The Japanese Society of Toxicology

Fundamental Toxicological Sciences

URL : http://www.fundtoxicolsci.org/index_e.html

Original Article

Validation of the statistical parameters and model selection criteria of the benchmark dose methods for the evaluation of various endpoints in repeated-dose toxicity studies

Mariko Matsumoto[1], Mutsuko Hirata-Koizumi[1,a], Tomoko Kawamura[1], Sawako Sakuratani[1,a], Atsushi Ono[2] and Akihiko Hirose[1]

[1]Division of Risk Assessment, Biological Safety Research Center, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa, 210-9501, Japan
[2]Division of Pharmaceutical, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 1-1-1 Tusimanaka, Kita-ku, Okayama, 700-8530, Japan

(Received April 4, 2019; Accepted April 12, 2019)

ABSTRACT — The benchmark dose (BMD) approach is one of the important techniques in dose-response assessment for the risk assessment of chemicals and adapted by various international organizations. We investigated the appropriateness of the statistical parameters and model selection criteria for BMD lower bound (BMDL) estimation by BMD software (BMDS) (developed by the US Environmental Protection Agency) and PROAST (developed by the National Institute for Public Health and the Environment of the Netherlands). Publicly available repeated-dose toxicity study data (226 dichotomous datasets and 151 continuous datasets) were used for the investigation. Our findings were applied to establish BMD technical guidance for BMDS for the evaluation of various endpoints in repeated-dose toxicity studies. Under the Japan Chemical Substance Control Law (CSCL), the DRA-BMDS guidance (i.e., Division of Risk Assessment-BMDS guidance) is used for the evaluation of a "Priority Assessment Chemical Substance." Namely, selecting of an extra risk of 10% (dichotomous data) or a level change of 1SD (continuous data) as a default benchmark response. Running all the models without or with parameter constraints. Selecting the model that calculated the lowest BMDL but excluding the one that estimated a BMD/BMDL ratio ≥ 10 or lowest dose/BMDL ratio ≥ 10. We believe that the DRA-BMDS guidance can assist risk assessors in the selection of the BMD model.

Key words: Benchmark dose, Benchmark response, BMD, BMDL, BMDS, PROAST

INTRODUCTION

In the risk assessments of chemicals, the maximum tolerable exposure levels, such as acceptable daily intake (ADI) and tolerable daily intake (TDI), have been established in the "dose-response assessment." This step involves the analysis of the relationship between the amount of exposure and health effects, and a different method is historically adopted for cancer and noncancer health effects (Environmental Protection Agency [EPA], 2012). For noncancer health effects, the no observed (adverse) effect level (NO[A]EL) or lowest observed (adverse) effect level (LO[A]EL) was traditionally used as the point of departure (POD) to establish tolerable intake levels in human, but several disadvantages were indicated in the NO(A)EL/LO(A)EL approach (EPA, 2012). For example, the NO(A)EL/LO(A)EL depends largely on the study design, such as dose selection, dose spacing, and

Correspondence: Mariko Matsumoto (E-mail: mariko@nihs.go.jp)
[a]*Previous affiliation address*

sample size in animal experiments. If the dose spacing is large, NOAEL may be quite different from the actual biological effect threshold. Occasionally, NO(A)EL cannot be determined owing to significant changes at all doses, and a further addition or adjustment of the uncertainty factor is applied for establishing the TDI or ADI.

To overcome these issues, the benchmark dose (BMD) approach was developed. BMD is a dose level that is expected to result in a predetermined change in effect (called as the benchmark response [BMR]). BMD lower bound (BMDL), which refers to the lower limit of a one-sided 95% confidence interval on the BMD, is increasingly used as an alternative to NOAEL to establish a more objective TDI or ADI. The BMD approach has several advantages over the NO(A)EL/LO(A)EL approach, as indicated in several papers (Oberg, 2010; Sand *et al*., 2008; Slob, 2014a, 2014b). The BMD approach is also applied to the dose-response assessment of genotoxic carcinogens, wherein a specific cancer risk level is derived from the calculated BMDL by using a linear extrapolation.

Regarding the BMD approach, some guidance documents are available. The US EPA released the latest version of the "Benchmark Dose Technical Guidance" in June 2012 (EPA, 2012). The International Programme on Chemical Safety (IPCS) published the "Principles for Modelling Dose-Response for the Risk Assessment of Chemicals" in 2009 (IPCS, 2009). The European Food Safety Authority (EFSA) also published the "Guidance of the Scientific Committee on Use of the Benchmark Dose Approach in Risk Assessment" in 2009 (EFSA, 2009) and has been recently updated (Hardy *et al*., 2017). However, the BMD analysis method recommended in these guidance documents had some differences, particularly on BMR selection, modeling constraint, goodness-of-fit evaluation, and model comparison. Thus, we had to establish a standard procedure to use the BMD approach for the evaluation of chemicals, particularly contaminants in foods or industrial chemicals in the environment in Japan.

We derived the BMDs and BMDLs under various conditions to evaluate their adequacy. The dose-response data employed in the US EPA Integrated Risk Information System (IRIS) and WHO guidelines for drinking water quality, as well as data obtained in the Japanese safety programmes for existing chemicals, were analyzed using BMDS and PROAST. The overall BMDLs (the BMDL we selected) were compared with the NO(A)EL or LO(A)EL. On the basis of the findings of this investigation, we established a BMD technical guidance documents for BMDS and PROAST. The guidance for BMDS (Division of Risk Assessment [DRA]–BMDS guidance) established at this time is now used for the evaluation of a "Priority Assessment Chemical Substance" under the Japan Chemical Substance Control Law (CSCL), which is available in Japanese in the website of DRA of the National Institute of Health Sciences (http://dra4.nihs.go.jp/bmd). The purpose of DRA-BMDS guidance used for CSCL is to select the most conservative BMDL from a feasible dose range. It is because the scope of the CSCL is to prevent environmental pollution by industrial chemical substances under the precautionary principle.

Recently, a new concept for BMDL derivation, namely, model averaging, was introduced by the EFSA (Hardy *et al*., 2017), and the EPA also released the Bayesian dichotomous model averaging feature in BMDS 3.0 (EPA, 2018). This investigation does not cover the new features of both aproaches; however, we still need guidance for the model selection criteria in the BMD approach if the model averaging method is inappropriate for data analysis. The new averaging approaches have not yet been validated by the current regulatory assessment. Meanwhile, our BMD approach is currently used in the regulation of existing industrial chemical substances under the CSCL in Japan. In the present paper, we describe the basis of the selection statistical parameters and the model selection criteria at the BMD analysis.

