Categorical Regression Analysis of Toxicity Data

(Received November 10, 2000; Revised December 18, 2000; In final form January 15, 2001)

Categorical regression analysis is a method by which regression analysis is applied to response data, in the form of ordered categories, for the purpose of predicting the probability of achieving a particular category of response as a function of one or more independent variables. This approach has been applied extensively to the evaluation of acute inhalation toxicity data, based on its ability to (1) use multiple independent variables to explain the response, (2) predict exposures related to various levels of effect severity, and (3) combine multiple studies. Categorical regression can predict a concentration associated with a specific probability of a particular response severity for a specified exposure duration, even without data for that exposure duration, as long as the data base contains exposure duration and response information on either side of the duration of interest. These principles are illustrated using the database for fluorine, which includes human and animal data, primarily of a descriptive nature. There have been fewer applications of categorical regression to oral data, and these applications have primarily been used to inform the interpretation of the RfD, rather than to derive the RfD. A number of issues related to the use of categorical regression, particularly for oral data, are presented.

Keywords: *Categorical regression; fluorine; dose-response analysis; meta-analysis*

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. The U.S. Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this article.

Comments On Toxicology
2001, Vol. 7, No. 5-6, pp. 437-452
Photocopying permitted by license only

INTRODUCTION

Categorical regression analysis is a method by which regression analysis is applied to response data, in the form of ordered categories, for the purpose of predicting the probability of achieving a particular category of response as a function of one or more independent variables. The response categories are ordered by severity, or strength, of response, but the spacing among the ordinal scores has no direct interpretation. Thus, an ordinal response of 2 is higher than a response of 1, but the difference is not necessarily the same as the difference between 3 and 2. McCullagh[1] was influential in the development of regression methods for ordinal data. Application of these methods to health risk assessment was suggested by Hertzberg and Miller[2] and by Hertzberg[3] as a method to empirically determine differences in sensitivity among species for subchronic and chronic toxicity data with responses ranked. In the application of categorical regression to toxicity data, effect data from toxicity studies are assigned to ordinal severity categories such as "no effect", "adverse effect", and "severe effect". Dose (or exposure concentration) can be used as one independent variable to explain response, while exposure duration can be used as another possible independent variable. The categorical regression model then develops the relationship of the ordinal response scores to the exposure information (e.g., concentration, or dose, and duration).

The advantages of the categorical regression approach for toxicity assessment include the abilities to (1) use multiple independent variables to explain the response, (2) predict exposures related to various levels of effect severity, and (3) combine multiple studies. The first ability is of particular interest for acute inhalation exposure, for which a key issue is how the predicted effects change with changes in exposure duration (e.g., from 1 hour to 4 hours). Traditional approaches have used Haber's law or a modification of Haber's law[4] to predict how effects change with exposure duration. These approaches are based on the assumption that the effect is constant if the product of concentration and time is a constant (i.e., $c \times t = \text{constant}$), or if $c^n \times t = \text{constant}$, respectively. However, even with the latter approach it is difficult to directly incorporate exposure-duration-response data into the assessment. Such analyses are easily done with categorical regression, which can predict toxicity based on two independent variables. While it is no special feat for a dose-response model to predict the probability of achieving a

response for a specific concentration (or dose) for which there are no experimental data, a categorical regression model that uses two independent variables can predict response based on either or both variables. For example, imagine that one requires a toxicity benchmark (e.g., a concentration associated with a specific probability of a particular response severity) for a three hour exposure duration. Categorical regression can predict this benchmark, even without data for a three hour duration, as long as the data base contains exposure duration and response information on either side of three hours.

The second advantage of categorical regression, the potential for predicting exposures related to responses of various severities, is also of particular interest to acute inhalation toxicity, but is also useful for other applications. This ability is of particular use for the development of emergency planning guidelines, which are defined in terms of the exposure limits needed to protect people from increasingly severe effects (e.g., protection from mild effects, severe effects, and death).

