Environmental Health Perspectives
Vol. 79, pp. 229-241, 1989

Characterization of a Developmental Toxicity Dose-Response Model

by Elaine M. Faustman,* Dorothy G. Wellington,† William P. Smith,† and Carole A. Kimmel‡

The Rai and Van Ryzin dose-response model proposed for teratology experiments has been characterized for its appropriateness and applicability in modeling the dichotomous response data from developmental toxicity studies. Modifications were made in the initial probability statements to reflect more accurately biological events underlying developmental toxicity. Data sets used for the evaluation were obtained from the National Toxicology Program and U.S. EPA laboratories. The studies included developmental evaluations of ethylene glycol, diethylhexyl phthalate, di- and triethylene glycol dimethyl ethers, and nitrofen in rats, mice, or rabbits. Graphic examination and statistical evaluation demonstrate that this model is sensitive to the data when compared to directly measured experimental outcomes. The model was used to interpolate to low-risk dose levels, and comparisons were made between the values obtained and the no-observed-adverse-effect levels (NOAELs) divided by an uncertainty factor. Our investigation suggests that the Rai and Van Ryzin model is sensitive to the developmental toxicity end points, prenatal deaths, and malformations, and appears to model closely their relationship to dose.

Introduction

Current methodology for quantitative developmental toxicity risk assessment is based on the identification of no-observed-adverse-effect levels (NOAELs) and the application of uncertainty factors (UFs) to them. Other articles in this issue discuss the problems and limitations inherent in this approach (*1,2*). In particular, emphasizing a single dose level, the NOAEL, to set regulatory target doses ignores most of the information on the relationship between dose and response and does not take into account the variability inherent in the data. This article presents the authors' initial efforts to characterize and modify a dose-response model that has been proposed as an alternate approach to current risk assessment in developmental toxicity.

Since models of carcinogenic risk cannot be used for data from developmental toxicity studies because they do not allow for the relatedness or nonindependence of observations from the same litter, new modeling approaches are needed that take these points into consideration. In particular, theoretical mathematical models are needed to provide estimates of risk levels at specified low doses and estimates of doses at specified low levels of risk that are not strongly dependent upon the choice of the lowest dose administered in a study and to provide estimates for which confidence (uncertainty) statements can be made.

A model proposed in 1985 by Rai and Van Ryzin (*3*) for characterizing dose response for dichotomous data in teratology studies was chosen for investigation because of its potential for adaptation and extension to reflect simple mechanistic aspects of the underlying biologic processes. This model can be described as a two-stage probability model with flexible distributional assumptions. The model was designed to estimate the probability of teratological outcomes for individual fetuses. However, as originally proposed, the probability statements assumed teratogenicity contingent on maternal toxicity. Moreover, the data that were used to demonstrate the model application were from a dominant lethal study in which males were exposed to X-irradiation prior to mating. In the present paper, the model probability statements are reinterpreted to fit the typical developmental toxicity study with exposure to the dam during the period of major organogenesis and evaluation of the offspring at term. The model was characterized using data from developmental toxicity studies conducted by the National Toxicology

*Department of Environmental Health, SC-34, University of Washington, Seattle, WA 98195.
†Statistical Policy Branch, PM 223, Office of Policy Planning and Evaluation, U.S. Environmental Protection Agency, Washington, DC 20460.
‡Reproductive and Developmental Toxicology Branch, RD 689, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC 20460.

Program and the U.S. EPA Health Effects Research Laboratory.

Model Description

Basically, the Rai and Van Ryzin (RVR) model is the product of two component probabilities—the first unconditional and the second conditional upon the first—that represent two related stages in the development of an adverse developmental effect (response) under study. The model estimates the probability that an individual fetus from a litter of specified size s, whose dam was exposed during the appropriate period to dose d, will exhibit the response of interest. The relatedness within litters (intralitter correlation) that is due to having the same dam is addressed by an effort to account for the variability among dams (litters) in the same dose group. This interdam or interlitter variability is represented by a variable for litter size, which is included in the second probability component of the model, structured upon the assumption that larger litter size indicates a healthier dam and hence a lower risk to the individual fetus.

Expression of the RVR model has been extended to the following general form:

Probability of effect on fetus =

Probability of effect on litter environment X Probability of effect on fetus conditional on litter effect

or $$P(d,s) = \lambda(d) \times P[d,s;\lambda(d)] \quad [1]$$

where $P(d,s)$ is the probability of an adverse developmental effect on a fetus from a litter of size s, whose dam was exposed to dose d; $\lambda(d)$ is the probability of a predisposing change (disturbance) in the litter environment; and $P[d,s;\lambda(d)]$ is the conditional probability of an adverse developmental effect on a fetus from a litter of size s, whose dam was exposed to dose d, given that a predisposing change occurred in its litter environment.

As an initial approach, Rai and Van Ryzin suggested the one-hit model to represent $\lambda(d)$, the probability that the toxicant will produce a change in the litter environment that can lead to a specific adverse effect in the individual fetus. However, other models may be used to represent this probability of a litter effect. The one-hit model assumes that a response can be induced after a single susceptible target has been hit by a single biologically effective unit of dose (*4*). The probability of this hit is taken to be proportional to the dose level, i.e., Pr(hit) = βd, for some positive value of β. If the number of hits is assumed to follow the Poisson distribution (a distribution used for small numbers of events or counts) with the parameter of the distribution (the mean number of hits) equal to βd, then the probability of a response at dose d equals the probability of at least one hit, which is one minus the probability of no hits, or $1 - \exp(-\beta d)$. This version of the one-hit model is often referred to as the linear nonthreshold dose-response model since at low doses (small values of βd) the probability of a hit is approximately equal to βd.

Background risk is incorporated into this model by adding an intercept, α, to the linear expression in the exponent. The first component of the RVR model, the probability of a predisposing effect on the litter environment, is thus expressed as $1 - \exp\{-(\alpha + \beta d)\}$. The tendency of the one-hit model toward linearity at very low doses produces higher estimates of risk, and hence more conservative safe dose estimates than risk models that have convex (concave upwards) shapes, such as the probit model.

Rai and Van Ryzin expressed the second component of the model as an exponential function, with the exponent modified by (multiplied by) litter size, s. Although it could be any function of dose, they chose the exponent to be linear in dose, and again incorporated an intercept term, θ_1, to allow for a background risk, i.e., the conditional risk of a fetal response in the no-dose group. The exponent in the second component is thus expressed as $(-s(\theta_1 + \theta_2 d))$, with the coefficient θ_2 indicating the importance of dose level in the conditional risk to the fetus.

