

US EPA ARCHIVE DOCUMENT

# Appendix 1

**Preliminary Cumulative  
Hazard and Dose-Response Assessment for  
Organophosphorus Pesticides:**

**Determination of Relative Potency  
and Points of Departure  
for Cholinesterase Inhibition**

## Appendix 1.

### Summary of Approach

The goal of the statistical methods was to estimate a quantity that would be proportional to the potency of each chemical, along with confidence intervals. The data for this study were in the form of dose-response studies which measured the effect of different concentrations of OP pesticides on cholinesterase activities in brain, red blood cells, and plasma. The mean and standard deviation of cholinesterase activity, and number of animals examined were available for several dosages in each data set. Females and males were analyzed separately in each study. Studies were nested: for each chemical there were several groups of studies, each with a different MRID; within MRID, one or more studies were conducted, each with measurements taken for several durations of exposure. It is possible that potency increases to an asymptotic value as exposure duration increases. Studies with the same MRID were conducted in the same laboratory. Thus several steps were required to analyze the collection of data sets for each compartment x sex combination for each OP pesticide:

1. Adequately model the relationship between dose and cholinesterase activity for each individual study, and estimate the absolute potency for that study.
2. Determine which exposure durations are likely to be long enough that potencies are close to the steady-state value
3. Combine potency estimates within MRID, resulting in a single estimate for each MRID, with standard error.
4. Combine potency estimates across MRIDs, resulting in a single potency estimate for the chemical x compartment x sex combination
5. Compute the relative potency by dividing all the potency measures within sex and compartment by that of the index compound, and estimate the standard error of the result.
6. Compute BMD and BMDLs for each data set for the index compound, and combine the estimates using the same methods as for potency.

The following sections describe each of these steps in greater detail.

## Dose-Response Modeling

The cumulative dose-response assessment for this analysis is based on the Relative Potency Factor (RPF) methodology (U.S. EPA, 2001). The RPF approach assumes that the dose-response for a combination of exposures, at least for relatively low levels of exposure, are dose-additive. That is, the response  $y$  for a combination of exposures  $D_i$  would be :

$$y = f\left(\sum_i m_i D_i; b\right) \text{ (1) (eq.1)}$$

where  $m_i$  is the absolute potency of the  $i^{th}$  exposure, and  $b$  is a vector of parameters of the dose-response function whose values are the same for all chemicals. In practice, we select one index chemical, (call it  $I$ ), and express all potencies as ratios to that of the index chemical;  $R_i = m_i/m_I$ . Then the expected response  $y$  for a combination of exposures  $D_i$  would be:

$$y = f\left(m_I \sum_i R_i D_i; b\right) \text{ (2) (eq. 2)}$$

The response is the same as it would be if a dose of the index compound equal to  $\sum_i R_i D_i$  had been given. Furthermore, any combination of dosages that give the same overall weighted sum should result in the same response. For example, suppose the following potencies apply to four chemicals:

Chemical	Potency (mg/kg/day) <sup>-1</sup> ( $m_i$ )	Relative Potency ( $R_i$ )
A	0.346	0.309
B	0.0082	0.0073
C	1.21	1.08
D [index]	1.12	1.00

Suppose two sets of doses:

Chemical	Set 1	Set 2
A	0.0137	0.00324
B	1.17	0.00162
C	0.0391	0.0486
D	0.0391	0.0405
$\sum D_i$	1.26	0.094
$\sum R_i D_i$	0.094	0.094

Even though the two sets of doses are very different, and result in quite different total doses (1.26 mg/kg/day in set 1, 0.094 mg/kg/day in set 2), the expected responses are identical. If  $BMDL_1$  is the lower confidence limit for a specified response (the benchmark response, BMR, in this case, a 10% reduction in cholinesterase activity) for the index chemical, then the response to the combination of exposures represented by the  $D_i$  is likely to be smaller than the BMR if  $\sum R_i D_i$  is smaller than  $BMDL_1$ .

In this analysis, the dose-response function had to accommodate two important features of the data. First, since the results of multiple studies, perhaps carried out in different laboratories and at different times, and even sometimes reporting AChE activities in different units, it seemed prudent to express activity at a given dosage as a fraction of control activity. Implicit in this formulation is the idea that the among-data-set component of variability follows a multiplicative error distribution. Second, it was observed that, as doses increased, AChE activity in quite a few data sets approached a lower non-zero asymptote. This asymptote varied among data sets, chemicals, and compartments. These two properties of the data were accommodated by fitting the model:

