NOVARTIS

Development IIS

Efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty

Request for CHMP Qualification Opinion
Response to Questions dated 11 June 2013

Author(s): Bretz F, Bornkamp B, Pinheiro J

Document type: Response to HA questions

Document status: Final

Release date: 28 June 2013

Table of contents

List of figures

1 Introduction

This response document addresses the questions raised by the Scientific Advice Working Party (SAWP) on 11 June 2013 in the context of the Qualification procedure EMEA/H/SAB/030/1/Q/2013 for the qualification of the statistical MCP-Mod method *"Efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty"*.

2 Response to Questions

2.1 Question 1

In order to further quantify gains in efficiency for MCP-Mod in relation to 'traditional' approaches based on pairwise comparisons it may also be of interest, if feasible, to extend the simulations conducted to compare.

    a. an optimised ANOVA approach based on a fixed sample size (e.g. n=150, 250 as per the simulation exercise) versus an optimised MCP-Mod approach based on the same fixed sample size. ('Optimised' would need to be defined)

    b. a traditional ANOVA approach based on a fixed sample size (e.g. n=150, 250 as per the simulation exercise, 2 dose levels plus placebo) verses an MCP-Mod approach based on the same fixed sample size but optimal number of dose levels.

Answer

The answers to Questions 1a and 1b will be presented jointly.

To further quantify gains in efficiency for MCP-Mod in relation to ANOVA-type approaches, additional simulations were conducted. The simulation assumptions were identical to those used in Section 3.2.3.2 of the Qualification Request document (and the PhRMA ADRS working group simulations, see Bornkamp et al. 2007). As requested we departed from the original assumptions in the number and location of doses for some of the investigated methods as described below in more detail. The same performance metrics as in the previous simulations were used.

2.1.1 ANOVA approaches

Optimized ANOVA approach (as referred to in Question 1a)

We defined "optimized" with respect to the employed design, that is, the doses under investigation and the sample size allocations at the doses. The main objective of traditional ANOVA approaches is to identify a significant pairwise comparison. For a known dose response relationship the optimal design, in terms of maximizing power, would allocate 50% of the patients on placebo and 50% on the dose with the highest effect. This approach is impractical for various reasons (among others the dose with highest effect might not be known before trial start). Thus we investigated two designs with 4 and 8 equally spaced active doses (subsequently named opt-ANOVA4, and opt-ANOVA8), with the same dose levels as

used in the previous simulations. However, the square-root allocation rule from Dunnett (1955) was used to determine the number of patients per treatment group. This allocation minimises the variance for the pairwise comparisons of active doses versus placebo. According to this rule the sample size for the placebo group should be selected slightly larger than in the other groups, namely proportional to the square root of k, where k is the number of active doses.

Traditional ANOVA approach (as referred to in Question 1b)

This approach is identical to the optimized ANOVA approach described above, except that 2 active dose levels are used and equal sample sizes are employed across all 3 groups. Since the performance of this traditional ANOVA approach is expected to depend strongly on the underlying true dose response curve and thus the location of the lower dose level, we investigated three different design options, namely selecting the lower of the two active doses as 2 (low), 4 (middle) and 6 (high). The three options are denoted as trad-ANOVA-low, trad-ANOVA-mid and trad-ANOVA-high, respectively.

For both the "optimized" and the "traditional" ANOVA methods, a dose response signal was declared if at least one dose was significantly different from placebo. The minimum effective dose (MED) was estimated by selecting the smallest dose with an observed point estimate of at least 1.3 units better than the placebo response and statistically significantly different from placebo. For estimating the dose response relationship, we used cubic spline interpolation of the observed means (i.e. using the option "fmm" in the R function spline), for both ANOVA approaches. Note that with this we deviated slightly from the ANOVA approach used in the simulations for Section 3.2.3.2 of the Qualification Request as well as Bornkamp et al. (2007), where an actual dose response model was fitted for the ANOVA approach to assess the dose response estimation metrics. We deviated here because the traditional ANOVA method with only two active doses does not allow fitting dose response models.

2.1.2   Optimal MCP-Mod (as referred to in Questions 1a and 1b)

Optimizing the MCP-Mod procedure could be done with respect to different objectives such as power of the MCP part, precision of MED estimation, or dose response estimation. Here, we focused on determining the dose response relationship and optimized the MCP-Mod procedure using D-optimality. The candidate model set included logistic, linear, quadratic and Emax models. We then calculated a Bayesian D-optimal design with equal prior weights using the DoseFinding R package (Bornkamp et al. 2013), which – after rounding – assigned 1/4, 1/6, 1/6, 1/6, and 1/4 of the patients to the doses 0, 0.54, 3.2, 4.8, and 8, respectively. Note that this design is robust in the sense that it is optimized simultaneously for all four, quite diverse dose response shapes and is subsequently abbreviated as opt-MCPMod. If in practice more prior information is available, the design could be further optimized/fine-tuned to specific scenarios.

2.1.3   Simulation Results

In all cases the function rndDesign from the DoseFinding package (implementing the algorithm by Pukelsheim and Rieder, 1992) will be used to round the ANOVA and MCP-Mod designs to integer samples sizes matching exactly 150 and 250. We used the DoseFinding package version 0.9-6 and R 2.15.2 to run 10'000 clinical trial simulations for each scenario.

