

EFSA-RIVM Meeting on divergent opinion (Art. 30 Regulation (EC) No 178/2002) on the Guidance on the use of the benchmark dose approach in risk assessment (doi.org/10.2903/j.efsa.2022.7584)

Date: 1 December 2023

Location: EFSA, Parma

Attendees:

- **Meeting Chair**
Matthew Wheeler (Scientist at the Biostatistics & Computational Biology Branch of the National Institute of Environmental Health Sciences - US)
- **EFSA staff**
Claudia Roncancio-Peña (Head of Methodology and Scientific Support Services Unit)
Bernard Bottex (Senior Scientific officer - Knowledge, Innovation and Partnership Management Unit)
José Cortiñas Abrahantes (Senior Scientific officer - Methodology and Scientific Support Services Unit)
- **EFSA experts**
Josef Slatter (Chair of the EFSA BMD Working Group)
Marc Aerts (Member of the EFSA BMD Working Group)
- **National Institute for Public Health and the Environment (RIVM) staff:**
Annelies Dijk (Head of Centre for Prevention, Lifestyle and Health)
Guangchao Chen (BMD expert - Centre for Prevention, Lifestyle and Health)
Bas Bokkers (BMD expert, Toxicologist & Risk Assessor - Centre for Safety of Substances and Products)
Martine Bakker (Senior Scientist and Project leader - Centre for Safety of Substances and Products)

Apologies:

- **EFSA staff**
Nikolaus Georg Kriz (Head of Risk Assessment and Services Department)

Background

EFSA published an update of the “Guidance on the use of the benchmark dose approach in risk assessment” on 25 October 2022 following a public consultation round in which RIVM raised concerns about some aspects considered in the methods proposed. In addition, RIVM addressed its concerns on two other occasions. RIVM, unconvinced with EFSA’s response to the points raised, asked EFSA to start an Article 30 (4) of Regulation 178/2002 procedure with respect to the diverging opinions of RIVM and EFSA on the matter of BMD modelling on 2 June 2023.

1. Opening and objectives of the meeting

The chair welcomed the participants, who shortly introduced themselves. Apologies for absence were received by Nikolaus Georg Kriz (Head of Risk Assessment and Services Department). EFSA presented the objectives of the meeting, which were to outline and discuss the different views regarding the EFSA “Guidance on the use of the benchmark dose approach in risk assessment”, to establish whether the divergence could be resolved and to identify the next steps to continue the long-standing collaboration between EFSA and RIVM on the topic of BMD modelling.



2. RIVM Presentation: "Comments on the EFSA BMD guidance 2022"

RIVM presented their main concerns regarding the latest EFSA guidance document on BMD modelling.

Bayesian BMD Tool

The new EFSA Bayesian BMD tool and the former EFSA BMD tool provide different outcomes for BMD estimation in a number of occasions. RIVM have found that in specific settings, the ratio between the two BMDs can be high for both continuous and quantal data. RIVM expresses its concern for the chemical risk assessment in Europe, as the new Bayesian approach may result in very different BMD(L)s compared to the former used frequentist approach. RIVM suspects that the differences between the two methods may be (partly) due to certain important principles in BMD modeling that RIVM considers are not followed by EFSA's updated guidance on BMD modelling. RIVM outlined these modeling principles, which RIVM believes EFSA needs to follow.

Model framework

RIVM presented their view on what the model framework should be, with a focus on continuous data. The first point RIVM outlined regards the definition of distance of doses in toxicological studies (e.g. the difference between two doses, or the ratio between two doses). RIVM considers that the distance between two doses must be looked at on a multiplicative scale, as their ratio. For example, an increase of dose from 1 mg to 2 mg is equivalent, on an additive scale, to an increase from 11 mg to 12 mg. However, in toxicology, these increases are not considered equivalent. Toxicologically, an increase from 1 mg to 2 mg is equivalent to an increase from 10 mg to 20 mg. I.e. these doses are equivalent on a multiplicative scale. A similar reasoning should be considered for continuous biological responses as well (e.g. body weight). As a result, the assumptions of the modelling framework should reflect the multiplicative scale. This leads to RIVM's point of view that models for the BMD modelling should be multiplicative according to the expression: $f(x) = ac^{F(x;b,d)}$. Where parameter a is the geometric mean background response; b the potency/sensitivity parameter; c the maximum fold change of response; and d the steepness parameter, where steepness relates to log-dose. In this way, the models defined by this expression are analogous, i.e. the same parameter (e.g. a) carries the same biological meaning in different models. Besides, to be consistent with the multiplicativity of continuous biological responses, the within-group variation of responses should be considered as lognormally distributed, and the BMR should be expressed as a percent change. In case there are arguments pointing to the endpoint being additive, then the model expressions and all other assumptions should also be defined in an additive way, because assumptions made in model framework should be consistent with each other, and should not conflict with each other.

