EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

23 January 2014
EMA/617236/2013
Committee for Medicinal Products for Human Use (CHMP)

Submission of comments on 'Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty' (EMA/CHMP/SAWP/592378/2013)

Comments from:

Name of organisation or individual	
1	IFAPP (International Federation of Associations of Pharmaceutical Physicians)
2	Biostatistics Specialist Group, H.Lundbeck A/S
3	IDEAL (Integrated DEsign and AnaLysis of small population group trials) FP7 Consortium (European Union Seventh Framework Programme under grant agreement n° 602552)
4	GSK
5	Alan Maloney
6	EFPIA - Pär Tellner (par.tellner@efpia.eu)
7	Nazneen Shariff
8	Qing Liu, Ph.D.

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7523 7040
E-mail info@ema.europa.eu **Website** www.ema.europa.eu
An agency of the European Union

1. General comments

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
1	We agree on the statement that Phase II studies, specifically intended for the generation of a dose-response curve, are frequently badly designed. In many cases results are not clear-cut, and the dose selection for the Phase III studies is sometimes based on wrong assumptions. Indeed it is not so infrequent the finding that some drugs modify their dosage schedule some time after the market authorization. We therefore agree on the proposed model, but with a note. The suggested model may be applied to conventional drugs such as small molecules. More innovative drugs (proteins, monoclonal antibodies, vaccines, gene and cellular therapies) cannot follow the proposed scheme.	Long acting biologics, vaccines, gene and cellular therapies are not in the scope of this Qualification Opinion.
2	The primary advantage of the MCP-Mod is the MCP step which allows for a more powerful approach to detection of a drug effect ("dose-response") than traditional pairwise dose versus placebo testing when there is uncertainty about the dose-response shape. The subsequent modelling step which tries to identify a "minimal effective dose" (MED) is weaker. The concept of MED is often not well defined and it will always, in practice, be a discussion about size of benefit versus risk. This approach looks at only the benefit side which is unrealistic. Another criticism of the modelling part is that statistical significance of a contrast does not	The questions expressed over the definition and hence the relevance of MED are shared. Of course, MED is just one of many possible quantities to estimate using MCP-Mod, hence specific reference to MED is deleted. In terms of trading of benefit and risk the use of a clinical utility index is discussed in response to Issue 11 from the SAWP discussions. Simultaneous investigation of efficacy and safety is identified as potential area for further development, though arguably 'benefit' is not always the subject of the modelling here, rather a PD parameter acting as surrogate for benefit and surrogate parameters for safety might equally be explored. Sec 3.1.2 of the submission dossier refers to the use of likelihood ratio tests instead of contrast tests, but research is limited so far.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	necessarily imply a good model-fit for the corresponding model. The focus of the qualification is on the use in the context of efficacy endpoints. It would be useful with an opinion about its potential use for safety endpoints, i.e. endpoints where the desired outcome would be absence of a drug effect, as well. As statistical analyses for dose-finding studies are traditionally considered "learning" rather than "confirming", and thus less regulated, it is unclear what is implicated by giving the approach status as qualified. Does it mean that the approach is recommended and thus that other approaches are less relevant? Does it mean that studies using the approach can be used as confirmatory studies?	See the response to Issue 11 raised by the SAWP. In response to the specific questions: No, this approach is not recommended above other approaches, only that it is a qualified approach in and of itself. This is stated explicitly in the Opinion. No, the scope of the documentation and assessment was limited to exploratory studies.
3	We welcome this statement. Dose finding is one of the most difficult steps in drug development and the note provides a realistic assessment of the value of one flexible and useful approach to dose finding (the MCP Mod approach). It would be helpful to future potential users of this approach if the agency could state whether it would be prepared to consider the use of doses in subsequent stages of development that were identified by this method although not actually studied in Phase II. For example, if the method suggested that the optimal dose would be one between two of the doses	In response to the specific question posed the Opinion indicates that a model based approach such as MCP-Mod would better support interpolation. A well constructed model might be interpreted without particular regard to the doses used in the generation of data to inform the model.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	actually studied would the agency be prepared to consider use of such a dose in Phase III studies?	
