offset, in the case of the maximum likelihood estimates, by at least two factors. First, it is known that the maximum likelihood estimates from the multistage model are much more sensitive to small changes in the data than are the lower bound estimates. Such sensitivity might entail more outliers in those cases in which few animals were tested or the length of observation was short. Moreover, for those same types of experiments (few animals or short duration) and for those carcinogenic responses that are not significantly related to dose, infinite RRD estimates are possible.

When the epidemiologically derived RRD estimates are finite, the loss is exacerbated. The sieve eliminates such data sets from consideration when better ones are available.

The results for the median lower bound predictor, L_{2Q}, are apparently the most stable, as reflected in the fact that the action of the sieve is neutral in a general sense. Note, however, that certain analyses demonstrate definite improvement when the sieve is applied, even with L_{2Q} as the predictor.

Predictors

The four predictors that have been selected, L_M, L_{2Q}, MLE_M, MLE_{2Q}, can be compared on the basis of the average loss suffered when each predictor is used in any particular analysis. Tables 2-3 through 2-5 clearly indicate that, no matter which loss function is employed, L_{2Q} is the best predictor to use. In only 2 to 6 (depending on the loss function) of the 76 analyses (38 pairs, with and without the sieve) does that predictor fail to yield smaller loss than does L_M. Analysis 2, for example, appears to yield better prediction when L_M is used instead of L_{2Q}. Comparison of the loss values among the analyses, however, reveals that Analysis 2 is not one of the better methods of calculating RRD estimates, so this observation has little bearing on the noted superiority of L_{2Q}. In three to five instances (again depending on the choice of loss function) L_{2Q} does not provide loss smaller than results from using MLE_{2Q}. Analyses 8 and 18 account for most of these exceptions. Recall that these analyses could be performed on only six and ten chemicals, respectively, so again, the fact that MLE_{2Q} yields the smaller loss should not be given too much weight.

The superiority of L_{2Q} is independent of the choice of loss function. Also independent of loss function is the fact that MLE_{2Q} is superior to MLE_M among those analyses employing a sieve; the MLE_{2Q} losses are smaller in 17 or 18 of the 22 such analyses for which MLE_M and MLE_{2Q} differ. (Note that the analyses that average over sex, study, and species, Analyses 12 through 24d, provide a single lower bound estimate and a single maximum likelihood estimate. Consequently, $L_M = L_{2Q}$ and

MLE_M = MLE_{2Q} for those analyses and the losses associated with use of L_M are identical to those associated with use of L_{2Q} and similarly for MLE_M and MLE_{2Q}.) Included in the set of 17 or 18 analyses are those that yield the smallest losses when a maximum likelihood predictor is used so the superiority of MLE_{2Q} over MLE_M is clear when the sieve is applied.

The case is not so obvious when no sieve is applied. In this instance, the results do depend on the choice of loss function. When fit is measured by the CAUCHY loss function (Table 2-4) MLE_{2Q} is superior to MLE_M in 15 of the 22 no-sieve analyses. When the TANH loss function is used (Table 2-5), MLE_{2Q} is superior in only 10 of 22 no-sieve analyses. However, with both loss functions, MLE_{2Q} is better than MLE_M for most of those analyses that yield the smallest losses (Analyses 2 and 11c being the exceptions). Thus, unless one wishes to analyze bioassays by method 2 (using only those experiments that dosed treated animals for at least 80% of the standard experiment length) or by 11c (use rat experiments only) without a sieve, one must conclude that in the case of the maximum likelihood estimates, as well as in the case of the lower bounds, a median predictor is a better choice than a minimum predictor. Interestingly, Analysis 2 (though not 11c) is also the method consistently yielding smaller losses with L_M than with L_{2Q}.

One final observation will conclude this comparison of the predictors. Since L_{2Q} is better than L_M and MLE_{2Q} is better than MLE_M, it is of interest to determine whether L_M is better than MLE_{2Q}. The answer depends to some extent on the choice of the two loss functions that can be used to compare these predictors, CAUCHY and TANH (Tables 2-4 and 2-5, respectively). Among the 22 pairs of analysis methods that do not average over sex, study, and species the CAUCHY loss function indicates that L_M produces smaller average loss than does MLE_{2Q} in 21 cases, i.e. less than half the time. Among those same analyses, L_M outperforms MLE_{2Q} for 40 analyses when measured by TANH loss. Both loss functions indicate that MLE_{2Q} is superior for Analyses 3b with and without the sieve (analyses providing some of the best correlations in the correlation analysis) and for Analysis 6 with or without the sieve (the method applicable to only six chemicals). For those 32 analyses that do average over sex, study, and species (Analyses 12 through 24d, with and without the sieve), L_M is better in every case but two (Analysis 18 with

and without the sieve) no matter which loss function is used. Thus, L_M outperforms MLE_{2Q} in the majority of cases, especially when assessed by the TANH loss function, but any conclusion about the superiority of L_M may depend strongly on the analysis methods that are of interest.

Comparison of Analysis Methods

The comparison of the analyses and the identification of the best ones are complicated because four separate predictors have been used and three different loss functions have been defined. If the different predictors or different loss functions result in distinct orderings of the methods, interpretation is more difficult. Table 2-6 presents the five best analyses (those giving the smallest average losses) by predictor and loss function.

Analyses 6, 18, and 19 dominate the list of methods giving smallest losses. This is true no matter which predictor or which loss function is used. (Analysis 18 does not appear in any TANH list, however.) Recall that these are the analyses cited in the discussion of correlation analysis results as those that yield relatively large correlation coefficients but that are applicable to few chemicals (six, ten, and nine chemicals for analyses 6, 18, and 19, respectively). So, as with the correlation results, the prediction results are suggestive for these analysis methods, but no firm conclusions are warranted.

Table 2-7 lists the analyses that yield the smallest average losses after eliminating Analyses 6, 18, and 19, which are based on relatively few chemicals. Analysis 17 appears on the list frequently. That method uses the response that is the combination of significant individual responses and is limited to experiments that dosed and observed the test animals for a suitably long period. Furthermore, RRD estimates are averaged over sex, study, and species. That this method should appear to provide good fit to the data is somewhat surprising since the correlation coefficients associated with it are on the order of 0.58, not among the better correlation results. Once again, however, the number of chemicals that can supply data meeting the requirements of this approach is limited; only 11 chemicals had studies that dosed and

observed animals long enough and for which the combination of significant responses could be defined.

Analysis 16 is similar to Analysis 17 in that the endpoint chosen is the combination of significant responses and the estimates are averaged over sex, study and species. It does not, however, exclude experiments on the basis of their length of observation and dosing, so that thirteen chemicals can be analyzed by this method. Analysis 16 also appears in the list given in Table 2-7, predominantly when the CAUCHY loss function is used. The analysis that does not average over sex, study, and species but that does use the combination of significant responses (Analysis 8a) does not yield average losses that are among the smallest, for any loss function or predictor. Thus, it appears that averaging at each level may be the most useful method when the combination of all significant responses is the endpoint used.

Analysis 20 is identified as a method yielding small losses, but only when a lower bound predictor is used. That method is best when loss is measured by the TANH function. This analysis method selects as the endpoint of interest total tumor-bearing animals and averages estimates over sex, study, and species. The corresponding unaveraged method (Analysis 8b) does yield losses not much larger than those associated with Analysis 20, 0.127 vs. 0.121 and 0.125 vs 0.121 for L_M and L_{2Q}, respectively (measured by TANH). Consequently, use of total tumor-bearing animals in conjunction with a lower bound estimate appears to be an appropriate technique, if TANH is a suitable measure of loss.

Note that, of the twenty analyses listed as providing the smallest average losses determined by CAUCHY and TANH (the two functions that consider the best epidemiological estimates of RRD) for the lower bound predictors (L_M and L_{2Q}), all but four use an endpoint that is a combination of individual responses, either total tumor-bearing animals or the combination of significant responses. Due to limitations in the data available for analysis, not inherent limitations of the methods themselves, some of these analyses were applicable to relatively few chemicals. Nevertheless, the consistency with which these endpoints yield small average loss indicates that they should be considered viable candidates for estimation of human risk.

The DISTANCE^2 loss function identifies a set of good methods that intersects with the sets identified by CAUCHY and TANH infrequently. All but one of the ten analyses listed in Table 2-7 under DISTANCE^2 use individual carcinogenic responses rather than a combined response. In two cases (Analysis 25, the best when L_M is the predictor, and Analysis 8c, also associated with the L_M predictor) the response is limited to those that are associated with human exposure. This response may be identifiable when human data exist, but when such data are absent, as would be the case for a new chemical, then the appropriate choice of endpoint is unknown and application of these methods problematical.

The DISTANCE^2 loss function identified Analysis 3b with the sieve as the best method (in terms of average loss) when L_{2Q} is the predictor. This analysis was also clearly superior in the correlation analysis ($\rho = 0.90$). It is not surprising that DISTANCE^2 would tend to match the results of the correlation analysis, especially when L_{2Q} is the predictor. First, L_{2Q} was one end of the interval of animal RRDs used in the correlation analysis. Second, DISTANCE^2 does not consider the location of the best estimate of RRD derived from the epidemiological data and so is concerned only with the position of the human interval. In any case, Analysis 3b yields the smallest loss with the DISTANCE^2 function and reasonably small losses with CAUCHY and TANH, 0.413 and 0.140, respectively. (Note: it is not appropriate to compare the loss values obtained using different loss functions. The fact that different formulations of loss are used entails that the values in the different columns of Table 2-7, for example, are not comparable.)

Since L_{2Q} is the predictor that produces the best fit of the animal results to the human results (a fact that is reinforced by examination of Table 2-7), we concentrate on those analyses that perform best with that predictor. Table 2-7 shows that Analyses 3b, 17, and 20 are the analyses yielding the smallest average losses for one of the three loss functions. One would like to have results that are independent of the choice of loss function. That is, a good analysis method should be robust with respect to differences in loss functions. To investigate the analyses in this manner, we have defined what is called "total

incremental normalized loss" as follows. For each loss function, the difference between the smallest average loss and the largest average loss among the analyses (still ignoring Analyses 6, 18, and 19) when L_{20} is used is known. For each analysis the difference between the average loss for that analysis and the minimum average loss, divided by the difference between the minimum and maximum average losses, is defined as the "incremental normalized loss". The sum of these across all three loss functions gives the total incremental normalized loss (Table 2-8). Normalization eliminates the difference in scale of the three loss functions and should allow an overall appraisal of the analyses.

Table 2-8 reveals that Analysis 17 (with the sieve) obviously adds least to the average loss incurred. Analysis 17 without the sieve is nearly as good. Analyses 3b with the sieve and 20 without the sieve, the other two analyses picked as best by one of the three loss functions, yield total incremental losses that are about the same, 0.555 and 0.558, respectively, and follow the pair of Analysis 17 results, as the next best methods of analysis.

Figures 2-35 through 2-38 display the plots of those four analysis methods. One thing that is clear from these figures is that Analysis 17 derives much of its good performance from the specific subset of chemicals to which it can be applied. For only three of those eleven chemicals does the best fitting line fail to pass through the interval of human RRD estimates, for any loss function. However, even when Analyses 3b with the sieve and 20 without the sieve are limited to the same eleven chemicals (Table 2-9), Analysis 17 with the sieve is better when measured by CAUCHY and TANH. On the other hand, Analysis 3b with the sieve, restricted to the seventeen chemicals to which Analysis 20 can be applied, yields smaller losses than does Analysis 20 as measured by all three loss function.

Other analyses that yield relatively good, robust results can be identified from Table 2-8. Those for which the total incremental normalized losses are less than 1.0, for example, for at least one member of the pair of results (with or without the sieve) include Analyses 4a through 4d (analyses that differ from the standard only with respect to the dose units used to extrapolate from animals to humans);

8b, utilizing total tumor-bearing animals as the endpoint (as does Analysis 20 discussed above); 8c, which is limited to carcinogenic responses that humans got; 9, which averages results over sex within a study; 11b, a method that averages the results from rats and mice; and 11c which uses rat data only. The best total incremental normalized losses among these analyses range from 0.698 to 0.963.

Note, in passing, that an alternative ranking procedure that can be applied is a minimax scheme. That is, for any analysis, the maximal loss over the three loss function can be determined. The analyses that have the smallest maximums are best in a minimax framework. Since the loss functions have different scales, this approach should also be based on the incremental normalized losses. If this is done, then Analyses 17, 3b, and 20 remain the best three, in order, and several of the others just cited, notably 8c and 8b, remain in the list of good analyses. Analysis 22 also satisfies the minimax criterion well.

Asymmetric Loss

The discussion up to this point has concerned observations relating to loss that is symmetric. That is, the loss functions employed reflected the assumption that it is no worse to overestimate RRDs by a given amount than to underestimate them by the same amount. In fact, it is reasonable to think otherwise, *i.e.* to think and base our decisions on the premise that overestimation is worse than underestimation. The health considerations involved in cancer risk assessment make this a prudent approach.

The effect of incorporating asymmetry into the loss calculations is investigated in the following manner. The TANH loss function has been used to fit a line to the results of each analysis method. In the definition of TANH is a factor, m, called the asymmetry constant, that reflects the degree of asymmetry thought to be pertinent. The symmetric version has m = 1. The fitting is performed now with m equal to 1.5, 2, 5, 10, 50, and 100. Larger values of m reflect stronger beliefs about the inadvisability of overestimating RRDs. Tables 2-10 and 2-11 display the results for the lower bound predictors.

Note that the losses incurred when m = 50 and m = 100 are identical (to three decimal places) for every analysis. Any high degree of asymmetry drives the line toward the chemicals in the lower right-hand corner of the plots. Figure 2-39 displays this phenomenon for Analysis 3b, the method applicable to the most chemicals.

As with the symmetric version of the TANH loss function (cf. Table 2-5), Analyses 6 and 19 perform well with moderate degrees of asymmetry. Analysis 6 continues to be among the five best for larger degrees of asymmetry while 19 does not. When the minimum lower bound is the predictor, Analyses 20 (for moderate degrees of asymmetry) and Analysis 17 (for all degrees, with the interesting exception of m = 5) perform well, as they did with symmetric loss. When the predictor is L_{20}, Analysis 17 is again good for all degrees of asymmetry, but Analysis 8b, not Analysis 20, moves to the top five for moderate asymmetry constants. For both predictors, Analysis 8c (which uses an endpoint that humans get) produces small losses for $m \geq 5$. Most notable, however, is the fact that Analysis 22 (using total malignancy-bearing animals and averaging over sex, study, and species), a method moderately good with no asymmetry, is second only to Analysis 6 (which is applicable to only six chemicals) for high degrees of asymmetry. This implies that, if it is deemed necessary or desirable not to overestimate any RRD, method 22 is the best analysis to use (once again ignoring the suggestive results of Analysis 6).

It is possible to characterize those analyses that will provide the smallest losses with an asymmetric loss function. If those chemicals that fall below the line fit with the symmetric function are nearly colinear with slope equal to one, then a great reduction in asymmetric loss can be achieved by moving the line to the right (decreasing the y-intercept). Of course, if those chemicals that lie above the symmetrically fit line also have this colinear relationship, then the increase in loss for those chemicals when the line moves to the right (as it always will with the introduction of asymmetry) can be minimized. Hence, those analyses that do not produce outliers falling in the upper left corner or, especially, in the lower right corner of the plots will suffer relatively less loss than analyses that do produce such outliers.

In this regard, compare Analyses 20 and 22 (Figures 2-40 and 2-41). With no asymmetry, Analysis 20 is superior. Any degree of asymmetry greater than 1.5 makes Analysis 22 preferable, since asbestos can be accounted for without having to move the line too far to the right.

An asymmetric approach to the prediction problem can be viewed as a means to find the conversion between animal and human risk estimates that is more apt to underestimate human RRDs but is not overly protective, given a specified degree of belief about the undesirability of overestimating RRDs. At one extreme are those conversions obtained when $m \geq 50$. In these cases the conversion is driven by the chemicals that overestimate the most (cf. asbestos in Figures 2-40 and 2-41). If one believes that it is not 50 times worse to overestimate than to underestimate RRDs, then conversions that are still protective (what has been called "conservative") can be obtained. These correspond to smaller values of m and tend to include more, though not necessarily all, chemicals above the fitted line, the region where bioassays predict larger risks than are obtained from the epidemiology for the given conversion.

Since the question of asymmetric loss is closely linked to degrees of belief about the relative desirability of underestimation and overestimation of RRDs, no further investigation of this issue is undertaken. It should be borne in mind, however, that all of the analyses reported in this document can be undertaken using asymmetric loss. The remainder of the results and the discussion focus solely on symmetric loss.

Animal-to-Human Conversion

In the previous discussion of asymmetry, conversion of animal RRD estimates to human RRD estimates was mentioned. This conversion is based on the best-fitting line that relates the two sets of RRD estimates. Specifically, it depends on the y-intercept, c, that defines the line

$$\text{Log}_{10}(\text{RRD}_H) = \text{Log}_{10}(\text{RRD}_A) + c, \qquad (2\text{-}1)$$

or

$$RRD_H = 10^c \cdot RRD_A \qquad (2\text{-}2)$$

which of course depends on the analysis method and predictor. This conversion is over and above those that are used to equate the units between humans and animals: recall that, for each of the possible choices of units used to extrapolate doses from animals to humans, species- and chemical-specific dose conversions were used to arrive at the units mg/kg/day in humans, the units in which all RRD estimates are expressed.

The conversion that is discussed here is a multiplicative factor that is the empirical result of fitting Eq. 2-1 to the ensemble of bioassay data. The fitted line will rarely pass through a data point. That is, for any given chemical used to fit the line, the conversion determined by Eq. 2-1 rarely describes the exact relationship between RRD_H and RRD_A for that chemical alone. Rather, all the study chemicals together determine c, and this factor may then be applied to estimate RRD_H for any other chemical without direct epidemiological estimates. Tables 2-12 through 2-15 display the y-intercept values for each analysis and each predictor.

It is of some interest to determine the conversion factor suggested by the data that applies to the standard analysis, which is modelled after the Carcinogen Assessment Group's (CAG's) usual procedure. That group uses the minimum lower bound as its predictor. Table 2-12 shows that Analysis 0 with L_M yields y-intercepts between 0.51 and 1.71 when no sieve is applied and between 0.83 and 1.07 when the full sieve is used. The ratios, RRD_H/RRD_A (which we will call conversion factors), with these intercepts range between 3.24 ($= 10^{0.51}$) and 51.7 or 6.71 and 11.7 without or with the sieve, respectively. These figures are uniform in suggesting that CAG's procedure is conservative, in the sense of underestimating RRDs or overestimating risk and so being protective of human health. Given that CAG screens its data to select the best available studies, a process that may act like our sieve, the degree of underestimation is likely to be about an order of magnitude for the

level of risk of interest here.

Since L_{2Q} was found to be the best predictor regardless of loss function, the remainder of the discussion of conversion factors focuses on that predictor (Table 2-13). Over all analyses and loss functions, the ratio, RRD_H/RRD_A, determined by Eq. 2-2 ranges from 0.184 to 151. Among those analyses that are based on extrapolation assuming $mg/m^2/day$ human-and-animal equivalence (which include almost all of the analyses since the standard analysis assumes such equivalence) the ratio ranges from 0.184 (from the CAUCHY loss function applied to Analysis 21 with the sieve) to 74.6 (from the TANH loss function line fit to Analysis 6 with or without the sieve). Since Analysis 6 results are based on only six chemicals, an alternative upper value that is more firmly supported can be obtained from the CAUCHY line fit to Analysis 3a without the sieve, 28.4. On the other hand, if we limit attention to those analyses that appear to yield the smallest average losses with the L_{2Q} predictor and the loss functions for which they are best (cf. Table 2-7) then the range is from 1.29 (Analysis 20 without the sieve, TANH loss function) to 16.7 (Analysis 3b with the sieve, $DISTANCE^2$ loss function).

At this point one can compare and contrast the results of the analyses that are identical except for choice of the dose units assumed to yield animal and human equivalence with respect to carcinogenic response (component 4, cf. Table 1-1). To facilitate this comparison, the supplemental analyses discussed in Section 1 are examined as well. The results for these analyses are presented in Tables 2-16 and 2-17.

It is possible to identify three sets of five analyses each such that the analysis within a set differ only with respect to the dose units assumed to yield equivalence. These sets are (0, 4a, 4b, 4c, 4d), (12, 24a, 24b, 24c, 24d), and (31, 30, 32, 33, 34). In each set the ordering of units is $mg/m^2/day$, mg/kg/day, ppm diet, ppm air, and mg/kg/lifetime. Tables 2-18 and 2-19 present the loss and intercept results for each of these analyses.

When the sieve is applied to these analyses, it is nearly uniformly the case that mg/kg/day are the units yielding smallest loss within each set. The exceptions are in the first set, (0, 4a-4d), with the

DISTANCE^2 loss function and in the second set, (12, 24a-24d), with the DISTANCE^2 and TANH loss functions. In all these instances, mg/kg/lifetime are the units producing the smallest average losses. The units mg/kg/lifetime are linear transformations of mg/kg/day dependent only on the length of experiment, so a weight-based extrapolation appears good when the sieve is used. For Analysis 30 with the sieve, the analysis that yields the smallest loss of any in Table 2-18, the y-intercepts (Table 2-19) indicate the ratio RRD_H/RRD_A is between 1.079 and 2.438, depending on which of the loss functions is used. If attention is restricted to the loss functions that base the fit of the lines on the location of the best epidemiological RRD estimates (*i.e.* CAUCHY and TANH) the range is narrowed to between 1.079 and 1.698. Thus, these calculations indicate that RRDs obtained from Analysis 30, the least restrictive analysis using mg/kg/day, very slightly underestimate human RRDs. This is interesting in light of the fact that Analysis 4a, which is like Analysis 30 in every way except that routes of exposure are limited to inhalation, gavage, oral, and the route that humans encounter, overestimates RRDs on average (note the negative intercepts in Table 2-19). This is an instance of a general phenomenon: no matter what units are used for extrapolation, the analysis that is less restrictive with respect to routes of exposure yields larger y-intercepts than the more restrictive analysis. The effect of including all routes of exposure appears generally to be to decrease the median lower bound. Using the restricted set of exposure routes but averaging results over sex, study, and species has the same effect. Conversion factors (ratios) for all units of extrapolation are given in Table 2-20.

To close out this discussion of conversion factors, it is of some interest to compare conversion from rats to humans and from mice to humans. The comparison can be made using Analyses 11c and 11d (rats alone and mice alone, respectively, restricted routes of exposure, with extrapolation based on mg/m^2/day) and Analyses 49 and 50 (rats alone and mice alone, respectively, any route of exposure, with extrapolation based on mg/kg/day). For the first pair of analyses, the rat bioassay conversion factor ranges from 0.81 to 1.85 with no sieve and from 1.43 to 1.92 with the sieve whereas the mouse bioassay conversion factor may vary between 1.78 and 11.67 without the sieve and between 3.72 and 4.30

with the sieve. These results indicate that, unadjusted, the rat results come closer to the direct epidemiological results. (Average losses are generally smaller with rat data also.) For the supplemental pair, Analyses 49 and 50, rat data fits better only when the sieve is not applied (Table 2-16) but tend to overestimate human RRDs whereas the mouse data underestimate (Table 2-17). When no sieve is applied, the degree of underestimation with mouse data is comparable to the degree of overestimation with rat data. However, when the sieve is applied, the underestimation with mouse data (conversion factors between 1.31 and 1.53) is less extreme than the overestimation with rat data (conversion factors between 0.32 and 0.58). [All conversion factors are based on the CAUCHY and TANH loss function, not DISTANCE^2.]

Uncertainty

It is important to characterize the sources and amount of uncertainty associated with any method of estimating human risks from animal data. As described in Section 1, two approaches are taken to investigate uncertainty. The first, which is referred to as residual uncertainty, is the analog of the residual error aspect of statistical analyses. It applies to each analysis method as a whole and delineates the degree of uncertainty that remains even when the best unit-slope line describes the data. The other uncertainty investigation attempts to say something about the uncertainty associated with each of the components of risk assessment. This investigation is more qualitative, but aids in identification of major sources of uncertainty and in the degree of variation attributable to those sources.

Residual Uncertainty. The DISTANCE^2 loss function is ideally suited to an investigation of residual uncertainty. This function finds the line that minimizes the squared distances to the intervals defining the range of epidemiologically derived RRD estimates. That being the case, the contribution to the total loss of any individual chemical indicates how far that line is from the chemical's interval and thus indicates uncertainty over and above that associated with the epidemiologically derived estimates. In this sense it is called residual uncertainty: it is uncertainty remaining after the epidemiological uncertainty is considered.

For any analysis method, the DISTANCE^2-fitted line determines a predicted dose, RRD_p, for each chemical. If RRD_p for any chemical lies between the upper and lower bounds of the epidemiologically derived estimates, $RRD_{H,U}$ and $RRD_{H,L}$, respectively, then no residual uncertainty exists for that chemical. Otherwise, residual uncertainty remains. The residual uncertainties are aggregated in two ways so as to indicate something about the uncertainty in terms of multiplicative factors that may be applied to the predictions to give a range of estimates about the predicted values which are consistent with the data (cf. the description of the methods in Section 1 of this volume). Of course, larger factors (wider ranges) indicate greater residual uncertainty.

When all the chemicals included in any analysis method, even those with no residual uncertainty, are used to characterize uncertainty, a single factor is estimated. This factor is the average amount by which the predicted RRDs must be multiplied or divided so as to eliminate residual uncertainty. Alternatively, two sets of chemicals, those whose epidemiological estimates lie completely above the line of predicted values and those whose epidemiologically derived estimates lie completely below that line, can be used separately to determine two multiplicative factors, one to accommodate underprediction and one to accommodate overprediction. Tables 2-21 and 2-22 present these factors for all analyses (including the supplemental analyses) using the L_{2Q} predictor.

Analyses 6, 18, and 19 are the analyses yielding the smallest factors. These are the analyses with the fewest numbers of chemicals. As in previous discussions, no more will be said about these analyses.

The only other analyses for which overall uncertainty factors (the factors based on all chemicals included in an analysis) are less than 2.0 are Analyses 45 and 47, with the sieve. These supplemental analyses average either over sex (45) or over all species (47). As can be seen from Table 2-12, these two analyses are two of the best of the supplemental, indeed of all, analyses. Of course, the overall uncertainty factors are closely tied to loss as determined by the DISTANCE^2 function. Consequently, those producing small average loss

(e.g. those in Table 2-6) also yield relatively small uncertainty factors.

The factors estimated using only those chemicals with positive residual uncertainty (those for which the line does not intersect their vertical interval) generally follow the same pattern as the overall uncertainty factors. Since fewer chemicals are used to estimate these values, they may be less stable than the overall factors, however. The usefulness of separate "above the line" and "below the line" estimates can be visualized if one considers that the chemicals completely below the line are the ones of primary concern. They are the chemicals for which bioassay data overestimate RRDs (even given the conversion factor suggested by the best-fitting line). As long as one accepts that the health implications are worse when RRDs are overestimated than when they are underestimated, it may be reasonable to want to eliminate residual uncertainty with respect to the former but not with respect to the latter. One approach mentioned earlier is to use asymmetric loss functions; high degrees of asymmetry do act to eliminate the residual uncertainty of concern. Another approach, embodied here, is to estimate an uncertainty factor tailored to those chemicals below the line.

That uncertainty factor can be seen to vary between 0.009 (Analysis 21 without the sieve) and 0.363 (Analysis 45 with the sieve), still ignoring Analyses 6, 18, and 19. Generally, among the better analyses, the values indicate that predictions would need to be divided by a factor of 3 to 5 to account for the chemicals that overpredict RRDs.

Component-Specific Uncertainty. The supplemental analyses consist of an alternative standard (Analysis 30) and 18 variations of the standard. Each variant differs from the standard in only one respect, *i.e.* in the approach taken to one of the components defining the analyses. This supplemental set is used to investigate the uncertainty associated with each of those components.

The alternative standard accepts any experiment and assumes a mg/kg/day equivalence between humans and animals. This alternative is used in place of Analysis 0 because the correlation analysis and certain of the prediction analysis results suggest that allowing all routes of exposure

(Analysis 3b) is preferable to restricting the routes to those that humans encounter, gavage, inhalation, or oral. Moreover, mg/kg/day rather than m/m^2/day may be the preferred units for extrapolation when L_{20} is the predictor. All component-specific uncertainty investigations are limited to this predictor.

In such investigations, one is interested in how the RRD estimates change when a component is changed. Consequently, it is not necessary that there be epidemiologically derived estimates to use for comparison and so all 44 chemicals (not only the 23 with human data) can be used to address this question.

It is usually the case that any change in an assumption underlying a quantitative risk assessment will result in a change in the risk estimates. A component-specific uncertainty investigation should then tell us two things: how the risk estimates change and how consistently they change. A histogram approach has been used to address these issues.

Figures 2-42 through 2-59 display the histograms resulting from this investigation. Each histogram corresponds to one of the variations of the alternative standard analysis. The entry for a chemical in any histogram indicates the magnitude of the ratio of the RRD estimates (L_{20}) from the variant to that from the alternative standard. In this way, the distribution of the changes among the chemicals can be visualized.

Table 2-23 displays the mode of the distribution of each histogram. Also presented in that table is a dispersion factor that is analogous to the uncertainty factor used in the residual uncertainty analysis and specifies the average factor by which the ensemble of chemicals differs from the mode (cf. the Section 1 description of the methodology). This factor is dependent on the specific cut points chosen for the histograms, but because those cut points are the same for each analysis method, it is a valid means of comparing the components with respect to uncertainty. The greater the factor, the less consistent is the change in RRD estimates that results from the component change corresponding to the histogram. Less consistency (more chemical dependency) indicates

more uncertainty is associated with the corresponding component.

Figures 2-42 through 2-45 pertain to the choice of dose units used for animal-to-human extrapolation. These figures show relatively little dispersion of the chemicals and hence indicate little uncertainty associated with the choice of dose units. That is not to say that the resulting RRD estimates are not dramatically affected by changes in units. These are the only analyses for which the mode is not in the interval from 0.8 to 1.25 (Table 2-23). However, the dispersion factors and the figures indicate that changing units has a relctively predictable effect, one that is not chemical dependent. The plot in Figure 2-42, for example, largely reflects the standard values (body weight, surface area coefficient; cf. Table 1-3) used in the conversion from mg/kg/day to $mg/m^2/day$ in rats and mice. When those standards are used, the ratio of the RRD estimates (in all cases, the standard, using mg/kg/day, is in the denominator) is about 0.09 for mice and 0.21 for rats. The chemicals falling between 0.1 and 0.2 in Figure 2-42 are the result of using experiment-specific body weights or of cases in which changing units also changes the ordering of the experiments (due to species-specific changes) and so changes the experiment yielding the median estimate. Hence this figure, showing the greatest dispersion of the four because of the fairly even split between use of mice and rats, may even exaggerate the uncertainty here.

The next two histograms (Figures 2-46 and 2-47) relate to criteria placed on the length of observation and length of dosing, respectively. The relatively large dispersion noted is due to extreme changes in one or two chemicals. Restriction to long experiments decreases the RRD estimate for cigarette smoke by a factor over 1000. Similarly, restriction to experiments that dosed the treated animals for at least 80% of the standard experiment length increases RRD estimates for asbestos and cadmium by over three orders of magnitude.

No extreme changes are noted when experiments are limited to those using the route of exposure by which humans encounter the chemicals in question (Figure 2-48). However, for only 24 of the 44 chemicals were there studies employing that route. When a less restrictive criterion is used (the route just mentioned plus gavage, inhalation, and oral

routes; Figure 2-49), moderately extreme values do appear, benzo(a)pyrene and arsenic which change by factors of 847 and 410, respectively. These chemicals were not included in Figure 2-48 but account for the majority of the dispersion seen in Figure 2-49. Otherwise Figure 2-49 is less disperse than Figure 2-48.

The next four histograms (Figures 2-50 through 2-53) relate to the choice of endpoints to be analyzed. These are among the most disperse of the figures in the sense of including several extreme changes (Figures 2-50 and 2-53) and also in the sense of having less dominant modes (Figures 2-51 through 2-53). If not for two extreme changes (cigarette smoke and saccharin for which the ratios are 9.77×10^{-4} and 8.69×10^{-6}, respectively), Figure 2-50 (malignant tumors only) would display much less dispersion, being on the order of 13 rather than 291.

All of the histograms discussed above, except for those that relate to choice of dose units, depict changes that occur because a subset of the data are used. They demonstrate the effect such selections have on the location of the median lower bound RRD. A chemical with a ratio greater than 1 is one for which the selection tends to eliminate smaller estimates. Components such as these that relate to bioassay or response inclusion criteria can be very sensitive to the data that are available, certainly more so than those components that relate to manipulation of whatever data are available (such as the component related to the choice of dose units). This sensitivity is reflected in the fact that for these histograms fewer than the maximum number of chemicals (44) are addressable once the inclusion criteria are applied and may contribute strongly to the appearance of the extreme changes that have been noted. One must be aware that some confounding due to data availability may be present in the histograms of Figures 2-46 through 2-53.

On the other hand, those analyses that dictate how the experiment-specific RRDs are averaged are all based on the same data. In the standard analysis no averaging is performed. Analyses 45 through 47 (Figures 2-54 through 2-56) average results over sex alone, study alone, and species alone, respectively. The uncertainty associated with any of these procedures is small, the dispersion factors indicate that the

average change in the RRD estimates is less than a factor of 2.2.

However, if we limit attention to rats and mice (Figures 2-57 through 2-59) uncertainty is again great. This, too, may be in part a reflection of dependence on data availability. Consider the case of saccharin which contains mouse and rat bioassays predominantly. The rat studies are of better quality (they get rank 1 by the quality screen) than are the mouse studies (rank 3) so that when both are considered, only the rat studies are analyzed (the full sieve is used in the analyses represented in the histograms). Therefore, no change is seen when rats alone or the average of rat and mouse data are used (Figures 2-47 and 2-46, respectively). The mouse results are over five orders of magnitude smaller than the rat results.

Nevertheless, some species-specific changes can be discerned. Cigarette smoke is apparently more potent in rats than in other species. Arsenic is less potent in rats and mice although this may also reflect some data dependence. Overall, the choice of species appears to be a highly uncertain component of risk assessment as indicated by the large dispersion factors for Analyses 49 and 50, and by the difference in dispersion between Analyses 47 and 48 (Figures 2-56 and 2-57) which differ only in that species other than rats and mice are included in Analysis 47. It is easy to see how data availability can affect the estimates from any given species, above and beyond the question of the most appropriate species for any given chemical.

Table 2-1

CORRELATION COEFFICIENTS AND ASSOCIATED p-VALUES, BY ANALYSIS METHOD AND SIEVE

Analysis	# of Chemicals	No Screens ρ	No Screens p-value	Quality Screen ρ	Quality Screen p-value	Significance Screen ρ	Significance Screen p-value	Quality and Significance Screen ρ	Quality and Significance Screen p-value
0	20	.68	.0002	.73	.0001	.78	<.0001	.78	.0001
1	18	.55	.0095	.55	.0083	.68	.0013	.68	.0015
2	19	.61	.0034	.55	.0075	.49	.0187	.49	.0153
3a	17	.62	.0041	.64	.0026	.74	.0005	.73	.0007
3b	23	.80	<.0001	.77	<.0001	.78	<.0001	.90	<.0001
4a	20	.70	.0008	.76	<.0001	.77	<.0001	.78	.0001
4b	20	.67	.0004	.73	.0001	.76	<.0001	.76	.0001
4c	20	.67	.0008	.71	.0004	.77	.0002	.78	<.0001
4d	20	.68	.0004	.74	.0002	.77	.0001	.78	<.0001
5	20	.69	.0003	.74	.0001	.78	.0001	.79	<.0001
6	6	.96	.0028	.79	.0317	.93	.0106	.79	.0342
7	19	.55	.0079	.64	.0015	.72	.0003	.76	.0001
8a	13	.50	.0379	.56	.0207	.50	.0435	.56	.0214
8b	17	.60	.0050	.60	.0052	.66	.0013	.66	.0022
8c	18	.76	.0004	.71	.0009	.76	.0002	.76	.0001
9	20	.69	.0004	.70	.0007	.77	<.0001	.76	.0003
10	20	.71	.0004	.73	<.0001	.75	.0001	.77	.0002
11a	20	.60	.0025	.73	.0002	.69	.0011	.76	<.0001
11b	20	.66	.0009	.72	.0001	.73	.0003	.73	<.0001
11c	19	.77	.0002	.74	.0001	.79	.0001	.79	<.0001
11d	13	.62	.0121	.69	.0046	.80	.0006	.76	.0023
12	20	.75	.0005	.75	.0001	.73	.0001	.75	<.0001
13	18	.48	.0240	.50	.0172	.43	.0368	.43	.0416
14	19	.71	.0005	.75	.0001	.70	.0007	.71	.0005
15	18	.48	.0267	.50	.0177	.45	.0321	.46	.0316
16	13	.48	.0489	.49	.0472	.49	.0470	.49	.0436
17	11	.57	.0358	.57	.0369	.58	.0280	.58	.0301
18	10	.79	.0036	.76	.0046	.74	.0090	.73	.0090
19	9	.79	.0062	.79	.0057	.79	.0060	.79	.0058
20	17	.67	.0020	.64	.0035	.65	.0024	.63	.0043
21	13	.43	.0715	.37	.1046	.43	.0698	.38	.1023
22	15	.34	.1078	.34	.1075	.35	.1001	.35	.1036
23	13	.18	.2832	.01	.4904	.18	.2744	.18	.2821
24a	20	.76	<.0001	.76	.0004	.72	.0006	.75	.0001
24b	20	.73	.0004	.75	.0001	.71	.0003	.74	.0001
24c	20	.74	.0001	.76	<.0001	.72	<.0001	.74	.0001
24d	20	.73	.0003	.76	.0002	.71	.0001	.75	<.0001
25	16	.69	.0023	.64	.0042	.79	.0001	.81	.0002

Table 2-2

ABBREVIATIONS FOR CHEMICALS INCLUDED IN THE STUDY

Those with Suitable Epidemiolgical Data		Others	
Abbreviation	Chemical	Abbreviation	Chemical
AB	Asbestos	AC	Acrylonitrile
AF	Aflatoxin	AL	Allyl Chloride
AS	Arsenic	AM	4-Aminobiphenyl
BN	Benzene	BA	Benzo(a)pyrene
BZ	Benzidine	CO	Chlordane
CB	Chlorambucil	CT	Carbon Tetrachloride
CD	Cadmium	DB	3,3-Dichloro-benzidine
CR	Chromium		
CS	Cigarette Smoke	DE	1,2-Dichloroethane
DS	DES	DL	Vinylidene Chloride
EC	Epichlorohydrin	DP	Diphenylhydrazine
EO	Ethylene Oxide	ED	EDB
ES	Estrogen	FO	Formaldehyde
IS	Isoniazid	HC	Hexachlorobenzene
MC	Methylene Chloride	HY	Hydrazine
ML	Melphalan	LE	Lead
NC	Nickel	MU	Mustard Gas
PC	PCBs	NA	2-Naphthylamine
PH	Phenacetin	NT	NTA
RS	Reserpine	TD	TCDD
SC	Saccharin	TE	Tetracholorethylene
TC	Trichloroethylene	TP	2,4,6-Trichloro-phenol
VC	Vinyl Chloride		
		TO	Toxaphene

Table 2-3

AVERAGE LOSS AS DETERMINED BY THE SYMMETRIC DISTANCE^2 LOSS FUNCTION, BY ANALYSIS METHOD, PREDICTOR, AND SIEVE

Analysis	Predictor			
	L_M		L_{20}	
	No Sieve	Sieve	No Sieve	Sieve
0	.650	.570	.146	.298
1	.750	.630	.261	.377
3a	.742	.364	.215	.310
3b	.779	.620	.266	.113
4a	.701	.552	.124	.273
4b	.684	.599	.166	.316
4c	.646	.578	.157	.272
4d	.767	.516	.134	.267
5	.640	.548	.142	.285
6	.240	.159	.001	.001
7	.495	577	.200	.331
8a	.514	.279	.430	.224
8b	.550	1.088	.678	.597
8c	.367	.246	.189	.243
9	.645	.529	.141	.298
10	.541	.412	.523	.268
11a	.639	.465	.272	.239
11b	.659	.488	.241	.274
11c	.732	.577	.253	.277
11d	.490	.256	.280	.228
12	.541	.310	.541	.310
13	.352	.459	.352	.459
14	.279	.368	.279	.368
15	.368	.430	.368	.430
16	.352	.234	.352	.234
17	.289	.290	.289	.290
18	.052	.064	.052	.064
19	.090	.075	.090	.075
20	.441	.752	.441	.752
21	1.100	.964	1.100	.962
22	.530	.574	.530	.574
23	1.181	1.170	1.181	1.170
24a	.605	.298	.605	.298
24b	.580	.318	.580	.319
24c	.526	.296	.526	.296
24d	.630	.276	.630	.277
25	.278	.191	.232	.190

Table 2-4

AVERAGE LOSS AS DETERMINED BY THE SYMMETRIC CAUCHY LOSS FUNCTION, BY ANALYSIS METHOD, PREDICTOR, AND SIEVE

	Predictor							
	L_M		L_{20}		MLE_M		MLE_{20}	
Analysis	No Sieve	Sieve	No Sieve	Sieve	No Sieve	Sieve	No Sieve	Sieve
0	.566	.509	.440	.457	.586	.415	.507	.482
1	.540	.508	.478	.506	.569	.513	.619	.428
3a	.547	.464	.487	.467	.606	.479	.551	.492
3b	.551	.477	.453	.413	.558	.483	.520	.423
4a	.597	.492	.423	.437	.570	.502	.509	.463
4b	.546	.528	.420	.466	595	.536	.492	.495
4c	.546	.493	.398	.440	584	.502	.471	.467
4d	.605	.485	.428	.454	.573	.492	.508	.480
5	.560	.511	.440	.456	.591	.515	.500	.476
6	.453	.359	.270	.309	.446	.360	.261	.297
7	.496	.523	.521	.491	.559	.535	.645	.510
8a	.467.	441	.430	.419	.478	.463	.453	.444
8b	.459	.442	.439	.447	.604	.674	.685	.651
8c	.511	.533	.432	.494	.604	.578	.711	.599
9	.559	.516	.433	.643	.589	.524	.535	.490
10	.530	.482	.499	.465	.563	.495	.538	.485
11a	.560	.486	.476	.448	.575	.492	.575	.488
11b	.477	.447	.381	.408	.500	.466	.517	.433
11c	.429	.421	.378	.399	.445	.449	.527	.428
11d	.528	.490	.488	.451	.529	.458	.527	.421
12	.524	.460	.524	.460	.566	.474	.566	.474
13	.490	.519	.490	.519	.572	.567	.572	.567
14	.498	.493	.498	.493	.564	.510	.564	.579
15	.500	.489	.500	.489	.618	.579	.618	.579
16	.413	.409	.413	.409	.437	.431	.437	.431
17	.375	.383	.375	.363	.412	.399	.412	.399
18	.366	.377	.366	.377	.366	.373	.366	.373
19	.344	.322	.344	.322	.354	.327	.354	.32?
20	.410	.456	.410	.456	.614	.664	.614	.664
21	.242	.452	.424	.452	.817	.766	.817	.766
22	.445	.432	.445	.432	.664	.694	.664	.694
23	.460	.443	.460	.443	.665	.680	.665	.680
24a	.541	.448	.541	.448	.559	.468	.559	.468
24b	.522	.477	.522	.477	.571	.491	.571	.491
24c	.516	.451	.516	.451	.560	.466	.560	.466
24d	.550	.430	.550	.450	.570	.468	.570	.468
25	.463	.494	.470	.506	.558	.530	.494	.561

Table 2-5

AVERAGE LOSS AS DETERMINED BY THE SYMMETRIC TANH LOSS FUNCTION, BY ANALYSIS METHOD, PREDICTOR, AND SIEVE

	Predictor							
	L_M		L_{2Q}		MLE_M		MLE_{2Q}	
Analysis	No Sieve	Sieve	No Sieve	Sieve	No Sieve	Sieve	No Sieve	Sieve
0	.193	.180	.159	.170	.232	.214	.224	.207
1	.198	.190	.183	.186	.237	.226	.330	.225
3a	.163	.179	.172	.184	.211	.269	.315	.302
3b	.191	.175	.151	.140	.199	.175	.185	.141
4a	.201	.177	.156	.164	.237	.213	.224	.203
4b	.197	.212	.157	.175	.240	.219	.223	.212
4c	.194	.179	.152	.167	.234	.214	.216	.206
4d	.201	.175	.153	.166	.237	.210	.223	.205
5	.193	.180	.159	.169	.231	.214	.223	.206
6	.149	.134	.110	.120	.142	.134	.111	.118
7	.183	.184	.181	.177	.226	.221	.269	.216
8a	.187	.178	.179	.172	.246	.244	.246	.241
8b	.127	.136	125	.125	.357	.426	.466	.414
8c	.169	.153	142	.152	.301	.294	.590	.296
9	.193	.181	.158	.171	.234	.216	.248	.209
10	.181	.173	.176	.169	.224	.209	.234	.206
11a	.190	.175	.167	.162	.230	.210	.246	.247
11b	.192	.171	.156	.160	.222	.208	.216	.199
11c	.179	.172	.156	.162	.214	.212	.268	.202
11d	.197	.173	.174	.163	.235	.213	.322	.205
12	.180	.169	.180	.169	.225	.204	.225	.204
13	.177	.187	.177	.187	.236	.281	.236	.281
14	.175	.176	.175	.176	.225	.214	.225	.215
15	.178	.176	.178	.176	.289	.280	.289	.280
16	.176	.172	.176	.172	.244	.243	.244	.243
17	.123	.121	.123	.121	.206	.203	.206	.203
18	.183	.185	.183	.185	.181	.185	.181	.185
19	.124	.120	.124	.120	.126	.121	.126	.121
20	.121	.130	.121	.130	.365	.421	.365	.421
21	.141	.149	.141	.149	.456	.516	.456	.516
22	.140	.137	.140	.137	.395	.398	.395	.398
23	.153	.152	.153	.152	.375	.379	.375	.379
24a	.186	.167	.186	.167	.227	.204	.227	.204
24b	.184	.173	.184	.173	.229	.208	.229	.208
24c	.179	.169	.179	.169	.223	.205	.223	.205
24d	.187	.166	.187	.166	.229	.209	.229	.203
25	.159	.152	.154	.151	.252	.247	.252	.251

Table 2-8

COMPARISON OF ANALYSES; FIVE BEST ANALYSES,
BY PREDICTOR AND LOSS FUNCTION

	Loss Function					
	DISTANCE^2		CAUCHY		TANH	
Predictor	Analysis	Avg.Loss	Analysis	Avg.Loss	Analysis	Avg.Loss
L_M	18	.052 (ns)[a]	19	.322 (s)	19	.120 (s)
	19	.075 (s)	6	.359 (s)	20	.121 (ns)
	6	.159 (s)	17	.363 (s)	17	.121 (s)
	25	.191 (s)	18	.366 (ns)	8b	.127 (ns)
	2	.209 (ns)	16	.409 (s)	6	.134 (s)
L_{2Q}	6	.001 (s)	6	.270 (ns)	6	.110 (ns)
	18	.052 (ns)	19	.302 (s)	19	.120 (s)
	19	.075 (s)	17	.363 (s)	20	.121 (ns)
	3b	.113 (s)	11c	.378 (ns)	17	.121 (s)
	4a	.124 (ns)	11b	.381 (ns)	8b	.125 (s)
MLE_M	--[b]		19	.327 (s)	19	.121 (s)
			6	.360 (s)	6	.134 (s)
			18	.366 (ns)	3b	.175 (s)
			17	.399 (s)	24d	.203 (s)
			16	.431 (s)	17	.203 (s)
MLE_{2Q}	--[b]		6	.261 (ns)	6	.111 (ns)
			19	.327 (s)	19	.121 (s)
			18	.366 (ns)	3b	.131 (s)
			17	.399 (s)	11b	.199 (s)
			11d	.421 (s)	11c	.202 (s)

[a]The loss given is the smaller of the two losses (with and without the sieve) for any analysis. The code in parentheses indicates whether it comes from the analysis without the sieve (ns) or with the sieve (s).