Since the second component of the model is also a probability and cannot be greater than 1, it reduces the primary probability, that of effect on litter environment, by modifying it through the influence of litter size and dosage to obtain the risk to the individual fetus. This formulation of the model, with the constraint that probability must be positive and no greater than 1, does not allow increased litter size to increase the risk to the fetus, but only to decrease the risk, or to have no effect. The complete model is given as

$$P(d,s) = [1-\exp\{-(\alpha+\beta d)\}] \exp\{-s(\theta_1+\theta_2 d)\} \quad [2]$$

This form of the model estimates the risk to the individual fetus when the litter size is known. In order to generalize the model across all observed litter sizes, while maintaining the adjustment for litter size, a theoretical distribution for litter size must be postulated. The overall risk to the fetus at a specific dose is then determined by integrating the risks for each litter size at that dose over the assumed distribution of litter sizes. This is similar to taking the weighted mean of the risks for each litter size, with weights equal to litter size frequencies based on the assumed distribution. The observed litter size frequencies for a specific dose group are used to estimate the distribution parameters at that dose.

The litter size distribution that is incorporated into this version of the model is the Poisson distribution for counts, and in this case the counts are the numbers of littermates. This model version assumes that the Poisson distributions parameter, $E(S)$, the expected

(mean) litter size, changes with dose, but not linearly. Instead, the expected litter size is assumed to have the following nonlinear form in dose d

$$E(S) = \phi_1 \exp(-\phi_2 d) \quad [3]$$

where the parameter ϕ_1 is the expected litter size in the controls, and the new parameter ϕ_2 represents the effect of dose on litter size. This formulation implies larger litter sizes at lower doses if ϕ_2 is estimated to be greater than zero, and smaller litters if its estimate is negative.

When the appropriate integration (summation here because litter size is a discrete variable) is carried out, the final model for estimating individual pup risk at dose d is

$$P(d) = [1-\exp\{-(\alpha+\beta d)\}] \exp[-\phi_1 \exp(-\phi_2 d)\{1-\exp(-(\theta_1+\theta_2 d))\}] \quad [4]$$

with the parameters, α, β, θ_1, θ_2, ϕ_1, and ϕ_2 appropriately constrained to ensure that neither of the two component probability estimates is greater than 1 nor negative. It can be seen that, while linear functions of dose are employed in the exponents, the risk model itself is nonlinear. If, instead of the one-stage model, a multistage model were used for the first component (the probability of a predisposing change in the litter environment), then the dose function in the exponent would itself be nonlinear. In the model as formulated above, the background (d = 0) probability of disturbance in the litter environment is equal to $1-\exp(-\alpha)$. The overall background risk is expressed as

$$P(0) = \{1-\exp(-\alpha)\}\exp[-\phi_1\{1-\exp(\theta_1)\}] \quad [5]$$

Thus, when exposure to the toxicant under study is zero, as in the controls, this model still allows for a nonzero probability of the predisposing litter disturbance and a nonzero conditional probability of an adverse effect on the fetus.

Parameter Estimation and Hypothesis Tests

Estimation of the model parameters was done using the maximum likelihood method that maximizes the fit of the model to the actual data set and assumes a binomial distribution for events within a litter, i.e., the number of affected fetuses out of a litter of size s. Thus, in this model, the variable for litter size functions in two capacities, one as a varying risk set (subsample) size, and another as a covariate to represent differences among litters.

Hypotheses on the parameters are tested by the conditional likelihood ratio test. In this version of the model, the hypotheses test the degree of exponential linearity of the dose effect in each of the two components. Testing the hypothesis that $\beta = 0$ indicates whether there is a relationship between dose and the probability of litter disturbance. Technically, testing $\theta_2 = 0$ indicates whether the relative advantage to the fetus of being from a litter one size larger varies with dose. Since the model is constructed on the assumption that larger litter size is associated with reduced risk, this test is an indirect way to investigate dose effect on individual fetus risk, conditional upon a predisposing change in litter environment. The parameters α and θ_1 could also be similarly tested to obtain information on the significance of background risk, as described in the previous section.

Estimation of Low-Risk Doses

Use of a mathematical model enables the estimation of risk at any specified dose and the estimation of dose for any specified risk, as well as the confidence limits on the risk or dose estimates. In particular, a model allows those low doses that are associated with very low specified risks, such as 1 in 10,000, to be estimated, as well as their lower 95% confidence limits. For their model, Rai and Van Ryzin provide two different methods for approximating the lower 95% confidence limit on these low-risk doses: the first is dependent on the distribution of litter sizes, and the second is not. Both methods estimate the lower limit in terms of the added risk over background, since this is the risk of interest at low doses.

In the first method, a low level of the added risk, π^*, is specified, say 10^{-4}, and the following equation is solved for d to obtain the estimated dose that will produce no greater added risk:

$$\pi^* = \pi(d) = P(d) - P(0) = 10^{-4} \quad [6]$$

with the maximum likelihood estimates of the parameters substituted into $P(d)$ and $P(0)$. The variance of the estimated dose is obtained from the Information matrix (the Information matrix is the negative inverse of the covariance matrix of the likelihood function). The approximate lower 95% confidence limit of the dose estimate is calculated by assuming that the dose estimate is normally distributed.

The second method uses the fact that the first probability component of the model is actually an upper bound on the model. $P(d,s)$, the risk estimate for a fetus from a litter of size s, decreases as litter size s increases, and thus has a maximum value when $s = 0$, i.e., where $P(d,s) = P(d,0) = \lambda(d)$, its upper bound. Thus, a conservative method for interpolating to a low-risk dose is to consider the added risk over background only for the first component probability. The formulation is then $\pi(d) = \lambda(d) - \lambda(0)$, which, for the one-hit model, is

approximately equal to $(\beta d)\exp(-\alpha)$. The equation to be solved for d, the low-risk dose in this case, is:

$$\pi^* = \pi(d) = (\beta d)\exp(-\alpha) = 10^{-4} \quad [7]$$

The approximate variance of the estimated dose is derived by taking its natural logarithm and applying the formula for the variance of a ratio. Again, the component variances are obtained from the Information matrix based on the likelihood function. The approximate lower 95% confidence limit is calculated for the estimated low-risk dose by assuming that its logarithmic transform is normally distributed.

Application of the Model to Data

The RVR model has the flexibility to allow the end point data to vary in accordance with the study design and the research goal. It can assess dose-response relationships when adverse effects are taken to be non-live implants, total or specific malformations, total adverse outcomes, etc. It requires only that the appropriate measure of litter size be used in each case, e.g., total implantations for non-live implants; number of live fetuses at term for malformations.

