ECOTOX.SCIENCE - TECHNICAL REPORT

Toward a Unified Framework for Toxicity Metrics: Introducing the Equivalent Effect Concentration (EEC)

David R. Fox, Jannicke Moe*

2026-01-07

Status: Working draft technical report - not peer reviewed and may change substantially **Version:** 0.3-draft

*Contributors: Walter Zobel, Raoul Wolf

Contents

Availability and Citation

This technical report is archived on Zenodo and can be cited as:

Fox, D. R., & Moe, J. (2026). *Toward a Unified Framework for Toxicity Metrics: Introducing the Equivalent Effect Concentration (EEC).*
DOI: https://doi.org/10.5281/zenodo.18168375

1 Introduction

Ecotoxicological risk assessment relies heavily on toxicity thresholds derived from controlled experiments. These thresholds provide the basis for the establishment of environmental quality guidelines and for regulatory decision-making. While not wishing to revisit the long debates and controversy surrounding the use of flawed toxicity metrics like the NOEC and LOEC, it suffices to say that the last 20 years has witnessed a growing awareness and acceptance of the superiority of statistical approaches which **model** the relationship between **response** and **dose** rather than ignoring it - as is done in ANOVA-based methods.

The regression approach has the advantage of providing quantitative links between chemical exposure and biological effects and as noted by Fox and Landis (2016) even in datasets deemed "problematic," regression models offer a richer, more robust framework for the estimation of point and interval toxicity thresholds than ANOVA-based methods (Fox & Landis, 2016). This, however is not a shared view with others arguing there are real-world scenarios where hypothesis-testing metrics like the NOEC are suitable and even preferable (Green et al., 2013).

A variety of statistical modelling methods is available and, collectively these fall under the umbrella **Dose–response (D-R)** modelling. Responses are typically *sigmoidal* and can be well-described mathematically by the logistic, log–logistic, Weibull, or Hill equations, enabling estimation of a variety of metrics (Hendriks et al., 2013; Ritz, 2010). Some of the more common of these are discussed in the next section.

2 Common toxicity metrics used in ecotoxicology

2.1 Categorical versus Ratio metrics

Toxicity metrics can be classified as having been derived using one of two common statistical methods - **Regression** or **ANOVA / T-tests**. While both techniques are variants of what statisticians refer to as the *general linear model* or *GLM* (as distinct from *generalised* linear models), their treatment of the *dose* variable is decidedly different. Regression-based methods (as used in dose-response modelling) treat dose as a **ratio** variable whereas ANOVA-based methods treat dose as a **categorical** variable (also referred to as a 'factor' in R). To distinguish these two estimation paradigms we will denote $R-type$ and $C-type$ for regression and categorical methods respectively. Apart from other advantages to be discussed, $R-type$ methods are infinitely flexible whereas $C-type$ methods are limited to a small number of rigid procedures such as *Dunnett's test* (Dunnett, 1955), *Jonckheere–Terpstra test* (Jonckheere, 1954; Terpstra, 1952), or *Williams test* (Williams, 1971).

$R-type$ estimates use model-based estimation that accounts for the entire dose–response curve. In this way, it is possible to incorporate *both* statistical significance and biological significance. These methods are statistically robust and have been widely adopted in regulatory practice. Nonetheless, quantification of *biological significance* is generally more difficult than the specification of statistical significance. Furthermore, calculation of the lower confidence bound for an $R-type$ metric usually relies on either numerical approximations (e.g. the **delta-method**) or computer simulation (e.g.bootstrapping). The results are also sensitive to the form of the model and to data sparsity near the target response level.

2.2 Formal Definitions

R − type **metrics**

- LC_x **(Lethal Concentration,** $x\%$**)**
  The concentration of a substance that is lethal to $x\%$ of test organisms over a specified exposure period. The LC_{50} (median lethal concentration) remains the most widely reported lethal toxicity endpoint (Hendriks et al., 2013).

- EC_x **(Effect Concentration,** $x\%$**)**
  The concentration that results in an $x\%$ *'effect'* (usually defined with reference to a zero or *control* dose, but not necessarily).

  Low-effect levels such as EC_{10} or EC_{20} are recommended for regulatory use because they are thought to be biologically meaningful yet precautionary. Derived from regression models, EC_x values utilise the full concentration–response curve (Ritz, 2010).

- **NEC (No-Effect Concentration)**
  A model-derived threshold parameter representing the concentration below which no effect is predicted. The NEC is estimated as a parameter of a dose–response mode - often via maximum likelihood or Bayesian methods (Fisher et al., 2024), and is accompanied by confidence or credible intervals. Being regression-based and not limited to tested concentrations, it is a significant improvement over the NOEC (Fox, 2010a).

- **NSEC (No-Significant-Effect Concentration)**
  Similar to the NEC, the NSEC is a regression-based analog of the NOEC (Fisher & Fox, 2023a). While the NEC assumes a threshold response mechanism, the NSEC does not. Although computationally possible, the NSEC should not be computed from threshold models as it is based on a different model parameter to the NEC.

  The NSEC retains the operational logic of "no significant effect" while addressing some of the more serious concerns with the NOEC (Fisher & Fox, 2023a).

- **BMD (Benchmark Dose)**
  The dose corresponding to a predefined *benchmark response (BMR)*, such as a 10% increase in incidence (quantal data) or a 5–10% relative change in mean response (continuous data). The BMD is an *inverse-regression* problem which requires either an analytic or numerical solution to the equation

$$\widehat{BMD} = \left\{ x : f\left(x; \hat{\Theta}\right) = y \right\} \tag{2.1}$$

  where $\hat{\Theta}$ is the vector of parameter estimates for the dose-response model $f(\cdot)$ and y is a response value defined by the BMR.

  Unlike the NOEC, NEC and NSEC which are estimates of *no (significant) effect*, the BMD and BMDL (see below) integrate biologically meaningful effects via the specification of the BMR. (Crump, 1984; Slob, 2002).

- **BMDL (Benchmark Dose Lower Bound)**
  The lower one-sided confidence (frequentist) or credible (Bayesian) interval bound on the BMD, typically at the 95% level. Regulatory agencies adopt the BMDL as the operative point of departure from the control response, ensuring precaution while explicitly incorporating statistical uncertainty (EFSA Scientific Committee, 2017, 2022; U.S. Environmental Protection Agency, 2012). In practice, the BMDL is now the international standard for human health and increasingly for ecotoxicological risk assessment.

- HC_x **(Population toxicity metric: Hazardous Concentration for x% of species)**
  Derived from a species sensitivity distribution (SSD), the HC_5 (hazardous to 5% of species) is most widely used. The HC_5 provides a probabilistic *population-level* threshold and underpins water-quality guidelines in many jurisdictions (Sánchez-Bayo & Goka, 2007).

- **PNEC (Predicted No-Effect Concentration)**
  A regulatory threshold intended to protect most species in an ecosystem. PNECs are typically derived from SSD-HC_5 with assessment factors when sufficient data are available, or from single-species EC_x/NOEC values with larger safety factors under data-limited conditions (European Chemicals Agency (ECHA), 2017).

Note: *Although sounding very similar, the **PNEC** and **NEC** are conceptually and operationally quite different in ecotoxicology. The **NEC** is a model-derived threshold parameter estimated directly from a single-species dose–response dataset whereas the **PNEC** is a regulatory benchmark concentration intended to be protective of an entire ecosystem. It is not tied to a single species but is derived from multiple lines of evidence.*

$C-type$ **metrics**

- **NOEC (No-Observed-Effect Concentration)**
  Defined as the highest **tested** concentration at which the mean response is not statistically different from the mean control response at a pre-specified level of significance (α). The NOEC is constrained to tested doses, depends strongly on replication and variance, and lacks uncertainty quantification. It has been heavily criticized as a toxicity metric (Fox, 2009, 2010b; Jager, 2012).

- **LOEC (Lowest-Observed-Effect Concentration)**
  The lowest tested concentration producing a statistically significant effect relative to control. Like NOEC, LOEC is design-dependent and subject to the same criticisms.

- **MATC (Maximum Acceptable Toxicant Concentration)**
  The MATC is calculated as the geometric mean of the NOEC and LOEC which is simply $\sqrt{NOEC \cdot LOEC}$. However the claim that this represents the highest concentration that does not harm aquatic organisms is naive. By definition, the MATC is the *mid-point* between the NOEC and LOEC **on a *log* scale**.

  The phrase "Maximum Acceptable Toxicant Concentration" can be traced back to the 1960s when NOEC/LOEC-based methods were in common use (Mount & Stephan, 1967).

  Being a mathematical conflation of two flawed metrics, the MATC is possibly even less useful and interpretable than either alone.

The MATC has historically been used in the United States to derive chronic aquatic life criteria, but its reliance on NOEC/LOEC inherits their statistical weaknesses.

2.3 Limitations of Traditional Metrics

Traditional metrics (NOEC, NSEC, ECx etc.) are often interpreted from a statistical significance or *effect magnitude* standpoint, but rarely combine both. They do not formally incorporate the notion of **equivalence**, i.e., demonstrating that an effect is *negligibly different* or *biologically indifferent* to the control. Moreover, they treat the problem asymmetrically: the burden is to prove harm, not to demonstrate safety within a biologically acceptable margin.

This asymmetry arises because classical null hypothesis tests are designed to detect differences from control and failure to detect such a difference is not the same as showing that the treatment is similar. In contrast, equivalence testing explicitly inverts the burden of proof: it allows researchers to declare that an effect is small enough to be considered safe, within a predefined margin of practical equivalence. This makes equivalence-based approaches like the EEC fundamentally more aligned with regulatory needs for affirming safety.

3 Jurisdictional Preferences

3.1 Australia and New Zealand

The **ANZECC/ARMCANZ (2000)** water-quality guidelines historically relied on NOEC and LOEC values. More recent practice, however, reflects international developments and emphasizes regression-based metrics (M. St. J. Warne et al., 2025). Current guidance favors the use of the NEC, low-effect EC_{10} or EC_{20} values, and SSD-derived HC_5 concentrations as the primary inputs to guideline derivation (M. St. J. Warne et al., 2025).

3.2 Canada

The Canadian Council of Ministers of the Environment (**CCME**) employs a tiered strategy. When sufficient high-quality data are available, SSD-based HC_5 values are the preferred metric for deriving water-quality objectives. In cases where data are more limited, chronic EC_x estimates are combined with assessment factors to provide protective concentrations. While NOEC values are still found in legacy guidance, recent Canadian practice discourages their use in favor of regression-based metrics (Canadian Council of Ministers of the Environment (CCME), 2007).

3.3 United States

The US EPA historically applied the **MATC** in aquatic life criteria derivation. However, the agency's **Benchmark Dose Technical Guidance** identifies the BMD framework, and particularly the BMDL, as the scientifically preferred point of departure. Although NOEC/LOEC remain embedded in some regulatory

programs, BMD and BMDL are now increasingly used in ecological as well as human health risk assessments (U.S. Environmental Protection Agency, 2012).

3.4 European Union

Under the REACH framework, the **European Chemicals Agency (ECHA)** discourages reliance on NOEC values. Instead, $SSD - \mathrm{HC}_5$ and $R - type$ metrics such as the EC_x or BMDL are used to derive Predicted No-Effect Concentrations (PNECs) (European Chemicals Agency (ECHA), 2017). Guidance from EFSA explicitly endorses BMDL as the point of departure for ecological and health risk assessments, reflecting a strong preference for $R - type$ approaches (EFSA Scientific Committee, 2017, 2022).

3.5 Historical and Regulatory Context

The *effect concentration* (EC_x) metric has long been used as a toxicity metric in ecotoxicology and pharmacology. An ECx is the concentration associated with a specified percentage change in response (e.g., EC10 for a 10% reduction in growth or reproduction), interpolated directly from a fitted dose–response model. This simplicity made ECx attractive in experimental toxicology, particularly as an alternative to $C - type$ metrics such as the *no-observed-effect concentration (NOEC)* and *lowest-observed-effect concentration (LOEC)*. However, EC_x values were historically often reported as point estimates without confidence intervals, limiting their usefulness in regulatory risk assessment (M. S. J. Warne et al., 2018). Although statistical methods such as the delta method or bootstrapping can be applied to estimate confidence bounds for EC_x, this has not been a consistent practice in the ecotoxicological literature.

In the 1990s, the *benchmark dose (BMD)* framework emerged within human health risk assessment and was subsequently adopted in ecological contexts. The *U.S. Environmental Protection Agency* (*Benchmark Dose Technical Guidance*, 2012) and later the *OECD* (*Guidance Document on the Use of the Benchmark Dose Approach in Risk Assessment*, 2020) promoted the BMD approach as a replacement for NOEC/LOEC. The BMD is mathematically equivalent to an ECx for the same benchmark response definition (see §4.1), but it is embedded in a regulatory framework that requires formal uncertainty analysis. Specifically, the *lower confidence bound (BMDL)* on the BMD is designated as the critical effect level for deriving guideline values, ensuring a conservative and statistically defensible basis for risk assessment. The BMD approach also introduced standardized benchmark responses, such as a 10% change or one standard deviation departure from control, improving cross-study consistency and comparability (EFSA Scientific Committee, 2017).

The practical distinction between ECx and BMD therefore lies less in the mathematical definition and more in the treatment of uncertainty. Both metrics are point estimates derived from the same concentration–response function, and both can be accompanied by confidence bounds estimated via profile likelihood, bootstrap, or delta method. The difference is that the *BMD framework institutionalized the use of confidence bounds*, with BMDL serving as the regulatory point of departure. This requirement is reinforced through guidance documents and dedicated software platforms such as BMDS (US EPA) and PROAST (RIVM) (National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, 2023), which standardize the implementation of BMD procedures.

In summary, the BMD_x for a given benchmark response (BMR) and the EC_x are mathematically equivalent when $x \equiv BMR$. The distinction is primarily procedural and historical: ECx arose as a descriptive toxico-

logical endpoint, while the BMD evolved into a regulatory framework designed to replace NOEC/LOEC by embedding uncertainty analysis. The adoption of the BMDL as the protective effect level reflects the broader shift in ecotoxicology and risk assessment from descriptive statistics toward inferential, uncertainty-based decision-making (*Benchmark Dose Technical Guidance*, 2012; EFSA Scientific Committee, 2017; *Guidance Document on the Use of the Benchmark Dose Approach in Risk Assessment*, 2020; M. S. J. Warne et al., 2018).

3.6 Summary

The balance of evidence from both statistical literature and regulatory practice supports a decisive move away from $C-type$ metrics. These metrics are strongly design-dependent, lack confidence intervals, and may misrepresent true effect thresholds (Fox, 2009; Jager, 2012). Although some authors defend their continued use (Green et al., 2013), most scholars and agencies have concluded that $R-type$ approaches provide more reliable and biologically meaningful estimates.

$R-type$ metrics offer transparency, reproducibility, and explicit uncertainty quantification. Given its pervasiveness in the United States and the European Uninion, it could be argued that the benchmark dose (BMD) approach, and its associated lower confidence bound (BMDL), represents the international standard for deriving points of departure. In parallel, community-level protection is most often established through SSD-based HC_5 values and derived PNECs. Together, these developments mark a paradigm shift in ecotoxicology away from $C-type$thresholds toward statistically rigorous, $R-type$ metrics.

4 Similarities and Differences between toxicity metrics[1]

4.1 BMD *versus* ECx

The similarities between the BMD and the EC_x are readily apparent from the definitions provided in section 2.1. Mathematically we we can show equivalence when $BMR \equiv x$ as follows.

Suppose we fit a parametric model $f(x;\Theta)$ to describe the mean response at concentration x where Θ is a vector of model parameters.

The benchmark response (BMR) is defined as an $x\%$ change from the control mean response, $f(0;\Theta)$.

The EC_x is similarly defined as the solution to:

$$f(EC_x;\Theta) = f(0,\Theta)\left(1 - {}^{x}/_{100}\right)$$

Thus,

$$EC_x = \{c : f(c,\Theta) = y^*\}$$

where

$$y^* = \left(1 - {}^{x}/_{100}\right) \cdot f\left(0,\hat{\Theta}\right)$$

[1] Without loss of generality, we assume a *decreasing* relationship between dose and response in the remainder of this document.

.

The BMD is defined *identically* with

$$y^* = \left(1 - \frac{BMR}{100}\right) f\left(0, \hat{\theta}\right)$$

.

So we see the EC_x and BMD are **identical** when $x = BMR$.

4.2 BMDL/ECx *versus* NSEC

Without loss of generality, let the first element of $\hat{\Theta}$, $(\hat{\theta}_1)$, be the estimated response-axis intercept (i.e. response at $dose = 0$) having standard error $SE\left(\hat{\theta}_1\right)$. Now, the $NSEC$ is the solution to:

$$\left\{x : f\left(x; \hat{\Theta}\right) = \hat{\theta}_1 - t_{\nu,1-\alpha} \cdot SE\left(\hat{\theta}_1\right)\right\}$$

where $t_{\nu,1-\alpha}$ is the $1-\alpha$ quartile of the t distribution having ν degrees of freedom.