$$y = B + (A - B)e^{-mD} \quad (\text{eq. 3})$$

where  $A$  is the background level of cholinesterase activity, and  $B$  is the limit of AChE activity for large doses. In practice, in some cases, it was not possible to estimate all three parameters for a data set, or this model failed to adequately fit the data. In these cases,  $B$  was often set to 0 and higher doses were dropped from consideration, as will be described below in more detail. To force all parameter estimates to be non-negative, what was actually estimated in each case was the natural logarithm of the parameter. So, for example, the parameters estimated were  $\ln A$ ,  $\ln B$ , and  $\ln m$ , with  $A = e^{\ln A}$ , for example. Standard errors of the log-parameters reported by the statistical software were transformed using the delta method to be on the appropriate scale. Parameters for this model for each data set were estimated using generalized nonlinear least squares (GNLS). GNLS was selected because it does not require distributional

assumptions about the individual data (which could not be checked, since only summary statistics were available), unlike maximum likelihood (ML) estimation, and it is more robust than ML (Davidian and Giltinan, 1995, pp. 31, 39, 59). GNLS is an extension of weighted nonlinear least squares. In weighted nonlinear least squares, the parameter estimates are the values for the components of  $\hat{a}$  (that is, in this case, the vector  $[A, B, m]'$ ) that minimize the weighted sum of squares

$SS = \sum w_i \left( y_i - f(Dose_i; b) \right)^2$ , where  $y_i$  is the observed activity,  $f()$  is the function described above in eq. 1-3, and the weights  $w_i$  are already known and are proportional to the reciprocal of the variance. In GNLS, the weights are taken to be a known function of the mean. In the OP data for this analysis, the variance among observations within dose groups is approximately proportional to the square of the mean of the group (see fig [1]), so regression weights based on the square of the estimated mean were used to improve the efficiency of the estimates over what would be obtained with unweighted regression.

Goodness of fit of each fitted model to the corresponding data was quantified through a global test of goodness-of-fit, specifically the Pearson chi-squared statistic, through visual inspection of graphs, and through examining tables of standardized residuals. The Pearson chi-squared statistic is  $\chi^2 = \sum_i \left( \bar{y}_i - f(Dose_i; \hat{b}) \right)^2 / (\hat{s}^2 / n_i)$ ,

where  $i$  indexes dose groups. If the model is true, then  $\chi^2$  will be distributed (approximately) as Chi-squared with degrees of freedom equal to the number of dose groups minus the number of parameters estimated.

The process for getting the final parameter estimates for a data set was as follows:

1. Estimate A, B, and m using GNLS for all dose groups in the dataset.
2. If the P-value for the  $\chi^2$  statistic was greater than 0.05, then the result is the estimate used. Otherwise (that is, if the P-value was less than 0.05, or no estimates resulted because the model did not converge), set B to zero, and try again with all the data.
3. If the P-value is still less than 0.05, or there is no model fit at all, then sequentially drop the remaining highest dose and refit the model with B set to zero until either the P-value exceeds 0.05, or there are only three doses remaining.

## Identifying Steady State

The sets of data for each chemical, sex, and compartment included a range of exposure durations. To determine which data sets had a sufficiently long exposure duration that potency was no longer changing with time, we regressed the estimated potency against exposure duration, weighting observations by the reciprocal of the squared estimated standard errors. The data for the shortest remaining exposure duration in the data set was repeatedly removed until a data set was derived in which the slope of potency versus time was not significant (that is, the P-value exceeded 0.05). In any case, the process was stopped when only three distinct durations remained. After a first pass through the data, a single duration was identified such that exposures exceeding that duration rarely showed a significant increase with time; all exposures less than that duration were removed from further consideration.

## Combining Potency Estimates and Computing Relative Potencies

Potency estimates were nested in two levels for each chemical  $\times$  sex  $\times$  compartment: generally several data sets, representing a range of exposure durations and some duplication within each MRID, and several MRIDs. Since the data sets representing exposure durations at which steady state had not been achieved were deleted from the study before this stage, it was reasonable to model the individual potency estimates as coming from a nested hierarchical sampling scheme:

1. First, assume there is an overall mean potency for a given chemical  $\times$  sex  $\times$  compartment combination; the procedure described below is designed to estimate this quantity, which will be used for computing the relative potency.
2. Because of small differences in husbandry, analytic procedures, and other laboratory procedures, the potency realized in MRIDs may vary among MRIDs. Model this as sampling a MRID-specific potency from a distribution centered about the overall mean potency, with variance  $S_{MRID}^2$ .
3. Again, because of differences among the studies that contributed the data sets within a given MRID, potencies realized within a given data set may vary among data sets within an MRID. Model this as sampling a data-set-specific potency from a distribution centered about the MRID-specific potency, with variance  $S_{DS}^2$ .
4. Finally, because a study is based on only a finite sample of animals, we can only estimate the data-set specific potency, with an error variance of  $S^2$ .

Our final goal is to estimate the overall mean potency for the chemical  $\times$  sex  $\times$  compartment combination, along with a standard error that reflects the uncertainty in the estimate due both to errors in estimation and the variances among the MRID-specific and data-set-specific potencies. Furthermore, once each such overall mean potency is computed and an index chemical selected, we want to compute the potency for each chemical relative to the index chemical in each sex  $\times$  compartment combination.