Figure 2-1   Power to detect dose response
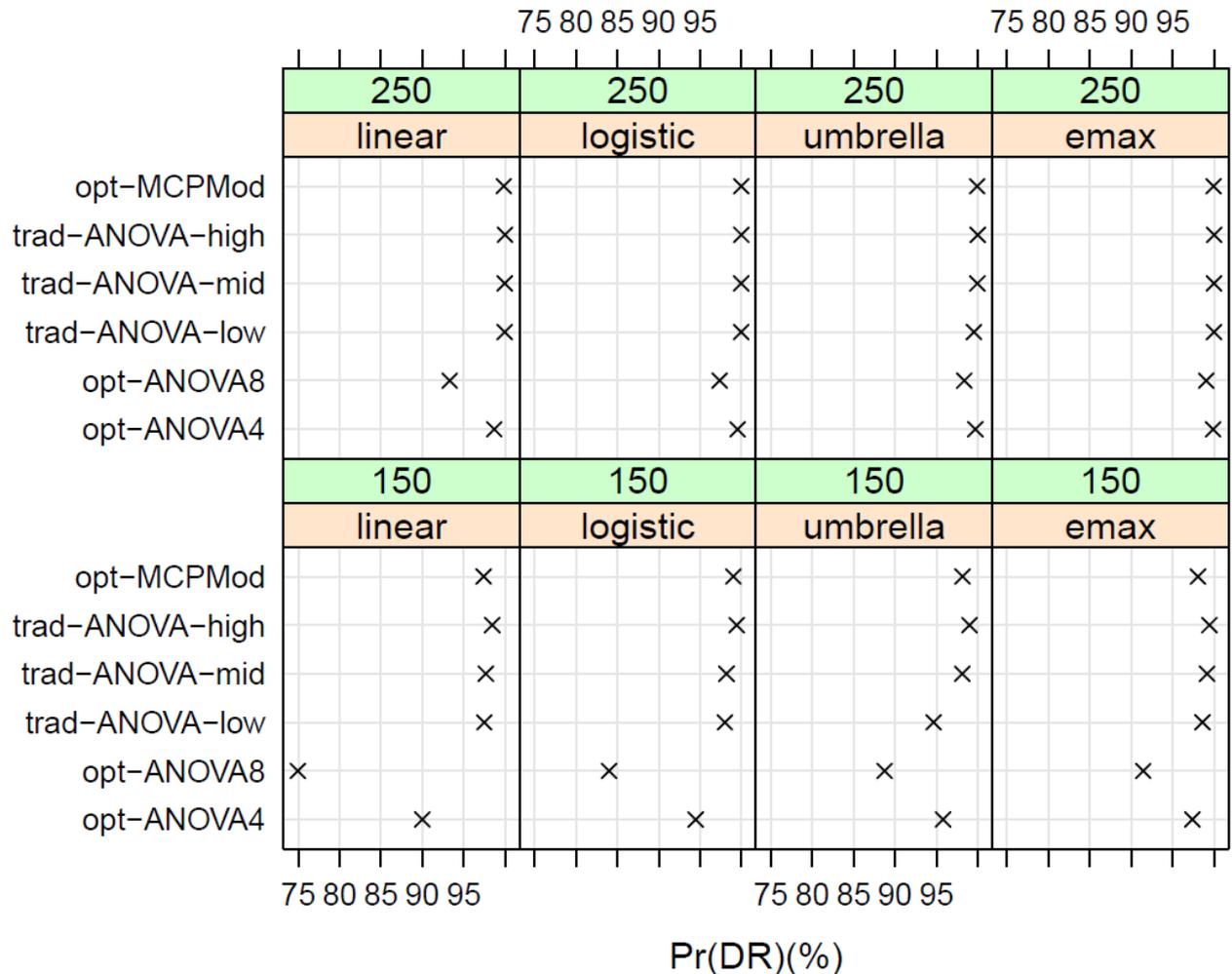
In Figure 1-1 one can observe the power to detect a dose response trend for the different methods. The three traditional ANOVA approaches perform best in most situations because only two active doses are used. Despite the fact that MCP-Mod uses 4 active doses, its power values are very close to that of the respective best approach. When using more than two dose levels with an ANOVA approach, one can observe that the performance quickly deteriorates, despite the fact that optimal allocations are used, see the results for opt-ANOVA4 and opt-ANOVA8.