But $\hat{\theta}_1 = f\left(0, \hat{\Theta}\right)$ and so the previous equation can be written as:

$$\left\{x : f\left(x; \hat{\Theta}\right) = f\left(0, \hat{\Theta}\right) - t_{\nu,1-\alpha} \cdot SE\left(\hat{\theta}_1\right)\right\}$$

Thus, both the BMD and EC_x are defined in terms of a *relative* change from the control whereas the $NSEC$ uses an *absolute* change. Equivalence between the BMD and $NSEC$ occurs for

$$\hat{\theta}_1 \cdot \frac{\delta}{100} = t_{\nu,1-\alpha} \cdot SE\left(\hat{\theta}_1\right)$$

or, equivalently:

$$\delta' = t_{\nu,1-\alpha} \cdot \ cv\left(\hat{\theta}_1\right) \ \ 0 < \delta' < 1$$

where $\delta' = \frac{\delta}{100}$ is either BMR or x and $cv\left(\hat{\theta}_1\right)$ is the *coefficient of variation* for $\hat{\theta}_1$. Thus:

$$NSEC = \begin{cases} BMD & if \quad BMR = \delta' \\ EC_x & if \quad x = \delta' \end{cases}$$

4.3 Reconciling $R-type$ and $C-type$ toxicity metrics

An often overlooked requirement of $C-type$ methods is that they **demand** *replication* at each dose level. This is critical to the way, for example, ANOVA works - it tests for the equality among treatment *means* via a comparison of *variances* - namely the **within dose** estimate of the error variance σ^2 and the **between dose** estimate of σ^2.

If there is no dose effect (implying, on average, the same response at all doses) then the two estimates will be approximately equal and the ratio $F = \frac{MSE_{between}}{MSE_{within}}$ will (within sampling error) be unity.

However, it can be shown that when there *is* a dose effect, $MSE_{between}$ will be greater than MSE_{within} and so $F = \frac{MSE_{between}}{MSE_{within}} > 1$. Whether the computed F statistic is *signicantly* greater than 1 is assessed by

reference to a critical value from the F distribution with a significant result leading to the rejection of the null hypothesis of no dose effect.

The mathematical details are as follows.

4.3.1 The ANOVA framework

We observe k distinct doses $\{x_1, \ldots, x_k\}$, and for simplicity, assume **equal** replication $n \geq 2$. The total sample size is therefore

$$N \;=\; n\,k$$

The standard, one-way ANOVA model is

$$y_{ij} \;=\; \mu(x_j) + \varepsilon_{ij}, \qquad i = 1, \ldots, n, \;\; j = 1, \ldots, k,$$

with independent errors satisfying $E(\varepsilon_{ij}) = 0$ and $\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$.

The group (dose) means are:

$$\bar{y}_{\cdot j} \;=\; \frac{1}{n}\sum_{i=1}^{n} y_{ij}, \qquad j = 1, \ldots, k.$$

Now, the **within-dose** estimate of σ^2 (also known as the 'pure error' estimate) is:

$$SS_{\mathrm{W}} = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_{\cdot j}\right)^2,$$

$$\mathrm{df}_{\mathrm{W}} = N - k,$$

$$MS_{\mathrm{W}} = \frac{SS_{\mathrm{W}}}{N-k}.$$

and the **between-dose** estimate of σ^2 is:

$$SS_{\mathrm{B}} = n\sum_{j=1}^{k}\left(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}\right)^2,$$

$$\mathrm{df}_{\mathrm{B}} = k - 1,$$

$$MS_{\mathrm{B}} = \frac{SS_{\mathrm{B}}}{k-1}.$$

A generic, *post-ANOVA* contrast is evaluated via the statistic T given by:

$$T_j = \frac{\bar{y}_{\cdot j} - \bar{y}_{\cdot 0}}{\hat{\sigma}_{\mathrm{ANOVA}}\sqrt{2/n}},$$

where the variance estimate is

$$\hat{\sigma}^2_{\mathrm{ANOVA}} = MS_{\mathrm{W}}, \quad df_{\mathrm{W}} = N - k.$$

Dunnett's test (Dunnett, 1955) which is one of many multiple comparison techniques used to tease out

differences between *pairs* of responses following the rejection of the null hypothesis of no dose effect. It differs from other comparison tests in that it only looks at differences between the control response and the response at the remaining dose levels. Importantly, it adjusts the level of significance of each of the pairwise comparisons so that the *overall* family-wise error rate ($FWER$) equals the nominated α.

The significance of each comparison is assessed by comparing the computed value of T for the sample data with a critical value from Dunnett's distribution having denominator $df = N - k$.

As previously noted, the ANOVA approach is model-agnostic - by reducing *dose* to a categorical variable, the structural form of the relationship between response and dose is ignored. In the case of $n = 1$ (i.e. a single observation per dose level), the ANOVA model is *saturated*, meaning there are as many model parameters as data values and hence $N = k$ leaving $N - k = 0$ degrees of freedom with which to estimate the error variance. With replication, the model is not saturated — it estimates one mean per dose group and leaves residual variation within groups. In most dose-response experiments both n and k are small, meaning df_w is also small and, as we shall see shortly, this has implications for the precision with which σ^2 is estimated as well as the power of Dunnett's test.

In contrast to ANOVA-based methods, regression-based toxicity estimates *model* the observed relationship between dose and response and aim to do so using as few parameters as possible.

4.3.2 The $R - type$ framework

We fit a parametric dose response mean function $\eta(x;\theta)$ with p estimable parameters. Fitted means at observed doses are:

$$\hat{\eta}_j \;=\; \eta(x_j;\hat{\theta}), \qquad j = 1, \ldots, k.$$

The adequacy of the fitted model is assessed using information contained in the *residual sum of squares* and the *mean squared residual* where:

$$SS_{\text{res}} = \sum_{j=1}^{k}\sum_{i=1}^{n}(y_{ij} - \hat{\eta}_j)^2,$$

$$\text{df}_{\text{res}} = N - p,$$

$$MS_{\text{res}} = \frac{SS_{\text{res}}}{N - p}.$$

Unlike the $C - type$ framework, replicating measurements at each dose is **not** a requirement. As argued by Fox (Fox et al., 2016), in the context of dose-response experimentation where resources are constraining and the objective is to get a good 'fix' on the functional form of the dose-response relationship, replication is wasteful. Rather than replicating at each dose, Fox (Fox et al., 2016) suggests a more useful approach would be to spread the experimental effort across the dose continuum, although how to do this in an 'optimal' manner requires advanced statistical skills.

When replication has been used in a dose-response experiment, advantage can be taken of the extra information provided **not** on the functional form of the relationship, but whether the relationship is **useful**. This 'extra' information comes about from a further partitioning of the residual sum of squares (SS_{res}) into a **'pure error'** component (SS_{pe}) and a **'lack of fit'** component (SS_{lof}).

It is shown in Appendix 5 that this decomposition of both the sums of squares and the degrees of freedom is a simple additive one, that is:

$$SS_{\text{res}} = SS_{\text{pe}} + SS_{\text{lof}}$$

with:

$$df_{pe} = N - k$$
$$df_{lof} = k - p$$
$$df_{res} = df_{pe} + df_{lof} = N - p$$

Two things follow from this decomposition: (i) the *significance* of the lack-of-fit can be assessed by comparing the ratio

$$MS_{lof}/MS_{pe}$$

with a critical value from $F_{k-p,N-k}$; and (ii) if the parametric model is correct, $SS_{lof} = 0$ and hence $MS_{res} = \frac{SS_{res}}{N-p} = \frac{SS_{pe}}{N-p} = \frac{N-k}{N-p} MS_{pe}$.

It is further shown in Appendix 5 that if the parametric model is correct, the *theoretical* variances of MS_{pe} and MS_{res} are, respectively:

$$\text{Var}(MS_{\text{pe}}) = \frac{2\sigma^4}{N-k}, \qquad \text{Var}(MS_{\text{res}}) = \frac{2\sigma^4}{N-p}.$$

and for $p < k$ it follows that $\text{Var}(MS_{\text{res}}) < \text{Var}(MS_{\text{pe}})$

4.3.3 Discussion: Implications for Dunnett's Test

The inequality

$$\text{Var}(MS_{\text{res}}) < \text{Var}(MS_{\text{pe}})$$

(for $p < k$) implies that *if* the parametric model is correct, the residual mean square from the fitted dose–response model provides a **more precise estimate of** σ^2 than the pure-error mean square from the one-way ANOVA treatment.

In the context of multiple comparisons, this matters because **Dunnett's test** relies on the ANOVA error mean square as its variance estimator. To see this, we have from Appendix 2 that for a one-sided test, the *per-comparison power* for treatment i is:

$$Power = 1 - \text{pt}\left(c_\alpha, \nu, \lambda\right)$$

where $\text{pt}\left(c_\alpha, \nu, \lambda\right)$ is the *cdf* of a noncentral T distribution evaluated at c_α (critical value of Dunnett's test) and the non-centrality parameter (NCP), λ and degrees of freedom ν are given as:

$$\lambda_i\left(\sigma\right) = \frac{\Delta_i}{\sigma\sqrt{1/n_i + 1/n_0}},\ \nu = \left(\sum_{j=0}^{k} n_j\right) - (k+1)$$

It is evident that this power varies inversely with σ. For a *regression* model, the estimate of σ^2 is $\hat{\sigma}_R = MS_{res}$, while for an *ANOVA* model it is $\hat{\sigma}_A = MS_{res}$, but for a correctly specified dose-response model

$MS_{res} = \frac{N-k}{N-p} MS_{pe} \Rightarrow MS_{res} < MS_{pe}$ for $p < k$ and so $\hat{\sigma}_R < \hat{\sigma}_A$. Thus, the power for detecting differences from the control will be greater for a correctly specified dose-response model than the power of a Dunnett's type test.

However, this gain depends critically on correct model specification. If the model is mis-specified, lack-of-fit inflates MS_{res}, reducing power and potentially compromising error control.

4.3.4 Power and threshold implications

- $C - type$ **toxicity metrics**
  Because MS_{pe} is estimated with only $N - k$ degrees of freedom, the variance estimate is less precise. This inflates the confidence intervals around the treatment means. As a result, the observed difference from control must be larger to reach statistical significance. In practice this means the NOEC will tend to be **higher** (less sensitive), which is an undesirable outcome (hence the claim that the NOEC 'rewards' poor experiments).

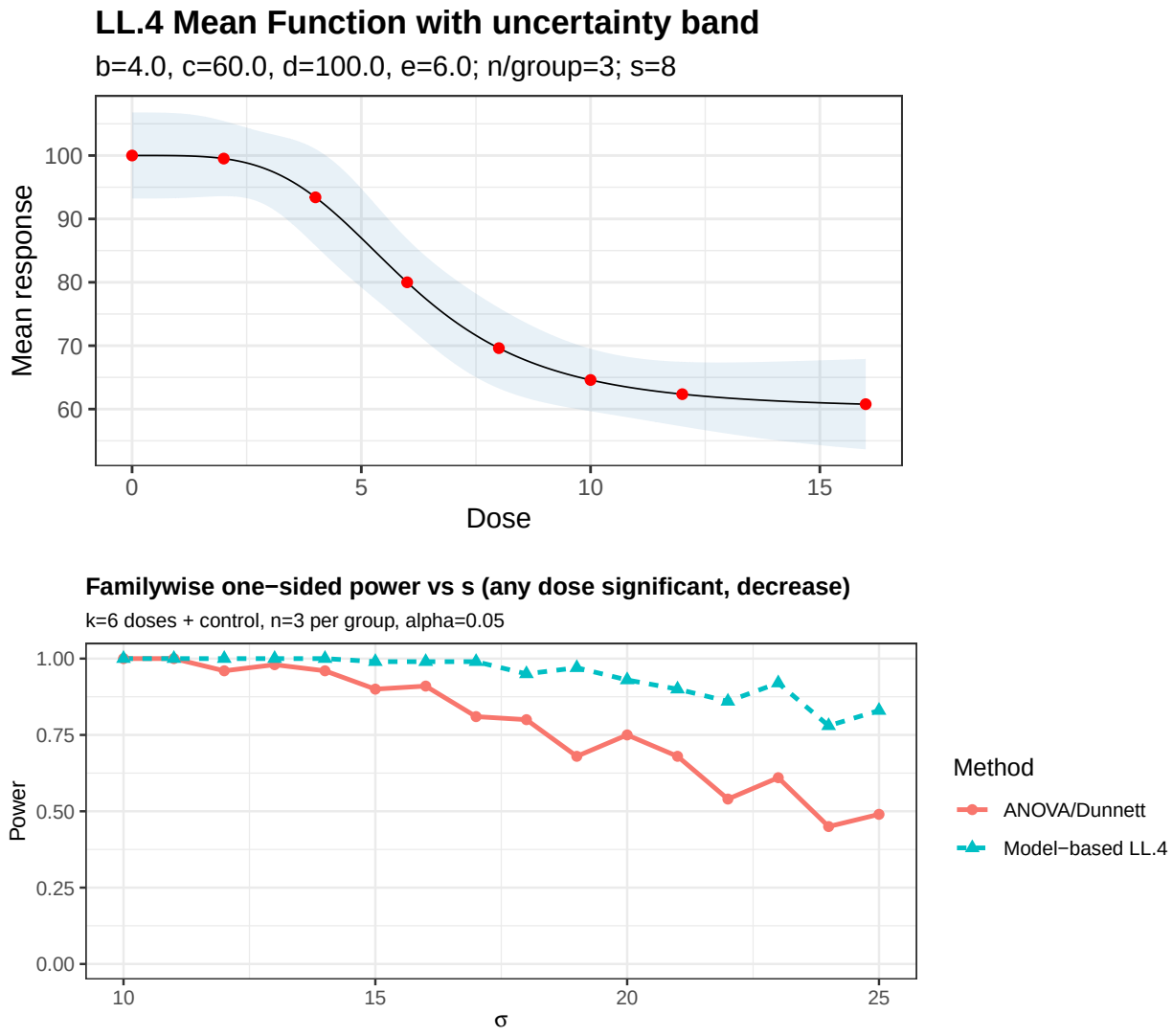
- $R - type$ **toxicity metrics**
  With $N - p$ degrees of freedom, MS_{res} is a more precise variance estimate *when the model lack-of-fit is low*. The lower bound *lower* for the intercept is therefore sharper, leading to a smaller (more sensitive) toxicity estimate. Under correct model specification, this provides a more precise threshold.

- **Trade-off:** If the parametric model is misspecified, $SS_{\text{lof}} \gg 0$ inflates MS_{res}, leading to an **inflated** toxicity estimate.

4.3.5 Example

In this example we evaluate the statistical power of two different approaches for determining effect concentrations in dose–response experiments. The R code simulates data under a specified model and then applies both the traditional ANOVA with Dunnett's test and a model-based benchmark dose approach.

By repeating the simulation many times, we can estimate the probability that each method will correctly identify a treatment effect at the chosen significance level. The output therefore provides a direct comparison of the operating characteristics of the two methods, highlighting situations in which the model-based approach may offer higher power due to its use of the full dose–response curve, while the ANOVA approach may be more limited by its reliance on pairwise contrasts and the associated variability in the pure-error mean square.



This example illustrates the difference how differences in underlying assumptions about variance estimation translate into differences in sensitivity for detecting ecotoxicologically meaningful effects.

---

5 Equivalence Testing - Introducing the Equivalent Effects Concentration (EEC)

5.1 Overview of Equivalence Testing

Equivalence testing reverses the traditional hypothesis testing framework. Rather than testing whether an effect differs from zero, equivalence testing evaluates whether an observed effect lies within a pre-specified margin (δ) of a reference value (Lakens, 2017; Schuirmann, 1987). It addresses the question: is the observed effect sufficiently close to the reference value to be considered biologically unimportant?

Mathematically:

$$\begin{aligned} H_0 : &\quad |effect| \geq \delta \\ H_A : &\quad |effect| < \delta \end{aligned}$$

This test is often implemented via the Two One-Sided Tests (TOST) procedure as follows.

Split the null hypotheses into two, one-sided tests:

$$\begin{aligned} H_{01} : &\quad effect \leqslant -\delta \\ H_{11} : &\quad effect > -\delta \end{aligned}$$

and

$$\begin{aligned} H_{02} : &\quad effect \geqslant +\delta \\ H_{12} : &\quad effect < +\delta \end{aligned}$$

The two one-sided tests procedure is that if we conclude that *effect* $> -\delta$ **and** *effect* $< +\delta$, then it has effectively been concluded that $-\delta < \textit{effect} < +\delta$.

We show in the Appendix 1 that the TOST procedure is operationally identical to the procedure of declaring equivalence only if the $1 - 2\alpha$ confidence interval for the response is completely contained within the equivalence interval $[-\delta, +\delta]$.

Equivalence condition (90% CI within margin):

$$CI_{(1-2\alpha)} \subseteq [-\delta, +\delta] \Rightarrow \text{Conclude equivalence}$$

5.2 Defining the EEC

For a monotonically decreasing dose-response relationship (common in ecotoxicology), where the response **decreases** with increasing dose (e.g., survival, growth,reproduction), the Equivalent Effects Concentration (EEC) is defined as:

$$\text{EEC} = \min\left\{d \mid CI_{\hat{f}(d)} \subseteq [R - \delta, R + \delta]\right\}$$

where R is the reference value - typically $f\left(0, \hat{\Theta}\right)$ in regression-based models.

That is, the **smallest concentration** at which the **entire confidence interval** for the predicted mean response lies within a pre-defined equivalence band around the nominal response R (typically the control response).

This ensures that the EEC reflects a **conservative and protective threshold**, marking the lowest concentration beyond which responses can no longer be confidently declared as biologically equivalent to control. It aligns with regulatory goals of defining a **safe concentration** with high statistical and ecological confidence.

5.3 Positioning the EEC Among Other Metrics

The EEC not only compliments, but as is shown in the next section, subsumes exisiting metrics. - Like NOEC/NSEC, it identifies a safe concentration. - Like BMDL and NSEC, it uses dose-response modeling. - Unlike either the NOEC or NSEC, it provides a direct test of equivalence rather than absence-of-effect or defined-effect. The EEC is thus especially suitable in regulatory or guideline contexts where the goal is to demonstrate safety with high confidence.