MATERIALS AND METHODS

Data selection

First, we collected repeated-dose toxicity data, which form the basis for the calculation of reference dose, TDI, or ADI in the US EPA IRIS (http://www.epa.gov/IRIS/) and WHO guidelines for drinking water quality (http://www.who.int/water_sanitation_health/publications/2011/dwq_guidelines/en/). We analyzed the assessment documents for approximately 700 repeated-dose toxicity datasets; however, the original study data for most pesticides were not publicly available. Consequently, only 228 original study datasets were obtained for the repeated-dose toxicity. Among these studies, we selected studies with ≥ 4 dose groups, including the control group. Thereafter, the subjects of our investigation were 62 endpoints (31 dichotomous datasets and 31 continuous datasets) that increased or decreased in a dose-dependent manner and reached a significant level in the dose groups. Similarly, we analyzed 259 studies that were performed in Japanese safety programmes for existing chemicals. We obtained 232 endpoints, including 112 dichotomous datasets and 120 continuous datasets, from the combined repeated-dose toxicity studies with the repro-

duction/developmental toxicity screening tests (OECD TG 422) and obtained 83 dichotomous datasets from the 28-day repeated-dose toxicity studies (OECD TG 407). In total, we obtained 226 dichotomous datasets and 151 continuous datasets.

Software

The US EPA software for BMD analysis, namely, BMDS 2.1.2, was used for all derivations. The mathematical models available in BMDS are the Gamma, Logistic, Log-Logistic, Log-Probit, Multistage, Probit, Weibull, and Quantal-Linear models for dichotomous data and the Exponential, Hill, Linear, Polynomial, and Power models for continuous data. For comparison purposes, PROAST 23.2 (for dichotomous data) and 28.1 (for continuous data), which were developed by RIVM, were also used. In PROAST, 10 models (i.e., One-Stage, Two-Stage, Log-Logist, Weibull, Log-Probit, Gamma, Logistic, Probit, E2, and H2 models) and 2 models (i.e., Hill and Exponential models) can be fitted to dichotomous and continuous data, respectively.

BMD approach

A total of 226 dichotomous datasets and 151 continuous datasets were fitted to all models available in BMDS or PROAST, and the overall BMDL for each endpoint was determined under the following conditions:

(1) BMR
Dichotomous data: 5% or 10%; continuous data: one control SD from the control mean (1SD), 5%, or 10%

(2) Modeling constraints
In BMDS, the option to constraining the model parameters existed for the Gamma, Log-Logistic, Log-Probit, Multistage, and Weibull models for dichotomous data and the Hill and Power models for continuous data. Constraining these parameters is the default approach of BMDS. In PROAST, the users can select either constraining or non-constraining the parameters of dichotomous data modeling. We fitted the data to all available models under constraining and non-constraining modes.

(3) Model selection
If more than one model is successfully fitted to the data, these models must be compared with each other to select the most appropriate BMDL. First, model selection was performed as recommended by the U.S. EPA guidance (EPA, 2000; EPA, 2012), i.e., if the BMDL was estimated by successfully fitted models, and range was within a factor of three (EPA, 2000), the model with the lowest Akaike's information criterion (AIC) was selected; otherwise, the lowest BMDL was selected. The *p*-values suggested by the BMDS user guide were used for testing the goodness of fit ($p > 0.1$) and other statistics ($p > 0.1$ or 0.05). Hereinafter, we called this method the EPA method. For comparison, we also performed the model selection as recommended by the EFSA guidance (EFSA, 2009), i.e., by simply selecting the lowest BMDL. Hereinafter, we called this method the EFSA method. PROAST automatically makes a judgment on whether the model was accepted, i.e., if the model passed the goodness of fit. The critical *p*-value for accepting a model is 0.05 in PROAST (EFSA, 2009).

(4) Additional criteria
The EFSA noted that the data were not sufficiently informative to derive a POD if the ratio of BMD to BMDL estimated from the individual model was extremely large or if the BMDLs estimated from different models varied widely (EFSA, 2011). Therefore, we considered excluding the model with a BMD/BMDL ratio ≥ 10. We also examined the necessity of additional criteria when the range of BMDLs estimated from different models was wide or when the overall BMDL was substantially lower than the lowest dose tested.

Overall, the BMDLs derived under various conditions were compared with the NO(A)ELs and LO(A)ELs established in the US EPA IRIS, WHO guidelines for drinking water quality, or the Japanese safety programmes for existing chemicals for each endpoint.

RESULTS

Dichotomous

Benchmark dose response

The $BMDL_{10}$s for an extra risk of 10% and $BMDL_{05}$s for an extra risk of 5% were obtained for 219 dichotomous datasets. BMD modeling was performed using BMDS with parameter constraints. The model with a BMD/BMDL ratio ≥ 10 was excluded. Model selection was performed by the EPA method. The selected overall BMDL values were compared with the NO(A)EL and/or LO(A)EL (Fig. 1). As a result, 8.3% (15/180) of the $BMDL_{05}$ and 27% (48/180) of the $BMDL_{10}$ were higher than the NOAEL, and 100% (219/219) of the $BMDL_{05}$ and 97% (213/219) of the $BMDL_{10}$ were lower than the LOAEL. The percentage of $BMDL_{10}$, which lies between the NOAEL and LOAEL (24%; 43/180), was higher than that of $BMDL_{05}$ (8.3%; 15/180). This trend was the same as PROAST ($BMDL_{10}$; 25.6%; 41/160 and $BMDL_{05}$; 7.5%; 12/160) under the same conditions.

