Statistical Review and Evaluation
Qualification of Statistical Approach

Proposed Qualification: MCP-Mod

Sponsor: Janssen Research and Development (work conducted by investigators at Janssen and Novartis)

Statistical Reviewers: Mohamed Alosh, Ph.D. Division of Biometrics III
Lei Nie, Ph.D. Division of Biometrics V
Dionne Price, Ph.D. Division of Biometrics IV

I. Background

A request has been submitted by Janssen Research and Development for qualification of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. The methodology uses principles of multiple comparisons with modeling techniques and is purported to be advantageous in its flexibility to search for and identify an adequate dose for use in confirmatory studies. Moreover, the submission states that the approach is efficient in that it uses available data better than traditional pairwise approaches.

Understanding the dose-response relationship and identifying the appropriate dose for Phase III clinical trials are probably the most critical, and yet most challenging, components of the clinical development program of a new therapy. Despite this importance, dose ranging studies have often been designed using a small number of doses and a narrow dose-range. Furthermore, the upper end of the dose response relationship has often been the focus, contributing to the lack of understanding of the dose-response relationship and selection of the appropriate dose. Inappropriate dose selection can result in unacceptable toxicity or adverse events when the dose is too high, and alternatively, selecting a low dose increases the likelihood that the therapy provides insufficient evidence of effectiveness. Even when evidence of effectiveness is shown for a selected dose, the need may arise to investigate alternative doses due to inadequate knowledge of the dose-response relationship.

Traditionally, the analysis of dose-response studies has been divided into two major strategies: multiple comparison procedures and model-based approaches. The latter approach assumes a functional relationship between the response and the dose. Consequently, a pre-specified parametric model is fit to the data, and the fitted model is then used to estimate an adequate dose to achieve a desired response. The validity of the conclusion is dependent on the correct choice of the 'unknown' dose response model. On the other hand, multiple comparison procedures generally do not make assumptions about the underlying dose-response relationship and have

been used to detect an overall dose related signal as well as to estimate target doses of interest. One approach that is often used for dose selection is based on assessing the significance of contrasts between different dose levels while preserving control of the Type I error rate. The procedure is robust as it does not require model specification; however, its inference is confined to the selection of a dose among the dose levels under investigation. In summary, modelling approaches and multiple comparison procedures each have merits as well as disadvantages for identifying the "right" dose.

To overcome some of the shortcomings of the individual approaches, MCP-Mod was introduced by Bretz, Pinheiro and Branson (2005) and further explored by others in series of papers in the statistical literature. The papers describe MCP-MOD as a general framework that facilitates the combination of confirming the existence of a drug effect and estimating the dose that provides a particular therapeutic response. In Section II, we provide a high-level summary of MCP-Mod and summarize the extensions, examples, and simulations presented in the submission. We briefly present comments and possible future work in Section III. We conclude with our perspective in Section IV.

II. MCP-Mod

Methodology

MCP-MOD aims to provide a comprehensive approach for dose-response testing and estimation so that adequate doses can be identified for future confirmatory clinical trials. The dose or doses are selected through a general framework based on confirming the existence of a drug effect and estimating dose(s) that provide a certain therapeutic response.

The methodology starts with the trial design stage. During the trial design stage, the study population, endpoint, number of doses, and required sample sizes are established. In addition, a set of plausible candidate parametric models that are likely to represent the dose-response shape are identified. The set of candidate models is pre-specified based on available information, including clinical information or prior information from similar compounds. This candidate set should be chosen to cover a sufficiently large and diverse set of dose response shapes when there is large model uncertainty. Once a set of candidate models has been identified, "optimum" contrast coefficients for each of the candidate models to be used for testing the presence of a dose-response signal are derived. The optimality here refers to maximizing the marginal power to detect a specific dose-response shape associated with the candidate model when that model is the correct one. For deriving the optimal contrast coefficients, the methodology considers that it is sufficient to work with standardized versions of the candidate models along with initial estimates of the parameters of the standardized models, which consequently provide estimates of the mean response for doses in the models. Prior estimates of the parameters are typically

derived from initial knowledge (or guesses) of the expected percentage of the maximum response associated with a given dose.