5.4 Common toxicity Metrics as Variants of the Equivalent Effects Concentration (EEC)

The Equivalent Effects Concentration (EEC) provides a unifying framework for interpreting traditional toxicity metrics. Defined as the lowest concentration at which the confidence interval for the predicted response lies within a biologically acceptable margin (δ) around a reference response (typically the control), the EEC captures both statistical certainty and biological relevance.

All commonly used metrics can be understood as variants or limiting cases of the EEC, differing only in how δ is defined, how uncertainty is incorporated, or whether the comparison is relative or absolute.

5.4.1 EC_x and BMD as Deterministic EECs

It has already been noted that the EC_x is equivalent to a BMD with x replaced by the benchmark response (BMR). Both metrics specify a fixed relative effect size but do not require that the confidence interval of the prediction be contained within an equivalence band. In this sense, the EC_x and BMD can be regarded as point-estimate versions of the EEC, representing deterministic thresholds that omit the statistical safeguard of equivalence testing.

5.4.2 BMDL as a Confidence-Bounded EEC

The BMDL is the lower confidence limit on the BMD. By reintroducing uncertainty, the BMDL is effectively a one-sided confidence-interval version of the EEC, with δ tied to the chosen BMR. Conceptually, the BMDL occupies a middle ground between the purely deterministic BMD and the fully conservative EEC, acknowledging variability while still linking the margin directly to a fixed percentage change.

5.4.3 NSEC as an Absolute-Change EEC

The NSEC is defined as the dose where the predicted response equals the lower confidence bound of the control mean. This formulation makes the NSEC an absolute-change metric, in contrast to the relative-change definitions of BMD and EC_x. The NSEC is therefore equivalent to an EEC in which δ as defined in §4 is determined by the product of the standard error of the control and an appropriate t-statistic. In this way, the NSEC can be understood as an absolute-margin EEC anchored on the variability of the fitted control response.

5.4.4 NOEC as a Discrete-Dose Approximation of the EEC

The No-Observed-Effect Concentration (NOEC) is determined by a sequence of Dunnett-style multiple comparisons of treatments to control, using the ANOVA error mean square (MS_{pe}) as the variance estimator. Within the EEC framework, the NOEC corresponds to a situation in which δ is implicit in the size of the confidence intervals around treatment means. Unlike the model-based approaches, the NOEC depends strongly on the experimental design, particularly dose spacing and replication, and it employs a noisier variance source (based on MS_{pe} rather than the model-based MS_{res}). Consequently, it functions as a low-power, categorical approximation to the EEC that is sensitive to design choices.

5.4.5 NEC as the Zero-Margin EEC

The NEC derives from threshold or segmented regression models, representing the concentration below which responses are indistinguishable from control. Within the EEC framework, this corresponds to the limiting case in which $\delta \to 0$. The NEC can therefore be regarded as the strictest variant of the EEC, demanding exact equivalence with control and tolerating no departure from the reference level.

5.4.6 Summary

The relationships among these metrics can be summarized in the following table.

Metric	Relation to EEC	Key Difference
EC_x / BMD	Deterministic point-estimate EEC	Ignores CI containment
BMDL	One-sided CI EEC	δ tied to chosen BMR
NSEC	Absolute-change EEC	$\delta = SE(\text{control}) \times t$
NOEC	Discrete-dose EEC	δ implicit in ANOVA contrasts
NEC	Zero-margin EEC	$\delta \to 0$

5.4.7 Implications

This reframing demonstrates that the EEC subsumes all traditional toxicity metrics. Each metric arises from specific assumptions about the equivalence margin δ, the treatment of uncertainty, and the modeling approach. By embedding all metrics within the EEC framework, regulators and practitioners gain a consistent and transparent basis for comparing and interpreting toxicity thresholds.

5.5 Strengths and weakenesses of the EEC

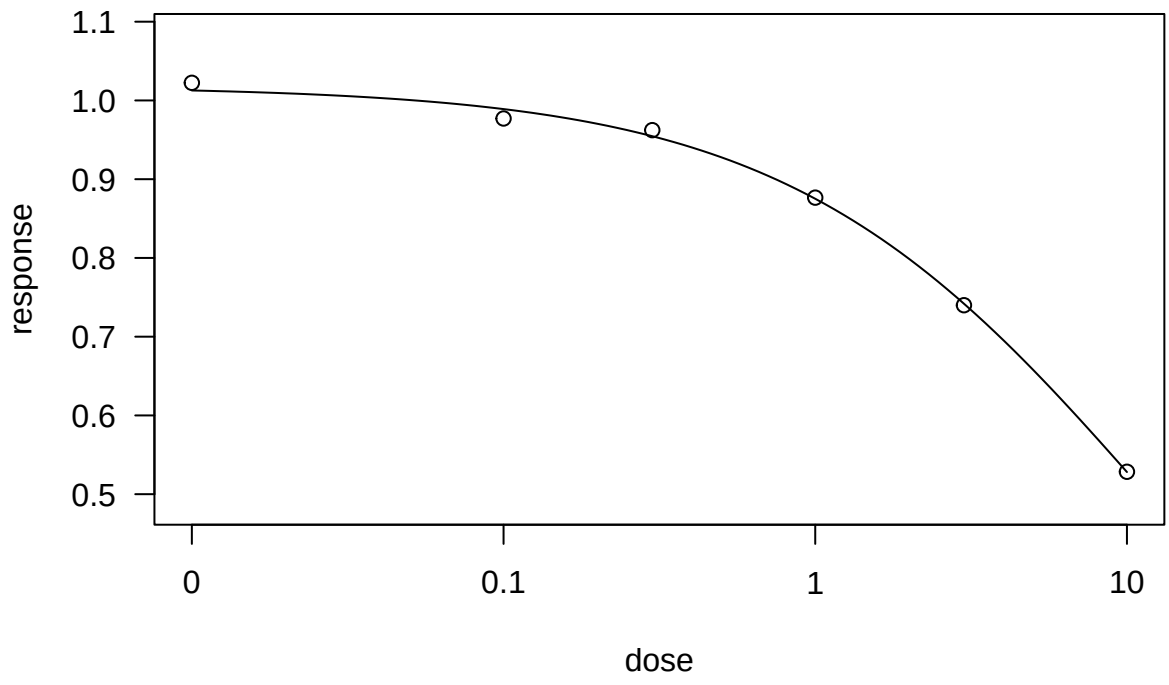
Strengths: *(as for BMDL) but additionally:* - Directly answers the regulatory question: *is the response at the EEC (biologically) equivalent to the control response?* - Avoids pitfalls of *p-values* although still uses a NHST framework - albeit in a slightly different manner (see § 2). - Grounded in effect size and margin of relevance.

Weaknesses: - Depends on selection of δ. - Requires root-finding and bootstrapping to estimate CIs.

6 Worked Example in R

We demonstrate the EEC approach using a synthetic dataset and the `drc` package in R.

```r
library(drc)

# Simulated data
set.seed(123)
data <- data.frame(
  dose = rep(c(0, 0.1, 0.3, 1, 3, 10), each = 6),
  response = c(
    rnorm(6, 1.00, 0.05),
    rnorm(6, 0.98, 0.05),
    rnorm(6, 0.96, 0.05),
    rnorm(6, 0.90, 0.05),
    rnorm(6, 0.75, 0.05),
    rnorm(6, 0.50, 0.05)
  )
)

# Fit dose-response model
fit <- drm(response ~ dose, data = data, fct = LL.4())
plot(fit)
```




```r
summary(fit)
```

```
##
## Model fitted: Log-logistic (ED50 as parameter) (4 parms)
##
```

```
## Parameter estimates:
##
##                Estimate Std. Error t-value   p-value
## b:(Intercept)  0.744313   0.229627  3.2414  0.002778 **
## c:(Intercept) -0.032973   0.665873 -0.0495  0.960814
## d:(Intercept)  1.017991   0.018136 56.1315 < 2.2e-16 ***
## e:(Intercept) 11.999005  19.148626  0.6266  0.535350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error:
##
##  0.0457373 (32 degrees of freedom)
```

```
ED(fit, 10, interval = "delta")  # Estimated dose for 10% effect
```

```
##
## Estimated effective doses
##
##        Estimate Std. Error    Lower    Upper
## e:1:10  0.62676    0.48405 -0.35921  1.61273
```

```
# Define equivalence margin
delta <- 0.1

# Estimate control response
alpha_hat <- predict(fit, newdata = data.frame(dose = 0))

# Predict over fine dose grid
dose_grid <- exp(seq(log(0.01), log(10), length.out = 500))
preds <- predict(fit, newdata = data.frame(dose = dose_grid), interval = "confidence")

# Identify EEC: highest dose where CI is within [alpha - delta, alpha + delta]
eec_pass <- preds[, 2] >= (alpha_hat - delta) & preds[, 3] <= (alpha_hat + delta)
EEC <- max(dose_grid[eec_pass], na.rm = TRUE)

cat("EEC:", round(EEC, 3), "\n")
```

```
## EEC: 0.366
```

6.1 Graphical Comparison of metrics

For this comparison, we use data in Table 1 (Fisher & Fox, 2023b).

Table 2: Fish growth across a range of concentrations of an unknown pollutant

Concentration	Rep 1	Rep 2	Rep 3
0	6.59	6.14	7.19
1	4.91	5.03	6.25
2	5.89	7.44	6.11
5	4.51	6.75	5.69
11	4.52	5.52	3.71
25	2.54	1.76	0.21
50	0.10	0.00	0.91

Note. There are three independent replicates for each concentration treatment. Rep = replicate.

6.1.1 Explanation of `R` code (see §`6.1.11` below)

To run: 1. copy the code to a file 2. use the `R` command `source(filename)` where `filename` is whatever you called the file in 1.

User Settings The top of the script sets key parameters: - `delta_prop`: δ as a fraction of the control response (e.g. 10%). - `alpha_one`: Significance level (e.g. 0.05). - `B`: Number of bootstrap replicates. - `endpoint_decreases`: Whether the endpoint is expected to decrease with increasing dose. - `data_file`: CSV file with columns `dose`, `response`, etc. - `eec_default`: Whether to use one-sided or two-sided equivalence.

6.1.2 EEC Type Selection

Allows interactive choice (if running in a console) between: 1. One-sided non-inferiority (EEC_95) 2. Two-sided equivalence via TOST (EEC_90) Sets internal flags to control which version of EEC to compute.

6.1.3 Load Data and Fit Model

- Reads in the dataset.
- Renames `dose` to `conc`.
- Fits a 4-parameter log-logistic dose–response model using `drc::drm()`.

6.1.4 Predicted Values and Root-Finding Helpers

Defines two utility functions: - `get_pred()`: safely extract fitted value and standard error for a given concentration. - `safe_root()`: a wrapper for `uniroot()` that returns `NA` if the function doesn't cross zero or contains `NA`s.

6.1.5 Basic Quantities

Extracts control response (`alpha_hat`) and standard error.
Computes: - `delta`: the effect size margin. - `tcrit95`: t-quantile for 95% confidence. Also defines a function `effect()` to calculate the treatment effect relative to control.

6.1.6 NSEC Calculation

- `NSEC_95` is computed as the concentration at which the predicted mean drops below the lower 95% CI of the control.
- `NSEC_delta` is where the predicted response equals control minus δ. Bootstrap functions (`boot_NSEC95`, `boot_NSECd`) are used to compute confidence intervals.

6.1.7 BMDL Calculation

Defines a bootstrap function that computes the concentration at which the response differs from control by δ. Returns the **5th percentile** of bootstrap estimates as the `BMDL_95`.

6.1.8 EEC Calculation

- For **EEC_95**, computes where the **upper** bound of the effect crosses δ.
- For **EEC_90**, computes where the **absolute effect ± margin** crosses δ (TOST logic). Uses root-finding over predicted values, then bootstraps this estimate.

6.1.9 Results Table

Outputs a summary table showing: - Point estimates for NSEC, BMDL, EEC - 95% bootstrap confidence intervals (except for BMDL, which uses 5%) Enables comparison of where each threshold falls along the concentration axis.

6.1.10 Plotting

The final plot: - Shows the fitted dose–response curve with 95% CI ribbon - Adds a dashed horizontal line at the control ± δ margin - Adds vertical lines for each metric - Includes a legend mapping line styles and colors to each metric This visualization helps highlight differences in how each threshold is defined and where it falls.