To facilitate this latter computation, all operations were conducted with logarithms of the potencies. Thus, in the end, the relative potencies were computed as  $e^{\log m - \log ml}$ , where “logm” is the logarithm of the potency for the chemical for which the relative potency is being calculated, and “logml” is the logarithm of the potency of the index chemical. The uncertainty of the relative potency is then expressed as a confidence interval, obtained by exponentiating the endpoints of the confidence interval for the difference “logm - logml”, whose standard error is  $\sqrt{se_{\log m}^2 + se_{\log ml}^2}$ , where  $se_Q^2$  is the square of the standard error for parameter Q, here one of the log potencies. The estimates of logm were constructed in two stages: first, estimates of data-set-specific logms were combined, using their estimated standard errors, to yield MRID-specific estimates of logm and standard errors; then, via the same process, the MRID-specific estimates of logm were combined to yield overall mean estimates. The procedure used to combine the estimates is known as the “global two-stage method” (Davidian and Giltinan, 1995, pp 138ff). The logic of the global two-stage method is simple, and is illustrated here for estimating a MRID-specific estimate of logm. Since we have individual estimates of logm for each data set, it is natural to estimate a MRID-specific estimate as the mean of the individual data-set-specific estimates. However, if we estimate the standard error in the usual way, based on the standard deviation of the data-set-specific estimates, it turns out that the resulting estimates of the standard error are biased upwards, because the procedure described here ignores the uncertainty of the individual estimates. The global two-stage method corrects this bias by explicitly taking into account that uncertainty.

### Estimating BMDs and BMDLs

Benchmark doses were estimated for a 10% reduction in activity from background. That is, the benchmark dose is the value  $D$  such that:

$$0.1 = \frac{1}{A} \left\{ A - \left[ B + (A - B)e^{-mD} \right] \right\}$$

which is:

$$D = -\frac{1}{m} \ln \left( \frac{0.9A - B}{A - B} \right).$$



The BMDL was based on the lower 95% confidence limit on the estimate of the BMD. The confidence limit was computed from the estimated standard error of the reciprocal of the BMD estimate, because simulations (described in the next subsection), indicated that such confidence limits came closer in practice to the theoretical coverage than did limits based on standard errors computed on the original scale or on the logarithm scale. That is, the lower 95% confidence limit was computed as:

$\frac{1}{BMD_{inv} + 1.645 se_{BMD_{inv}}}$ , where  $BMD_{inv}$  is the reciprocal of the BMD estimate:

$$BMD_{inv} = \frac{m}{\ln\left(\frac{A - B}{0.9A - B}\right)} = g(\mathbf{b}),$$

where  $\hat{\mathbf{a}}$  is the vector of parameters,  $[A, B, m]'$ , and  $se_{BMD_{inv}}$  is its estimated standard error, computed as

$$se_{BMD_{inv}} = \left( \frac{\nabla g}{\nabla \hat{\mathbf{b}}} \right)' \hat{\Sigma} \left( \frac{\nabla g}{\nabla \hat{\mathbf{b}}} \right),$$

where  $\hat{\Sigma}$  is the estimated covariance matrix for the parameters, and  $\frac{\nabla g}{\nabla \hat{\mathbf{b}}}$  is the gradient of  $g()$  evaluated at the parameter estimates.

## Simulations to Support the BMDL Computations

### Simulation Methods

The performance of the statistical methods used in OPCumRisk, as well as the behavior of the estimation process, was checked by simulation. The simulation process was as follows:

1. Generate simulated data sets
  - a. For a set of values of  $A$ ,  $B$ , and  $m$ , different levels of maximum inhibition observed (that is, given  $A$ ,  $B$ , and  $m$ , pick the dose that yields the specified maximum level of inhibition), and particular dose regimen, compute mean cholinesterase activity levels ( $\mu$ ) using:

$$\mu(\text{dose}) = B + (A - B) e^{-m \times \text{dose}}$$

- b. Compute the standard deviation of data from each dose ( $\sigma$ ) by:

$$\sigma(\text{dose}) = \text{CV} * \mu(\text{dose})/100.$$

- c. For each of 500 replicate data sets, sample means from a normal random number generator with mean given as in (a) and standard deviation given as  $\sigma/\sqrt{\text{Sample Size}}$ . Sample standard deviations by generating Chi-squared random numbers with degrees of freedom equal to (Sample Size - 1) using a Chi-squared random number generator, then multiplying by the square of the required standard deviation, dividing by Sample Size, and taking the square root.
- d. The parameter values used for the simulation were based on values observed in a pilot sample of studies, and are given in the table below:

level of A:	2000
levels of B (B/A):	0, 500, 1000 (0, 0.25, 0.5)
levels of m:	0.03, 0.20, 1.0, 5.0
Highest dose selected to give activity at highest dose = A - (1 - ActMaxF)*(A - B)	ActMaxF=(0.5, 0.85, 0.95)
Dose Regimens (fractions of highest dose):	{0, 0.05, 0.20, 1.0}, and {0, 0.01, 0.067, 0.3, 1.0}
CV:	20%, 40%
Sample Sizes:	6, 10

A total of 216 unique combinations of parameter values were simulated, with each represented by 500 simulated data sets, for a total of 108,000 different simulated data sets available for parameter estimation.