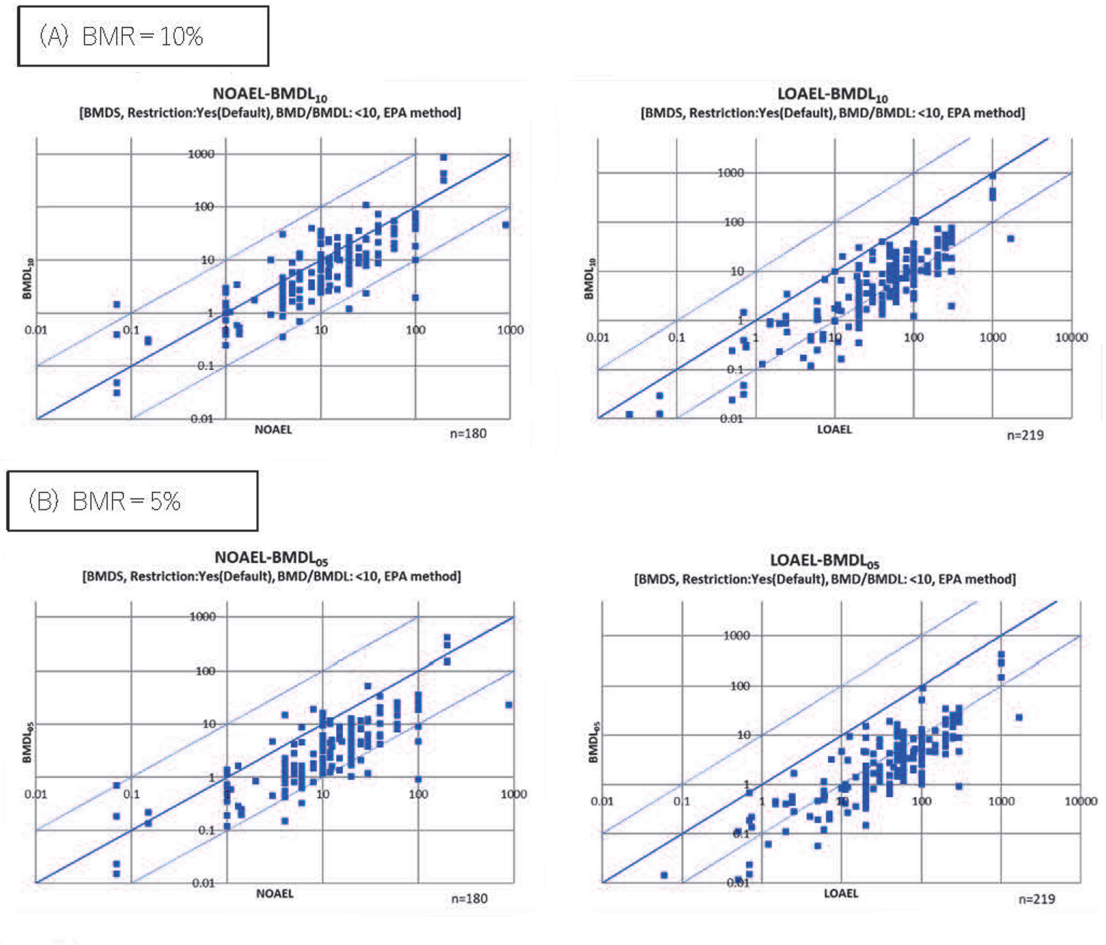

Fig. 1. Comparison of the $BMDL_{10}$ (A) or the $BMDL_{05}$ (B) with the NOAEL or the LOAEL for dichotomous data. BMD modeling was performed using the BMDS with parameter constraints. The model with BMD/BMDL ratio of ≥ 10 was excluded. Model selection was performed by the EPA method.

Constraints of model parameters

To compare the BMDLs selected from EPA default models (parameter-constraining) with those selected from parameter-constraining and non-constraining models, the data were fitted by Gamma, Log-Logistic, Log-Probit, Multistage, and Weibull models with parameter constraints in BMDS. Fig. 2 illustrates the comparison of the $BMDL_{10}$ selected from the values calculated by constraining and non-constraining models with the NOAEL or LOAEL. The $BMDL_{10}$s for 74 datasets (34%) dropped to a lower value, and the $BMDL_{10}$s of 23 datasets (11%) were less than half by adding non-constraining models for the selection of the overall $BMDL_{10}$. There were a few cases with increased values, but most of the other $BMDL_{10}$s were unchanged by the addition of non-constraining models. This trend was the same as PROAST (41% of data decreased by the addition of non-constraining models). Data with an overall $BMDL_{10}$ that lay between NOAEL and LOAEL (NOAEL < $BMDL_{10}$ < LOAEL) were compared for the BMDL estimated from constraining models only and that from constraining and non-constraining models. By adding the non-constraining models, it was slightly decreased as follows: PROAST (from 45 to 42 datasets) and BMDS (from 43 to 35 datasets).

Model comparison and BMDL selection

The $BMDL_{10}$ values selected by the EFSA method and the EPA method were compared. BMD modeling was performed using BMDS with parameter constraints. The model with a BMD/BMDL ratio ≥ 10 was excluded. Only 14% (31/216) of the data resulted in different values because the range of the $BMDL_{10}$s (maximum $BMDL_{10}$/

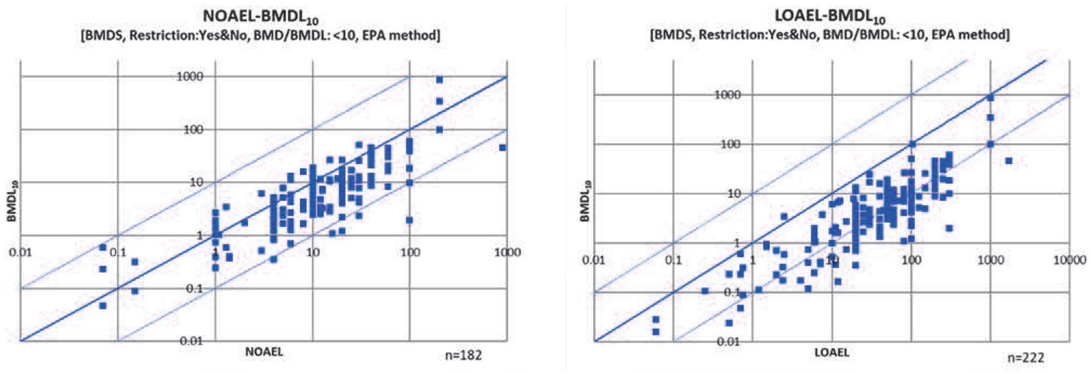

Fig. 2. Comparison of the $BMDL_{10}$ selected from the values calculated by constraining and non-constraining models with the NOAEL or the LOAEL for dichotomous data. BMD modeling was performed using the BMDS with or without parameter constraints. The model with BMD/BMDL ratio of ≥ 10 was excluded. Model selection was performed by the EPA method.

minimum $BMDL_{10}$) calculated from the accepted models was > 3 for most data. For these data, the model with the lowest $BMDL_{10}$ value was selected by the EPA method, which is the same as the EFSA method. The differences in the $BMDL_{10}$s selected according to the EFSA method and EPA method were all < 3 times.

Additional criteria

The $BMD_{10}/BMDL_{10}$ ratio or the lowest dose/$BMDL_{10}$ ratio reflects the uncertainty in the BMD estimation. Fig. 3A shows the comparison of the selected $BMDL_{10}$s by PROAST with non-constraining models and the corresponding NOAEL or LOAEL. When the models were excluded by the criterion of $BMD_{10}/BMDL_{10}$ ratio ≥ 10 (Fig. 3B), the distribution of $BMDL_{10}$ values converged. Fig. 4 shows the comparison of the highest $BMDL_{10}$/lowest $BMDL_{10}$ ratio and $BMD_{10}/BMDL_{10}$ ratio. Correlations were observed between these two ratios. Thus, the outlier $BMDL_{10}$ was excluded by rejecting the models with a $BMD_{10}/BMDL_{10}$ ratio ≥ 10. We also examined the necessity of an additional criterion for a condition that overall BMDL is substantially lower than the lowest dose tested. There were 17 out of 226 datasets whose lowest dose/$BMDL_{10}$ ratio ≥ 100, but these data were also excluded by rejecting the models with $BMD_{10}/BMDL_{10}$ ratio ≥ 10.