6.1.11 R Code for comparing metrics

```
# source(knitr::purl("toxicity_metrics.Rmd", output = tempfile()))
############################################################
## Unified toxicity metrics - choose 1-sided or 2-sided EEC
############################################################
library(knitr)
library(kableExtra)
library(tidyverse)

## ---------- USER SETTINGS --------------------------------------------
delta_prop   <- 0.10      # delta as fraction of control mean
alpha_one    <- 0.05      # one-sided alpha  (95 % bounds)
B            <- 10     # bootstrap replicates
endpoint_decreases <- TRUE
#data_file    <- "~/Budapest_2024/OSLO/Equivalence/NSEC_paper.txt"
data_file   <- "~/Budapest_2024/OSLO/Equivalence/growth.csv"

## ---------- CHOOSE EEC TYPE ------------------------------------------
eec_default <- "1"  # "1" = one-sided 95 %, "2" = two-sided 90 %
if (interactive()) {
  cat("\nWhich EEC?\n",
      "  1. one-sided 95 % non-inferiority (upper bound .LE. delta)\n",
      "  2. two-sided 90 % TOST equivalence (|effect| .LE. delta)\n", sep = "")
  ans <- menu(c("One-sided 95 %", "Two-sided 90 %"), graphics = FALSE)
  choice <- if (ans == 0) eec_default else as.character(ans)
} else {
  choice <- eec_default
}
use_two_sided <- identical(choice, "2")

## ---------- PACKAGES -------------------------------------------------
suppressPackageStartupMessages({
  library(drc); library(boot); library(ggplot2)
})
boot_quiet <- function(...) suppressWarnings(boot(...))

# ------------------------------------------------------------------
#  PROGRESS-BAR bootstrap wrapper (quiet + bar)
# ------------------------------------------------------------------
boot_pb <- function(data, statistic, R, title = "Bootstrapping", ...) {
  if (.Platform$OS.type == "windows") {
    pb <- winProgressBar(title = title,
                         label = "Progress...",
                         min = 0, max = R, width = 300)
    i <- 0L
    stat_wrap <- function(d, idx) {
      i <<- i + 1L
      setWinProgressBar(pb, i, label = sprintf("%s: %d of %d", title, i, R))
      statistic(d, idx)
    }
    res <- suppressWarnings(boot::boot(data, stat_wrap, R = R, ...))
    close(pb)
    cat("\n")
```

```
    return(res)
  } else {
    pb <- txtProgressBar(min = 0, max = R, style = 3)
    i <- 0L
    stat_wrap <- function(d, idx) {
      i <<- i + 1L
      setTxtProgressBar(pb, i)
      statistic(d, idx)
    }
    res <- suppressWarnings(boot::boot(data, stat_wrap, R = R, ...))
    close(pb)
    cat("\n")
    return(res)
  }
}

## ---------- DATA & MODEL ------------------------------------------
mydata <- read.csv(data_file)
#mydata$response <- with(mydata, r / n)
names(mydata)[names(mydata) == "dose"] <- "conc"
mod <- drm(response ~ conc, data = mydata, fct = LL.4())

## ---------- HELPERS -----------------------------------------------
get_pred <- function(model, x) {
  pr <- tryCatch(predict(model, data.frame(conc = x), se.fit = TRUE),
                 error = function(e)
                   tryCatch(predict(model, data.frame(conc = x), seFit = TRUE),
                            error = function(e2) NA))
  if (is.list(pr) && !is.null(pr$fit))              list(fit = pr$fit,  se = pr$se.fit)
  else if (is.matrix(pr) || is.data.frame(pr))      list(fit = pr[,1],  se = pr[,2])
  else if (is.numeric(pr) && length(pr) == 2)       list(fit = pr[1],   se = pr[2])
  else                                              list(fit = NA_real_, se = NA_real_)
}
safe_root <- function(fun, interval) {
  fL <- fun(interval[1]); fR <- fun(interval[2])
  if (anyNA(c(fL, fR)) || fL * fR > 0) return(NA_real_)
  uniroot(fun, interval)$root
}

## ---------- BASIC QUANTITIES --------------------------------------
rng        <- range(mydata$conc)
control_c  <- min(mydata$conc)
pred0      <- get_pred(mod, control_c)
alpha_hat  <- pred0$fit
se_alpha   <- pred0$se
delta      <- delta_prop * alpha_hat
tcrit95    <- qt(1 - alpha_one, df = mod$df.residual)

effect <- function(c) {
  pr <- get_pred(mod, c)
  if (anyNA(c(pr$fit, pr$se))) return(list(g = NA, se = NA))
  g <- if (endpoint_decreases) alpha_hat - pr$fit else pr$fit - alpha_hat
  list(g = g, se = pr$se)
}
```

```r
g_upper <- function(c) { with(effect(c), g + tcrit95 * se) }

## ---------- NSEC & BMDL ESTIMATES + CIs -------------------------------
root_NSEC95 <- safe_root(function(c) get_pred(mod, c)$fit -
                           (alpha_hat - se_alpha * tcrit95), rng)
L_alpha_delta <- (alpha_hat - delta) - tcrit95 * se_alpha
root_NSECd <- safe_root(function(c) get_pred(mod, c)$fit - L_alpha_delta, rng)

boot_NSECd <- function(data, idx) {
  m <- tryCatch(drm(response ~ conc, data = data[idx, ], fct = LL.4()),
                error = function(e) return(NA_real_))
  ctl <- get_pred(m, min(data$conc)); if (anyNA(ctl$fit)) return(NA_real_)
  a <- ctl$fit; seA <- ctl$se
  L_ad <- (a - delta_prop * a) - tcrit95 * seA
  safe_root(function(c) get_pred(m, c)$fit - L_ad, range(data$conc))
}
boot_NSEC95 <- function(data, idx) {
  m <- tryCatch(drm(response ~ conc, data = data[idx, ], fct = LL.4()),
                error = function(e) return(NA_real_))
  ctl <- get_pred(m, min(data$conc)); if (anyNA(ctl$fit)) return(NA_real_)
  a <- ctl$fit; seA <- ctl$se
  safe_root(function(c) get_pred(m, c)$fit - (a - seA * tcrit95),
            range(data$conc))
}
#cat("Bootstrapping NSEC_delta …\n")
#NSECd_CI  <- quantile(na.omit(boot_pb(mydata, boot_NSECd,  R = B)$t), c(0.025, 0.975), names = FALSE)
cat("Bootstrapping NSEC_delta …\n")
NSECd_CI <- quantile(
  na.omit(boot_pb(mydata, boot_NSECd, R = B, title = "Bootstrapping NSEC_delta")$t),
  c(0.025, 0.975),
  names = FALSE
)

cat("Bootstrapping NSEC_95 …\n")
#NSEC95_CI <- quantile(na.omit(boot_pb(mydata, boot_NSEC95, R = B)$t), c(0.025, 0.975), names = FALSE)
NSEC95_CI <- quantile(
  na.omit(boot_pb(mydata, boot_NSEC95, R = B, title = "Bootstrapping NSEC_95")$t),
  c(0.025, 0.975),
  names = FALSE
)

boot_BMDL <- function(data, idx) {
  m <- tryCatch(drm(response ~ conc, data = data[idx, ], fct = LL.4()),
                error = function(e) return(NA_real_))
  ctl <- get_pred(m, min(data$conc)); if (anyNA(ctl$fit)) return(NA_real_)
  a <- ctl$fit
  gfun <- function(c) {
    pr <- get_pred(m, c); if (is.na(pr$fit)) return(NA_real_)
    if (endpoint_decreases) a - pr$fit else pr$fit - a
  }
  safe_root(function(c) gfun(c) - delta, range(data$conc))
}
cat("Bootstrapping BMDL_95 …\n")
#BMDL <- quantile(na.omit(boot_pb(mydata, boot_BMDL, R = B)$t), 0.05, names = FALSE)
```

```r
BMDL <- quantile(
  na.omit(boot_pb(mydata, boot_BMDL, R = B, title = "Bootstrapping BMDL_95")$t), 0.05,names = FALSE)

## ---------- EEC ESTIMATE & CI ----------------------------------------
if (use_two_sided) {
  tcrit_eec <- qt(1 - 0.10/2, df = mod$df.residual)
  eec_label <- "EEC_90"
  g_eec <- function(c) {
    ef <- effect(c); if (anyNA(unlist(ef))) return(NA_real_)
    abs(ef$g) + tcrit_eec * ef$se
  }
} else {
  tcrit_eec <- tcrit95
  eec_label <- "EEC_95"
  g_eec     <- g_upper
}
root_EEC <- safe_root(function(c) g_eec(c) - delta, rng)

boot_EEC <- function(data, idx) {
  m <- tryCatch(drm(response ~ conc, data = data[idx, ], fct = LL.4()),
                error = function(e) return(NA_real_))
  ctl <- get_pred(m, min(data$conc)); if (anyNA(ctl$fit)) return(NA_real_)
  a <- ctl$fit
  gfun <- function(c) {
    pr <- get_pred(m, c)
    if (anyNA(c(pr$fit, pr$se))) return(NA_real_)
    g <- if (endpoint_decreases) a - pr$fit else pr$fit - a
    if (use_two_sided) abs(g) + tcrit_eec * pr$se else g + tcrit_eec * pr$se
  }
  safe_root(function(c) gfun(c) - delta, range(data$conc))
}
cat("Bootstrapping", eec_label, "…\n")
#EEC_CI <- quantile(na.omit(boot_pb(mydata, boot_EEC, R = B)$t),
#                   c(0.025, 0.975), names = FALSE)
EEC_CI <- quantile(
  na.omit(boot_pb(mydata, boot_EEC, R = B, title = eec_label)$t),
  c(0.025, 0.975),
  names = FALSE
)

## ---------- RESULTS TABLE --------------------------------------------
results <- data.frame(
  Metric   = c("NSEC_95",  "NSEC_delta", "BMDL_95%", eec_label),
  Estimate = c(root_NSEC95, root_NSECd,  BMDL,       root_EEC),
  CI_lower = c(NSEC95_CI[1], NSECd_CI[1], NA,        EEC_CI[1]),
  CI_upper = c(NSEC95_CI[2], NSECd_CI[2], NA,        EEC_CI[2])
)
# print(results, digits = 4)
kable(results, format = "latex", booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "hold_position"))

#results %>%
#  gt() %>%
#  fmt_number(
```

```r
#    columns = where(is.numeric),  # or use specific names like columns = c("EEC", "NSEC", "BMD")
#    decimals = 3
#  )

## ---------- LEGEND PREP ------------------------------------------------
order_levels <- results$Metric[order(results$Estimate, na.last = TRUE)]
grid <- data.frame(conc = seq(rng[1], rng[2], length.out = 200))
grid_pred <- t(sapply(grid$conc, function(x) unlist(get_pred(mod, x))))
grid$fit  <- grid_pred[, 1]
grid$upr  <- grid$fit + tcrit95 * grid_pred[, 2]
grid$lwr  <- grid$fit - tcrit95 * grid_pred[, 2]
hline <- if (endpoint_decreases) alpha_hat - delta else alpha_hat + delta

vlines <- data.frame(
  Metric = factor(order_levels, levels = order_levels),
  x      = results$Estimate[match(order_levels, results$Metric)]
)
legend_segs <- data.frame(
  Metric   = vlines$Metric,
  x        = rng[1],
  xend     = rng[1] + diff(rng)*0.08,
  y        = hline,
  yend     = hline,
  row.names = NULL
)
cols <- setNames(grDevices::rainbow(length(order_levels), v = .9, s = .85),
                 order_levels)
lts_base <- c(`NSEC_95` = "11", `NSEC_delta` = "31",
              `BMDL_95%` = "22", `EEC_95` = "44", `EEC_90` = "44")
lts <- lts_base[order_levels]
label_vec <- setNames(
  sprintf("%s (%.2f)", order_levels,
          results$Estimate[match(order_levels, results$Metric)]),
  order_levels
)

## ----------SET UP POSITIONS OF CONFIDENCE BANDS FOR EACH METRIC----------

band_data <- subset(results,
                    Metric %in% c("EEC_95", "NSEC_95", "NSEC_delta") &
                      !is.na(CI_lower) & !is.na(CI_upper))

# Drop unused factor levels
band_data$Metric <- droplevels(factor(band_data$Metric))

names(band_data)[names(band_data) == "Metric"] <- "Metric"  # ensure matching

# Define bounds for the bands
band_data$xmin <- band_data$CI_lower
band_data$xmax <- band_data$CI_upper
band_data$ymin <- 0
band_data$ymax <- max(mydata$response) * 1.05

# Full set of ordered metric names
```

```
order_levels <- results$Metric[order(results$Estimate)]

# Color palette for all metrics (includes BMDL, but 'well subset later)
cols <- setNames(rainbow(length(order_levels), v = .9, s = .85), order_levels)

# Only fill for metrics in band_data
fill_cols <- cols[names(cols) %in% levels(band_data$Metric)]

## ---------- PLOT ------------------------------------------------------
delta_pct    <- sprintf("%.0f%%", delta_prop * 100)
legend_title <- bquote("Metric ("*delta*" = "*.(delta_pct)*")")
plot_main    <- bquote("Generalized Toxicity metrics ("*delta*" = "*.(delta_pct)*")")

ggplot(mydata, aes(conc, response)) +
  geom_point() +
#  geom_rect(
#    data = band_data,
#    aes(xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax, fill = Metric),
#    alpha = 0.3, color = "white", inherit.aes = FALSE, show.legend = FALSE) +
  geom_line(data = grid, aes(x = conc, y = fit)) +
  geom_ribbon(data = grid,
              aes(x = conc, ymin = lwr, ymax = upr),
              inherit.aes = FALSE, alpha = .15) +
  geom_hline(yintercept = hline, linetype = "dashed",linewidth = 0.6) +
  geom_vline(data = vlines,
             aes(xintercept = x, colour = Metric, linetype = Metric),
             linewidth = 0.6, show.legend = FALSE) +
  geom_segment(data = legend_segs,
               aes(x = x, xend = xend, y = y, yend = yend,
                   colour = Metric, linetype = Metric),
               inherit.aes = FALSE, linewidth = 0.6) +
  scale_colour_manual(values = cols, labels = label_vec, name = legend_title) +
#  scale_fill_manual(values = fill_cols) +
  scale_linetype_manual(values = lts, labels = label_vec, name = legend_title) +
  labs(y = "Response", x = "Concentration", title = plot_main) +
  theme_bw() +
  theme(legend.position = "bottom",
        legend.direction = "horizontal",
        legend.title    = element_text(size = 10),
        legend.text     = element_text(size = 9))
```

```
## Warning: package 'readr' was built under R version 4.5.2
```

```
## Warning: package 'stringr' was built under R version 4.5.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.0     v stringr   1.6.0
## v lubridate 1.9.4     v tibble    3.3.0
## v purrr     1.1.0     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()     masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()        masks stats::lag()
```

```
## x dplyr::select()     masks MASS::select()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

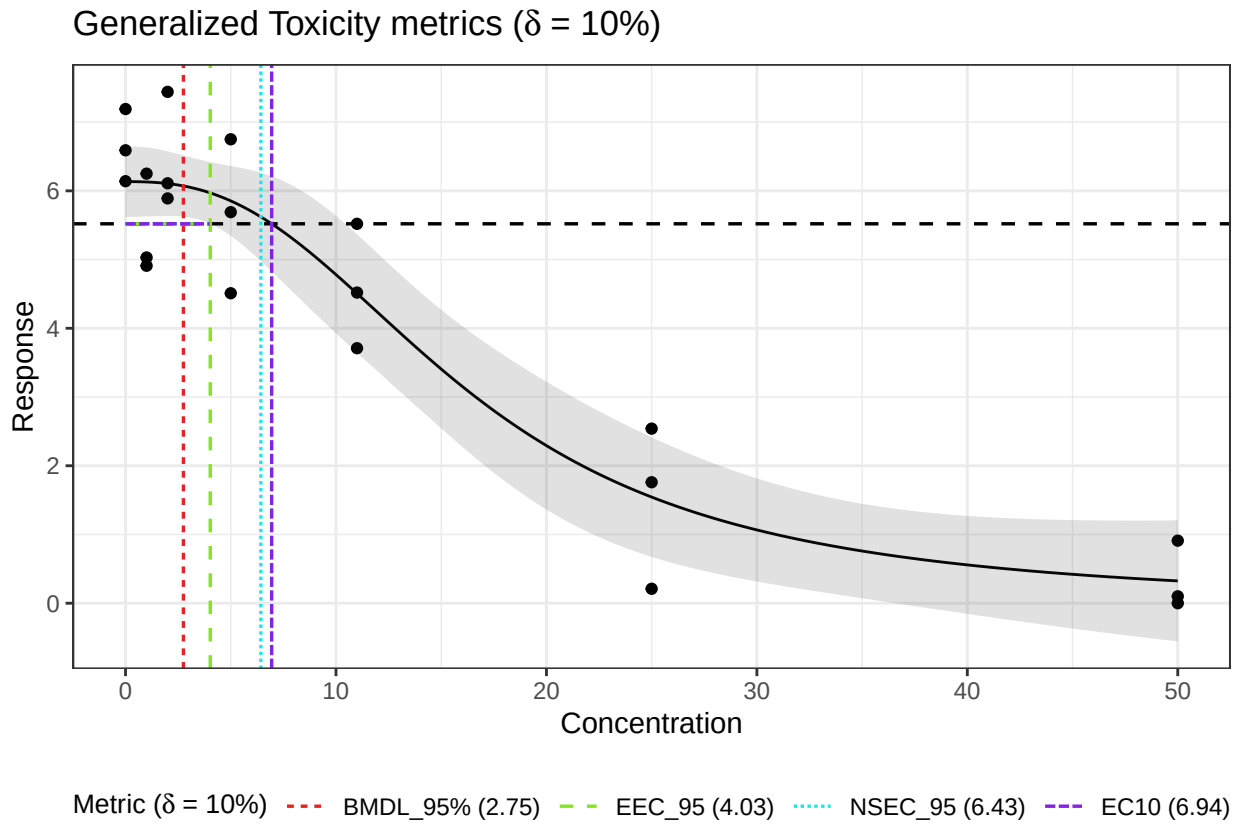
```
## Bootstrapping NSEC_95 …
```

```
## Bootstrapping BMDL_95 …
```

```
## Bootstrapping EC10 …
```

```
## Bootstrapping EEC_95 …
```

Metric	Estimate	CI_lower	CI_upper
NSEC_95	6.432298	2.202298	10.964365
EC10	6.944245	2.310202	12.287782
BMDL_95%	2.750583	NA	NA
EEC_95	4.025899	1.139059	8.417817



6.2 Comparison with USEPA BMDP Program

It is noted that the Hill model used in the USEPA BMD program (and others) is equivalent to the 4-parameter log-logistic model, `LL.4` in the `drc` package via a simple mapping of parameters as follows.

6.2.1 Hill Model (as in USEPA BMD)

$$f(x) = g + \frac{\nu x^n}{k^n + x^n} \quad \nu < 0 \text{ for monotonically decreasing response}$$

or, equivalently:

$$f(x) = g + \frac{\nu}{1 + \left({}^{k}/_{x}\right)^n} = g + \frac{\nu}{1 + \left({}^{x}/_{k}\right)^{-n}}$$

6.2.2 Four parameter Log-logistic Model (LL.4) as in `drc`

$$\mathrm{g}(x) = c + \frac{d - c}{1 + \left({}^{x}/_{e}\right)^b}$$

To match a **decreasing** Hill model, we use *positive* b and map: - $c = g + \nu$ (lower asymptote, high dose) - $d = g$ (upper asymptote, control) - $e = k$ (half-max dose) - $b = n$ (Hill power, ≥ 1) **NB:** Using the alternative mapping $b = -n$ and $c = g$, $d = g + \nu$ will flip the direction (i.e result in an increasing curve) and strange looking g estimates (often near 0), even though ν, k, n look correct.

6.2.3 Standard errors of Hill Model parameter estimates

The standard error for each of the Hill Model parameter estimates is readily obtained using the output of `LL.4` fit from `drc` since the former are simply linear combinations of the latter via the mapping above. This is conveniently handled using matrix notation.

6.2.3.1 `drc::LL.4()` :

$$g(x; b, c, d, e) = \mathrm{g}(x) = c + \frac{d - c}{1 + \left({}^{x}/_{e}\right)^b}, \text{ let}\Theta = (b, c, d, e)^T$$

6.2.3.2 Hill (decreasing response):

$$f(x;) = g + \tfrac{\nu x^n}{k^n + x^n}, \quad \Phi = (g, \nu, k, n)^T$$

For a monmotonically decreasing curve:

$$g = d; \nu = c - d; k = e; n = b$$

This is a linear reparameterisation:

$$\Phi = A\,\Theta \quad \text{where} \quad A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

(rows correspond to g, ν, k, n; columns to b, c, d, e). Let $\hat{\Theta} = \left(\hat{b}, \hat{c}, \hat{d}, \hat{e}\right)^T$ be the `LL.4` estimator. `drc` reports the estimated covariance matrix $\widehat{Var}\left(\hat{\Theta}\right) \equiv \hat{V}_\Theta$ obtained from the inverse observed information/Hessian, scaled by the residual variance estimate.

6.2.3.3 Variance of mapped Hill parameter estimates.

From standard statistical theory,

$$\widehat{Var}\left(\hat{\Phi}\right) = A\,V_\Theta\,A^T = \hat{V}_\Phi$$

Hence, the standard errors are:

$$SE\left(\hat{\phi}_i\right) = \sqrt{\left(\hat{V}_\Phi\right)_{ii}}, \quad i \in \{g, \nu, k, n\}$$

6.2.3.4 Wald confidence intervals for Hill parameter estimates

The Wald $100\,(1-\alpha)\,\%$ CIs are:

$$\hat{\phi}_i \;\pm\; t_{df,(1-\alpha/2)}\,SE\left(\hat{\phi}_i\right)$$

where $df = n - 4$.

6.3 Example

We again use the data from (Fisher & Fox, 2023b):

Table 3: Fish growth across a range of concentrations of an unknown pollutant

Concentration	Rep 1	Rep 2	Rep 3
0	6.59	6.14	7.19
1	4.91	5.03	6.25
2	5.89	7.44	6.11
5	4.51	6.75	5.69
11	4.52	5.52	3.71
25	2.54	1.76	0.21
50	0.10	0.00	0.91

Using `drc` the following results are obtained:

```
##
## Model fitted: Log-logistic (ED50 as parameter) (4 parms)
##
## Parameter estimates:
##
##               Estimate Std. Error t-value   p-value
```