Figure 2-2   Relative bias in dose estimate

Relative Bias in Dose Estimate (%)
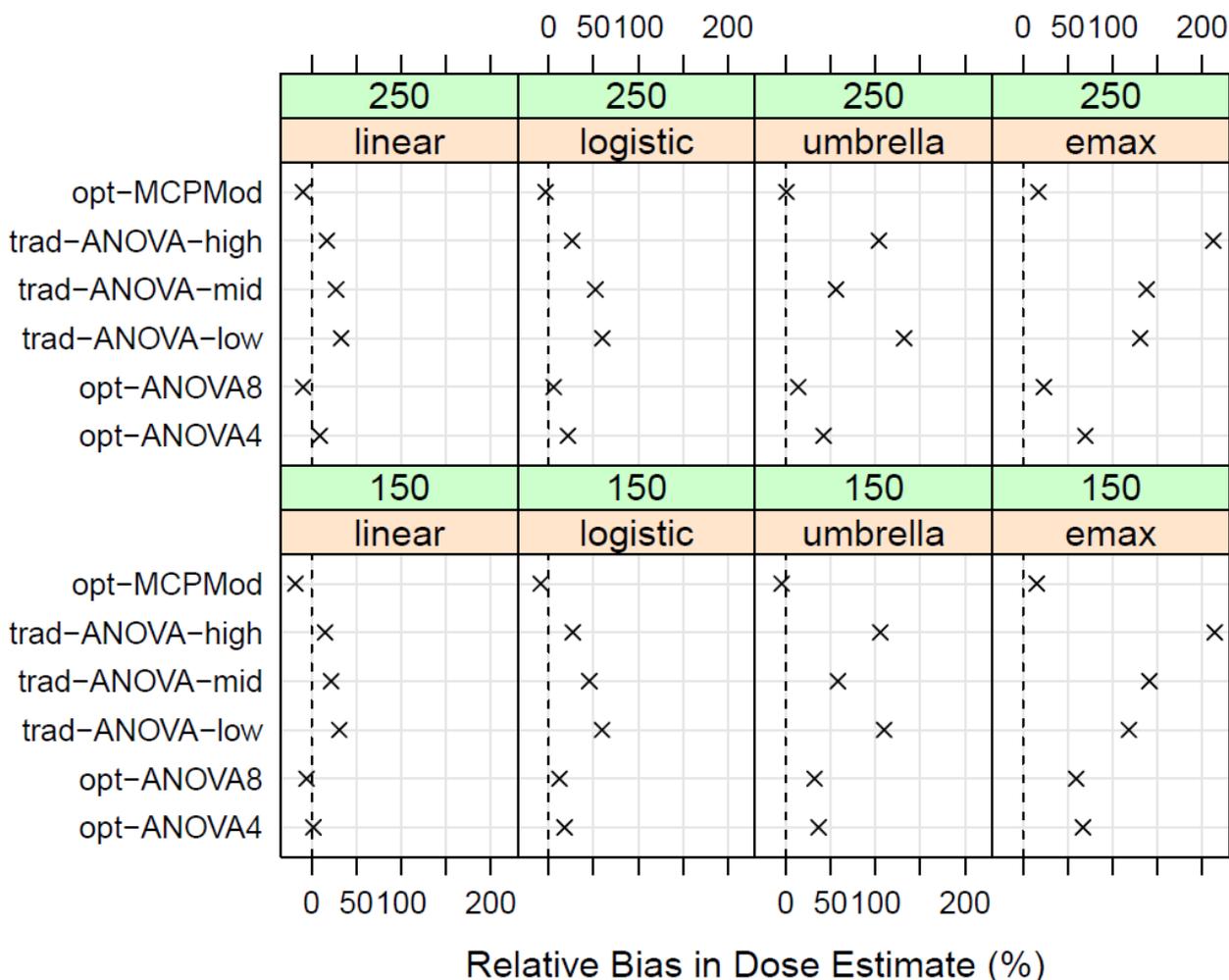
When it comes to the bias in the dose estimate (Figure 1-2) one can observe that MCP-Mod provides roughly unbiased estimates of the target dose in almost all scenarios. The ANOVA type approaches are all slightly upwards biased (in particular the traditional ANOVA approach), which results from the fact that only the doses studied in the design can be chosen for ANOVA. Only the opt-ANOVA8 approach with 8 active doses provides performance similar to MCP-Mod. This approach however was able to identify a dose response signal in far fewer situations than MCP-Mod, i.e. it had a considerably smaller power than MCP-Mod (see Figure 1-1).