Several other strengths of categorical regression apply to both its application for evaluation of acute inhalation toxicity, and its application for other types of data, such as oral data. The meta-analytical use of categorical regression (the third advantage listed above) can be invaluable when individual studies provide only minimal useful dose-response data. For example, several individual single-dose studies may (if they are of similar design) together define a useful dose-response curve. Classifying toxic effects into severity categories provides a way to put different endpoints on a common scale for analysis and also allows the use of all types of data in the analysis. Continuous, descriptive and categorical data can be used, in addition to incidence data, as long as these effects can be classified into severity categories. This means that, for example, studies evaluating liver histopathology and studies evaluating only serum biochemistry measures can be combined into an overall evaluation of the dose-response for liver toxicity. Such an approach has the advantage in that more of the database can be included, in comparison to a strictly quantitative approach, such as benchmark dose (BMD) modeling. A disadvantage of combining studies is that considerable toxicological judgment is needed regarding when it is and is not appropriate to model together different studies and endpoints. As described in greater detail in the section on oral exposure, categorical regression has also been used to evaluate the risk above the RfD[5–6]. Finally, categorical regression is the only quantitative modeling approach that allows

one to incorporate graded data (i.e., taking into account the severity of a response in the modeling, not just that there was a response).

APPLICATION TO ACUTE INHALATION TOXICITY

A meta-analytical approach is particularly useful for acute inhalation exposures due to the more variable nature of acute experimental designs. Acute inhalation studies often vary exposure concentrations or durations, but both components of exposure are rarely examined in the same study. By combining information from multiple studies, the contribution of both concentration and duration to toxicity can be considered. Moreover, the combined analysis allows the analyst to investigate the variation among experiments, species, sexes and other covariates.

Guth and colleagues[7–12] adapted the categorical regression approach for use in acute inhalation toxicity, and found it extremely useful because acute toxicity often depends on both exposure concentration and duration. Led by Guth and colleagues, the National Center for Environmental Assessment (NCEA), U.S. Environmental Protection Agency, began the development of categorical regression as one approach to acute inhalation toxicity assessment. NCEA developed software, CatReg, especially adapted for application to the wide variety of data available on inhalation acute toxicity[13–14].

The structure of the model used in CatReg is:

$$\Pr(Y \geq s|C,T) = H\{\alpha_s + \beta_1 C + \beta_2 T\} \qquad \text{Equation 1}$$

Where s is the response represented by severity category, C is exposure concentration, T is exposure duration and H is a link function that keeps Pr between 0 and 1. The probability statement reads: The probability of Y being greater than or equal to a certain severity, at a particular concentration and duration, is a function of the concentration and duration of exposure. The model calculates the parameters α, β_1 and β_2 to best fit the data set to which the model is applied. If all of the input data are incidence data, this probability is equivalent to the risk – i.e., the estimated percent response at severity s, given exposure to concentration C for time T. However, if some of the input data are continuous or descriptive data, the probability reflects the probability that exposure to concentration C for time T will result in a response at severity s (e.g., the probability that the exposure is an adverse effect level).

CatReg includes a number of special features to facilitate application to the various types of data available for acute toxicity assessment. User-specified covariates (e.g., species, sex, target organ, etc.) can be incorporated into the model so that their impact on the exposure-response relationship can be assessed. For example, the data analyst may want to stratify some or all of the model parameters with respect to the species and target organ of the exposed individual with the expectation that doing so will reduce variability in the model estimates. (Stratification refers to the process of allowing specified model parameters to vary with specified covariates. Modeling may be done with only some parameters stratified, so that the data for one stratum still influence the fitted model for other strata.) CatReg also allows severity scores for a single response to be reported as a single severity category or as a range of severity categories. Reporting as a range of categories (or "censoring") allows for the use of responses for which the severity is uncertain. "Censoring" data means that the model fitting takes the censored data into account by making the likelihood of *any* of the severity categories included in the censoring range as great as possible (as opposed to making one specific severity category as likely as possible). Since categorical regression facilitates the combination of data from multiple studies, CatReg provides an option in which the model adjusts confidence limits in order to account for "clustering" among data where response rates within one laboratory are more similar to one another than to response rates from another laboratory. In addition to analyzing incidence data, CatReg can also analyze data reported as group means[13–14]. To aid in performing sensitivity analyses, CatReg allows the exclusion of user-specified data from the analysis without changing the data input file. CatReg also includes customized hypothesis tests, diagnostics, and graphical displays to assist in judging model fit.