```
## b:(Intercept)  2.562908   1.052760  2.4345 0.0262286 *
## c:(Intercept) -0.007844   0.898266 -0.0087 0.9931343
## d:(Intercept)  6.133842   0.295012 20.7918 1.585e-13 ***
## e:(Intercept) 16.375463   3.341412  4.9008 0.0001349 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error:
##
##  0.8902487 (17 degrees of freedom)
```

From USEPA BMD (*see Appendix B for detailed BMDP output*) we obtain:

Parameter	Estimate	Std Error
g	6.135	0.266
v	-6.165	0.924
k	16.445	3.057
n	2.544	0.949

Using methods described above and the output from `LL.4` fit we obtain the following Hill parameter estimates, stndard errors and approximate 95% CIs.

Parameter	Estimate	SE	LCL95	UCL95
g (background)	6.134	0.295	5.511	6.756
nu (max change)	-6.142	1.007	-8.266	-4.018
k (ED50)	16.375	3.341	9.326	23.425
n (power)	2.563	1.053	0.342	4.784

Appendix 1: Proof of Equivalence Between TOST and Confidence Interval Inclusion Criterion

In section 2.1 it was stated that the TOST procedure is operationally identical to the procedure of declaring equivalence only if the $1 - 2\alpha$ confidence interval is completely contained within the equivalence interval $[-\delta,\ +\delta]$. A formal proof of this assertion follows. Let $\hat{\theta}$ be an estimator of a parameter θ, with standard error SE, and suppose we wish to test whether θ lies within an equivalence margin $(-\delta, \delta)$ using the Two One-Sided Tests (TOST) procedure at significance level α. We aim to prove the following equivalence:

$$\text{TOST declares equivalence at level } \alpha \iff \text{CI}_{1-2\alpha} \subset (-\delta, \delta),$$

where $\text{CI}_{1-2\alpha}$ denotes a $(1 - 2\alpha)$ confidence interval for θ.

Definitions

The hypotheses for TOST are:

$$H_{01} : \theta \leq -\delta,$$
$$H_{02} : \theta \geq \delta.$$

We reject H_{01} if:

$$T_1 = \frac{\hat{\theta} + \delta}{SE} > z_{1-\alpha},$$

and reject H_{02} if:

$$T_2 = \frac{\hat{\theta} - \delta}{SE} < -z_{1-\alpha}.$$

TOST concludes equivalence if and only if both H_{01} and H_{02} are rejected. The $(1 - 2\alpha)$ confidence interval for θ is:

$$\text{CI}_{1-2\alpha} = \left[\hat{\theta} - z_{1-\alpha}SE,\ \hat{\theta} + z_{1-\alpha}SE\right].$$

Proof

6.3.0.1 (⇒) TOST ⇒ CI is contained within $(-\delta, \delta)$: Assume both H_{01} and H_{02} are rejected:

$$\frac{\hat{\theta} + \delta}{SE} > z_{1-\alpha} \quad \Rightarrow \quad \hat{\theta} > -\delta + z_{1-\alpha}SE,$$

$$\frac{\hat{\theta} - \delta}{SE} < -z_{1-\alpha} \quad \Rightarrow \quad \hat{\theta} < \delta - z_{1-\alpha}SE.$$

Then:

$$\hat{\theta} - z_{1-\alpha}SE > -\delta \quad \text{and} \quad \hat{\theta} + z_{1-\alpha}SE < \delta,$$

which implies:

$$\left[\hat{\theta} - z_{1-\alpha}SE,\ \hat{\theta} + z_{1-\alpha}SE\right] \subset (-\delta, \delta).$$

6.3.0.2 ($\Leftarrow$) CI $\subset (-\delta, \delta) \Rightarrow$ TOST: Assume:

$$\hat{\theta} - z_{1-\alpha} SE > -\delta \quad \text{and} \quad \hat{\theta} + z_{1-\alpha} SE < \delta.$$

Then:

$$\hat{\theta} + \delta > z_{1-\alpha} SE \quad \Rightarrow \quad \frac{\hat{\theta} + \delta}{SE} > z_{1-\alpha} \quad \Rightarrow \quad \text{Reject } H_{01},$$

$$\hat{\theta} - \delta < -z_{1-\alpha} SE \quad \Rightarrow \quad \frac{\hat{\theta} - \delta}{SE} < -z_{1-\alpha} \quad \Rightarrow \quad \text{Reject } H_{02}.$$

Thus, both null hypotheses are rejected and TOST declares equivalence.

Conclusion

$$\text{TOST rejects both } H_{01} \text{ and } H_{02} \iff \text{CI}_{1-2\alpha} \subset (-\delta, \delta).$$

Appendix 2: Power for the One-sided Dunnett's Test

Setup and Notation

We consider a one-way layout with one control group and k treatment groups.
- Groups: $i = 0, 1, \dots, k$, where $i = 0$ is control.
- Sample sizes: $n_0, n_1, \dots, n_k$.
- Group means: $\bar{Y}_i$.
- Model:

$$Y_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_i, \sigma^2)$$

We want to test familywise one-sided hypotheses

$$H_{0i} : \mu_i \leq \mu_0 \qquad \text{vs.} \qquad H_{1i} : \mu_i > \mu_0$$

for increasing alternatives (reverse the sign for decreasing responses).

Test Statistic

The ANOVA mean square error is

$$S_{pe}^2 = \frac{\sum_{i=0}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{N - (k+1)}, \qquad N = \sum_{i=0}^{k} n_i$$

For treatment i, the difference from control is

$$\hat{\theta}_i = \bar{Y}_i - \bar{Y}_0$$

Its estimated standard error is

$$\text{SE}(\hat{\theta}_i) = S_{pe} \sqrt{\tfrac{1}{n_i} + \tfrac{1}{n_0}}$$

Thus the one-sided Dunnett test statistic is

$$T_i = \frac{\hat{\theta}_i}{S_{pe} \sqrt{\tfrac{1}{n_i} + \tfrac{1}{n_0}}}$$

Null Distribution

Under the global null $H_0 : \mu_0 = \mu_1 = \cdots = \mu_k$, the vector

$$\mathbf{T} = (T_1, \dots, T_k)$$

follows a **multivariate t distribution** with
- degrees of freedom $\nu = N - (k+1)$,
- correlation (for $i \neq j$)

$$\rho_{ij} = \frac{\frac{1}{n_0}}{\sqrt{\left(\frac{1}{n_0} + \frac{1}{n_i}\right)\left(\frac{1}{n_0} + \frac{1}{n_j}\right)}}$$

Critical Value

Let c_α be the one-sided critical value such that

$$\Pr\left(\max_{i=1,\ldots,k} T_i \leq c_\alpha \,\middle|\, H_0\right) = 1 - \alpha$$

This value is obtained from the distribution of the maximum of correlated t-statistics (tabulated by Dunnett, implemented in software such as `multcomp` in R).

Power Function

Let the true mean difference be $\delta_i = \mu_i - \mu_0$. Then under H_1,

$$T_i \sim t_\nu(\lambda_i), \quad \lambda_i = \frac{\delta_i}{\sigma\sqrt{1/n_i + 1/n_0}}$$

The familywise **power** is

$$\Pr\left(\max_{i=1,\ldots,k} T_i > c_\alpha \,\middle|\, \delta_1, \ldots, \delta_k\right)$$

where the joint distribution is multivariate noncentral t with correlations as above.

Per-Comparison vs Familywise Power

- **Per-comparison power**:
  For a specific treatment i, the probability of rejecting H_{0i} when δ_i is true is

  $$1 - F_{t_\nu(\lambda_i)}(c_\alpha)$$

  where $F_{t_\nu(\lambda)}$ is the CDF of a noncentral t distribution.

- **Familywise power**:
  The probability of rejecting *at least one* false null is the multivariate probability

$$1 - \Pr(T_1 \leq c_\alpha, \ldots, T_k \leq c_\alpha)$$

Sample-Size Targeting (Any-Pair Power)

Suppose all treatments have the same sample size $n_i = n$ and the same effect size δ. Then

$$\lambda = \frac{\delta}{\sigma\sqrt{2/n}} = \sqrt{\tfrac{n}{2}}\,\frac{\delta}{\sigma}$$

To achieve familywise power $1 - \beta$, choose n such that

$$\Pr\left(\max_{i=1,\ldots,k} T_i > c_\alpha \,\middle|\, \lambda\right) \geq 1 - \beta$$

This requires evaluating the multivariate noncentral t distribution. Numerical solutions (root-finding over n) can be implemented in R using the `mvtnorm` or `multcomp` packages.

Bottom Line

- Both S^2_{pe} (ANOVA error mean square) and S^2_{res} (residual variance from a fitted dose–response model) are unbiased estimators of σ^2.

- With a correctly specified model,

$$\mathrm{Var}(S^2_{res}) < \mathrm{Var}(S^2_{pe})$$

- Therefore, model-based tests yield more precise estimates of variance and smaller standard errors for contrasts $\{\hat{\mu}(d_i) - \hat{\mu}(d_0)\}$, leading to higher power than the classical Dunnett procedure at the same sample size.

Appendix 3: Computation of BMDL using profile likelihood

Model and target level

- LL.4 mean function (log-dose on the x-axis):

$$f(x;\, b, c, d, e) \;=\; c \;+\; \frac{d-c}{1+\exp\big(b(\log x - \log e)\big)}, \qquad b \neq 0,\; e > 0.$$

Let the benchmark be specified by a target response $y^\star$ that lies strictly between the asymptotes $c < y^\star < d$. Two common specifications: **Proportion ("extra/relative") BMR.** Pick $q \in (0, 1)$ and set

$$y^\star \;=\; c + q(d-c).$$

Absolute-change BMR. Pick $k \in (0,\, d-c)$ and set

$$y^\star \;=\; c + k.$$

The **BMD** is the dose $x^\star$ solving $f(x^\star; b, c, d, e) = y^\star$.

- Explicit solution for $x^\star$ Solve $f(x) = y^\star$:

$$\frac{d-c}{1+\exp\big(b(\log x - \log e)\big)} \;=\; y^\star - c \;\Longrightarrow\; \exp\big(b(\log x - \log e)\big) \;=\; \frac{y^\star - c}{d - y^\star}.$$

Define

$$s(b, c, d;\, y^\star) \;:=\; \frac{y^\star - c}{d - y^\star} \;>\; 0.$$

Then

$$\log x^\star - \log e \;=\; \frac{1}{b}\log s \quad\Longrightarrow\quad x^\star \;=\; e\, s^{1/b}. \tag{1}$$

Two immediate corollaries: **Proportion BMR.** If $y^\star = c + q(d-c)$, then

$$s = \frac{q}{1-q} \quad\Rightarrow\quad x^\star \;=\; e\left(\frac{q}{1-q}\right)^{1/b}. \tag{1a}$$

So the BMD depends only on b and e. **Absolute-change BMR.** If $y^\star = c + k$, then

$$s = \frac{k}{(d-c)-k} \quad\Rightarrow\quad x^\star \;=\; e\left(\frac{k}{(d-c)-k}\right)^{1/b}, \tag{1b}$$

so the BMD depends on b, e and also c, d.

Below, denote the BMD by ξ (i.e., $\xi := x^\star$).

- Profile likelihood for a derived quantity $\xi = \phi(\theta)$

Let $\theta = (b, c, d, e)$ and $\phi(\theta) = x^\star$ given by (1). With data $\{(x_i, y_i)\}_{i=1}^n$ and a parametric error model (e.g., Normal with variance σ^2, or whatever family you fit in `drc`), the full log-likelihood is $\ell(\theta, \eta)$ where η collects

any nuisance parameters (e.g., σ^2). **Unconstrained MLE:**

$$(\hat{\theta}, \hat{\eta}) = \arg\max_{\theta,\eta} \ell(\theta, \eta), \qquad \hat{\ell} = \ell(\hat{\theta}, \hat{\eta}).$$

We want a CI for $\xi = \phi(\theta)$, a smooth function of θ. The **profile log-likelihood** for a proposed value ξ_0 is

$$\ell_p(\xi_0) = \sup_{\theta,\eta:\, \phi(\theta)=\xi_0} \ell(\theta, \eta).$$

The **LR statistic**

$$W(\xi_0) = 2(\hat{\ell} - \ell_p(\xi_0)) \xrightarrow{d} \chi^2_1 \quad \text{(Wilks)}.$$

Thus, the one-sided lower $100(1-\alpha)\%$ limit ξ_L solves

$$W(\xi_L) = \chi^2_{1,\,1-2\alpha}, \qquad \xi_L < \hat{\xi}, \quad \hat{\xi} = \phi(\hat{\theta}).$$

(Upper limits are analogous.) All that remains is how to compute $\ell_p(\xi_0)$ efficiently. Two equivalent routes:

Reparameterize to eliminate the constraint

Proportion BMR (most common): Let $r = \dfrac{q}{1-q}$ (a constant). From (1a),

$$\xi = e\, r^{1/b} \iff e = h(b, \xi) := \xi\, r^{-1/b} = \xi\, \exp\!\big(-(\log r)/b\big). \tag{2a}$$

Use (b, c, d, ξ) as the parameter vector and replace e everywhere by $h(b, \xi)$. The fitted mean becomes

$$f(x;\, b, c, d, \xi) = c + \frac{d-c}{1 + \exp\!\big(b\{\log x - \log h(b, \xi)\}\big)}.$$

For any fixed $\xi = \xi_0$, maximize $\ell(b, c, d, \xi_0, \eta)$ over b, c, d, η. That maximized value is $\ell_p(\xi_0)$. Useful partials (for gradient-based solvers):

$$\frac{\partial e}{\partial b} = e\, \frac{\log r}{b^2}, \qquad \frac{\partial e}{\partial \xi} = \frac{e}{\xi}.$$

Absolute-change BMR: Here $y^\star = c + k$ and $s = \dfrac{k}{(d-c) - k}$. From (1b),

$$e = h(b, c, d, \xi) := \xi\, s^{-1/b}. \tag{2b}$$

Reparameterize as (b, c, d, ξ) and substitute $e = h(b, c, d, \xi)$ in the mean. For fixed ξ_0, maximize over b, c, d, η. With $s = s(c, d)$,

$$\frac{\partial e}{\partial b} = e\, \frac{\log s}{b^2}, \quad \frac{\partial e}{\partial c} = -\,e\, \frac{1}{b}\frac{1}{s} \cdot \frac{1}{(d-c) - k}, \quad \frac{\partial e}{\partial d} = +\,e\, \frac{1}{b}\frac{1}{s} \cdot \frac{1}{(d-c) - k}, \quad \frac{\partial e}{\partial \xi} = \frac{e}{\xi}.$$

Constrained maximization via a Lagrangian

Keep (b, c, d, e) and, for a given ξ_0, solve

$$\max_{b,c,d,e,\eta} \ell(b, c, d, e, \eta) \quad \text{s.t.} \quad \phi(b, c, d, e) - \xi_0 = 0.$$

Form

$$L(b, c, d, e, \eta, \lambda) = \ell(b, c, d, e, \eta) + \lambda(\phi(b, c, d, e) - \xi_0),$$

and solve the KKT system. Numerically this is often less stable than reparameterisation; for LL.4 the closed-form substitution in reparameterisation is preferable.

- The likelihood-ratio CI Fit the unconstrained model to get $\hat{\ell}$ and $\hat{\theta}$, hence $\hat{\xi} = \phi(\hat{\theta})$. For a grid or a root-finder over ξ_0 (one side at a time): 1. Compute $\ell_p(\xi_0)$ by maximizing over the free parameters (Route A recommended). 2. Compute $W(\xi_0) = 2\{\hat{\ell} - \ell_p(\xi_0)\}$. The **lower one-sided** $100(1-\alpha)\%$ **BMDL** is the smallest $\xi_0 < \hat{\xi}$ with

$$W(\xi_0) = \chi^2_{1,\,1-2\alpha}.$$

(Upper works analogously on $\xi_0 > \hat{\xi}$.)

By Wilks' theorem, the χ^2_1 cutoff is valid; profiling in ξ is parameterization-invariant.

- Practical notes (LL.4 and `drc`) - Monotonicity / direction. The formulae hold regardless of the sign of b; the exponent $1/b$ handles increasing vs decreasing curves automatically. Ensure $0 < q < 1$ (or $0 < k < d - c$) so $y^\star$ is between c and d.

- **Unconstrained** Replacing e with $h(\cdot)$ removes the constraint; you maximize an unconstrained likelihood in (b, c, d) at each ξ_0.
- **Small-sample caution.** LR cutoffs are asymptotic. With very small n, LR CIs can be unreliable.

- Summary formulas General:

$$x^\star(\theta) = e\left(\frac{y^\star - c}{d - y^\star}\right)^{1/b}.$$

Proportion BMR $y^\star = c + q(d-c)$:

$$x^\star = e\left(\frac{q}{1-q}\right)^{1/b}, \qquad e = \xi\left(\frac{q}{1-q}\right)^{-1/b}.$$

Absolute-change BMR $y^\star = c + k$:

$$x^\star = e\left(\frac{k}{(d-c)-k}\right)^{1/b}, \qquad e = \xi\left(\frac{k}{(d-c)-k}\right)^{-1/b}.$$

$$W(\xi_0) = 2\left[\hat{\ell} - \ell_p(\xi_0)\right] \overset{d}{\sim} \chi^2_1.$$

Example

We simulate data and fit the LL.4 model by least squares (equivalent to MLE under Normal errors with σ^2 profiled out), computes $\hat{\xi}$ at a specified BMR q, and then profiles $\ell_p(\xi_0)$ over (b, c, d) with e eliminated via $e = h(b, \xi_0)$. NB: The simulated data can be replaced with your own (x, y).