The RVR model in the version described above was applied to several data sets from standard Segment II developmental toxicity studies conducted by Research Triangle Institute under contract to the National Toxicology Program (NTP). Chemical data sets modeled were diethylhexyl phthalate (DEHP), mice (*5*) and rats (*6*); ethylene glycol (EG), mice (*7*) and rats (*8*); triethylene glycol dimethyl ether (TGDM), mice (*9*) and rabbits (*10*); diethylene glycol dimethyl ether (DYME), mice (*11*) and rabbits (*12*); and theophylline (THEO), mice (*13*) and rats (*14*). Data from a postnatal study on nitrofen exposure to rats during organogenesis were obtained from the Health Effects Laboratory, U.S. EPA, Research Triangle Park, NC (*15*).

In all the NTP data sets analyzed, the risk for two categories of end points were considered: (a) non-live, i.e., the total number of deaths and resorptions out of the total number of implantations; and (b) mal, i.e., the number of fetuses with at least one malformation out of the total number of live fetuses. For the nitrofen data (perinatal toxicity study), a single adverse outcome was analyzed, i.e., the sum of the pre- and perinatal deaths and the cyanotic pups at birth out of the total number of implantations.

Results of Model Application

The model-estimated risk probabilities and their two components are displayed in Table 1 for both nonlive and mal in 10 of the data sets. The second column gives the first component, i.e., the probability of change in the litter environment. The next column gives the second component, i.e., the probability that a fetus displays the effect, given that the precipitating change in the litter environment has occurred. The last column is the overall or final (unconditional) probability of a fetus response and is the product of the two component columns. It should be noted that the two component risk probabilities are not determined independently, since the four parameter estimates upon which they are based are jointly (simultaneously) estimated in the process of maximizing the likelihood function.

In Table 1 a pattern emerges among these studies in which the risk to the litter environment tended to increase in small increments as dose increased, but the conditional risk to an individual litter member increased in sizeable increments only at the top dose(s). While the overall risk to the fetus tended to combine these two probability patterns into one with more moderate, but still notable, increases at the highest dose, in individual studies the final fetus risk pattern was closer either to the litter risk (e.g., EG, mice, mal) or to the conditional fetus risk (e.g., TGDM, mice, nonlive). The actual size of the final risk is a product of both components and will not necessarily be close to either one. However, one can think of the primary probability as being the risk to the litter environment, which is then modified (decreased) by the conditional probability of risk to an individual fetus in that litter.

The results of fitting the Rai and Van Ryzin model to these data sets can be observed more easily when graphed. Figures 1*a–k* show three point estimates at each dose in each study for the two end points modeled: one is the RVR model estimate of fetus risk, the second is the average proportion per litter of affected littermates, and the third is the average percentage of litters with at least one member affected. The second and third estimates are from calculations presented in the original reports on these studies (*3–13*). All three are estimates of risk probabilities. The first two estimate in different ways the probability at a specific dose that an individual fetus is affected. The third estimates the probability at a specific dose that a litter is affected with the criterion being that at least one litter member is affected.

The point estimates of risk from the model are connected in the graphs in order to represent the continuous property of the RVR model. However, the actual graph of the model would be represented more accurately by a curved line of the continuous function expressed by $P(d)$. Since the first two point estimates named previously (RVR model estimate and average proportion per litter of affected littermates) express the risk to an individual fetus, and the third (average percentage of litters with at least one member affected) reflects the risk to a litter, the third point is markedly higher than the other two because it represents a broad range of risk scenarios for any one affected litter, from only one affected fetus to the entire litter affected. Only when every affected litter contains no unaffected fetuses

Table 1. Model-estimated risk probabilities.

Data set[a]/ dose, mg/kg	Probability of predisposing change in litter environment[b]	Conditional probability of fetus effect[c]	Overall probability of fetus effect[d]
DEHP, mice			
Non-live			
0	0.7187	0.1879	0.1351
44	0.7396	0.2419	0.1789
91	0.7602	0.3168	0.2408
191	0.7988	0.5616	0.4486
292	0.8314	0.9999	0.8314
Mal			
0	0.2220	0.0944	0.0210
44	0.4725	0.1520	0.0718
91	0.6517	0.2379	0.1550
191	0.8560	0.5147	0.4406
292	0.9410	0.9139	0.8599
DEHP, rats			
Non-live			
0	0.4468	0.0262	0.0122
357	0.4704	0.0770	0.0362
666	0.4735	0.2210	0.1046
856	0.4753	0.4497	0.2137
1055	0.4773	1.0000	0.4773
Mal			
0	0.0373	0.1877	0.0070
357	0.0374	0.3361	0.0126
666	0.0374	0.5485	0.0205
856	0.0374	0.7371	0.0276
1055	0.0375	1.0000	0.0375
EG, mice			
Non-live			
0	0.2860	0.3716	0.1063
750	0.3223	0.3988	0.1285
1500	0.3568	0.4269	0.1523
3000	0.4205	0.4858	0.2043
Mal			
0	0.0046	0.6523	0.0030
750	0.2090	0.7333	0.1533
1500	0.3714	0.8158	0.3030
3000	0.6031	0.9835	0.5931
EG, rats			
Non-live			
0	0.1105	0.3820	0.0422
1250	0.1319	0.4879	0.0644
2500	0.1528	0.6214	0.0949
5000	0.1930	1.0000	0.1930
Mal			
0	0.1547	0.0939	0.0145
1250	0.6248	0.1652	0.1032
2500	0.8335	0.2813	0.2345
5000	0.9672	0.7441	0.7197
DYME, mice			
Non-live			
0	0.2797	0.1659	0.0464
62.5	0.3103	0.2077	0.0644
125	0.3395	0.2600	0.0883
250	0.3943	0.4074	0.1606
500	0.4907	1.0000	0.4907
Mal			
0	0.0564	0.0355	0.0020
62.5	0.3981	0.0648	0.0258
125	0.6160	0.1110	0.0684
250	0.8438	0.2739	0.2311
500	0.9741	0.9632	0.9383

(*Continued on next page*)

Table 1. Model-estimated risk probabilities (continued).