Figure 2-3 Relative absolute error in dose estimate
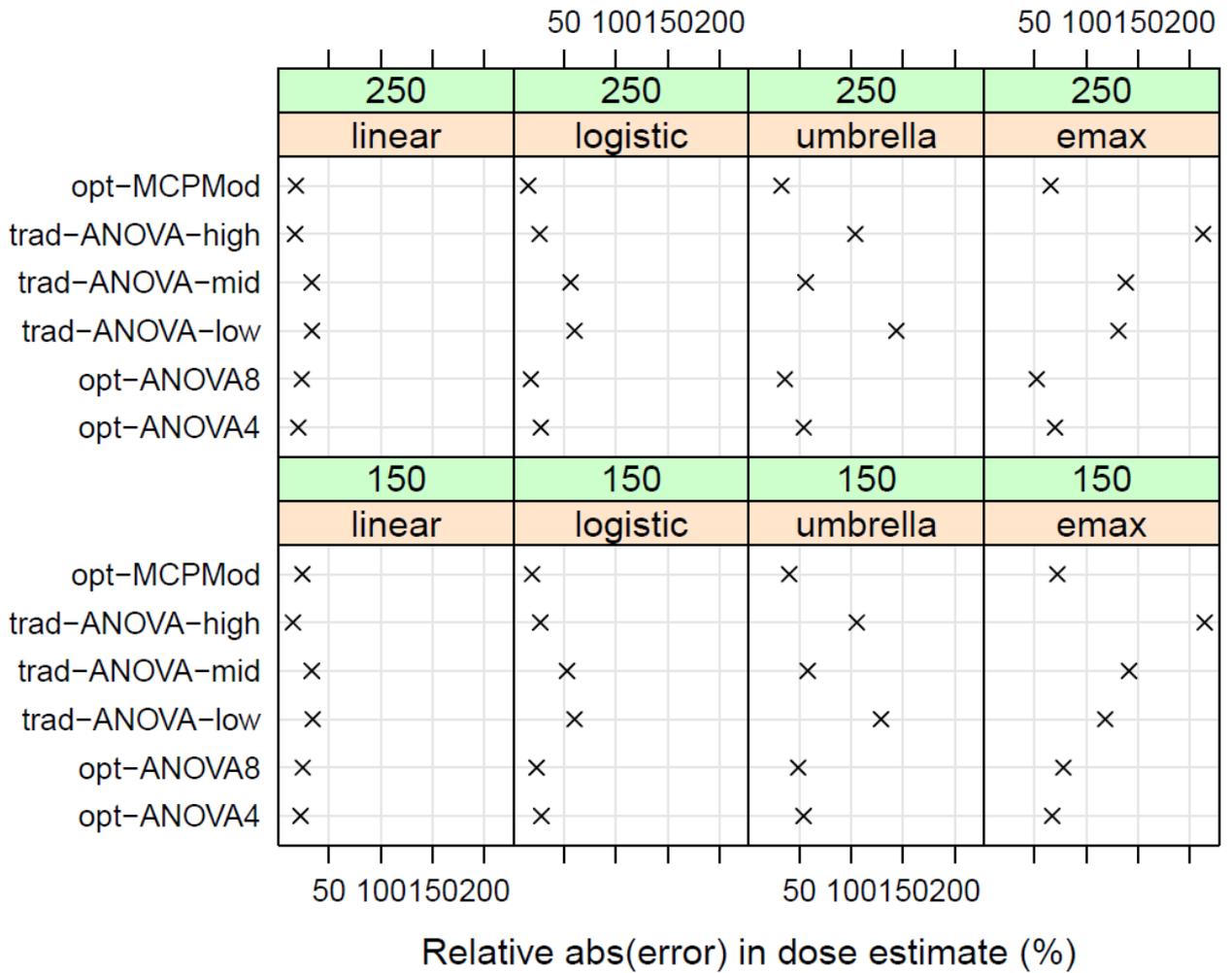
For the absolute estimation error (Figure 1-3) MCP-Mod is either the best or close to the best approach. The performance of the ANOVA approach depends on the underlying design used, in particular for the traditional ANOVA approaches.

Figure 2-4   **Average prediction error in estimating the dose response function**

	250 linear	250 logistic	250 umbrella	250 emax
opt–MCPMod	×	×	×	×
trad–ANOVA–high	×	×	×	×
trad–ANOVA–mid	×	×	×	×
trad–ANOVA–low	×	×	×	×
opt–ANOVA8	×	×	×	×
opt–ANOVA4	×	×	×	×

	150 linear	150 logistic	150 umbrella	150 emax
opt–MCPMod	×	×	×	×
trad–ANOVA–high	×	×	×	×
trad–ANOVA–mid	×	×	×	×
trad–ANOVA–low	×	×	×	×
opt–ANOVA8	×	×	×	×
opt–ANOVA4	×	×	×	×