```r
# LL.4 mean
f_ll4 <- function(x, b, c, d, e) {
  c + (d - c) / (1 + exp(b * (log(x) - log(e))))
}

# RSS-based negative log-likelihood (up to additive constant) with sigma profiled out:
# nll(theta) = n/2 * log(RSS/n), so maximizing ll <=> minimizing RSS.
nll_from_rss <- function(residuals) {
  n <- length(residuals)
  rss <- sum(residuals^2)
  0.5 * n * log(rss / n)
}

# Unconstrained fit: parameters are (b, c, d, loge)
fit_unconstrained <- function(x, y, par0 = c(b=1, c=min(y), d=max(y), loge=log(median(x)))) {
  obj <- function(p) {
    b <- p[1]; c <- p[2]; d <- p[3]; e <- exp(p[4])
    mu <- f_ll4(x, b, c, d, e)
    nll_from_rss(y - mu)
  }
  opt <- optim(par0, obj, method = "BFGS", control = list(reltol = 1e-10))
  c(opt$par, value = opt$value, convergence = opt$convergence)
}

# Eliminate e for PROPORTION BMR (q in (0,1))
# e = h(b, xi) = xi * r^(-1/b), where r = q/(1-q)
h_prop <- function(b, xi, q) {
  r <- q / (1 - q)
  xi * r^(-1 / b)
}

# Profile nll at fixed xi0, optimizing over (b, c, d)
profile_nll_prop <- function(x, y, xi0, q, par0 = c(b=1, c=min(y), d=max(y))) {
  obj <- function(p) {
    b <- p[1]; c <- p[2]; d <- p[3]
    e <- h_prop(b, xi0, q)
    mu <- f_ll4(x, b, c, d, e)
    nll_from_rss(y - mu)
  }
  # light boxing for stability on (d - c) > 0:
  obj_boxed <- function(p_raw) {
    # reparam: c = a, d = a + exp(t) to enforce d > c
    b  <- p_raw[1]
    a  <- p_raw[2]
    t  <- p_raw[3]
    c  <- a
    d  <- a + exp(t)
    p  <- c(b, c, d)
```

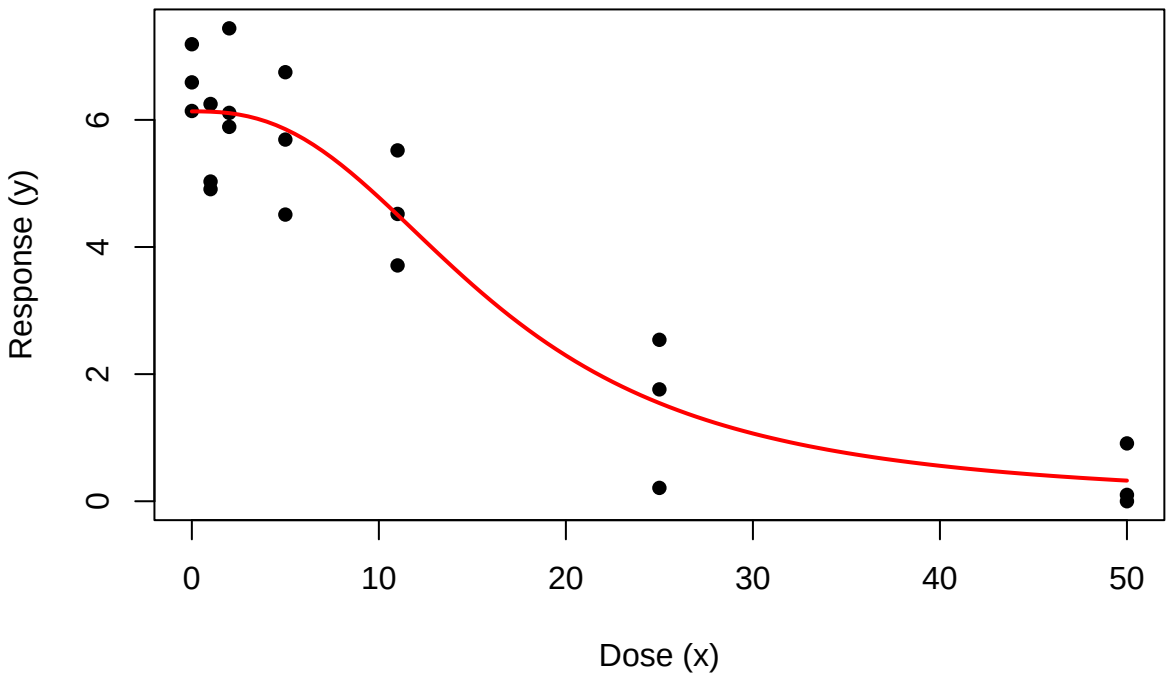
```
    obj(p)
  }
  # initial raw params
  b0 <- par0[1]; a0 <- par0[2]; t0 <- log(max(1e-6, par0[3]-par0[2]))
  p0 <- c(b0, a0, t0)
  opt <- optim(p0, obj_boxed, method = "BFGS", control = list(reltol = 1e-10))
  list(opt = opt, nll = opt$value)
}
```

```
# LL.4 mean
f_ll4 <- function(x, b, c, d, e) {
  c + (d - c) / (1 + exp(b * (log(x) - log(e))))
}

# RSS-based negative log-likelihood (up to additive constant) with sigma profiled out:
# nll(theta) = n/2 * log(RSS/n), so maximizing ll <=> minimizing RSS.
nll_from_rss <- function(residuals) {
  n <- length(residuals)
  rss <- sum(residuals^2)
  0.5 * n * log(rss / n)
}

# Unconstrained fit: parameters are (b, c, d, loge)
fit_unconstrained <- function(x, y, par0 = c(b=1, c=min(y), d=max(y), loge=log(median(x)))) {
  obj <- function(p) {
    b <- p[1]; c <- p[2]; d <- p[3]; e <- exp(p[4])
    mu <- f_ll4(x, b, c, d, e)
    nll_from_rss(y - mu)
  }
  opt <- optim(par0, obj, method = "BFGS", control = list(reltol = 1e-10))
  c(opt$par, value = opt$value, convergence = opt$convergence)
}

# Eliminate e for PROPORTION BMR (q in (0,1))
# e = h(b, xi) = xi * r^(-1/b), where r = q/(1-q)
h_prop <- function(b, xi, q) {
  r <- q / (1 - q)
  xi * r^(-1 / b)
}

# Profile nll at fixed xi0, optimizing over (b, c, d)
profile_nll_prop <- function(x, y, xi0, q, par0 = c(b=1, c=min(y), d=max(y))) {
  obj <- function(p) {
    b <- p[1]; c <- p[2]; d <- p[3]
    e <- h_prop(b, xi0, q)
    mu <- f_ll4(x, b, c, d, e)
    nll_from_rss(y - mu)
  }
  # light boxing for stability on (d - c) > 0:
  obj_boxed <- function(p_raw) {
    # reparam: c = a, d = a + exp(t) to enforce d > c
    b  <- p_raw[1]
    a  <- p_raw[2]
    t  <- p_raw[3]
```

```
    c  <- a
    d  <- a + exp(t)
    p  <- c(b, c, d)
    obj(p)
  }
  # initial raw params
  b0 <- par0[1]; a0 <- par0[2]; t0 <- log(max(1e-6, par0[3]-par0[2]))
  p0 <- c(b0, a0, t0)
  opt <- optim(p0, obj_boxed, method = "BFGS", control = list(reltol = 1e-10))
  list(opt = opt, nll = opt$value)
}
```

```
# Simulate a decreasing LL.4 curve (b > 0)
 n  <- 40
 x  <- sort(exp(seq(log(0.5), log(200), length.out = n)))
bT <- 2.0; cT <- 10; dT <- 100; eT <- 30
mu <- f_ll4(x, bT, cT, dT, eT)
y  <- mu + rnorm(n, sd = 5)

plot(x, y, pch = 16, xlab = "Dose (x)", ylab = "Response (y)")
curve(f_ll4(x, bT, cT, dT, eT), ,col="red",add = TRUE, lwd = 2)
```

```
library(drc)
load("fisher.Rdata")
x <- fisher.dat$dose; y <- fisher.dat$y
plot(x, y, pch = 16, xlab = "Dose (x)", ylab = "Response (y)")
p<-as.numeric(coef(drm(y ~x,fct=LL.4())))
curve(f_ll4(x,p[1],p[2],p[3],p[4] ), ,col="red",add = TRUE, lwd = 2)
```




```
# Unconstrained fit
fit0 <- fit_unconstrained(x, y)
b.h <- fit0[1]; c.h <- fit0[2]; d.h <- fit0[3]; e.h <- exp(fit0[4])
nll0 <- fit0["value"]

cat(sprintf("Unconstrained fit (b, c, d, e-hat) = (%.3f, %.3f, %.3f, %.3f)\n",
            b.h, c.h, d.h, e.h))
```

```
## Unconstrained fit (b, c, d, e-hat) = (2.544, -0.030, 6.135, 16.445)
```

```
# Choose a proportion BMR
q <- 0.10
# BMD (xi-hat) from fitted parameters (proportion case)
xi_hat <- e.h * (q/(1-q))^(1 / b.h)
xi_hat
```

```
##     loge
## 6.933244
```

```
alpha <- 0.05     # one-sided lower 95%
crit  <- qchisq(1 - 2*alpha, df = 1)

# Profile over xi0 on a grid to find lower limit
# Work below xi_hat
grid <- sort(unique(c(
```

```
  seq(max(1e-4, 0.1*xi_hat), xi_hat, length.out = 60),
  xi_hat
)))

prof_vals <- numeric(length(grid))
for (i in seq_along(grid)) {
  xi0 <- grid[i]
  # Warm start near unconstrained
  par0 <- c(b = b.h, c = c.h, d = d.h)
  pr   <- profile_nll_prop(x, y, xi0, q, par0 = par0)
  prof_vals[i] <- pr$nll
}

W <- 2 * (nll0 - prof_vals)

# Find smallest xi0 with W >= crit (on lower side)
idx <- which(W >= crit & grid <= xi_hat)
if (length(idx)) {
  bmdl <- min(grid[idx])
} else {
  bmdl <- NA_real_
}

list(
  xi_hat = xi_hat,
  BMDL_lower_1s = bmdl,
  LR_crit = crit
)
```

```
## $xi_hat
##     loge
## 6.933244
##
## $BMDL_lower_1s
## [1] NA
##
## $LR_crit
## [1] 2.705543
```

```
plot(grid, W, type = "l", xlab = expression(xi[0]), ylab = expression(W(xi[0])),
     main = "LR profile for BMD (proportion BMR)")
abline(h = crit, lty = 2)
abline(v = xi_hat, lty = 3)
if (!is.na(bmdl)) abline(v = bmdl, col = 2, lty = 2)
legend("topleft",
       legend = c("W", "chi^2_1 cutoff", "xi-hat", "BMDL (lower 1-sided)"),
       lty = c(1,2,3,2), col = c(1,1,1,2), bty = "n")
```

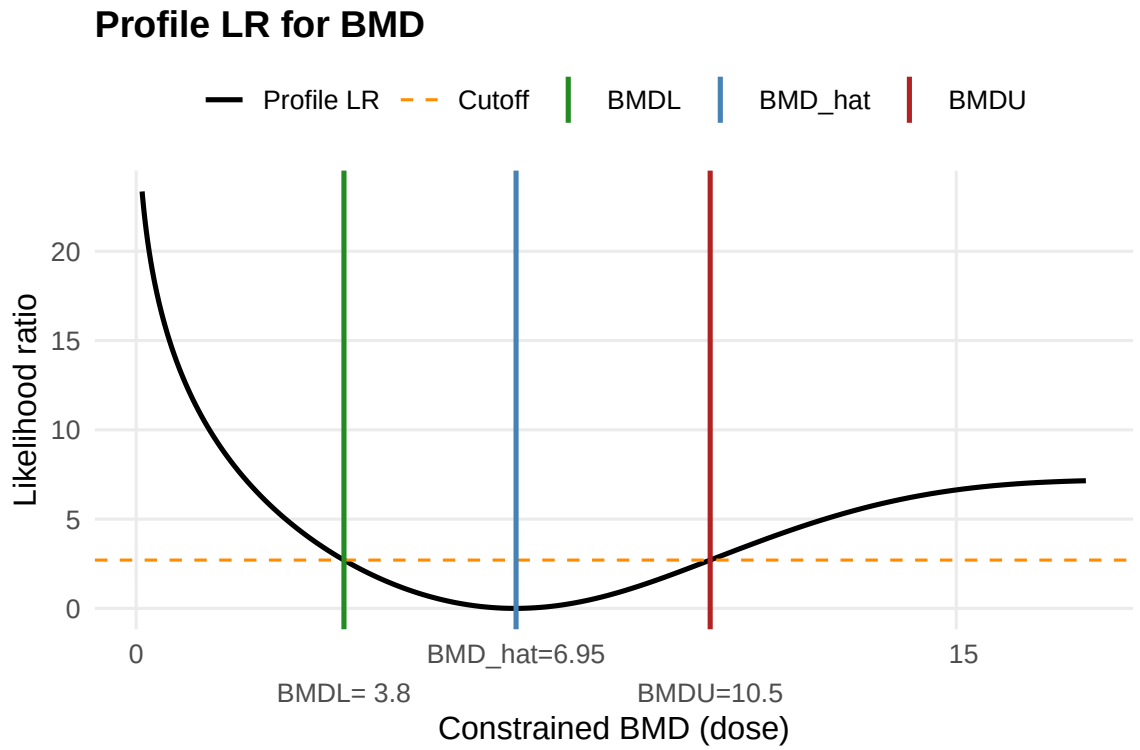
```
source("my_bmd_profile_ll.R")
 out <- bmd_profile_CI(fisher.dat, bmr = 0.1, type = "extra", alpha = 0.05, verbose = TRUE)
```

```
## Check: LR(BMD_hat)   -0.0007 (should be ~0)
```

```r
list(
  BMD_hat = out$BMD_hat,
  BMDL_lower_1s =  out$BMDL,
  BMDL_upper_1s =  out$BMDU,
  AIC  = AIC(out$fit),
  LR_crit = crit
)
```

```
## $BMD_hat
## [1] 6.948093
##
## $BMDL_lower_1s
## [1] 3.800303
##
## $BMDL_upper_1s
## [1] 10.49865
##
## $AIC
## [1] 60.27524
##
## $LR_crit
## [1] 2.705543
```

```r
source("plot_profile_gg.R")
plot_lr_profile_gg(out, dat = fisher.dat)
```



R code for finding BMDL by profile likelihood

```r
# install.packages("drc")  # if needed
library(drc)
library(ggplot2)

# ---- Model definition (LL.4 as in drc) ----
LL4_fun <- function(x, b, c, d, e) {
  # LL.4 parameterization used by drc:
  # f(x) = c + (d - c) / (1 + (x/e)^b)
  c + (d - c) / (1 + (x / e)^b)
}

# ---- Unconstrained fit with drc ----
fit_LL4 <- function(dat) {
  # dat must have columns: dose, y
  stopifnot(all(c("dose", "y") %in% names(dat)))
  drm(y ~ dose, data = dat, fct = LL.4(names = c("b","c","d","e")))
}

# ---- Unconstrained log-likelihood (Gaussian with sigma^2 profiled out) ----
loglik_gaussian_profiled <- function(residuals) {
  n <- length(residuals)
  RSS <- sum(residuals^2)
  # MLE sigma^2 = RSS/n; profiled loglik (up to an additive constant) is:
  #   l = -n/2 * [1 + log(2*pi*RSS/n)]
  -0.5 * n * (1 + log(2 * pi * RSS / n))
}

# ---- BMD from parameter vector (extra or additional risk) ----

bmd_from_pars <- function(pars, bmr, type = c("extra","additional"), add_amount = NULL) {
  type <- match.arg(type)
  pars <- setNames(as.numeric(pars), sub(":.*","", names(pars)))  # normalize names to b,c,d,e
  b <- pars["b"]; c <- pars["c"]; d <- pars["d"]; e <- pars["e"]

  # effect fraction r_eff in (0,1)
  if (type == "extra") {
    r_eff <- bmr
  } else {
    if (is.null(add_amount)) stop("For type='additional', provide add_amount (same units as response).")
    r_eff <- add_amount / (d - c)
  }
  if (!(is.finite(r_eff) && r_eff > 0 && r_eff < 1)) {
    stop("BMR mapping produced r not in (0,1). Check bmr/add_amount and (d-c).")
  }

  # Direction-aware mapping to g-target:
  # decreasing (b>0): effect = 1 - g  => g_target = 1 - r_eff
  # increasing (b<0): effect = g      => g_target = r_eff
  g_target <- if (b > 0) 1 - r_eff else r_eff

  # Invert g(x) = 1 / (1 + (x/e)^b) for x:
  # x = e * ((1 - g)/g)^(1/b)
```

```
  e * ((1 - g_target) / g_target)^(1 / b)
}

# ---- Constrained re-fit for a given candidate BMD (d0) via parameter substitution ----
# Key trick: enforce the constraint by solving for 'e' in terms of (b, c, d, d0, r),
# so the optimizer only searches over (b, c, d). This avoids a constrained optimizer.
# ---- Constrained re-fit for a given candidate BMD (d0) via parameter substitution ----
# Enforce the benchmark by solving for 'e' so that g(d0) = g_target
constrained_fit_loglik <- function(dat, d0, bmr, type = c("extra","additional"), add_amount = NULL,
                                   start_pars, penalty_big = 1e9) {
  type <- match.arg(type)
  x <- dat$dose; y <- dat$y

  # starting values for (b,c,d) from the unconstrained fit:
  par0 <- c(b = start_pars["b"], c = start_pars["c"], d = start_pars["d"])
  if (any(!is.finite(par0))) stop("Non-finite starting values for (b,c,d); check upstream fit.")

  obj <- function(par) {
    b <- par[1]; c <- par[2]; d <- par[3]
    if (!is.finite(b) || !is.finite(c) || !is.finite(d)) return(penalty_big)
    if (d <= c) return(penalty_big)

    # effect fraction r_eff
    r_eff <- if (type == "extra") bmr else add_amount / (d - c)
    if (!(is.finite(r_eff) && r_eff > 0 && r_eff < 1)) return(penalty_big)

    # direction-aware target for g at the benchmark
    g_target <- if (b > 0) 1 - r_eff else r_eff
    if (!(is.finite(g_target) && g_target > 0 && g_target < 1)) return(penalty_big)
    if (abs(b) < 1e-6) return(penalty_big)

    # Solve 1/(1 + (d0/e)^b) = g_target     e = d0 * (g_target/(1 - g_target))^(1/b)
    e <- d0 * (g_target / (1 - g_target))^(1 / b)
    if (!is.finite(e) || e <= 0) return(penalty_big)

    mu <- LL4_fun(x, b, c, d, e)
    res <- y - mu
    -loglik_gaussian_profiled(res)
  }

  opt <- optim(par = par0, fn = obj, method = "L-BFGS-B",
               control = list(maxit = 1000))
  if (opt$convergence != 0) warning("Constrained optimization may not have fully converged (code=",
     opt$convergence, ").")