Data set[a]/ dose, mg/kg	Probability of predisposing change in litter environment[b]	Conditional probability of fetus effect[c]	Overall probability of fetus effect[d]
DYME, rabbits			
Non-live			
0	0.1249	0.4137	0.0517
25	0.2182	0.4469	0.0975
50	0.3015	0.4838	0.1459
100	0.4424	0.5696	0.2520
175	0.6023	0.7378	0.4444
Mal			
0	0.1293	0.2574	0.0333
25	0.2882	0.3018	0.0870
50	0.4182	0.3517	0.1471
100	0.6112	0.4706	0.2876
175	0.7876	0.7018	0.5527
TGDM, mice			
Non-live			
0	0.6179	0.1350	0.0834
250	0.6822	0.1474	0.1005
500	0.7357	0.1606	0.1182
1000	0.8172	0.1898	0.1551
Mal			
0	0.0538	0.0271	0.0015
250	0.0689	0.0627	0.0043
500	0.0838	0.1505	0.0126
1000	0.1128	1.0000	0.1128
TGDM, rabbits			
Non-live			
0	0.3856	0.2611	0.1007
75	0.4168	0.3968	0.1654
125	0.4367	0.5204	0.2273
175	0.4559	0.6787	0.3094
250	0.4835	1.0000	0.4835
Mal			
0	0.2317	0.3237	0.0750
75	0.3767	0.4938	0.1860
125	0.4578	0.6265	0.2868
175	0.5283	0.7708	0.4073
250	0.6173	1.0000	0.6173
THEO, mice			
Non-live			
0	1.0000	0.0975	0.0975
282	1.0000	0.2050	0.2050
372	1.0000	0.2586	0.2586
396	1.0000	0.2752	0.2752
Mal			
0	0.0122	0.2379	0.0029
282	0.0199	0.6652	0.0133
372	0.0224	0.9177	0.0205
396	0.0230	1.0000	0.0230
Nitrofen, rats			
Deaths and cyanotic			
0.000	0.1719	0.4223	0.0726
0.150	0.1745	0.4245	0.0741
0.460	0.1799	0.4292	0.0772
1.390	0.1959	0.4435	0.0869
4.170	0.2419	0.4890	0.1183
12.500	0.3646	0.6532	0.2381
25.000	0.5124	1.0000	0.5124

[a]DEHP, diethylhexyl phthalate; EG, ethylene glycol; DYME, diethylene glycol dimethyl ether; TGDM, triethylene glycol dimethyl ether; THEO, theophylline.

[b]$\lambda(d) = 1 - \exp[-(\alpha + \beta d)]$.

[c]$P[d,s;\ \lambda(d)] = \exp[-\phi_1 \exp(-\phi_2 d)\ \{1 - \exp(-(\theta_1 + \theta_2 d))\}]$.

[d]$P(d) = \lambda(d) \times P[d,s;\ \lambda(d)]$.

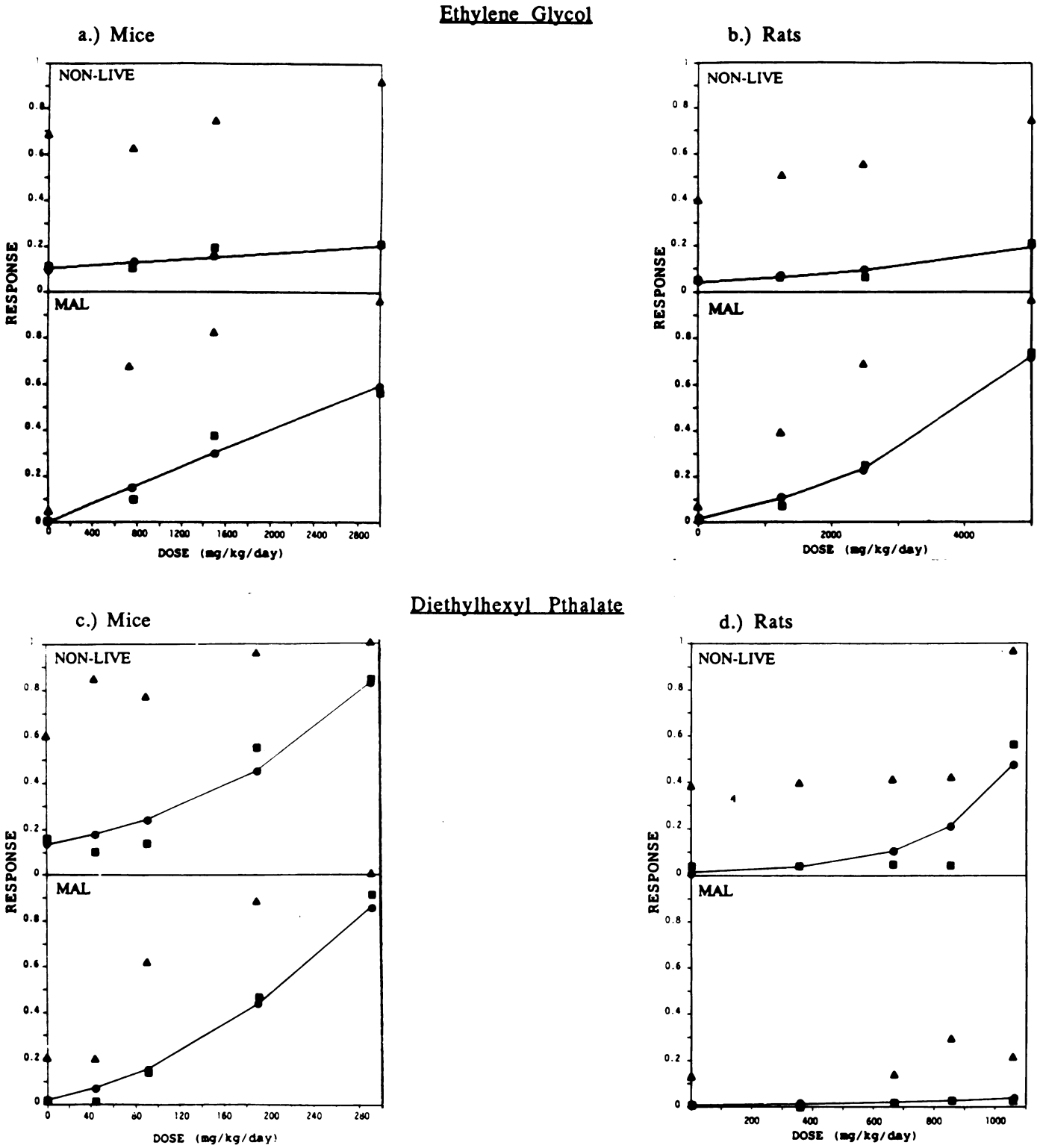

FIGURE 1. (*a-j*) Show, for the end points nonlive (total prenatal deaths) and mal (malformations), a graphic comparison of the NTP study risk estimates versus the Rai and Van Ryzin model risk estimates. (*k*) Shows the same comparison for the perinatal toxicity end point in the nitrofen study. (▲) Percentage of litters with at least one affected fetus; (■) average percentage of affected fetuses in a litter, as calculated in the NTP studies. (●) Rai and Van Ryzin model risk predictions, connecting lines indicate the continuous nature of the modeled dose-response relationship.