Average prediction error relative to target effect (%)
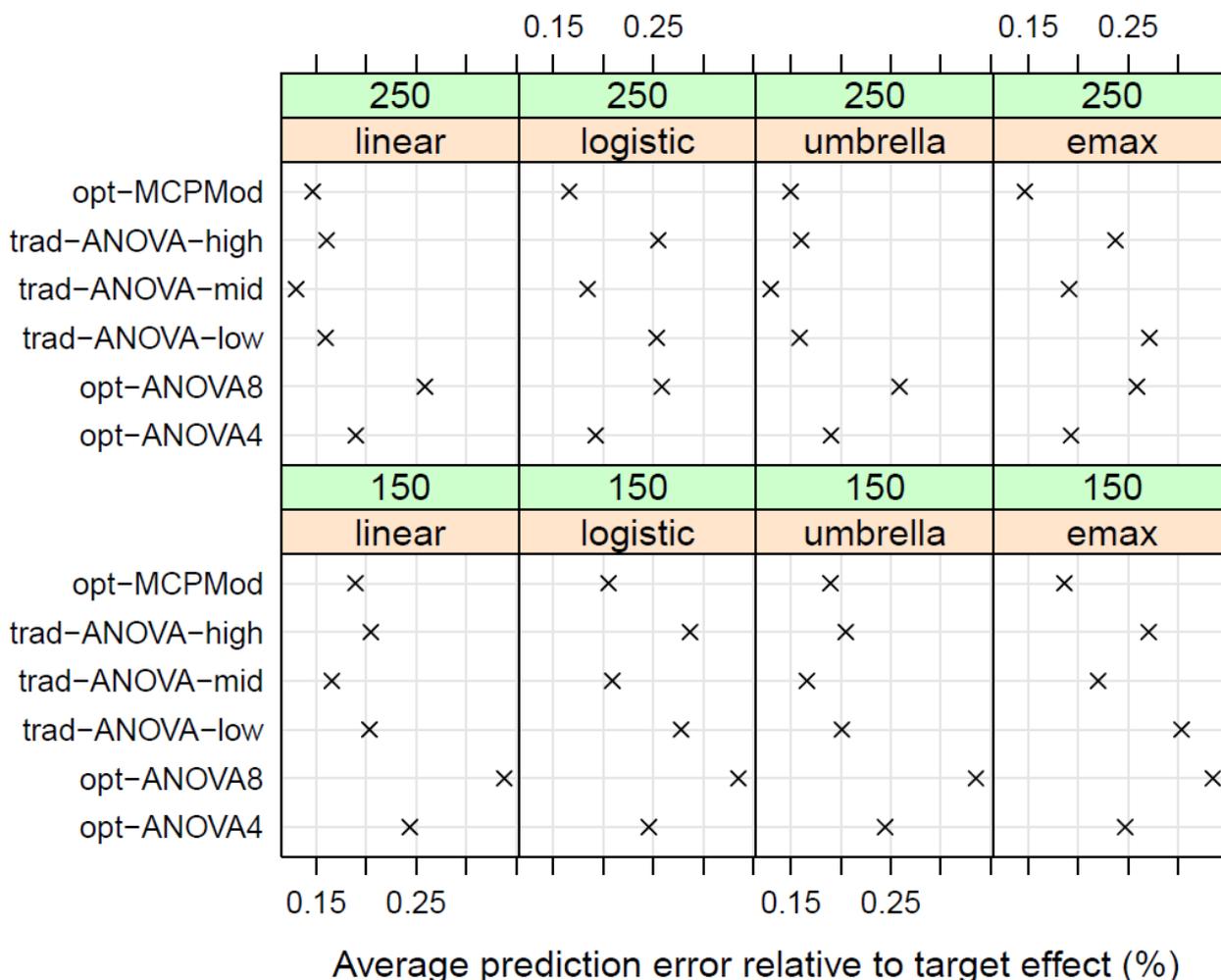
For the average prediction error in estimating the dose response function (Figure 1-4) MCP-Mod again performs best or is close to the best approach. The traditional ANOVA approach performs reasonably well in some of the scenarios because the spline interpolation can work well with a small number of doses, depending on the true dose response curve and the employed design. For example, the trad-ANOVA-high approach performs quite well for a linear model, but bad for the emax dose response model.

2.1.4   Summary

The simulations provide evidence that MCP-Mod is a robust methodology for dose response modeling. We compared MCP-Mod with a total of 5 traditional or optimized ANOVA approaches. While some of the ANOVA approaches occasionally give comparable or even slightly better performance, no single ANOVA approach demonstrates a robust performance across all metrics and scenarios as compared to MCP-Mod. For example, opt-ANOVA8 performs well for estimating the target dose (Figures 1-2 and 1-3), but performs worse in terms of power or dose response estimation (Figures 1-1 and 1-4). As another example, trad-ANOVA-high performs well across all metrics if the true dose response model is linear. If the

true dose response model follows an Emax shape, however, its performance is always among the worst methods in the dose-response and dose estimation metrics.

In general the performance of the ANOVA approaches is sensitive to the underlying scenario and the employed design, in particular when the used number of dose levels is small. When the number of dose levels is larger, the performance of the ANOVA approaches with respect to dose response estimation and power deteriorates. However, including a sufficiently large number of doses in a clinical dose finding study is important to reliably estimate dose response not only for the main efficacy endpoint (as studied in this simulations), but also important safety or tolerability variables, which will also influence dose selection for Phase II

2.2   Question 2

What aspects of technical performance and ease of application should the user consider when deciding between different criteria for model selection criteria, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC)? Is any single approach universally preferable in terms of the accuracy of the resulting information? Similarly, for conducting an efficient trend test, how should a user decide which approach might be preferred (e.g. a multiple contrast test or a likelihood ratio test)?

Answer

Many common information criteria are based on a specific form of a penalized log-likelihood function. The main objective is to find a balance between a good model fit and a reasonably small model complexity, where complexity refers to the number of parameters in model. The most widely used criteria are the AIC or the BIC, which differ in the penalization term to measure model complexity. The AIC uses the length of the parameter vector (i.e. number of model parameters) multiplied by 2 as penalty and the BIC uses log(n) multiplied with the length of the parameter vector, where n denotes the number of observations.

There are different possibilities to compare the quality of model selection criteria. In the following we briefly describe the asymptotic properties of consistency and efficiency to compare the AIC and BIC against each other. In general, no information criterion is uniformly better with respect to all other criteria and it depends on the situation at hand which criterion to use. Simulations are then necessary to better understand the relative performance in a particular situation, see further below.

One way to compare different model selection criteria is the consistency property. Loosely defined, consistency means that the information criterion picks the "true" model with probability tending to one, if there exists a candidate model that is closest to the "true" model. Whether an information criterion is consistent or not depends in general very sensitively on the structure of the penalty term. In case that there is only one best model, a sufficient condition for consistency is given by the condition that the strictly positive penalty term divided by the sample size converges to zero. Therefore both the AIC (penalty: twice the length of the parameter) and the BIC (penalty: log(n) multiplied with the length of the parameter) are consistent in this situation. However, if the best approximating model amongst the candidate models is not unique and there exist several "best" models with different complexities, information criteria are not consistent anymore if their penalty is fixed and does not depend on the sample size. In this situation, the AIC is therefore not a consistent model

selection criterion, but the BIC is. Another disadvantage of the AIC caused by its fixed penalty is that it will not necessarily select the model with the smallest number of parameters in the set of "best" models and therefore has the tendency of overfitting.