EXAMPLE APPLICATION TO ACUTE INHALATION TOXICITY

Categorical regression analysis is especially useful for application to the acute inhalation toxicity data for fluorine. Most of the available effect data are descriptive data that are reported for an entire exposure group, which prevents analysis by more quantitative techniques. In addition, individual experimental protocols encompass a limited range of possible human exposure durations. Categorical regression allows studies to be

combined so that benchmark toxicity values can be developed for a larger range of exposure concentrations and durations than individual studies would allow. Also, combination of animal and human data allows the exposure-response curve to be extrapolated to humans in a way that uses the available human data to the fullest extent possible.

There is only one human study with adequate exposure and effect information. In this study volunteers reported no irritation to severe irritation during exposure to 10–15 ppm fluorine for 1–15 minutes.[15] The animal database consists of three published studies. Two studies report gross and histopathological results and signs of irritation in rats, mice, guinea pigs, and rabbits exposed to a range of concentrations for 5 minutes to 1 hour[15–16] and one study reports lethality in mice and rats exposed from 5 minutes to 7 hours[17]. Virtually all of the data on sublethal effects are reported descriptively and at the group level, and therefore are not amenable to quantitative methods such as the BMD approach,[18–19] which requires incidence or continuous data. Categorical regression provides an approach to combine the data from this limited number of studies and apply a single regression model to the combined information.

Response information was categorized according to an ordinal severity scale, with a severity category assigned to each experimental group, since the available studies presented response information only by group. No incidence data were reported. The severity categories included no adverse effect, mild adverse effect, severe adverse effect, and lethal, or likely to be lethal had the animals not been sacrificed (based on comparison with LC_{50} values from the same lab).

The data were analyzed with the CatReg software. In order to evaluate species differences, the model was applied with the intercept term stratified on species. This allows the intercept parameter (α in Equation 1) to be unique for each species, while the parameters that determine the slope (i.e., β_1 and β_2) are the same for all species. The results from this model, the *common-slope model*, are shown in Figure 1 (top) for mild adverse effects. Each line delineates species-specific concentration-duration coordinates predicted to result in a 10% chance (i.e., EC-T10) of observing an effect equal to or greater than mild adverse effects. CatReg also simultaneously calculates EC-T10 estimates for severe adverse and lethal effects (not shown). Figure 1 (top) shows separate EC-T10s for each species because the model includes stratum-specific values for α; the lines are parallel because the same values of β_1 and β_2 (the slope parameters) are