Software

To compare two software tools, $BMDL_{10}$s from PROAST and BMDS were selected by using the same methodology, namely, the optional model parameters were both constrained and non-constrained. Model selection was performed by the EFSA method. Models with a $BMD_{10}/BMDL_{10}$ ratio ≥ 10 or the lowest dose/$BMDL_{10}$ ratio ≥ 100 were excluded. The comparison results of the $BMDL_{10}$s estimated using BMDS and PROAST under the same conditions show that the value was almost identical (Fig. 5).

Continuous data

Benchmark dose response

The BMDL estimation was performed with three BMR settings (1SD, 10%, and 5%). We obtained $BMDL_{1SD}$s for 104 endpoints, $BMDL_{10}$s for 98 endpoints, and $BMDL_{05}$s for 102 endpoints for BMDS. These estimates were compared with the corresponding NO(A)EL and/or LO(A)EL (Fig. 6). Fifty-two percent of $BMDL_{1SD}$s, 50% of $BMDL_{10}$s, and 30% of $BMDL_{05}$s were within the range between the NOAEL and LOAEL. One SD as a BMR seemed like a good selection in PROAST ($BMDL_{1SD,}$ $BMDL_{10,}$ and $BMDL_{05}$) meet condition of NOAEL< BMDL < LOAEL were 43%, 33%, and 34.2%, respectively). The data were analyzed by using four endpoints (i.e., bodyweight, blood biochemistry, hematology, and others), but no specific tendency was observed among them.

BMD modeling was performed using BMDS with parameter constraints. Model selection was performed by the EPA method. The model with $BMD_{10}/BMDL_{10}$ ratio ≥ 10 was excluded.

Constraints of model parameters

We fitted the data to the Hill and Power models with parameter constraints in BMDS. The overall $BMDL_{1SD}$s selected from the values estimated from the constraining and non-constraining models were compared with the NO(A)EL or LO(A)EL (Fig. 7). In comparison with

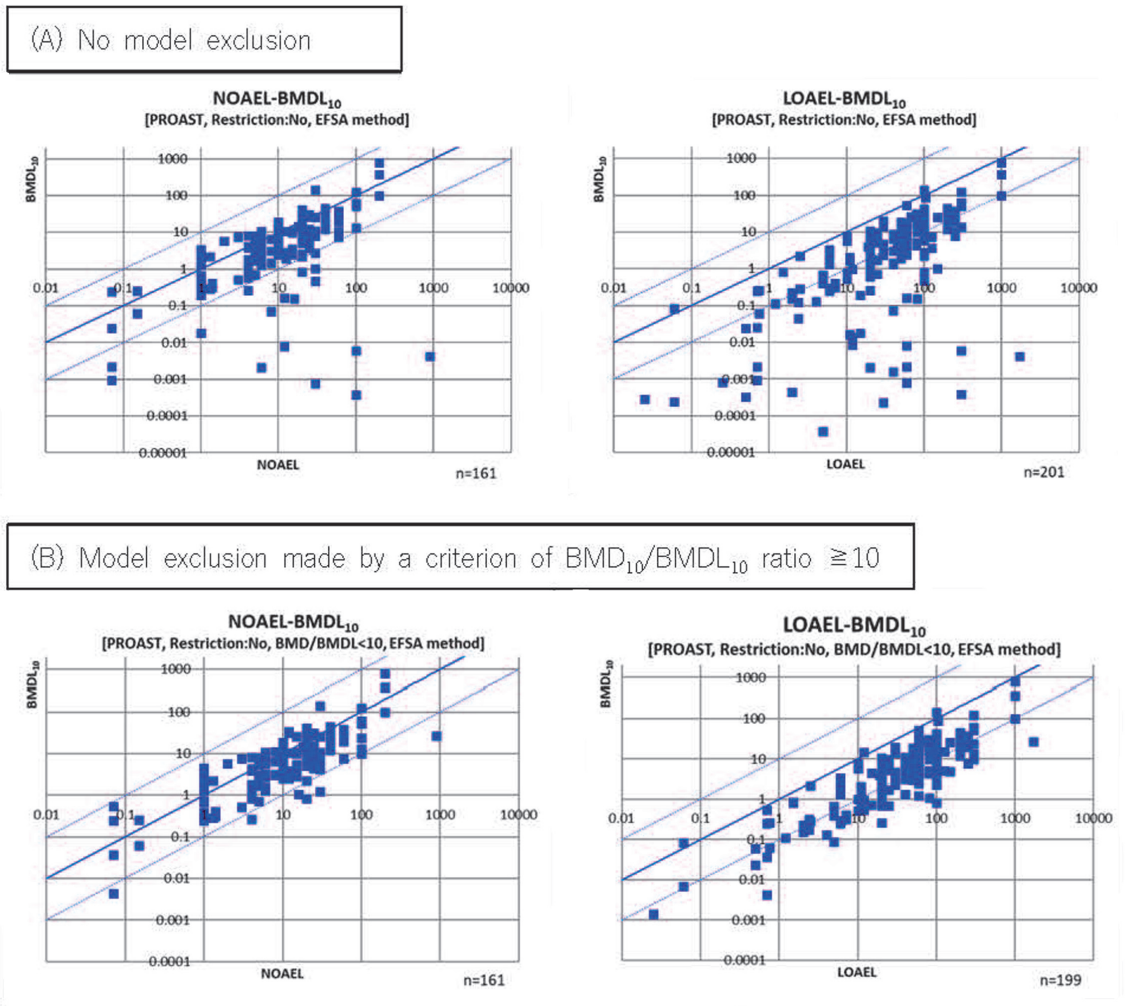

Fig. 3. Comparison of the $BMDL_{10}$ selected from the values calculated by constraining models with the NOAEL or the LOAEL for dichotomous data. BMD modeling was performed using the PROAST without parameter constraints, and model selection was performed by the EPA method (A). BMD modeling and model selection were performed by the same methodology, but the model with $BMD_{10}/BMDL_{10}$ ratio of ≥ 10 was excluded (B).

the results of the $BMDL_{1SD}$s with parameter constraints (Fig. 6A) and those with or without parameter constraints (Fig. 7), the latter values were slightly low. More specifically, the selected overall $BMDL_{1SD}$ for 39 of the 104 endpoints (38%) decreased to a lower value by adding the values estimated from the non-constraining models. The $BMDL_{1SD}$s for 24 endpoints (26%) were less than half of the $BMDL_{1SD}$ selected from the values estimated under the default condition (i.e., all of the optional model parameters were constrained). By adding non-constraining models, data in a range between NOAEL and LOAEL decreased, such as dichotomous data (from 53 to 39 datasets).

