A Multi-tiered Hierarchical Bayesian New Approach Method to Derive Toxic Equivalency Factors for Dioxin-Like Compounds

**Supplementary Technical Information A**

**Bayesian Dose-Response Modeling**

Caroline Ring,1a Alexander Blanchette\*2; William D. Klaren2; Seneca Fitch3; Laurie Haws1; Matthew W. Wheeler4; Michael DeVito5, Nigel Walker4, Daniele Wikoff2

1. ToxStrategies, Austin, TX, USA

2. ToxStrategies, Asheville, NC, USA

3. ToxStrategies, Katy, TX, USA

4. National Institute of Environmental Health Sciences/National Institutes of Health, Research Triangle Park, NC, USA

5. Environmental Protection Agency, Center for Computational Toxicology and Exposure, Research Triangle Park, NC, USA

a. Current affiliation with Environmental Protection Agency, Center for Computational Toxicology and Exposure, Research Triangle Park, NC, USA

Corresponding Author:

Daniele Wikoff\*

dwikoff@toxstrategies.com  
31 College Place, Suite B118

Asheville, NC 28801, USA

Phone: 828.348.6833

# Likelihood function

Different formulations of the likelihood function (different error models) were required for continuous responses and dichotomous responses. Within a group of datasets with a common reference dataset, responses are either all continuous or all dichotomous. The continuous and dichotomous likelihood functions both use the following definitions: Each group of datasets includes non-reference congener datasets (where may be different for each group), and one dataset for the reference compound (usually TCDD, sometimes PCB126). The compounds in a dataset may be indexed as , where indexes the reference compound. In the dataset for the *j*th congener, there are dose/concentrations, indexed . For each of these dose/concentrations, responses are typically reported in summary form. For continuous responses, the following are reported: sample mean response, , sample response standard deviation, , and number of subjects, . (If responses are reported individually, they can be treated as equivalent to “summary” reporting, with the individual response, and .) For dichotomous responses, the following are reported: number of occurrences of the outcome, , and number of subjects, .

For continuous responses, the residual errors were modeled as independent and identically distributed (i.i.d.), obeying a zero-mean normal distribution. The continuous likelihood function was formulated using a parameterization of the normal density function suitable for data reported in summary form (EPA, 2012)(Equation A1).For computational efficiency, the (natural) log transformation of the normal density function is used to give the log-likelihood.

Equation A1

In Equation A1, is the constant error variance for the group of datasets.

For dichotomous responses, the likelihood function is formulated by assuming that data are i.i.d following a binomial distribution (Equation A2). For computational efficiency, the (natural) log transformation of the binomial density function is used to give the log-likelihood.

Equation A2

In both Equation A1 and Equation A2, is given by Equation A3. For continuous responses, is the model-predicted response for dose/concentration for the dataset for compound j. For dichotomous responses, is the model-predicted probability of the dichotomous outcome occurring for dose/concentration for the dataset for compound .

Equation A3

In both Equation A1 and Equation A2, the inner summation (index ) gives the log-likelihood for one dataset within a group; the outer summation (index ) gives the total log-likelihood across all datasets in the group. In the outer summation, runs from 1 to to index the datasets for the non-reference congeners as well as the reference congener.

### Prior probabilities for parameters

Hierarchical priors were implemented for the Hill model parameters (Supplemental Table 1). Constraints were placed on some parameters. Hillslope was constrained to be greater than 0.6 to allow for the possibility of supralinear responses while still preventing situations where the slope of the curve becomes infinitely large as dose or concentration approaches zero, which may not make physical sense. No endpoint allowed a physically meaningful negative response value in the REP2004 database, so both Bottom and Top were constrained to be non-negative. Other constraints depended on the type of response (continuous or dichotomous) and the direction of the curve (increasing or decreasing with dose/concentration). No constraints were enforced for Log10EX50. Since it is a log-scale dose/concentration variable, it may be positive, negative, or zero. For computational reasons, Stan performs better when estimating parameters that are unconstrained. Therefore, for parameters with constraints, priors were specified for “raw” versions of the Hill model parameters. These “raw” versions were then transformed to yield parameters with the needed constraints. For parameters whose value depends strongly on the scale of the data, the transformed parameters were further scaled using the range of dose/concentration values or the range of response values in each dataset.