The trial analysis stage follows the design stage. It entails the following steps: (i) test for the presence of a dose response signal, (ii) select the best dose-response model for the observed data out of the pre-specified set of candidate models, and (iii) estimate targeted dose(s) of interest (e.g. the minimum effect dose) via modelling. In the trial analysis stage, contrast tests for the mean response for the doses under investigation along with the pooled variance estimator are used to detect an overall trend. The final test statistic is based on the maximum contrast test, and a dose response signal is confirmed if at least one single contrast test is statistically significant at a pre-specified significance while controlling the familywise error rate for dose response signal detection. Once a model is selected based on the contrast test, its original form (not the standardized version) can be used for the dose estimation. The next step of the analysis stage involves selecting the "best" model from a set of reference dose-response models established from the models with statistically significant dose-response signals. The "best" model may be selected using Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), or it may be the model associated with the most significant contrast test. The submission also states, "Alternatively, multiple significant models can be selected if model averaging is preferred (Buckland et al., 1997)." Lastly, the selected dose response model is used for making inference on the targeted doses within the dose range under investigation using inverse regression techniques. The precision of the estimated dose can be assessed using, for example, bootstrap methods.

Software, Extensions, and Simulations

In subsequent sections of the submission, the applicant discussed a software package to implement MCP-Mod, provided an example to illustrate the methodology, explored further investigations and extensions of the methodology, summarized the MCP-Mod experience of Janssen and Novartis, and presented simulation results. Of note, an assessment of the statistical package is beyond the scope of this review.

Following an example of the use of MCP-Mod for normally distributed responses, the applicant listed several extensions of the methodology for non-normally distributed responses and other types of study designs to illustrate the generality of the methodology. In particular, the applicant described two variants for using the methodology in response-adaptive dose finding studies and for general parametric models (e.g., time to event endpoints and longitudinal data modeling). An example which considers a mixed effects model for repeated measurements over time (Pinheiro et al, 2014) was additionally provided.

Of note, the applicant stated that the scope of MCP-Mod is the analysis of an efficacy response variable that is ideally predictive of the clinical Phase III efficacy outcome. The response can be normally distributed, binary, count or a time to event endpoint. The observations can be cross sectional (i.e., from a single time point) or longitudinal, with the dose being a typical dose level, or more broadly the "dose" could be any univariate, continuous, quantitative measurements as long as an ordering of measurements is possible. Further, four doses, at a minimum, are suggested for the application of the methodology.

The submission additionally included a summary of Novartis' and Janssen's experiences with using MCP-Mod methodology in several studies and presented results from two simulation studies to allow a quantitative assessment of the methodology. The first simulation study from Branson et al. (2003) compared the MCP-Mod approach with competing trend tests including the likelihood ratio test (*LRT*) and the "*step contrasts*". The simulation study considered nine alternative dose-response models with different sample sizes and assessed the performance of MCP-Mod with respect to its power to detect the existence of a dose-response signal and its dose selection performance through assessing the ability to choose a dose close to the desired level. The results of the simulation experiments show that the Type I error rate is well controlled at its nominal level 0.05 for all sample sizes The ability of the contrast tests in the MCP-Mod approach to discriminate among the models in the candidate set was assessed by comparing the simulation probabilities of correctly identifying the response model for the models in the candidate set and for the different sample sizes. Findings for the comparison of the power values for the MCP-Mod with the *LRT* and the *step contrasts* were dependent on the assumed dose-response shape. The power for MCP-Mod is generally comparable to that of *LRT* test. While the *LRT* test has a slight power advantage for the convex shape, the MCP-Mod is more powerful than the *LRT* for the quadratic and double-logistic shapes. Both the MCP-Mod and the *LRT* are generally more powerful than the *step contrasts*. As for the dose selection using the MCP-Mod, the results of the simulation study show that the precision of the dose selection varies by the underlying dose-response shape.

