HOKKAIDO UNIVERSITY

Title	Simulation-based assessment of model selection criteria during the application of benchmark dose method to quantal response data [an abstract of dissertation and a summary of dissertation review]
Author(s)	吉井, 啓太
Citation	北海道大学. 博士(医学) 甲第15239号
Issue Date	2022-12-26
Doc URL	http://hdl.handle.net/2115/88261
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	theses (doctoral - abstract and summary of review)
Note	配架番号：2743
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	YOSHII_Keita_abstract.pdf (論文内容の要旨)

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学位論文内容の要旨

博士の専攻分野の名称　博士（医　学）　　氏 名　吉 井　啓 太

学 位 論 文 題 名

Simulation-based assessment of model selection criteria during the application of benchmark dose method to quantal response data

(ベンチマークドーズ法の用量反応関係への適用にかかるモデル選択基準に関するシミュレーション研究)

【Introduction】 To determine the reference dose of the chemical substances, such as the acceptable daily intake, NOAEL method has been popular in use. However, the determination of NOAEL imposes a serious statistical limitation that involves non-negligible sampling errors. The benchmark dose (BMD) method as an alternative can address the problems because it can account for the response data across all aspects of dose-response data considering the underlying biological uncertainties. This approach determines the threshold dose BMD associated to the benchmark response (BMR), a specified level of response from the background response, by fitting various statistical models to the dose-response curve. The benchmark dose lower bound (BMDL), which is the lower (one-sided) limit of the 95% confidence interval (CI) of BMD, can yield a point of departure that is empirically comparable to that based on NOAEL. However, no uniform guideline of technical use of the BMD method has been established worldwide due to technical issues in the usage of the BMD methods. To employ the BMD method in toxicological risk assessment, it is critical to understand how BMDL for reference dose calculation is selected following statistical fitting procedures of multiple mathematical models. Although several technical problems exist, we believe that the biggest obstacle to the wide application of the BMD method in various governmental settings is the lack of uniform guidelines that specify the steps required to scrutinize fitting results and identify a single BMDL value for determining the acceptable daily intake. Among the remaining technical issues, model selection methods including model averaging, as well as model exclusion criteria, have been discussed and developed in previous articles, but very limited number of them were verified. While all issues surrounding the use of the BMD method for quantal response data cannot be fully and immediately resolved, a simulation-based evaluation might help to identify a possible well-performing pathway of model exclusion and selection. To support the formulation of technical guidelines for risk assessment practices for food safety in Japan, we conducted a simulation study to compare the performance of each and various combinations of model exclusion and selection criteria, as applied to three qualitatively different types of quantal response datasets.

【Materials and Methods】 The BMD method has been employed using simulation-based evaluation of model exclusion and selection processes by comparing validity, reliability, and other model performance parameters. Three different empirical datasets for different chemical substances were analyzed for the assessment, each having qualitatively different characteristics of the dose-response pattern. Briefly, our analysis goes by: (i) identification of a “reference model” for each dataset by AIC (Akaike Information Criteria), (ii) generation of a total of 1,000 simulated datasets (each dataset includes fittings by 9 individual model) from the “reference model”, (iii) application of model exclusion criteria if available, (iv) application of one of the model selection criteria including methods using model averaging, and select or calculate one of the representative BMDL value from each dataset, and (v) BMDL values were evaluated in two aspects, the validity and the reliability. First, for each quantal dataset, we first identified the best-fit model by selecting the model with the lowest AIC value out of the total of nine

different distributions that consist of 2–4 unknown parameters. During the simulations, we regarded the identified best model for each chemical substance as the "reference model". Such a true model is accompanied by the known lower bound of the benchmark dose with response level at 10% (unbiased $BMDL_{10}$) as derived from the maximum likelihood estimates of the parameters. The statistical estimation was conducted using the maximum likelihood method, and the likelihood function was defined under the assumption that the quantal response data at a given dose follows a binomial distribution. Computation of the 95% CI, including BMDL and BMD upper bound (BMDU) (i.e. one-sided upper 95% CI of BMD) was conducted using the Monte Carlo algorithm. We used validity and reliability as evaluators of model selection criteria out of 1000 simulated values: Briefly, validity is the proportion that the simulated $BMDL_{10}$ value is the dose lower than the known benchmark dose with response level at 10% (unbiased BMD_{10}), and reliability is the relative distance between the simulated $BMDL_{10}$ value and the unbiased $BMDL_{10}$ value. We also assessed calculability of BMDL, the proportion of simulated datasets that yielded convergence, and the proportion of the same statistical model, out of the nine candidate models, that was recovered to be identical to the original. We considered a total of four possible model exclusion criteria and six possible model selection criteria, including criteria using model averaging. Avoiding excessive combinations of the two, we tested and compared a total of 18 possible combinations.

【Results】 The best performing criteria of model exclusion and selection were different across the different datasets. Generally, however, one of the model selection criteria, model averaging over the three models with the lowest three AIC values (MA-3) did not produce the worst performance in any of the three quantal datasets we used, and MA-3 without model exclusion produced the best results among the model averaging. Model exclusion including the use of the Kolmogorov-Smirnov test in advance of model selection did not necessarily improve the validity and reliability of the models.

【Discussion】 As part of the technical assessment for possible improvements in the guidelines, we conducted a simulation-based experiment to assess the model exclusion and selection process by comparing the validity, reliability, and other model performance indicators using all possible combinations of model exclusion and selection criteria. There are two take-home messages. First, although we did not identify the best exclusion and selection criteria for the qualitatively differently distributed datasets, we have shown that model averaging over three models with the lowest three AIC values (MA-3) did not yield the worst result, and MA-3 without prior model exclusion produced the best results among all the model averaging results. Second, we found that model exclusion using the KS test and the ratios of BMD or BMDU to BMDL did not necessarily yield better validity and reliability than non-exclusion. Particularly in model averaging, we found that all the model averaging options that we tested performed well overall. Furthermore, we found that averaging over some of the models might yield a better performance than averaging over all converged models, considering that the uncertainties of well-fitted models might be far smaller than those of badly fitted models. While numerous technical issues have yet to be explored in applying the BMD methods to risk assessment, we concluded that MA-3 can be considered the best guiding option to derive the reference dose when the guidelines are expected to specify a single model exclusion and selection method.

【Conclusion】 We conducted a simulation-based experiment to assess the model exclusion and selection process by comparing the validity, reliability, and other model performance indicators using all possible combinations of model exclusion and selection criteria. If a uniform methodological suggestion for the guideline is required to choose the best performing model for exclusion and selection, our results indicate that using MA-3 is the recommended option whenever applicable.