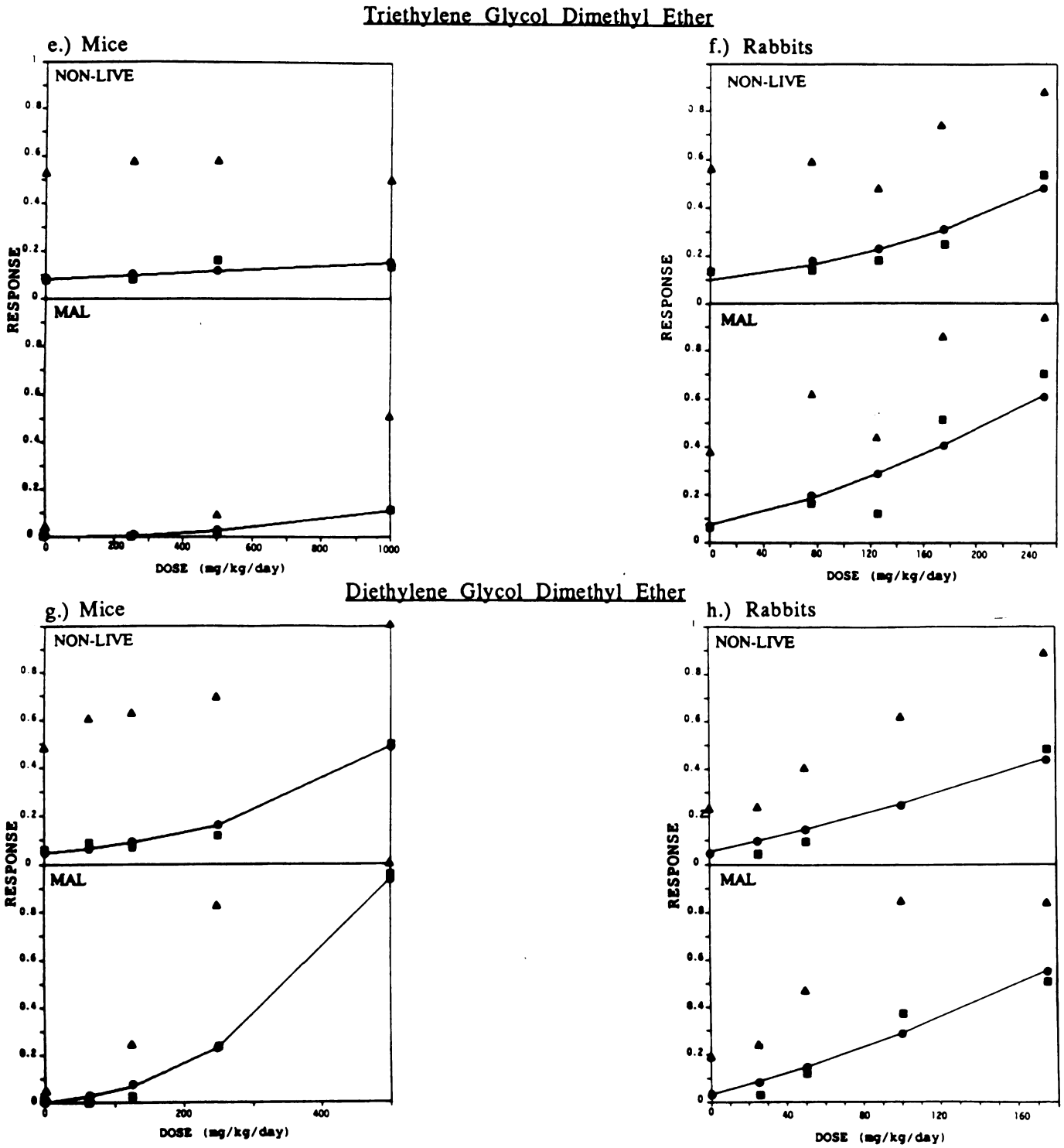

Figure 1. *Continued*

(*Continued on next page*)

two components of the model in the shape of the overall model. For example, the study of EG in mice (Table 2) shows no dose relationship in either litter effect or fetus effect for the non-live end point, but significant dose relationships in both components for the mal end point. Similarly, Figure 1*b* shows a strong dose-response relationship in the overall model for mal, but a weaker one for non-live. In the DEHP studies, Figure 1*d* (rats) shows a much weaker trend with dose in point. Similarly, Figure 1*a* shows a weak dose-response relationship for non-live and a strong one for mal. The study of EG in rats (Table 2) shows a significant dose relationship in litter and fetus risk for the mal end point, but only in the fetus risk for the non-live end general than does Figure 1*c* (mice). This is also reflected in Table 2, where the modeled dose relation-