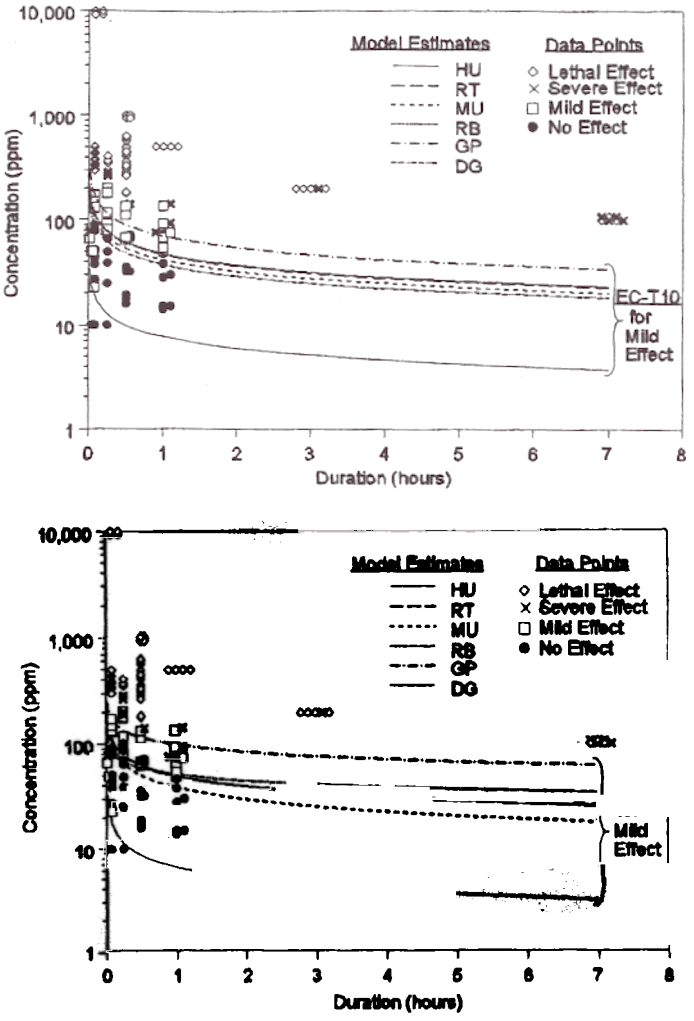

FIGURE 1 Stratified categorical regression model of acute fluorine toxicity in the respiratory system. The common-slope model is shown in the top half; the intercept parameter (α) is stratified on species. The variable-slope model is shown in the bottom half; the intercept parameter (α) and the concentration parameter (β_1) are stratified on species. Lines represent model predictions for EC-T10 values for mild adverse effects in each species. Model estimates for other severity categories are omitted for clarity

used for all species. The results show a clear difference between the model predictions for humans and experimental animals. Based on a likelihood ratio test, the model fit for the stratified analysis was significantly better than the unstratified analysis (not shown).

To further evaluate species differences, the model was fit with both the intercept (α) and the concentration parameters (β_1) stratified on species. This model estimates species-specific parameters for both intercept and slope (concentration parameter) and is referred to as the *variable-slope model*. The resulting model predictions for mild adverse effects are shown in Figure 1b. Although there is some difference in slope for the various species, the fit of the overall model is not improved significantly by this further degree of stratification and model complexity.

In evaluating a hierarchy of models applied to the same data set, it is reasonable to use the simplest form of the model that improves the fit as the most appropriate model to describe the data. The stratified common-slope model, therefore, is judged to be the best description of the exposure-response relationship and could be used in risk assessment. Figure 1a shows the EC-T10 results for mild adverse effects in the human stratum, which is the most relevant species for risk assessment.

In the case of fluorine toxicity, the human data alone are clearly inadequate to address the need for toxicity assessment over a range of durations, and, taken separately, the animal studies also are limited. The categorical regression model allows the human data to define the intercept, while using the animal data to characterize the slope, and combines limited data at durations longer than 1 hour with the data at 1 hour or less to develop health criteria estimates for 1 to 7 hours. Current methods for toxicity assessment would deal with these data gaps through the use of uncertainty factors or default duration adjustments. The categorical regression approach allows data gaps to be addressed by use of related information at different durations or in different species. The results of categorical regression can be adopted in risk assessment as a description of the exposure-response relationship in the range of the observed data, and as the basis for extrapolation to derive a health criterion value in risk assessment for acute exposures.

RECENT DEVELOPMENTS FOR CATREG

In recognition of the fact that constraining the probability functions for the various severity categories to parallelism (i.e., α varies with severity,

but β_1 and β_2 are the same for all severity categories) may not be biologically plausible for all toxins, CatReg now incorporates other options[13–14]. In addition to the original cumulative odds regression, CatReg now has three more options for the formulation of the probability calculation: unrestricted cumulative odds regression, conditional odds regression, and unconstrained conditional odds. The unrestricted version of the cumulative odds regression model allows the probability functions for the various severities to deviate from parallelism (i.e., each severity category can have a different β_1 and β_2). While this added flexibility can be useful in the modeling, it can lead to illogical results (e.g., the probability of a severity 2 level response being greater than the probability of a severity level 1 or greater response). Care needs to be taken to ensure that any such results are outside the concentration-duration range of interest. The conditional odds models (both the restricted and unrestricted versions) express the conditional probability that severity level s occurs, *given that* severity level s or *lower* occurs. This conditional construct allows the probability curves to deviate somewhat from parallelism even for the restricted version (where β_1 and β_2 are not allowed to vary with severity), but keeps them in the appropriate order so that the probability curves for the various severities do not cross. The unrestricted conditional odds model is similar to the conditional odds model, but also allows β_1 and β_2 to vary with severity. Along with allowing the severity curves to deviate from parallelism, CatReg also includes a diagnostic test to determine whether the data would best fit a simpler, parallel model.