  # Reconstruct best (b,c,d,e) with the same mapping
  b <- opt$par[1]; c <- opt$par[2]; d <- opt$par[3]
  r_eff <- if (type == "extra") bmr else add_amount / (d - c)
  g_target <- if (b > 0) 1 - r_eff else r_eff
  e <- d0 * (g_target / (1 - g_target))^(1 / b)
  mu <- LL4_fun(x, b, c, d, e)
  res <- y - mu
  ll <- loglik_gaussian_profiled(res)
```

```
  list(par = c(b = b, c = c, d = d, e = e), logLik = ll, value = opt$value)
}


# ---- Profile-likelihood scan & BMDL solver ----
bmd_profile_CI <- function(dat, bmr = 0.1, type = c("extra","additional"),
                           add_amount = NULL, alpha = 0.05,
                           grid_n = 80, expand_low = 1e-3, expand_high = 1e3,
                           verbose = TRUE) {
  type <- match.arg(type)

  # 1) Unconstrained fit
  fit <- fit_LL4(dat)
  phat <- coef(fit)                          # names: b,c,d,e
  nm   <- sub(":.*", "", names(phat))               # strip suffixes like ":(Intercept)"
  phat <- setNames(as.numeric(phat), nm)
  yhat <- fitted(fit)
  ll_uncon <- loglik_gaussian_profiled(dat$y - yhat)

  # 2) BMD at MLE (closed form)
  bmd_hat <- bmd_from_pars(phat, bmr = bmr, type = type, add_amount = add_amount)

  # 3) LR threshold (EPA one-sided; reduce log-lik by chi^2_ {1,1-2})
  #    => LR(d0) = 2[ell_uncon - ell_constr(d0)] must equal qchisq(1-2*alpha, 1)
  q <- qchisq(p = 1 - 2 * alpha, df = 1)

  # helper: LR at candidate d0
  LR_at <- function(d0) {
    cf <- constrained_fit_loglik(dat, d0 = d0, bmr = bmr, type = type,
                                 add_amount = add_amount, start_pars = phat)
    LR <- 2 * (ll_uncon - cf$logLik)
    c(LR = LR, logLik = cf$logLik)
  }

  # ---- Upper crossing (BMDU): bracket to the RIGHT of bmd_hat and solve LR(d) = q ----
  BMDU <- NA_real_

  hi <- bmd_hat                       # use the existing lower-case variable
  up <- hi
  LR_up <- LR_at(up)["LR"]            # take only the LR scalar
  tries <- 0L
  expand <- 2.5
  max_tries <- 30L

  # Expand right until LR(up) >= q (or give up)
  while (is.finite(LR_up) && LR_up < q && tries < max_tries) {
    up    <- up * expand
    LR_up <- LR_at(up)["LR"]
    tries <- tries + 1L
  }

  if (is.finite(LR_up) && LR_up >= q) {
    # Root-find on log-scale: solve LR(exp(u)) = q on [log(hi), log(up)]
    f_u2  <- function(u) LR_at(exp(u))["LR"] - q
```

```
    root2 <- uniroot(f_u2, lower = log(hi), upper = log(up), tol = 1e-8, maxiter = 200L)
    BMDU  <- exp(root2$root)
  } else if (isTRUE(verbose)) {
    warning("Could not bracket the upper LR crossing; returning NA for BMDU.")
  }

  # 4) Find the lower one-sided bound (BMDL): search left of bmd_hat
  #    Start from a wide bracket on the *dose* axis (log scale helps in practice).
  lo <- max(min(dat$dose[dat$dose>0], na.rm=TRUE) * expand_low, .Machine$double.eps)
  hi <- bmd_hat

  # Ensure LR(lo) > q and LR(hi) = 0 (approximately)
  LR_hi <- LR_at(hi)["LR"]
  if (verbose) message(sprintf("Check: LR(BMD_hat)  %.4f (should be ~0)", LR_hi))
  # Expand search to the left until LR(lo) > q or we hit numerical floor
  LR_lo <- LR_at(lo)["LR"]
  expand <- 2.5
  tries <- 0
  while (is.finite(LR_lo) && LR_lo <= q && lo > .Machine$double.eps && tries < 30) {
    lo <- lo / expand
    LR_lo <- LR_at(lo)["LR"]; tries <- tries + 1
  }

  bmdl <- NA_real_
  if (is.finite(LR_lo) && LR_lo > q) {
    # Root find LR(d0) - q = 0 on [lo, hi]
    froot <- function(d0) LR_at(d0)["LR"] - q
    # guard: uniroot needs finite values
    f_lo <- froot(lo); f_hi <- froot(hi)
    if (is.finite(f_lo) && is.finite(f_hi) && sign(f_lo) != sign(f_hi)) {
      bmdl <- uniroot(froot, lower = lo, upper = hi, tol = 1e-6)$root
    } else {
      warning("Could not bracket the LR root cleanly; returning NA for BMDL.")
    }
  } else {
    warning("LR at very small dose did not exceed the cutoff; BMDL may be below explored range.")
  }

  list(
    alpha  = alpha,
    LR_at  = LR_at,
    fit = fit,
    pars_hat = phat,
    logLik_uncon = ll_uncon,
    BMR = bmr,
    type = type,
    BMD_hat = as.numeric(bmd_hat),
    BMDL = as.numeric(bmdl),
    BMDU = BMDU,
    cutoff = q,
    note = sprintf("One-sided BMDL uses LR = qchisq(1-2*alpha, df=1). For alpha=%.2f, cutoff ~ %.4f.", alpha,
     q)
  )
}
```

Appendix 4: USEPA BMDP output for Worked Example in § 4

U.S. Environmental Protection Agency. (2025). BMDS Online (25.1; pybmds 25.1; bmdscore 25.1) [Software]. Available from https://bmdsonline.epa.gov. Accessed August 07, 2025.

Fisher Table_1

Report Generated: 2025-Sep-22 04:09 UTC
Analysis URL: View / Update
BMDS Online Version: 25.1 (pybmds 25.1; bmdscore 25.1)

Session for Dataset #1

Dataset

Name: Dataset #1

Dose	Response
0	6.59
1	4.91
2	5.89
5	4.51
11	4.52
25	2.54
50	0.1
0	6.14
1	5.03
2	7.44
5	6.75
11	5.52
25	1.76
50	0
0	7.19
1	6.25
2	6.11
5	5.69
11	3.71
25	0.21
50	0.91

Test 1 Dose Response: <0.0001
Test 2 Homogeneity of Variance: 0.7088
Test 3 Variance Model Selection: 0.7088

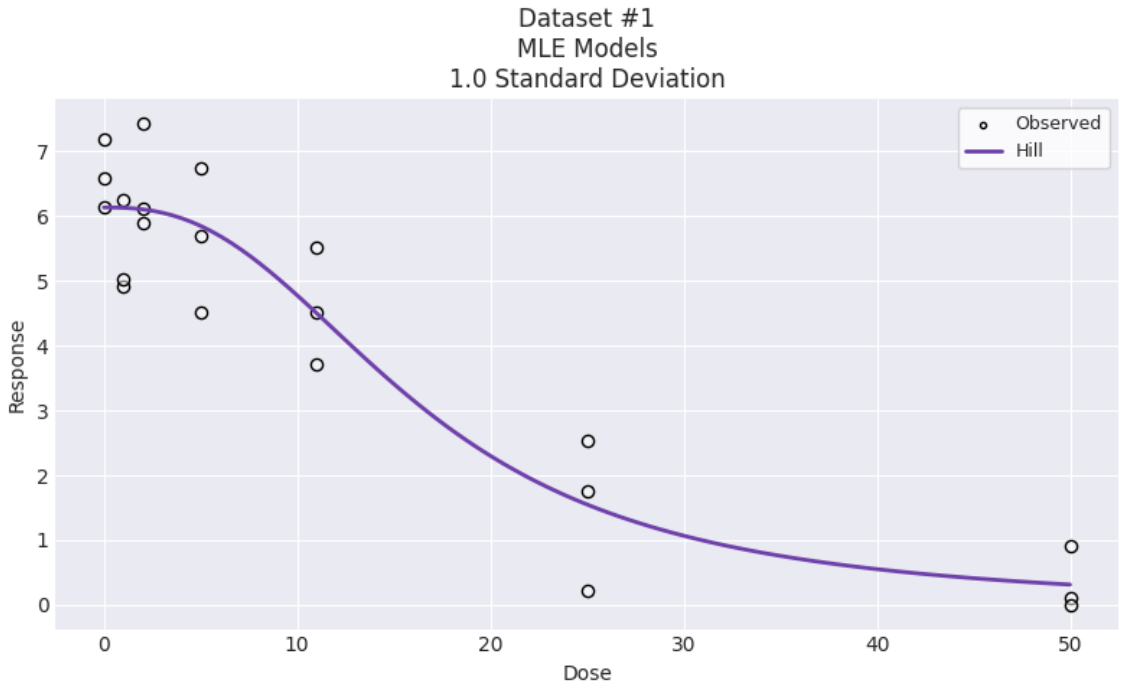
Settings

Setting	Value
BMR	1.0 Standard Deviation
Distribution	Normal + Constant variance
Adverse Direction	Down (↓)
Maximum Polynomial Degree	3
Confidence Level (one sided)	0.95

Maximum Likelihood Approach

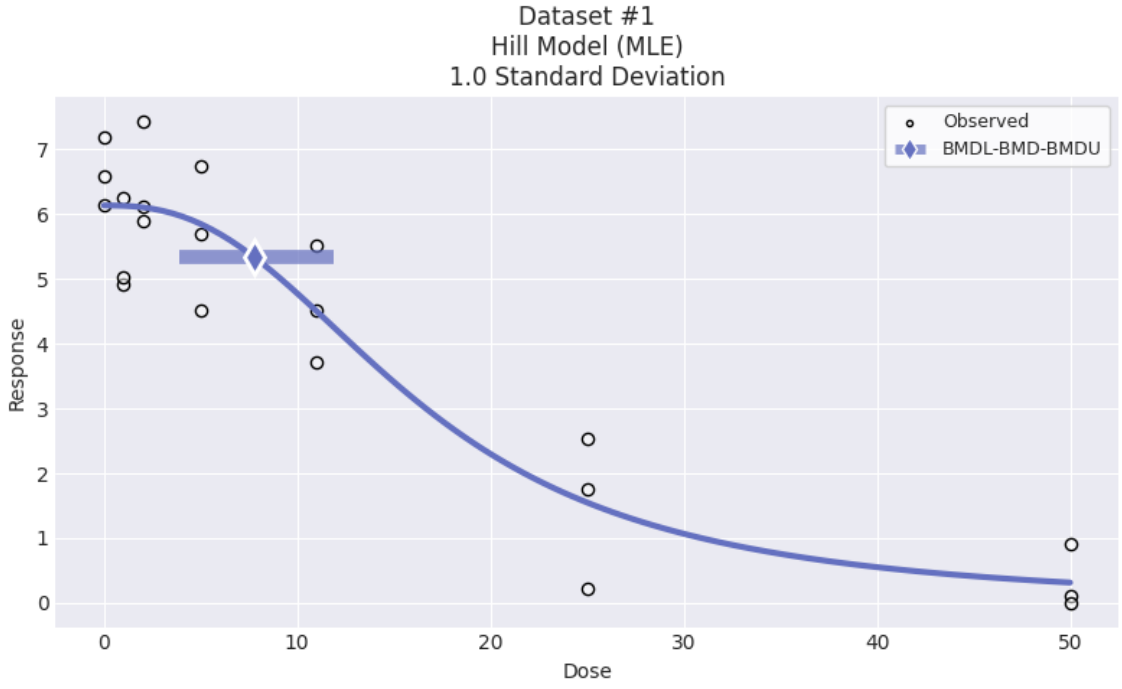
Model	BMDL	BMD	BMDU	*P*-Value	AIC	Scaled Residual at Control	Scaled Residual near BMD	Recommendation and Notes
Hill[a]	3.868	7.787	11.891	0.159	60.274	1.092	-0.434	**Recommended - Lowest AIC** Control stdev. fit > 1.5

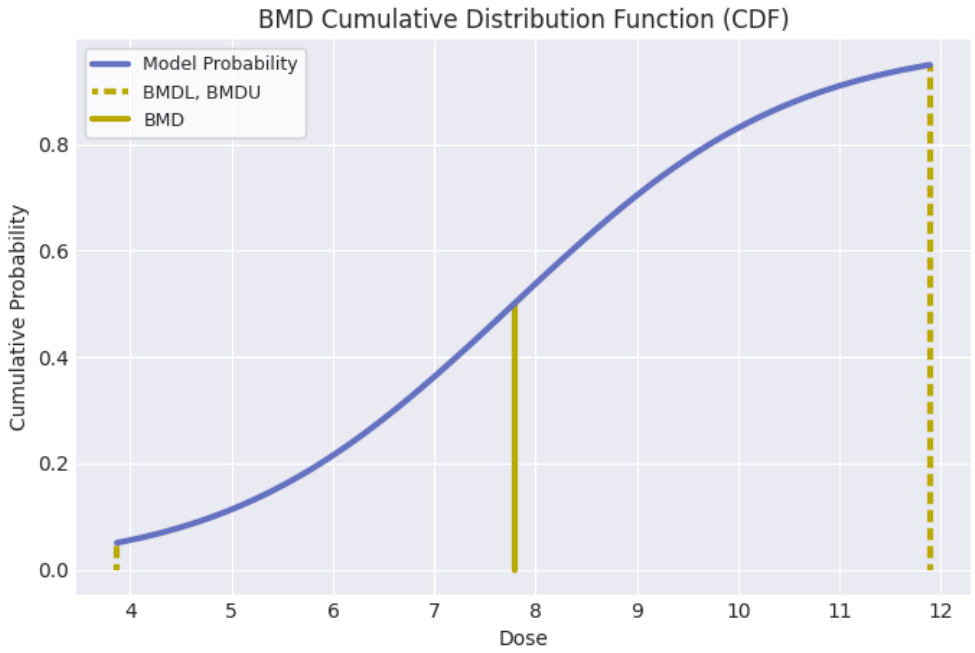
[a] BMDS recommended best fitting model



Individual Model Results

Hill Model





Hill Model

Version: pybmds 25.1 (bmdscore 25.1)

Input Summary:

BMR	1.0 Standard Deviation
Distribution	Normal + Constant variance
Modeling Direction	Down (↓)
Confidence Level (one sided)	0.95
Modeling Approach	MLE

Parameter Settings:

Parameter	Initial	Min	Max
g	0	-100	100
v	0	-100	100
k	0	0	5
n	1	1	18
alpha	0	-18	18

Modeling Summary:

BMD	7.7873
BMDL	3.86756
BMDU	11.8912
AIC	60.2745
Log-Likelihood	-25.1372
P-Value	0.159089
Model d.f.	3

Model Parameters:

Variable	Estimate	On Bound	Std Error
g	6.13508	no	0.266171
v	-6.1654	no	0.923721
k	16.4454	no	3.05689
n	2.54393	no	0.948536
alpha	0.641559	no	0.126998

Goodness of Fit:

Dose	N	Sample Mean	Model Fitted Mean	Scaled Residual
0	3	6.64	6.13508	1.09186
1	3	5.39667	6.13011	-1.58602
2	3	6.48	6.10622	0.808268
5	3	5.65	5.8506	-0.433783
11	3	4.58333	4.50472	0.169992
25	3	1.50333	1.54964	-0.10014
50	3	0.336667	0.31363	0.0498146

Dose	N	Sample SD	Model Fitted SD
0	3	0.526783	0.800974
1	3	0.74144	0.800974
2	3	0.83863	0.800974
5	3	1.12054	0.800974
11	3	0.906661	0.800974
25	3	1.18602	0.800974
50	3	0.499032	0.800974

Likelihoods:

Model	Log-Likelihood	# Params	AIC
A1	-22.5473	8	61.0946
A2	-20.666	14	69.332
A3	-22.5473	8	61.0946
fitted	-25.1372	5	60.2745
reduced	-48.1872	2	100.374

Tests of Mean and Variance Fits:

Name	-2 * Log(Likelihood Ratio)	Test d.f.	P-Value
Test 1	55.0424	12	1.77855e-07
Test 2	3.76261	6	0.708765
Test 3	3.76261	6	0.708765
Test 4	5.17989	3	0.159089

Test 1: Test the null hypothesis that responses and variances don't differ among dose levels (A2 vs R). If this test fails to reject the null hypothesis (p-value > 0.05), there may not be a dose-response.

Test 2: Test the null hypothesis that variances are homogenous (A1 vs A2). If this test fails to reject the null hypothesis (p-value > 0.05), the simpler constant variance model may be appropriate.

Test 3: Test the null hypothesis that the variances are adequately modeled (A3 vs A2). If this test fails to reject the null hypothesis (p-value > 0.05), it may be inferred that the variances have been modeled appropriately.

Test 4: Test the null hypothesis that the model for the mean fits the data (Fitted vs A3). If this test fails to reject the null hypothesis (p-value > 0.1), the user has support for use of the selected model.

Appendix 5: Comparing error estimates from categorical and continuous dose-response modelling

Setup

Let doses be x_j for $j = 1, \dots, k$ with $n_j \geq 2$ replicates and total $N = \sum_{j=1}^{k} n_j$. Observations satisfy

$$y_{ij} = \mu(x_j) + \varepsilon_{ij}, \qquad \mathbb{E}[\varepsilon_{ij}] = 0,\ \mathrm{Var}(\varepsilon_{ij}) = \sigma^2,\ i = 1, \dots, n_j.$$

Define group means $\bar{y}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$. Fit a parametric dose-response mean $\eta(x; \theta)$ with p estimable parameters and denote $\hat{\eta}_j = \eta(x_j; \hat{\theta})$.

Sums of squares and mean squares

Pure (within-dose) error:

$$SS_{\mathrm{pe}} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\cdot j})^2, \quad df_{\mathrm{pe}} = N - k, \quad MS_{\mathrm{pe}} = \frac{SS_{\mathrm{pe}}}{N - k}.$$

Model residuals:

$$SS_{\mathrm{res}} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \hat{\eta}_j)^2, \quad df_{\mathrm{res}} = N - p, \quad MS_{\mathrm{res}} = \frac{SS_{\mathrm{res}}}{N - p}.$$

Lack-of-fit decomposition (using $y_{ij} - \hat{\eta}_j = (y_{ij} - \bar{y}_{\cdot j}) + (\bar{y}_{\cdot j} - \hat{\eta}_j)$ and $\sum_i (y_{ij} - \bar{y}_{\cdot j}) = 0$):

$$\boxed{SS_{\mathrm{res}} = SS_{\mathrm{pe}} + SS_{\mathrm{lof}}}, \qquad \boxed{SS_{\mathrm{lof}} = \sum_{j=1}^{k} n_j (\bar{y}_{\cdot j} - \hat{\eta}_j)^2 \geq 0}.$$

Therefore

$$\boxed{MS_{\mathrm{res}} = \frac{N - k}{N - p} MS_{\mathrm{pe}} + \frac{SS_{\mathrm{lof}}}{N - p}}.$$

Consequences

- Under homoscedasticity and a correct mean model, $\mathbb{E}[MS_{\mathrm{pe}}] = \mathbb{E}[MS_{\mathrm{res}}] = \sigma^2$ (both unbiased for σ^2).
- A standard lack-of-fit F-test compares

$$F = \frac{SS_{\mathrm{lof}}/(k - p)}{SS_{\mathrm{pe}}/(N - k)} \approx F_{k-p,\ N-k}$$

under the null that the parametric mean is correct (requires replication at doses).

With replication at each dose

The **pure-error** estimator uses only within-dose variation and is **model-agnostic**:

$$\hat{\sigma}^2_{\text{pe}} = MS_{\text{pe}}, \quad \mathbb{E}[\hat{\sigma}^2_{\text{pe}}] = \sigma^2 \text{ regardless of mean-model correctness.}$$

If the parametric curve is correct, $\hat{\sigma}^2_{\text{model}} = MS_{\text{res}}$ is also unbiased and typically **more precise**:

$$\text{Var}(MS_{\text{pe}}) = \frac{2\sigma^4}{N-k}, \qquad \text{Var}(MS_{\text{res}}) = \frac{2\sigma^4}{N-p}, \quad (N-p > N-k \text{ if } p < k).$$

If the mean is mis-specified, $SS_{\text{lof}} > 0$ inflates MS_{res}.

Equality conditions (numerical equality vs expectation)

- **Equal in expectation:** If the D–R model is correct, $\mathbb{E}[MS_{\text{pe}}] = \mathbb{E}[MS_{\text{res}}] = \sigma^2$.
- **Exactly equal in-sample:** Only if $\hat{\eta}_j = \bar{y}_{\cdot j}$ for all j (i.e., the **saturated** ANOVA treatment-means model, or by chance). Otherwise $SS_{\text{lof}} > 0$ and $MS_{\text{res}} \neq MS_{\text{pe}}$.

Clarification

When a "correct" model **passes through the group means**, that corresponds to the **saturated** one-way ANOVA model: $\hat{\eta}_j = \bar{y}_{\cdot j}$ for all j. In this case,

$$SS_{\text{lof}} = 0, \qquad SS_{\text{res}} = SS_{\text{pe}}, \qquad MS_{\text{res}} = MS_{\text{pe}} \text{ (numerically).}$$

For any **constrained** parametric D–R family (e.g., log-logistic, Hill), $\hat{\eta}_j$ generally differs from $\bar{y}_{\cdot j}$, so SS_{lof} is typically nonzero and the MSEs differ in the sample—despite both being unbiased if the family is truly correct in expectation.

Consider the mathematical relationship between the **mean squared error (MSE)** derived from a fitted **dose–response (D–R) model** and the MSE obtained from a **one–way ANOVA** in which the **dose levels are treated as categorical treatments** (with replication at each dose): We assume independent, homoscedastic errors across observations and replication at each dose.

MSE comparison

We investigate the mathematical relationship between the **mean squared error (MSE)** derived from a fitted **dose–response (D–R) model** and the MSE obtained from a **one–way ANOVA** in which the **dose levels are treated as categorical treatments** (with replication at each dose).

Notation and setup

Let doses be x_j for $j = 1, \dots, k$, with $n_j \geq 2$ replicates at each dose and total $N = \sum_{j=1}^{k} n_j$. We observe

$$y_{ij} \;=\; \mu(x_j) + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{iid}}{\sim} (0, \sigma^2), \quad i = 1, \dots, n_j.$$

Define group means $\bar{y}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$.

We fit a parametric D–R mean function $\eta(x; \theta)$ with p estimable parameters and denote $\hat{\eta}_j := \eta(x_j; \widehat{\theta})$.

Sums of Squares and Mean Squares

Pure Error (within–dose)

The **pure–error** (within–group) sum of squares and its MSE are

$$SS_{\text{pe}} \;=\; \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\cdot j})^2, \qquad \text{df}_{\text{pe}} \;=\; N - k, \qquad MS_{\text{pe}} \;=\; \frac{SS_{\text{pe}}}{N - k}.$$

Model Residual and Lack–of–Fit

The **model residual** sum of squares and MSE are

$$SS_{\text{res}} \;=\; \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \hat{\eta}_j)^2, \qquad \text{df}_{\text{res}} \;=\; N - p, \qquad MS_{\text{res}} \;=\; \frac{SS_{\text{res}}}{N - p}.$$

Using the identity

$$y_{ij} - \hat{\eta}_j = (y_{ij} - \bar{y}_{\cdot j}) + (\bar{y}_{\cdot j} - \hat{\eta}_j)$$

and the fact that $\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\cdot j}) = 0$, we obtain the **lack–of–fit decomposition**

$$\boxed{SS_{\text{res}} \;=\; SS_{\text{pe}} \;+\; SS_{\text{lof}}, \qquad SS_{\text{lof}} \;=\; \sum_{j=1}^{k} n_j (\bar{y}_{\cdot j} - \hat{\eta}_j)^2 \;\geq\; 0.}$$

As a result,

$$\boxed{MS_{\text{res}} \;=\; \frac{N - k}{N - p}\, MS_{\text{pe}} \;+\; \frac{SS_{\text{lof}}}{N - p}.}$$

Under the usual regularity conditions with a correctly specified mean model,

$$\mathbb{E}[MS_{\text{pe}}] \;=\; \sigma^2, \qquad \mathbb{E}[MS_{\text{res}}] \;=\; \sigma^2,$$

so both are unbiased for σ^2 (but have different variances due to different dfs). When the parametric mean is misspecified, $SS_{\text{lof}} > 0$ inflates MS_{res}.

Lack–of–Fit Test

A standard lack–of–fit F-statistic is

$$F \;=\; \frac{SS_{\text{lof}}/(k-p)}{SS_{\text{pe}}/(N-k)},$$

which is approximately $F_{k-p,\,N-k}$ under the null that the parametric mean is correct (homoscedastic, independent errors, replication at doses).

Worked Algebraic Example (Two Doses)

Consider two doses x_1, x_2 with $n_1, n_2 \geq 2$ replicates and group means $\bar{y}_{\cdot 1}, \bar{y}_{\cdot 2}$, $N = n_1 + n_2$.

For dose j,

$$\sum_{i=1}^{n_j}(y_{ij} - \hat{\eta}_j)^2 = \sum_{i=1}^{n_j}(y_{ij} - \bar{y}_{\cdot j})^2 + n_j(\bar{y}_{\cdot j} - \hat{\eta}_j)^2.$$

Summing over $j = 1, 2$ yields

$$SS_{\text{res}} = SS_{\text{pe}} + \sum_{j=1}^{2} n_j(\bar{y}_{\cdot j} - \hat{\eta}_j)^2,$$

i.e., the same lack–of–fit decomposition.

Numerical Illustration