[b]The DISTANCE^2 loss function is not used with MLE predictors.

Table 2-7

COMPARISON OF ANALYSES; FIVE BEST ANALYSES, EXCLUDING ANALYSES 6, 18, AND 19, BY PREDICTOR AND LOSS FUNCTION

	Loss Function					
	DISTANCE^2		CAUCHY		TANH	
Predictor	Analysis	Avg.Loss	Analysis	Avg.Loss	Analysis	Avg.Loss
L_M	25	.191 (s)[a]	17	.363 (s)	20	.121 (ns)
	2	.209 (ns)	16	.409 (s)	17	.121 (s)
	16	.234 (s)	20	.410 (ns)	8b	.127 (ns)
	8c	.254 (s)	11c	.421 (s)	22	.137 (s)
	11d	.256 (s)	21	.424 (ns)	21	.141 (ns)
L_{20}	3b	.113 (s)	17	.363 (s)	20	.121 (ns)
	4a	.124 (ns)	11c	.378 (ns)	17	.121 (s)
	3a	.130 (s)	11b	.381 (ns)	8b	.125 (s)
	4d	.134 (ns)	16	.409 (s)	22	.137 (s)
	9	.141 (ns)	20	.410 (ns)	3b	.140 (s)
MLE_M	--[b]		17	.399 (s)	3b	.175 (s)
			16	.431 (s)	24d	.203 (s)
			11c	.445 (ns)	17a	.203 (s)
			11d	.458 (s)	24a	.204 (s)
			11b	.466 (s)	12	.204 (s)
MLE_{20}	--[b]		17	.399 (s)	3b	.141 (s)
			11d	.421 (s)	11b	.199 (s)
			3b	.423 (s)	11c	.202 (s)
			11c	.428 (s)	24d	.203 (s)
			16	.431 (s)	17	.203 (s)

[a]The loss given is the smaller of the two losses (with and without the sieve) for any analysis. The code in parentheses indicates whether it comes from the analysis without the sieve (ns) or with the sieve (s).

[b]The DISTANCE^2 loss function is not used with MLE predictors.

Table 2-8

TOTAL INCREMENTAL NORMALIZED LOSSES,
BY ANALYSIS AND SIEVE[a]

Analysis	Total Incremental Normalized Loss	
	No Sieve	Sieve
0	1.019	1.418
1	1.693	1.997
2	1.543	2.175
3a	1.719	1.390
3b	1.079	0.555
4a	0.920	1.198
4b	0.900	1.559
4c	0.698	1.258
4d	0.853	1.313
5	1.015	1.385
7	1.835	1.736
8a	1.534	1.176
8b	0.996	0.963
8c	0.758	1.239
9	0.961	1.466
10	1.944	1.417
11a	1.450	1.194
11b	0.746	0.983
11c	0.741	0.968
11d	1.627	1.215
12	2.156	1.430
13	1.751	2.158
14	1.695	1.767
15	1.836	1.804
16	1.324	1.133
17	0.259	0.166
20	0.558	1.231
21	1.553	1.700
22	1.117	1.043
23	2.004	1.888
24a	2.398	1.325
24b	2.242	1.591
24c	2.084	1.369
24d	2.484	1.301
25	1.183	1.292

[a]Calculated using L_{20} as the predictor.

Table 2-9

AVERAGE LOSS FOR RESTRICTED SETS OF CHEMICALS
FOR ANALYSES 3b, 17, and 20, BY LOSS FUNCTION[a]

Sets of Chemicals	Analysis	Loss Function		
		DISTANCE^2	CAUCHY	TANH
11 to which Analysis 17 is applicable	3b with sieve	0.053	0.414	0.131
	17 with sieve	0.290	0.363	0.121
	20 w/o sieve	0.161	0.409	0.133
17 to which Analysis 20 is applicable	3b with sieve	0.082	0.360	0.100
	20 w/o sieve	0.441	0.410	0.121

[a]The L_{2Q} predictor is used.

Table 2-10

AVERAGE LOSS AS DETERMINED BY THE ASYMMETRIC TANH LOSS FUNCTION FOR L_M, BY ANALYSIS AND DEGREE OF ASYMMETRY

Analysis[a]	Asymmetry Constant (m)					
	1.5	2	5	10	50	100
0	.207	.224	.279	.312	.319	.319
1	.220	.244	.296	.328	.332	.332
2	.204	.222	.264	.284	.290	.290
3a	.200	.217	.260	.293	.321	.321
3b	.191	.205	.247	.283	.336	.336
4a	.201	.218	.272	.306	.312	.312
4b	.211	.228	.282	.312	.319	.319
4c	.205	.222	.275	.305	.312	.312
4d	.199	.218	.270	.305	.310	.310
5	.207	.225	.276	.310	.316	.316
6	.144	.149	.161	.173	.173	.173
7	.213	.232	.281	.310	.313	.313
8a	.200	.218	.282	.290	.296	.296
8b	.157	.170	.215	.276	.352	.352
8c	.177	.186	.211	.232	.280	.280
9	.208	.226	.279	.311	.314	.314
10	.197	.214	.266	.294	.301	.301
11a	.200	.218	.273	.305	.309	.309
11b	.192	.209	.264	.296	.300	.300
11c	.191	.207	.260	.292	.297	.297
11d	.201	.220	.271	.291	.291	.291
12	.192	.208	.260	.287	.288	.288
13	.211	.224	.256	.267	.267	.267
14	.199	.216	.265	.292	.301	.301
15	.200	.219	.263	.276	.285	.285
16	.189	.205	.275	.293	.312	.312
17	.140	.157	.218	.234	.250	.250
18	.198	.210	.279	.335	.363	.363
19	.134	.146	.215	.267	.288	.288
20	.154	.169	.210	.270	.340	.340
21	.168	.177	.225	.293	.340	.340
22	.149	.160	.196	.225	.230	.230
23	.170	.187	.257	.342	.377	.377
24a	.188	.204	.255	.283	.287	.287
24b	.195	.211	.259	.288	.289	.289
24c	.189	.205	.253	.281	.282	.282
24d	.188	.204	.254	.282	.287	.287
25	.177	.196	.228	.253	.305	.305

[a]Analyses have been performed using the sieve.

Table 2-11

AVERAGE LOSS AS DETERMINED BY THE ASYMMETRIC TANH LOSS FUNCTION FOR L_{2Q}, BY ANALYSIS AND DEGREE OF ASYMMETRY

Analysis[a]	Asymmetry Constant (m)					
	1.5	2	5	10	50	100
0	.188	.203	.249	.283	.290	.290
1	.208	.220	.267	.299	.304	.304
2	.207	.223	.264	.279	.287	.287
3a	.195	.207	.237	.264	.293	.293
3b	.156	.169	.210	.235	.283	.285
4a	.184	.198	.240	.275	.281	.281
4b	.193	.205	.250	.281	.288	.288
4c	.185	.199	.243	.274	.281	.281
4d	.186	.200	.240	.275	.281	.281
5	.188	.202	.247	.282	.288	.288
6	.126	.130	.144	.149	.149	.149
7	.199	.214	.260	.290	.293	.293
8a	.189	.204	.269	.277	.283	.283
8b	.148	.162	.201	.261	.339	.339
8c	.170	.179	.205	.227	.272	.272
9	.189	.205	.257	.290	.292	.292
10	.190	.207	.251	.279	.286	.286
11a	.184	.199	.247	.270	.284	.284
11b	.181	.197	.246	.272	275	.275
11c	.182	.196	.236	.269	.275	.275
11d	.189	.207	.258	.279	.279	.279
12	.192	.208	.260	.287	.288	.288
13	.211	.224	.256	.267	.267	.267
14	.199	.216	.265	.292	.301	.301
15	.200	.219	.263	.276	.285	.285
16	.189	.205	.275	.293	.312	.312
17	.140	.157	.218	.234	.250	.250
18	.198	.210	.279	.335	.363	.363
19	.134	.146	.215	.267	.288	.288
20	.154	.169	.210	.270	.340	.340
21	.168	.177	.225	.293	.340	.340
22	.149	.160	.196	.225	.230	.230
23	.170	.187	.257	.342	.377	.377
24a	.188	.204	.255	.283	.287	.287
24b	.195	.211	.259	.288	.289	.289
24c	.189	.205	.253	.281	.282	.282
24d	.188	.204	.254	.282	.287	.287
25	.176	.189	.220	.244	.297	.297

[a]Analyses have been performed using the sieve.

Table 2-12

Y-INTERCEPT VALUES FOR BEST-FITTING LINES,
L_M PREDICTOR, BY ANALYSIS, SIEVE, AND LOSS FUNCTION

	No Sieve			Sieve		
Analysis	DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
0	1.587	0.510	1.714	0.827	1.066	1.067
1	1.450	0.931	1.067	0.821	1.357	1.067
2	0.927	0.824	1.067	0.404	0.537	0.665
3a	1.683	2.004	1.922	1.117	1.284	1.374
3b	1.528	1.820	2.164	1.976	0.530	1.260
4a	0.890	-0.462	1.517	0.087	-0.095	0.072
4b	1.352	0.308	1.032	0.609	0.546	0.742
4c	1.309	0.344	1.079	0.596	0.599	0.725
4d	2.563	1.230	2.919	1.582	1.621	1.617
5	1.546	0.440	1.610	0.765	1.084	1.071
6	2.453	1.519	2.086	2.300	1.488	2.086
7	1.151	1.723	1.417	0.730	0.960	1.067
8a	1.120	1.135	1.493	0.907	1.045	1.555
8b	0.869	0.071	0.302	0.455	-0.174	0.233
8c	0.860	1.732	1.391	0.583	1.247	0.813
9	1.565	0.493	1.714	0.770	0.956	0.966
10	1.162	0.665	1.208	0.611	0.771	0.955
11a	1.370	0.340	1.439	0.734	0.839	0.874
11b	1.322	0.359	1.391	0.701	0.462	0.603
11c	1.235	0.361	0.519	0.748	0.291	0.447
11d	1.301	0.228	1.097	0.197	0.849	0.788
12	0.939	0.664	0.939	0.444	0.540	0.749
13	0.319	0.709	0.727	0.168	-0.045	0.272
14	0.482	1.003	0.929	0.456	0.360	0.731
15	0.298	0.693	0.744	0.237	0.291	0.272
16	0.905	0.613	0.710	0.749	0.679	0.631
17	0.709	0.443	0.467	0.772	0.451	0.447
18	1.274	0.518	0.634	1.308	0.616	0.988
19	1.188	0.374	0.447	1.254	0.450	0.447
20	0.450	-0.161	-0.110	0.159	0.038	0.233
21	0.106	-0.159	0.233	0.053	-0.736	0.233
22	0.233	-0.679	-0.549	0.286	-0.657	-0.549
23	-0.073	-0.564	-0.058	-0.120	-0.715	-0.613
24a	0.245	0.113	0.361	-0.305	-0.372	-0.212
24b	0.691	0.182	0.498	0.226	0.249	0.470
24c	0.634	0.328	0.545	0.209	0.261	0.471
24d	1.812	1.943	2.179	1.201	1.296	1.364
25	1.033	1.758	1.714	0.722	1.381	0.955

Table 2-13

Y-INTERCEPT VALUES FOR BEST-FITTING LINES,
L_{2Q} PREDICTOR, BY ANALYSIS, SIEVE, AND LOSS FUNCTION

	No Sieve			Sieve		
Analysis	DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
0	0.528	0.380	0.532	0.474	0.199	0.315
1	0.416	-0.270	0.272	0.550	0.135	0.583
2	0.630	0.258	0.315	0.201	-0.067	0.272
3a	0.635	1.454	1.339	0.695	0.624	0.822
3b	1.314	0.612	1.067	1.223	0.927	1.080
4a	-0.257	-0.428	-0.143	-0.358	-0.566	-0.401
4b	0.225	0.115	0.226	0.236	-0.230	0.069
4c	0.150	0.211	0.338	0.186	-0.079	0.024
4d	1.282	1.310	1.598	1.183	1.017	1.222
5	0.493	0.334	0.434	0.435	0.154	0.315
6	1.667	1.670	1.873	1.667	1.415	1.873
7	0.370	0.872	0.977	0.428	0.210	0.555
8a	1.045	0.686	0.742	0.793	0.683	0.742
8b	0.145	-0.081	0.233	0.072	-0.036	0.233
8c	0.258	0.289	0.413	0.478	-0.357	0.654
9	0.607	0.433	0.532	0.494	0.278	0.449
10	0.925	0.292	1.065	0.356	0.461	0.459
11a	0.664	0.447	0.841	0.161	0.183	0.283
11b	0.515	-0.064	0.173	0.347	0.293	0.447
11c	0.047	-0.093	0.267	0.297	0.155	0.283
11d	0.808	0.251	1.067	0.085	0.633	0.571
12	0.939	0.664	0.939	0.444	0.540	0.749
13	0.319	0.709	0.727	0.168	-0.045	0.272
14	0.482	1.003	0.929	0.456	0.360	0.731
15	0.298	0.693	0.744	0.237	0.291	0.272
16	0.905	0.613	0.710	0.749	0.679	0.631
17	0.769	0.443	0.447	0.772	0.451	0.447
18	1.274	0.518	0.634	1.308	0.616	0.988
19	1.188	0.374	0.447	1.254	0.450	0.447
20	0.450	-0.161	-0.110	0.159	0.038	0.233
21	0.106	-0.159	0.233	0.053	-0.736	0.233
22	0.233	-0.679	-0.549	0.286	-0.657	-0.549
23	-0.073	-0.564	-0.058	-0.120	-0.715	-0.613
24a	0.245	0.113	0.361	-0.305	-0.372	-0.212
24b	0.691	0.182	0.498	0.226	0.249	0.470
24c	0.634	0.328	0.545	0.209	0.261	0.471
24d	1.817	1.943	2.179	1.201	1.296	1.364
25	0.516	0.492	0.571	0.651	0.722	0.813

Table 2-14

Y-INTERCEPT VALUES FOR BEST-FITTING LINES,
MLE_M PREDICTOR, BY ANALYSIS, SIEVE, AND LOSS FUNCTION

	No Sieve			Sieve		
Analysis	DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
0	-1.481	0.327	1.366	-1.814	0.994	0.928
1	-1.617	1.059	1.366	-1.895	1.186	1.201
2	-2.307	0.257	0.514	-4.060	0.567	0.514
3a	-1.684	1.811	1.762	-2.009	1.159	1.201
3b	2.116	0.783	1.708	1.815	0.768	1.124
4a	-2.192	-0.496	0.478	-2.556	-0.081	-0.069
4b	-1.666	-0.065	1.487	-2.020	0.624	0.610
4c	-1.685	-0.014	1.534	-2.014	0.487	0.592
4d	-0.710	1.238	2.002	-1.117	1.557	1.476
5	-1.528	0.278	1.299	-1.860	0.994	0.933
6	2.129	1.470	1.946	2.129	1.319	1.946
7	-1.936	1.113	0.928	-2.154	0.846	0.882
8a	-2.600	1.026	1.366	-2.681	0.917	1.008
8b	-4.428	0.444	0.352	-6.527	0.633	0.449
8c	-4.192	1.231	0.882	-4.292	1.261	0.882
9	-1.522	0.292	1.366	-1.864	0.783	0.752
10	-1.903	0.338	0.759	-2.079	0.680	0.859
11a	-1.689	0.519	0.894	-1.920	0.784	0.737
11b	-1.713	0.294	0.505	-1.958	0.374	0.505
11c	-1.852	0.162	0.365	-2.056	0.177	0.299
11d	-2.861	0.663	0.908	-3.358	0.752	0.908
12	-2.150	0.132	0.480	-2.215	0.423	0.555
13	-2.828	-0.768	-0.081	-4.380	0.200	0.299
14	-2.520	0.460	0.455	-2.441	0.267	0.528
15	-4.601	0.372	0.514	-4.502	0.068	0.031
16	-2.932	0.466	0.471	-2.917	0.539	0.733
17	-3.238	0.308	0.299	-3.222	0.308	0.299
18	1.098	0.343	0.387	1.176	0.446	0.739
19	1.027	0.210	0.299	1.094	0.276	0.299
20	-4.892	-0.352	-0.378	-6.720	0.744	0.306
21	-6.075	-4.276	0.023	-8.220	-1.706	0.549
22	-7.516	-0.969	-0.499	-7.438	-1.086	-0.192
23	-8.007	-0.300	-0.294	-8.063	-0.902	-0.255
24a	-2.874	-0.670	-0.429	02.961	-0.482	-0.353
24b	-2.348	-0.085	0.209	-2.409	-0.183	-0.279
24c	-2.381	-0.015	0.210	-2.416	0.169	0.281
24d	-1.371	0.980	1.411	-1.519	1.174	1.178
25	-2.502	1.341	0.882	-2.581	1.302	0.882

Table 2-15

Y-INTERCEPT VALUES FOR BEST-FITTING LINES, MLE_{2Q} PREDICTOR, BY ANALYSIS, SIEVE, AND LOSS FUNCTION

	No Sieve			Sieve		
Analysis	DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
0	-4.168	0.010	0.080	-2.253	0.074	0.372
1	-5.846	-0.057	0.579	-2.302	0.154	5.793
2	-5.944	0.035	0.199	-5.788	0.041	0.199
3a	-6.189	1.016	0.643	-2.417	0.596	0.064
3b	-1.873	-0.045	0.643	1.075	0.744	0.946
4a	-4.923	-0.785	-0.594	-2.999	-0.693	-0.543
4b	-4.358	-0.303	-0.226	-2.484	-0.324	0.115
4c	-4.365	-0.241	-0.179	-2.488	-0.204	-0.065
4d	-3.482	0.933	1.179	-1.544	0.905	1.179
5	-4.203	-0.027	-0.029	-2.303	0.034	0.273
6	1.524	1.310	1.454	1.524	1.180	1.454
7	-8.983	0.537	0.574	-2.444	0.114	0.372
8a	-2.775	0.563	0.604	-2.845	0.561	0.603
8b	-8.556	1.231	0.145	-6.719	1.166	0.248
8c	-15.326	-0.697	-0.564	-4.436	0.463	0.574
9	-5.517	0.387	0.207	-2.218	0.147	0.372
10	-3.880	-0.046	0.500	-2.294	0.310	0.520
11a	-4.041	0.240	0.363	-4.031	0.156	0.251
11b	-2.631	-0.162	-0.095	-2.327	0.168	0.299
11c	-4.437	-0.986	-0.465	-2.366	0.009	0.131
11d	-5.660	0.451	0.643	-3.491	0.535	0.643
12	-2.150	0.132	0.480	-2.215	0.423	0.555
13	-2.828	-0.768	-0.087	4.380	0.200	0.299
14	2.520	0.460	0.455	-2.440	0.267	0.528
15	-4.601	0.372	0.514	-4.502	0.068	0.031
16	-2.933	0.466	0.471	-2.917	0.539	0.733
17	-3.238	0.308	0.299	-3.222	0.308	0.299
18	1.098	0.343	0.387	1.176	0.446	0.739
19	1.027	0.210	0.299	1.094	0.276	0.299
20	-4.982	-0.352	-0.378	-6.720	0.744	0.306
21	-6.075	-4.264	0.022	-8.220	-1.706	0.549
22	-7.516	-0.969	-0.499	-7.438	-1.086	-0.192
23	-8.007	-0.300	-0.294	-8.063	-0.902	-0.255
24a	-2.874	-0.679	-0.429	-2.961	-0.482	-0.353
24b	-2.349	-0.085	0.209	-2.409	0.183	0.279
24c	-2.380	-0.015	0.021	-2.416	0.169	0.281
24d	-1.371	0.980	1.411	-1.518	1.174	1.178
25	-10.801	-0.227	-0.047	-2.735	0.644	0.602

Table 2-16

AVERAGE LOSS FOR SUPPLEMENTAL ANALYSES WITH THE L_{20} PRECICTOR, BY ANALYSIS, SIEVE, AND LOSS FUNCTION

	No Sieve			Sieve		
Analysis	DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
30	0.224	0.441	0.149	0.107	0.390	0.137
31	0.266	0.453	0.151	0.113	0.413	0.140
32	0.263	0.434	0.152	0.124	0.420	0.142
33	0.273	0.442	0.154	0.141	0.437	0.149
34	0.288	0.458	0.150	0.131	0.400	0.138
35	0.213	0.468	0.176	0.738	0.511	0.198
36	0.370	0.457	0.173	0.549	0.502	0.182
37	0.195	0.489	0.181	0.129	0.445	0.174
38	0.124	0.434	0.156	0.273	0.437	0.164
41	0.330	0.555	0.183	1.303	0.535	0.223
42	0.622	0.462	0.174	0.567	0.461	0.172
43	0.185	0.403	0.110	0.253	0.435	0.118
44	0.229	0.461	0.146	0.887	0.577	0.182
45	0.163	0.406	0.142	0.072	0.376	0.134
46	0.220	0.454	0.154	0.234	0.411	0.145
47	0.271	0.469	0.163	0.084	0.378	0.133
48	0.367	0.486	0.166	0.792	0.509	0.180
49	0.172	0.404	0.149	0.953	0.509	0.199
50	0.496	0.538	0.168	0.711	0.448	0.154

Table 2-17

Y-INTERCEPT VALUES FOR BEST-FITTING LINES,
AMONG SUPPLEMENTAL ANALYSES,[a]
BY ANALYSIS, SIEVE, AND LOSS FUNCTION

	No Sieve			Sieve		
Analysis	DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
30	0.431	-0.147	0.072	0.387	0.033	0.230
31	1.314	0.612	1.067	1.223	0.927	1.080
32	1.099	0.250	0.575	0.950	0.655	0.774
33	1.056	0.157	0.557	0.868	0.277	0.820
34	2.217	1.606	1.784	2.015	1.863	1.901
35	0.011	-1.063	-0.226	0.475	-0.362	-0.180
36	-0.177	-0.578	-0.402	-0.597	-0.910	-0.650
37	-0.113	0.673	0.350	-0.076	-0.308	-0.209
38	-0.257	-0.428	-0.143	-0.358	-0.566	-0.401
41	-0.021	0.921	0.072	0.549	-0.529	-0.180
42	0.803	0.034	0.476	0.689	0.005	0.476
43	-0.238	-0.747	-0.545	-0.015	-0.338	-0.295
44	-0.437	-0.308	-0.217	0.475	-0.030	0.106
45	-0.561	-0.014	0.148	0.467	0.063	0.230
46	0.799	0.161	0.467	0.588	-0.028	0.148
47	0.531	-0.076	0.230	0.249	0.000	0.230
48	0.321	-0.363	-0.007	0.229	-0.120	0.230
49	-0.105	-0.516	-0.344	0.257	-0.490	-0.238
50	0.446	0.628	0.230	0.443	0.184	0.117

[a]The L_{2Q} predictor is used.

Table 2-18

AVERAGE LOSS, BY DOSE UNITS, SIEVE AND LOSS FUNCTION[a]

Units	Analysis	No Sieve			Sieve		
		DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
mg/m^2/day	0	0.146	0.440	0.159	0.298	0.457	0.170
	12	0.541	0.524	0.180	0.310	0.460	0.169
	31	0.266	0.453	0.151	0.113	0.413	0.140
mg/kg/day	4a	0.124	0.434	0.156	0.273	0.437	0.164
	24a	0.605	0.541	0.186	0.298	0.448	0.167
	30	0.224	0.441	0.149	0.107	0.390	0.137
ppm diet	4b	0.166	0.420	0.157	0.316	0.466	0.175
	24b	0.580	0.522	0.184	0.319	0.477	0.173
	32	0.263	0.434	0.152	0.124	0.420	0.142
ppm air	4c	0.157	0.398	0.152	0.272	0.440	0.167
	24c	0.526	0.516	0.179	0.296	0.451	0.169
	33	0.273	0.442	0.154	0.141	0.437	0.149
mg/kg/life	4d	0.134	0.428	0.153	0.267	0.454	0.166
	24d	0.630	0.550	0.187	0.277	0.450	0.166
	34	0.288	0.458	0.150	0.131	0.400	0.138

[a]The L_{2Q} predictor is used.

Table 2-19

Y-INTERCEPTS BY DOSE UNITS, SIEVE, AND LOSS FUNCTION[a]

Units	Analysis	No Sieve			Sieve		
		DISTANCE^2	CAUCHY	TANH	DISTANCE^2	CAUCHY	TANH
mg/m^2/day	0	0.528	0.380	0.532	0.474	0.199	0.315
	12	0.939	0.664	0.939	0.444	0.540	0.749
	31	1.314	0.612	1.067	1.223	0.927	1.080
mg/kg/day	4a	-0.257	-0.428	-0.143	-0.358	-0.566	-0.401
	24a	0.245	0.113	0.361	-0.305	-0.372	-0.212
	30	0.431	-0.147	0.072	0.387	0.033	0.230
ppm diet	4b	0.225	0.115	0.226	0.236	-0.230	0.069
	24b	0.691	0.182	0.498	0.226	0.249	0.470
	32	1.099	0.250	0.575	0.950	0.655	0.774
ppm air	4c	0.150	0.211	0.338	0.186	-0.079	0.024
	24c	0.634	0.328	0.545	0.209	0.261	0.471
	33	1.056	0.157	0.557	0.868	0.277	0.820
mg/kg/life	4d	1.282	1.310	1.598	1.183	1.017	1.222
	24d	1.817	1.943	2.179	1.201	1.293	1.364
	34	2.217	1.606	1.784	2.015	1.863	1.901

[a]The L_{20} predictor is used.

Table 2-20

CONVERSION FACTORS[a] FOR ALL DOSE UNITS,
BY METHOD OF ANALYSIS AND SIEVE[b]

Units	Analysis Method	No Sieve	Sieve
$mg/m^2/day$	Restricted routes, unaveraged (0)	2.40 - 3.40	1.58 - 2.07
	Restricted routes, averaged[c] (12)	4.61 - 8.69	3.47 - 5.61
	Unrestricted routes, unaveraged (31)	4.09 - 11.67	8.45 - 12.02
mg/kg/day	Restricted routes, unaveraged (4a)	0.37 - 0.72	0.28 - 0.40
	Restricted routes, averaged[c] (24a)	1.30 - 2.30	0.43 - 0.61
	Unrestricted routes, unaveraged (30)	0.72 - 1.18	1.08 - 1.70
ppm diet	Restricted routes, unaveraged (4b)	1.30 - 1.68	0.59 - 1.17
	Restricted routes, averaged[c] (24b)	1.52 - 3.15	1.77 - 2.95
	Unrestricted routes, unaveraged (32)	3.76 - 8.91	4.52 - 5.94
ppm air	Restricted routes, unaveraged (4c)	1.62 - 2.18	0.83 - 1.06
	Restricted routes, averaged[c] (24c)	2.13 - 3.51	1.82 - 2.96
	Unrestricted routes, unaveraged (33)	1.43 - 3.61	1.89 - 6.61
mg/kg/life	Restricted routes, unaveraged (4d)	20.42 - 39.63	10.40 - 16.67
	Restricted routes, averaged[c] (24d)	87.70 - 151.01	19.63 - 23.12
	Unrestricted routes, unaveraged (34)	40.36 - 60.81	72.95 - 79.62

[a]The factor by which a bioassay-based RRD estimate is multiplied to give best fit, on average, to the human RRD estimates (RRD_H/RRD_A).
[b]The range given is that suggested by the CAUCHY and TANH loss functions, the two that use point estimates of human RRDs.
[c]Averaged analyses average over sex, study, and species, in that order.

Table 2-21

UNCERTAINTY FACTORS FOR ANALYSES WITHOUT THE SIEVE[a]

Analysis	All Chemicals		Chemicals Below Line[b]		Chemicals Above Line[b]	
	n[c]	Factor	n[d]	Factor	n[e]	Factor
0	20	2.257	3	0.188	5	3.292
1	18	3.381	4	0.185	5	5.194
2	19	6.852	3	0.068	4	8.248
3a	17	2.862	3	0.146	6	3.242
3b	23	4.216	6	0.210	4	8.251
4a	20	2.046	4	0.291	4	3.752
4b	20	2.488	3	0.171	6	3.228
4c	20	2.504	3	0.155	6	2.735
4d	20	2.131	4	0.269	5	3.336
5	20	2.239	3	0.194	6	2.950
6	6	1.048	1	0.874	1	1.144
7	19	2.800	4	0.204	6	3.642
8a	13	8.005	2	0.071	2	16.426
8b	17	29.008	4	0.043	4	10.066
8c	18	2.936	4	0.252	5	4.246
9	20	2.227	3	0.185	5	3.111
10	20	10.240	5	0.127	3	29.165
11a	20	3.745	4	0.183	3	11.055
11b	20	5.040	5	0.315	2	21.608
11c	19	4.530	4	0.137	3	6.038
11d	13	4.570	3	0.233	1	36.415
12	20	10.065	5	0.119	3	29.894
13	18	5.731	4	0.159	5	7.409
14	19	3.918	4	0.160	5	5.541
15	18	6.018	4	0.150	5	7.688
16	13	5.871	2	0.092	2	12.588
17	11	4.174	1	0.047	2	6.959
18	10	1.467	2	0.462	1	3.529
19	9	1.790	2	0.428	1	5.442
20	17	7.113	5	0.132	4	8.968
21	13	62.713	2	0.009	3	24.689
22	15	8.561	5	0.141	5	8.539
23	13	89.156	3	0.017	4	16.448
24a	20	11.292	5	0.100	3	34.730
24b	20	9.954	5	0.107	3	32.919
24c	20	10.541	5	0.126	3	29.441
24d	20	10.455	4	0.067	4	21.383
25	16	3.032	3	0.180	3	6.725
30	23	3.275	6	0.231	4	6.391
31	23	4.216	6	0.210	4	8.251
32	23	4.328	8	0.278	4	8.818
33	23	4.408	7	0.249	4	9.259
34	23	4.728	6	0.190	4	8.649
35	20	2.858	6	0.255	4	5.403

Table 2-21 (continued)

UNCERTAINTY FACTORS FOR ANALYSES WITHOUT THE SIEVE[a]

Analysis	All Chemicals n[c]	All Chemicals Factor	Chemicals Below Line[b] n[d]	Chemicals Below Line[b] Factor	Chemicals Above Line[b] n[e]	Chemicals Above Line[b] Factor
36	19	5.954	4	0.119	4	8.894
37	17	2.662	3	0.161	6	3.068
38	20	2.046	4	0.219	4	3.752
41	20	4.604	4	0.110	4	7.186
42	16	17.063	3	0.039	3	14.234
43	17	2.817	3	0.199	4	4.362
44	19	3.164	3	0.148	4	5.718
45	23	2.557	5	0.264	4	5.261
46	23	3.104	6	0.235	5	5.397
47	23	3.856	7	0.211	5	6.016
48	23	4.623	8	0.190	6	6.973
49	21	2.657	4	0.239	3	6.728
50	18	6.807	4	0.073	6	6.186

[a]The L_{20} predictor is used.
[b]The line is the best-fitting line determined by the DISTANCE^2 loss function.
[c]Number of chemicals in analyses.
[d]Number of chemicals with human RRD intervals completely below line.
[e]Number of chemicals with human RRD intervals completely above line.

Table 2-22

UNCERTAINTY FACTORS FOR ANALYSES WITH THE SIEVE[a]

Analysis	All Chemicals n[c]	All Chemicals Factor	Chemicals Below Line[b] n[d]	Chemicals Below Line[b] Factor	Chemicals Above Line[b] n[e]	Chemicals Above Line[b] Factor
0	20	5.300	3	0.076	5	4.623
1	18	6.676	3	0.066	4	6.407
2	19	10.029	4	0.077	4	15.975
3a	17	2.119	4	0.307	3	4.190
3b	23	2.008	4	0.249	4	3.342
4a	20	4.552	3	0.091	4	5.534
4b	20	5.454	3	0.072	5	4.968
4c	20	4.616	3	0.086	5	4.327
4d	20	4.448	3	0.092	4	5.322
5	20	5.060	3	0.081	5	4.462
6	6	1.048	1	0.874	1	1.144
7	19	5.422	3	0.078	5	5.275
8a	13	3.406	1	0.044	2	5.471
8b	17	22.834	3	0.031	4	8.148
8c	18	3.235	3	0.149	4	5.786
9	20	5.426	3	0.073	6	3.917
10	20	4.493	3	0.088	7	3.432
11a	20	3.611	4	0.143	5	4.175
11b	20	4.535	3	0.094	4	5.777
11c	19	4.444	3	0.099	4	5.589
11d	13	3.066	2	0.115	3	4.282
12	20	5.393	3	0.079	4	6.250
13	18	8.254	5	0.145	4	12.263
14	19	6.022	3	0.073	3	10.427
15	18	7.670	4	0.117	4	10.793
16	13	3.508	1	0.049	2	6.523
17	11	4.181	1	0.047	2	6.931
18	10	1.560	2	0.433	1	4.128
19	9	1.675	3	0.529	1	4.675
20	17	31.128	4	0.040	4	13.087
21	13	36.455	3	0.030	4	22.798
22	15	9.010	5	0.128	4	11.914
23	13	82.564	3	0.018	4	17.378
24a	20	5.162	3	0.084	4	6.150
24b	20	5.518	3	0.072	5	5.054
24c	20	5.101	3	0.082	5	4.869
24d	20	4.718	3	0.090	4	5.639
25	16	2.645	4	0.250	3	5.322
30	23	2.026	5	0.283	4	3.097
31	23	2.008	4	0.249	4	3.432
32	23	2.216	4	0.231	4	3.578
33	23	2.256	5	0.272	4	4.362
34	23	2.293	3	0.172	4	3.647
35	20	78.202	6	0.146	4	42.184

Table 2-22 (continued)

UNCERTAINTY FACTORS FOR ANALYSES WITH THE SIEVE[a]

Analysis	All Chemicals n[c]	All Chemicals Factor	Chemicals Below Line[b] n[d]	Chemicals Below Line[b] Factor	Chemicals Above Line[b] n[e]	Chemicals Above Line[b] Factor
36	19	10.196	5	0.111	4	16.301
37	17	2.076	3	0.247	4	3.319
38	20	4.552	3	0.091	4	5.584
41	20	96.444	5	0.034	4	91.020
42	16	14.512	3	0.048	3	13.458
43	17	4.200	4	0.219	3	8.401
44	19	84.246	5	0.089	3	87.708
45	23	1.670	5	0.363	4	2.608
46	23	3.594	7	0.256	4	7.166
47	23	1.770	4	0.298	3	3.354
48	23	67.502	5	0.059	4	42.975
49	21	129.229	6	0.074	4	57.072
50	18	23.545	4	0.062	4	20.743

[a]The L_{2Q} predictor is used.
[b]The line is the best-fitting line determined by the DISTANCE^2 loss function.
[c]Number of chemicals in analyses.
[d]Number of chemicals with human RRD intervals completely below line.
[e]Number of chemicals with human RRD intervals completely above line.

Table 2-23

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]	Number of Extremes[d]
31	44	.05 - .1	2.3	0
32	44	.2 - .5	1.7	0
33	44	.2 - .5	1.8	0
34	44	.02 - .05	1.3	3
35	40	.8 - 1.25	28.5	1
36	34	.8 - 1.25	86.0	4
37	24	.8 - 1.25	5.3	0
38	40	.8 - 1.25	33.7	2
41	39	.8 - 1.25	290.6	3
42	29	.8 - 1.25	75.6	1
43	31	.8 - 1.25	39.6	1
44	37	.8 - 1.25	54.1	4
45	44	.8 - 1.25	1.2	0
46	44	.8 - 1.25	1.7	0
47	44	.8 - 1.25	2.2	0
48	43	.8 - 1.25	23.2	2
49	39	.8 - 1.25	39.6	3
50	36	.8 - 1.25	335.6	3

[a]The ratios are of the chemical-specific RRD estimates from the indicated analysis to those of Analysis 30, the alternative standard.

[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 chemicals differ from the mode.

[d]The number of chemicals for which the ratios are greater than 100 or less than 0.01.