Although CatReg allows the use of response data assigned to an exposure group rather than to individuals, incidence data are preferred because they best represent the number of animals in the exposure group. Continuous data often are reported as a mean value, with a measure of dispersion, such as the standard error or standard deviation, for each treatment group. To convert these data to severity levels for CatReg, each severity level needs to be equated to an interval of values on the continuous scale. The entire group is then assigned to the appropriate severity category based on the mean response. The *CatReg Software User Manual* now includes a method to estimate the individual responses for each severity level from group data reported as a mean and standard deviation (or standard error)[14]. The estimated incidence figures may not be whole numbers, but can still be treated as individual data for input to CatReg. Incidence estimation is not possible if the mean is reported without a measure of dispersion.

Other new additions to CatReg aid in determining model fit. CatReg now provides, upon user request, generalized analysis of variance and R^2 statistics that are derived from deviance statistics[13–14]. The R^2 statistic, which signifies the proportion of variation accounted for by the model, is a measure of the model's explanatory power. This statistic, which ranges between 0 and 1, is computed as the ratio of the model and total sums of squares. These sums of squares, along with degrees of freedom, and F-tests, are reported in the form of an analysis of variance table.

An additional plotting function was also added to help the analyst assess how well individual observations are explained by the exposure-response curve. Rather than making plots of the differences between observed and predicted responses, as done for linear regression analyses, CatReg plots the individual contributions to the deviance statistic to measure how well individual observations are explained by the exposure-response curve. Observations that contribute to any lack of fit of the exposure-response curve can be identified by their large generalized deviance residuals. This function provides a representation of the relative effectiveness of the model in fitting the different observations or strata. If one stratum has large deviances, the model may be inadequate for this stratum. Plots of deviance versus concentration or time can be performed to study the adequacy of the functional form of the regression relationship (e.g., log concentration versus raw concentration). Trends in the deviances, rather than a consistently random pattern, would suggest a problem with the functional form.

APPLICATION TO ORAL DATA

Only a limited number of complete evaluations of oral data have been done using categorical regression. Unlike the situation with inhalation data, where the primary application of categorical regression is for deriving risk values, categorical regression as applied to oral data has primarily been used to inform the interpretation of the RfD, rather than to derive the RfD. For example, Dourson et al.[5] applied categorical regression to human (acute exposure duration) clinical data for aldicarb ingestion exposures. Four categories of effect levels were used (no effect, nonadverse effect, adverse effect, and frank effect), based on clinical signs and cholinesterase inhibition. Two analyses were con-

ducted. In the first analysis, RBC and whole blood cholinesterase inhibition >20% were considered adverse, while these effects were considered nonadverse in the second analysis. Clinical signs were always considered adverse, and plasma cholinesterase inhibition was always considered nonadverse. Based on these categorizations, the authors calculated the cumulative probability of response for adverse and frank effects (Figure 2). The data also were compared with the existing RfD on U.S. EPA's IRIS, which is based on a human NOEL for cholinergic effects following an acute exposure. The authors used the modeling results to estimate doses corresponding to specific probabilities of observing effects of different severities.

Since the RfD may be considered the NOAEL for a sensitive population, there is often interest in determining the risk at doses somewhat above the RfD. The application of categorical regression to the aldicarb data set allowed such risks to be calculated. Confidence in this approach was enhanced by the close proximity of the data to the RfD. (A total uncertainty factor of 10 was used, so a dose 10 times the RfD was in the range of the data.) Greater caution would be needed in the estimation of risks further from doses at which data exist, as described in the next section. The approach used for this oral data set differs from the approach described above for inhalation data, in that this analysis did not consider bounds on those estimates. However, the difference is a function of the modeling programs used, rather than being related to differences between modeling of oral and inhalation data.