RIVM also illustrated their view on toxicologically important properties of the models defined with the expression above:

1. They never predict negative values, as biological responses are not negative;
2. The outcome of the BMD analysis does not depend on the measurement unit used for the response;
3. The ratio of BMDs in different compounds or populations is the same when the shape of the curves in both populations is the same whatever the value of the BMR.

RIVM also presented the main issues with the model framework in the latest EFSA guidance document on BMD modelling:

Issue #1: Family 1 models



- The EFSA BMD models allow the use of both the log-normal and normal distributions for the within-group variation in the response. As a consequence, the process of model averaging results in a mix of multiplicative and additive assumptions in these models, which is inappropriate from RIVM's standpoint;

-The EFSA models in Table 2 in the 2022 guidance are not analogous since the parameters (a vs. α (alpha), and c vs. γ (gamma)) have different meanings in the two models of expressions (1) and (2). While parameter c is the maximum fold change, parameter γ reflects a maximum fold change on the log-scale, which has no biological meaning. According to RIVM the parameters in both expressions are related to each other by:

$$\alpha = \ln(a)$$
$$\gamma = 1 + \ln(c) / \ln(a)$$

which shows that parameter γ is an expression of parameters a and c, which lacks biological interpretation and shows that model parameters are not analogous in all model definitions.

- Lognormal Family 1 models exhibit a chaotic behaviour. As a consequence of the behaviour of the lognormal family 1 models, the fitting of the data depends on the measurement unit of the response, which is (obviously) not correct. E.g., if a=1 mg/L, it is not possible to model a trend, but when expressing the concentration in a different unit (a=1000 μ g/L) a dose-response can be obtained. The issues related to Family 1 models were raised by RIVM during the public consultation phase of the EFSA guidance document on BMD (comment #38). However, RIVM assesses that RIVM's concern was not addressed by EFSA.

Issue #2: Family 2 models

Family 2 models are based on other expressions than $f(x) = ac^{F(x;b,d)}$. They are also not analogous to the family 1 models (and all Family 2 models are not analogous within their own family due to the original vs. log-response scale). Examples of non-analogy of model parameters and their consequences were presented:

- In EFSA's increasing Family 2 models: $\mu(0) \neq a$, hence parameter a is not the background response.
- In a covariate analysis differences in parameter a between subgroups cannot be interpreted in a biological meaningful way.
- In EFSA's Family 1 models $c = \mu(\infty)/\mu(0)$, which is the maximum fold change, and in Family 2 models $c = \mu(\infty)$ which is the maximum response. Hence, parameter c in Family 1 is not analogous with parameter c in Family 2.

The issues related to the analogy of Family 2 models were also raised by RIVM during the public consultation phase of the latest EFSA guidance document on BMD (comment #38). RIVM considers that RIVM's concerns were not addressed by EFSA's response. RIVM also has the impression that the 'natural parameters' are of the most interest in the guidance; while models in Table 2 are only reparametrized expressions of the 'natural parameters'. However, EFSA did not use models that directly include these 'natural parameters', but presented other model expressions and then took the burden of an extra step for reparametrization. This makes the BMD models in Table 2 of EFSA's latest guidance only intermediate expressions rather than the actual models used in the BMD modelling, RIVM considers that the models in the guidance should be updated accordingly. Besides, in a covariate analysis with the EFSA Bayesian tool, using the Family 2 models, RIVM has found that the estimated BMDs may depend on the measurement unit used. An example was given to illustrate this point, a covariate analysis was carried out considering 4 chemicals as a covariate. While the model averaging results for both analysis (using the original units and the transform units) provide essentially the same results, for the logistic model in model Family 2 considering the lognormal distribution (one out of the 16 models fitted) resulted in different BMD credible intervals for the four chemicals. The resulting logistic model (L4_LN) selected a different model