2. Loop through the simulated data sets:
  - a. Attempt to fit a model, estimating  $B$ , to all data points.
  - b. If the P-value for the Pearson goodness-of-fit chi-square was less than 0.05, or the model did not converge, try to fit a model with  $B$  set to 0.
  - c. Repeat step (b), dropping the high dose group with each iteration, until the Pearson chi-square is greater than 0.05, or only three doses remain.
  - d. Record parameter estimates, the covariances of the estimates, the estimated benchmark dose for a 10% reduction of activity, and 95% lower confidence limits based on standard errors for (i) BMD computed on the original dose scale, (ii) the natural logarithm of BMD, and (iii)  $1/\text{BMD}$ .
3. Summarize the simulation results. For each combination of parameter values, collect:
  - a. The bias  $[\text{mean}(m) - \text{true}(m)]$  and relative bias  $[(\text{mean}(m) - \text{true}(m))/\text{true}(m)]$  of the estimate of  $m$ .
  - b. The proportion of each of the approaches for computing confidence limits in which the calculated limit is less than the true BMD. The true BMD is computed from the values of  $A$ ,  $B$  and  $m$  from which the data were generated.
  - c. Analyze the simulation results to examine relationships between parameter values and levels of bias or BMD coverage, and to determine which of the approaches to computing the BMDL comes closest to providing the nominal 95% coverage. The analysis is based largely on the use of regression trees (Breiman et al., 1984) as implemented in the Rpart package version 3.0.1 for R version 1.3.0.

The simulations showed that the bias on the absolute potency estimate in an individual data set depends upon the true value of  $B$ , and on the degree of cholinesterase inhibition at the high dose. Three different levels of  $B$  were considered in the simulations: 50% of background, 25% of background, and 0. Note that here,  $B$  refers to the value of the dose-response model used to generate the data, NOT the value estimated by fitting the data. Figure 2 shows that, if  $B$  was 50% of background, then  $m$  was underestimated by about 28%. None of the other factors in the simulation was systematically associated with variation around this level. If  $B$  was a smaller fraction of background, either 0 or 25% of background, then the bias depended upon

how much of the dose-response curve was captured in the individual dataset. If the cholinesterase activity at a high dose level was at least 85% of the estimated B, then on average absolute potency was over estimated by about 3%. If the cholinesterase activity at the high dose level was only 50% of the estimated B, then again, the degree of bias depended upon the true value of B. For example, if the true value of B was 25% of background, then the absolute potency was overestimated by around 11%. When B was 0, potency was overestimated by over 50%.

If the true value of B was known for a given dataset, the degree to which the estimate of absolute potency for that data set is likely to be biased could be evaluated, since both conditions that influence bias in this simulation depended upon the true value of B. However, B is unknown for real data sets. To try to evaluate the magnitude of bias in the estimates of absolute potency that might be expected, we have estimated B/A for each chemical  $\times$  sex  $\times$  compartment combination by combining estimates of B/A from data sets in those compartments which were adequately fit by the full model (that is, while estimating B). Also, the ratio of the model-predicted activity at the highest dose actually used in the fitting to the predicted control activity was computed for each data set (using predicted values smooths out some of the statistical fluctuations). That ratio was used to estimate ActMaxF for each data set. The degree to which potency estimates might be biased by cross-tabulating the estimate of B/A and the estimate of ActMaxF can be evaluated. Although the simulation was run at discrete values of B/A and ActMaxF, of course the real data are distributed continuously across the possible ranges of these two variables. To do the cross-tabulation, both variables were broken at the midpoints between discrete points used in the simulation: B/A at 0.125 and 0.375, and ActMaxF at 0.675. Altogether, B could be estimated for 1135 data sets. The breakdown by B/A and ActMaxF and the of the bias in the absolute potency estimate the simulations predict would be operative if every dataset in the corresponding group of the table had values of B/A and ActMaxF at the corresponding simulation value is:

B/A	ActMaxF	
	$\leq 0.675$	$> 0.675$
$\leq 0.125$	73 (6.4%)	117 (10.3%)
	bias: 50%	bias: 3%
$0.125 < B/A \leq 0.375$	199 (17.5%)	382 (33.7%)
	bias: 11%	bias: 3%
$0.375 < B/A \leq 1$	139 (12.2%)	225 (19.8%)
	bias: -30%	bias: -30%

About 61.5% of these data sets fall into the 3% or 11% bias groupings, while about 32% of the data sets fall into groupings where the simulations predict that absolute potency would be underestimated by about 30%.