COMPARISON OF BMD AND CATEGORICAL REGRESSION

It is interesting that, although both categorical regression and BMD modeling are essentially curve-fitting dose-response modeling techniques, the published literature on these two techniques are very different with regard to the appropriateness of extrapolating to low doses. Categorical regression has been used for extrapolating below the data, such as particularly for estimating the risk above the RfD[5–6]. Teuschler et al.[6] used categorical regression to evaluate the risk above the RfD for five pesticides. A possible application of their work would be in the case of a regulator who has multiple pesticides where exposure exceeds the RfD, and the regulator wishes to set priorities for action based on the pesticide with the greatest likelihood of harm. In this case, the categori-

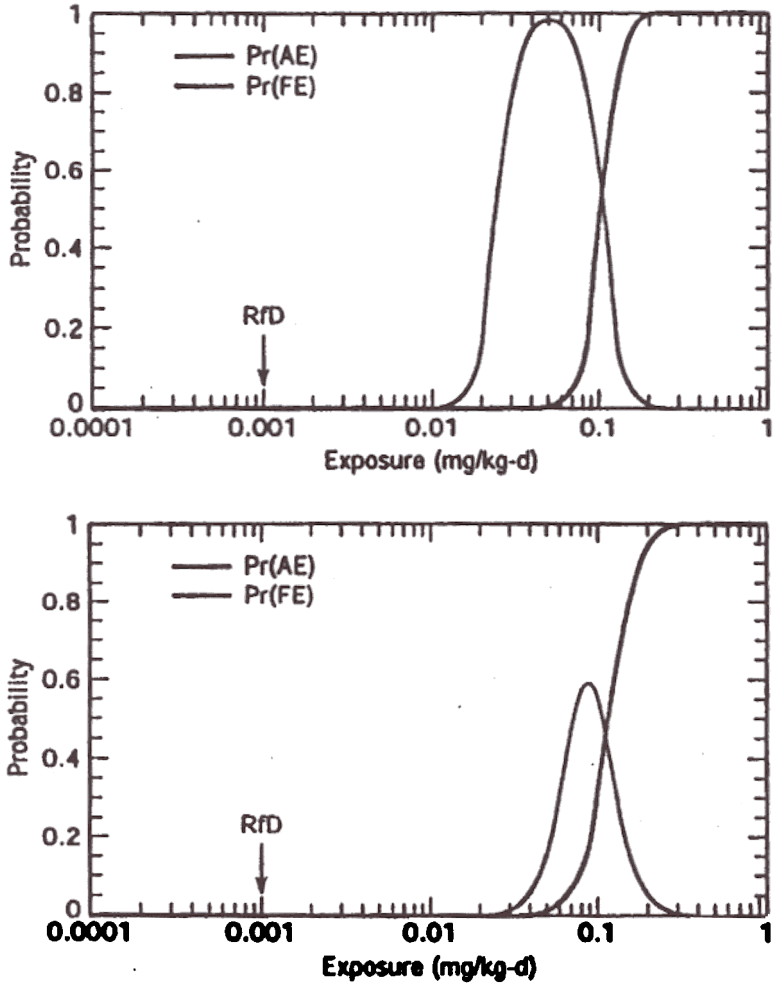

FIGURE 2 Probability of either an adverse effect (AEL) or frank effect (FEL) with aldicarb exposure, where whole blood or RBC cholinesterase inhibition of 20% or greater is considered to be an AEL (top) or a NOAEL (bottom). Reprinted, with permission, from Dourson et al.[2]

cal regression probability statement can be interpreted as asking "what is the probability of being wrong that x exposure is safe?"

The study authors noted several problems with their approach to using categorical regression to estimate the risk just above the RfD. The method used predicts a finite (but small) risk just above RfD. This is somewhat contradictory to the concept that the RfD is a subthreshold dose. A second issue is the choice of the appropriate response level to be used as a basis for an assessment. Current methods are to use a 10% response[14,11–12], analogous to the 10% response typically used for BMD modeling, but the implications of this choice have not been investigated in detail. The choice of response level is not unique to categorical regression, and is also encountered in BMD modeling.