when considering the response variables in different units, producing then different BMDL, BMD and BMDU, for each of the measurement unit of the responses. Changing the unit of the response from cm to mm, which should not influence the BMD, leads to a change of BMD of the four chemicals, ranging from a small change to up to a factor of 50 for one of the 16 models used in the model averaging, while the model averaging results produce the same results for both unit of measurements.

Issue #3: Priors

The guidance on setting the priors is not clear. Specifically, there is no clear description provided to users on how to set the priors or how to check the prior distribution's influence on the response. This could be an issue since an incorrect use of priors could impact the results. The issues related to priors were also raised by RIVM during the public consultation phase of the latest EFSA guidance document on BMD (comment #3).

Issue #4: Quality of the BMDL as Reference Point

The 2022 guidance suggests that the BMD/BMDL ratio is used to assess the quality of the BMD analysis. Another criterion is that the BMD should not be lower than 10 times the lowest non-zero dose. RIVM argues that the BMD/BMDL ratio is not an appropriate measure of uncertainty because the BMD itself is uncertain as characterized by its confidence interval (BMDL to BMDU range). RIVM considers the BMDU/BMDL ratio as the criterion to judge the quality of the data. The precision of the BMD is reflected by the BMDU/BMDL ratio. RIVM believes that suggesting more criteria is confusing and unnecessary.

Issue #5: Use of NOAEL

The latest EFSA guidance document on BMD suggests the use of NOAEL as an alternative to the BMDL when data are considered of insufficient quality to derive a BMDL. RIVM disagrees with this suggestion. When the toxicological data are not sufficiently informative to derive a (narrow) BMD confidence interval, then RIVM considers that such data are also insufficiently informative to derive a NOAEL (or LOAEL). According to RIVM, deriving a NOAEL from poor data is not justified because it hides the lack of dose-response information and ignores the uncertainties associated with the data. The issue related to the use of NOAEL was raised by RIVM during the public consultation phase of the latest EFSA guidance document on BMD (comment #66).

Issue #6: Suitability of data for modelling/pairwise testing

The latest EFSA guidance document on BMD suggests the use of pairwise testing for the selection of the datasets to be analysed.

One main reason for developing the BMD approach, is that it avoids relying on (pairwise) significance testing, which has several documented drawbacks. The 2022 guidance re-introduces the use of such testing to determine the suitability of data for modeling. Pairwise testing to test data suitability is not recommended because this will result in low power of detecting the effect and therefore many important endpoints may be unjustly discarded in the risk assessment. RIVM expressed the view that pairwise comparison should be avoided since it has several documented drawbacks. This issue was raised by RIVM during the public consultation phase of the latest EFSA guidance document on BMD (comment #79).

Other issues

Several issues related to the tool accompanying the 2022 guidance were listed in the presentation, but were not discussed due to lack of time.



Conclusions

RIVM remarked once again how differences between the outcomes of the two tools are often present, for reasons that are partially unknown, but that should be addressed. The 2022 EFSA guidance document on BMD modelling proposes the application of models, assumptions and procedures that do not correspond to the fundamental characteristics of toxicological and biological data and are different with respect to existing practices. Although the Bayesian approach is considered a step forward in BMD modelling, it should be combined with a statistically coherent model framework and clear guidance on setting priors. RIVM also considers that in contrast to the aim of harmonization, the EFSA guidance moves away from it by applying other models, assumptions and procedures compared to existing practice in USA and Europe. Furthermore, RIVM believes that harmonization is only justified when ensured that methods and assumptions are scientifically correct.