An alternative way of comparing different model selection criteria is the efficiency property. A model selection criterion is called efficient if it selects a model such the ratio of the expected loss function (e.g. prediction error, squared estimation error) at the selected model and the expected loss function at its theoretical minimizer tends to one in probability. Therefore the efficiency depends on the choice of the loss function. Comparing the AIC with BIC with respect to efficiency, it can be shown that the AIC performs better than the BIC because of the fixed penalty term. In some cases the risk function of the BIC criterion is even unbounded (e.g. Claeskens and Hjort, 2008, p. 104).

Experience gained from simulations in recent years shows that the BIC compared to the AIC sometimes tends to penalize model selection too strongly for the sample sizes and variances encountered in typical Phase II dose finding trials. For example, the simulations in Bornkamp (2006) showed that the BIC favored the linear model even in situations where it was not appropriate.

Although the AIC and BIC are standard model selection criteria that are often employed in practice, there are alternative information criteria that could be considered as well. AIC and BIC select one "best" model, which is then used to explain all aspects of the true model, but there are many cases where one model has good properties for one particular objective (such as prediction), whereas other competing models are more appropriate for other objectives (such as estimating a target dose, like the MED), which might be more relevant in multi-objective experiments like clinical dose finding studies. In order to address different objectives at the model selection stage the focused information criterion (FIC) has been introduced (Claeskens and Hjort, 2003), which selects the "best" model for estimating a particular focus parameter. The main advantage of this criterion consists in the fact that it selects the model for a particular statistical inference, which might be useful in the context of, for example, MED estimation. However, the application of the FIC to clinical dose finding studies is subject to further research; see also the answer to Question 4a.

For the purpose of performing a trend tests in the context of MCP-Mod, a likelihood ratio test for the comparison of several competing dose response models against a flat model of no dose-related response has been described only very recently (Dette et al., 2013). In comparison to a multiple contrast test, such a likelihood ratio test has at least two advantages. On the one hand, it makes better use of the available information, because it uses the complete structure of the regression models. On the other hand, it does not require guesstimates of the parameters of the competing models which are needed to construct the contrast tests for dose response signal detection. However, the proposed likelihood ratio tests are difficult to implement, because of the problem of non-identifiability of some model parameters under the null hypothesis of no dose response. Consequently, standard asymptotic theory for likelihood ratio tests is not applicable here and likelihood ratio tests become more difficult to implement numerically than multiple contrast tests (Dette et al., 2013). While the MCP-Mod procedure allows for applying either of these two trend tests derived from a pre-specified set of competing dose response models, future research is necessary to better understand the properties of the likelihood ratio tests and how to apply it in general parametric models. On

the other hand the use of multiple contrast tests is well understood in the case of general parametric models and a software implementation is available with the DoseFinding R package; see Pinheiro et al. (2013) as well as the answer to Question 4b.

2.3   Question 3

With regard to the handling of statistical model uncertainty in the context of dose selection prior to phase III, please summarise the pros and cons of MCP-Mod versus other possible approaches, in terms e.g. of biases and overfitting, and possible ways to minimize the other approaches drawbacks, and discuss how the user could select one approach or another depending on circumstances.

Answer

The common way of performing a model-based approach is to fit several dose response models once the data have been observed and select the best fitting model. However, such a naïve approach does not account for model uncertainty and can lead to undesirable effects due to data dredging, such as overfitting, biased treatment effect estimates and over-optimistic analysis results. For example, it is common practice to compute the variance of a parameter estimate without acknowledging the model selection step, while the variance of the parameter estimate acknowledging the model selection will typically be quite different. In the following we elaborate further on the problem post-model-selection inference.

Classical statistical theory grants validity of statistical tests and confidence intervals by separating the model selection process from the analysis of the data being modeled. In practice this separation is rarely appropriate since a model is often selected after a data driven selection process. However, model selection has an important impact on subsequent statistical inference. As pointed out by Leeb and Pötscher (2005) there are many difficulties if statistical inference is based on a preliminary model selection step regardless of the particular model selection procedure (due to e.g. hypothesis testing, Akaike or Bayesian information criterion, final prediction error, or cross-validation). In particular, the sampling properties of post-model-selection estimators are typically quite different from the nominal distributions that arise in a fixed model, and as a consequence, inference procedures not taking into account the model selection step can have to undesirable properties.

It is mathematically correct that the use of a consistent model selection procedure guarantees that the point-wise asymptotic distributions of the post-model-selection estimators coincide with the asymptotic distributions which would arise if the selected model would be chosen a-priori. However, this does not support the frequently used argument in the literature that consistent model selection procedures allow the application of the standard asymptotic distributions in a fixed model, because the finite-sample distributions of a post-model-selection estimator are typically not uniformly close to the respective point-wise asymptotic distributions. Therefore – even for large sample sizes – these asymptotic distributions cannot be used directly to replace a complicated finite-sample distribution of a post-model-selection estimator. Moreover, the asymptotic distributions of post-model-selection estimators are typically very complicated and depend on unknown parameters whose estimation seems to be impossible (even in large samples).