Model comparison and BMDL selection

We compared the overall $BMDL_{1SD}$s selected according to the EPA method and the EFSA method by using BMDS. Unlike the $BMDL_{10}$s for dichotomous data, the range of $BMDL_{1SD}$s (maximum $BMDL_{1SD}$/minimum $BMDL_{1SD}$) estimated from the accepted models were mostly (80/104 datasets) less than three; however, the EPA and the EFSA methods led to different $BMDL_{1SD}$ for only 36 datasets, and the differences in the $BMDL_{10}$s selected according the EFSA method and the EPA method were all less than three.

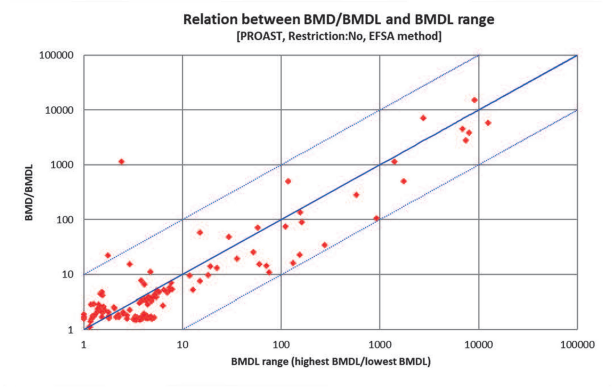

Fig. 4. Comparison of the ratio of highest $BMDL_{10}$/lowest $BMDL_{10}$ and the ratio of BMD_{10}/$BMDL_{10}$ for dichotomous data. BMD modeling was performed using the PROAST without parameter constraints, and model selection was performed by the EFSA method.

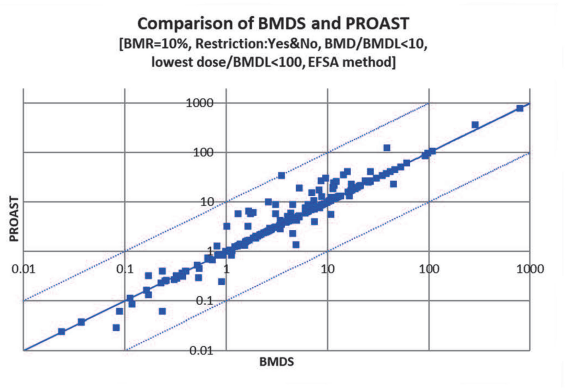

Fig. 5. Comparison of the $BMDL_{10}$ calculated by the BMDS and the PROAST for dichotomous data. BMD modeling was performed with and without parameter constraints. Model selection was performed by the EFSA method. The model with BMD_{10}/$BMDL_{10}$ ratio of ≥ 10 or with the lowest dose/$BMDL_{10}$ ratio of ≥ 100 was excluded.

Additional criteria

Fig. 8A shows the comparison of the selected $BMDL_{1SD}$s by PROAST with non-constraining models and the corresponding NOAEL or LOAEL. When the models were rejected by the criterion of BMD_{1SD} / $BMDL_{1SD}$ ratio ≥ 10, the models distributed on the lower side were excluded (Fig. 8A). Given that $BMDL_{1SDS}$ were widely distributed on the lower side, an additional criterion was considered. Five models were further excluded by rejecting the models with the lowest dose/$BMDL_{1SD}$ ratio ≥ 100.

Software

The $BMDL_{1SD}$s were also calculated by using PROAST. BMD modeling was performed with both constraining and non-constraining parameters. Model selection was performed by the EFSA method. The model with a BMD_{1SD}/ $BMDL_{1SD}$ ratio ≥ 10 or the lowest dose/$BMDL_{1SD}$ ratio ≥ 100 was excluded. The overall $BMDL_{1SD}$s selected from the values estimated by BMDS and PROAST under the same condition were generally similar (Fig. 9).

DISCUSSION

On the basis of the analysis of 226 dichotomous dose-response datasets, we consider it appropriate to select 10% extra risk as the first choice of BMR because the percentage of $BMDL_{10}$, which lies between the NOAEL and LOAEL, was higher than that of $BMDL_{05}$. This default value is the same as the EFSA and the EPA recommendations (EFSA, 2009; EPA, 2012). For continuous data, we analyzed 151 data, and a BMR defined as a change of the mean by one standard deviation of the control group (1SD) seemed to be the best selection. The percentage of $BMDL_{10}$ that was between the NOAEL and LOAEL was 1SD > 10% > 5% (BMDS). The US EPA also recommends reporting the estimate of the BMD corresponding to the BMR of 1SD for continuous data (EPA, 2012), whereas the EFSA recommends the BMR of 5% change (EFSA, 2009). A BMD that is associated with the BMR of 1SD may depend on group variations. When the variation of the group is small, it is recommended to consider a 10% BMR as a second choice. These recommendations are described in the DRA-BMDS guidance.

In this study, the use of modeling constraints was examined. The US EPA recommended that the model parameters should be constrained (EPA, 2012), and it is the default in BMDS. The US EPA noted that the power parameters should be generally constrained to ≥ 1. However, when the observed data suggest a superliner dose-response relationship, there is a need to investigate whether unconstrained models or models that contain an asymptote term can support reasonable BMD/BMDL values (EPA, 2012). The EFSA guidance also mentioned the constraining of model parameters, but no further advice was provided. There is no mention of modeling constraints in the IPCS guidance (IPCS, 2009). We compared the overall BMDLs derived by constraining models and those derived by both constraining and non-constraining models. Approximately one-third of the overall BMDLs