Two complications in the calculation of the relative potency factors need to be considered for a total evaluation of bias. First, since the RPFs are based on average absolute potencies, data sets with both high and low values of ActMaxF would be combined to get an overall average. This should mean that bias in the overall average potency should fall somewhere between that of the individual data sets. Secondly, the denominator of the RPF, the absolute potency of methamidophos, itself must have some bias. This has the effect of reducing the overall bias of the RPF of chemicals whose absolute potencies were overestimated, and increasing the bias of the RPF of chemicals whose absolute potencies were underestimated. In fact, it is likely this is a small effect, since, in the RBC compartment in males, the ratio B/A is about 0.14, so the bias in its absolute potency would range between 3% and 11% based on the simulation.

It should be clear that the overall bias of the relative potency estimates that is due to the estimation procedure is likely to be relatively small. However, this whole analysis should be taken as suggestive, rather than determinative, of the levels of bias likely to exist in the estimates of absolute potency. The real data sets have a range of dose-placements and sample sizes, while the simulations, while based on the distribution seen in the data, used a much smaller range. We have estimated B/A, but many data sets do not allow B to be estimated. By dropping those data sets, we may have biased the estimate of B. Finally, this is not a very quantitative analysis of any bias that might result from combining data sets.

### ***Summary of Simulation Results***

1. Bias of Potency Estimates (Figure 2). The primary determinants of the relative bias in potency estimates is the value of B and the activity at the highest dose (ActMax). In models with  $B = 1000$  (that is, B is half of A), the potency is underestimated by about 28%. In models with smaller B, if doses are large enough that the activity at the highest dose is close to B, the bias is about 3% of the true value. The largest bias, around 54%, occurs in models with  $B=0$ , and where the activity at the highest dose is half the background.
2. Coverage of Nominal 95% Confidence Interval (Figure 3). In general, confidence limits for the BMD in individual data sets have lower than their nominal coverage; that is, the BMDL is too high. Only if  $B=0$ , where the average coverage is about 97% or, for larger B, the activity at the highest dose is only half way to the horizontal asymptote, where the average coverage is about 93%, are coverages close to nominal.

## Software

To facilitate modeling the large number of datasets evaluated in this study, special purpose software was written using version 1.2.1 of the open source statistical programming language R, (Ihaka and Gentleman, 1996; <http://cran.r-project.org>). A graphical user interface using the *tcltk* package for R was constructed to facilitate all phases of the analysis. Model parameters and their standard errors were estimated using the function *gnls* in the R package *nlme* (version 3.1-10; see Pinheiro and Bates, 2000).

### ***Consistency of results with dose-additivity assumptions: Observation of maximal response for cholinesterase inhibition***

In an ideal world, chemicals in a common mechanism group would demonstrate the same theoretical maximal response for the common mechanism endpoints. Some argue that the achievement of the theoretical maximal response may have important implications for dose-additivity, particularly at high dose levels (Putzrath, 1997). The theoretical maximal response for cholinesterase activity would be 100% inhibition of the enzyme.

Achieving the theoretical maximal response implies that the value of  $B$  is zero or practically zero. Because the actual value of  $B$  is affected by the background activity and the method of analysis, it is more appropriate to compare the ratios of  $B/A$  (i.e., the y-asymptote divided by the background) across studies than comparing the actual  $B$  values. This ratio of  $B/A$  represents the percent of background cholinesterase activity achieved at the estimated maximal response. For example if  $B/A$  is 0.25 then it is estimated that cholinesterase plateaus at 75% enzyme inhibition.

A discussion of  $B/A$  ratios is complicated by the fact that the model fitting procedure included setting  $B$  to zero and also dropping high doses. Therefore, *all* dose-response models in addition to dose-response models with *non-zero*  $B$  values are reported below. As shown below, 75% of dose-response models exhibited  $B/A$  ratios of 0-0.20 and 0-0.39 for all dose-response models and models with positive  $B$  values, respectively. Distributions of  $B/A$  for each chemical-sex-compartment combination are given in Appendix 3. Larger  $B/A$  ratios (> 30%) were observed for acephate, bensulide, malathion, methidathion, mevinphos, naled, phosmet, pirimiphos-methyl, and tetrachlorvinphos.

These results indicate that the value of  $B$  was not the same for all datasets but was rather heterogeneous. In addition, these results also indicate that for many datasets, the estimated  $B$  value was larger than the theoretical maximum of zero.



Distributional analyses of dose-response models for all single cholinesterase measurements.								
	Total Number of datasets	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	NA**s
Distribution of B/A <sup>a</sup> over all models	1312*	0	0	0	0.116	0.203	0.890	6*
Distribution of B/A over models with B > 0	535	0.009	0.144	0.249	0.283	0.399	0.890	--

\*Includes brain anatomical sections.