By contrast, BMD modeling is discussed almost solely in terms of modeling in the range of the data, with the recommendation that extrapolation to low doses is not appropriate[8,20,19]. Two primary reasons have been advanced as to why BMD modeling should be done only in the range of the data. The first is that BMD modeling is only a curve-fitting exercise with no biological basis (lacking even the initial connection to biology that is in the multistage model). The second reason for limiting BMD modeling to the doses in the range of the data is that extrapolating beyond the data increases the model dependence; models that may fit equally well in the range of the data can have widely different results at doses several orders of magnitude lower.

Both of the reasons for limiting BMD modeling to the range of the data would also appear to apply to categorical regression. Like BMD modeling, categorical regression fits generic flexible mathematical models to the data, without a biological basis for the models. Similarly, the concern about model dependence increasing as one gets further away from the data would also apply to categorical regression. Indeed, Dourson et al.[5] noted the possibility for model dependence as a possible limitation to using categorical regression for estimating the risk above the RfD, but did not explore the issue in greater depth.

Consideration of these issues suggests that the dichotomy of approaches to categorical regression and BMD modeling may be due more to historical reasons, and related to different emphases of the researchers for each approach, rather than being an inherent difference between the two approaches. Nevertheless, there may be valid reasons why low-dose extrapolation using categorical regression may be more appropriate than similar extrapolation using BMD modeling. As noted

above, categorical regression can be used to combine data from multiple studies in multiple species. This means that the resulting ED10 (or EC-T10) might no longer be based solely on the most sensitive strain/species/sex, and may be more predictive of the actual human risk. This is because effects in animals are not perfectly predictive of the targets in humans, and the scatter of data in categorical regression can provide information on the uncertainty in the target, as well as uncertainty in the dose-response (Dourson, personal communication). As discussed above, the CatReg software allows for "stratification" by species, so that the dose-response curves of different species can be separated partially or completely; statistical features allow for the identification of the simplest model that fits the data well. Still, considerable judgement is needed in determining when the most sensitive species should be used, and when it is appropriate to combine data from different species. Mechanistic and other data on the relevance to humans of the most sensitive species are considered in the determination. This issue is addressed in EPA's RfC Guidelines[21], but additional research needs to be done in this area.

The meta-analytical aspects of categorical regression may also have implications on the appropriate choice of UF_A, the uncertainty factor for extrapolation from animals to humans. To date, categorical regression has used standard uncertainty factors for deriving any values, such as acute reference exposures (AREs) (Strickland et al., unpublished), or RfDs (Haber et al., unpublished). Thus, a factor of 3 has been used for extrapolation from animal studies when inhalation dosimetry is used, and a factor of 10 has been used for extrapolation from oral animal studies. However, the meta-analytical aspects might argue that a higher factor should be used (because the most sensitive species/strain/sex is no longer being used), or that a lower factor should be used (because a broader range of data are being considered, and there is less overall uncertainty). Further research in this area is needed.

Several other issues related to the application of categorical regression would also benefit from more in-depth investigation. An advantage of categorical regression is that it is the only quantitative modeling approach that takes into account increasing severity of response, not only increasing incidence of response. As noted by Hertzberg and Dourson[22], the implications of a BMD (10% response) is very different for a mild response than for a severe response. It would be useful to systematically compare how categorical regression and BMD modeling handle graded data.

While exposure duration is an integral aspect of categorical regression modeling of inhalation data, there has been little research on how to take duration into account for oral data. Hertzberg and Dourson[22] evaluated the dose-duration response for chronic human exposure to manganese, but saw no clear effect of duration. The RfD methodology uses a default uncertainty factor of 10 for extrapolation from subchronic to chronic exposure. Categorical regression might be used to evaluate the appropriate size of this uncertainty factor. Although this factor is unlikely to be needed if a good quality chronic study is available, a subchronic study may have evaluated key endpoints not considered in the chronic study. Additional research is needed on how to combine subchronic and chronic data for categorical regression modeling.

Acknowledgements

One of the authors (LH) is grateful to Michael Dourson and Bruce Allen for many enlightening conversations that helped frame my thoughts on the issues presented here.