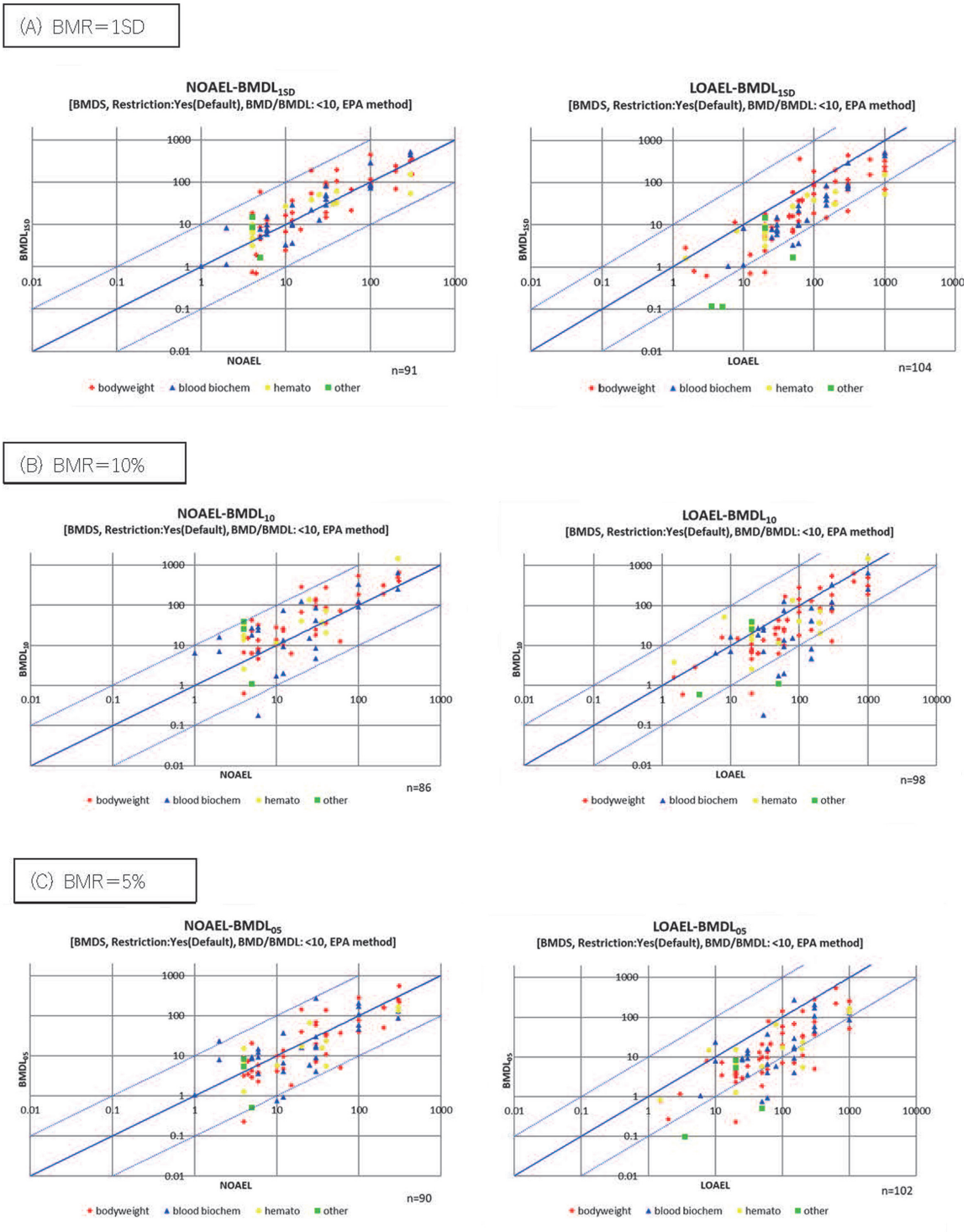

Fig. 6. Comparison of the $BMDL_{1SD}$ (A), the $BMDL_{10}$ (B) or the $BMDL_{05}$ (C) with the NOAEL or the LOAEL for continuous data. BMD modeling was performed using the BMDS with parameter constraints. Model selection was performed by the EPA method. The model with $BMD_{10}/BMDL_{10}$ ratio of ≥ 10 was excluded.

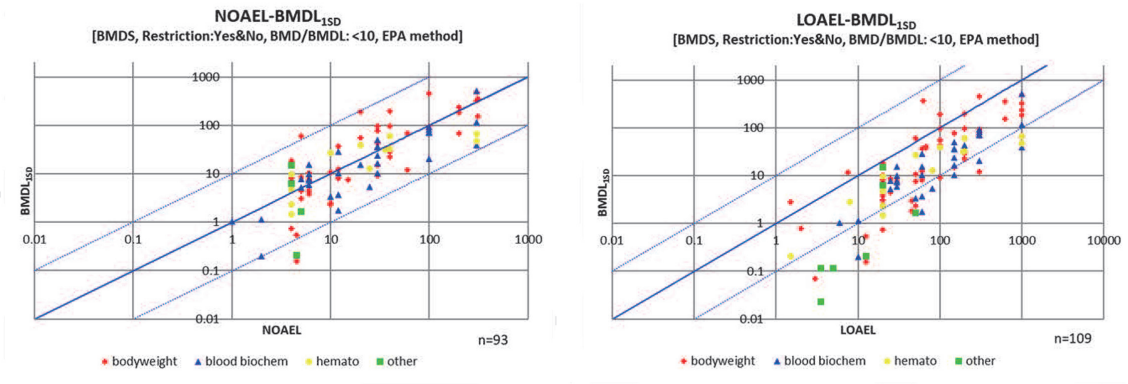

Fig. 7. Comparison of the $BMDL_{1SD}$ selected from the values calculated by constraining and non-constraining models with the NOAEL or the LOAEL for continuous data. BMD modeling was performed using the BMDS with and without parameter constraints. Model selection was performed by the EPA method. The model with $BMD_{1SD}/BMDL_{1SD}$ ratio ≥ 10 was excluded.