```r
# Example data: two doses, three replicates each
y1 <- c(10, 12, 8)  # dose x1
y2 <- c( 7,  9, 8)  # dose x2

n1 <- length(y1); n2 <- length(y2)
N  <- n1 + n2
k  <- 2            # number of doses

ybar1 <- mean(y1)
ybar2 <- mean(y2)

# Pure error SS and MS
SS_pe <- sum( (y1 - ybar1)^2 ) + sum( (y2 - ybar2)^2 )
df_pe <- N - k
MS_pe <- SS_pe / df_pe

list(SS_pe = SS_pe, df_pe = df_pe, MS_pe = MS_pe)
```

```
## $SS_pe
## [1] 10
##
## $df_pe
## [1] 4
##
## $MS_pe
```

```
## [1] 2.5
```

Case A: Saturated (ANOVA Treatment–Means) Model

```
eta1_sat <- ybar1
eta2_sat <- ybar2

SS_lof_sat <- n1*(ybar1 - eta1_sat)^2 + n2*(ybar2 - eta2_sat)^2
SS_res_sat <- SS_pe + SS_lof_sat

p_sat  <- k   # saturated model has one parameter per dose mean
df_res <- N - p_sat
MS_res_sat <- SS_res_sat / df_res

list(SS_lof = SS_lof_sat, SS_res = SS_res_sat, df_res = df_res, MS_res = MS_res_sat)
```

```
## $SS_lof
## [1] 0
##
## $SS_res
## [1] 10
##
## $df_res
## [1] 4
##
## $MS_res
## [1] 2.5
```

Case B: Constrained Model

```
eta1 <- 9.6
eta2 <- 8.4

SS_lof <- n1*(ybar1 - eta1)^2 + n2*(ybar2 - eta2)^2
SS_res <- SS_pe + SS_lof

p_constr <- 1
df_res   <- N - p_constr
MS_res   <- SS_res / df_res

list(SS_lof = SS_lof, SS_res = SS_res, df_res = df_res, MS_res = MS_res)
```

```
## $SS_lof
## [1] 0.96
##
## $SS_res
## [1] 10.96
##
## $df_res
## [1] 5
```

```
##
## $MS_res
## [1] 2.192
```

Summary

- $SS_{\text{res}} = SS_{\text{pe}} + SS_{\text{lof}}$.
- $MS_{\text{res}} = \frac{N-k}{N-p} MS_{\text{pe}} + \frac{SS_{\text{lof}}}{N-p}$.
- Equality holds if the model passes through all group means (saturated).

References

Benchmark dose technical guidance. (2012). *EPA/100/R-12/001*. https://www.epa.gov/risk/benchmark-dose-technical-guidance

Canadian Council of Ministers of the Environment (CCME). (2007). *Canadian environmental quality guidelines*. Canadian Council of Ministers of the Environment.

Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundamental and Applied Toxicology*, *4*(5), 854–871. https://doi.org/10.1016/0272-0590(84)90164-6

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, *50*(272), 1096–1121. https://doi.org/10.1080/01621459.1955.10501294

EFSA Scientific Committee. (2017). Update: Guidance on the use of the benchmark dose approach in risk assessment. *EFSA Journal*, *15*(1), 4658. https://doi.org/10.2903/j.efsa.2017.4658

EFSA Scientific Committee. (2022). Guidance on the use of the benchmark dose approach in risk assessment. *EFSA Journal*, *20*(10), e07584. https://doi.org/10.2903/j.efsa.2022.7584

European Chemicals Agency (ECHA). (2017). *Guidance on information requirements and chemical safety assessment: Chapter r.10 – characterisation of dose [concentration]–response for environment*. European Chemicals Agency. https://echa.europa.eu/documents/10162/13632/information_requirements_r10_en.pdf

Fisher, R., Barneche, D. R., Ricardo, G. F., & Fox, D. R. (2024). Bayesnec: An R package for concentration-response modeling and estimation of toxicity metrics. *Journal of Statistical Software*, *110*(5), 1–41. https://doi.org/10.18637/jss.v110.i05

Fisher, R., & Fox, D. R. (2023a). Introducing the no-significant-effect concentration. *Environmental Toxicology and Chemistry*, *42*(9), 2019–2028. https://doi.org/10.1002/etc.5610

Fisher, R., & Fox, D. R. (2023b). Introducing the no-significant-effect concentration. *Environmental Toxicology and Chemistry*, *42*(9), 2019–2028. https://doi.org/10.1002/etc.5610

Fox, D. R. (2009). Is the EC_x a legitimate surrogate for the NOEC? *Integrated Environmental Assessment and Management*, *5*(3), 351–353. https://doi.org/10.1897/IEAM_2008-043.1

Fox, D. R. (2010a). A bayesian approach for determining the no-effect concentration and hazardous concentration in ecotoxicology. *Ecotoxicology and Environmental Safety*, *73*(2), 123–131. https://doi.org/10.1016/j.ecoenv.2009.09.012

Fox, D. R. (2010b). Statistics and ecotoxicology: Shotgun marriage or enduring partnership? *Integrated Environmental Assessment and Management*, *6*(3), 501–502. https://doi.org/10.1002/ieam.158

Fox, D. R., Blasco, J., Campana, O., Hampel, M., & Chapman, P. M. (2016). Contemporary methods for statistical design and analysis. In *Marine ecotoxicology: Current knowledge and future issues* (pp. 39–66). Academic Press. https://doi.org/10.1016/B978-0-12-803371-5.00002-9

Fox, D. R., & Landis, W. (2016). Don't be fooled — a no-observed-effect concentration is no substitute for a poor concentration–response experiment. *Environmental Toxicology and Chemistry*, 1–8. https://doi.org/10.1002/etc.3459

Green, J. W., Springer, T. A., & Staveley, J. P. (2013). The drive to ban the NOEC/LOEC in favor of ECx) is misguided and misinformed. *Integrated Environmental Assessment and Management*, *9*(1), 12–16. https://doi.org/10.1002/ieam.1367

Guidance document on the use of the benchmark dose approach in risk assessment (OECD Series on Testing

and Assessment, No. 329). (2020). OECD. https://doi.org/10.1787/9153016e-en

Hendriks, A. J., Awkerman, J. A., Zwart, D. de, & Huijbregts, M. A. J. (2013). Sensitivity of species to chemicals: Dose–response characteristics for various test types (LC50, LR50 and LD50) and modes of action. *Ecotoxicology and Environmental Safety*, *97*, 10–16. https://doi.org/10.1016/j.ecoenv.2013.06.020

Jager, T. (2012). Bad habits die hard: The NOEC's persistence reflects poorly on ecotoxicology. *Environmental Toxicology and Chemistry*, *31*(2), 228–229. https://doi.org/10.1002/etc.746

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, *41*(1/2), 133–145. https://doi.org/10.2307/2333011

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362.

Mount, D. I., & Stephan, C. E. (1967). A method for establishing acceptable toxicant limits for fish–malathion and the butoxyethanol ester of 2,4-d. *Transactions of the American Fisheries Society*, *96*(2), 185–193. https://doi.org/10.1577/1548-8659(1967)96%5B185:AMFEAT%5D2.0.CO;2

National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. (2023). *Manual for the PROASTweb application, version 70.1.* PROAST web application manual. https://proastweb.rivm.nl/

Ritz, C. (2010). Toward a unified approach to dose–response modeling in ecotoxicology. *Environmental Toxicology and Chemistry*, *29*(1), 220–229. https://doi.org/10.1002/etc.7

Sánchez-Bayo, F., & Goka, K. (2007). Simplified models to analyse time- and dose-dependent responses of populations to toxicants. *Ecotoxicology*, *16*(7), 511–523. https://doi.org/10.1007/s10646-007-0158-9

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, *15*(6), 657–680.

Slob, W. (2002). Dose–response modeling of continuous endpoints. *Toxicological Sciences*, *66*(2), 298–312. https://doi.org/10.1093/toxsci/66.2.298

Terpstra, T. J. (1952). The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, *14*, 327–333.

U.S. Environmental Protection Agency. (2012). *Benchmark dose technical guidance* (EPA/100/R-12/001). U.S. EPA Risk Assessment Forum. https://www.epa.gov/risk/benchmark-dose-technical-guidance

Warne, M. S. J., Batley, G. E., Braga, O., Chapman, J. C., Fox, D. R., Hickey, C. W., Stauber, J. L., & Van Dam, R. A. (2018). Revisions to the derivation of the australian and new zealand guidelines for toxicants in fresh and marine waters. *Environmental Toxicology and Chemistry*, *37*(3), 622–637. https://doi.org/10.1002/etc.3996

Warne, M. St. J., Batley, G. E., Dam, R. A. van, Chapman, J. C., Fox, D. R., Hickey, C. W., Stauber, J. L., & Fisher, R. (2025). *Method for deriving australian and new zealand water quality guideline values for protecting aquatic ecosystems from toxicants – update of 2018 version* [Technical Report].

Williams, D. A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, *27*(1), 103–117. https://doi.org/10.2307/2528930