Figure 2-1

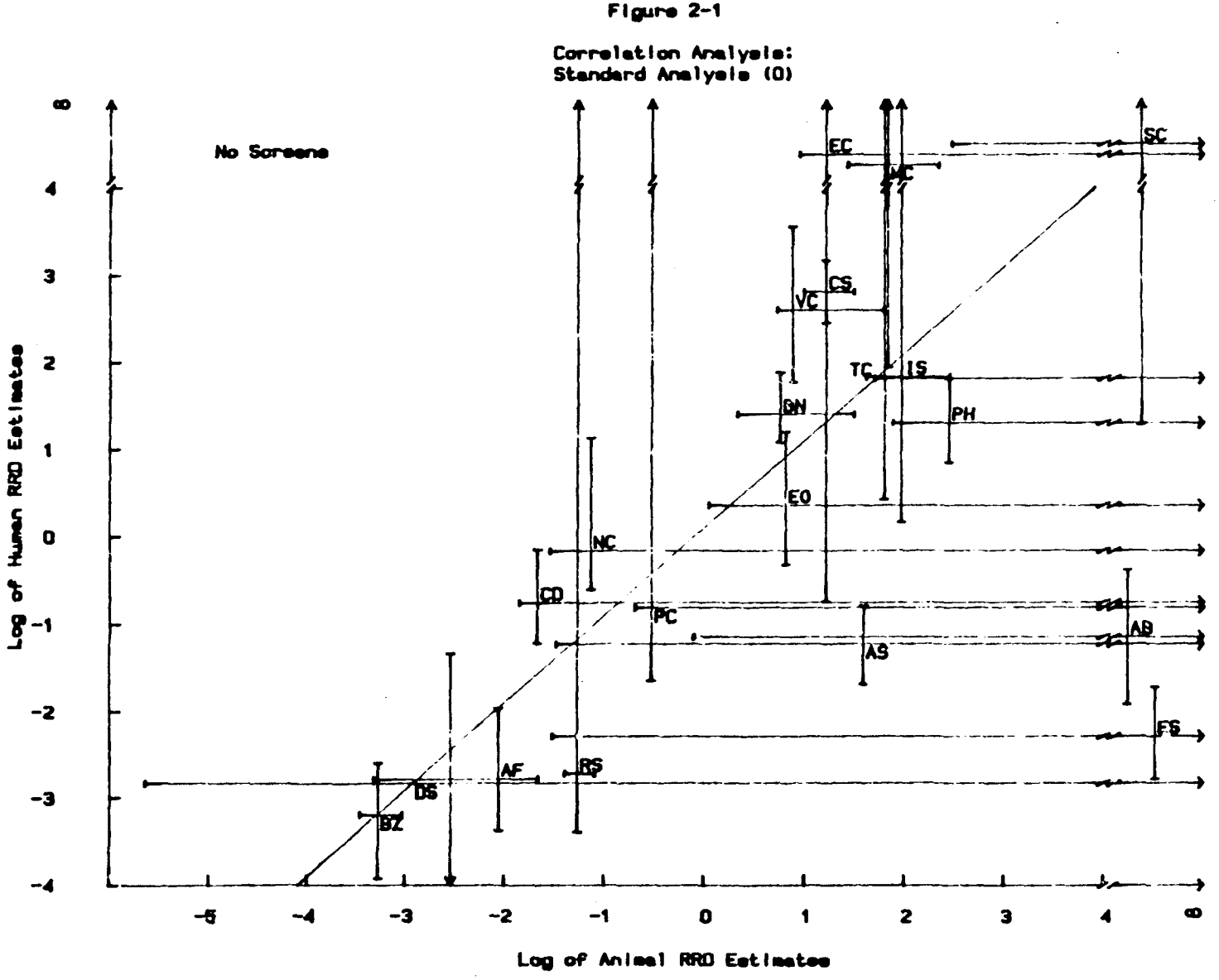

Correlation Analysis:
Standard Analysis (0)
No Screens
Log of Human RRD Estimates
Log of Animal RRD Estimates
∞
4
3
2
1
0
-1
-2
-3
-4
-5
-4
-3
-2
-1
0
1
2
3
4
∞
EC
MC
SC
CS
VC
TC
IS
SN
PH
EO
NC
CD
PC
AS
AB
FS
RS
AF
DS
BZ

Figure 2-2

Correlation Analysis:
Standard Analysis (0)

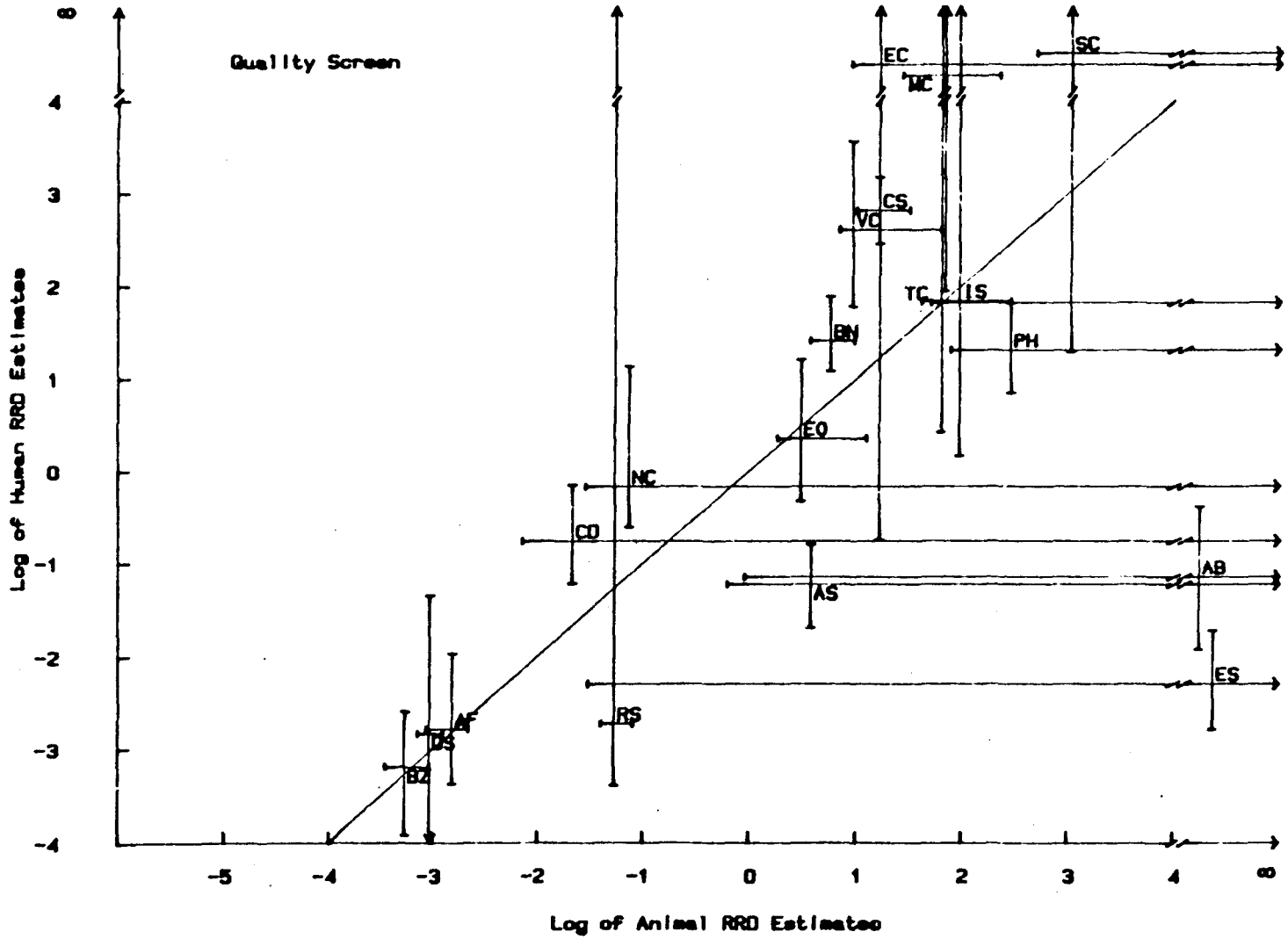


Figure 2-3

Correlation Analysis:
Standard Analysis (0)

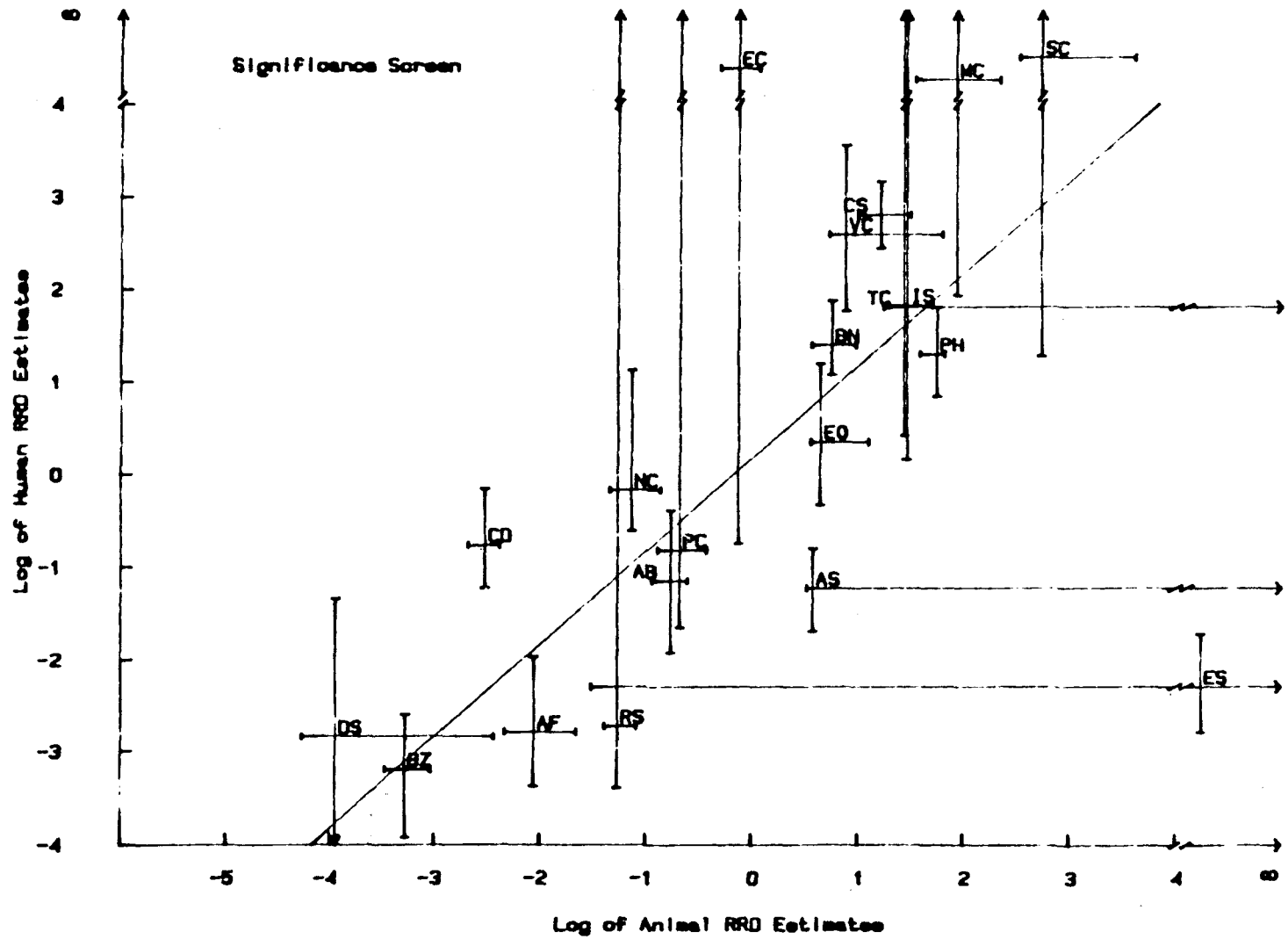


Figure 2-4

Correlation Analysis:
Standard Analysis (0)

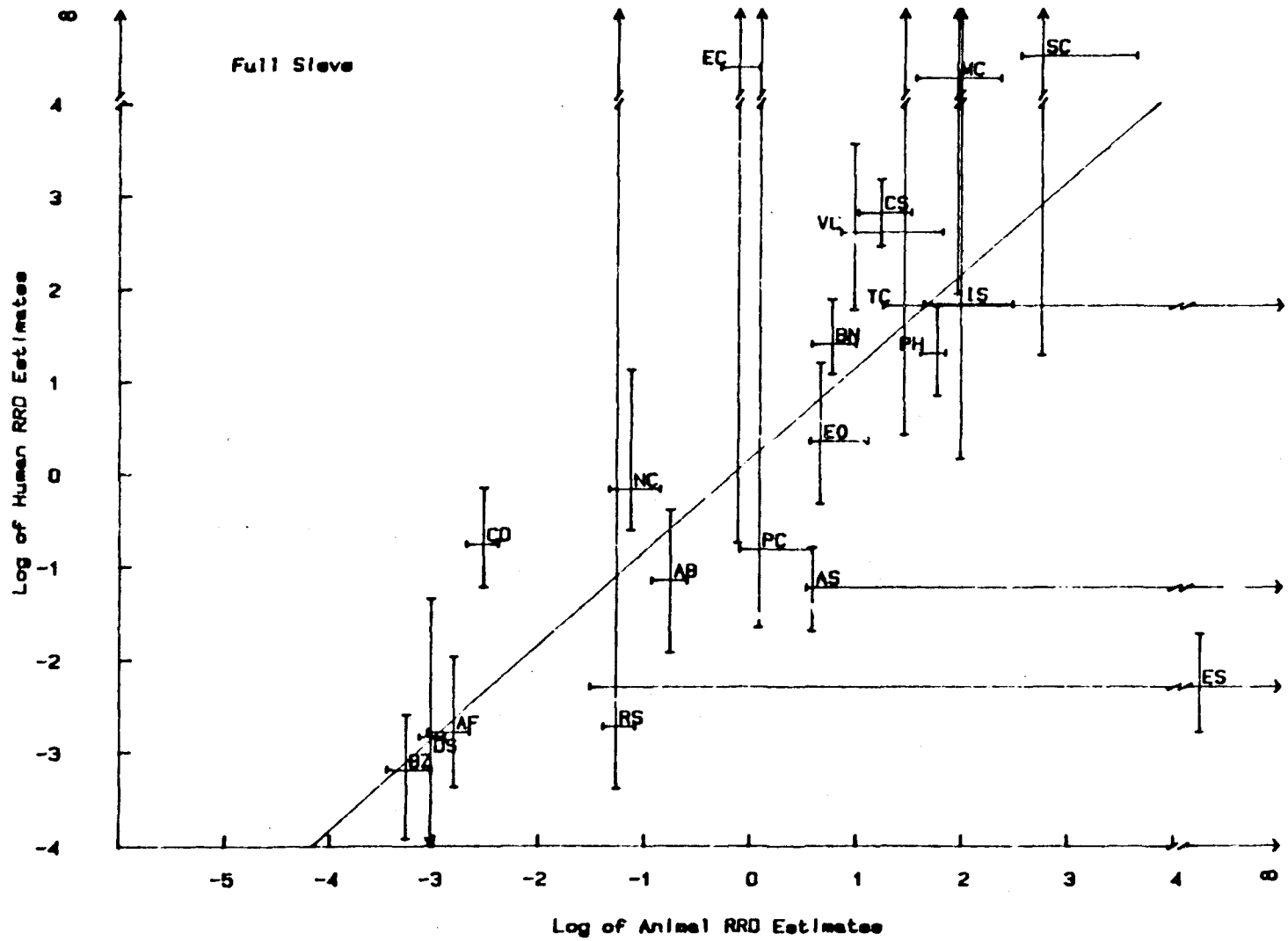


Figure 2-5

Correlation Analysis:
Long Experiments Only (1)

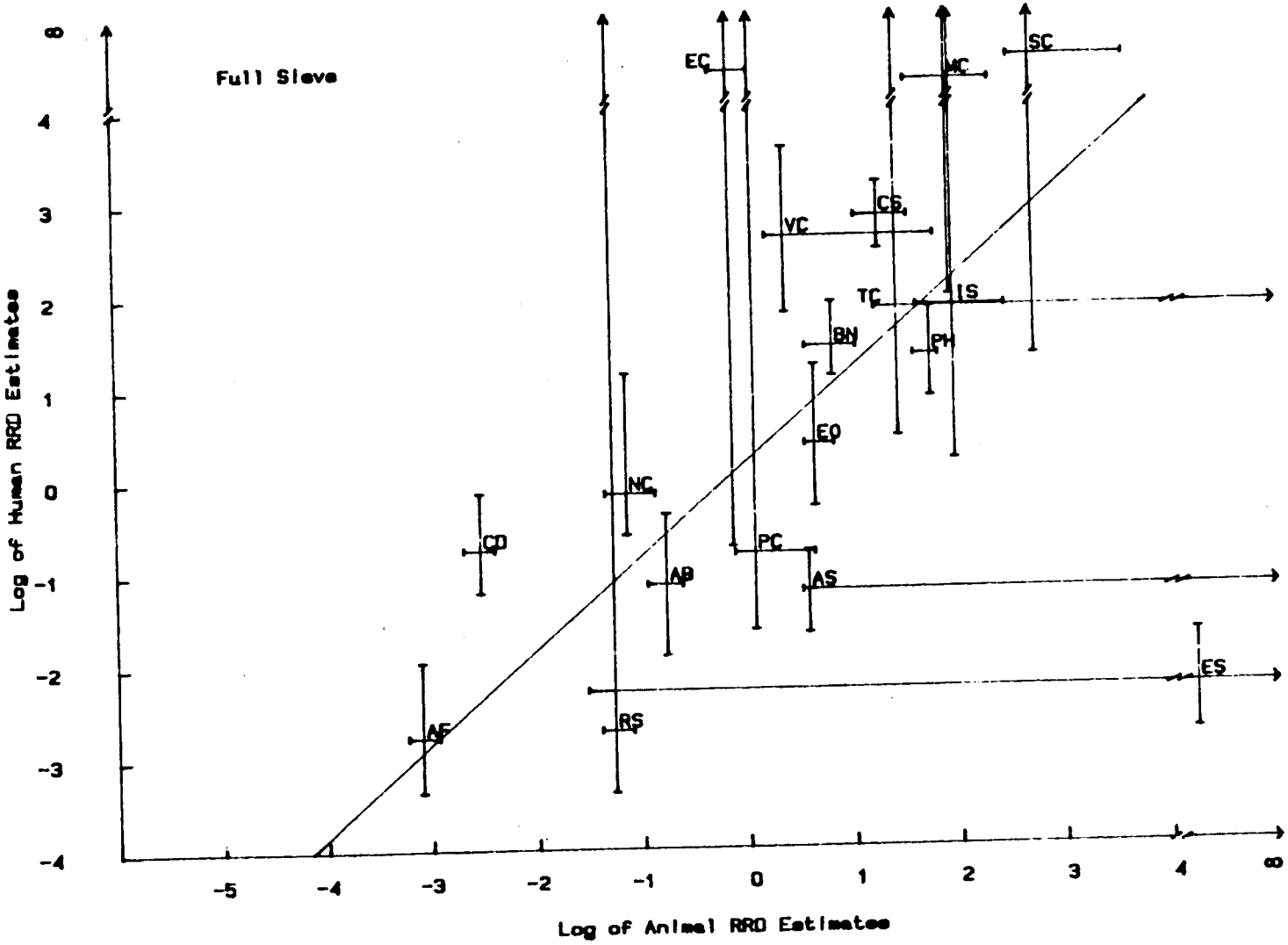


Figure 2-6

orrelation Analysis:
ung Dosing Only (2)

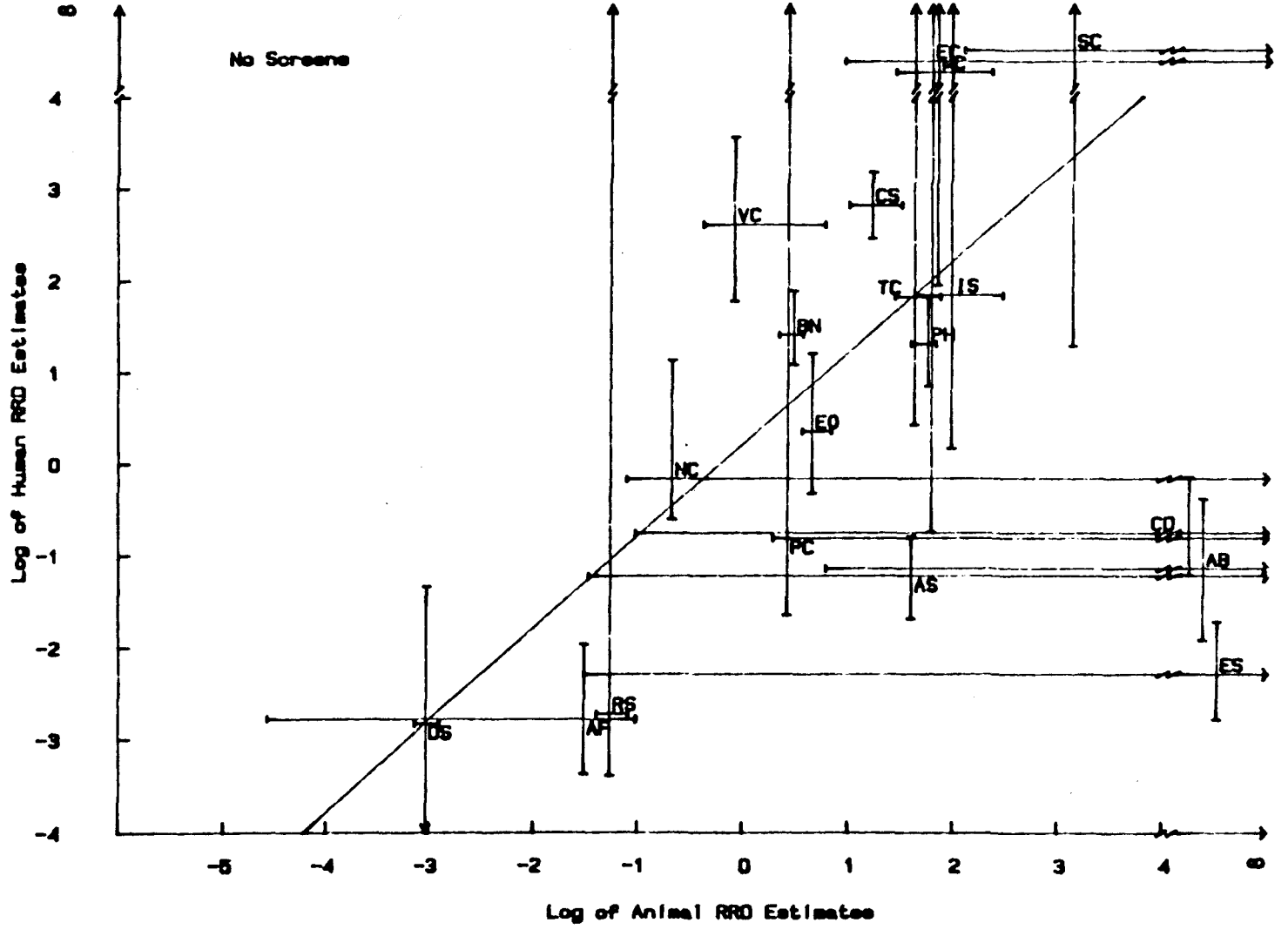


Figure 2-7

Correlation Analysis:
Long Dosing Only (2)

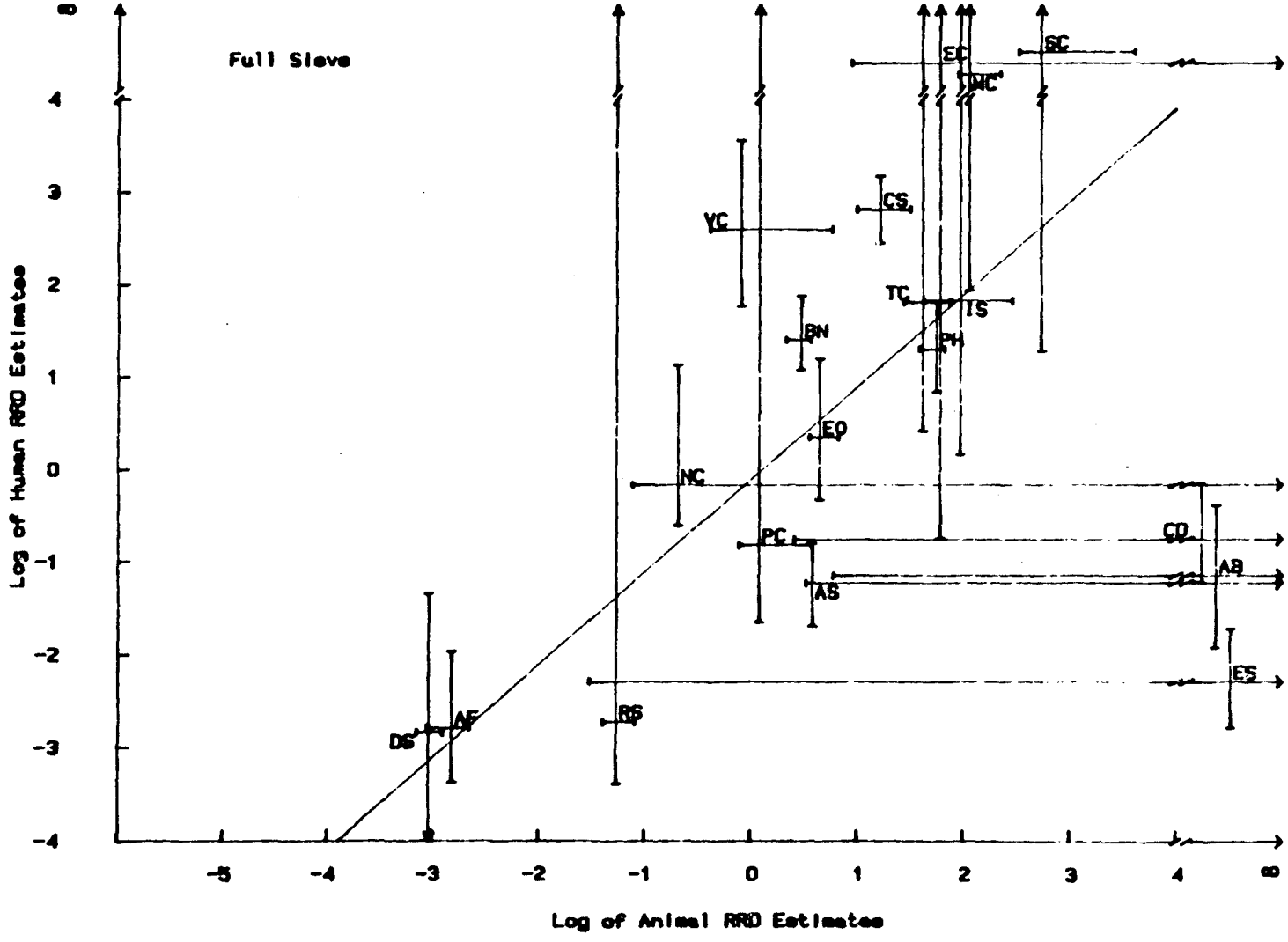


Figure 2-8

Correlation Analysis:
Route That Humans Encounter (3a)

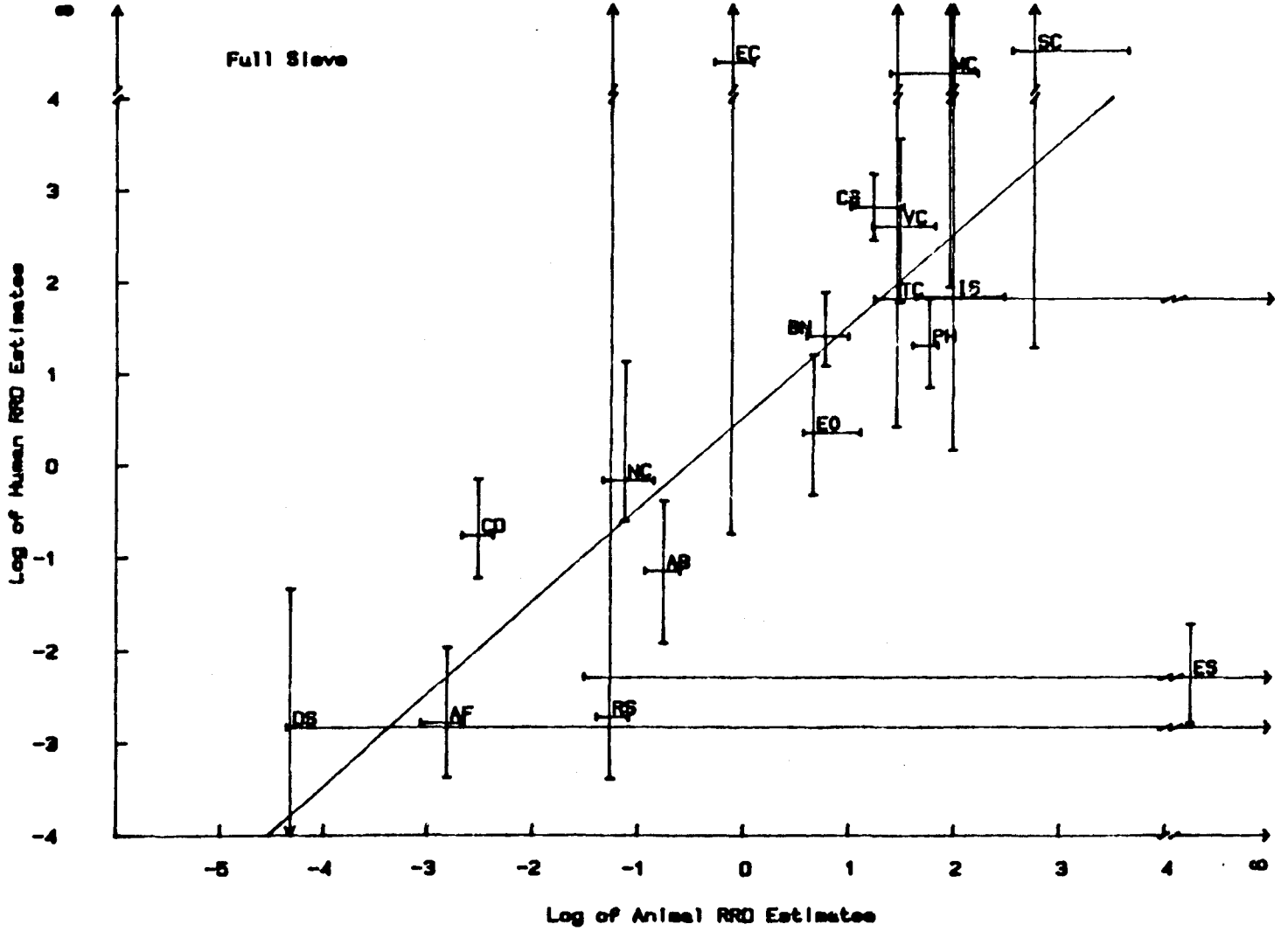


Figure 2-8

Correlation Analysis:
Any Route of Exposure (3b)

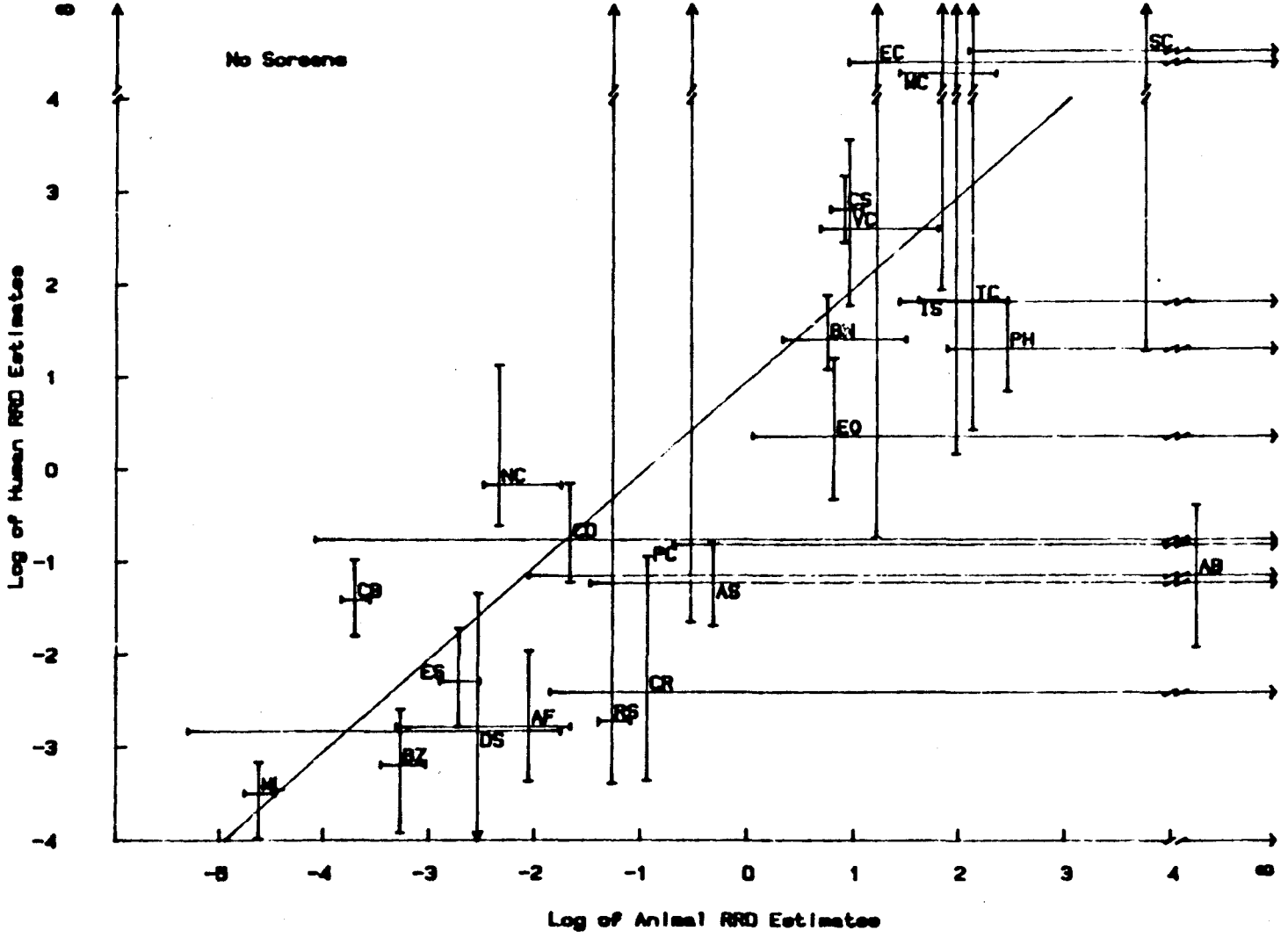


Figure 2-10

Correlation Analysis:
Any Route of Exposure (3b)

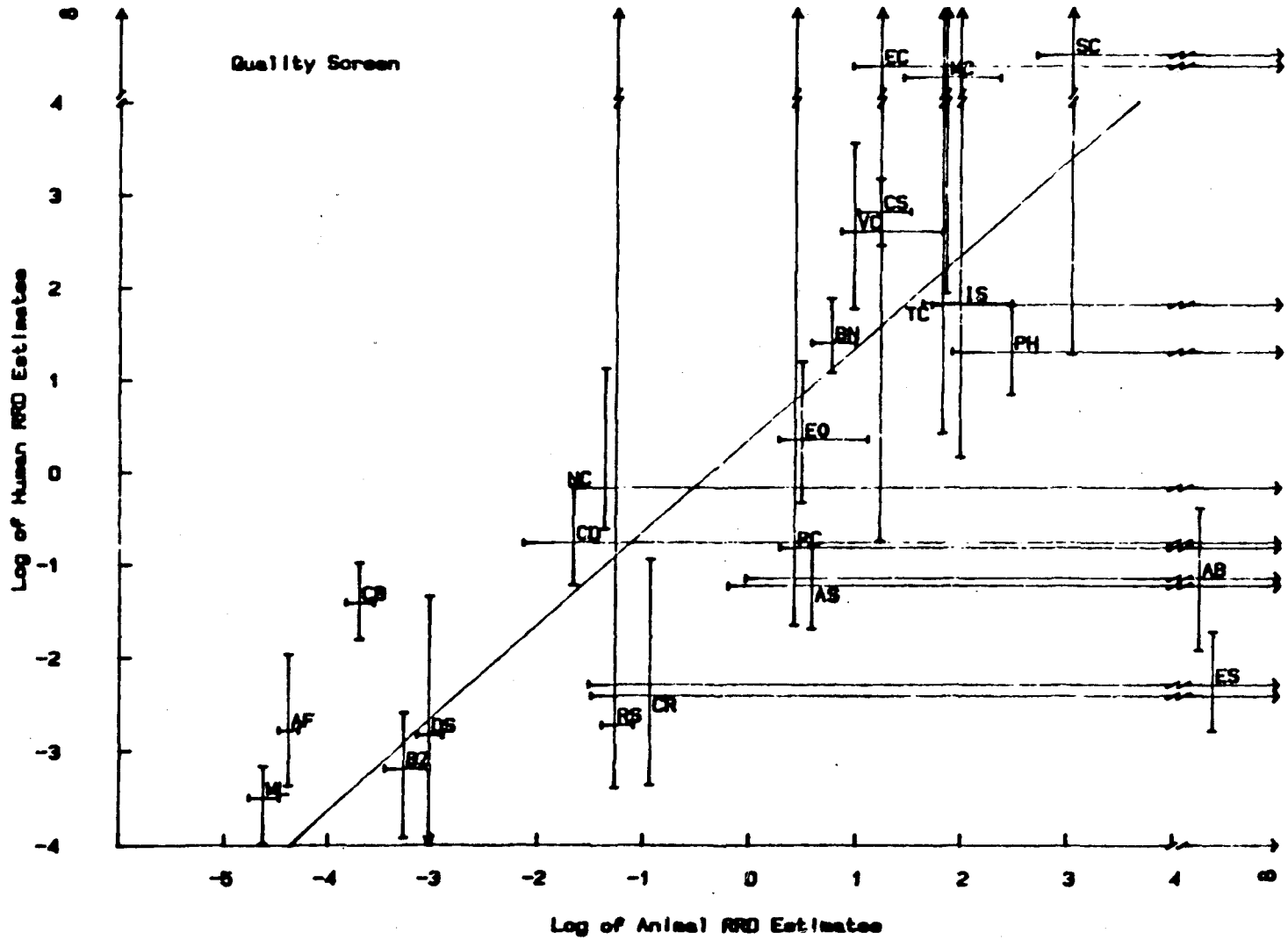


Figure 2-11

Correlation Analysis:
Any Route of Exposure (3b)

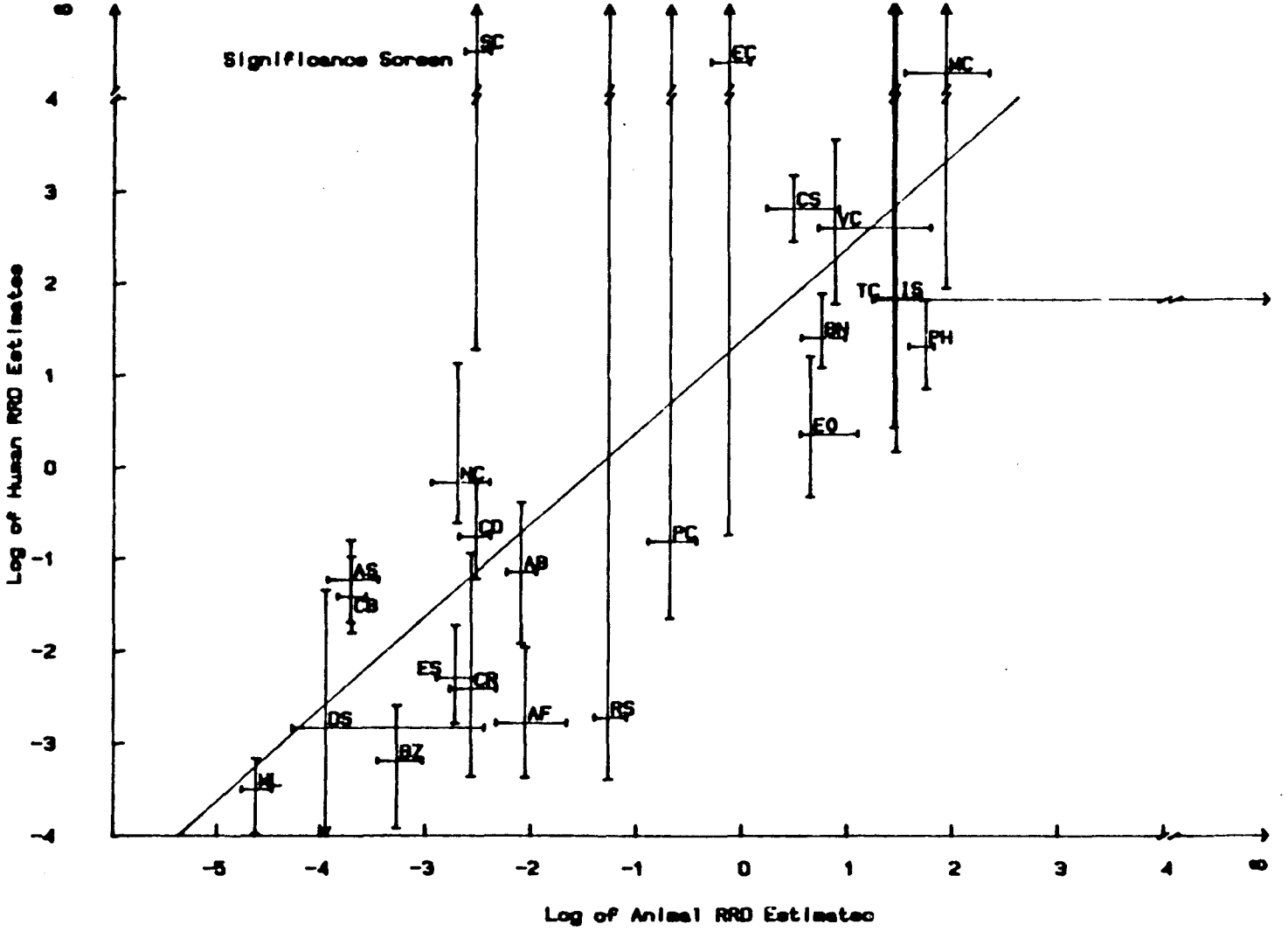


Figure 2-12

Correlation Analysis:
Any Route of Exposure (3b)

