|  |
| --- |
|  |
| Appendix A |
| **RESPONSE CURVE MODELING RATIONALE AND GUIDELINES** |
|  |
| David R. Mount |
|  |

|  |
| --- |
|  |

**RESPONSE CURVE MODELING RATIONALE AND GUIDELINES**

The following text was prepared to guide model selection for regression modeling and ECx estimation of chronic toxicity data, in particular for chronic data analyses supporting the OPP/OW Aquatic Effects Harmonization effort. The primary modeling software for this analysis is the Toxicity Response Analysis Program (TRAP) v1.22, developed by Russell J. Erickson and available from U.S. EPA at: <http://archive.epa.gov/med/med_archive_03/web/html/trap.html> The present document was prepared to supplement the guidance provided with TRAP, because there are many toxicity data sets that show interpretable exposure response information sufficient to make reasonable estimates of ECX values, but do not meet the strict assumptions and requirements of the regression models provided in TRAP. It is recognized that real world data sets present a wide variety of analysis challenges, some of which may not neatly fit within the categorizations and guidance below. Final model selection requires some level of professional judgement, which should be applied in a way that is conceptually consistent with the principles below. These guidelines should not be used to justify application of a modeling approach that is inconsistent with overall professional judgement.

1. Threshold sigmoid regression with log transformed exposure data is the default model. This selection is based on the observation that many response curves have an indication of a shoulder at low levels of effect, and that most exposure-response curves show greater symmetry when plotted as a function of log exposure rather than a linear exposure scale. The following is an example of the threshold sigmoid model:

**Log Exposure Concentration**

**Biological Variable**

1. When the control exposure concentration is reported as 0 or below analytical detection, one-tenth of the lowest treatment concentration is used for control to allow it to be plotted on a log axis (a value of 0 cannot be plotted on a log axis). The presumption underlying this approach is that control exposures are generally far lower than the experimental treatments, and that placement of the control at one-tenth of the lowest treatment generally should not unduly influence the regression result. For example, in the figure above, placement of the control response (leftmost point) at 0.5x, 0.1x, or 0.01x of the lowest treatment would not substantively affect the overall regression model. The analyst should conduct a visual evaluation of models for any indication that control placement is influencing regression and adjust the modeling approach according to professional judgement if necessary. One important check is to verify that the modeled response plateau at low exposures is reasonable in light of the control response (see use of 2-parameter models below for more discussion).
2. If growth is assessed as both weight and length, weight is the preferred variable unless there is a reason not to use it. If only length is reported, use the cube of the length (L3) as the response variable. The rationale for plotting the cube of length is rooted in allometry, with a desire to keep comparable levels of effect between growth measurements based on weight and on length. Given a three dimensional shape of fixed proportions and density, the mass of the shape will vary directly with the cube of any one of its three dimensions (length, height, width). As a result, a 20% reduction in length can be expected to cause a nearly 50% reduction in weight. Modeling length response as L3 results in an approximate normalization of levels of effect between length and weight (i.e., the EC20 for L3 should be roughly equivalent to the EC20 based on weight).
3. Response curves that are very sharp and contain one or zero treatments with intermediate effects cannot be fit with a unique solution using the threshold sigmoid model. In cases where there are insufficient data to inform the regression algorithm as to the slope and position of the response curve, piecewise regression and/or direct interpolation can be used to estimate ECx values. Examples of and approaches to such situations are described as follows:
	1. If the response is completely “all or none”, then interpolation between the points bracketing the response can be used to estimate ECx values. Example:

**Log Exposure Concentration**

**Biological Variable**

* 1. If there is a single intermediate response that does not provide much information as to the response curve shape near the EC20, then interpolation can be used by creating a piecewise fit that has the shallowest slope that still passes through all of the data points that either indicate a partial response or that bracket the range of the response curve. Example:

**Log Exposure Concentration**

**Biological Variable**

* 1. If, using a piecewise regression approach, the highest concentration in the “no response” portion of the response curve has a value that lies above the y intercept (i.e., the “no response” value for the biological variable, or Y0), then the interpolation is done using an inflection point with an X value equal to the highest tested concentration in the no-response portion of the curve, and a Y value equal to Y0.

**Log Exposure Concentration**

**Biological Variable**

1. ECx estimation should not involve substantial extrapolation beyond the documented responses. At the same time, we would like to avoid eliminating data sets that show evidence of an exposure-response but fall just short of a particular level of effect. To strike this balance, the following rules were applied to ECx estimation:
	1. If the data indicate a plausible exposure response curve but the highest exposure concentration causes an effect of ≥15 % (relative to the Y0 estimate) but <40%, an EC20 value is reported, but not an EC50. The following example shows a 15% effect at the highest concentration, which could be used to estimate an EC20 value, but not an EC50:

**Log Exposure Concentration**

**Biological Variable**

* 1. The above notwithstanding, ECx estimates should not be made if the variability within the “no response” portion of the curve casts doubt on whether the perceived downturn in the response curve is reasonably distinct from overall variability. For example, an EC20 value should not be inferred from the curve below, even though it could be considered as meeting the “at least 15%” criterion from the previous bullet:

**Log Exposure Concentration**

**Biological Variable**

* 1. If the data indicate a reasonable exposure response curve and the highest exposure concentration causes an effect of at least 40% (relative to the Y0 estimate), an EC50 can be estimated by extrapolation if the overall data set provides reasonable information regarding the apparent response curve.
	2. In general, interpolation of ECx values between the lowest tested concentration and the control is considered undesirable. If all of the treatment concentrations cause more than 15% effect, EC10 values should not be reported, unless the shape of the remaining response curve suggests strongly that the EC10 value can be reasonably estimated based on the response data at higher exposures. Likewise, EC20 values should not be reported where all treatment concentrations cause more than 25% effect, unless the shape of the remaining response curve suggests strongly that the EC20 value can be reasonably estimated based on the response data at higher exposures. EC50 values should not be interpolated between the lowest exposure concentration and control. The following graph shows a case where an EC10 may be reasonably estimated even though the EC10 is interpolated between the lowest concentration and the control:

**Log Exposure Concentration**

**Biological Variable**

1. Many data sets contain data for both negative control and a solvent or carrier control. Cases where responses in these two controls vary substantially cause uncertainty in how regression estimates should be made. Decisions of whether to estimate ECx values in these cases should be based on a judgement of how well the overall data set seems to present a reasonable exposure response curve. If the interpretation of a response curve is highly dependent on what is assumed regarding control response, ECx values should not be reported, at least within the range of effects influenced heavily by the ambiguity.
	1. This example shows a case in which regression analysis would be used despite substantial variability in the control data. The defining characteristic here is that even though there are disparities in control response, the low levels of exposure appear to define a response plateau that is not inconsistent with the control responses. Very similar models would result if neither control was used, or if either one was used alone, so the exact control response is not very important to the overall interpretation of the response curve:

**Log Exposure Concentration**

**Biological Variable**

* 1. The following three graphs show a data set with the same control variability shown in (6.a) above, but for which no regression estimates would be made because interpretation of the response curve is too dependent upon what is assumed to be “zero response.” Each of the three graphs shown are plausible depending on how one interprets the control data:

**Log Exposure Concentration**

**Biological Variable**

**Log Exposure Concentration**

**Biological Variable**

**Log Exposure Concentration**

**Biological Variable**

1. Particularly in cases where a response curve is shallow and does not have a clear plateau in response at lower levels of exposure, a sigmoid curve might be fit such that the control response is on a sloped portion of the response curve (rather than on the plateau) and that the maximum (plateau) value of the response variable is at a level higher than the maximum observed within the data set. In such cases, a two parameter model should be considered, in which case the maximum value for the response variable is set equal to the control response (or some other value suggested by the data set) and the regression fits only the slope and EC50. This approach forces the curve to adopt the control response (or other value selected as an appropriate estimate of the maximum response) as the plateau for the response curve. See the examples below for identical data sets. The first was analyzed with a three parameter model. The second was analyzed as a two parameter model. Note how the plateau is more fully formed in the two parameter model. It should also be noted that the second example (two parameter) provides a more conservative estimate of the EC20 relative to the control value.

**Biological Variable**

**Log Exposure Concentration**

**Mean Control Response**

**EC20**

**Biological Variable**

**EC20**

**Mean Control Response**

**Log Exposure Concentration**

8. Regressions generated by TRAP assume that the biological variable (e.g., survival, weight) is asymptotic to zero, meaning the value of the biological variable declines from a “control” level to zero as concentration increases. Particularly for weight (or length) data, this assumption is generally not the case, because in long-term exposures, most organisms must achieve a certain level of growth or they die. Even if the endpoint is actual growth and not weight, zero growth may not be an achievable result of exposure. In addition, treatments that experience substantial mortality may show growth that is inconsistent with that occurring at lower levels of exposure, either because of size-selective mortality, or because the reduced number of fish increases the per-fish consumption of food as a result of either greater ration per fish or decreased competition and associated behavioral effects.

Because of the above, decisions must be made regarding what data to include when modeling data related to growth. While including data for treatments with high mortality can be misleading, censoring too aggressively can limit the ability to model the growth response and derive associated EC20 values (if censoring leaves too few treatments to define a robust response curve).

The paired figures below illustrate the issue. In this case, the two highest concentrations caused 100% mortality, and therefore had zero weight. In the upper panel, the two treatments with 100% mortality were censored from the regression calculation (open symbols), and the resulting growth regression was a fairly shallow curve. In the lower panel, the two highest concentrations were included; the result is a very sharp curve, which does not have a shallow shoulder as suggested by the response profile in the upper panel. This difference occurs because the regression models in TRAP are symmetrical, and must therefore have the same degree of “shoulder” at both the top and bottom of the curve. The best compromise (mathematically) in this case is to use a steep slope and ignore the evidence for a shoulder at lower exposures.

As a general guideline for regressions of length or weight, treatments having high levels of mortality should be censored if including them appears to bias the shape of curve in a way that seems to misrepresent the response curve at lower levels of effect (e.g., EC10, EC20).