\*Datasets which did not converge to exponential model

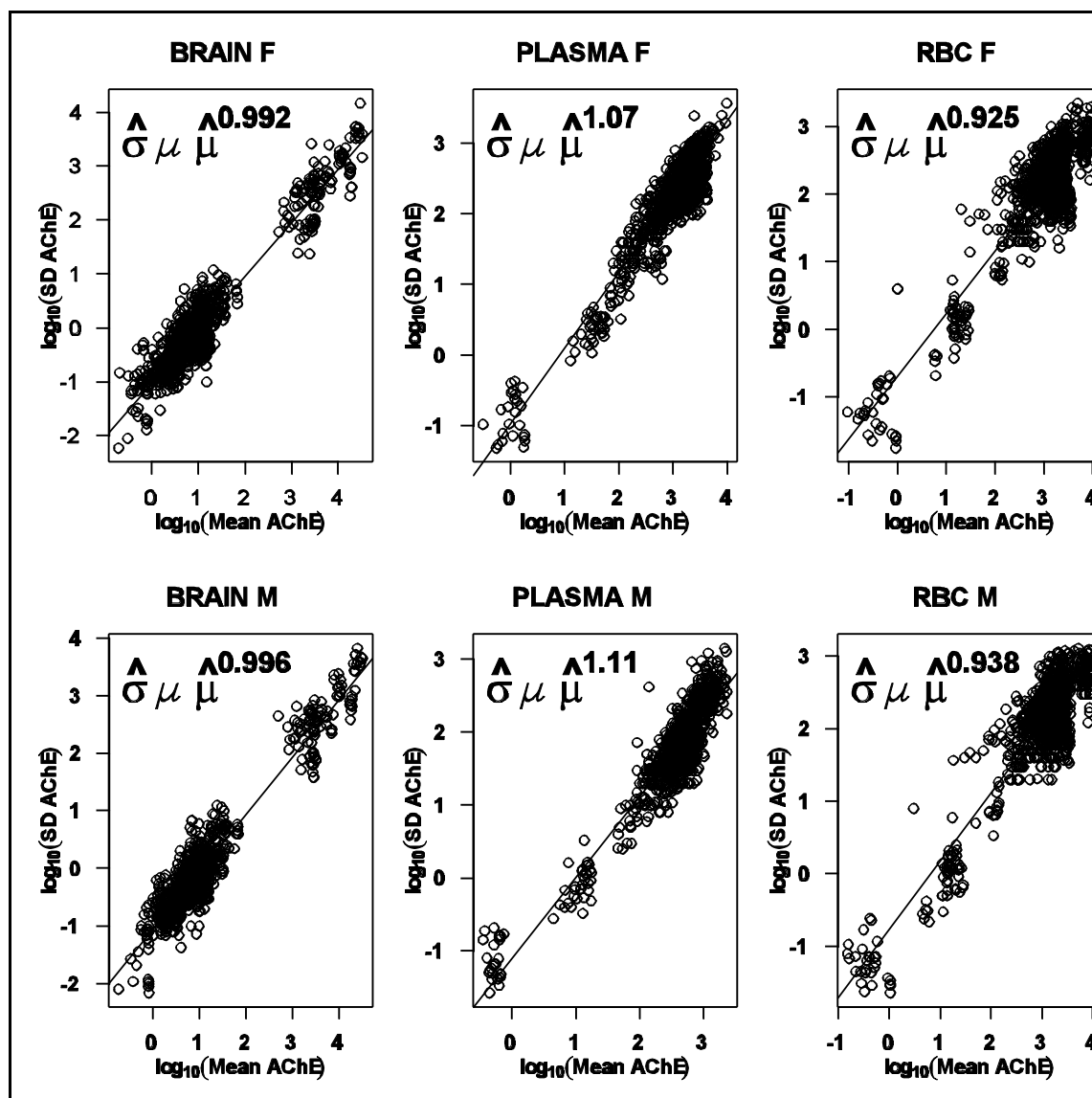
<sup>a</sup> B/A = Ratio of y-asymptote to background.

\*NA, not available

In general, although the value of estimate maximal response was not the same for all the datasets, the results of the analysis did not contradict the assumption of dose-additivity, especially at low exposure levels appropriate for extrapolation of the cumulative risk to humans. Thus, there was not a sufficient basis to depart from dose additivity.