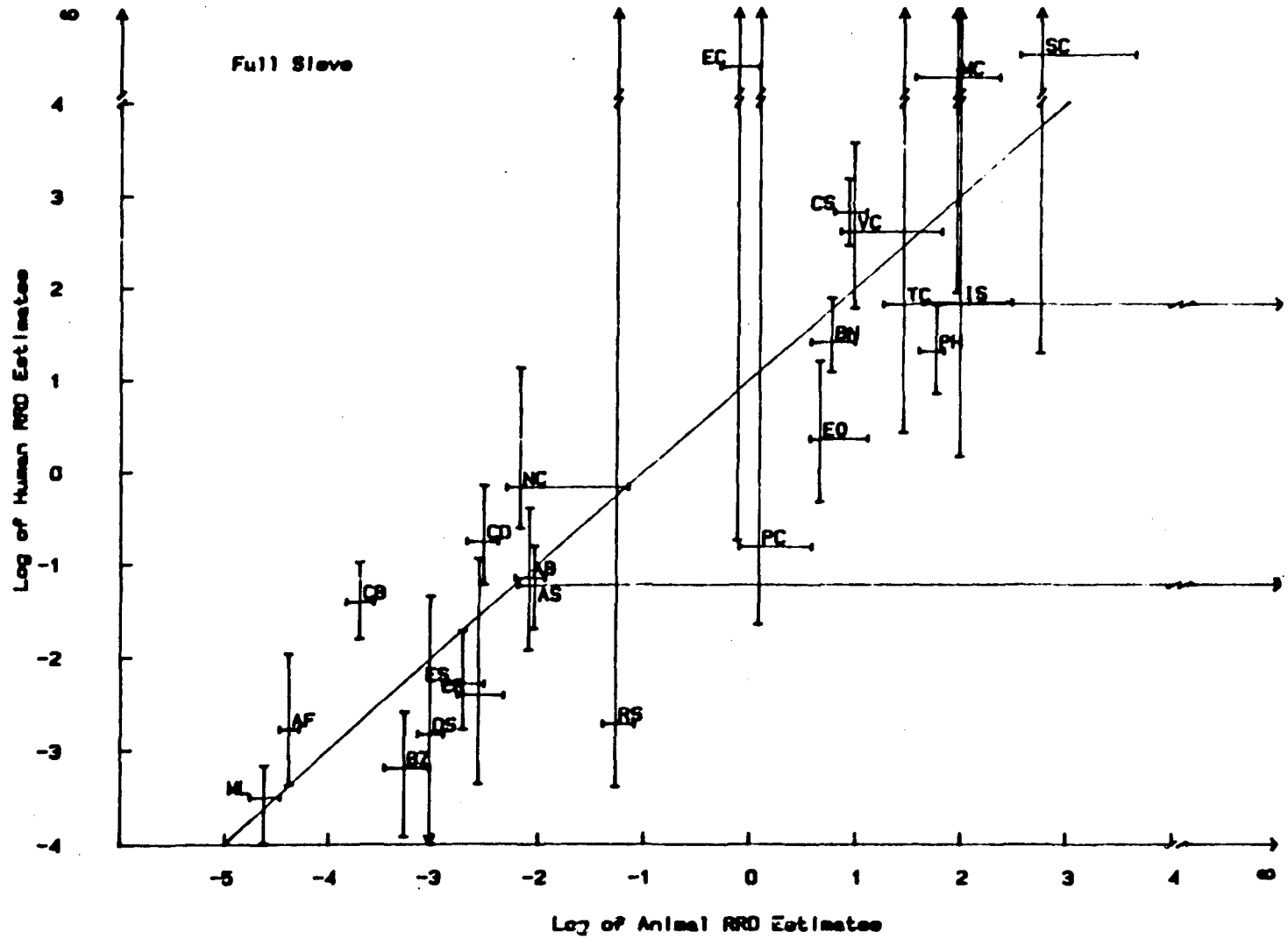


Figure 2-13

Correlation Analysis:
Average Dose over 80% of Experiment (5)

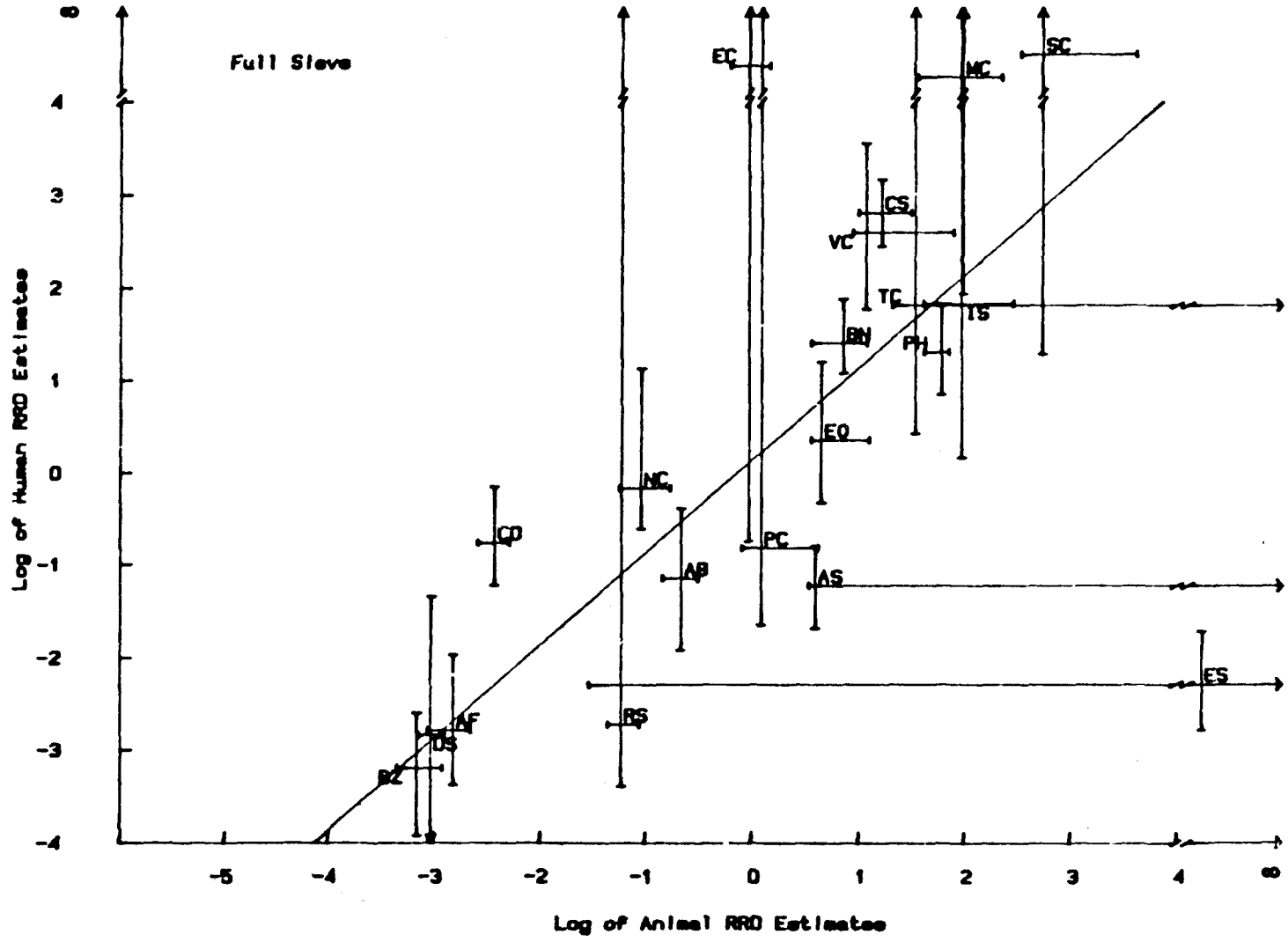


Figure 2-14

Correlation Analysis:
Malignant Tumors Only (7)

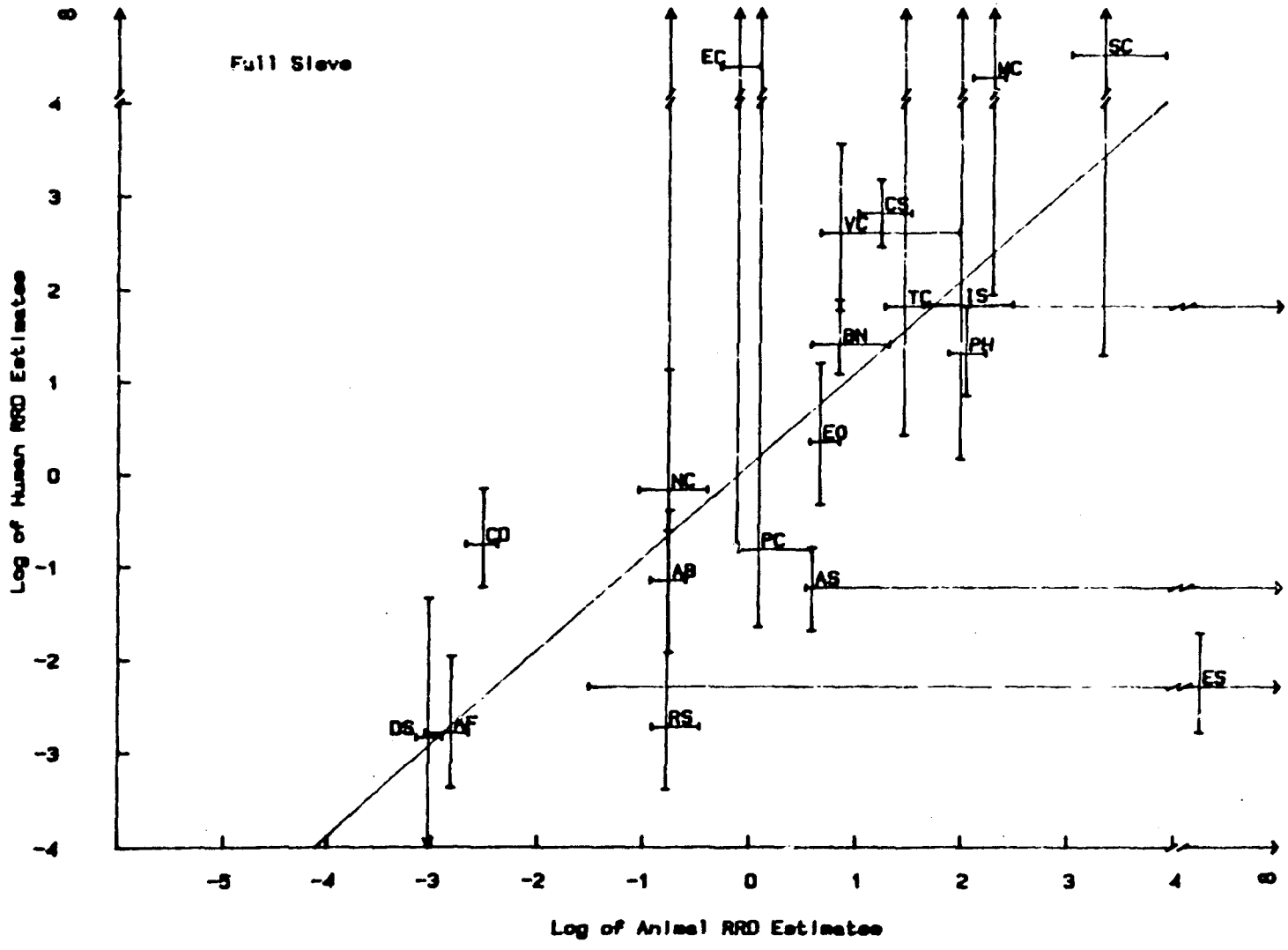


Figure 2-15

Correlation Analysis:
Combination of Significant Responses (Sa)

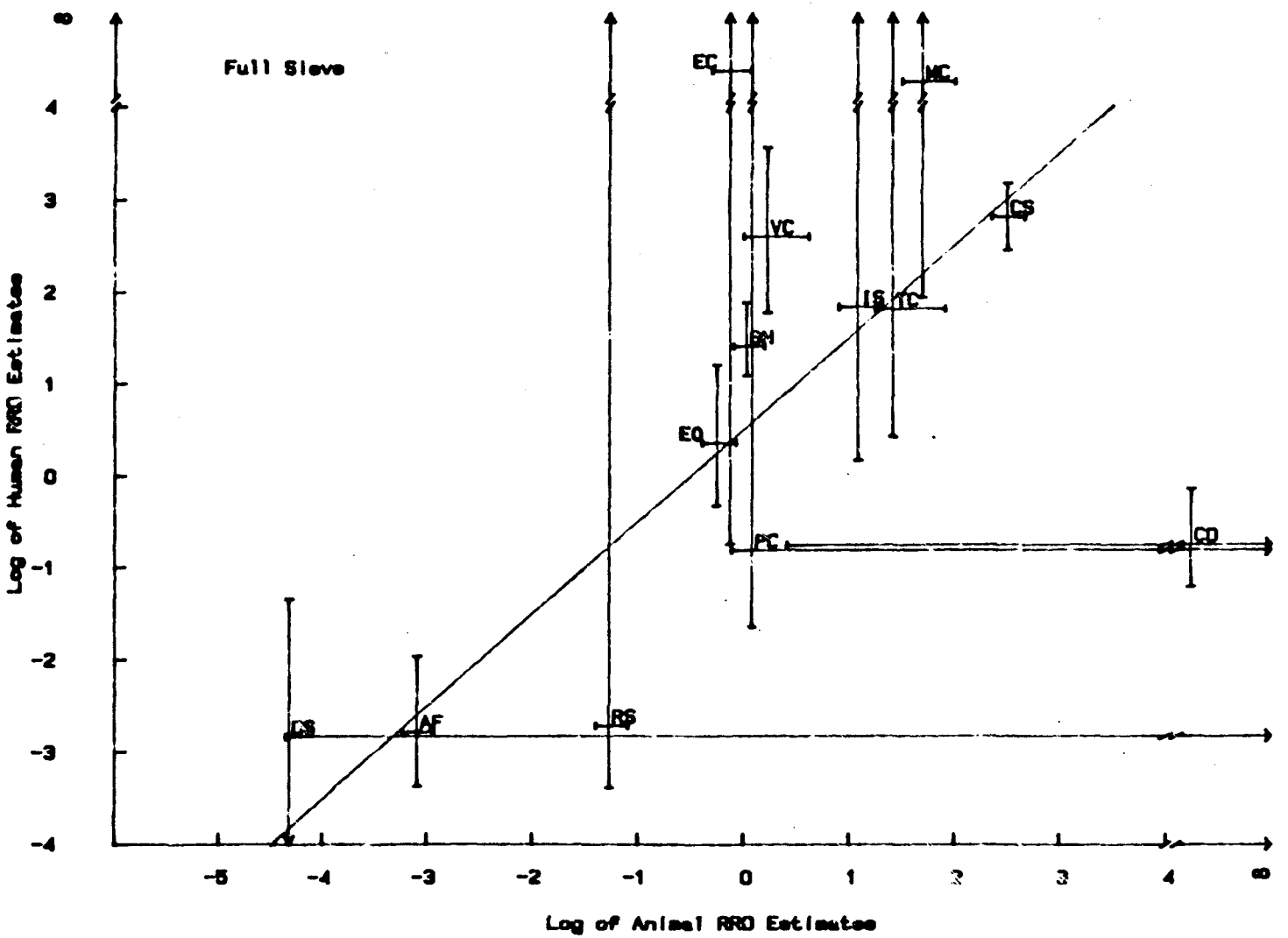


Figure 2-16

Correlation Analysis:
Total Tumor-Bearing Animals (8b)

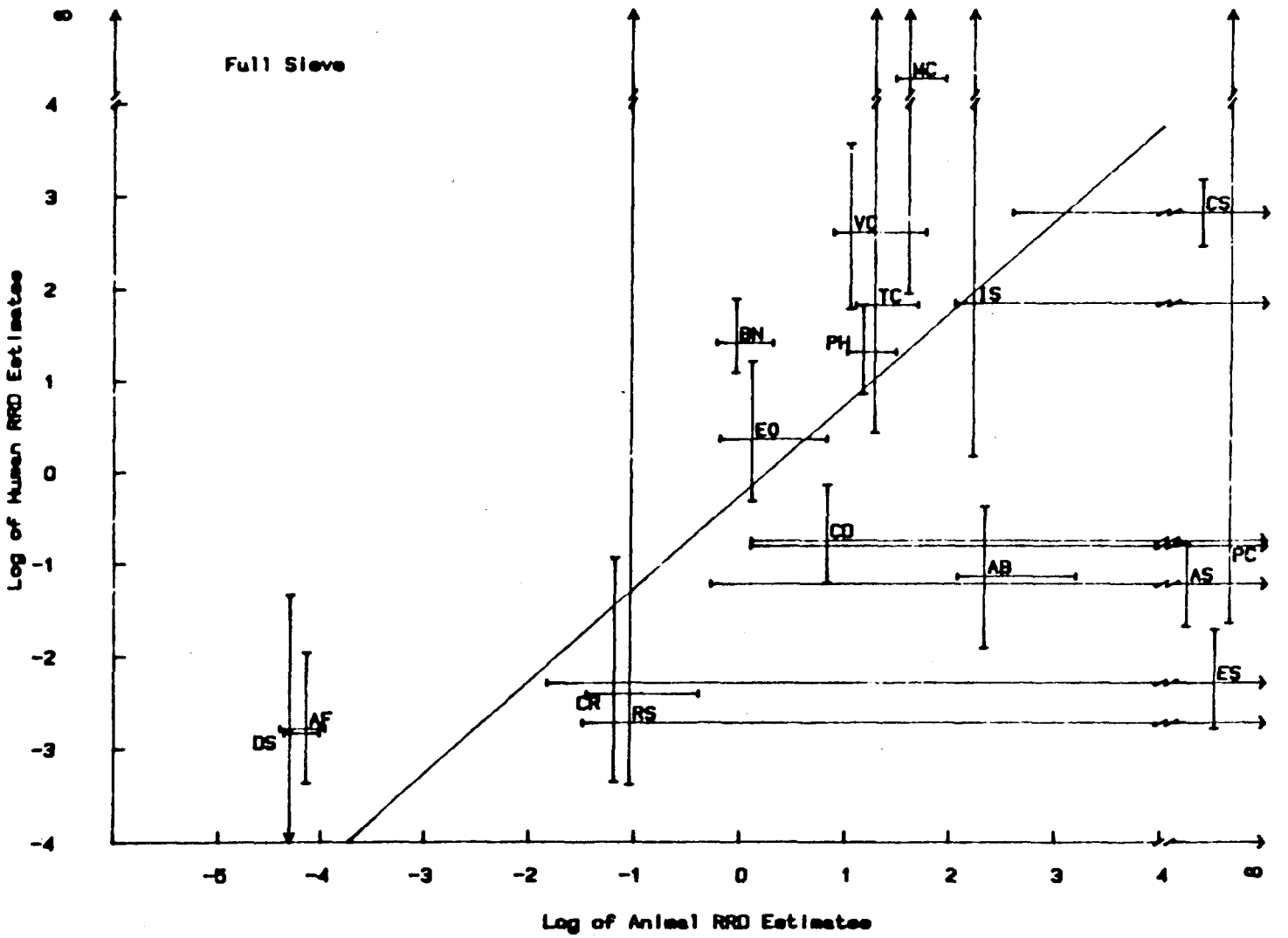


Figure 2-17

Correlation Analysis:
Response that Humans Get (Bc)

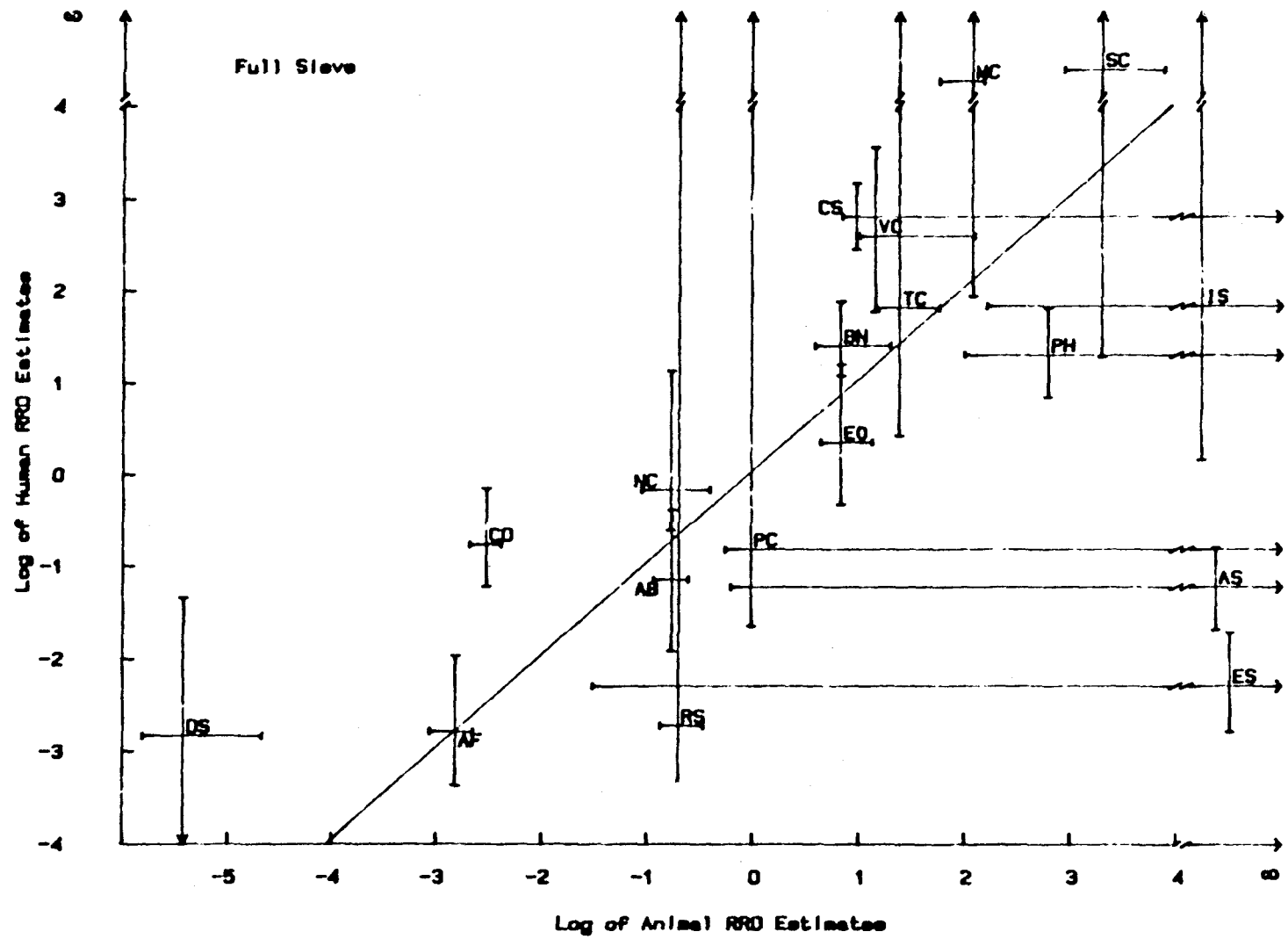


Figure 2-18

Correlation Analysis:
Average over Sex (9)

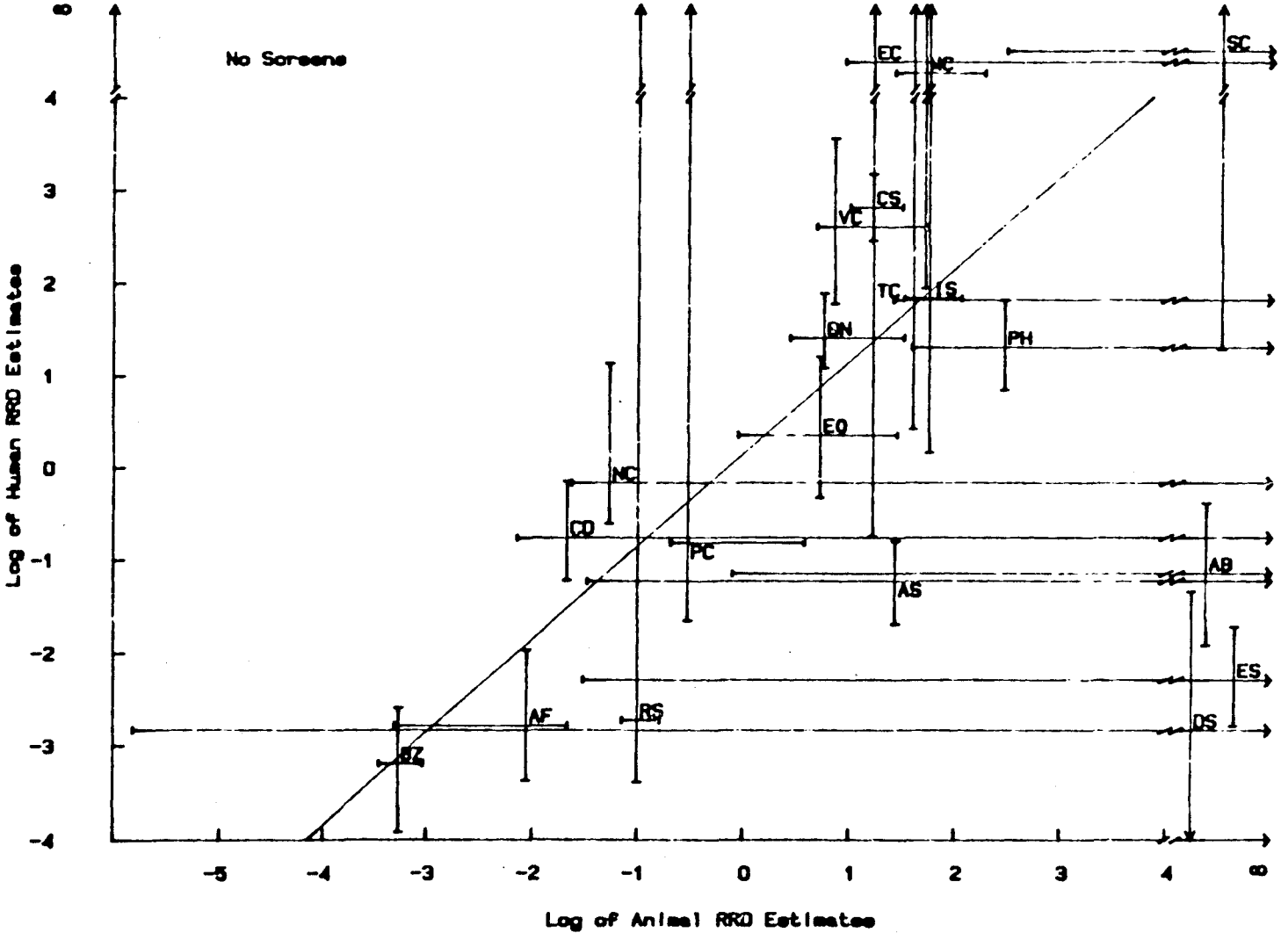


Figure 2-19

Correlation Analysis:
Average over Sex (9)

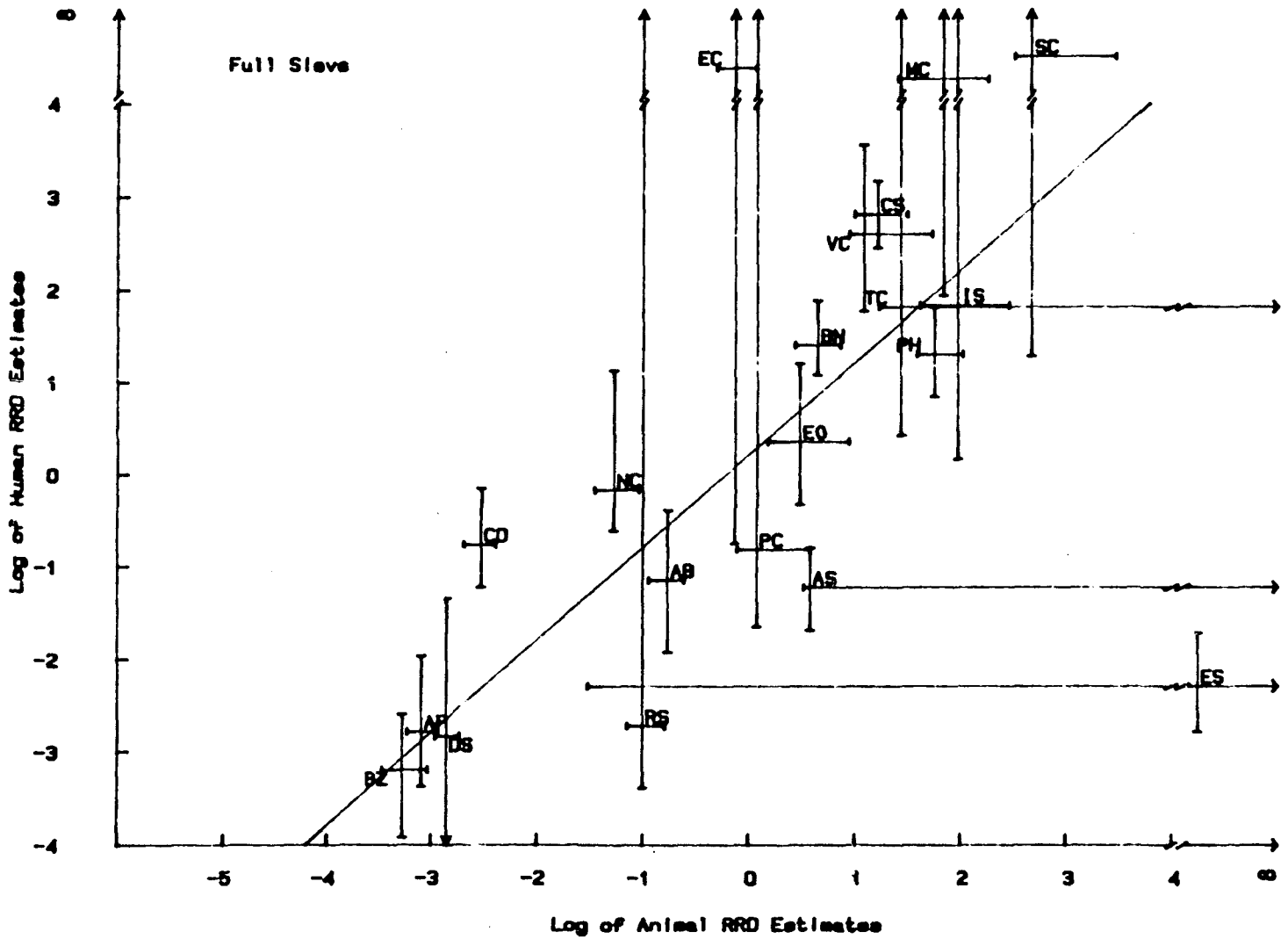


Figure 2-20

Correlation Analysis:
Average over Study (10)

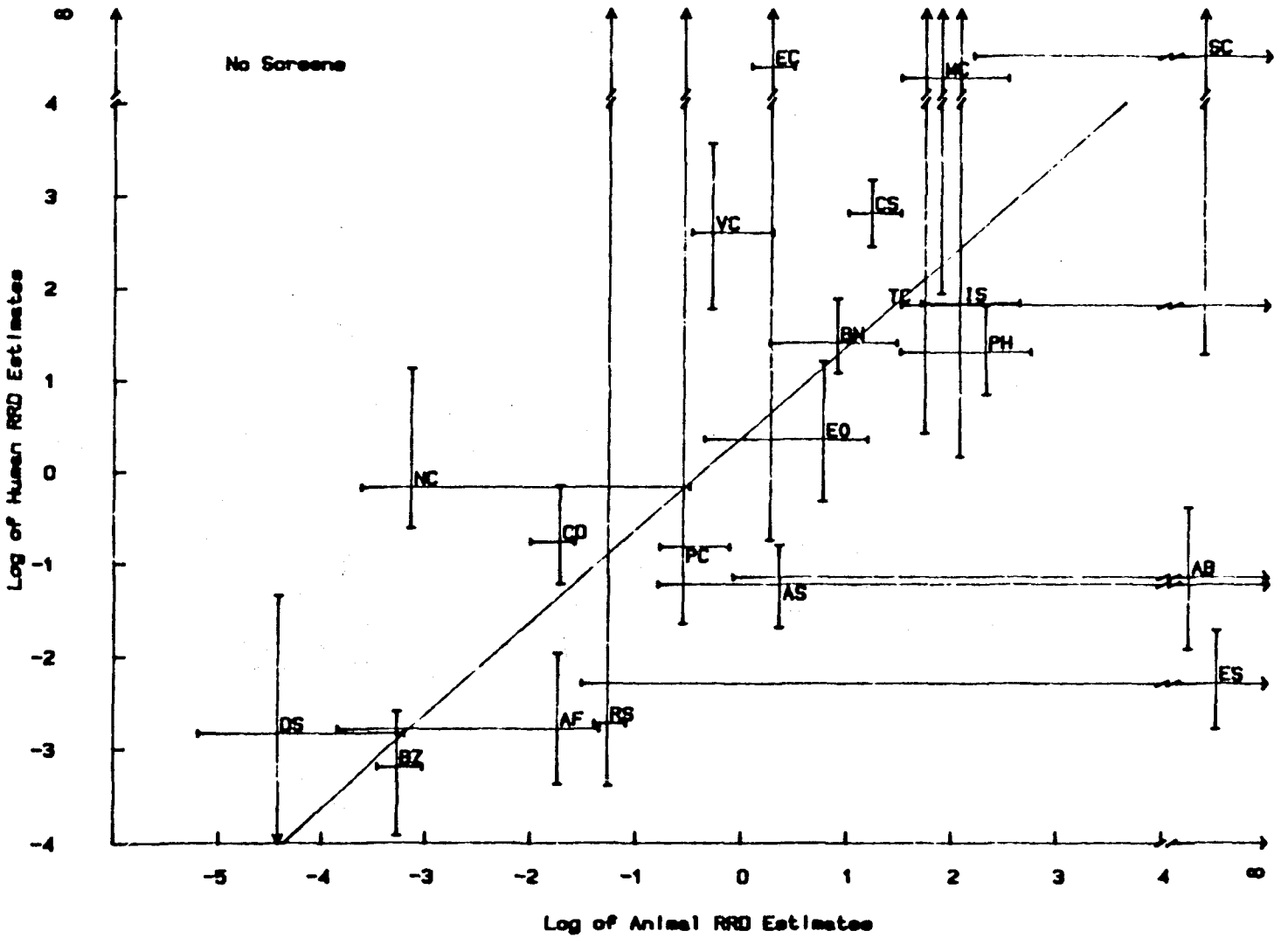


Figure 2-21

Correlation Analysis:
Average over Study (10)

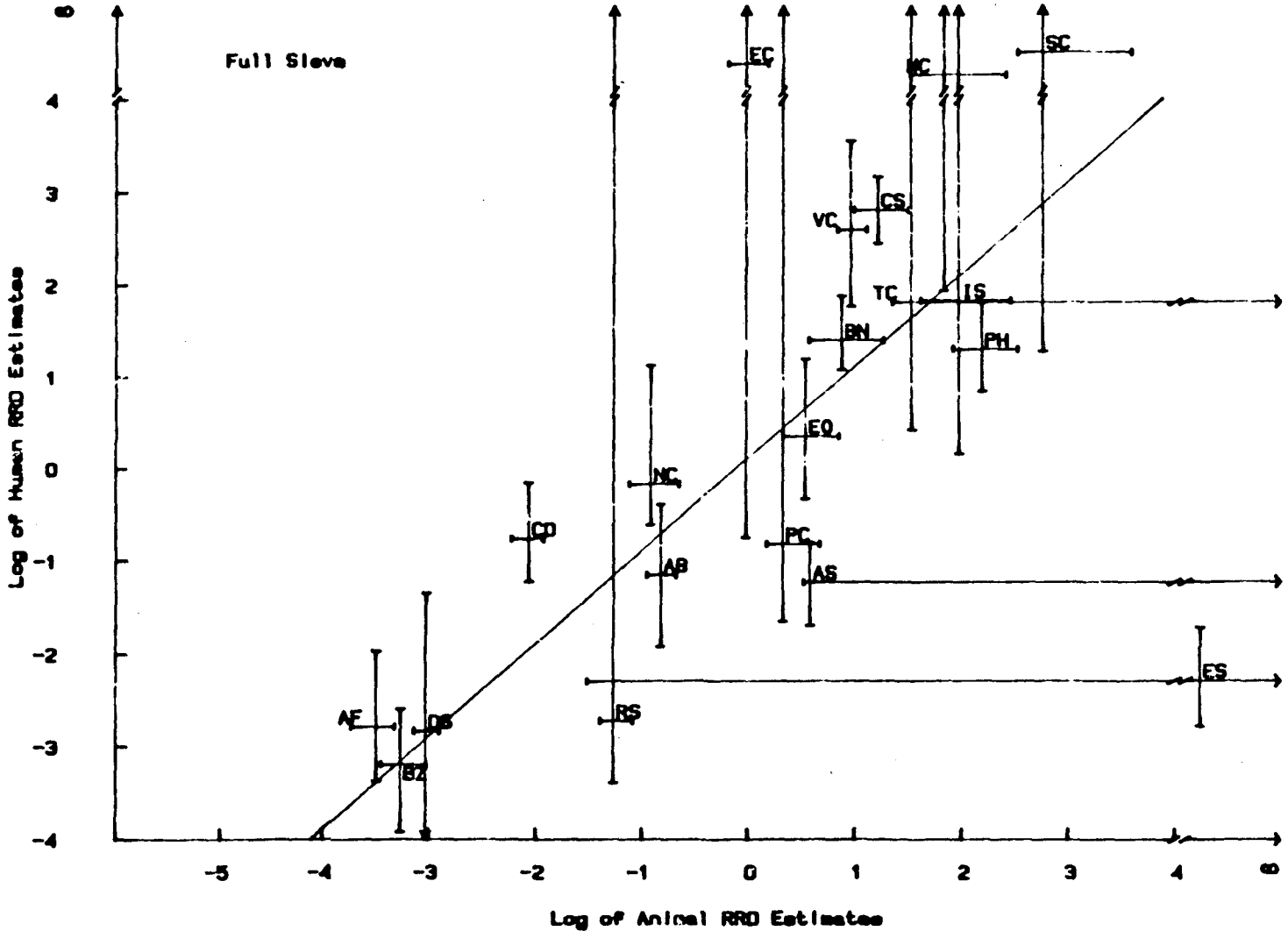


Figure 2-22

Correlation Analysis:
Average over All Species (11a)

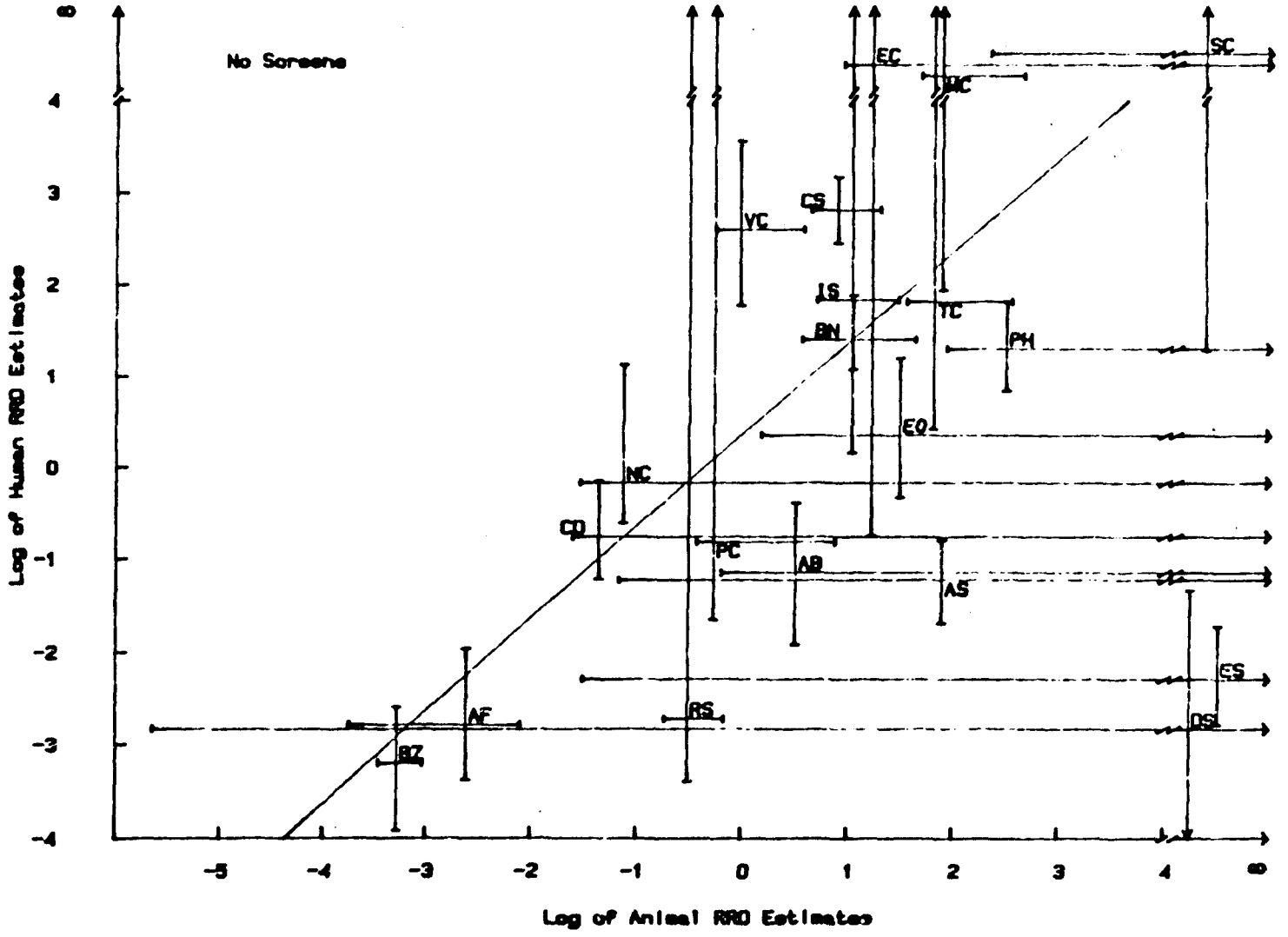


Figure 2-23

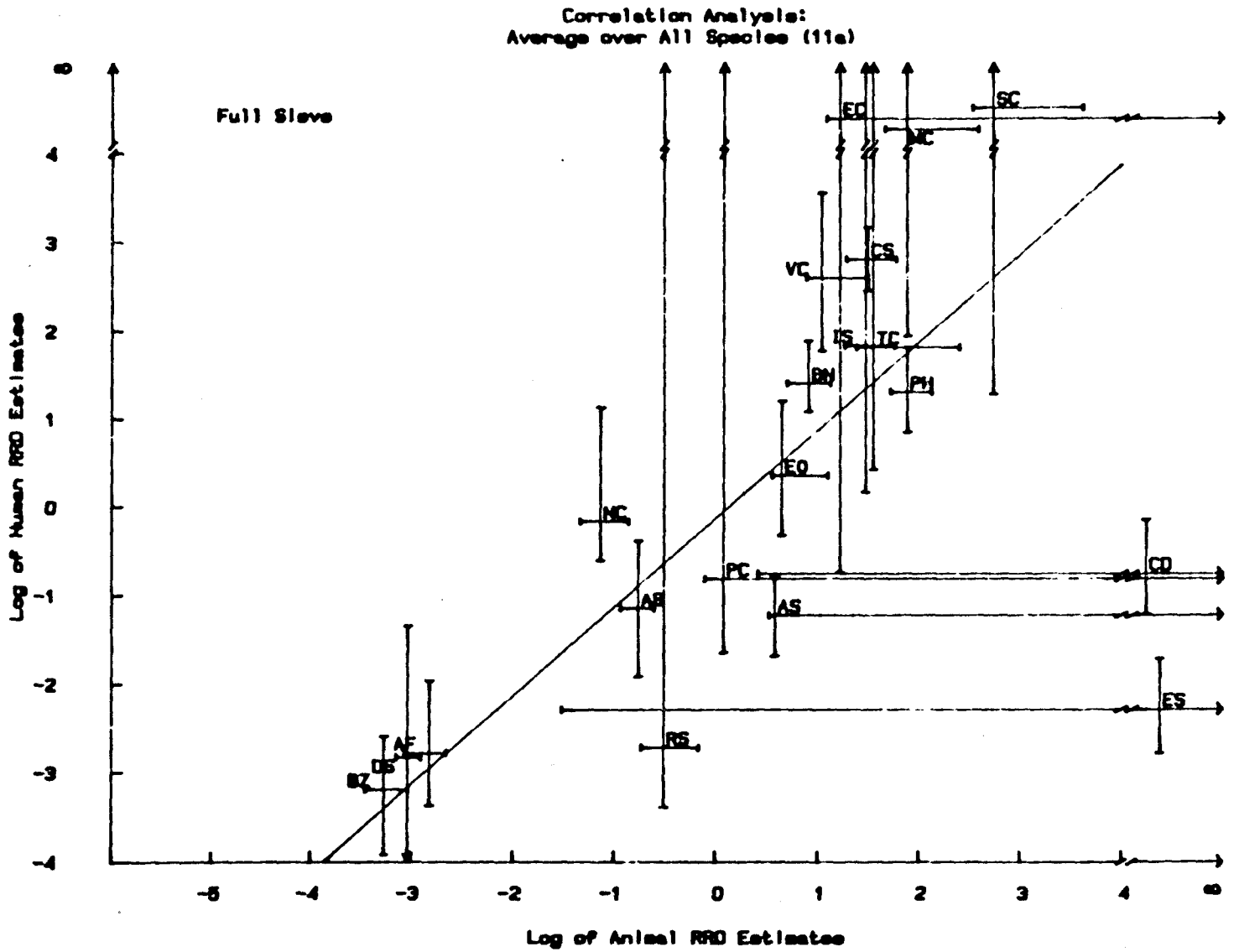

Correlation Analysis:
Average over All Species (11a)
Full Sieve
Log of Human RRD Estimates
Log of Animal RRD Estimates
ED
SC
MC
CS
VC
IS
TC
BN
PH
EO
NC
PC
CD
AF
AS
ES
RS
AF
DS
BZ