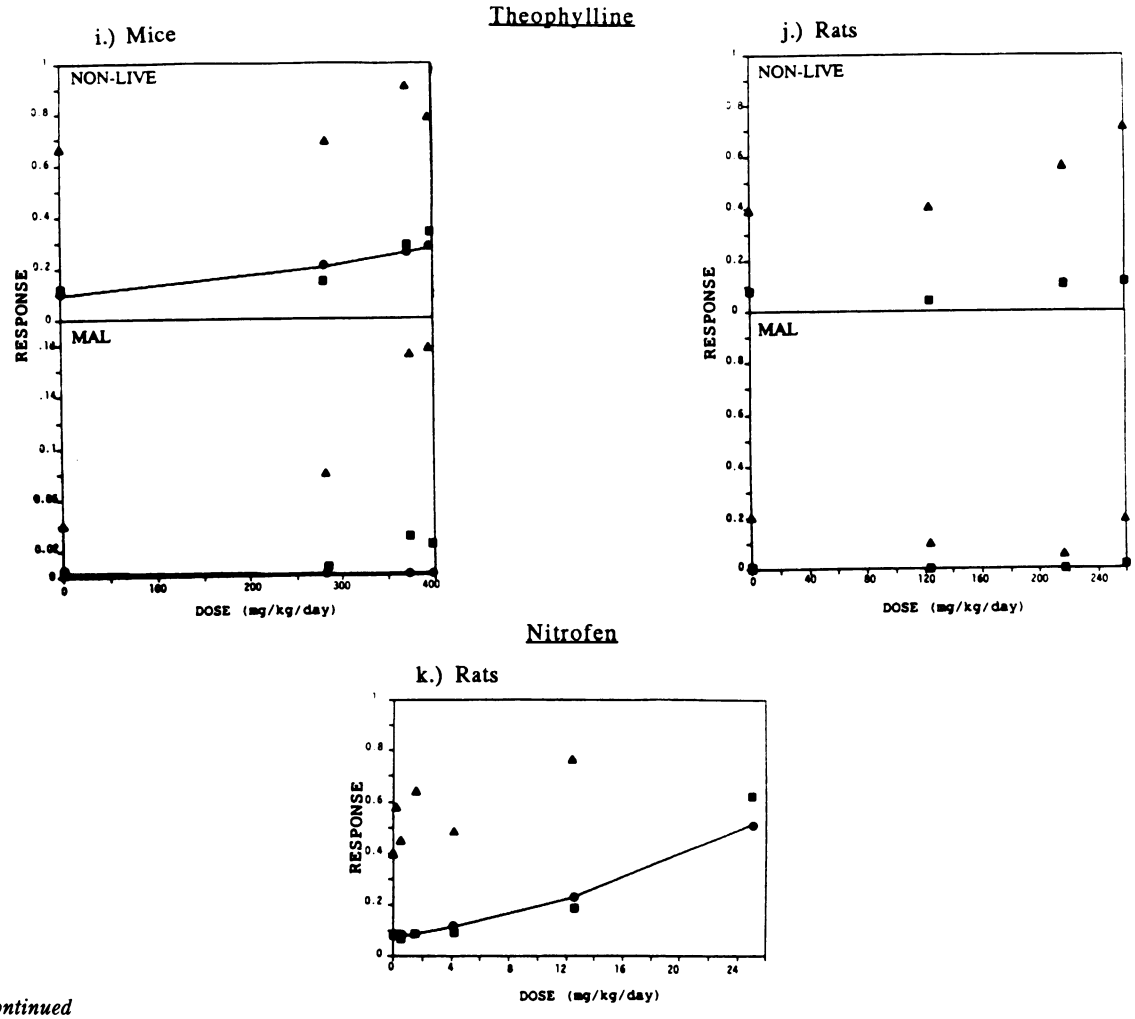

Figure 1. *Continued*

ship in litter and fetus risk for both end points in the DEHP studies was significant in mice, but only in the fetus risk for the non-live end point in rats.

Comparison of Low-Risk Dose Estimates with the NOAEL/UF Approach

Interpolations were done as part of this analysis in order to compare the model dose estimates at very low risks with the results of the method currently in use, i.e., the NOAEL/UF approach. The model-estimated low-risk doses were determined for a risk of 1 in 10,000 by two methods described earlier. A comparison of doses that are estimated to be acceptable by specified risk standards is given in Table 3 for the non-live end point and in Table 4 for malformations. The NOAEL in each case was determined from the original study reports to be the highest experimental dose tested that did not produce a statistically significant (p = 0.05) adverse effect when compared to controls. The NOAELs were then divided by a UF of 100, i.e., a 10-fold factor for interspecies variability and a 10-fold factor for within-species variability (*16*). In those cases when all doses showed a statistically significant effect compared to controls, the lowest dose was considered the lowest-observed-adverse-effect level (LOAEL) and was divided by an additional factor of 10.

The model-estimated doses were within an order of magnitude or were no more than one order of magnitude smaller than the uncertainty factor-adjusted NOAELs in all but three cases: TGDM mice, non-live by method 2; DYME mice, mal by method 2; and TGDM rabbits, mal by both methods. In the case of TGDM mice, non-live, there was no significant increase by pairwise comparison at any dose tested, so that it was not possible to determine a NOAEL accurately. However, this was not the case for the DYME mice, mal end point. The greatest discrepancy was between the NOAEL/UF and the model-estimated low risk doses for TGDM rabbits, mal. In this case, it can be seen from the graph (Fig. 1*f*) that the dose chosen as the NOAEL (125

is the individual fetus risk equal to the litter risk. To the extent that only a portion of fetuses in a litter are affected, the risk probabilities between litter and fetus will diverge.

The graphs show that the model reflected strong dose-response relationships, such as those observed for both categories of end points, non-live and mal (Fig. 1*c*, *f*, *g*, and *h*), for mal only in ethylene glycol-treated mice and rats (Figs. 1*a* and *b*), and for perinatal toxicity in nitrofen-treated rats (Fig. 1*k*). The model also reflected the shallow dose-response relationships observed for nonlive in ethylene glycol-treated mice and rats (Figs. 1*a* and *b*), for mal in DEHP-treated rats and theophylline-treated mice, and for both responses in mice following exposure to TGDM (Fig. 1*e*). Figure 1*j* shows examples of instances where there was no dose-response relationship and the data could not be modeled.

Goodness-of-Fit Test

The test proposed by Rai and Van Ryzin for determining the fit of the model to any particular set of data is appropriate only for large sample sizes. Their suggested remedy, to group litters of different but similar sizes, was not considered to be satisfactory, and a different test that did not require large samples was used. This goodness-of-fit test is based on the pure error obtained from replications, using the litters of the same size from the same dose group as the replicates. The difference between the total sum of squares of residuals (the deviations of the data points from their model predictions) and the replicates' sum of squares provides the sum of squares due to lack of fit. The ratio of the mean square due to lack of fit to the mean square for pure error has an F distribution with the appropriate degrees of freedom.

Of the 19 data sets, this version of the RVR model was found to fit 15; the fit for each of the remaining 4 data sets was rejected at the $p = 0.01$ significance level. These data sets were the nonlive response for DEHP mice and EG rats, and the mal response for TGDM rabbits and THEO mice (a data set with minimal dose response). However, the results of the model evaluations for these four data sets are included for comparison.

Further study of these data sets with respect to the deviations between the observed data points and their model predictions could provide insight into possible alternate versions of the RVR model that would better represent these four dose-response relationships. Interactive investigation of a hypothesized model and its actual fit to a specific data set can work toward both model improvement and increased understanding of the information in the data.

Hypothesis Tests of Dose-Response Relationships

The results of testing the significance of the parameters that estimate the relationship, as modeled, between the risk probabilities and dose in these data sets are given in Table 2. For the RVR model, this involves testing whether the parameters β and θ_2 could be significantly different from zero, with the significance level set at $p = 0.05$. If a significant difference from zero is found, it means that a relationship with dose is likely to exist. But if a significant difference from zero is not found, it means only that the relationship between dose and risk that is postulated in the model could not be confirmed by the data. There still may exist a dose-response relationship in the data that was not adequately modeled.

In Table 2, the tests of dose-related effects are given for four risk estimates: (a) the component of the model that estimates risk of predisposing change in the litter environment; (b) the model component that estimates conditional risk to the individual fetus; (c) the risk to the litter as estimated by the proportion of litters with at least one affected fetus; and (d) the risk to the fetus as estimated by the average proportion affected per litter. The last two estimates are from the original study reports and are based on the results of linear trend tests.

The closest agreement in identification of a significant dose-related effect was between the two estimates of individual fetus risk, i.e., the model estimate and the mean proportion affected per litter. In only three cases was there a disagreement between these estimates, i.e., non-live for EG mice, mal for DEHP rats, and mal for THEO mice. In all three of these cases the model-defined dose-response relationship was not found to be significant, although there was a slight though sufficient dose-related increase that could be identified by the linear trend test in the original study report. However, in none of these cases did the pairwise comparisons show a statistically significant increase in response above control values for any of the doses tested.