3. EFSA Presentation: "Bayesian BMD: Comparison with the Frequentist approach in PROAST"

The presentation aimed at comparing the two different approaches (Bayesian vs. Frequentist in PROAST) applied by EFSA and RIVM respectively as well as at answering the issues raised by RIVM.

Motivation to update the BMD Guidance

The main aim of the EFSA guidance document on BMD was to adhere to specific concepts contained in the update of Chapter 5¹ on dose response assessment and derivation of health-based guidance values of the WHO Environmental Health Criteria 240 on dose-response models. In particular, the specific concepts are:

- The BMR should be based on biological relevancy;
- As there are cases where the data are more suggestive of one distributional form over the other, for continuous data, two distributional assumptions at least should be considered: normal and log normal;
- A general model family of dose-response models, for quantal and continuous data, should be used when doing Bayesian model averaging, which is the preferred approach;
- The d parameter to be constrained or not, given that at control models might induce infinite slope if it is not constrained;
- Use prior distribution for parameter d that allows the framework to take values that are below one but with a reduced probability;
- The Bayesian framework allows to include extra information with the use of informative priors, which may prevent problems in the analysis in data poor situations;
- The results of the modelling need to be useful and applicable in the regulatory context.

Model framework description

Model components

The components of the models were listed including what concepts are considered to be able to define the models. An explanation on how the parameters a , b , c , and d are all parametrized in terms of background response, maximum response, and BMD (natural parameters) and how this different parametrization of the model provide deterministic relationships with respect to the natural parameters was given. An example, in the LOGN model, the background parameter corresponds to e^a ($a' = e^a$).

The prior distribution for the parameter d currently used by EFSA was presented with an explanation on how this prior has a larger probability mass for values of d to be below one, with respect to the prior used by US-EPA (in particular, d has a probability of 15% of taking values

¹chapter5-dose-response.pdf (who.int)



below 1, while the one used by U.S. Environmental Protection Agency (US-EPA) for quantal models is only having 5% probability of being below 1).

RIVM is not constraining the parameter d and, for specific data patterns, this will result in the model favouring the parameter to be estimated below 1, which will induce confidence intervals for the Frequentist approach to have lower bounds that are unnecessarily small, close to zero.

Illustrations of divergences with examples

BMDL-BMDU intervals

Examples of BMD modelling performed with other Bayesian software were shown underlining how these methods, although different, provide very similar results, while the results of the frequentist approach often differ in specific data patterns. Most of the time, these discrepancies appear due to the fact that the d parameter is estimated to be below 1.

The effect of d

An example to illustrate how, in many cases (40% of the times in the simulation example considered), the Confidence Interval (CI) generated with the frequentist approach PROAST does not contain the true BMD and the lower bound of the confidence interval is estimated to be very close to zero was also given. The example is based on simulated data, considering 1000 simulations. It was reported how instead, with EFSA's Bayesian framework, the CI lower bound results closer to the true BMD and the upper bound is estimated always higher than the true BMD. The credible intervals obtained in each simulation always contains the true BMD.

In relation to the use of N distributions, when generating data from a LOGN distribution, the models identified with higher weights are those considering LOGN and the model averaging results are solely based on LOGN models. This illustrates that EFSA's model averaging tool automatically upweights an appropriate distribution, and downweights an inappropriate distribution.

Conclusions

EFSA remarked that the assumptions made by the two paradigms (Frequentist and Bayesian) influence the resulting confidence/credible intervals, and differences in results should be therefore carefully considered and that the update of the EFSA Guidance on the use of the benchmark dose approach in risk assessment was triggered by the objective of harmonization, i.e. by the update of Chapter 5: Dose-Response Assessment and Derivation of Health-Based Guidance Values of the EHC 240: Principles and methods for the risk assessment of chemicals in food. As many points of divergence as time allowed were discussed during the meeting.

4. Discussion and way forward

1. Requirements of BMD/allowable models

RIVM

The BMD should be defined in the model in such a way that it is not dependent on the incidental background response of the study population and conclude that model family 2 is unsuitable for deriving a BMD for continuous endpoints.