Figure 2-24

Correlation Analysis:
Average over Rats and Mice (11b)

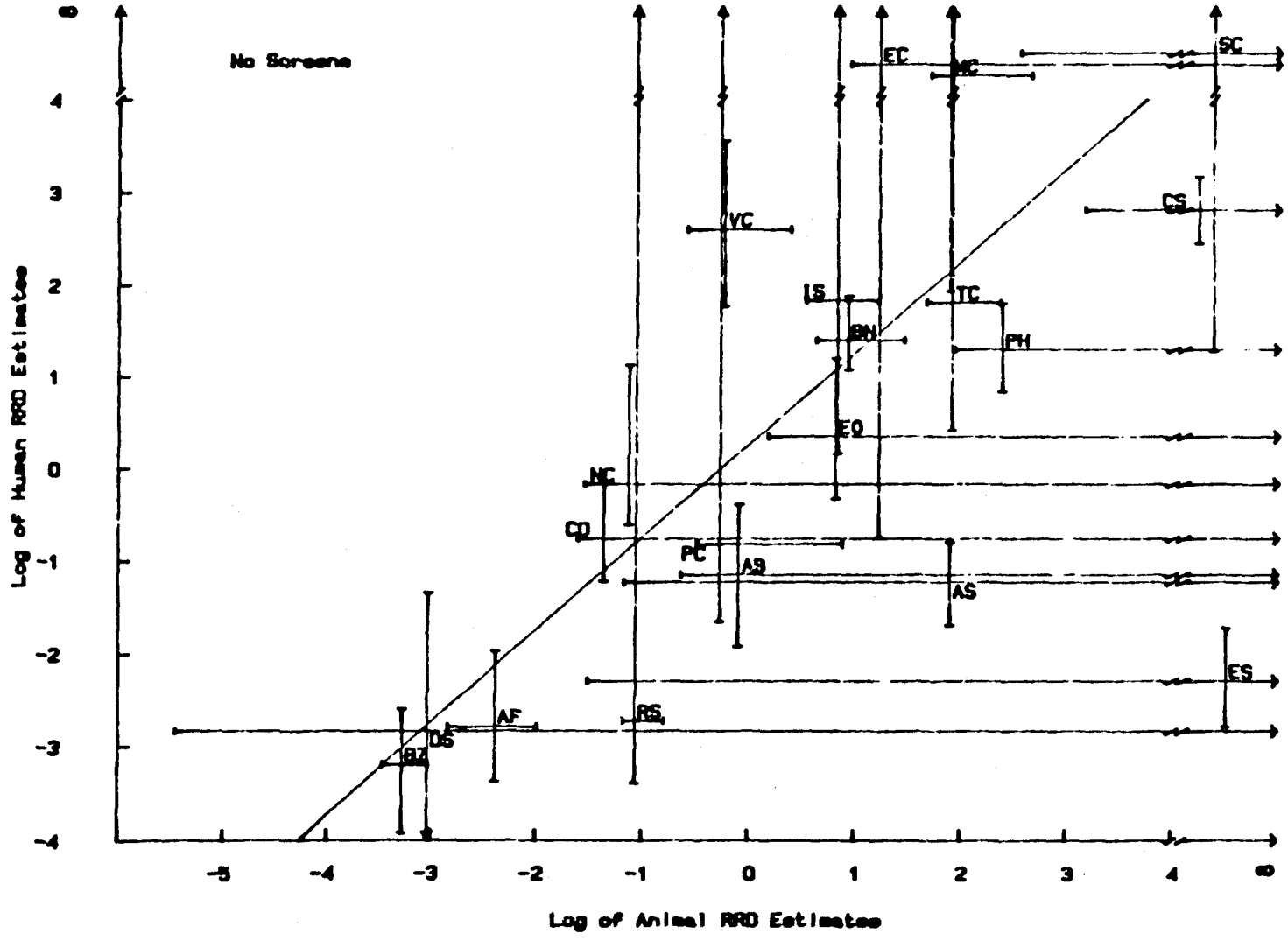


Figure 2-25

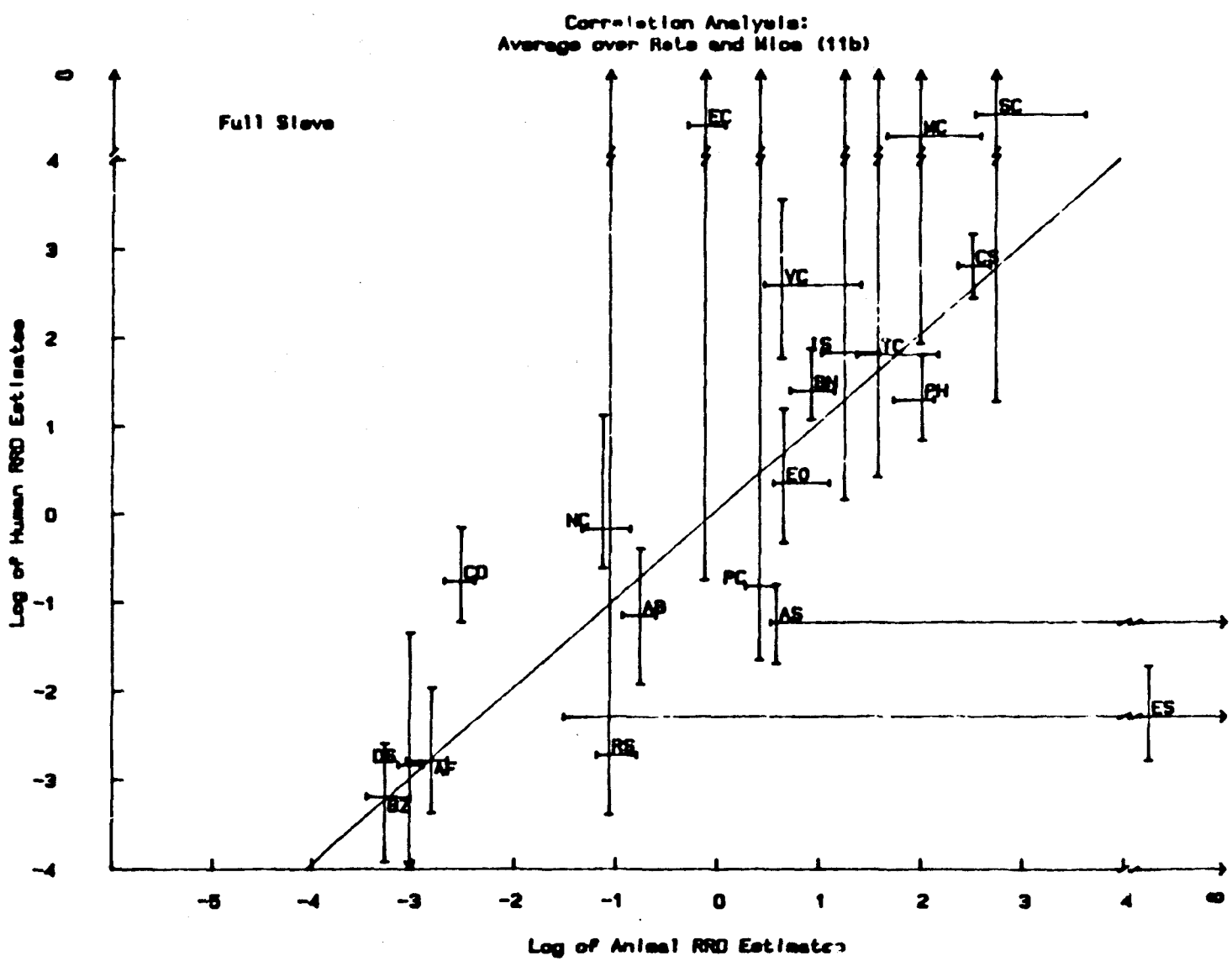

Correlation Analysis:
Average over Rats and Mice (11b)
Full Sieve
Log of Human RRD Estimates
Log of Animal RRD Estimates
EC
MC
SC
VC
IS
TC
BN
PH
EO
NC
CD
PC
AB
AS
ES
BS
DE
AF
BZ
4
3
2
1
0
-1
-2
-3
-4
-5
-4
-3
-2
-1
0
1
2
3
4

Figure 7 26

Correlation Analysis:
Rat Data Only (11c)

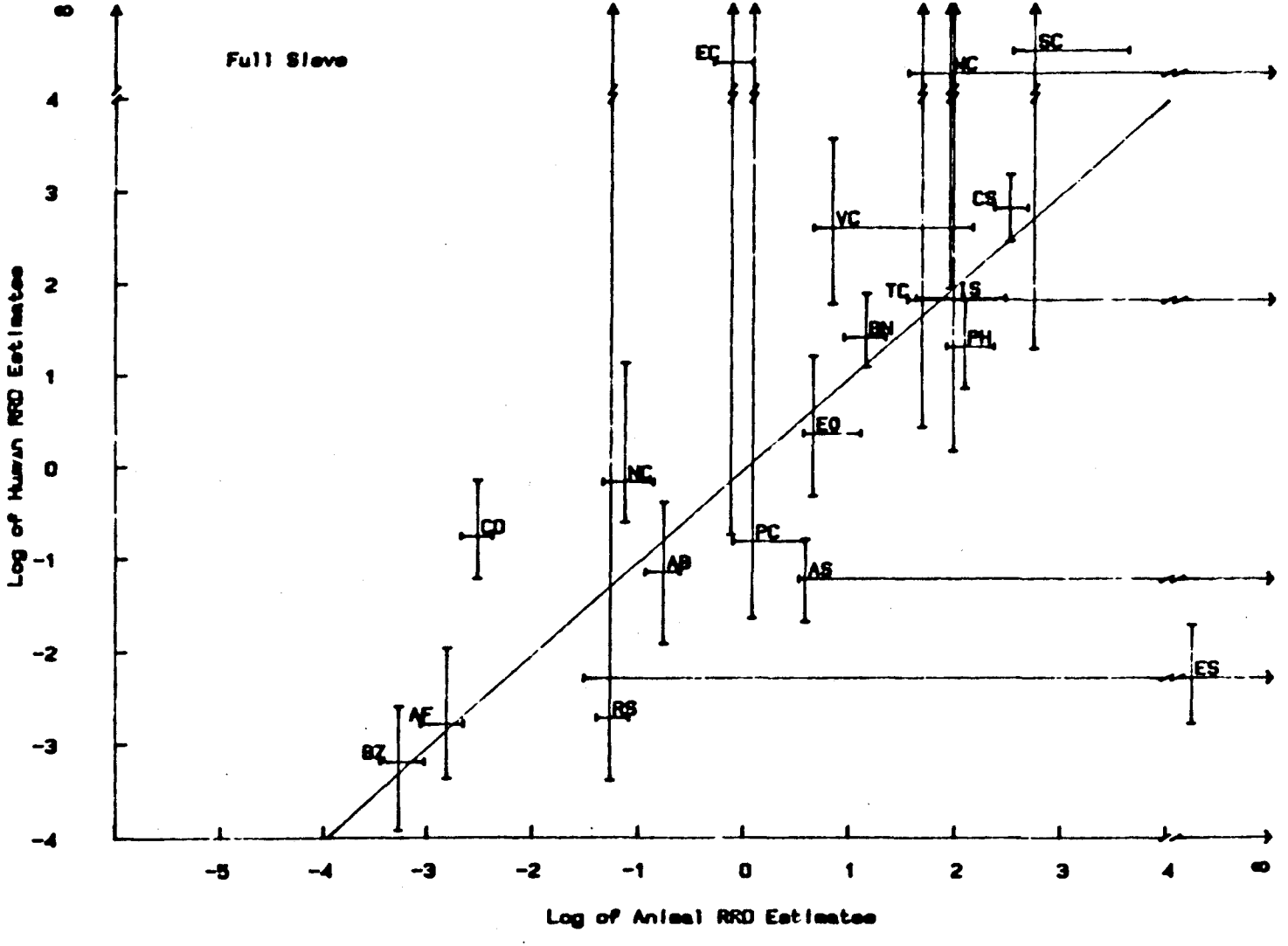


Figure 2-27

Correlation Analysis:
Mouse Data Only (11d)

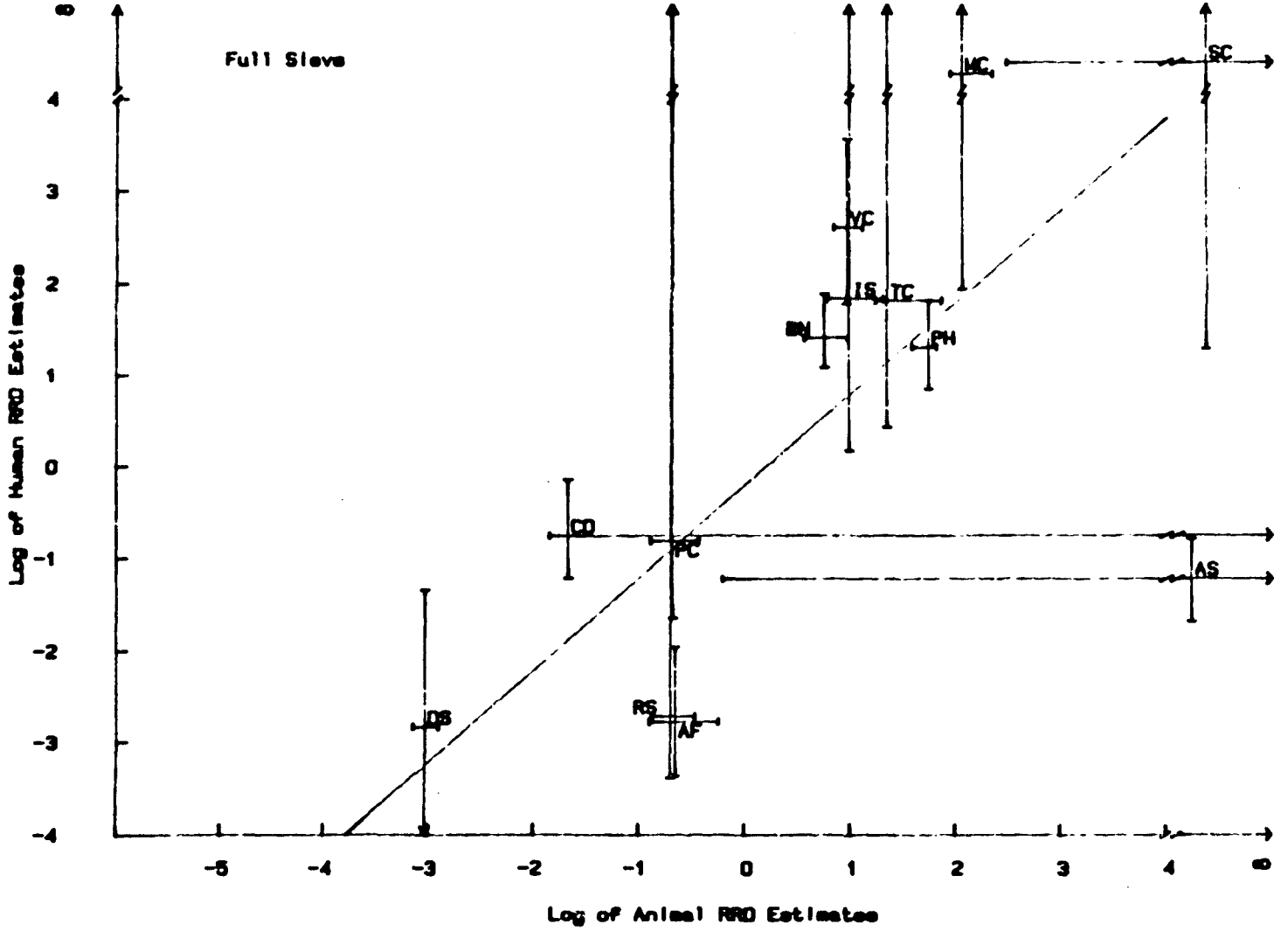


Figure 2-28

Correlation Analysis:
Average over Sex, Study, and Species (12)

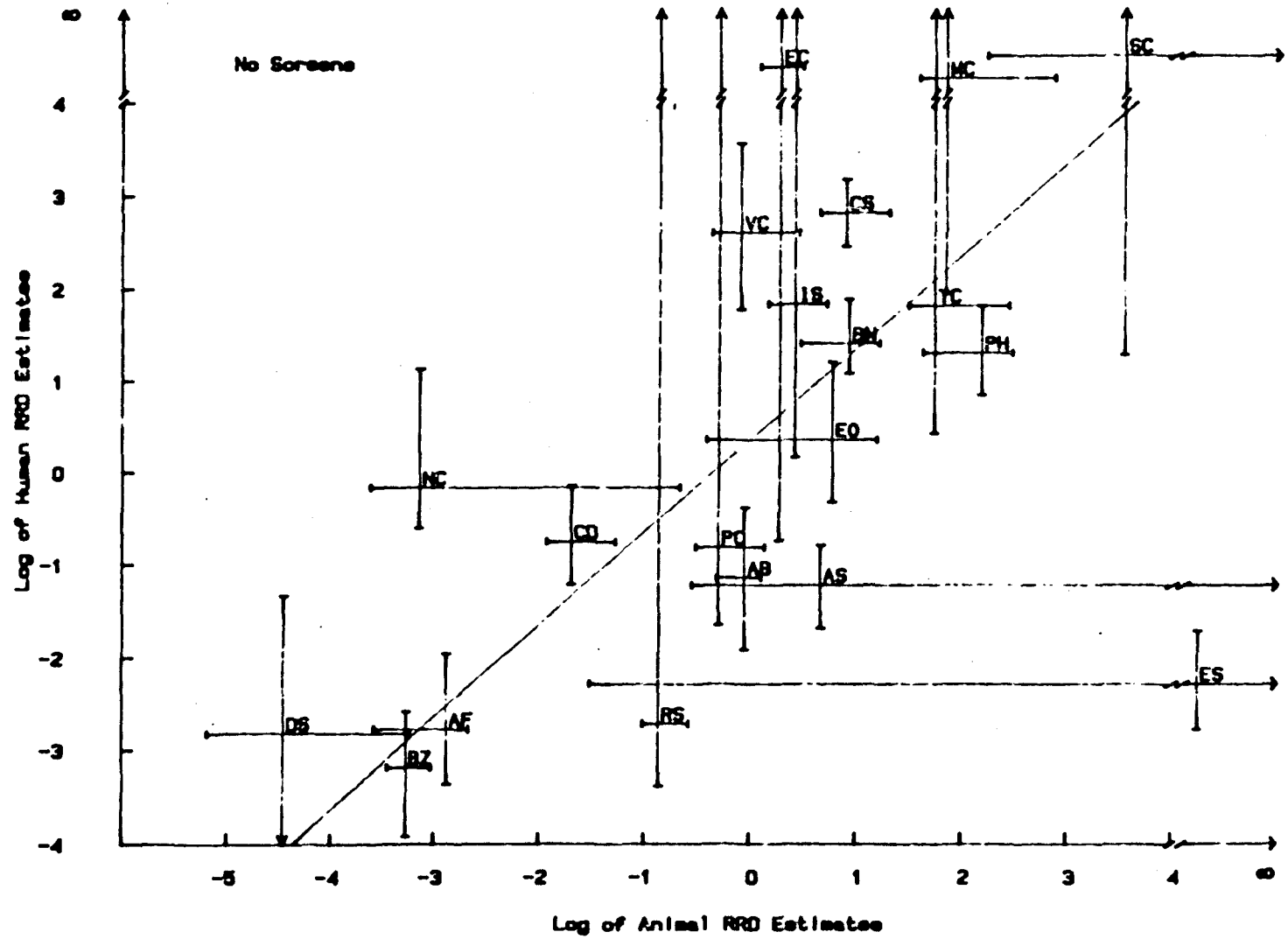


Figure 2-28

Correlation Analysis:
Average over Sex, Study, and Species (12)

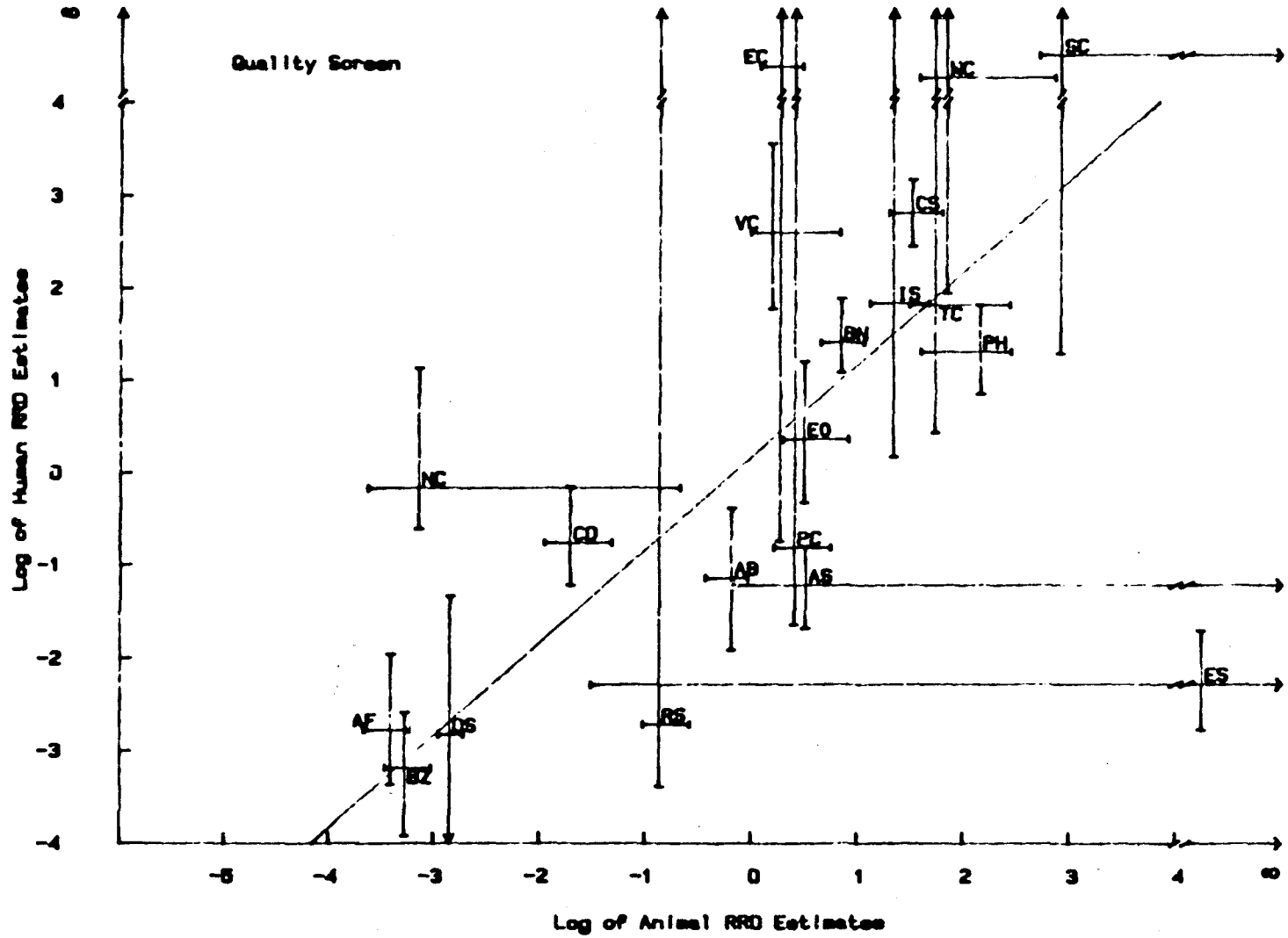


Figure 2-30

Correlation Analysis:
Average over Sex, Study, and Species (12)

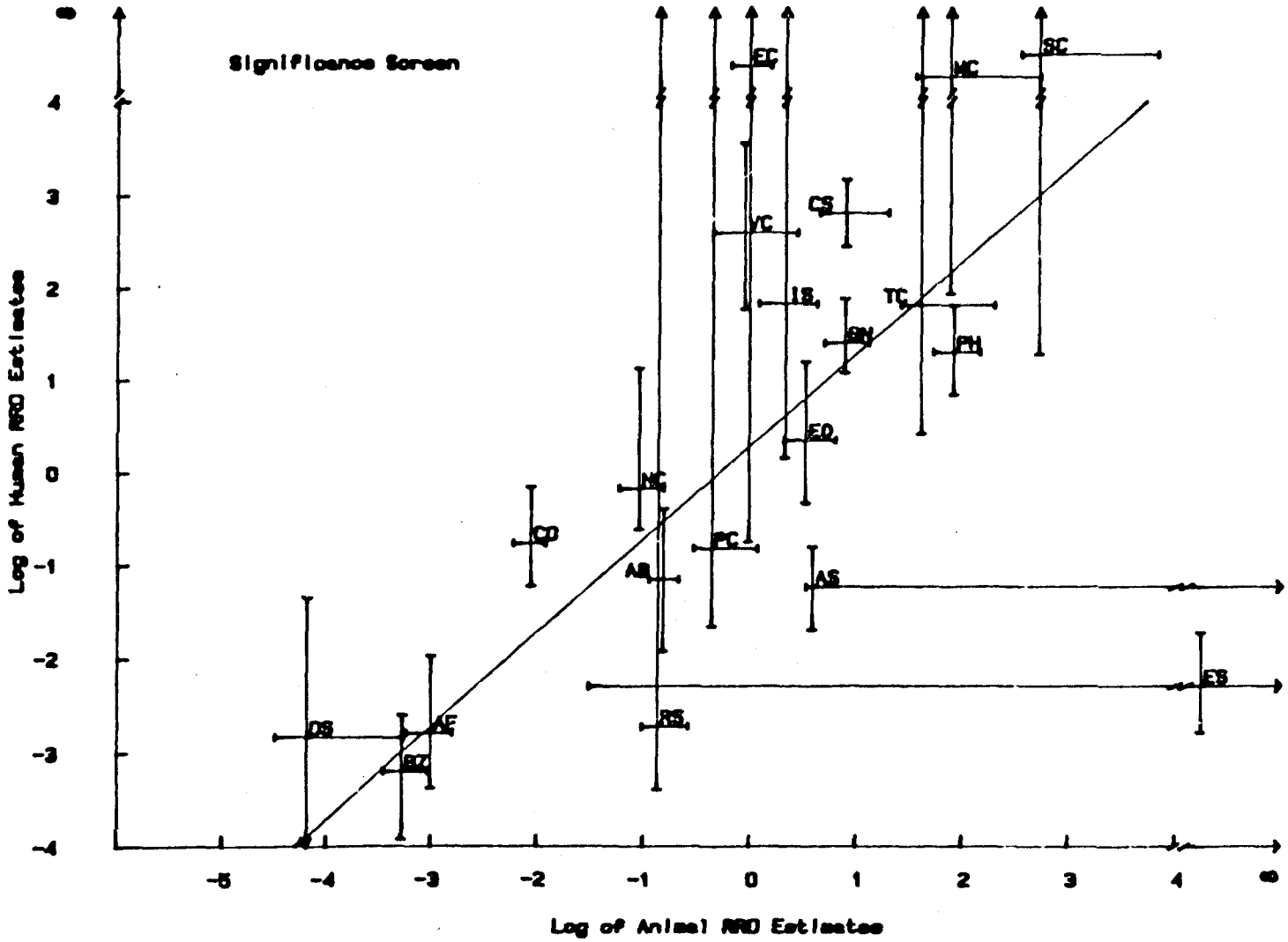


Figure 2-31

Correlation Analysis:
Average over Sex, Study, and Species (12)

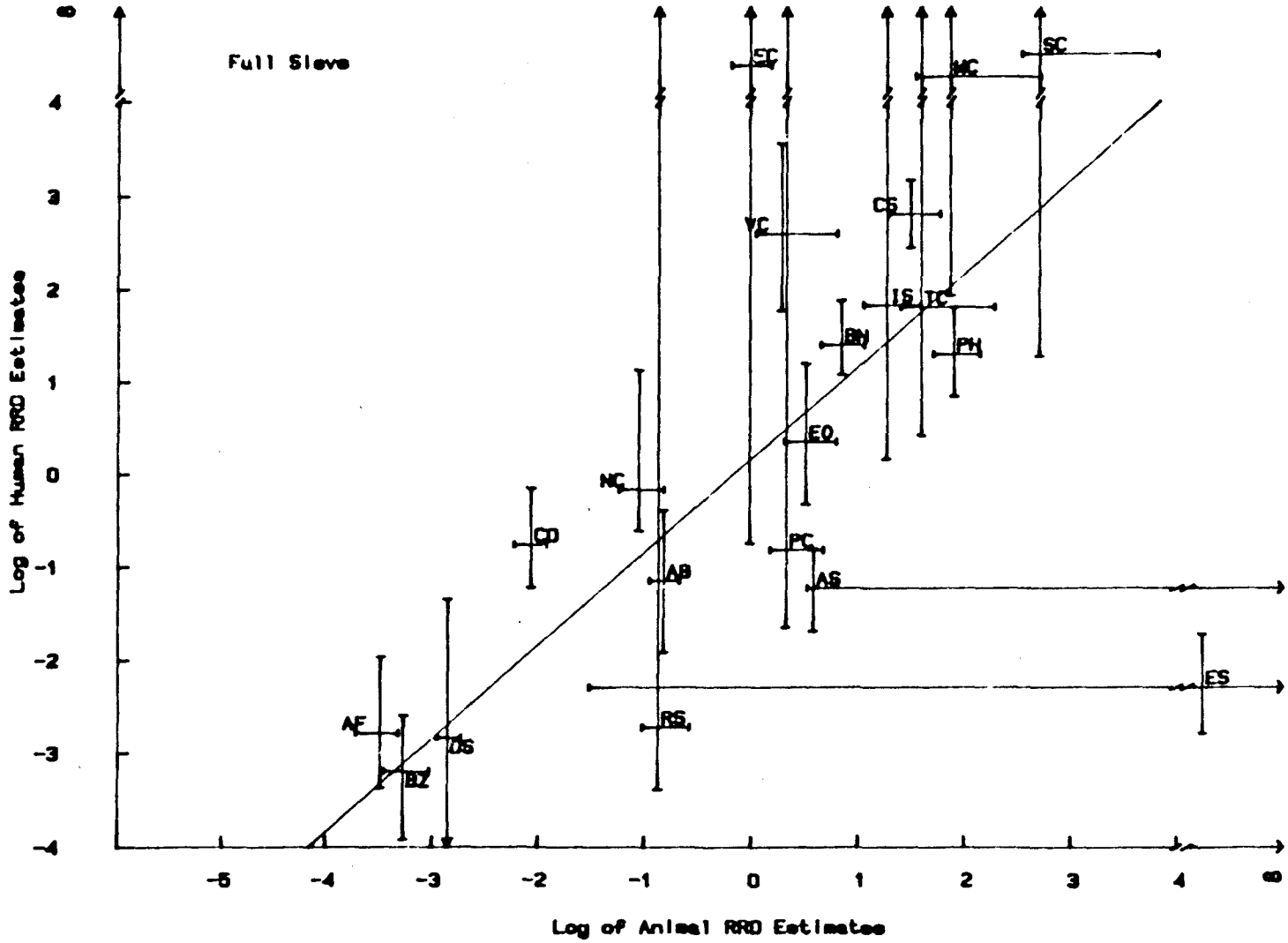


Figure 2-32

Correlation Analysis:
Average over All; Combination of Significant Responses (16)

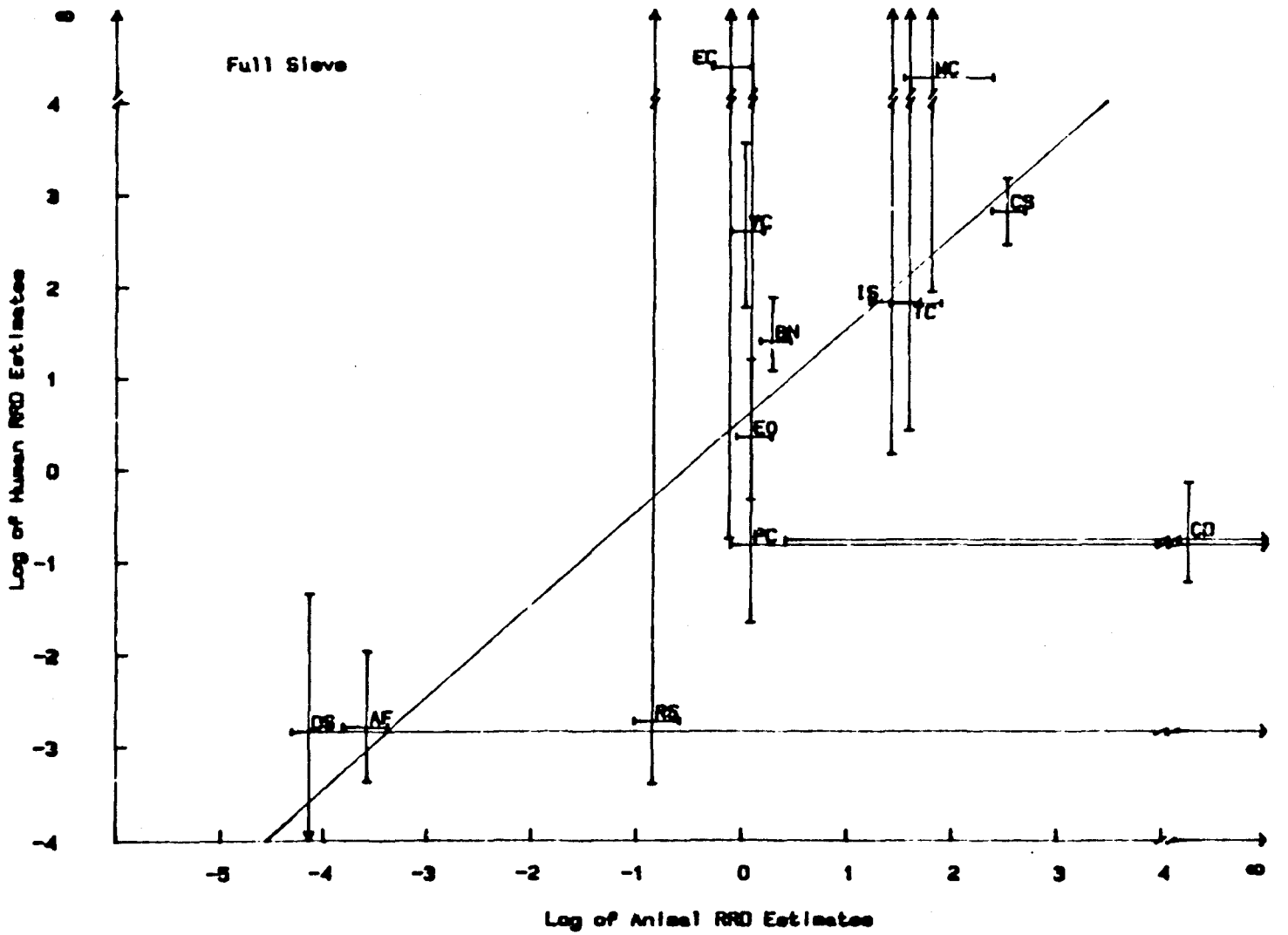


Figure 2-33

Correlation Analysis:
Average over All; Total Tumor-Bearing Animals (20)

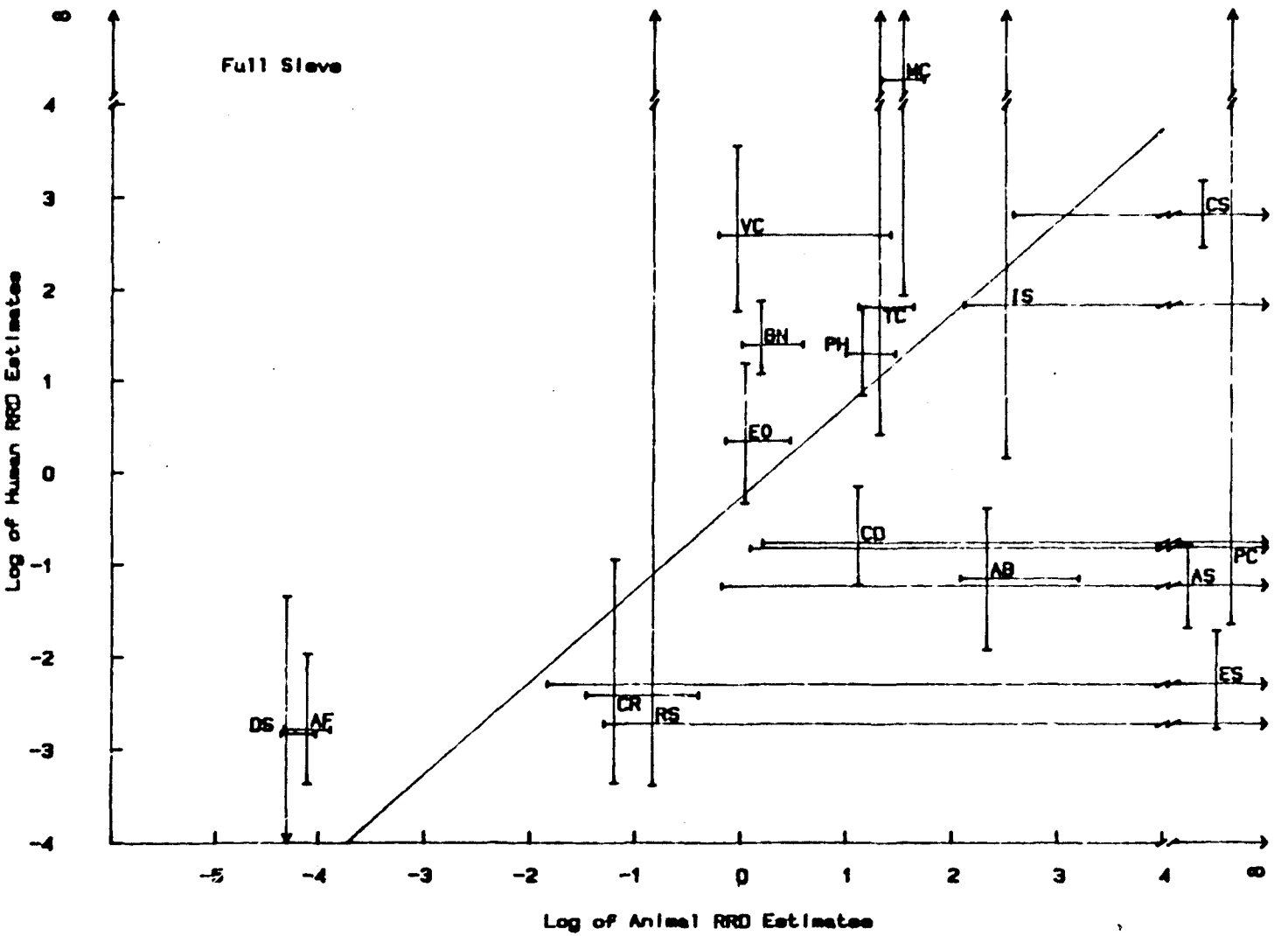


Figure 2-34

Correlation Analysis:
Route and Response Like Humans (25)