The second simulation study investigated the performance of several innovative approaches for dose- ranging methods and was originally the subject of a white paper conducted by the PhRMA "Adaptive Dose-Ranging Studies" (ADRS) working group (Bornkamp et al. 2007). The study compared the response-adaptive MCP-Mod version, an ANOVA approach, a Bayesian model-averaging approach and a nonparametric dose-response modelling approach. The study considered five underlying candidate models for the dose-response shapes and assumed a normally distributed response for dose. The performance of the method in this simulation study was assessed by the following: (a) detecting dose-response, (b) identifying clinical relevance, (c) selecting a target dose and (d) estimating the dose-response. While the results of the simulation experiment demonstrated that all methods were capable of controlling the Type I error rate, no clear best method dominated the others for the above criteria; however, the MCP-Mod performed

favorably compared to the other methods and outperformed the benchmark ANOVA approach in many cases.

III. General comments

Based on our review of the submission, we have the following general comments. Many of the comments are only intended for future extensions or improvements to the approach.

1. The MCP-MOD approach focuses on a univariate response of efficacy or safety. However, in practice, selection of the appropriate doses for the confirmatory clinical trials should generally be based on consideration of both of efficacy and safety jointly. The authors state that multivariate problems are out of the scope of the current request.

2. The method currently does not provide general guidance on selection of the dose levels and sample sizes in each dose level, which could depend on knowledge of the dose-response curve.

3. The method currently does not provide guidance on how the results should be interpreted or utilized when none of the candidate models are correct. Perhaps the results can be still used (e.g. models meeting criteria that are less stringent than the pre-specified criteria could be considered as candidate models for future dose-ranging studies).

4. The package did not provide details on the model averaging approach. .

5. When none of the candidate models are significant, the user should explore possible reasons such as small sample sizes (e.g. a sample size targeted to a different model than what the data suggests) or poorly chosen initial candidate models.

6. The type I error rate and power of MCP-MOD are comparable to other alternatives explored in the submission and independently by the review team.

IV. Conclusions

Multiple comparison procedures and modelling approaches, traditionally used for the analysis of dose ranging studies and ultimately selecting a dose for future confirmatory trials, each have merits and shortcomings. The proposed MCP-Mod strategy aims to provide a comprehensive approach that combines merits of multiple comparisons and modelling in the analysis of t dose-response studies. The sequential approach underlying MCP-Mod entails learning the likely dose-response model through multiple testing across candidate models after adjusting for multiplicity and then using the "best" models to select the appropriate doses. MCP-Mod is logical and avoids some of the pitfalls of traditional approaches. MCP-Mod is efficient in that it uses the available data better than traditional pairwise comparisons. The methodology, as the

applicant noted, is intended for use in Phase II dose findings studies to support dose selection for Phase III trials. The distribution for response, taken to be a function of the dose, can be normal, binary, count or time to event endpoint. The measurements can be cross-sectional (i.e., from a single time point) or longitudinal data. Further, the methodology can be applied to a more broad definition of the dose, to cover any univariate variable where ordering of the measurements is possible and the difference between the measurements is interpretable. For application of the methodology, the applicant recommended that the number of doses in the study should be four, at a minimum including placebo.

Based on MCP-Mod analysis findings as summarized in the applicant's submission and the relevant statistical literature, endorsement of the MCP-Mod as a method for analysis of dose-ranging Phase II trial can be justified. However, this endorsement does not preclude the availability and use of other statistical methods for dose selection. Although the applicant outlined results of simulation studies that compared the performance of MCP-Mod with other methods for dose selection, this review did not seek to address in detail how MCP-Mod fares in comparison with other methods. Generally, findings from simulation studies are driven by the models considered and their parameters.

The MCP strategy is advantageous in that it considers model uncertainty. It adjusts for multiplicity across the candidate models considered and controls the nominal Type I error rate for dose response signal detection. In conclusion, MCP-Mod is a principled approach that advocates for a better understanding of the dose response relationship.