EFSA

For quantal endpoints, (logit and probit models) are not complying with this requirement and have been used in all previous EFSA guidance. Also, for other definitions of the BMD, the expressions obtained might depend as well on all parameters in the model. Another point brought forward was the fact that as the model averaging is the preferred approach, which was also acknowledge by RIVM, even if every individual model is satisfying the requirements, when building the average curve, the requirements for the average curve will not be further satisfied. The BMD derived from the average curve for each bootstrap will depend on all estimated backgrounds in each of the models included in the model averaging curve, as well as all other parameters included in the models used to build the average curve.



2. Analogous models:

RIVM

Models used for continuous endpoints should be analogous, i.e. the regression parameters across different models should have the same meaning (interpretation). Thus models defined in family 2 should not be used. From the 2022 guidance it is clear that model parameters have different interpretations, as illustrated in RIVM's presentation and according to the guidance itself (e.g. on page 13 of the guidance, in the continuous family 1 models parameter c is defined as the fold change compared to background ($c = \mu(\infty)/\mu(0)$), while in family 2 parameter c is not expressed relative to background but as the (absolute) maximum response ($c = \mu(\infty)$). RIVM reiterates that dose-response models require to be analogous to facilitate correct model averaging, interpretation of covariate analysis results and results in general. The requirement of analogy will exclude particular models, resulting in a reduction of the number of dose-response models that can be used in model averaging. This may conflict with the premise that model averaging will be more reliable when more models are included, although no evidence is available considering the optimal number of models. Hence, RIVM advocates the use of a suit of interpretable, analogous models over a larger number of non-analogous models which hamper statistical and toxicological interpretation and comparison.

EFSA

All models from family 1 and family 2 are defined in terms of the natural biological parameters (background response, maximum fold change and BMD) on the original scale, so in EFSA's view these parameters do have the same meaning across models and so they are all analogous. Furthermore, the parameter d is not having the same meaning/value in any model. For that reason in EFSA's opinion the models RIVM advocate are not analogous neither. PERT prior distributions are defined on these natural, biological parameters. Model-specific estimates of the natural parameters can be interpreted in exactly the same way for all models.

The parameters are deterministically derived one from the other, posing no issues. When using model averaging, the idea is to bring into the MA a sufficiently large set of sufficiently different models that are reflecting different natures of dose-response effects, as the true response relationship is unknown and any mathematical representation is limited. According to the principle of MA, the more diverse the models are, the higher the chance that a candidate model provide estimates closer to those of the true underlying model.

Chair

In the view of the Chair, for model averaging, the more curve shapes you add, the better the dose response space can be described and the better the BMD can therefore be described. Given that the true dose-response is unknown, we add more curvature to better get the BMDL and BMDU.

3. Priors:

RIVM

The guidance on setting the priors is not clear. Choice of priors is important as they may determine the outcome of the analysis. In EFSA 2022 there is no clear guidance for setting the priors. The PERT-distribution appears only explorative. Although this was recognized by EFSA in the public consultation, no guidance was added to the final document. In addition, RIVM suggests to include method(s) to check possible dependence of results on the choice of prior(s).

EFSA

The information on how the default weakly informative prior distributions were chosen and the procedure followed to derive this distribution providing detail on the simulations performed to assess the impact and the reliability of the result for the final priors used in the EFSA BMD tool are published in the Report from the Hasselt University (2022, which can be found [here](#)) available in EFSA's website. In the Knowledge junction repository (<https://zenodo.org/records/7118583>) several html files are stored providing a description of the simulation results for the settings considered in the development of the framework.



Additionally, it was mentioned that in section 2.6.4 of the guidance considerations about the prior are given, providing as well comments on the example provided in appendix C, in which the effect of using informative priors is presented, especially in the context of datasets containing scarce information on how the dose affect the responses, in which only control and maximum effect is observed. The example illustrating this is presented in pages 51 up to 55.

EFSA has an ongoing project (December 2022-December 2024) aimed at creating a repository of priors based on data used in previous risk assessments. These priors will be obtained using the Bayesian modelling framework for each of the models included in the EFSA guidance. The final aim is to better inform the users on how to use priors and build informative priors aiding the BMD modelling process.