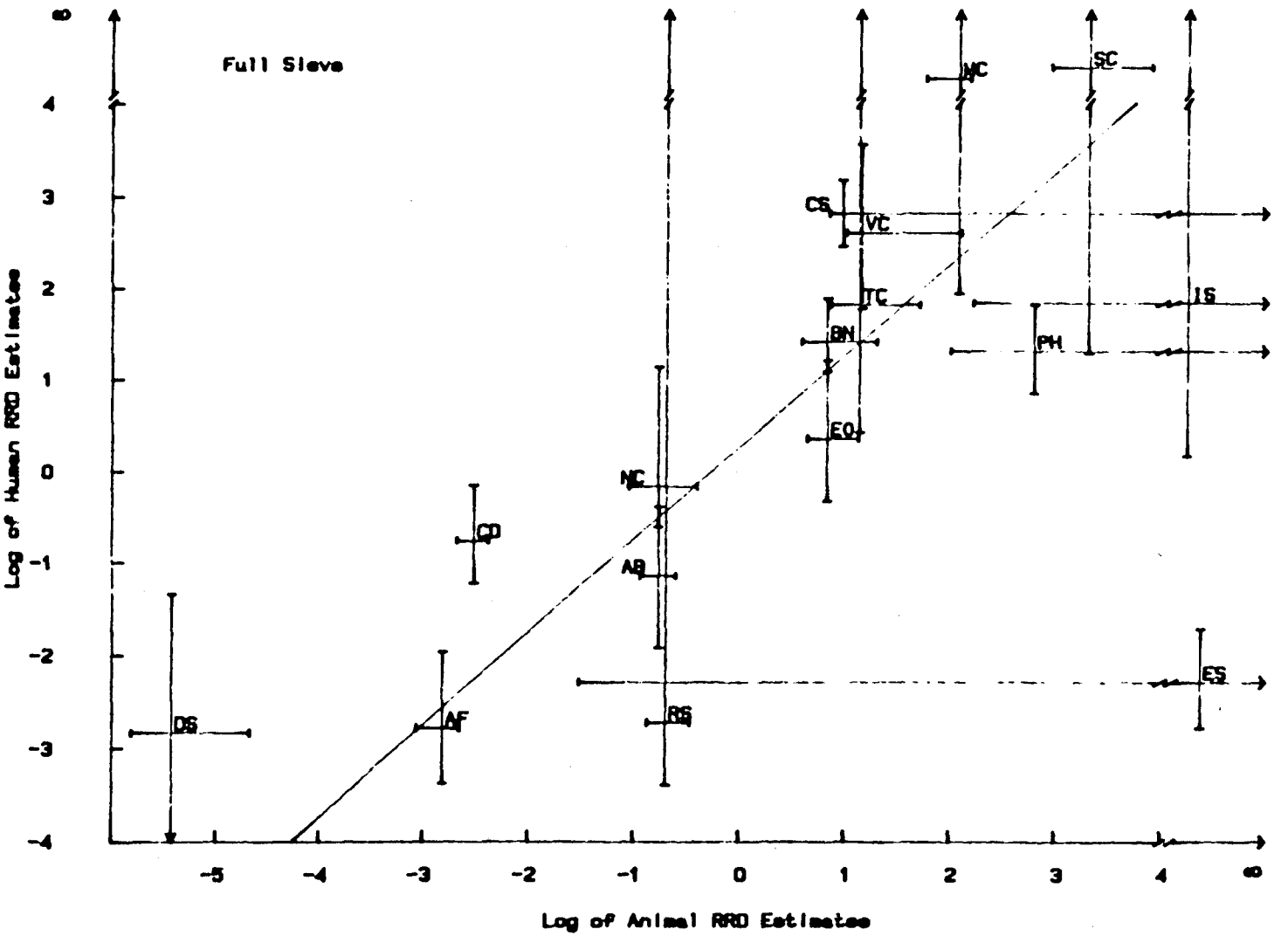


Figure 2-36

Prediction Analysis:
Analysis 17, Median Lower Bound Predictor

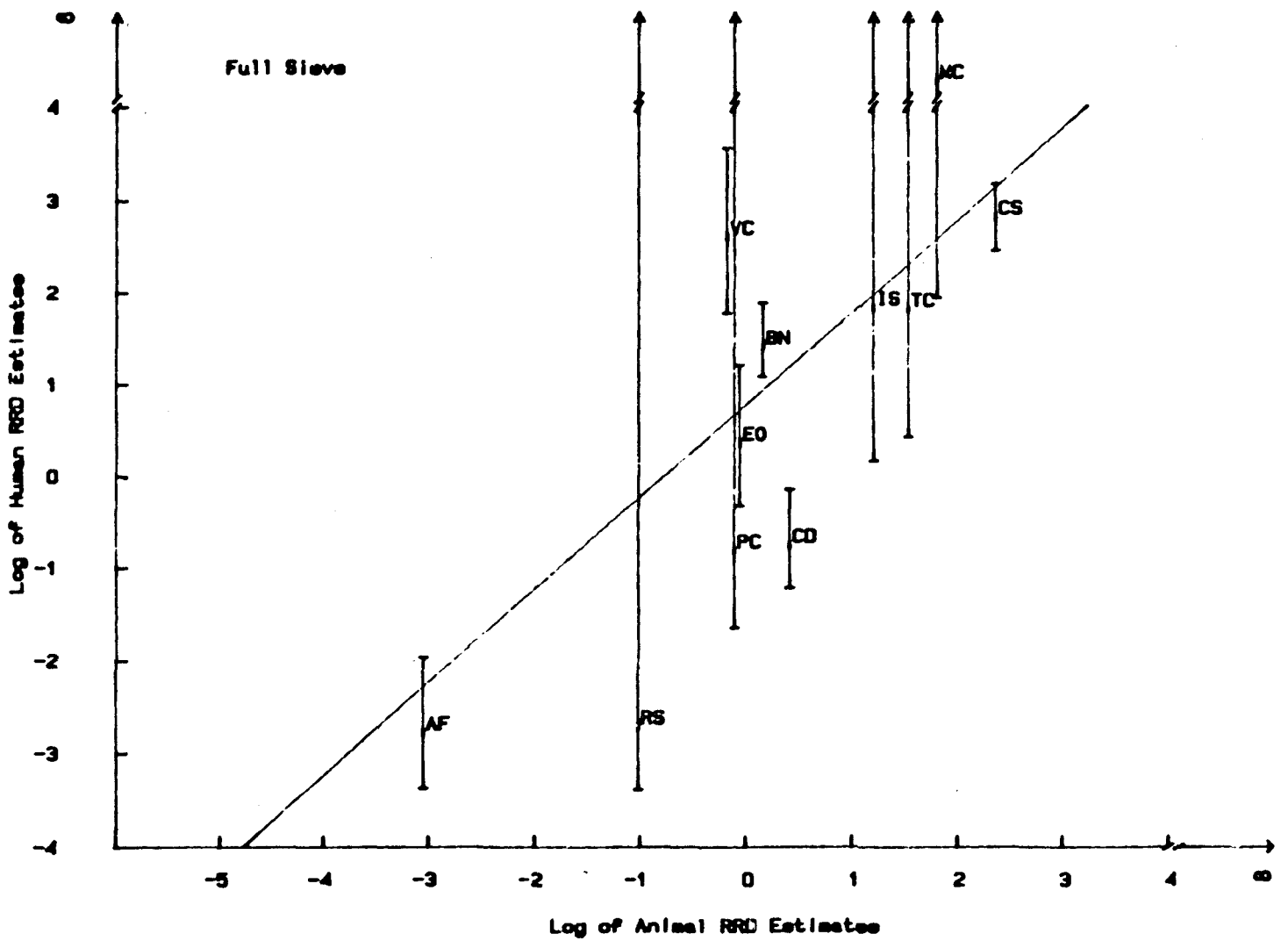


Figure 2-36

Prediction Analysis:
Analysis 17, Median Lower Bound Predictor

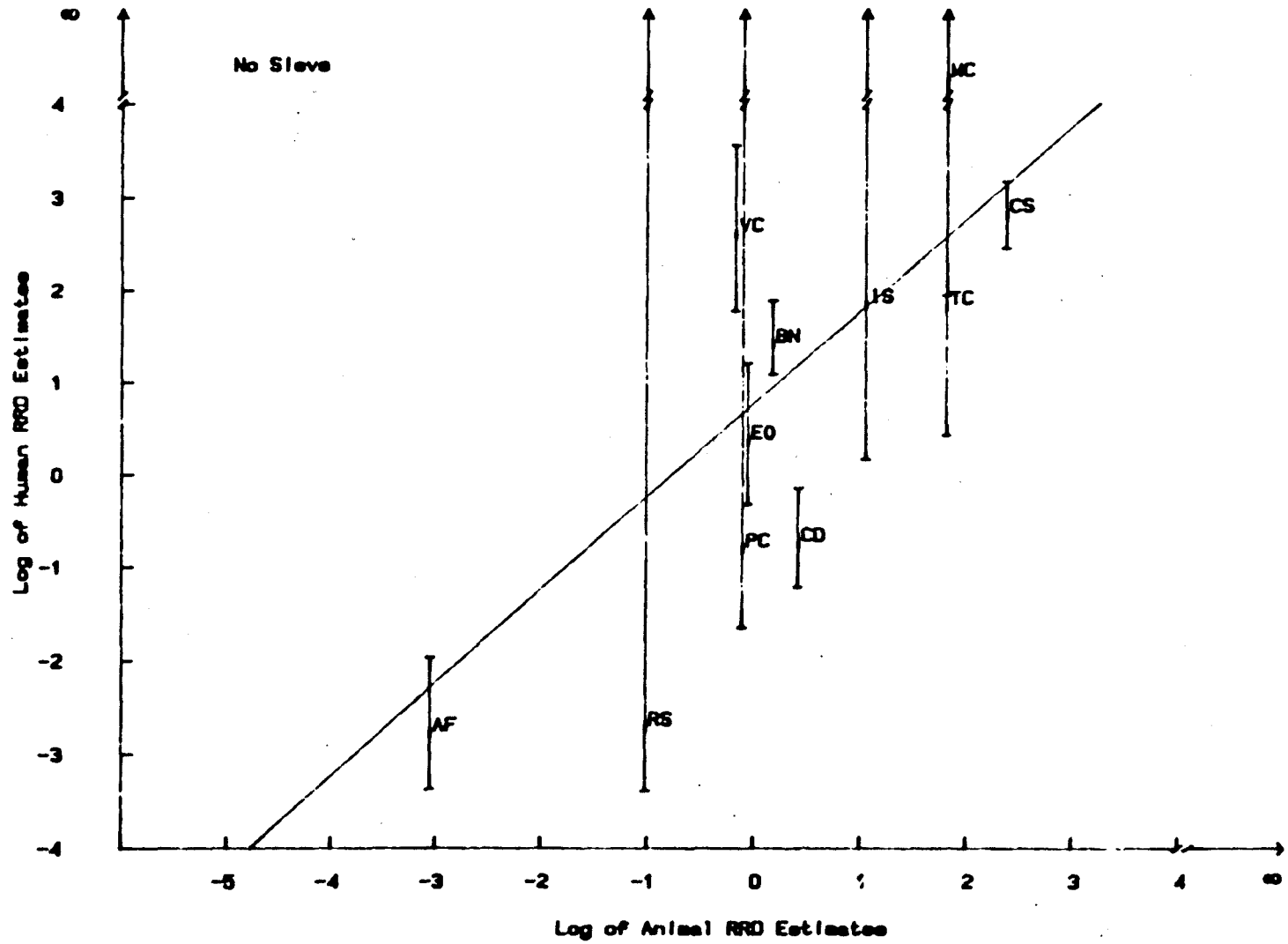


Figure 2-37

Predictor Analysis:
Analysis 3b, Median Lower Bound Predictor

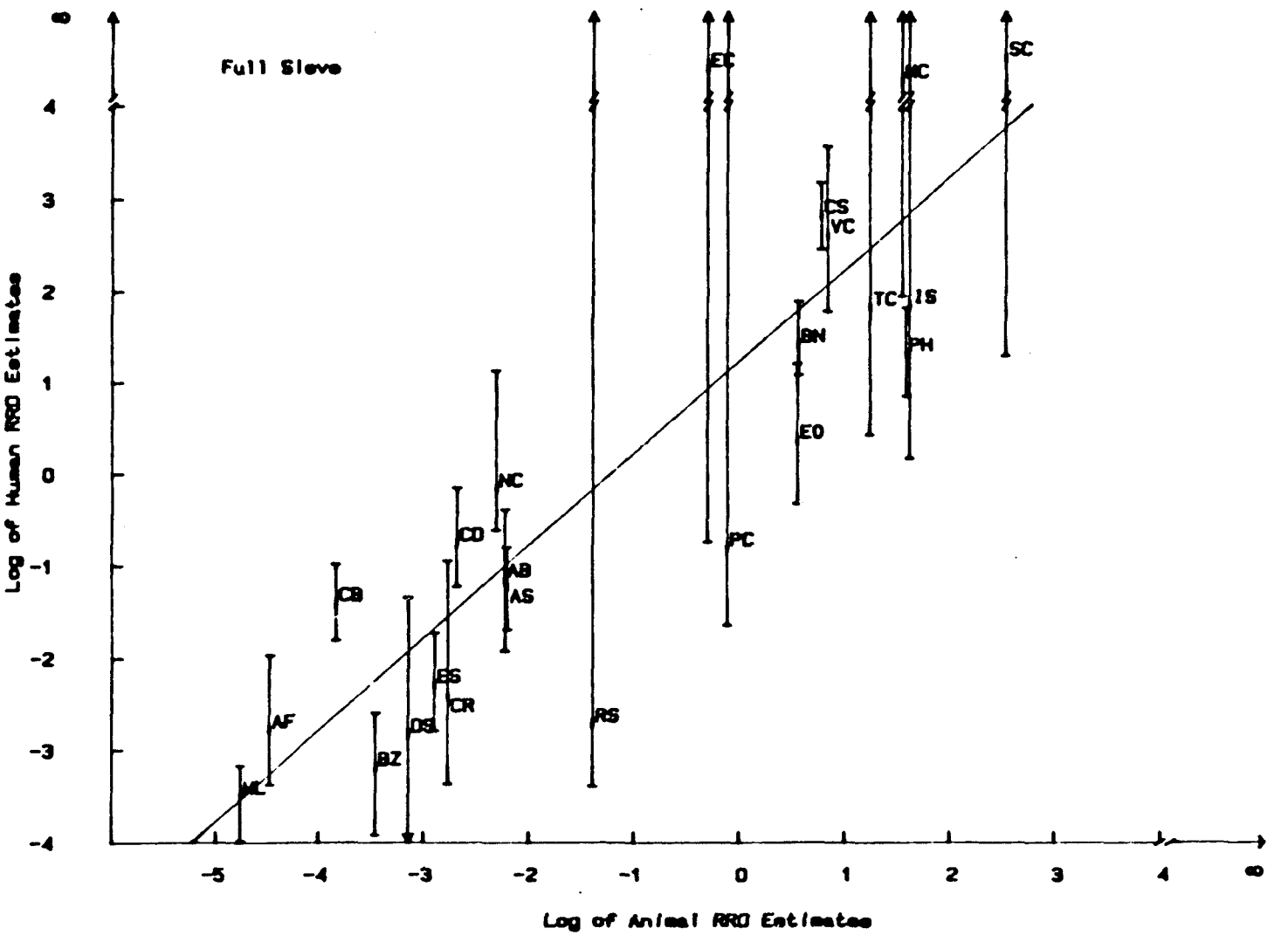


Figure 2-38

Prediction Analysis:
Analysis 20, Median Lower Bound Predictor

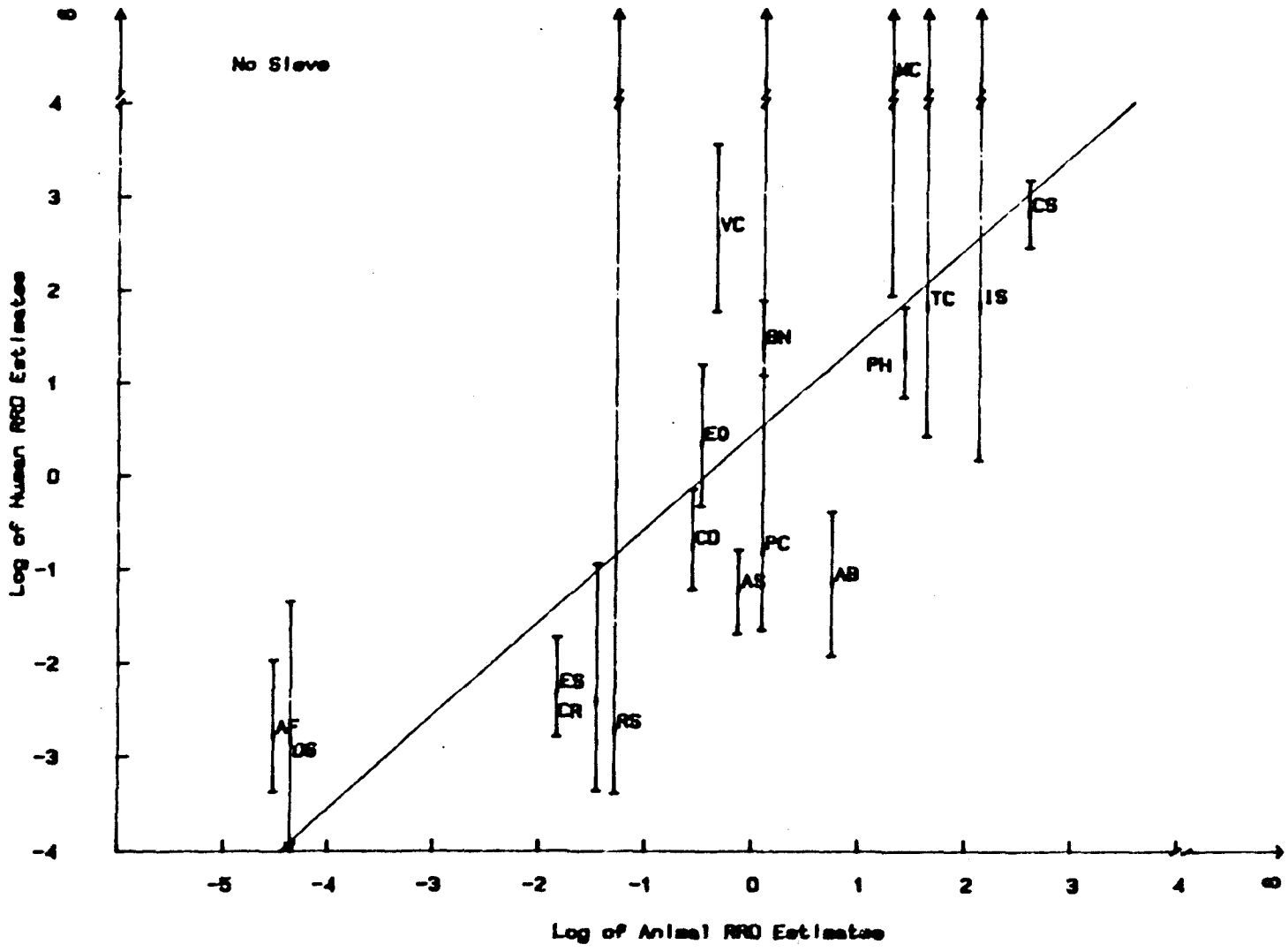


Figure 2-38

Prediction Analysis:
Analysis 3b, Median Lower Bound; Best-Fitting Lines with Increasing Degrees of Asymmetry

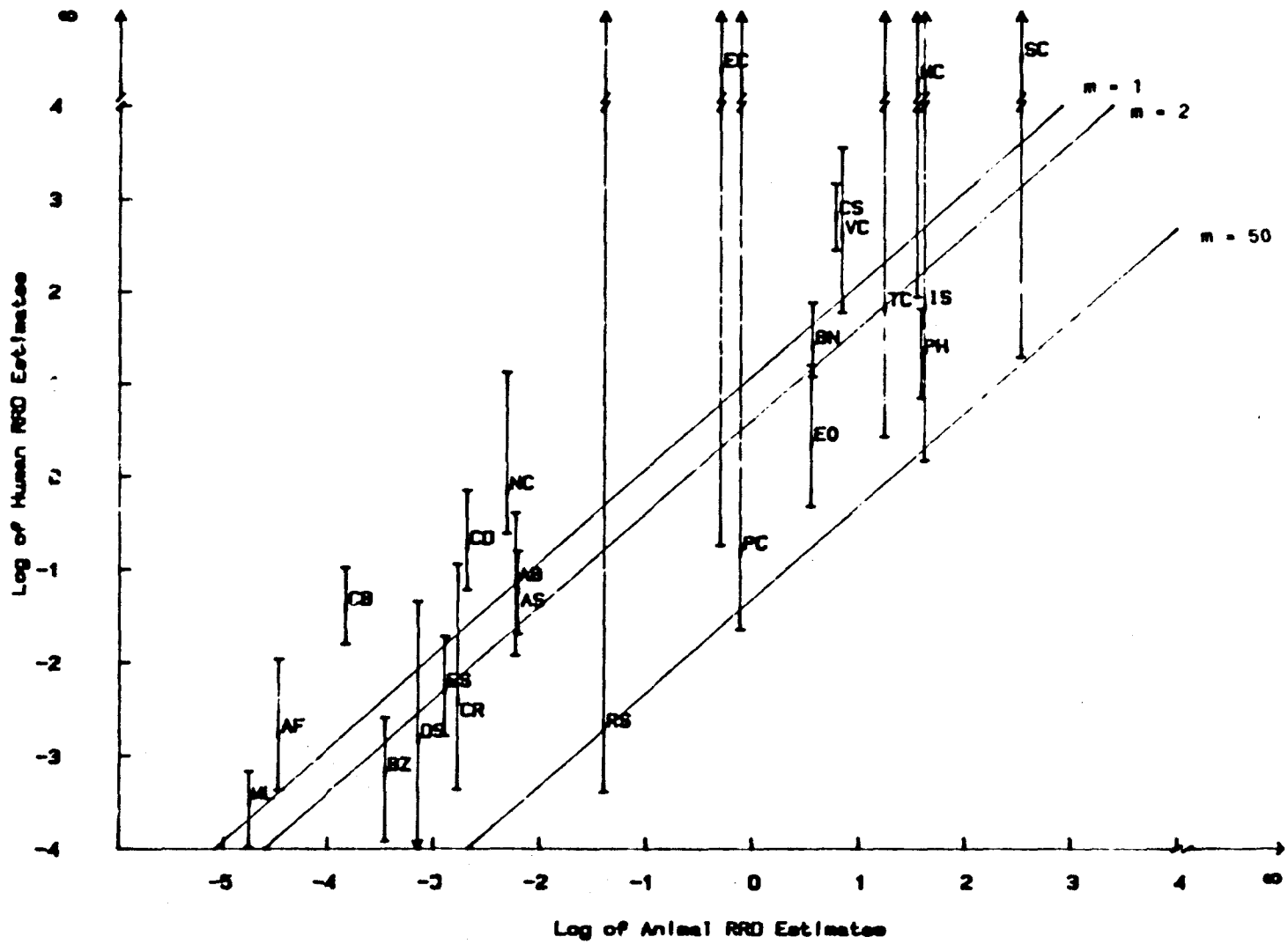


Figure 2-40

Prediction Analysis:
Analysis 20, Median Lower Bound; Best-Fitting Lines with Increasing Degrees of Asymmetry

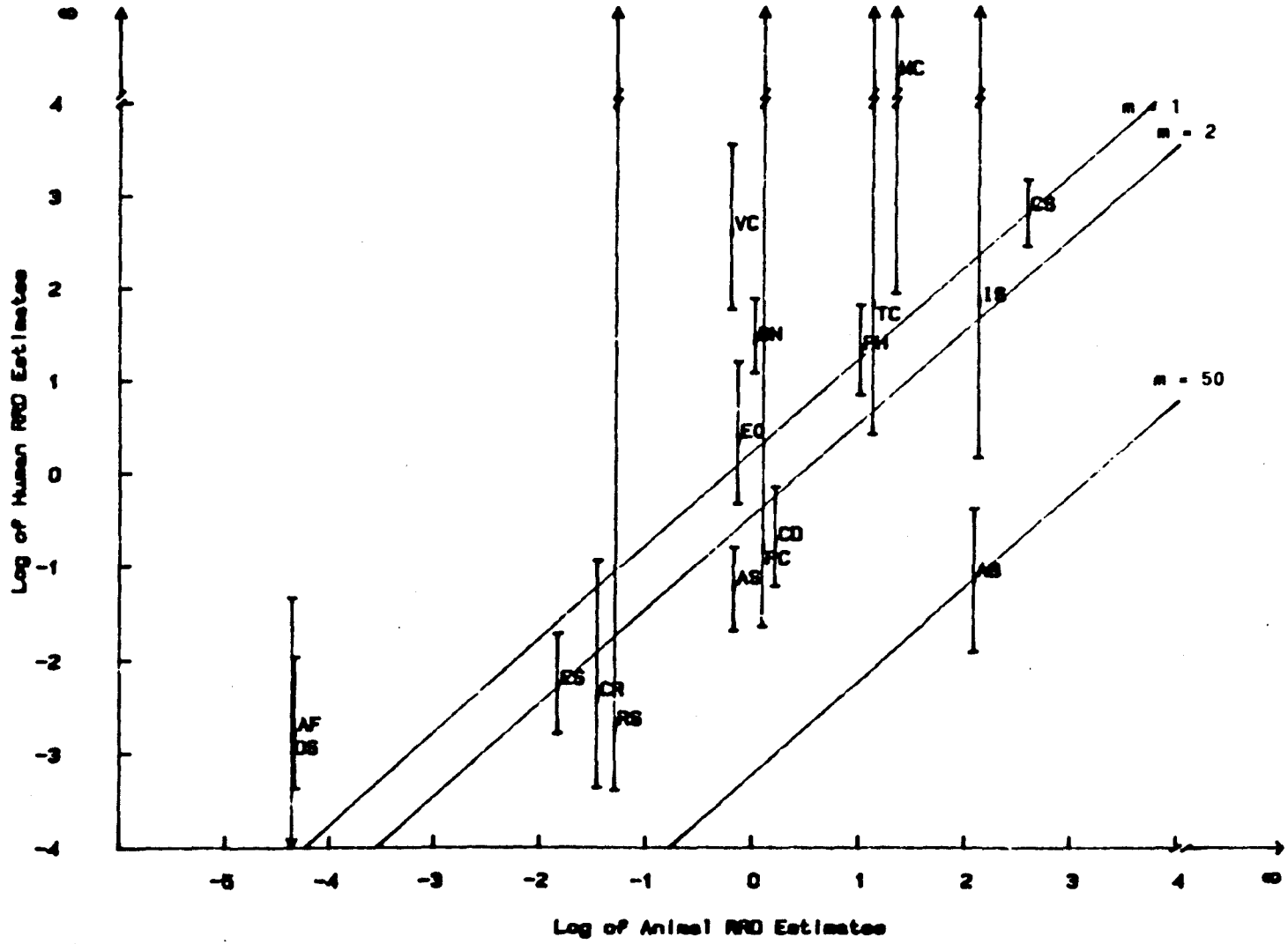


Figure 2-41

Prediction Analysis:
Analysis 22, Median Lower Bound; Best-Fitting Lines with Increasing Degrees of Asymmetry

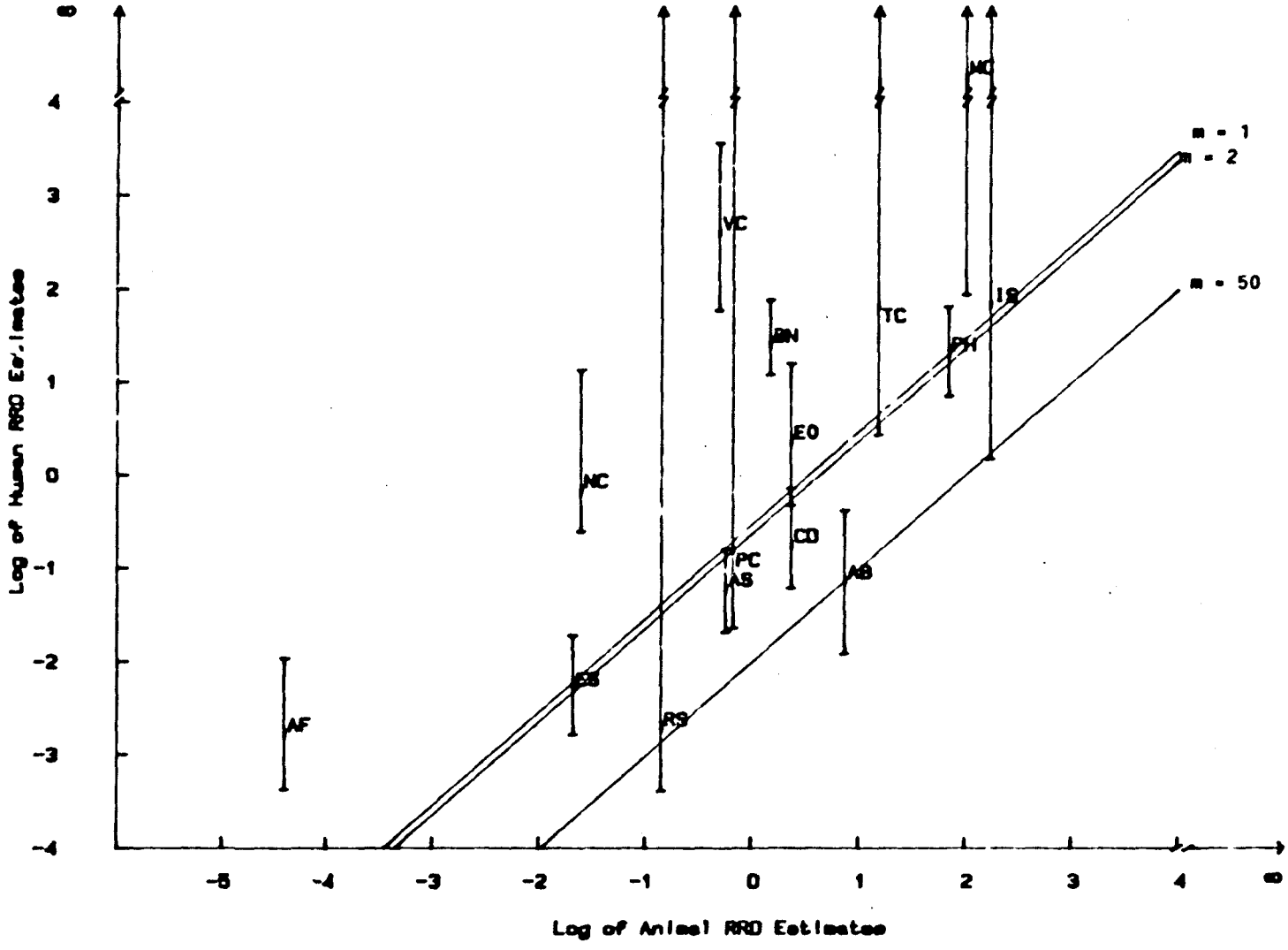


Figure 2-42

Component-Specific Uncertainty: Ratios of RRDs for Analysis 31 ($mg/m^2/day$) to RRDs for Analysis 30

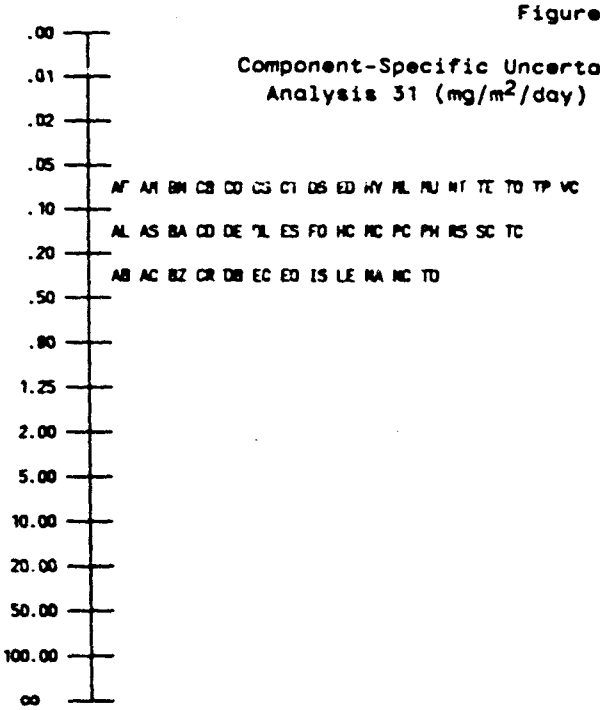


Figure 2-43

Component-Specific Uncertainty: Ratios of RRDs for Analysis 32 (ppm diet) to RRDs for Analysis 30

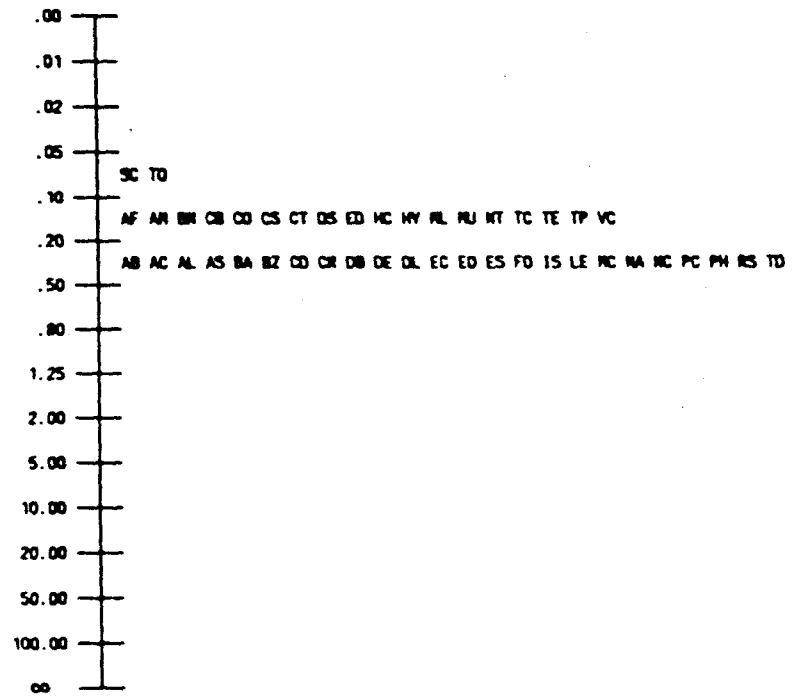


Figure 2-44

Component-Specific Uncertainty: Ratios of RRDs for Analysis 33 (ppm air) to RRDs for Analysis 30

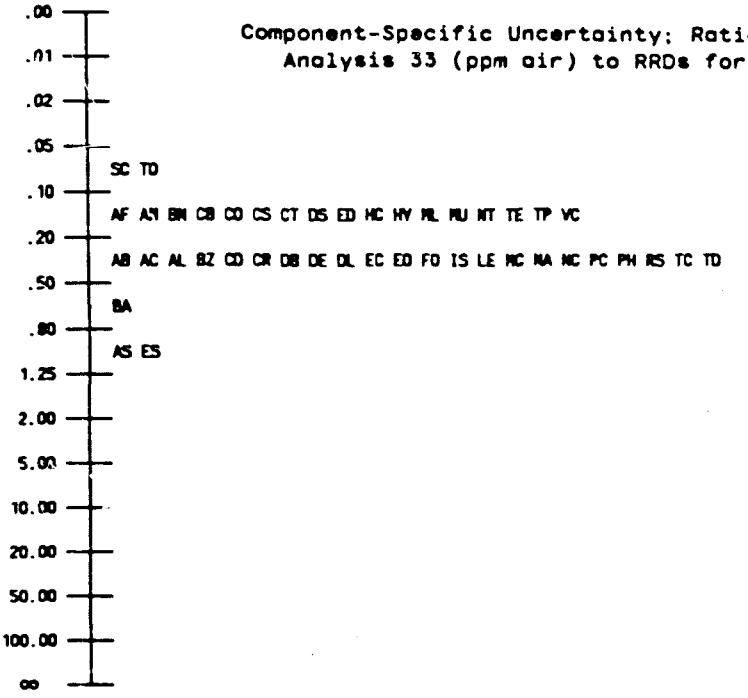


Figure 2-45

Component-Specific Uncertainty: Ratios of RRDs for Analysis 34 (mg/kg/lifetime) to RRDs for Analysis 30

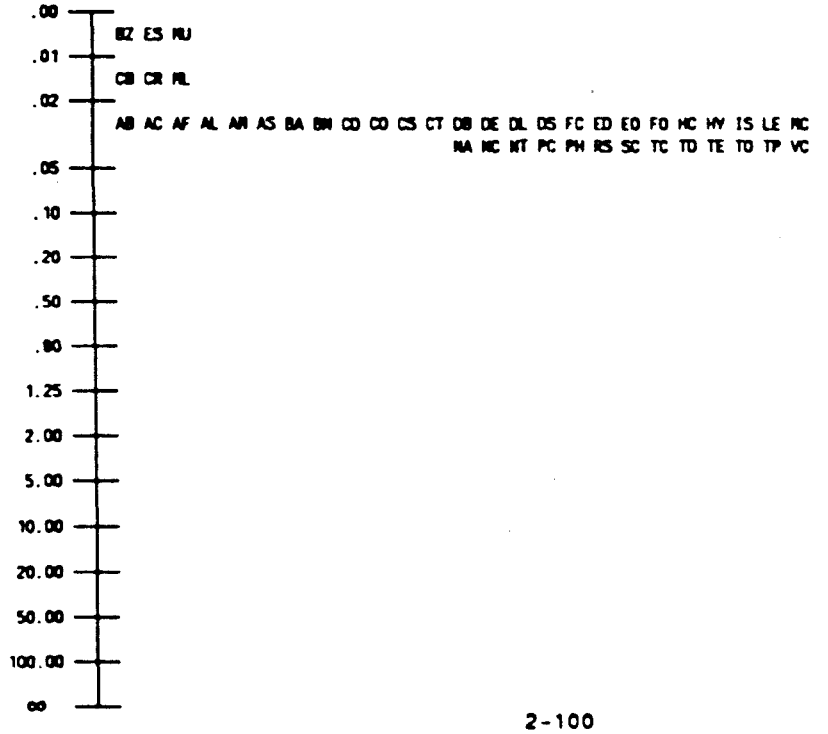


Figure 2-46

Component-Specific Uncertainty; Ratios of RRDs for Analysis 35 (Long Experiments Only) to RRDs for Analysis 30

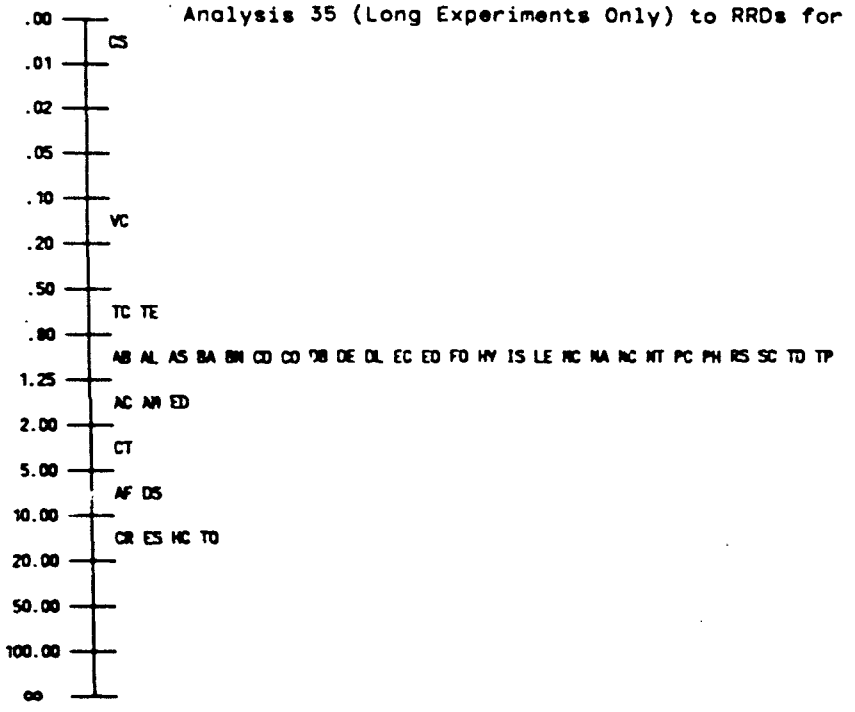


Figure 2-47

Component-Specific Uncertainty; Ratios of RRDs for Analysis 36 (Long Dosing Only) to RRDs for Analysis 30

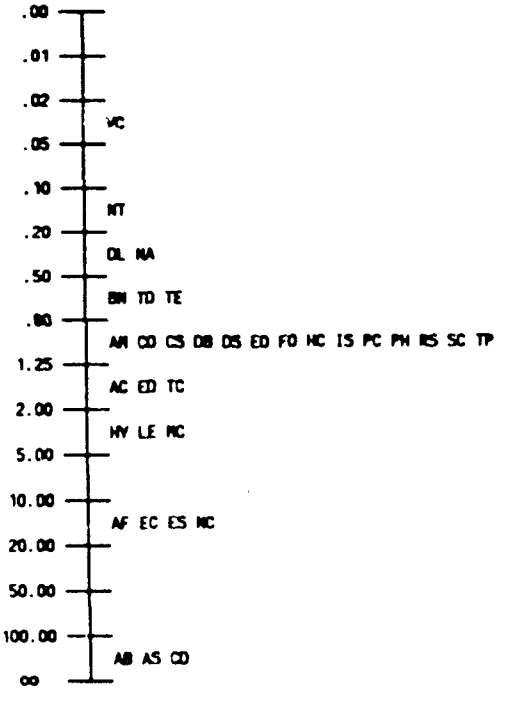


Figure 2-48

Component-Specific Uncertainty; Ratios of RRDs for Analysis 37 (Route Like Humans) to RRDs for Analysis 30

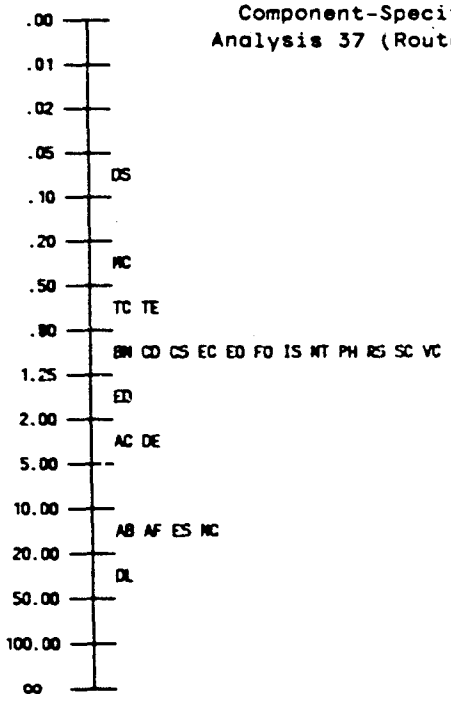


Figure 2-49

Component-Specific Uncertainty; Ratios of RRDs for Analysis 38 (Inhalation, Oral, Gavage, Route Like Humans) to RRDs for Analysis 30