The prior distributions for the “raw” versions of the parameters, and all transformation and scaling applied to yield the “final” versions of the parameters (used in the log-likelihood function), are detailed in Supplemental Table 1. In Supplemental Table 1, we use “tilde” notation for density functions: means that the probability of is given by the Normal density function, with mean and standard deviation , evaluated at . We also use notation for truncated distributions: means that the probability of is given by a truncated normal distribution with mean and standard deviation , truncated below at and truncated above at , evaluated at .

# Standardization of inferred parameters

After inference, each of the 9000 sampled sets of parameters for each group of datasets is standardized, relative to the reference compound parameters.

The standardized congener for the jth non-reference congener is given in Equation A4.

Equation A4

The standardized congener Top for the jth non-reference congener is given in Equation A5.

Equation A5

The standardized reference is always 0, and the standardized reference Top is always 1. Reference and congener Hillslopes remain the same for the standardized curves as for the non-standardized curves.

Supplemental Table 1. Priors and hierarchical priors for Hill model parameters. These apply to one group of datasets that share a common reference dataset. There are J non-reference congeners in the group, indexed j = 1, 2, …, J. The first column indicates which compound the parameter applies to (reference and all congeners; reference only; all non-reference congeners; or non-reference congener j.)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **For which compound in this group of datasets** | **Parameter** | **Raw Prior\*** | **Transformation & Scaling\*\*** | | **Effective bounds after transformation** |
| **Response Type** |  |  |
| Reference and all congeners |  |  | Continuous increasing |  |  |
| Continuous decreasing |  |  |
| Dichotomous increasing |  |  |
| Dichotomous decreasing |  |  |
| Reference only |  |  | Continuous increasing |  |  |
| Continuous decreasing |  |  |
| Dichotomous increasing |  |  |
| Dichotomous decreasing |  |  |
| Non-reference congener *j* |  |  | Continuous increasing |  |  |
| Continuous decreasing |  |  |
| Dichotomous increasing |  |  |
| Dichotomous decreasing |  |  |
| All non-reference congeners |  |  | — | | — |
| All non-reference congeners |  |  | — | | — |
| Reference only |  |  |  | |  |
| Non-reference congener *j* |  |  |  | |  |
| All non-reference congeners |  |  | — | | — |
| All non-reference congeners |  |  | — | | — |
| Reference only |  |  |  | |  |
| Non-reference congener *j* |  |  |  | |  |
| All non-reference congeners |  |  | — | | — |
| All non-reference congeners |  |  | — | | — |
| Reference and all congeners |  |  |  | |  |

\* Superscript prime (‘) indicates “raw” versions of parameters that are later transformed. The transformed version is used in the log-likelihood function and/or Hill equation.

Distribution notation is as follows:

represents the logistic density function with location parameter and scale parameter ,

represents the normal density function with mean and standard deviation ,

appended to a distribution means that distribution was truncated with lower bound and upper bound .

\*\* Additional scaling parameters are defined as follows.

: The average response at the minimum dose/concentration for the group of datasets (control response if available).

: The full range of responses for the reference dataset in the group.

: The full range of responses for the jth congener (non-reference) dataset in the group.

: half the range of log10-transformed doses/concentrations in the reference dataset in the group.

: The midpoint of the range of log10-transformed doses/concentrations in the reference dataset in the group.

: half the range of log10-transformed doses/concentrations in the jth congener (non-reference) dataset in the group.

: The midpoint of the range of log10-transformed doses/concentrations in the jth congener (non-reference) dataset in the group t.  
Although scaling by these parameters may not affect the bounds of the transformed variables, it improves computational performance by transforming the central-tendency location of the “raw” distribution to a value on a scale that at least roughly matches the data, making it more likely that the higher-density parts of the posterior distribution will correspond to higher-density parts of the “raw” prior distributions.

**References (Supplementary Technical Information A)**

EPA, U. S. (2012). *Benchmark Dose Technical Guidance*. https://www.epa.gov/risk/benchmark-dose-technical-guidance