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

**Supplementary Technical Information B**

**Bayesian Meta-Analysis**

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

The likelihood and priors for this Bayesian hierarchical model are defined as follows. In the following, we use JAGS-like or Stan-like “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 JAGS-like or Stan-like 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 .

## Congener-specific level

At the congener-specific level (top row in Figure 7 of the main text), the priors for the congener-specific standardized Hill model parameters are given by Equation B1.

Equation B1

Note that these apply to standardized Hill model parameters on the log scale (see Equation B2). These priors have the following implications:

* Congener-specific EX50 is always positive, and is equally likely to be anywhere between and
* Congener-specific Top is most likely to be 1. It is between 0 and 1 with 40% probability, and greater than 1 with 60% probability. It is always positive.
* Congener-specific Hillslope and TCDD Hillslope are both log-normally distributed, meaning that they are always positive. Hillslope is equally likely to be less than 1 or greater than 1.

These priors were selected for mathematical plausibility: to ensure that standardized Hill model parameters are positive (as required to ensure that predicted standardized responses are always positive), and to reflect the reasonable assumption that standardized Hill model parameters are unlikely to take on extreme values (e.g. Hillslope is unlikely to take a value of ), while still allowing a wide range of mathematically-plausible parameter values (*i.e.* to be as weakly-informative as possible).

## REP-specific level

At the REP-specific level (second row in Figure 7 of the main text), the priors for the REP-specific standardized parameters are given by Equation B2. Because study quality category is probabilistic (probabilities of each category estimated by the random forest model in step 1), the prior distribution for author-derived REP from study *k* of congener *j* is a weighted mixture of the distributions for each study quality category, , where the weights are the RF-model-predicted probabilities that study *k* is assigned to quality category *m*.

Equation B2

The study-quality-specific standard deviation for parameter p, where , is given by the piecewise definition in Equation B3.

Equation B3

The prior for each of the study-quality-specific standard deviations is a wide half-Cauchy distribution as given in Equation B4 (Gelman, 2006).

Equation B4

## Sample-specific level

At the sample-specific level (third row in Figure 7 of the main text), the likelihood functions are formulated separately depending on the type of REP: DR REP or author-derived REP.

 For DR REPs, the likelihood for the *i*th MCMC sample from the posterior distribution of each (ln-transformed) parameter for REP study *k* of congener *j* (Equation B5) is given by assuming that the residuals between the MCMC samples and the REP-specific parameter values are normally distributed, with a REP-specific standard deviation for each parameter. If the reference compound was PCB126 instead of TCDD, then the means of the distributions for and are adjusted appropriately to account for the fact that the parameters were standardized to PCB126 instead of to TCDD.

Equation B5

Priors for the REP-specific standard deviation for each parameter are uninformative wide uniform distributions (Equation B6).

Equation B6

For author-derived REPs, the likelihood function was defined by comparing the reported REP to , the REP predicted by evaluating Equation 14 (in the main text) using the REP-specific sDR Hill model parameters , , and for the method of REP derivation reported for (REP BMDF or REP EDF, with reported fraction ).

If was derived as a REP BMDF, then the REP-specific sDR parameters must produce a that matches the quantifiability of the observed author-derived REP: otherwise, the likelihood is zero. That is, if was quantifiable, then should be equal to or greater than . If if was non-quantifiable, then should be less than .

If was derived as a REP BMDF and quantifiability of matches quantifiability of , or if was derived as a REP EDF (which are always quantifiable), then the likelihood function is defined by assuming that obeys a normal distribution whose mean is with a REP-specific standard deviation .

Quantifiability for sample *i* of author-derived REP *k* of congener *j* was encoded as a binary value: if the REP was quantifiable, and if it was NQ. The likelihood function is defined in Equation B7. (Subscript *i* is used in Equation B7 to emphasize that author-derived REPs are indexed at the sample level, even though for author-derived REPs, always.)

Equation B7

where

The prior for the REP-specific error standard deviation is given by a wide half-Cauchy distribution (Equation B8).

Equation B8

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515-534. https://doi.org/10.1214/06-ba117a