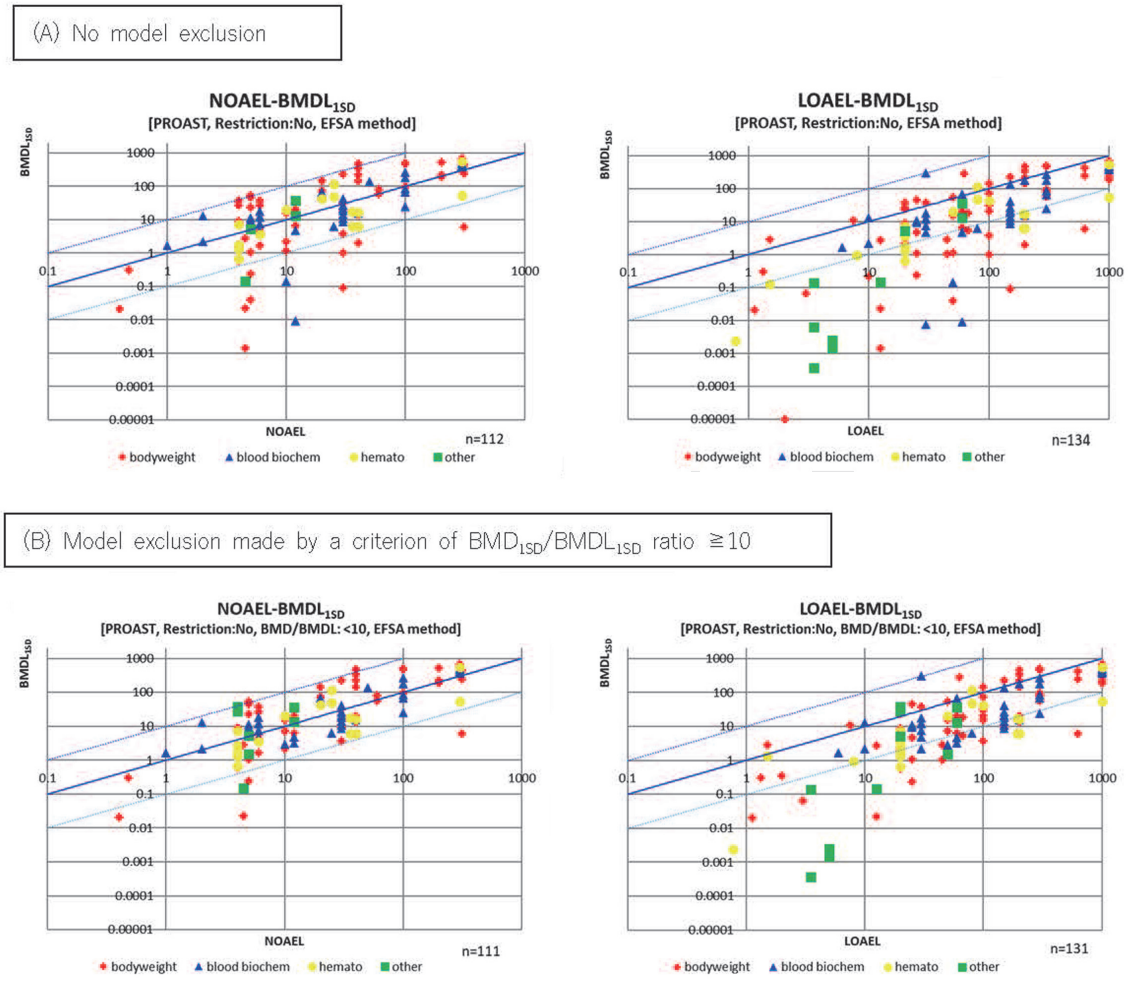

Fig. 8. Comparison of the $BMDL_{1SD}$ selected from the values calculated by non-constraining models with the NOAEL or LOAEL for continuous data. BMD modeling was performed using the PROAST with parameter constraints. Model selection was performed by the EFSA method (A). The model with $BMD_{1SD}/BMDL_{1SD}$ ratio ≥ 10 was excluded (B)

became lower when the non-constraining models were added for both dichotomous and continuous data. We considered it better to run both the constraining and non-constraining models for conservative risk assessment unless scientifically appropriate justification is not found. These recommendations are described in the DRA-BMDS guidance.

If more than one model is successfully fitted to the data, these models must be compared with each other to select the most appropriate BMDL (i.e., the overall BMDL). The goodness-of-fit statistic (*p*-value) was used in BMDS and PROAST as an important criterion for model adequacy. Given that the *p*-value is not designed to compare different models, it is used only as a cutoff criterion. The likelihood ratio test can be used to judge whether the goodness of fit is improved by the addition of a parameter within the family of dose-response models. This was noted by the IPCS, US EPA, and EFSA (EFSA, 2011; EPA, 2012; IPCS, 2009). On the contrary, AIC can be used to compare models from different families by using a similar fitting method, and the US EPA guidance recommended the use of AIC when the models estimated sufficiently close BMDL values (EPA, 2012). The IPCS guidance noted that the models in the same family or models with the same assumptions about the underlying probability distributions can be compared by AIC (IPCS, 2009). Meanwhile, the EFSA guidance recommend the selection of the lowest BMDL until more advanced methods are fully developed and validated (EFSA, 2009). Whether it is a good way to compare various models by the AIC remains controversial. Therefore, we also considered that a model with the lowest BMDL should be selected for conservative risk assessment. These recommendations are described in the DRA-BMDS guidance.

Considering the uncertainty, the use of additional criteria such as the BMD/BMDL ratio and the lowest dose/BMDL ratio was investigated in this study. When the models were rejected by using criteria such as BMD/BMDL ratio ≥ 10 and lowest dose/BMDL ratio ≥ 100, BMDL distribution converged, and the outlier $BMDL_{10}$ was excluded. Given that we decided to select the low-

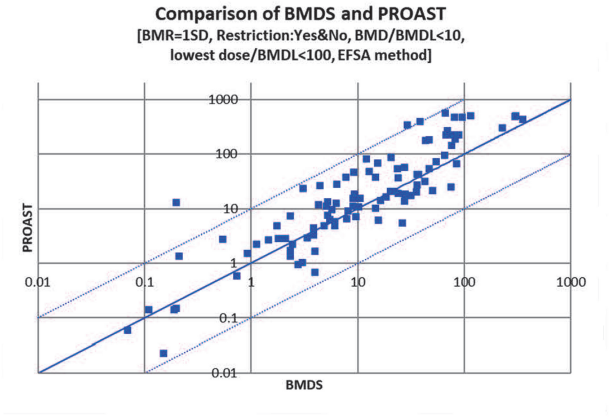

Fig. 9. Comparison of the $BMDL_{1SD}$ estimated with the BMDS and PROAST for continuous data. BMD modeling was performed with and without parameter constraints. Model selection was performed by the EFSA method. The model with $BMD_{1SD}/BMDL_{1SD}$ ratio ≥ 10 or the lowest dose level/$BMDL_{1SD}$ ratio ≥ 100 was excluded.

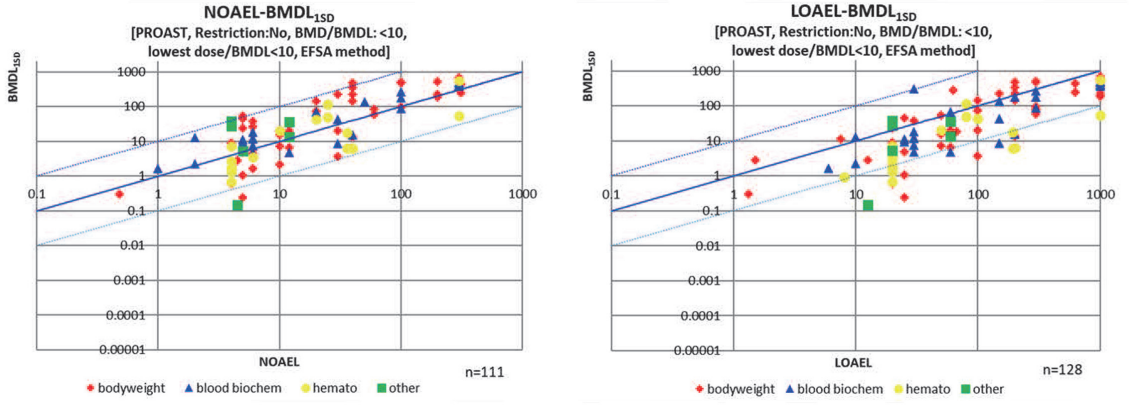

Fig. 10. Comparison of the $BMDL_{1SD}$ selected from the values calculated by non-constraining models with the NOAEL or LOAEL for continuous data. BMD modeling was performed using the PROAST with parameter constraints. Model selection was performed by the EPA method. The model with BMD_{1SD} /$BMDL_{1SD}$ ratio ≥ 10, or lowest dose/$BMDL_{1SD}$ ratio ≥ 10 was excluded.

est BMDL from all accepted models for conservative risk assessment, these criteria are important to avoid lower values of BMDL derivation without constraints. It is also recommended to see the graphical display for confirming the fitting and dose-response curve. The models with a biologically unrealistic dose-response curve, e.g.) non-monotonic curve for multistage models, should be rejected. These recommendations have been described in the DRA-BMDS guidance, but they have been partially modified recently, as discussed below.