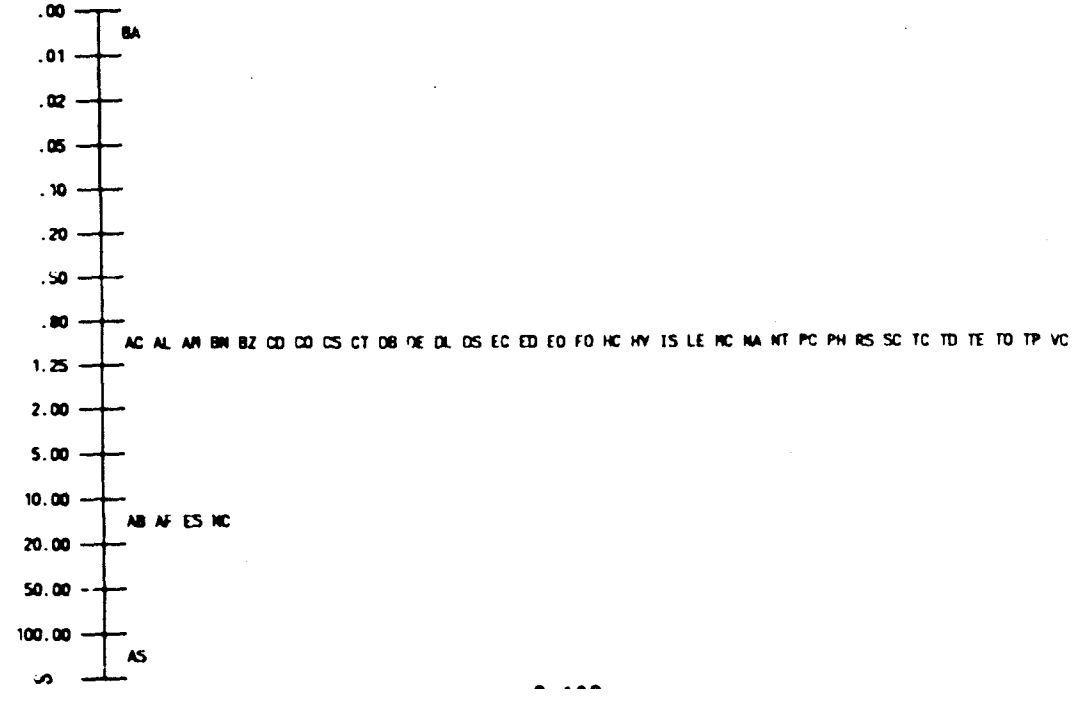


Figure 2-50

Component-Specific Uncertainty; Ratios of RRDs for Analysis 41 (Malignant Tumors Only) to RRDs for Analysis 30

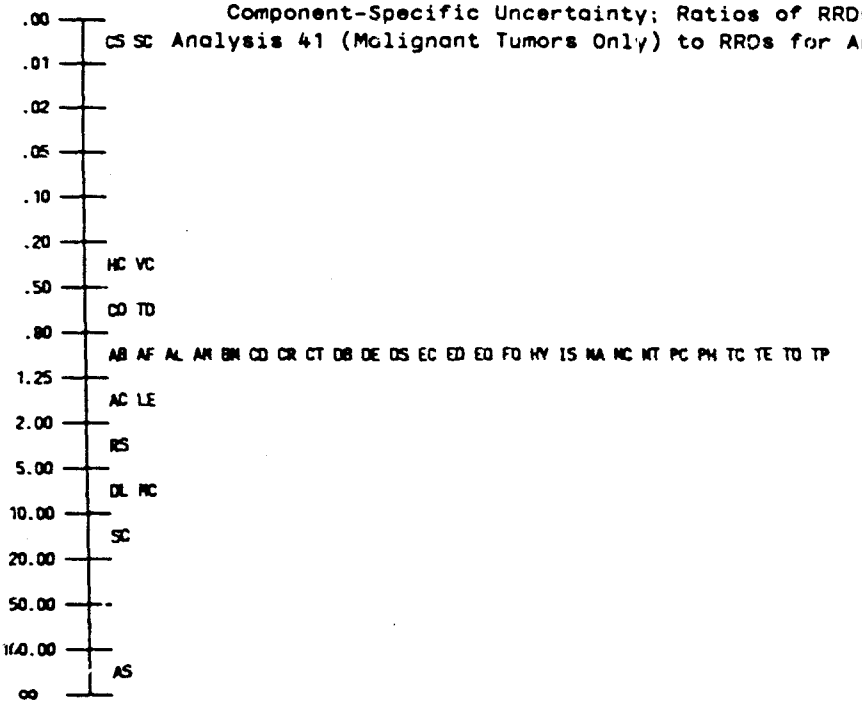


Figure 2-51

Component-Specific Uncertainty; Ratios of RRDs for Analysis 42 (Combination of Significant Responses) to RRDs for Analysis 30

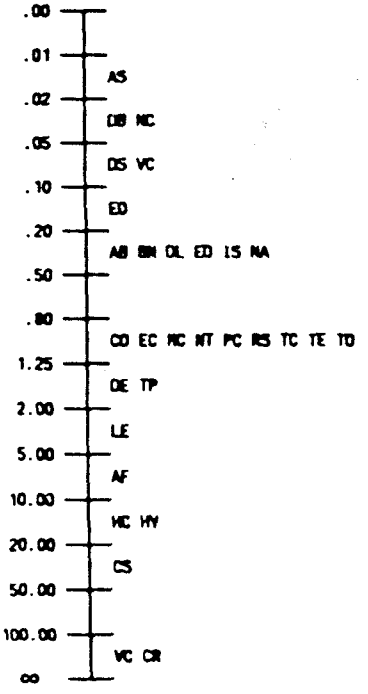


Figure 2-52

Component-Specific Uncertainty; Ratios of RRDs for Analysis 43 (Total Tumor-Bearing Animals) to RRDs for Analysis 30

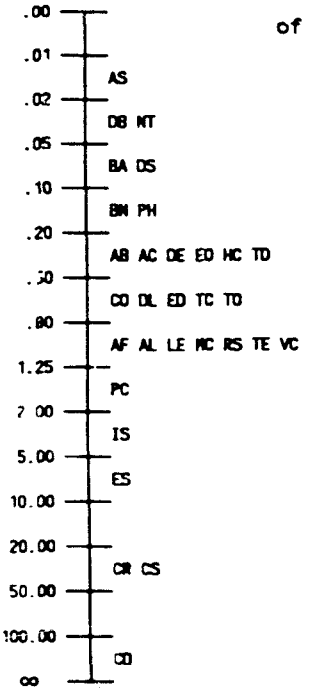


Figure 2-53

Component-Specific Uncertainty; Ratios of RRDs for Analysis 44 (Response Like Humans) to RRDs for Analysis 30

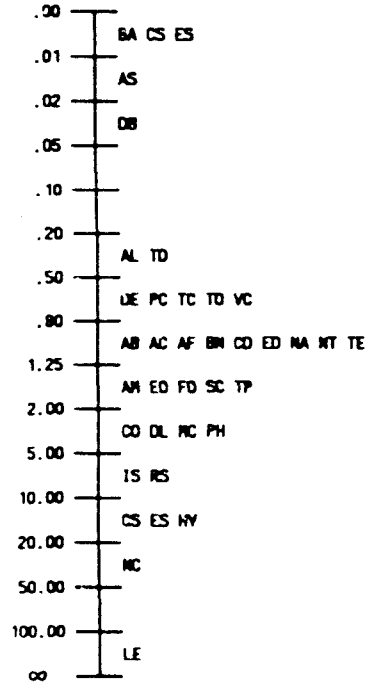


Figure 2-54

Component-Specific Uncertainty; Ratios of RRDs for Analysis 45 (Average Over Sex) to RRDs for Analysis 30

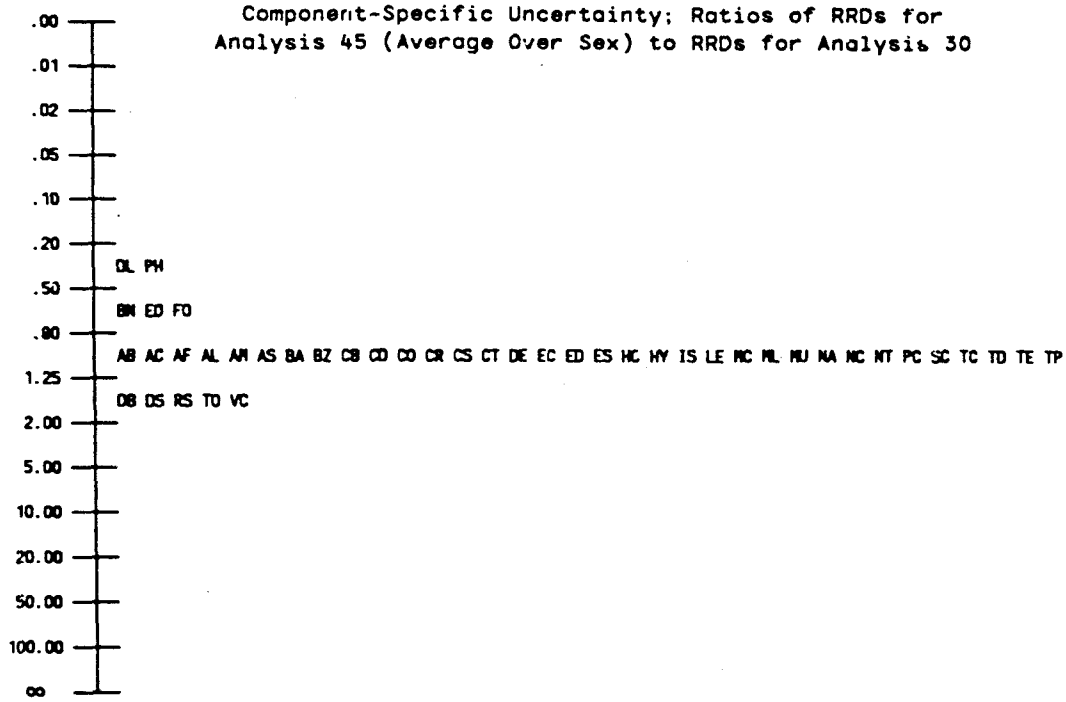


Figure 2-55

Component-Specific Uncertainty; Ratios of RRDs for Analysis 46 (Average Over Study) to RRDs for Analysis 30

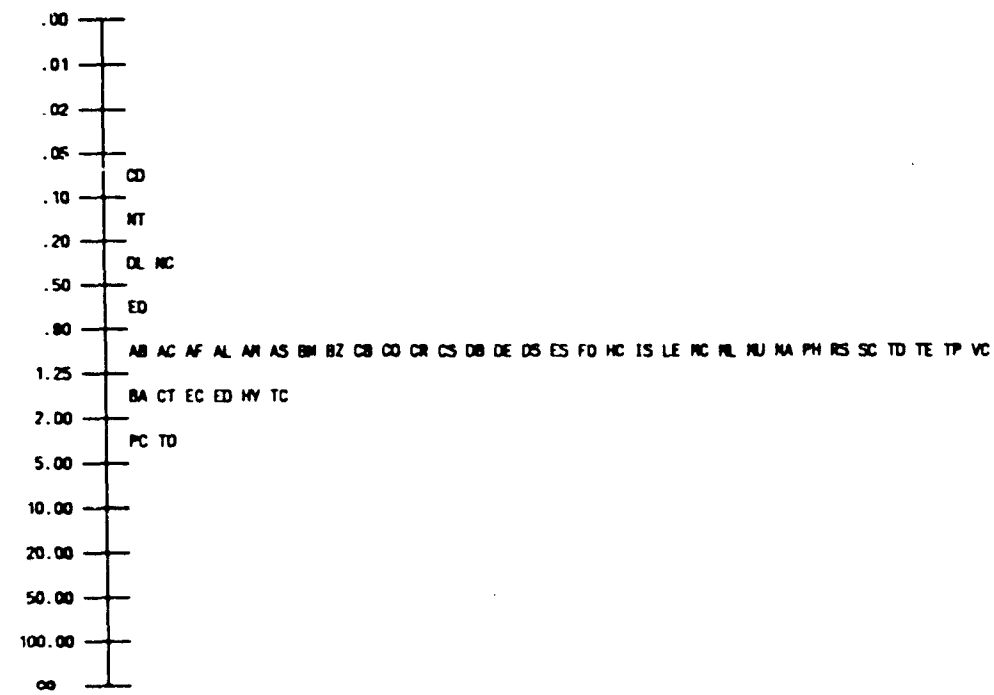


Figure 2-56

Component-Specific Uncertainty; Ratios of RRDs for Analysis 47 (Average Over All Species) to RRDs for Analysis 30

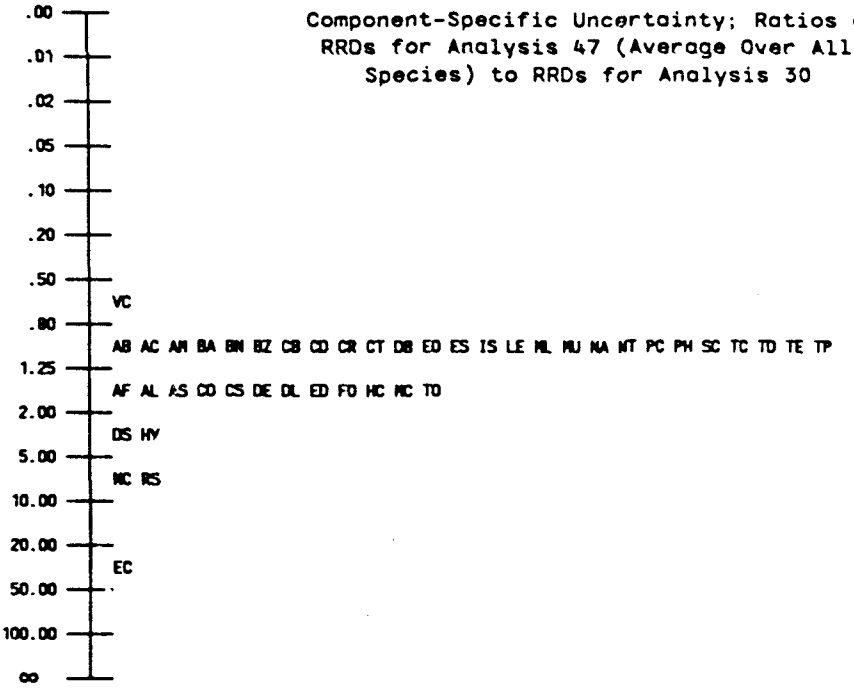


Figure 2-57

Component-Specific Uncertainty; Ratios of RRDs for Analysis 48 (Average Over Rats and Mice) to RRDs for Analysis 30

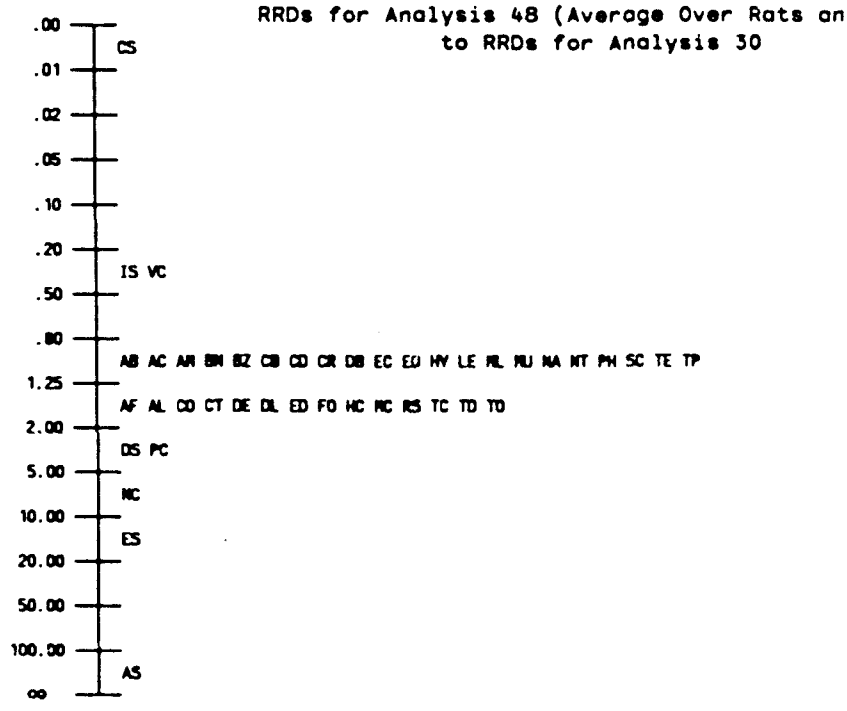


Figure 2-58

Component-Specific Uncertainty; Ratios of RRDs for Analysis 49 (Rat Data Only) to RRDs for Analysis 30

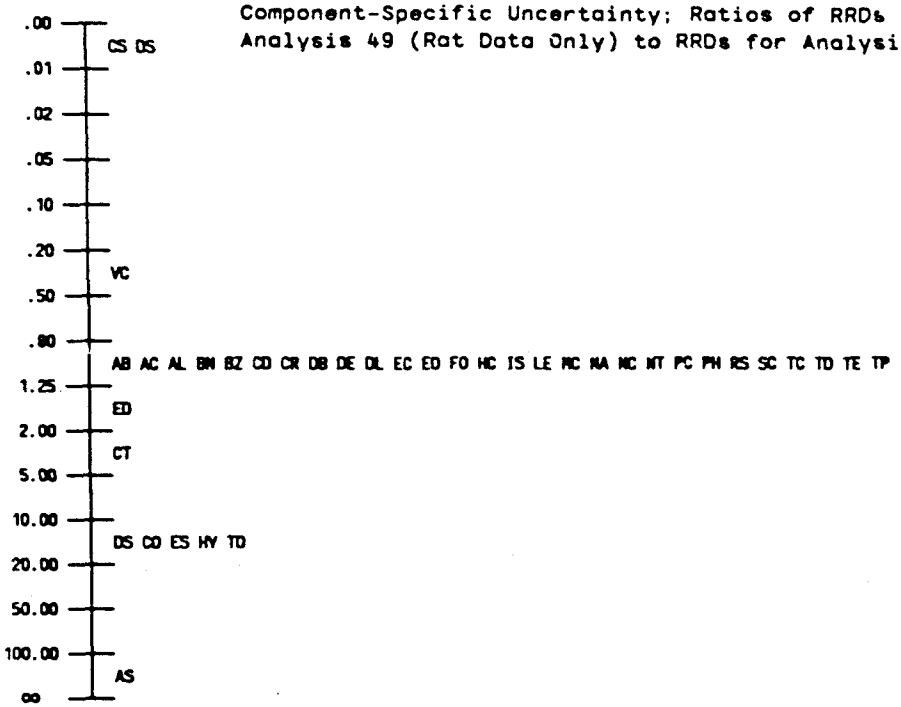
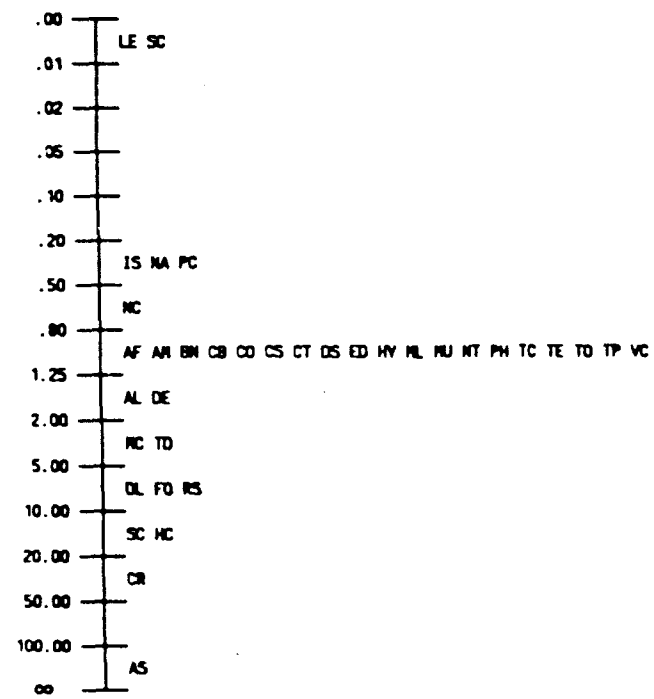


Figure 2-59

Component-Specific Uncertainty; Ratios of RRDs for Analysis 50 (Mouse Data Only) to RRDs for Analysis 30



Section 3

DISCUSSION

POSITIVE CORRELATION

The results presented in the previous section reveal that estimates of risk-related doses from animal bioassay data are generally highly correlated with the estimates derived directly from epidemiological data. Of the thirty-eight initial analysis methods investigated, 35 had p-values less than 0.05, when the full sieve was applied, and with that same sieve, 17 had p-values of 0.0001 or smaller. Even with no sieve, so that data from experiments of highly variable quality are included, thirty-five analyses have p-values less than 0.05. Not only do most of the analyses yield correlation coefficients that are statistically significantly positive, but the coefficients are large in an absolute sense as well. Twenty-six of the analyses have coefficients larger than 0.7.

The strongly positive result of the correlation analysis was obtained even though a number of uncertainties had to be accounted for. First, the uncertainty of the human RRD estimates is explicitly incorporated since the ranking that underlies the correlation analysis is based on the lengths and positions of the intervals of RRD estimates derived from the epidemiology. Those intervals reflect the uncertainty in the exposure estimates and statistical variability. Analysis of the epidemiological data, including the exposure uncertainty derivation, was conducted prior to the analysis of the animal data and, therefore, without knowledge of its outcome. Moreover, since the criteria used to determine the exposure uncertainty values were consistent across all chemicals, the subjective aspect of their derivation should not affect the correlations. That is, if the individual subfactors corresponding to sources of uncertainty in exposure estimation were to be altered, the bounds on exposure estimates would change in a predictable and largely consistent manner for all chemicals. The relative rankings of the chemicals should be minimally affected. This is one advantage of using

a nonparametric (rank-based) approach in the correlation analysis.

The second uncertainty accounted for in the correlation analysis pertains to the bioassay data. The intervals defined for the animal results also incorporate statistical variability; statistical lower bound and upper bound estimates define the endpoints of the intervals. In addition, the entire ensemble of data for a given chemical is considered in the sense that that ensemble defines the median lower bound and the median upper bound. So, while discounting extreme values, the intervals defined take into consideration the RRD estimates that can be obtained from each experiment in the data base.

These various uncertainties and the methods used to account for them will tend to wash out any real correlations that may exist, in the sense of producing small correlation coefficients that may not be significantly different from zero (no correlation). Despite this, strong correlations are obtained. The positive correlations exist for chemicals whose RRDs (and, therefore, potencies) span several orders of magnitude. Indeed, the strong correlations obtained despite these factors make it highly unlikely that the positive results are due to chance or to other factors not incorporated in the analysis.

The fact that these positive correlations exist is very important. The assumption that risk estimates derived from bioassay data are relevant to the estimation of human risk is crucial to all risk assessments for which epidemiological data is limited. Heretofore, it has been a largely untested assumption. The correlations determined in this investigation strongly support that assumption and thereby strengthen the scientific support for quantitative risk assessment.

The thirty-eight initial analysis methods represent a wide variety of approaches to bioassay-based risk assessment. Although a few of them appear to be less-well correlated with the epidemiological assessment results, the fact that most are highly correlated makes it reasonable to attempt to determine which methods are best when point estimates of risk are desired. The variety of methods ensures that a variety of point estimates will be available to discriminate between different predictors and different acceptable analysis approaches. This is the

subject of the prediction analysis, the interpretation of which will follow a discussion of data quality and the data screenings.

DATA QUALITY AND DATA SCREENING

As discussed in the second volume of this report, the extent and quality of the bioassay data varies greatly from chemical to chemical. Some chemicals have few acceptable experiments (*e.g.*, estrogen has two), some have experiments testing only one species (*e.g.*, chlorambucil with mouse data only), and some only have experiments of short duration or dosing (*e.g.*, benzidine and chromium). Still other limitations exist that affect the calculation of RRDs, such as the number of animals on test (which can greatly affect computation of the statistical confidence limits) as well as the actual conduct of the experiment (animal husbandry and care, adherence to protocol, etc., which we have not used to rate experiments) and, most importantly, data reporting limitations. Aside from reports like those produced by the National Toxicology Program, rarely were full details of the bioassay results available.

Even though correlation coefficients were large and significant for many of the unscreened (no sieve) analyses, as mentioned above, it was felt that some attempt should be made to use the "best" data that was available. On the other hand, it would not be appropriate to eliminate chemicals from the analyses on the basis of "quality" considerations. First, the maximum number of chemicals is 23, so that elimination of chemicals could lead to very small sample size. This is seen, for instance, when very restrictive criteria on carcinogenic endpoint and/or experimental protocol define an analysis method (*e.g.*, Analysis 19 with only nine chemicals) or when the data requirements are not satisfied by the published bioassay results available to us (*e.g.*, Analysis 6 with only six chemicals). Second, in any future risk assessment on a particular chemical, the data will undoubtedly be limited in certain respects. Part of our task is to try to determine how best to proceed even with those limitations.

Consequently, the data screening (sieve) selects the better data ("better" being defined solely on the basis of the definition of the

sieves) for use in the calculation of RRD estimates. In the present investigation, two screenings have been defined: a screening based on the significance of the dose-dependency of the carcinogenic responses and a screening based on the number of dosed animals and length of observation. The use of these two screenings yields three possible sieve approaches. Of course many reasonable alternatives, whether based on these criteria or others, are possible. No examination of other sieves has been undertaken.

The goal of screening the data is to produce a data base that will perform better when compared to the epidemiological estimates. The sieves defined here appear generally to achieve that goal. The significance screen, in particular, worked to increase the correlation coefficients for 25 of the 38 initial analyses. Although the quality screen does not provide substantial improvement over the significance screen in most cases, certain analyses are much better correlated when both screens (the full sieve) is applied. Rarely does the addition of the quality screen to the significance screen decrease the correlation. Thus, we have selected the full sieve to represent the action of data screening in the prediction and uncertainty analyses.

However, in the prediction analysis, it is frequently the case that the average loss for an analysis method is greater when the full sieve is applied than when no sieve is applied (cf. Tables 2-3 through 2-5 and Table 2-13), especially when the median lower bound estimate is the predictor. One might be tempted to conclude that only those methods yielding smaller loss when the sieve is applied should be considered as good risk assessment procedures. Conversely, one might conclude that either the sieve is not working correctly to select the better data or that it is working but the data it selects are not, in actuality, better for risk assessment purposes. We argue that any of these conclusions is unwarranted.

First, the results of the correlation analysis strongly indicate that the better data are being selected by the sieve and that these data are, in actuality, better for risk assessment purposes. This is in accordance with common sense: if too few animals are tested or if the period of observation is too short, then it is difficult to elicit an

observable (dose-related) carcinogenic response. Similarly, those responses that are significantly related to dosing tell us more about the carcinogenicity of a chemical than the endpoints that lack a significant relationship with dose (unless all responses lack a significant relationship, but then the significance screen does not eliminate any of the responses from consideration). Had the results of the correlation analysis been less consistent in indicating the benefit of the sieve, then one might have reason to suspect that the "common sense" reaction is not supported and may be in error.

Secondly, we prefer the correlation analysis results over the prediction analysis results as indicators of the action of the sieve since the former does not select a single estimate from each analysis method and it is not dependent on the specification of loss. The correlation analysis utilizes a range of estimates consistent with the ensemble of data available for each chemical and employs a general measure of the degree of similarity between the animal and human estimates. This framework is less sensitive to variations in the data and results that are due to unintentional changes (confounders) in the data base. It is entirely possible that application of the sieve may tend to eliminate certain routes of exposure, for example, although such a result is not the intended result of the sieve. Unless the elimination entails substantial change in the RRD estimates (as in the case of arsenic or estrogen, as discussed in Section 2) the generalized ranking scheme is not unduly affected.

Indeed, we feel that such confounding changes in the data base and random variation may largely explain the occurrence of average losses that are greater when the sieve is applied than when it is not. For any experiment, random factors affect the response rates and, consequently, the estimation of RRDs. For all the bioassays of a particular chemical, then, the changes seen when a sieve is applied depend on these random variations. [Again, this is one reason for preferring the correlation analysis over the prediction analysis as a test of the sieve: the correlation analysis accounts for the random variation by using upper and lower confidence limits instead of a single point.]

As a consequence of this observation, it is appropriate to compare the analysis methods in the prediction analysis both with and without the sieve. An analysis that yiel s smcll average loss under either of these conditions should bc considered a viable option in the sense of determining the best risk assessment procedures. Thus, for example, Analysis 20 without the sieve is the best approach, as measured by the TANH loss function, when L_{20} is the predictor (cf. Table 2-7). Average loss for that analysis is increased when the sieve is applied so that, even among the analyses employing the sieve, Analysis 20 is no longer the best. We wish to continue to consider Analysis 20 as a good potential procedure since it is not known whether the increase in average loss may be due to random variation or to data base changes confounded with the application of the sieve, though we suspect that it is. This procedure is followed throughout this discussion; those analyses cited as being good are those with small average losses for at least one of the pair (with sieve and without sieve). However, in the suggested guidelines for presenting risk estimates, and in the examples provided, screening of the data is always performed, no matter which analysis method is applied.

APPLICATION OF ANALYSIS RESULTS IN EXTRAPOLATING FROM ANIMALS TO HUMANS

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. Dose units that have been applied in the past include mg/kg body weight/day, mg/m^2 surface area/day, ppm in diet or air, and mg/kg body weight/lifetime. Because of differences between animals and humans in body weights, life spans, etc., use of different units produce different estimates of human risk. There is limited scientific support for use of any particular dose units (1). Results from the present study can be used empirically to determine appropriate methods for animal-to-human extrapolation. Specifically, multiplication of the animal RRD by the 10^c, where c is the y-intercept from the best-fitting line, 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 has 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). The bias correction factor 10^c corrects for any overestimation or underestimation by the analysis method used.

IDENTIFICATION OF GOOD METHODS

Predictors

Each analysis method was run with the four predictors examined in this investigation, the median and the minimum of the lower bound RRDs and of the maximum-likelihood RRDs. Three loss functions were defined that determine the lines of unit slope that minimize the total loss for the collection of chemicals being analyzed. Despite the fact that the three loss functions calculate loss in different ways, all three are consistent in indicating that the median lower bound RRD predictor, L_{2Q}, yields the smallest average losses for most analysis methods. It should be emphasized that this is a strong result not only because of the consistency of the loss functions but primarily because it is not dependent on the particular data that were available for analysis. For any given analysis (with rare exceptions) the loss when L_{2Q} is used is smaller than losses with other predictors even though the very same data are used to calculate the estimates and, hence, the losses for each predictor.

It is important to note that the predictor, L_{2Q}, which is derived from lower bound RRDs, yielded smaller losses than any of the predictors based on maximum likelihood estimated RRDs. This is probably related to the fact that small changes in the bioassay data can result in sizable changes in MLE estimates of RRDs, which suggests that the desirable large-sample theoretical statistical properties possessed by MLE estimates (such as consistency and asymptotic efficiency) are not operative to any practical extent in this situation given the usual sample sizes encountered in bioassays. This lack of stability of the MLE estimates is a much more severe problem when extrapolating to low doses. Regulatory agencies have in the past relied more on lower bound RRDs than maximum likelihood estimates, mainly in the interest of being protective of human health. This study shows that lower bound RRDs are,

in fact, better predictors of the human data than are the MLE estimates, and thus provides additional rationale for emphasizing lower bound RRDs in risk assessment. This study also estimates the level of conservatism (or anti-conservatism) that may be inherent in specific analyses that use lower bound RRDs and estimates compensating or bias-removing factors, i.e. the conversion factors (10^c). This issue will be discussed further later.

One potential problem with use of lower bound RRDs is that they are always finite, even when the data show no evidence of carcinogenicity (consistent with infinite maximum likelihood RRDs). To some this might imply that use of lower bound RRDs will lead a regulatory agency to treat every chemical as a carcinogen, irrespective of bioassay results. This need not be the case. For most purposes, there must be reasonably convincing evidence of carcinogenicity from bioassay results before an agency will undertake the assessment of risk. Moreover, the problem may be further mitigated if we recall that the correlation analysis demonstrated the strong positive correlation between ranges of human and ranges of animal RRDs. This result does not depend on the position of the best epidemiological or bioassay estimates, only on the bounds for estimates. Consequently, we know that those chemicals that tend to have larger RRD estimates (lower bounds) from epidemiological analyses also tend to have larger bioassay-based estimates (lower bounds) so that chemicals with large L_{20}'s (in a relative sense, compared to other chemicals) are those that may be of less concern when it comes to regulation and/or control. One corollary of this line of reasoning is that the degree of correlation, in addition to the average losses calculated for specific predictors, is an important factor in comparing the analysis methods and deciding which are better.

At any rate, there will always be the possibility that a noncarcinogen may be regulated as a carcinogen on the basis of false-positive data. Use of MLEs would not remove this problem; MLE RRDs from bioassays of noncarcinogens will be finite about 50% of the time. In this regard, it is of interest to note that in this study chemicals with infinite RRD estimates based on the epidemiological analyses did not in general have infinite maximum likelihood RRD estimates based on the animal data.

However, this was to some extent prearranged because for a chemical to be included in the analysis, positive evidence of carcinogenicity (implying finite RRDs) had to exist for either animals or humans.

Use of L_{2Q} as the predictor is in a sense less conservative than use of the minimum lower bound, L_M. The y-intercepts for the analyses are almost always larger when L_M is used in place of L_{2Q} (compare Tables 2-12 and 2-13). This means that L_M is more conservative, in fact, generally overconservative. Of course, if one applies the conversion factors suggested by the y-intercepts, then no approach is more or less conservative than another; estimates obtained by using the conversion factors are those that come closest to the epidemiological estimates. In this sense, the remaining error, expressed as average loss, is the primary determinant of good or bad analyses or predictors. As previously mentioned, L_{2Q} is preferable to L_M in this regard. But it is also the case that the conversion factors are less extreme with L_{2Q} than with L_M.

Analysis Methods

Given the superiority of L_{2Q} over the other predictors examined, one can compare the analysis methods on the basis of how they perform with L_{2Q}. This has been done for each loss function separately (cf. Table 2-7) and for the three functions combined (Table 2-8) for the initial 38 analyses. The supplemental analyses (Table 2-16) should also be considered, especially since their template is Analysis 3b which is a method producing excellent correlation and which is also identified as resulting in small average loss.

Analyses 6, 18, and 19, which are applicable to limited numbers of chemicals (six, ten, and nine, respectively), will not be examined in detail. Although both the correlation and prediction analyses suggest that these methods may be beneficial, the data are not sufficient to warrant detailed examination of these methods. In order to use the methods routinely, data availability would have to be improved. To perform Analysis 6, one must have available the number of animals alive in each dose group at the time of first occurrence of each tumor type. For Analyses 18 and 19 (as well as any other method that uses an

endpoint that is a combination of individual carcinogenic responses) one must know which animals got which tumors in order to combine responses. Detailed reporting procedures like those in many National Toxicology Program reports are ideally suited for these purposes. Bioassay results published in peer-reviewed journals rarely contain such detail. Nevertheless, some other means must be found to disseminate the full results before analyses like 6, 18 or 19 can be more thoroughly investigated. The incomplete but suggestive results of Analyses 6, 18, and 19 indicate that this may be worthwhile.

Comparing the results in Table 2-16 to those in Table 2-7 reveals that several analysis methods from the supplemental list are as good as or better than the best of the initial analyses. With the DISTANCE^2 loss function, Analyses 30, 45, and 47 yield the smallest losses of any analyses. Similarly, Analysis 43 results in the smallest loss as measured by the TANH loss function; Analyses 45 and 47, as well as 30, also perform well. Only the CAUCHY function failed to find a supplemental analysis that outperformed the best of the initial analyses (Analysis 17, CAUCHY average loss 0.363). However, analyses 45, 47, 30, and 43 (the latter without the sieve) yield CAUCHY losses comparable to the five best initial analyses.

Using the criterion of total incremental normalized loss, Analyses 3b, 17, and 20 were determined to be the best of the initial analyses. Note that these are the only analyses among the best five for more than one loss function. Analysis 3b is the same as Analysis 30 except for a change of units (mg/m^2/day and mg/kg/day, respectively; Analysis 3b is the same as Analysis 31). Analyses 17 and 20 have no counterparts among the supplemental analyses because those two average over sex, study, and species, differing from the initial standard in more than one component. However, the other supplemental analyses mentioned in the previous paragraph in connection with small average loss (methods 43, 45, and 47) have analogs in Analyses 8b, 9, and 11a, respectively (though the latter use mg/m^2/day, not mg/kg/day, and are restricted to certain routes of exposure). Analyses 8b and 9 are among the top five initial analyses for one of the loss functions and were mentioned in the previous section in connection with analyses that have relatively small total incremental normalized loss. It appears, then, that Analyses 17, 20, 30, 31 (= 3b),

43, 45, and 47 yield good predictions, that encompass different approaches to several of the components, and therefore constitute prime candidates for selection of appropriate bioassay analysis methods.

These and other analyses of interest are compared in Table 3-1. Analyses 17 and 20 are included because they appear to outperform all other initial analyses as assessed by one of the loss functions. Recall that Analyses 17 and 20 are methods that average RRD estimates over sex, study, and species. The RRD estimates that are averaged correspond to specific endpoints, the combination of significant responses and total tumor-bearing animals for 17 and 20, respectively. Analysis 17, but not 20, is limited to experiments that dosed for at least 80% and observed for at least 90% of the standard length of experiment. In addition, both Analyses 17 and 20 consider only those experiments administering the test chemical orally or via gavage, inhalation, or the route by which humans encounter the chemical. Similarly, Analysis 31 (35) was the best of the initial analyses when measured by one of the loss functions. Also like Analyses 17 and 20, Analysis 31 extrapolates estimates to humans on a surface-area basis (*i.e.* using $mg/m^2/day$). Analyses 0, 7, 11c, and 11d, all presented in Table 3-1, are surface-area-based extrapolative procedures as well, and, like 17 and 20, consider only oral, gavage, or inhalation studies unless another route is commonly encountered by humans. They are included in Table 3-1 because of their general interest. Analysis 0 is the one modeled after the EPA Carcinogen Assessment Group's approach (although the median, not minimum, lower bound is the predictor used here). Analysis 7 is the same except that malignant carcinogenic responses are the only ones considered. Similarly, Analyses 11c and 11d differ from Analysis 0 in the species considered; Analysis 11c is limited to rat experiments while Analysis 11d is limited to mouse tests.

The remaining analyses presented in Table 3-1 (30, 43, 45, and 47) extrapolate risks from animals to humans on a body-weight basis using mg/kg/day. In every other respect Analysis 30 is identical to Analysis 31: the RRD estimates are not averaged and all experiment and all individual carcinogenic responses found in those experiments are allowed. Analysis 43 differs from Analysis 30 only in that total tumor-bearing animals is the single endpoint evaluated in the former

case. Similarly, Analyses 45 and 47 differ from Analysis 30 because some averaging of RRDs does take place; for Analysis 45, estimates are averaged over bioassay identical except for the sex of the test species (*i.e.* over sex within study) and for Analysis 47, results obtained for each species are averaged to yield the ultimate RRD estimates.

The Base Analysis (Analysis 0) employing the minimal lower bound estimator, L_M (second row of Table 3-1) 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. However, this method is perhaps most like that presently employed by EPA. Modification of this method by using the median lower bound estimator, L_{2Q}, rather than L_M, as represented in the first row of Table 3-1, 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 use of minimum estimates.

Although Analyses 0, 7, 11c, and 11d were associated with generally good correlation values, their normalized loss values do not compare with the best of the remaining methods (*e.g.* Analyses 43, 45, and 47). Moreover, the residual uncertainty factors associated with these analyses are among the largest presented. Analyses 0, 7, 11d, and 11d therefore are not considered to be among the better methods for predicting human risk on the basis of bioassay results.

The case is somewhat more complex for Analyses 17 and 20. As previously mentioned, Analysis 17 is the best method determined by the CAUCHY loss function. In part because of that result, the total incremental normalized loss for Analysis 17 is nearly the smallest. Nevertheless, its correlation coefficient is also the smallest of those presented. Even if one notes that Analysis 17 is applicable to only 11 chemicals and that consequently the coefficient would be less stable, the importance attached to the correlation results when using the L_{2Q} predictor (as described above) tends to make the use of method 17 less desirable than use of the other methods. In addition, the residual

uncertainty factor is relatively large, in the range of those associated with Analyses 0, 7, 11c, and 11d.

Analysis 20 also has a large uncertainty factor. That fact, plus the large incremental loss value, makes Analysis 20 less appealing than the remaining five analyses, 30, 31, 43, 45, and 47. Note also that Analysis 43 is the only other method in Table 3-1 that uses the endpoint used by Analysis 20, total tumor-bearing animals. Analysis 43 is superior in all respects to Analysis 20. This is one other reason why Analysis 20 should not be considered among the better methods for extrapolating human risk.

There is another reason not to recommend Analyses 17 and 20 for use in extrapolating between humans and animals. Aside from Analyses 0, 7, 11c, and 11d, which have already been deemed inappropriate, only Analyses 17 and 20 are restricted to specific routes of exposure. It is likely, given the pattern seen for other analysis methods, that methods identical to 17 and 20 but without this restriction on route would do even better. This would seem to be the case because the supplemental analyses generally yield smaller losses than those analyses in the initial set that differ only with respect to allowable routes of exposure and units of extrapolation, the latter having little effect on average loss.

Analysis 20 and the other method using total tumor-bearing animals as the endpoint, Analysis 43, are the only two that overestimate RRDs on average. On the other hand, Analysis 31, a method extrapolating risk on the basis of $mg/m^2/day$, underestimates RRDs by roughly an order of magnitude. Analyses should not be compared on the basis of these conversion factors, however. When the estimates from any analysis method are multiplied by the indicated conversion factors, a line fit to the converted estimates (on the x-axis) and the epidemiological results (on the y-axis) would pass through the origin. The conversion factor represents a degree of freedom in the prediction analysis corresponding to the estimation of the intercept. A conversion factor estimated here for any method can be used to adjust the results obtained for a particular risk assessment on a single chemical when the bioassay data is analyzed by that method.

One might be tempted to conclude that Analysis 45, which extrapolates from animals to humans on a mg/kg/day basis, accepts all routes of exposure, and averages results over pairs of experiments that differ only with respect to the sex of the test animals (*i.e.* the experimenters, protocol, and strain of the test animal are the same), is the best of the analyses presented in Table 3-1. Its correlation coefficient is as good as any other, its incremental loss is smallest, and its residual uncertainty factor is smallest (Figure 3-1). While this analysis is certainly a good one by all these criteria, it is not possible to conclude unequivocally that Analysis 45 is better than some of the others listed. For one thing, the ranking of analyses differs depending on the choice of loss function. Analysis 45 is best with the DISTANCE^2 loss function, Analysis 17 is best with CAUCHY and Analysis 43 with TANH. A minimax criteria would select Analysis 17, followed by Analysis 43. Moreover, it is not clear which of the loss functions is most appropriate for determining fits of RRDs and no statistical development that would allow us to test for lack of fit or to test differences in values of average (or total) loss is available. For these reasons, the loss functions have been used in this investigation as a method of ranking the analyses. Since no one loss function is obviously more appropriate, an overall measure such as total incremental normalized loss has been employed to find analyses that are fairly robust with respect to calculation of loss. The analyses in Table 3-1 remaining after elimination from consideration of Analyses 0, 7, 11c, 11d, 17, and 20 (*i.e.* Analyses 30, 31, 43, 45, and 47) demonstrate such robustness.

Let us call these five analyses the set of recommended analyses. RRD estimates derived using these analyses for each of the chemicals included in this investigation but lacking epidemiological data sufficient for quantitative assessment are presented in Table 3-2. The values in Table 3-2 have not been adjusted by the conversion factors. When this is done the range of RRD estimates for each chemical appears as in Table 3-3.

