

Current challenges to the use of benchmark dose modelling in regulatory toxicology

In this guide

[In this guide](#)

1. [Benchmark dose modelling in a UK chemical risk assessment framework - cover](#)
2. [Benchmark Dose Modelling in a UK Chemical Risk Assessment Framework - Executive Summary](#)
3. [Benchmark dose modelling in a UK chemical risk assessment framework - Recommendations](#)
4. [Benchmark dose modelling in a UK chemical risk assessment framework - Use](#)
5. [Benchmark dose modelling in a UK chemical risk assessment framework - Advantages of BMD modelling](#)
6. [Current challenges to the use of benchmark dose modelling in regulatory toxicology](#)
7. [Benchmark dose modelling in a UK chemical risk assessment framework - COT's discussion](#)
8. [Benchmark dose modelling in a UK chemical risk assessment framework - Conclusions](#)
9. [Annex A - Introduction and Background](#)
10. [Annex A- Benchmark dose modelling](#)
11. [Annex A - Selected Previous Publications](#)
12. [Annex A - COT previous discussions](#)
13. [Annex A - NOAEL approach vs BMD approach](#)
14. [Annex A - Modelling the data](#)
15. [Annex A - Fitting the model to the data](#)
16. [Annex A - Bayesian vs frequentist approach](#)
17. [Annex A Case Study \(FSA Computational Fellow\)](#)
18. [Annex A - User experience](#)

19. [Annex A - Conclusions](#)
20. [Annex A - Questions on which the views of the Committee are sought](#)
21. [Annex A - List of Abbreviations](#)
22. [Annex A - Technical terms](#)
23. [Annex A - References](#)

BMD modelling, while recognised as a scientifically more sophisticated method for chemical risk assessment, faces challenges that affect its implementation, interpretation, and harmonisation across regulatory frameworks.

One of the most prominent challenges is the divergence in guidance and methodological preferences between regulatory bodies. EFSA has taken a leading role in promoting model averaging and Bayesian approaches, for instance, arguing that these methods better capture the uncertainty inherent in dose-response modelling. Their 2022 guidance formalised a shift from frequentist to Bayesian paradigms, introducing credible intervals and prior distributions as standard components of BMD analysis. In contrast, the US EPA's most recent guidance continues to favour a frequentist framework, advocating for the selection of a single best-fitting model based on statistical criteria such as the AIC. The EPA's rationale is rooted in concerns about the interpretability and complexity of model averaging, particularly when it comes to communicating results to non-specialist stakeholders. This fundamental difference in statistical philosophy creates inconsistencies in how BMD results are derived and interpreted, potentially leading to different health-based guidance values (HBGVs) for the same substance depending on the approach used. The challenge is that the US EPA has now made available a standalone software package named "The Model Averaging for Dichotomous Response Benchmark Dose (MADr-BMD) Tool" for download, which is currently not as user friendly as BMDS. However, Bayesian model averaging is also available in recent versions of BMDS. At the present time, the EPA does not provide technical guidance on Bayesian modelling or Bayesian model averaging.

Similar, an area of disagreement lies in the selection BMR values. EFSA recommends default BMRs of 5% for continuous data and 10% for quantal data, based on empirical comparisons with NOAELs, standard guideline study designs, and considerations of biological relevance. The US EPA, however, recommends defining BMRs for continuous data based on standard deviations rather than percentage changes, arguing that this provides a more standardised and statistically grounded basis for analysis. However, it does mean that choice of BMR becomes highly study dependent. These differing conventions can lead to

divergent BMD and BMDL estimates, even when using the same data set, complicating efforts to harmonise risk assessments across jurisdictions. Moreover, the lack of consensus on what constitutes a biologically meaningful or adverse effect further complicates BMR selection, especially in the context of novel endpoints such as gene expression or high-throughput screening data.

Technical challenges also persist in the practical implementation of BMD modelling. One such issue is model selection and fitting. While both EFSA and EPA provide suites of recommended models, the criteria for choosing among them are not always clear-cut and there is ongoing debate on which set of models should be used. In many cases, the underlying biological mechanisms of toxicity are poorly understood, making it difficult to justify the use of one model over another. This leads to a reliance on statistical fit rather than biological plausibility, which can undermine confidence in the resulting estimates. Additionally, the process of fitting models to data is not always straightforward. Problems such as non-convergence, over-parameterisation, and biologically implausible parameter estimates are common, particularly when dealing with sparse or noisy data. Similarly, the use of constraints in model fitting is a contentious issue. The EPA recommends constraining certain parameters to avoid unrealistic dose-response curves, such as steep supralinear slopes at low doses. EFSA, on the other hand, cautions against overly restrictive constraints, arguing that they can bias results and exclude plausible models. This disagreement reflects broader tensions between statistical rigour and biological realism and underscores the need for flexible yet principled approaches to model fitting.

Software variability adds another layer of complexity. Different BMD software packages—such as BMDS, PROAST, BMDEpress, and EFSA R4EU implement different models, algorithms, and default settings. Although efforts have been made to try to ensure consistency in the algorithms used, the same dataset analysed using the different tools can still yield different BMD and BMDL values. This variability is particularly problematic in regulatory contexts, where consistency and reproducibility are paramount. While efforts continue to be made to standardise software outputs and improve transparency, significant differences remain in terms of usability, data input requirements, and support for advanced features such as model averaging or Bayesian inference.