The MCP-Mod approach tries to alleviate these problems in two ways (i) using a pre-specified set of clinically meaningful dose response models and thus avoiding an extensive search for a well-fitting or overfitting model post hoc and (ii) when model averaging techniques are used in MCP-Mod (and not model selection), one formally incorporates model uncertainty in the statistical inference, by using weighted inference procedures. In several of the recent applications, MCP-Mod has been used in combination with model-averaging techniques.

More formally and along similar lines, a few articles have recently been published to address this problem in a similar fashion and in more generality. For example, Hjort and Claeskens (2003) proposed a frequentist model average estimation procedure, which addresses the additional uncertainty introduced by the model-selection step. Their approach is based on an asymptotic theory where the limiting distributions (and corresponding risk properties) of post-selection estimators can be precisely described. In a different direction some progress has been made to circumvent these problems by introducing a valid post-selection inference method based on simultaneous inference (see Berk et al. 2013).

2.4   Question 4

Please summarise any plans for further development in terms of:

   a.   Technical aspects

   b.   Breadth of application to different experimental situations

   c.   Dissemination of methodology to drug developers

Answer

2.4.1   Technical aspects

Further development of the MCP-Mod approach will focus on two main aspects: (i) investigation of suitable model selection criteria, and (ii) construction of robust designs for model selection.

   (i) The task of selecting an appropriate model is an important problem in statistical inference. Classical model selection criteria, such as AIC or BIC, do not address the problem that a model is selected for a particular purpose (such as the estimation of the minimum effective dose) ; see also the answer to Question 2. Claeskens and Hjort (2003) introduced the focused information criterion (FIC) which selects a model from a given class of candidate models for a particular focus. Unfortunately, their original methodology is not applicable in the context of dose finding studies, because it requires that the competing models are nested. This is rarely the case for the models commonly used in clinical dose response studies. Therefore, we plan to investigate the application of a focused information criterion for non-nested models and to use this criterion for the second step of the MCP-Mod approach. From a methodological point of view, this requires the investigation of the maximum likelihood estimates under model misspecification.

   (ii) It is well known that the application of optimal or efficient designs can substantially improve the performance of statistical procedures. While efficient or optimal designs have been constructed either for a fixed model (see, for example, Dette, Kiss, Bevanda

and Bretz (2010)) of for a class of competing models (e.g. Bornkamp, Bretz, Dette and Pinheiro (2011), Dette, Bretz, Pepelyshev and Pinheiro (2008)), optimal designs for model selection with respect to classical information criteria have been rarely discussed in the literature. There are references regarding discrimination designs, which refer to testing with likelihood ratio tests. On the other hand – to our best knowledge – no methodology is available for constructing optimal designs with respect to the commonly used information criteria (AIC or BIC) or the FIC criterion. New optimality criteria for the construction of efficient designs for model selection with respect to these information criteria, in particular in dose-response studies, have to be developed in the future.

A particular challenge in the construction of optimal designs consists in the fact that optimal designs in non-linear models usually depend on the unknown model parameters. Here, a Bayesian approach might be useful, which allows the experimenter to specify the uncertainty regarding the unknown parameters in the optimality criteria. The prior distribution can directly be used to reflect the knowledge about these parameters. If such knowledge is sparse, the construction of informative priors is a challenge. One possibility in these situations is to investigate optimal designs with respect to Jeffrey's prior.

An alternative to the Bayesian approach is the construction of adaptive designs for model selection. Here, experiments are conducted in several stages. In each stage the information from the previous stages is used to quantify probabilities for the competing models in the model selection criterion and to update the information regarding the unknown parameters in all models. It is an open problem if such a procedure leads to consistent model selection and to consistent estimates in the identified model.

2.4.2   Breadth of application to different experimental situations

The MCP-Mod methodology has been developed to cover clinical dose finding studies as they often occur in Phase II to support the dose selection for Phase III. The methodology available as of today and implemented in the DoseFinding package on CRAN covers a broad range of parametric models, where the response variable might be normally distributed, or even a count, binary or time-to-event variable. In addition, the analysis can be adjusted for relevant covariates (e.g. region, age, etc.) and account for measurements recorded over time (necessitating the use of longitudinal models); see Pinheiro et al. (2013) for further details. Examples of trial designs and modeling approaches where we have only limited experience, but in the future plan to broaden our understanding about applicability and performance of MCP-Mod, include among others:

* Exposure-response analyses are possible, but the MCP part would need to be performed using likelihood ratio tests. The use of the multiple contrast tests will be challenging due to the many different distinct values of the input variables.

* Regimen finding for long acting biologics where there is no steady state. Unlike conventional small drug molecules, which are commonly given as tablets once daily, biologics are typically injected at much longer time intervals, i.e. weeks or months. Hence both the dose and the time interval have to be optimized during the drug development process for biologics. Research on how to design clinical trials with biologics is sparse and we currently have a PhD student working on this problem as part of the MEDIASRES

program funded by the European Community's Seventh Framework program FP7/2011; see http://www.mediasres-itn.eu/ for more information.