When β is not significantly different from zero, but θ_2 is, the dose relationship is indicated to be stronger for the individual fetus than it is for the litter environment. This is seen clearly for the non-live end point in the modeled studies. A significant dose relationship was not often found in the modeled risk to litter environment (β) for the non-live end point. In five cases where the dose relationship was significant in the conditional risk to the fetus (θ_2), it was nonsignificant in the risk to the litter environment. This may be related to the fact that the background incidence of prenatal deaths is relatively high, and the dose-related effect must be relatively strong to represent a significant increase. For the mal end point, in every case except one (TGDM mice), when the dose relationship was significant in the conditional risk to the fetus, it was also significant in the risk to the litter environment. Conversely, this may reflect the fact that the background incidence of malformations is relatively low so that a relatively moderate increase with dose may be found to be significant.

Comparisons of the results in Table 2 with Figures 1*a*–*k* illustrate the roles of the dose relationships in the

Table 2. Tests of dose-response relationships.[a]

Compound[b]	Species	End point	Dose effect on litter environment[c]	Dose effect on fetus	% Litters affected[d]	% Affected per litter[d]
EG	Mice	Non-live	NS	NS	S	S
		Mal	S	S	S	S
	Rats	Non-live	NS	S	S	S
		Mal	S	S	S	S
DEHP	Mice	Non-live	S	S	ND	S
		Mal	S	S	ND	S
	Rats	Non-live	NS	S	ND	S
		Mal	NS	NS	ND	S
TGDM	Mice	Non-live	NS	NS	NS	NS
		Mal	NS	S	NS	S
	Rabbits	Non-live	NS	S	NS	S
		Mal	S	S	NS	S
DYME	Mice	Non-live	NS	S	S	S
		Mal	S	S	S	S
	Rabbits	Non-live	S	S	S	S
		Mal	S	S	S	S
THEO	Mice	Non-live	NS	S	NS	S
		Mal	NS	NS	S	S
	Rats	Non-live	—[e]	—[e]	NS	NS
		Mal	—[e]	—[e]	NS	NS
Nitrofen	Rats	Death, cyanotic	S	S	ND	ND

[a]$p \leq 0.05$.
[b]DEHP, diethylhexyl phthalate; EG, ethylene glycol; DYME, diethylene glycol dimethyl ether; TGDM, triethylene glycol dimethyl ether; THEO, theophylline.
[c]NS, nonsignificant; S, significant.
[d]Taken from results of linear trend tests in original study reports; ND = not done.
[e]Model parameters were not estimable due to absence of consistent relationship between dose and response.

Table 3. RVR model-estimated doses for excess risk of $1/10^4$ compared with NOAELs adjusted by uncertainty factor of 100.

Compound tested[a]	Species tested	Non-live, NOAEL/100: % Response per litter	Non-live, NOAEL/100: % Litters with response	Model-estimated dose for excess risk of $1/10^4$, mg/kg/day (lower 95% CI)[b]: Method 1	Method 2
EG	Mice	> 30[c]	> 30[c]	3.5 (0)	2.0 (3.4×10^{-2})
	Rats	25	> 50[c]	6.7 (0)	5.8 (2.9×10^{-1})
DEHP	Mice	9.1×10^{-1}	9.1×10^{-1}[d]	1.2×10^{-1} (8.4×10^{-2})	2.0×10^{-1} (9.7×10^{-4})
	Rats	8.56	8.56[d]	2.9 (1.5)	10.0 (0)
TGDM	Mice	> 10[c]	> 10[c]	1.5 (0)	3.6×10^{-1} (1.8×10^{-3})
	Rabbits	1.75	1.75	1.5×10^{-1} (1.1×10^{-1})	2.3×10^{-1} (2.7×10^{-3})
DYME	Mice	1.25	2.5	4.0×10^{-1} (1.4×10^{-1})	2.0×10^{-1} (1.9×10^{-2})
	Rabbits	5×10^{-1}	5×10^{-1}	5.6×10^{-2} (0)	2.5×10^{-2} (1.3×10^{-2})

[a]DEHP, diethylhexyl phthalate; EG, ethylene glycol; DYME, diethylene glycol dimethyl ether; TGDM, triethylene glycol dimethyl ether.
[b]Dose level calculated by model at an excess risk of 10^{-4}. The values in parentheses are lower 95% confidence limits on the dose estimate. Methods 1 and 2 refer to methods used to do the low-dose interpolation. See text for specific information.
[c]Value here is highest dose tested; no statistically significant LOAEL was found. True NOAEL could be much higher.
[d]Tests of significance for percentage of litters with response were not conducted in the original study. Values here are taken from tests done on the number of litters with response.

Table 4. RVR model-estimated doses for excess risk of $1/10^4$ compared with NOAELs adjusted by uncertainty factor of 100.

Compound tested[a]	Species tested	Mal, NOAEL/100		Model-estimated dose for excess risk of $1/10^4$, mg/kg/day (lower 95% CI)[b]	
		% Response per litter	% Litters with response	Method 1	Method 2
EG	Mice	7.5×10^{-1} [c]	7.5×10^{-1} [c]	5.0×10^{-1} (0)	3.3×10^{-1} (1.9×10^{-1})
	Rats	12.5	1.25[c]	1.7 (0)	1.8×10^{-1} (6.5×10^{-2})
DEHP	Mice	4.4×10^{-1}	4.4×10^{-1} [d]	1.1×10^{-1} (0)	1.5×10^{-2} (6.3×10^{-3})
	Rats	> 10.55[e]	> 10.55[d,e]	8.6 (4.6)	—[f] —[f]
TGDM	Mice	5.0	5.0	15.1 (0)	1.6 (1.6×10^{-3})
	Rabbits	1.25	1.25	8.6×10^{-2} (2.1×10^{-3})	4.7×10^{-2} (2.1×10^{-2})
DYME	Mice	1.25	1.25	3.8×10^{-1} (0)	1.5×10^{-2} (6.5×10^{-3})
	Rabbits	5.0×10^{-1}	5.0×10^{-1}	4.9×10^{-2} (0)	1.4×10^{-2} (7.1×10^{-3})

[a]DEHP, diethylhexyl phthalate; EG, ethylene glycol; DYME, diethylene glycol dimethyl ether; TGDM, triethylene glycol dimethyl ether.
[b]Dose level calculated by model at an excess risk of 10^{-4}. The values in parentheses are lower 95% confidence limits on the dose estimate. Methods 1 and 2 refer to methods used to do the low-dose interpolation. See text for specific information.
[c]Value here is lowest dose tested; and all doses were statistically significant when compared to control values. This value is assumed to be the LOAEL and an additional 10-fold UF (total of 1,000) was used.
[d]Tests of significance for percentage of litters with response were not conducted in the original study. Values here are taken from tests done on the number of litters with response.
[e]Value here is highest dose tested; no statistically significant LOAEL was found. True NOAEL could be much higher.
[f]Insufficient dose-response relationship.

mg/kg/day) was associated with a response level (percentage affected per litter) that was low compared to the next lower dose and was also much lower than the model estimate. If the response level had been as high as expected, based on the incidence at adjacent doses, then a lower NOAEL would have been chosen that, when divided by a UF of 100, would be closer to the model-estimated low risk dose.