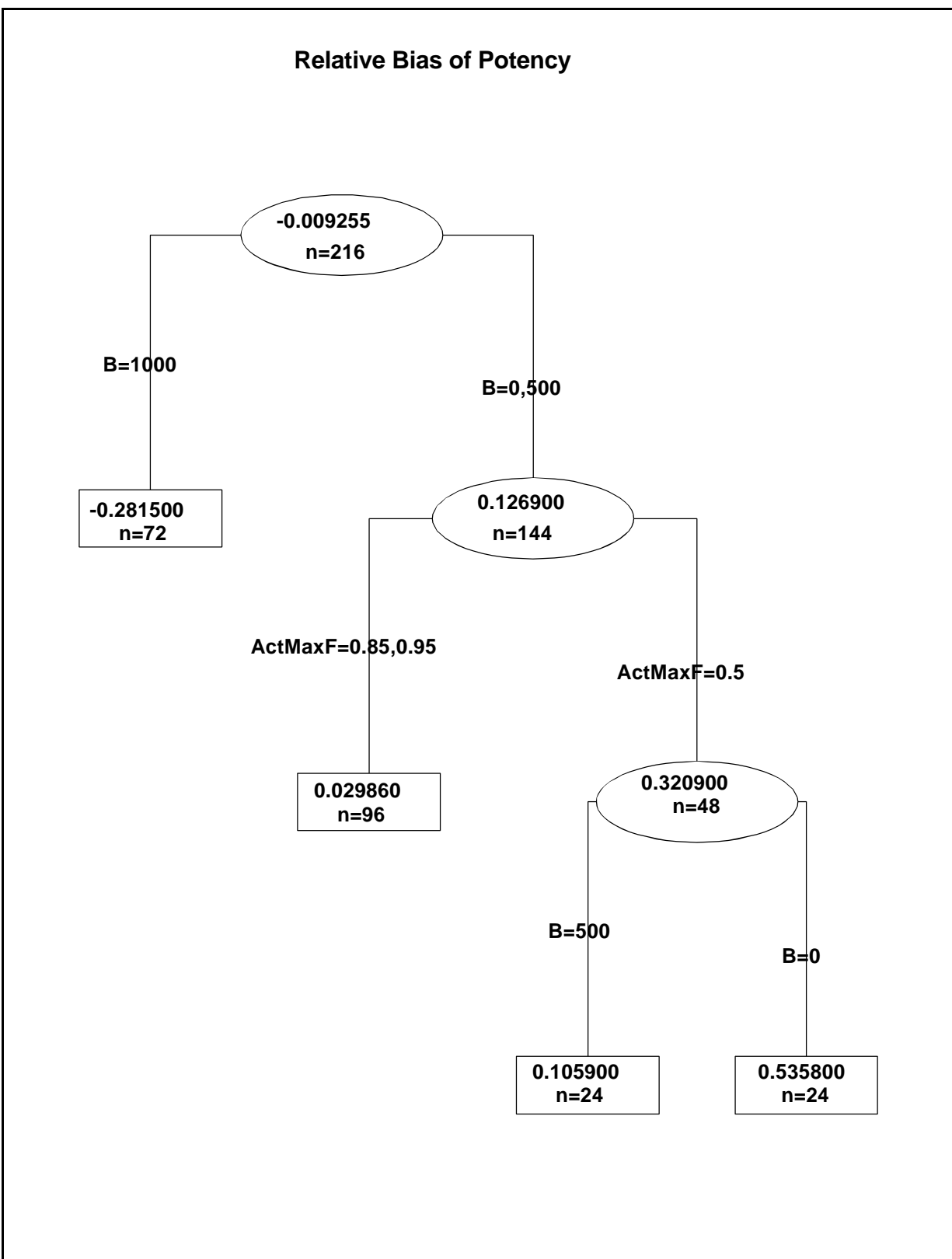
### Conclusion

The present approach to determining relative potency has several advantages. As opposed to another method, such as maximum-likelihood, the generalized least squares method used here for estimation of the parameters of the individual dose-response curves is generally more robust to misspecified data distributions which is important since actual data distributions were not directly available for checking. A novel aspect of this analysis was the use of a hierarchical statistical model to combine estimates of potency for the oral studies (average absolute potency values) and to combine estimates of benchmark dose for the oral, dermal, and inhalation routes (average BMD<sub>10</sub>s for the index chemical). Historically, OPP has selected single data sets or data points (such as reference doses [RfD] or NOAELs) for use in single chemical risk assessment. Aggregating over multiple data sets from studies with relatively well-defined study design has the advantage of being able 1) to increase the precision of the estimates when there is little additional variability among data sets and 2) to incorporate the variability among data sets into the overall estimate of uncertainty (standard errors or confidence limits). By combining potency estimates across data sets within studies and across studies, maximizes the use of the available information; almost all of the available dose-response data was used. Finally, this approach allows for a test of the dose-additivity assumption based on the similarly shaped dose-response curves because it generally forces an examination of each dose-response curve.