LYNNE HABER[a*],
JUDY A. STRICKLAND[b†] and DANIEL J. GUTH[c‡]

[a]*Toxicology Excellence for Risk Assessment (TERA), Cincinnati, OH 45223,*
[b]*National Center for Environmental Assessment MD 52, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, currently at Integrated Laboratory Systems, Inc., Research Triangle Park, NC 27709 and*
[c]*The Boeing Company, P.O. Box 3707, Seattle, WA 98124, work done while at EPA/NCEA*

References

1. P. McCullagh, J.R. Statist. Soc. B. **42**, 109 (1980).
2. R.C. Hertzberg and M. Miller, Toxicol. Ind. Health **1**, 43 (1985).
3. R.C. Hertzberg, Health Physics **57**, Suppl. 1, 405 (1989).
4. W.F. ten Berge, A. Zwart, and L.M. Appelman, "Concentration-Time mortality response relationship of irritant and systemically acting vapours and gases," *J. Haz. Mat.*, **13**, 301–309 (1986).

* 513–542–7475 x17 Haber@TERA.org
† 919–316–4639 strickl z@niehs.nih.gov
‡ 425–865–6935 daniel.j.guth@boeing.com

5. M.L. Dourson, L.K. Teuschler, P.R. Durkin, *et al.*, "Categorical regression of toxicity data: A case study using aldicarb," *Reg. Toxicol. Appl. Pharmacol.*, **25**, 121–129 (1997).
6. L.K. Teuschler, M.L. Dourson, W.M. Stiteler, P. McClure and H. Tully. 1999. Health risk above the reference dose for multiple chemicals. Reg. Toxicol. And Pharmacol., 30: S19- S26.
7. D.J. Guth, A.M. Jarabek, L. Wymer and R.C. Hertzberg, Proc. 84th AWMA, 91–173.2 (1991).
8. B.D. Beck, R.B. Conolly, M.L. Dourson, D. Guth, D. Hattis, C. Kimmel and S.C. Lewis, Fund. Appl. Tox., **20**, 1 (1993).
9. D.G. Simpson, R.J. Carroll, M. Xie, and D.J. Guth, Commun. Statist.-Theory Meth., **25**, 2615 (1996a).
10. D.G. Simpson, R.J. Carroll, H. Zhou and D.J. Guth, J. Agr. Biol. Environ. Statistics **1**, 354 (1996b).
11. D.J. Guth, 1996; Proc. 89th AWMA, 96-TP6.02 (1996).
12. D.J. Guth, R.J. Carroll, D.G. Simpson, and H. Zhou, Risk Anal., **17**, 321 (1997).
13. U.S. Environmental Protection Agency. *CatReg software documentation.* Research Triangle Park, NC: Office of Research and Development, National Center for Environmental Assessment; EPA /600/R-98/053 (2000a).
14. U.S. Environmental Protection Agency. *CatReg software user manual.* Research Triangle Park, NC: Office of Research and Development, National Center for Environmental Assessment; EPA /600/R-98/052 (2000b).
15. M.L. Keplinger and L.W. Suissa. Am. Ind. Hyg. Assoc. J. **29**, 10 (1968).
16. M.L. Keplinger, Toxicol. Appl. Pharmacol. **14**, 192 (1969).
17. N. Eriksen, DOE/EV/03490-T3, U.S. Atomic Energy Commission, Pharmacology Report 435. NTIS DE85–010190 (1945).
18. K.S. Crump, Risk Analysis **15**, 79 (1995).
19. U.S. Environmental Protection Agency, *The Use of the Benchmark Dose Approach in Health Risk Assessment*, EPA/630/R-94/007, Office of Research and Development, Washington, DC (1995).
20. D.G. Barnes, G.P. Datson, J.S. Evans, *et al.*, "Benchmark dose workshop: Criteria for use of a benchmark dose to estimate reference dose," *Reg. Toxicol. Pharmacol.*, **21**, 296–306 (1995).
21. U.S. Environmental Protection Agency. *Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry*, National Center for Environmental Assessment; EPA/600/8–90/066F (1994).
22. R.C. Hertzberg, and M.L. Dourson. Using categorical regression instead of a NOAEL to characterize a toxicologist's judgment in noncancer risk assessment. In: Toxicology of Chemical Mixtures: Case Studies, Mechanisms and Novel Approaches, R.S.H. Yang, Ed. Academic Press, San Diego, CA. (1993).