4. Log normal distribution:

RIVM

The log normal distribution is the most reasonable distribution for continuous data encountered in toxicological studies, since biological response is multiplicative in nature. Assumptions in the model framework should be consistent with this choice of scale. The choice of log normal distribution is based on toxicological rationale. It allows to define the models in the appropriate metric (ratios instead of differences). The models should define the effects in a multiplicative fashion.

EFSA

The framework allows to only use the log-normal distribution, by putting the prior on the normal distribution to be 0, based on objective, scientific and verifiable evidence. There is no proof that the log-normal is the only "true" distribution in this field of application and there is no field in data science, where the distribution is *a priori* known with certainty. Moreover, simulations have shown that the weights used in model averaging account for the appropriate distribution: the appropriate distribution gets the highest weights, the inappropriate distribution the lower ones; if data contain no clear evidence against one or the other distribution, weights are more evenly distributed among both distributions. So, there is no need to exclude any distribution and to be not open to other choices than the log-normal distribution. Also, the update of the BMD guidance was triggered to align with the updated chapter 5 of the EHC 240 that foresees other distributional assumptions than the lognormal, and the normal distribution is mentioned as a potential distribution that can be used to describe the nature of the scatter (variability of the response) that may be observed in an experiment.

5. Use of NOAEL:

RIVM

As explained during the presentation (section 2 of the minutes), the use of the NOAEL approach as an alternative to the BMD is not appropriate and the derivation of a NOAEL from poor data is not justified.

EFSA

In the guidance, it is stated that NOAEL approach would be considered only in very specific circumstances, which are described in the guidance in Section 2.6.5, and alongside the BMD credible interval characterizing the uncertainty, as recommended in the WHO Chapter 5. In some RA situations, there is no response up to the highest dose tested. This is the rationale to include the use of the NOAEL in the guidance: to deal with those extreme cases.



6. Suitability of data for modelling/pairwise testing:

RIVM

The EFSA BMD guidance suggests the use of pairwise testing for the selection of the datasets to be analysed. Pairwise comparison should be avoided since it has several documented drawbacks. The presence of dose-effect should be assessed using trend tests instead.

EFSA

The pairwise comparison described in the guidance is provided as an aid to establish if the data contain enough information to build the dose response with a certain level of accuracy. This will ensure the estimation of the BMD with a certain accuracy. The guidance states that the pairwise comparison is only used to flag "data poor" situations and not to define stopping rules to conduct the BMD evaluation. The procedure described in the guidance to assess dose-effect is not based on the pairwise comparison, the Bayesian factor comparing the NULL with the saturated model is proposed for this purpose. This is analogous to the procedure used in the frequentist paradigm, in which the AIC from the NULL and the saturated models are compared to establish if there is or no dose-response effect.

General discussion

EFSA

Regarding the results differences between the approaches (Frequentist PROAST and Bayesian EFSA), an example illustrating how the use of the Frequentist approach in PROAST will fail if the data doesn't have enough information about parameter d was presented. It was shown how the profile likelihood of the parameter d behaves, not allowing to build a reliable confidence interval. It was also mentioned how the use of the Bayesian approach would address this situation by adding information through the use of priors, even when using weakly informative priors. It was underlined how one of the main issues raised was the divergence between the BMDs obtained with the two approaches. The discussion eluded to the fact that differences were obtained due to the estimation of the parameter d , and according to EFSA this has been illustrated to be ill-posed in the frequentist framework.

RIVM

Bringing the discussion back to the Diverging Opinion on the BMD Guidance 2022, RIVM emphasizes that it is unclear from the guidance how and based on what information the prior for parameter d currently used by EFSA is derived (section 2, issue #3). EFSA's assumption that it is unlikely that parameter d has a value below 1 apparently stems from the assumption that doses should be viewed on an additive scale, whereas RIVM argued that doses should be considered on a multiplicative scale (see section 2, "model framework"). RIVM stresses that RIVM has no objections against the use of the Bayesian approach. RIVM believes that the focus of the discussion should be on how the Bayesian BMD model framework should be developed.