On the basis of the experience of employing the BMD analysis for the evaluation of Priority Assessment Chemical Substances with the DRA-BMDS guidance, we encountered a problem that an extremely low health-based guidance value was derived from a low BMDL when the lowest dose/BMDL ratio was nearly 100. An uncertainty factor of 10 has conventionally been applied to a LOAEL to estimate the NOAEL. Therefore, the lowest dose/BMDL ratio of 10 seemed to be a better cut-off value to consider uncertainty. The BMDLs derived from most of the appropriate models being >10 times lower than the LOAEL indicate that the dataset is inadequate for the BMD analysis. Fig. 10 depicts the comparison of the $BMDL_{1SD}$ selected from the model with a BMD_{1SD} /$BMDL_{1SD}$ ratio < 10 or lowest dose/$BMDL_{1SD}$ ratio < 10 with the NOAEL or LOAEL. The distribution of $BMDL_{1SD}$ was well-converged for comparison (Fig. 8A). For the dichotomous data, 20 models were rejected by the criterion of lowest dose/BMDL ratio ≥ 10 (data not shown). The DRA-BMDS guidance was then updated from "the models with the lowest dose/BMDL ratio ≥ 100 be excluded" to "the models with lowest dose/BMDL ratio ≥ 10 be excluded." The EFSA guidance was last updated in 2017, and the upper bound (BMDU)/lower bound (BMDL) ratio, which reflects the uncertainty of the BMD estimate, is discussed (Hardy *et al*., 2017). Readers may want to considere the BMDU/BMDL ratio for model selection on a case by case basis.

In our analysis, both of BMDS and PROAST provided almost identical value in their calculated BMDL, and the reliability of the calculated BMDS values were supported by comparison with experimental NOAEL; thus, we concluded that either software tool can be used for risk assessment. PROAST was originally developed as a package in R, and knowledge of R software was required to use PROAST. Therefore, BMDS was selected as the default software for the evaluation of the Priority Assessment Chemical Substances under CSCL. It is noted that currently web applications of PROAST are also available at https://efsa.openanalytics.eu/ and https://proastweb.rivm.nl/. To date, approximately 10 Priority Assessment Chemical Substances were evaluated with the BMD approach. Health-based guidance values for Priority Assessment Chemical Substance we established using the DRA-BMDS guidance are comparable those established by other international institutes.

BMDS 3.0 recently released the Bayesian dichotomous model averaging feature (EPA, 2018). However, the EPA BMD technical guidance that includes the model averaging has not yet been published. Moreover, these new feature bases on the different statistical prediction algorithm could yield different results, and there is no information about the manner of treating different BMDLs under BMDS software. The new EFSA guidance recommends model averaging as the preferred method for calculating the BMD confidence interval, and model averaging is now an option in PROAST. Furthermore, the AIC was introduced to characterize the relative goodness of fit of different mathematical models to a dose-response data set by the new guidance of the EFSA (Hardy *et al*., 2017).

Presently, a new investigation is underway to review and search for a better methodology of the BMD approach, including model averaging. The findings of our new investigation may be reflected in the next update of the DRA-BMDS guidance. However, if the model averaging method is inappropriate for data analysis, we need to select a single model for BMDL estimation (EFSA, 2017). Moreover, the new averaging approaches remain to be validated by the current regulatory assessment. We also believe that our DRA-BMDS guidance as a single model selection procedure can assist risk assessors in the selection of the model because they should select the BMDL on a case by case basis by considering all available information. Lastly, the BMD should be derived as conservatively as possible within the range of practical management under the CSCL, because the scope of the CSCL is to prevent environmental pollution by industrial chemical substances under the precautionary principle.

In conclusion, we summarize the current DRA-BMDS guidance as follows:

- Selecting an extra risk of 10% (dichotomous data) or a level change of 1SD (continuous data) as the default BMR
- Running all models without or with parameter constraints
- Using *p*-values suggested by BMDS user guide as a default for model selection
- Conducting visual inspection of plots for model selection
- Selecting the model that calculated the lowest BMDL but excluding the model that estimated the BMD/BMDL ratio ≥ 10 or lowest dose/BMDL ratio ≥ 10.

ACKNOWLEDGMENTS

This study was supported by a grant from the Food Safety Commission, Cabinet Office, Government of Japan (Research Program for Risk Assessment Study on Food Safety, No 1007).

Conflict of interest---- The authors declare that there is no conflict of interest.

REFERENCES

EFSA (2009): SCIENTIFIC OPINION: Use of the benchmark dose approach in risk assessment1 Guidance of the Scientific Committee (Question No EFSA-Q-2005-232) Adopted on 26 May 2009

EFSA (2011): TECHNICAL REPORT: Use of BMDS and PROAST software packages by EFSA Scientific Panels and Units for applying the Benchmark Dose (BMD) approach in risk assessment.

EPA (2000): Benchmark dose technical guidance document, External Review Draft, EPA/630/R-00/001. Risk Assessment Forum, US Environmental Protection Agency, Washington, DC 20460, October 2000.

EPA (2012): Benchmark dose technical guidance, EPA/100/R-12/001, Risk Assessment Forum, US Environmental Protection Agency, Washington DC, June 2012.

EPA (2018): Benchmark Dose Software (BMDS) Versuib 3.0 User Guide.

Hardy, A., Benford, D., Halldorsson, T., *et al*. (2017): Update: use of the benchmark dose approach in risk assessment. EFSA J., **15**, e04658.

IPCS (2009): Principles for modelling dose-response for the risk assessment of chemicals, Environmental Health Criteria 239, International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland.

Oberg, M. (2010): Benchmark dose approaches in chemical health risk assessment in relation to number and distress of laboratory animals. Regul. Toxicol. Pharmacol., **58**, 451-454.

Sand, S., Victorin, K. and Filipsson, A.F. (2008): The current state of knowledge on the use of the benchmark dose concept in risk assessment. J. Appl. Toxicol., **28**, 405-421.

Slob, W. (2014a): Benchmark dose and the three Rs. Part I. Getting more information from the same number of animals. Crit. Rev. Toxicol., **44**, 557-567.

Slob, W. (2014b): Benchmark dose and the three Rs. Part II. Consequences for study design and animal use. Crit. Rev. Toxicol., **44**, 568-580.