- Application of MCP-Mod in confirmatory studies to better use all available patient information in, for example, small population groups. The MCP-Mod methodology would have to be extended by employing the closed testing principle to develop confirmatory tests for the global trend assessment (i.e. whether there is any statistical evidence for a dose-related drug effect) as well as for the pairwise comparison of the individual doses against placebo. We plan to have a PhD student working on this problem as part of the IDEAL program funded by the European Community's Seventh Framework program FP7/2013.

- The MCP-Mod methodology can be extended to certain multivariate problems, such as the joint modeling of efficacy and toxicity, the presence of two primary endpoints, or drug combination trials. So far, we have limited experiences but we consider applying the MCP-Mod approach, if feasible, in future trials where such problems occur.

2.4.3   Dissemination of methodology to drug developers

We envisage disseminating the MCP-Mod methodology to drug developers (statisticians, pharmacologists, clinicians, etc.) in different ways. First, we plan to maintain and expand the DoseFinding package on CRAN (the Comprehensive R Archive Network available at http://cran.r-project.org/). This ensures public access to the most recent implementation of the MCP-Mod methodology and its variants. Furthermore, this enhances the reproducibility and transparency of the underlying calculations. Second, we have developed a Graphical User Interface (GUI) on top of the DoseFinding R package for Novartis-internal purposes. Such GUIs greatly enhance the use of novel methodologies, such as MCP-Mod, in particular by drug developers who are not familiar with R or related software solutions. The GUI is mainly used for planning purposes to facilitate the interactive discussions within clinical teams and help them bring together the information that is relevant to design the dose finding trials at hand. Third, we continue developing educational seminars that are targeted at different audiences (statisticians, clinicians, etc.). For statisticians, tutorial courses have been and will be given at scientific conferences and Novartis internally. For clinicians, for example, Novartis is currently developing (with external expertise from Hibernia College) an extensive course for GPTs (Global Program Teams), and those who impact dose decisions, to foster better dose finding practices (including a module on MCP-Mod) and encourage teams to look for desired information during early drug development, before moving into pivotal trials. This course will cover 4-6 weeks on-line learning for participants with assistance from tutors (subject matter experts), and three workshops during GPT meetings to work on specific case studies as well as to develop and share experience in a group setting.

References

- Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013). Valid post-selection inference, Annals of Statistics 41, 802-837.

- Bornkamp, B. (2006) Comparison of Model-Based and Model-Free Approaches for the Analysis of Dose-Response Studies, Diploma Thesis, University of Dortmund, http://www.statistik.tu-dortmund.de/~bornkamp/diplom.pdf

- Bornkamp, B., Bretz, F., Dmitrienko, A., Enas G., Gaydos, B., Hsu, C.H., Koenig, F., Krams, M., Liu, Q., Neuenschwander, B., Parke, T., Pinheiro, J., Roy, A., Sax, R., and Shen, F. (2007) Innovative approaches for designing and analyzing adaptive dose-ranging trials (with discussion). Journal of Biopharmaceutical Statistics 17, 965-995.

- Bornkamp, B., Bretz, F., Dette, H. and Pinheiro, J. (2011). Response-adaptive dose-finding under model uncertainty. Annals of Applied Statistics, Vol. 5(2B), 1611-1631

- Bornkamp, B., Pinheiro, J. and Bretz, F. (2013). DoseFinding: Planning and Analyzing Dose Finding experiments. R package version 0.9-6. http://CRAN.R-project.org/package=DoseFinding

- Claeskens, G. and Hjort, N.L. (2003). The focused information criterion. Journal of the American Statistical Association 98, 900-916 (with discussion)

- Claeskens, G and Hjort, N.L. (2008) Model selection and model averaging. Cambridge University Press, 2008.

- Dette, H., Bretz, F., Pepelyshev, A. and Pinheiro, J. (2008). Optimal designs for dose finding studies. Journal of the American Statistical Association, Vol. 103(483), 1225-1237

- Dette, H., Kiss, C., Bevanda, M., and Bretz, F. (2010) Optimal designs for the Emax, log-linear and exponential models. Biometrika 97, 513-518.

- Dette, H., Titoff, S., Volgushev, S., and Bretz, F. (2013) Model identification for dose response signal detection. (submitted) http://www.ruhr-uni-bochum.de/imperia/md/content/mathematik3/publications/dettitbre290812.pdf

- Dunnett, C.W. (1955) A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association 50, 1096-1121.

- Kabaila, P. (2009). The coverage properties of confidence regions after model selection, International Statistical Review 77, 405-414

- Kabaila, P., and Leeb, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection, Journal of the American Statistical Association 101, 619-629.

- Leeb, H. and Pötscher, B.M. (2005). Model selection and inference: facts and fiction, Econometric Theory 21, 21-59.

- Pinheiro, J., Bornkamp, B., Glimm, E., and Bretz F. (2013) Multiple comparisons and modeling for dose finding using general parametric models (submitted) http://arxiv.org/pdf/1305.0889

- Pukelsheim, F. and Rieder, S. (1992) Efficient rounding of approximate designs, Biometrika, 79, 763-770