5. Concluding remarks

During the discussion, attention was drawn by the Chair on the importance to differentiate between *potential misinterpretation* of the guidance document and *scientific methodological issues*; bilateral discussion with RIVM will be held to understand these two different situations:

- *Potential misinterpretation* on some of the aspects referred to above (e.g. the use of pairwise comparison to evaluate dose-effect) could be clarified by adding even more information/clarification in the guidance.
- The *scientific methodological disagreements* should be further discussed between EFSA and RIVM in the coming months, envisaging a possible collaboration between both institutions.

From the presentations and discussion, it is clear that RIVM and EFSA have different perspectives on the BMD modelling as proposed in the 2022 guidance. The current meeting did not resolve the



issues brought up by RIVM. As a consequence, it is concluded that there is a divergent opinion between EFSA and RIVM regarding BMD modelling as described in the 2022 guidance. Nevertheless, both parties agree to continue the discussion in a future meeting with the aim to solve the scientific divergence. Both parties expressed their willingness to continue the long-standing collaboration on BMD modelling and identified some specific activities. Firstly, as described above, the meeting participants agreed to have a discussion on the issues presented by RIVM. In addition, the Chair has proposed some questions, identifying opportunities for follow up discussions (see Annex A). In addition, EFSA proposed that activities related to the research on informative priors for parameters in the models are considered as suitable topics for collaboration. In particular, a suggestion from EFSA for RIVM to follow up on the activities of the project related to priors², which started in December 2022 and ends in December 2024. Furthermore, the Chair of the meeting invites EFSA and RIVM to continue the discussion on BMD modelling during an international workshop with an extended group of toxicologists and statisticians. The chair thanked all participants for their availability and contributions and expressed his willingness to collaborate more in the future on this topic. Both EFSA and RIVM agreed on 4 June 2024 that these minutes reflect the opinions of both parties and that the art. 30 procedure will be closed once the minutes of this meeting are published on the EFSA website.

References:

Hasselt University (2022). EFSA Platform for Bayesian Benchmark Dose Analysis. EFSA Supporting publication 2022:EN-7740. 91 pp. doi:10.2903/sp.efsa.2022.EN-7740

²Outsourcing activity: Informative priors repository for BMD (OC/EFSA/AMU/2020/02 – SC2)
Page 10 of 12



Abbreviations

BMD = Benchmark Dose
BMDL = Benchmark-Dose Lower bound
BMDU = Benchmark-Dose Upper bound
BMR = Benchmark Response
CI = Confidence Interval
FC = Fold Change
LOGN = Log-normal
MA = Model Averaging
N = Normal
NOAEL = No Observed Adverse Effect Level
PC = Public Consultation
RA = Risk Assessment



Annex A - Set of questions proposed by the chair for starting further discussion

The chair proposed the following set of questions to reframe the issues raised and to find a collaborative way forward. It should be noted that these questions were not discussed in the current meeting and the statements do not necessarily reflect the opinion of both parties.

- EFSA's Guidance seems incomplete but a good step forward, how can we complete it together and make it better? Can we do this on an expedited timeline to prevent confusion?
- In the EFSA tool there seems to be problems if the original data is multiplied by a factor, representing scale change, when considering covariate analysis, the BMD results changes (shouldn't happen). In PROAST, the constraint can lead to unintuitive fits and the bootstrap may have theoretical problems. How can we fix both?
- Grey box: Model averaging appears to be a black box to toxicologists. Is there a way to develop statistics to make it understandable, but still statistically, correct?
- Can we develop a way to have a biological BMD, and then attempt to provide information on approximation of parameters a , b , c , and d to have biological meaning, such as background response, maximum fold change, benchmark dose and a curvature parameter (d)?
- Mathematical issues for d' do exist in certain datasets, how do we set priors to be biologically meaningful? Model framework for prior setting?
- What are some better ways to identify a data-poor settings?
- Chemical companies just want to know "what do I need to do for you to regulate this?" and they want that to be stable. How can we help them be confident in this?
- What do we do for newer bioassays that are only useful when they are normalized?
- How do we communicate the deficiencies of unconstrained maximum likelihood while maintaining confidence in past decisions?