Thus, based on comparisons with the RVR model, the current approach for developmental toxicity risk assessment (NOAEL/UF) in most cases results in the identification of acceptable exposure levels that compare with model dose estimates for 10^{-3} to 10^{-4} excess risk. This is not surprising, as the NOAEL is likely to be in the response range of 1 to 10% and a 100-fold UF would reduce the risk to between 1 in 1,000 to 1 in 10,000 (*17*).

Summary

Preliminary studies have been conducted to evaluate the applicability of a dose-response model for dichotomous responses in developmental toxicity. The flexibility of the RVR model has been discussed, as well as the potential extension of this model to reflect mechanistic aspects of biological processes. We have shown that multiple hypothesis testing is possible with this model to determine the relationships of risk to dose. An important strength of this model is that low-dose interpolation is carried out using information from individual fetuses and litters. In addition, confidence limits can be calculated both for the estimated risk at any dose, and for the estimated dose at any specified risk level.

Our studies have shown that the RVR model as described here closely models the data from standard developmental toxicity studies for the proportion of non-live implants and the proportion of malformationed fetuses. Although the model may not account for all the variation due to litter differences (*18*), it uses litter size to account for a portion. Future direction for these evaluations include characterization of additional data sets, the addition of covariates other than litter size, the incorporation of intralitter correlation, and the application of other modeling functions, e.g., multihit, multistage, to express litter or individual fetus risk.

Portions of this work were presented at the Annual Meeting of the Teratology Society, Palm Springs, CA, June 1987 [Teratology 35:67A (1987)]. The views in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

REFERENCES

1. Gaylor, D. W. Quantitative risk analysis for quantal reproductive and developmental effects. Environ. Health Perspect. 79: 245–248 (1988).
2. Kimmel, C. A., Wellington, D. G., Farland, W., Ross, P., Manson,

J. M., Chernoff, N., Young, J. F., Selevan, S. G., Kaplan, N., Chen, C., Chitlik, L. D., Siegel-Scott, D. L., Valaoras, G., and Wells, S. Overview of a workshop on quantitative models for developmental toxicity risk assessment. Environ. Health Perspect. 79: 209–215 (1988).

3. Rai, K., and Van Ryzin, J. A dose response model for teratological experiments involving quantal responses. Biometrics 41: 1–10 (1985).
4. Hoel, D. G., Gaylor, D. W., Kirschenstein, R. L., Saffiotti, U., and Schneiderman, M. A. Estimation of risks of irreversible, delayed toxicity. J. Toxicol. Environ. Health 1: 133–151 (1975).
5. Tyl, R. W., Jones-Price, C., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of diethylhexyl phthalate (CAS No. 117-81-7) in CD-1 mice. NTIS No. PB85105674, National Technical Information Service, Springfield, VA, 1983.
6. Tyl, R. W., Jones-Price, C., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of diethylhexyl phthalate (CAS No. 117-81-7) in Fischer 344 rats. NTIS No. PB85105658, National Technical Information Service, Springfield, VA, 1983.
7. Price, C. J., Tyl, R. W., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of ethylene glycol (CAS No. 107-21-1) administered to CD-1 mice on gestational days 6 through 15. NTIS No. PB85105385/GAR, National Technical Information Service, Springfield, VA, 1984.
8. Price, C. J., Tyl, R. W., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of ethylene glycol (CAS No. 107-21-1) administered to CD rats on gestational days 6 through 15. NTIS No. PB85104594/GAR, National Technical Information Service, Springfield, VA, 1984.
9. George, J. D., Price, C. J., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of triethylene glycol dimethyl ether (CAS No. 112-49-2) administered to CD-1 mice on gestational days 6 through 15. NTIS No. PB86103215, National Technical Information Service, Springfield, VA, 1985.
10. George, J. D., Price, C. J., Marr, M. C., and Morrissey, R. E. Teratologic evaluation of triethylene glycol dimethyl ether (CAS No. 112-49-2) administered to New Zealand White rabbits on gestational days 6 through 19. NTIS No. PB87181657/AS, National Technical Information Service, Springfield, VA, 1987.
11. Price, C. J., George, J. D., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of diethylene glycol dimethyl ether (CAS No. 111-96-6) administered to CD-1 mice on gestational days 6 through 15. NTIS No. PB86135233/AS, National Technical Information Service, Springfield, VA, 1985.
12. Price, C. J., George, J. D., Marr, M. C., Morrissey, R. E., Schwetz, B. A., and Kimmel, C. A. Teratologic evaluation of diethylene glycol dimethyl ether (CAS No. 111-96-6) administered to New Zealand White rabbits on gestational days 6 through 15. NTIS No. PB87209532/AS, National Technical Information Service, Springfield, VA, 1987.
15. Gray, L. E., Jr., Kavlock, R. J., Chernoff, N., Ostby, J., and Ferrell, J. Postnatal developmental alterations following prenatal exposure to the herbicide 2,4-dichlorophenyl-*p*-nitrophenyl ether: A dose response evaluation in the mouse. Toxicol. Appl. Pharmacol. 67: 1–14 (1983).
13. George, J. D., Price, C. J., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of theophylline (CAS No. 58-55-9) administered to CD-1 mice on gestational days 6 through 15. NTIS No. PB86103223, National Technical Information Service, Springfield, VA, 1985.
14. George, J. D., Price, C. J., Marr, M. C., and Kimmel, C. A. Teratologic evaluation of theophylline (CAS No. 58-55-9) administered to CD rats on gestational days 6 through 15. NTIS No. PB86108172, National Technical Information Service, Springfield, VA, 1985.
16. Barnes, D. Reference dose (RFD): Description and use in health risk assessments. In: Integrated Risk Information System, Appendix A, U.S. Environmental Protection Agency, Washington, DC, EPA/600/8-86/032a, 1987.
17. Kimmel, C. A., and Gaylor, D. W. Issues in qualitative and quantitative risk analysis for developmental toxicology. Risk Anal. 8: 15–20 (1988).
18. Williams, D. A. Dose-response models for teratological experiments. Biometrics 43: 1013–1016 (1987).