One final comment will conclude this discussion of recommended analysis methods. Four of the five members of the recommended set make no

restrictions on the carcinogenic endpoints considered, and data are generally available for conducting these analyses. Analysis 43 utilizes total tumor-bearing animals as an endpoint. While, in theory, this should pose no restrictions on the analysis of bioassay data, in practice these endpoints often cannot be defined. Availability of needed data is an important consideration when assessing human risk and expressing the results as a range of RRDs consistent with the data but quantitatively incorporating uncertainties.

COMPONENT-SPECIFIC UNCERTAINTY

The discussion to this point has not considered uncertainty associated with any specific components of the risk assessment process. Rather, we have emphasized the analysis methods as wholes and examined the uncertainty remaining after the predictions have been obtained and compared to the human RRDs (residual uncertainty). This course has been followed because of the apparent interaction of the components. This interaction takes two forms. First, certain components are not mutually independent. A component that defines approaches to length of dosing obviously also influences choices concerning length of observation; a study cannot dose animals for 80 weeks without also observing the animals for at least 80 weeks. Moreover, as discussed above, altering the approach to some components can also, unintentionally, affect the make-up of the underlying base of data and, hence, changes attributed to changing those components may be confounded by changes that may be partially explicable by changes in other components. If, for example, limiting experiments to those that last at least 90 percent of the standard length also, unintentionally, excludes routes of exposure besides inhalation, oral, or gavage, then the change in RRDs attributed to changing requirements on the length of observation is confounded by changes due to restricting routes of exposure.

It is also the case that a component-specific investigation is not sufficient to characterize the best approaches because of the second type of interaction, the empirical interaction of the components on the results. Consider, for example, the components relating to choice of dose units (specifically, the approaches specifying use of $mg/m^2/day$ and

mg/kg/lifetime) and allowable routes of exposure (unrestricted versus restriction to an oral route, gavage, inhalation, or the route that humans encounter). These two components are not inherently interrelated. Nevertheless, the effect on the RRD estimates and on the estimation of loss (*i.e.* the adequacy of the predictions) resulting from selection of approaches to the indicated components is not readily attributable to one component or the other. Note that, when L_{20} is the predictor, Analysis 0 ($mg/m^2/day$, restricted routes) yields average loss of 0.298 as measured by the DISTANCE[2] function (sieve applied). If the units are changed (Analysis 4d: mg/kg/lifetime, restricted routes) or if the allowable routes are augmented (Analysis 31: $mg/m^2/day$, unrestricted routes) the loss decreases, to 0.267 or 0.113, respectively. When both components are changed, however (Analysis 34: mg/kg/lifetime, unrestricted routes), the decrease is intermediate between the two; average loss in that case is 0.131. Hence, the two components do not act independently on the estimates for some or all of the chemicals. In this sense, it is pointless to debate whether $mg/m^2/day$ is an appropriate dose measure for animal-to-human extrapolation without taking into consideration the approaches taken for other components. Evaluation of risk assessment methods should focus on the complete process rather than on individual components.

Consequently, one must be cautious in interpreting results of component-specific changes in analysis methods and should not evaluate analysis methods solely on the basis of component-specific changes. Nevertheless, such an examination may be useful in determining sources of uncertainty in risk assessment and in suggesting means of improving risk estimation through additional research or data acquisition. The results of our component-specific uncertainty investigation may also be useful for presenting a range of human risk estimates and so can be incorporated into the guidelines for determining that range.

The components can be divided into two sets. First are those that do not change the data base underlying the assessment. Included in this group are the components dictating the dose units used for extrapolation and those specifying the manner in which results are averaged. Such components are not susceptible to confounding due to unintentional changes in the data base. These are the components that show very

consistent changes when approaches to them are altered (*cf.* Table 2-23). The components related to averaging results have relatively little effect on the RRD estimates; the modes are in the interval 0.8 to 1.25 and the dispersion factors are between 1.2 and 2.2. Interestingly, two of the analyses included in the recommended set (45 and 47) differ from the standard, Analysis 30 (also in the recommended set), only in the way they average results. It appears that Analysis 30 is a satisfactory method of bioassay analysis and RRD prediction; the analyses that differ from it only in the approach to a component that produces consistent changes in RRDs also tend to be satisfactory. Changing dose units also produces consistent changes in RRD estimates (dispersion factors between 1.3 and 2.3) although the modes of the distributions are shifted, often substantially. Again, the analyses that differ from 30 only with respect to dose units yield relatively good predictions; Analysis 31 is included in the recommended set.

The second category of components includes those that change the data base on which a risk assessment is based. These display the least amount of consistency with respect to RRD changes and so are the most uncertain aspects of quantitative risk assessment. This conclusion is not diminished by the fact that these components are subject to confounding due to unintentional data changes. In any assessment of a particular chemical, which may have more limited data than many of the chemicals in our data base, such confounding remains a potential problem.

With one exception (Analysis 43), the analyses that incorporate alternative approaches to these components are relatively poor methods of human risk prediction; the predictive power and good correlation noted for Analysis 30 are diminished by altering one component. It seems likely that the high degree of chemical-specific change (lack of consistency) is responsible. That is not to say that some degree of chemical-specificity is not desired. One would like the RRDs that are too large (in Analysis 30) to be reduced and those that are too small to be increased. At this point, however, we have not identified approaches (or combinations of approaches to different components) that do this. Analysis 43, which selects total tumor-bearing animals as its endpoint is the exception; it is included in the recommended set despite the

large dispersion factor, 39.6, associated with the change in choice of endpoint.

A corollary of these observations is that these highly uncertain components -- related to length of observation and dosing, route of exposure, carcinogenic responses to use, and species to use -- deserve much more investigation (certainly more than choice of dose units for extrapolation). The goals of such an investigation include elucidation of the reasons behind the observed changes in RRDs and identification of new approaches that would produce the desired changes in RRDs, that is, ones that improve the predictiveness of the bioassay analyses. Potentially useful studies of the high degree of chemical-specific changes may start with identification of groups of chemicals (e.g., aromatic hydrocarbons, epigenetic carcinogens, early-stage carcinogens, etc.) and examination of patterns within the groups. For some components, notably the one associated with choice of species, other considerations, such as pharmacokinetic or genetic differences, may need to be examined. The empirical approach adopted for the present investigation may not be sufficient to explicate all of the changes seen. But above all else, availability of good data sets presenting information sufficient for studying these components with a minimum of confounding is essential.

OPTIONS FOR PRESENTING A RANGE OF RISK ESTIMATES

In this section we discuss three options for presenting a range of risk estimates suggested by the data. These options are derived from the five recommended analyses discussed in the previous section. Option 1 requires selection of a single analysis method from among the five, while Options 2 and 3 involve combining results from more than one analysis.

Regardless of the option selected, it seems reasonable to screen the data that are going to be used. The correlation analysis indicates that data screening improves the correlations in general. Consequently, a process akin to the sieve that has been defined here, one that selects the best of the available bioassays, is recommended. In applications to

a single chemical, a less automated, more customized procedure could be applied. On the other hand, if a consistent and uniform approach is desired for many chemicals, some automated sieve may be preferable.

Option 1

This option involves selecting one from among the five analyses discussed in the previous section. The selected analysis method is *applied to each of the eligible data sets, the median of the resulting* lower bound estimates is used as the predictor, and the conversion factor, 10^C (cf. Table 3-1), is applied to the predictor to correct for bias. The resulting estimate is multiplied and divided by the residual uncertainty factor (cf. Table 3-1) and the resulting range of RRDs is the desired range. Analysis-specific results are shown in Table 3-4 for the twenty-one chemicals in the data base for which human data are not available. Any one of the five intervals displayed for each chemical can be used to represent the range of risk estimates. Note that for several chemicals, it was not possible to apply Analysis 43.

Options 2 and 3

For these options, all five analyses must be performed, using the appropriate endpoints and dose units for extrapolation. For each analysis, select as the predictor the median of the lower bounds resulting from the analysis and apply the corresponding conversion factors. The values obtained in this manner represent the results of the methods of bioassay analysis that appear to be most appropriate for estimating human risk and form the basis for determining the range of those human estimates consistent with the data.

It is always possible to determine estimates via Analyses 30, 31, 45, and 47. It may, however, be the case that Analysis 43 cannot be completed given the data available (*cf.* Table 3-2 in which several chemicals lack estimates associated with this analysis). When this occurs, additional uncertainty is associated with the risk estimates: the full characterization of the range of estimates consistent with the recommended analyses is not possible. To account for this, one may wish to impute values for the missing estimates. The component-specific

uncertainty analysis, with its dispersion factor, provides the means to do so.

Analysis 43 is a single-component variant of Analysis 30. The histogram associated with this variation (Figure 2-52) indicates the mode lies between 0.8 and 1.25 (the geometric mean of these values being 1.0). The RRDs for Analysis 43 are imputed by taking the RRDs from Analysis 30 and multiplying them by a factor indicating the average ratio of the RRD pairs. We have used the geometric mean of the interval which contains the mode, i.e. 1.0. (Another reasonable factor could be based on the median ratio.) In doing this, the uncertainty is increased (reflecting the uncertainty due to lack of the complete ensemble of results) which is estimated by the dispersion factor. The imputed Analysis 43 results are multiplied and divided by the dispersion factor (39.6) since that factor is the average amount by which the ratios differ from the mode. The imputation of predictions for Analysis 43 is completed by applying the conversion factors for Analysis 43 just as if the estimates were not imputed.

At this stage, the assessment (i.e. prediction) of risk is completed. One has derived the best predictions of human RRDs that are possible from the data available: for an analysis that could be performed, a short interval (derived from the range of conversion factors pertinent to that method) of predictions is available, whereas for an analysis whose results have had to be imputed a generally much wider interval of predictions is the best that can be obtained. However, because of the variability characterized by the residual uncertainty factor, these intervals are not sufficient indicators of the range of risk estimates consistent with the data. The converted predictions must be multiplied and divided by the residual uncertainty factors to derive upper and lower uncertainty bounds for the risk estimates from each analysis method.

Since the recommended set of analyses contains methods that are good with respect to prediction of human risk, the ranges of estimates associated with those analyses characterize the human RRDs for the chemical in question. The ranges extending from the lower to the upper uncertainty bounds for each individual method can be considered as self-

contained results (this is Option 1) or they may be considered as a whole to present overall ranges of risks. Two lines of reasoning dictate how this might be accomplished. First (Option 2), one may reason as follows: to be most certain of including the true RRD in the overall range, one must consider each analysis (since the best one for any given chemical is not known) and characterize the range of estimates by the interval from the smallest lower bound to the largest upper bound (the "full range"). On the other hand (Option 3), one might argue that any of these methods is adequate and that the range of human estimates is suitably represented by the estimates from the method(s) that are most consistent with the entire ensemble of results. In this context, consistency can only be determined modulo the degree of uncertainty. [This is analogous to a statistical argument concerning the difference between point estimates, for example, which can only be resolved to the extent that the statistical variability allows.] Consequently, the Option 3 characterization of the range of risk estimates is defined as the union of the intervals from the lower to the upper bounds associated with some subset of the analyses such that the union contains the predictions from all the analyses (i.e. the values, like those in Table 3-3, that have been adjusted by the conversion factors but have not had the residual uncertainty factors applied) and is the smallest union satisfying that condition. This is a reasonable representation of the range of estimates consistent with the results from all recommended analyses, given our present degree of uncertainty. We will call it the smallest consistent range.

Comparison of Options

All three options presented define ranges of estimates by utilizing one or more of the methods that have been shown empirically in this investigation to do well with respect to prediction of human risk. Moreover, they all incorporate quantitatively those aspects of uncertainty that are summarized by the residual uncertainty factor. Several of the advantages and disadvantages of the options are discussed below.

Option 1 requires analysis of the bioassay data by a single method only. The selection of the single method may be somewhat problematical,

however. It has been argued that the ability of the analyses in the recommended set to predict human risk is not clearly distinguishable by the empirical approach adopted for the present investigation. Nevertheless, other factors, based on toxicological considerations for example, may dictate the choice of one of the analyses methods. In that case, there is no question about the method that should underlie Option 1.

It is hard to conceive of other factors that could clearly dictate the choice of a single analysis method, however. Aside from Analysis 43, all the analyses in the recommended set use exactly the same experiments and carcinogenic responses to estimate risk. If Analysis 43 is deemed inappropriate because it uses total tumor-bearing animals, for instance, one is left with four other methods one of which must, *a priori*, supply the range of risk estimates, if one follows the procedure of Option 1. *A priori* selection of a method may suit regulatory purposes very well.

Options 2 and 3, however, consider the intervals of estimates derived from all of the methods. No *a priori* decision is made about the particular method to use. Rather, the results of all the methods are examined for consistency and the summary range of estimates reflects that consistency as well as the analysis-specific uncertainty. (In this sense, these options reflect across-method uncertainty in addition to within-method uncertainties.) Greater consistency across analysis methods yields smaller ranges. Of course, the overall range is no smaller than the smallest range associated with any given method (which, given the conversion and residual uncertainty factors, must be from Analysis 45); an overall range can reflect no more certainty than the method with least uncertainty.

It may be the case that the full range estimated via Option 2 overestimates uncertainty. It is true that inclusion of more analysis methods in the preferred set can never diminish the Option 2 range. So, for example, should further investigation reveal other analysis methods warranting inclusion in the recommended set, their inclusion could not shrink the range determined by Option 2 and the current set of analyses. Furthermore, no particular use is made of the analyses with least residual uncertainty. If all analyses predicted the same RfDs, the method with the largest uncertainty factor, not the one with the

smallest factor, determines the full range.

Option 3 does not share these disadvantages with Option 2. Because the third option selects the smallest range that is consistent with all the predictions, priority is given to the methods with least uncertainty. Moreover, it is entirely possible that additional methods could reduce the smallest consistent range, even if they have larger uncertainty factors, if the added methods "cover" more of the original predictions. In this manner, additional information of comparable quality (i.e. as good in terms of predicting human risk) can refine our estimates of human health effects.

At first glance, it appears that the necessity of imputing values when particular analyses cannot be performed is a major disadvantage of Options 2 and 3. Indeed, the need to impute adds greatly to the uncertainty and may provide some justification for dropping from the recommended set those analyses for which imputation may be required. (Note that Options 2 and 3 are equally applicable no matter how many analysis methods are considered.) It must be emphasized, however, that the problem with imputation is not a methodological one, i.e. there is nothing inherent in Options 2 or 3 that makes them suffer from this difficulty. (In fact, Option 1 would have the same difficulty if the single method selected by that option was 43.) The increased uncertainty that results from imputation is caused by inadequacies in data reporting or data dissemination. If complete results, especially those allowing definition of the responses need, total tumor-bearing animals or the combination of significant responses, were available, then no imputation would be required and uncertainty caused by lack of data considerably reduced. The need to impute values is not a legitimate criticism of Option 2 or Option 3.

In closing this comparison, it should be noted that the uncertainties discussed in connection with all three of the options may not completely characterize uncertainty. In particular, there is uncertainty about the shape of the dose-response curve that is not quantitatively estimated. Moreover, the residual uncertainty factors represent only that part of the uncertainty that is not explainable by uncertainty in the human estimates. The uncertainties not quantified fall outside the definition

of those that are particularly associated with any given analysis method, but they should be borne in mind when considering the ranges of estimates of human risk derived from any option.

Examples

The chemicals included in this investigation but lacking epidemiological data sufficient for quantitative risk assessment (Table 2-2) can serve as examples of the application of the three options. Tables 3-2 through 3-4 present the median lower bound RRDs, the converted predictions, and intervals of estimates derived by application of the analysis-specific residual uncertainty factors, respectively, that underlie the application of these options. As mentioned earlier, any one column of Table 3-4 represents the output from Option 1. Table 3-5 contains the two overall ranges from Options 2 and 3.

Several interesting features are illustrated by the ranges in Table 3-5. First, as an example of the procedure for determining the smallest consistent range (Option 3), consider acrylonitrile. The interval (Table 3-4) associated with Analysis 43 does not intersect with the intervals derived for the other analyses. Consequently, the smallest consistent range is the union of the Analysis 43 interval and the smallest interval (from Analysis 47) that contains all the other predictions, as shown. For other chemicals (EDB, hexachlorobenzene), the intervals are not disjoint so that a more standard-looking interval of values is obtained. In any case, it is apparent that the smallest consistent range can provide improvement (in terms of a narrower range of estimates) over the full range (Option 2).

Second, imputation of values for Analysis 43 has been necessary in several instances. When this is so, the uncertainty bounds for the imputed values completely determine both ranges of estimates. It is easy to understand why this is the case with the Option 2 range: the uncertainty of the imputation carries with it a multiplicative factor of 39.6 that is used to determine the range of imputed values and which is much larger than any of the residual uncertainty factors. The reason why the imputed Analysis 43 uncertainty bounds determine the smallest consistent range is also linked to the imputation uncertainty, but

of those that are particularly associated with any given analysis method, but they should be borne in mind when considering the ranges of estimates of human risk derived from any option.

Examples

The chemicals included in this investigation but lacking epidemiological data sufficient for quantitative risk assessment (Table 2-2) can serve as examples of the application of the three options. Tables 3-2 through 3-4 present the median lower bound RRDs, the converted predictions, and intervals of estimates derived by application of the analysis-specific residual uncertainty factors, respectively, that underlie the application of these options. As mentioned earlier, any one column of Table 3-4 represents the output from Option 1. Table 3-5 contains the two overall ranges from Options 2 and 3.

Several interesting features are illustrated by the ranges in Table 3-5. First, as an example of the procedure for determining the smallest consistent range (Option 3), consider acrylonitrile. The interval (Table 3-4) associated with Analysis 43 does not intersect with the intervals derived for the other analyses. Consequently, the smallest consistent range is the union of the Analysis 43 interval and the smallest interval (from Analysis 47) that contains all the other predictions, as shown. For other chemicals (EDB, hexachlorobenzene), the intervals are not disjoint so that a more standard-looking interval of values is obtained. In any case, it is apparent that the smallest consistent range can provide improvement (in terms of a narrower range of estimates) over the full range (Option 2).

Second, imputation of values for Analysis 43 has been necessary in several instances. When this is so, the uncertainty bounds for the imputed values completely determine both ranges of estimates. It is easy to understand why this is the case with the Option 2 range: the uncertainty of the imputation carries with it a multiplicative factor of 39.6 that is used to determine the range of imputed values and which is much larger than any of the residual uncertainty factors. The reason why the imputed Analysis 43 uncertainty bounds determine the smallest consistent range is also linked to the imputation uncertainty, but

involves other features of the data as well. In particular, note in Table 3-2 that the median lower bound estimates from Analysis 43 are generally smaller than those from the other analyses. Moreover, the conversion factors for Analysis 43 are 0.18 and 0.29; i.e. the converted values are even smaller than the raw values, whereas the factors for other analyses in the restricted recommended set are greater than or equal to one. (This also explains why, even when no imputation is necessary, Analysis 43 uncertainty bounds always determine the lower end of the smallest consistent range.) The difference in conversion factors and the reason why they tend to separate the predictions of Analysis 43 from those of the other analyses can be explained by reference to Figure 2-52. This histogram depicts the chemical-specific ratios of RRD estimates derived from Analysis 43 to those derived from Analysis 30. The six chemicals whose ratios are greater than 1.25 are only from the set that have epidemiological data suitable for estimating the conversion factors. Given that the conversion factor for Analysis 30 is roughly unity, these six chemicals, especially, have shifted the best fitting line to the right, decreasing the y-intercept, and entailing conversion factors substantially less than one. But, as already noted, the chemicals represented in Table 3-2 generally have ratios in Figure 2-52 that are less than 0.80. Hence the divergence of the RRD predictions. The dichotomy displayed in Figure 2-52 between those chemicals with suitable epidemiological data and those without is undoubtedly fortuitous. Nevertheless, it does show that a dispersion factor as large as 39.6 used in the context of the imputation of values is necessary to cover such occurrences. It is important to note that this added uncertainty is unnecessary: imputation is dictated solely by data availability (having the ability to define the total tumor-bearing animal response). Better data reporting procedures can substantially reduce the ranges of risk estimates.

For purposes of comparison, Table 3-6 presents the ranges of estimates that are obtained from Options 2 and 3 when Analysis 43 is not considered. This eliminates the wide ranges produced as a results of imputation. However, by simply ignoring a method of extrapolation that has been deemed to be of a value comparable to those of the other methods in the recommended set, these ranges may be too narrow to the extent that across-method uncertainty is underestimated. Certainly, the

ranges presented in Table 3-5 are to be preferred over those in Table 3-6 when no imputation is necessary. When Analysis 43 cannot be performed and imputation is necessary, it is not clear which range is more appropriate; those based on fewer analyses (Table 3-6) may be too narrow while those based on an *ad hoc* imputation procedure may be too wide. It bears repeating that this dilemma could be avoided entirely if some better means of data dissemination were to be found.

At present, there are no quantitative estimates of RRDs derived from the epidemiological literature to which these predictions can be compared. It might be possible to qualitatively compare the predictions to the epidemiology in a couple of ways. The predictions could be used to rank the chemicals in order of their RRDs (reverse order of their carcinogenic potencies). Another ordering could be based on a comparative examination of the epidemiology. The degree of correspondence of the two orders might provide information about the predictions. Of course, without quantitative estimates, the epidemiologically based ordering would be subject to considerable uncertainty in and of itself. A chemical-specific examination of the epidemiology might be useful in uncovering predictions that are way off the mark. Such a comparison would probably be quite crude and may be limited to identifying those chemicals for which the predictions (being finite) indicate carcinogenicity but the epidemiology indicates no carcinogenicity. Neither type of comparison has been undertaken for this project.

GENERAL CONSIDERATIONS AND MAJOR CONCLUSIONS

It is apparent that 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 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.

Certain features of this investigation must be borne in mind. Primary among these is the fact that the level of risk for which RRDs have been determined is 0.25. This value is a compromise between the need to use a value high enough to be fairly independent of the choice of dose-response model in the bioassay analyses and the desire not to greatly exceed the risk found in most epidemiologically studied cohorts. A risk level of 0.25 is higher than that which exists in most human exposure situations. While we would not expect some of the results to be altered if the investigation had utilized a different risk level (*e.g.* 10^{-6}), it is not certain that all the conclusions would remain the same. In particular, the evaluation of uncertainty in this report does not encompass that related to the shape of the dose-response curve. It may be worthwhile to check some of the results at lower levels of risk, although it must be noted that the increased uncertainty associated with the shape of the dose-response curve at low doses may make interpretation of results concerning other components of risk assessment difficult.

Also recall that the bioassay data, though extensive, is rather crude in many respects. We have already noted the problems associated with data deficiencies, mostly caused by incomplete reporting of results. Over and above that, however, the analyses performed did not use time-to-tumor data, i.e. a quantal model has been used to estimate RRDs. Time and data constraints dictated that choice, but it is of interest to determine if time-to-tumor analyses, which utilize more of the information obtained from a bioassay, could refine our results and conclusions.

It must be recalled that when several forms of a suspected carcinogen have been tested in animal bioassays, the results for all forms have been grouped together. This primarily influences the data and results for the metals. Since all forms are individually identified, it is possible to perform the analyses on each form separately. Of course, the number of experiments for the affected chemicals would be reduced. Moreover, it is often not known, even for substantiated human carcinogens, which particular moieties cause cancer or which are most potent.

Other reasonable approaches to the components of risk assessment could have been defined. Thus, for example, the component related to length of dosing did not have to include only two approaches, one including all experiments and the other including only those experiments lasting at least 90 percent of the standard length. Short experiments by themselves could have been studied. This may have led to an examination of the correction factor that has been used to adjust for short observation periods.

As discussed in Volume 1 of this report, the epidemiological data used in this study are of variable quality. The bounds determining ranges of exposure are somewhat arbitrary and, for each chemical, one cancer endpoint from a single study was selected to represent the range of RRD estimates and to be the target of the bioassay analyses. It would be interest to determine how robust our findings are with respect to these choices. Moreover, the pattern of exposure for which RRDs have been estimated (45 years of constant exposure starting at age 20) is not realistic for some of the study chemicals (such as DES (estrogen). This choice, too, is a compromise between the usual lifetime exposure administered in bioassays (that is, lifetime after start of exposure which may be several weeks after the birth of the test animals) and the less consistent exposures which humans encounter.

In the prediction analysis, three loss functions have been defined. None of them is the standard squared-error loss routinely applied, since the latter is clearly not appropriate for the estimates derived here. No statistical development of these loss functions exists to inform us about lack of fit, significance of differences in loss, etc. If a statistical underpinning did exist, it would be possible to use the loss functions in some capacity besides as ranking procedures and, thereby, to be able to better differentiate between the analysis methods and refine the conclusions.

Finally, only 55 distinct bioassay analysis methods have been defined. This is only a small fraction of those that could be considered, even fixing the approaches to the components at those defined here.

Despite the caveats just presented, the following major conclusions have emerged from the present investigation.

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

- In the majority of cases considered, analysis methods for bioassay data that utilize lower statistical confi- :e limits as predictors yield better predictions of human risk than do the same methods using maximum likelihood estimates.

- Analysis methods for bioassay 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 lower bound RRDs (assuming approaches to other risk assessment components are chosen appropriately.

- 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.

- 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.

- 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.

- 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.

- Evaluation of risk assessment methods should focus on the complete risk assessment process rather than on individual components.

- The data base and methods used in this study can provide a useful basis for the evaluation of various risk assessment methods.

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.

REFERENCE

1. Crump, K., Silvers, A., Ricci, P., and Wyzga, R. (1985). Inter-species Comparison for Carcinogenic Potency to Humans. Principles of Health Risk Assessment. Ricci, P. (ed.). Prentice Hall.

Table 3-1

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	1.0 - 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_{20}, 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.

Table 3-2

MEDIAN LOWER BOUND RRD ESTIMATES, BY CHEMICAL AND ANALYSIS METHOD[a]

Chemical	Analysis 30	31	43	45	47
Acrylonitrile	4.39	9.29E-1	1.01	3.57	4.39
Allyl Chloride	6.92E+1	1.02E+1	6.71E+1	7.27E+1	1.11E+2
4-Aminobiphenyl	2.17E+1	2.03	--[b]	2.42E+1	2.17E+1
Benzo(a)pyrene	5.21E-1	7.02E-2	4.80E-2	5.87E-1	5.21E-1
Carbon tetrachloride	3.10E+1	2.89	--	3.57E+1	3.10E+1
Chlordane	2.36	1.96E-1	1.56	1.99	4.43
3,3-Dichlorobenzidine	1.24E+1	2.62	5.59E-1	1.91E+1	1.24E+1
1,2-Dichloroethane	2.79E+1	4.62	1.28E+1	3.34E+1	4.31E+1
EDB	3.77	2.94E-1	2.96	3.29	4.82
Formaldehyde	1.88	3.09E-1	--	1.15	3.16
Hexachlorobenzene	1.30	2.00E-1	2.84E-1	1.30	2.48
Hydrazine	1.8[illegible]	1.74E-1	--	1.87	9.15
Mustard Gas	1.40E-7	1.31E-8	--	1.40E-7	1.40E-7
Lead	6.14	1.30	6.09	6.14	6.14
2-Napthylamine	1.20E+1	2.54	--	1.20E+1	1.20E+1
NTA	6.24E+2	5.66E+1	2.23E+1	6.24E+2	6.97E+2
2,4,6-Trichlorophenol	1.73E+2	1.61E+1	--	2.11E+2	1.90E+2
TCDD	7.32E-5	1.55E-5	2.56E-5	6.87E-5	9.05E-5
Tetrachloroethylene	9.22E+1	8.25	8.06E+1	8.70E+1	1.13E+2
Toxaphene	2.58	1.89E-1	1.37	4.23	4.74
Vinylidene chloride	1.34	2.45E-1	7.26E-1	5.56E-1	2.34

[a]The full sieve was used to screen the data; the estimates have not been adjusted by the appropriate conversion factors.

[b]A "--" indicates that the data were not available to apply the method to the chemical.

Table 3-3

RRD PREDICTIONS[a], BY CHEMICAL AND ANALYSIS METHOD

Chemical	Analysis 30	31	43	45	47
Acrylonitrile	[4.74, 7.46][b]	[7.85, 1.12E+1]	[1.82E-1, 2.93E-1]	[4.14, 6.07]	[4.39, 7.46]
Allyl Chloride	[7.47E+1, 1.18E+2]	[8.62E+1, 1.23E+2]	[1.21E+1, 1.95E+1]	[8.43E+1, 1.24E+2]	[1.11E+2, 1.89E+2]
4-Aminobiphenyl	[2.34E+1, 3.69E+1]	[1.72E+1, 2.44E+1]	--[c]	[2.81E+1, 4.11E+1]	[2.17E+1, 3.69E+1]
Benzo(a)pyrene	[5.63E-1, 8.86E-1]	[5.93E-1, 8.44E-1]	[8.64E-3, 1.39E-2]	[6.81E-1, 9.98E-1]	[5.21E-1, 8.86E-1]
Carbon Tetrachloride	[3.35E+1, 5.27E+1]	[2.44E+1, 3.47E+1]	--	[4.14E+1, 6.07E+1]	[3.10E+1, 5.27E+1]
Chlordane	[2.55, 4.01]	[1.66, 2.36]	[2.81E-1, 4.52E-1]	[2.31, 3.38]	[4.43, 7.53]
3,3-Dichlorobenzidine	[1.34E+1, 2.11E+1]	[2.21E+1, 3.15E+1]	[1.01E-1, 1.62E-1]	[2.22E+1, 3.25E+1]	[1.24E+1, 2.11E+1]
1,2-Dichloroethane	[3.01E+1, 4.74E+1]	[3.90E+1, 5.55E+1]	[2.30, 3.71]	[3.87E+1, 5.68E+1]	[4.31E+1, 7.33E+1]
EDB	[4.07, 6.41]	[2.48, 3.53]	[5.33E-1, 8.58E-1]	[3.82, 5.59]	[4.82, 8.19]
Formaldehyde	[2.03, 3.20]	[2.61, 3.71]	--	[1.34, 1.96]	[3.16, 5.37]
Hexachlorobenzene	[1.40, 2.21]	[1.69, 2.40]	[5.11E-2, 8.24E-1]	[1.51, 2.21]	[2.48, 4.22]
Hydrazine	[2.02, 3.18]	[1.47, 2.09]	--	[2.17, 3.18]	[9.15, 1.56E+1]
Mustard Gas	[1.51E-7, 2.38E-7]	[1.11E-7, 1.57E-7]	--	[1.62E-7, 2.38E-7]	[1.40E-7, 2.38E-7]
Lead	[6.63, 1.04E+1]	[1.10E+1, 1.56E+1]	[1.10, 1.77]	[7.12, 1.04E+1]	[6.14, 1.04E+1]
2-Napthylamine	[1.30E+1, 2.04E+1]	[2.15E+1, 3.05E+1]	--	[1.39E+1, 2.04E+1]	[1.20E+1, 2.04E+1]
NTA	[6.74E+2, 1.06E+3]	[4.78E+2, 6.80E+2]	[4.01, 6.47]	[7.24E+2, 1.06E+3]	[6.97E+2, 1.18E+3]
2,4,6-Trichlorophenol	[1.87E+2, 2.94E+2]	[1.36E+2, 1.94E+2]	--	[2.45E+2, 3.59E+2]	[1.90E+2, 3.23E+2]
TCDD	[7.91E-5, 1.24E-4]	[1.31E-4, 1.86E-4]	[4.61E-6, 7.42E-6]	[7.97E-5, 1.18E-4]	[9.05E-5, 1.54E-4]
Tetrachloroethylene	[9.96E+1, 1.57E+2]	[6.97E+1, 9.92E+1]	[1.45E+1, 2.34E+1]	[1.01E+2, 1.48E+2]	[1.13E+2, 1.92E+2]
Toxaphene	[2.79, 4.39]	[1.60, 2.27]	[2.47E-1, 3.97E-1]	[4.91, 7.19]	[4.74, 8.06]
Vinylidene Chloride	[1.45, 2.28]	[2.07, 2.94]	[1.31E-1, 2.11E-1]	[6.45E-1, 9.45E-1]	[2.34, 3.98]

[a] The predictions are derived from the values in Table 3-2 by application of the appropriate conversion factors.
[b] The intervals are the result of applying the two conversion factors given in Table 3-1 for each analysis method.
[c] A "--" indicates that the data were not available to apply the method to the chemical.

Table 3-4

UNCERTAINTY INTERVALS FOR RRD PREDICTIONS[a], BY CHEMICAL AND ANALYSIS METHOD

Chemical	Analysis 30	31	43	45	47
Acrylonitrile	[2.37, 1.49E+1]	[3.93, 2.24E+1]	[6.50E-2, 8.20E-1]	[2.44, 1.03E+1]	[2.44, 1.34E+1]
Allyl Chloride	[3.74E+1, 2.36E+2]	[4.31E+1, 2.46E+2]	[4.32, 5.46E+1]	[4.96E+1, 2.11E+2]	[6.17E+1, 3.40E+2]
4-Aminobiphenyl	[1.17E+1, 7.38E+1]	[8.60, 4.88E+1]	--[b]	[1.65E+1, 6.99E+1]	[1.21E+1, 6.64E+1]
Benzo(a)pyrene	[2.82E-1, 1.77]	[2.96E-1, 1.69]	[3.09E-3, 3.89E-2]	[4.01E-1, 1.70]	[2.89E-1, 1.59]
Carbon Tetrachloride	[1.68E+1, 1.05E+2]	[1.22E+1, 6.94E+1]	--	[2.44E+1, 1.03E+2]	[1.72E+1, 9.49E+1]
Chlordane	[1.28, 8.02]	[8.30E-1, 4.72]	[1.00E-1, 1.27]	[1.36, 5.75]	[2.46, 1.36E+1]
3,3-Dichlorobenzidine	[6.70, 4.22E+1]	[1.10E+1, 6.30E+1]	[3.61E-2, 4.54E-1]	[1.31E+1, 5.53E+1]	[6.89, 3.80E+1]
1,2-Dichloroethane	[1.51E+1, 9.48E+1]	[1.95E+1, 1.11E+2]	[8.21E-1, 1.04E+1]	[2.28E+1, 9.66E+1]	[2.39E+1, 1.32E+2]
EDB	[2.04, 1.28E+1]	[1.24, 7.06]	[1.90E-1, 2.40]	[2.25, 9.50]	[2.68, 1.47E+1]
Formaldehyde	[1.02, 6.40]	[1.30, 7.42]	--	[7.82E-1, 3.33]	[1.76, 9.67]
Hexachlorobenzene	[7.00E-1, 4.42]	[8.45E-1, 4.80]	[1.82E-2, 2.31]	[8.88E-1, 3.76]	[1.38, 7.60]
Hydrazine	[1.01, 6.36]	[7.35E-1, 4.18]	--	[1.28, 5.41]	[5.08, 2.81E+1]
Mustard Gas	[7.55E-8, 4.76E-7]	[5.55E-8, 3.14E-7]	--	[9.53E-8, 4.05E-7]	[7.78E-8, 4.28E-7]
Lead	[3.32, 2.08E+1]	[5.50, 3.12E+1]	[3.93E-1, 4.96]	[4.19, 1.77E+1]	[3.41, 1.87E+1]
2-Napthylamine	[6.50, 4.08E+1]	[1.08E+1, 6.10E+1]	--	[8.18, 3.47E+1]	[6.67, 3.67E+1]
NTA	[3.37E+2, 2.12E+3]	[2.39E+2, 1.36E+3]	[1.43, 1.81E+1]	[4.26E+2, 1.80E+3]	[3.87E+2, 2.12E+3]
2,4,6-Trichlorophenol	[9.35E+1, 5.88E+2]	[6.80E+1, 3.88E+2]	--	[1.44E+2, 6.10E+2]	[1.06E+2, 5.81E+2]
TCDD	[3.96E-5, 2.48E-4]	[6.55E-5, 3.72E-4]	[1.65E-6, 2.08E-5]	[4.69E-5, 2.01E-4]	[5.03E-5, 2.77E-4]]
Tetrachloroethylene	[4.98E+1, 3.14E+2]	[3.48E+1, 1.98E+2]	[5.18, 6.55E+1]	[5.94E+1, 2.52E+2]	[6.28E+1, 3.46E+2]
Toxaphene	[1.40, 8.78]	[8.00E-1, 4.54]	[8.82E-2, 1.11]	[2.89, 1.22E+1]	[2.63, 1.45E+1]
Vinylidene Chloride	[7.25E-1, 2.56]	[1.04, 5.88]	[4.68E-2, 5.91E-1]	[3.79E-1, 1.61]	[1.30, 7.16]

[a]The intervals are derived from the values in Table 3-3 by application of the residual uncertainty factors (cf. Table 3-1).

[b]A "--" indicates that the data were not available to apply the method to the chemical.

Table 3-5

RANGES OF HUMAN RRDS DERIVED FROM THE RECOMMENDED SET OF ANALYSES[a]

Chemical	Option 2: Full Range[b]	Option 3: Smallest Consistent Range[c]
Acrylonitrile	[6.50E-2, 2.24E+1]	[6.50E-2, 8.20E-1] U [2.44, 1.34E+1] (43,47)[d]
Allyl Chloride	[4.32, 3.40E+2]	[4.32, 2.11E+2] (43, 45)
4-Aminobiphenyl	[3.52E-2, 6.98E+2]*	[3.52E-2, 6.98E+2] (43)
Benzo(a)pyrene	[3.09E-3, 1.77]	[3.09E-3, 3.89E-2] U [4.01E-1, 1.70] (43, 45)
Carbon Tetrachloride	[5.03E-2, 9.97E+2]*	[5.03E-2, 9.97E+2] (43)
Chlordane	[1.00E-1, 1.36E+1]	[1.00E-1, 1.27] U [1.28, 8.02] (43, 30)
3,3-Dichlorobenzidine	[3.61E-2, 6.30E+1]	[3.61E-2, 4.54E-1] U [6.89, 3 80E+1] (43, 47)
1,2-Dichloroethane	[8.21E-1, 1.32E+2]	[8.21E-1, 1.04E+1] U [2.28E+1, 9.66E+1] (43, 45)
EDB	[1.90E-1, 1.47E+1]*	[1.90E-1, 9.50] (43, 45)
Formaldehyde	[3.05E-3, 6.05E+1]*	[3.05E-3, 6.05E+1] (43)
Hexachlorobenzene	[1.82E-2, 7.60]	[1.82E-2, 4.42] (43, 30)
Hydrazine	[3.04E-3, 6.01E+1]*	[3.04E-3, 6.01E+1] (43)
Mustard Gas	[2.27E-10, 4.50E-6]*	[2.27E-10, 4.50E-6] (43)
Lead	[3.93E-1, 3.12E+1]	[3.93E-1, 1.77E+1] (43, 45)
2-Napthylamine	[1.95E-2, 3.86E+2]*	[1.95E-2, 3.86E+2] (43)
NTA	[1.43, 2.12E+3]	[1.43, 1.81E+1] U [4.26E+2, 1.80E+3] (43, 45)
2,4,6-Trichlorophenol	[2.81E-1, 5.56E+3]*	[2.81E-1, 5.56E+3] (43)
TCDD	[1.65E-6, 3.72E-4]	[1.65E-6, 2.08E-5] U [4.69E-5, 2.01E-4] (43, 45)
Tetrachloroethylene	[5.18, 3.46E+2]	[5.18, 1.98E+2] (43, 31)
Toxaphene	[8.82E-2, 1.45E+1]	[8.82E-2, 1.11] U [1.40, 8.78] (43, 30)
Vinylidene Chloride	[4.68E-2, 7.16]	[4.68E-2, 5.88] (43, 45, 31)

[a]Values of RRDs are in mg/kg/day.

[b]The full range extends from the smallest lower bound to the largest upper bound among analyses in the recommended set.

[c]The smallest consistent range is the union of intervals from analyses in the recommended set such that the union includes all predictions (from Table 3-3) and is the smallest union that does so.

[d]When the union is of disjoint parts, both parts are shown, connected by the union symbol, "U". In parentheses are the analyses whose union defines the smallest consistent range.

*An asterisk marks those intervals that are the result of imputing values for Analysis 43.

Table 3-6

RANGES OF HUMAN RRDS DERIVED FROM THE RECOMMENDED SET OF ANALYSES IGNORING ANALYSIS 43[a]

Chemical	Option 2: Full Range[b]	Option 3: Smallest Consistent Range[c]
Acrylonitrile	[2.37, 2.24E+1]	[2.44, 1.34E+1] (47)
Allyl Chloride	[3.74E+1, 3.40E+2]	[4.96E+1, 2.11E+2] (45)
4-Aminobiphenyl	[8.60, 7.38E+1]	[1.65E+1, 6.99E+1] (45)
Benzo(a)pyrene	[2.82E-1, 1.77]	[4.01E-1, 1.70] (45)
Carbon Tetrachloride	[1.22E+1, 1.05E+2]	[2.44E+1, 1.03E+2] (45)
Chlordane	[8.30E-1, 1.36E+1]	[1.28, 8.02] (30)
3,3-Dichlorobenzidine	[6.70, 6.30E+1]	[6.89, 3.80E+1] (47)
1,2-Dichloroethane	[1.51E+1, 1.32E+2]	[2.28E+1, 9.66E+1] (45)
EDB	[1.24, 1.47E+1]	[2.25, 9.50] (45)
Formaldehyde	[7.82E-1, 9.67]	[1.30, 7.42] (31)
Hexachlorobenzene	[7.00E-1, 7.60]	[1.38, 7.60] (47)
Hydrazine	[7.35E-1, 2.81E+1]	[1.28, 2.81E+1] (45, 47)
Mustard Gas	[5.55E-8, 4.76E-7]	[9.53E-8, 4.05E-7] (45)
Lead	[3.32, 3.12E+1]	[4.19, 1.77E+1] (45)
2-Napthylamine	[6.50, 6.10E+1]	[8.18, 3.47E+1] (45)
NTA	[2.39E+2, 2.12E+3]	[4.26E+2, 1.80E+3] (45)
2,4,6-Trichlorophenol	[6.80E+1, 6.10E+2]	[1.06E+2, 5.81E+2] (47)
TCDD	[3.96E-5, 3.72E-4]	[4.69E-5, 2.01E-4] (45)
Tetrachloroethylene	[3.48E+1, 3.46E+2]	[5.94E+1, 2.52E+2] (45)
Toxaphene	[8.00E-1, 1.45E+1]	[1.40, 8.78] (30)
Vinylidene Chloride	[3.79E-1, 7.16]	[3.79E-1, 5.88] (31, 45)

[a]Values of RRDs are in mg/kg/day.

[b]The full range extends from the smallest lower bound to the largest upper bound among analyses in the recommended set.

[c]The smallest consistent range is the union of intervals from analyses in the recommended set such that the union includes all predictions (from Table 3-3) and is the smallest union that does so.

Figure 3-1

RRD Plot For Analysis 45 (All Routes, Average over Sex)