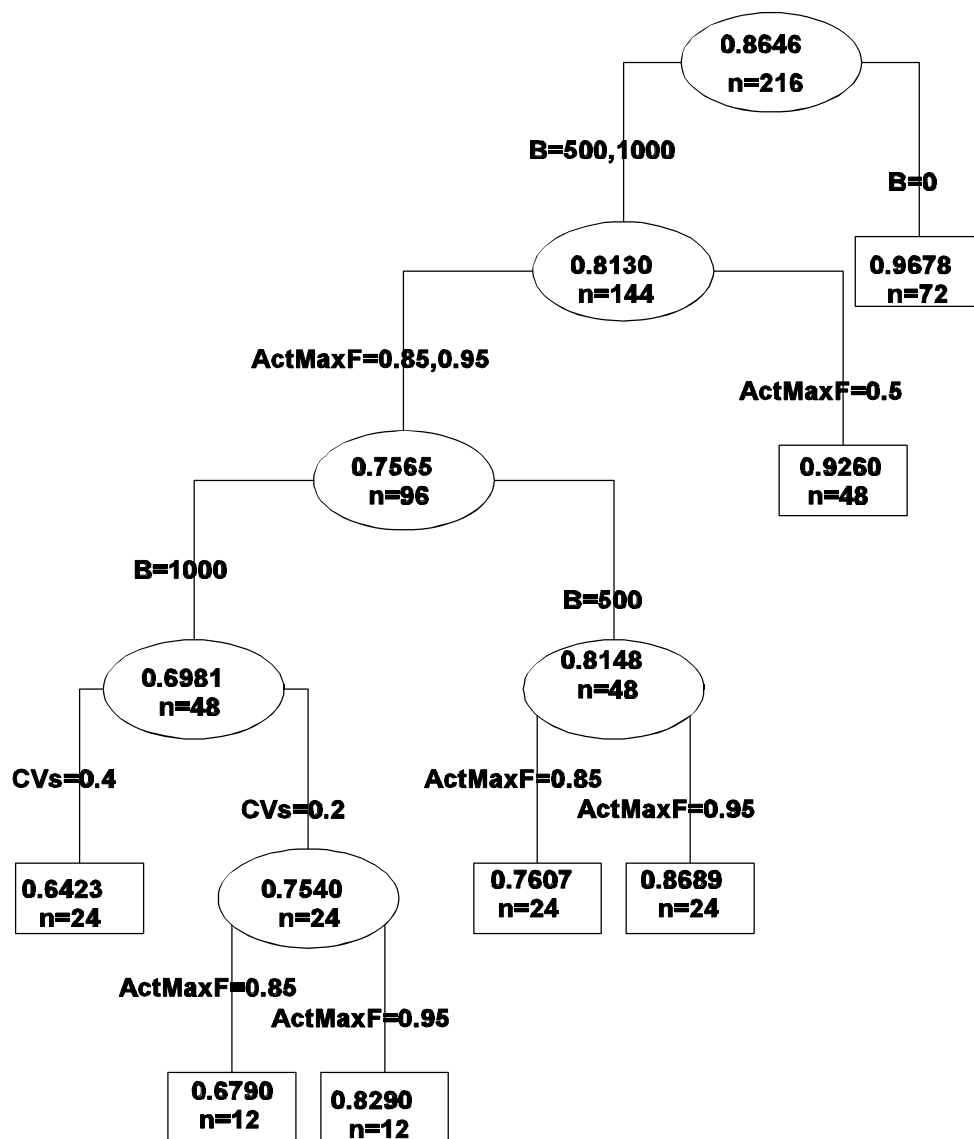
**Figure 1.** Log<sub>10</sub> of the standard deviation of cholinesterase activities plotted against the log<sub>10</sub> of the corresponding means, for each compartment and sex. Each point is a single dose group. The plotted line is a regression line fitted to the data in each panel. The regression suggests the indicated relationship between estimated standard deviation and mean.





**Figure 2.** Regression tree relating relative bias in potency estimates to model parameter values. The overall average relative bias is -0.009255 over all 216 combinations of conditions. If  $B < 1000$  (that is  $B/A < 0.5$ ), and the activity at the highest dose is nearly B, then the relative bias is only 0.0299.

### Coverage of BMDL based on inverse



**Figure 3.** Regression tree relating Coverage of the nominal 95% confidence limit for the BMD to the simulation conditions. The top number in each node is the average coverage at that node; n gives the number of observations that contribute to the average.

## References

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. A. 1984. *Classification and Regression Trees*. Wadsworth. New York.

Davidian, M. and Giltinan, D. M. 1995. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall. New York.

Ihaka, R. and Gentleman, R. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5: 299–314.

Pinheiro, J. and Bates, D. M. 2000. *Mixed Effects Models in S and S-Plus*. Springer. Berlin.

Putzrath RM. (1997) Estimating relative potency for receptor-mediate toxicity: reevaluating the toxic equivalence factor (TEF) method. *Regulatory Toxicology and Pharmacology*. 25: 68-78.