4	We welcome the opportunity to comment on this Draft Qualification Opinion and support the need for more model based approaches to dose finding. The benefit of the MCP-Mod approach when compared to ANOVA methods is clear and well described. Agreements • In general we agree with the EMA opinion and fully endorse the direction, provided the specific Novartis methodology is a possible choice from several approaches, rather than the only acceptable method. Ultimately, it would be helpful for MCP-Mod to be approved as one approach for model based dose response with researchers able to apply models themselves separate to MCP-Mod and not restrict methods to MCP-Mod only. As an example, MCP-Mod captures techniques for model averaging. There are other approaches to this (e.g. Bayesian model averaging) that could be equally suited. • MCP-Mod is a nice approach combining modelling and multiple comparison procedure together to give type I error control due to multiple model fitting while keeping the benefit of modelling. The linear contrast concept could be a useful option for testing existence of dose response in a potentially more	This is clearly stated in the Opinion document. Qualification is limited to the methodology and scope proposed by the applicant.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	efficient way- perhaps in the case with combined phase IIA/B trials, where hypothesis (PoC) is tested first and then try to find the "right dose".	
	Risk of over reliance on MCP-Mod • There is a risk that researchers become reliant on one approach (MCP-Mod) and don't play a critical role in the difficult arena of model selection. MCP-Mod could be used as a black-box unless methods and details of model fitting are well described and understood by the user.	It is fully accepted that any model based approach, including MCP-Mod, requires some competence to implement and to explain the approach. It is viewed as a tool to complement other practices of good design, analysis and inference of exploratory clinical trials.
	• It would be beneficial to gain qualification in model based approaches in general with MCP-Mod being one of them. The authors state that the MCP-Mod approach overcomes issues with model uncertainty and make this a key benefit, however this seems very similar to approaches where a small set of potential models are explored and compared (e.g via AIC/BIC) and best model selected. In many situations similar models will give identical answers, especially when the full dose range hasn't been studied and so modellers should always take careful attention to model selection.	Whilst this is in principle agreed, Qualification is limited to the methodology and scope proposed by the applicant. Discussion: Whilst there are pros and cons to different approaches, in contrast to a direct application of modelling dose response and selecting a best fitting model post-hoc (which will always lead to some fitted model, regardless of the absolute goodness-of-fit), MCP-Mod has a filtering process which (i) forces clinical teams to think ahead about model uncertainty and removes (at least partially) some of the arbitrariness and other issues of post-hoc model selection and (ii) may also lead to the conclusion that none of the pre-specified model is good enough and we should not continue the analysis
	• For phase IIb study, it could be overkill by introducing multiplicity adjustment in practice. The objective of phase IIb is finding the dose response curve instead of testing hypothesis.	The Opinion is clear that the multiplicity adjustment is set by the user according to the level of control they desire on the probability of an incorrect decision for dose response under a flat model (= continue developing an inefficacious drug).
	• Unless the dose response curve is non-monotonic,	This is agreed in principle.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	the 4-parameter Emax model can be quite flexible and can model a lot of different shapes defined by other parametric models. In this case, it is not necessary to pre-specify many models. Greater than 90% medicines have monotonic dose response. To accommodate that, we can put a back up model (normal dynamic or other spline types) for 4-parameter Emax just in case.	If deemed adequate from previous information, MCP-Mod can of course be used with a single model, such as the 4-parameter Emax model or just two models (such as a sigmoid Emax model and a flexible spline model). Regarding the sigmoid Emax model one has to keep in mind that 3 doses and PBO are a very common design and the sigmoid Emax model requires at least 4 active doses in order to have a chance to be fit reliably (and non-convergence will remain prevalent even with 4 doses).
	Development of Vaccines • In clinical trials for vaccines, the dose selection is a very important issue and unfortunately the choice cannot be based on PK/PD modelling. Multiple comparison methods used in a broader framework should be considered as well or just mentioned in the document when no modelling approaches are appropriate. • The Draft Qualification Opinion seems to focus one dimension (dose), however, in vaccine trials, we have more dimensions: vaccine schedule, adjuvant, more than 1 antigen (thus interaction may occur between antigens, also between antigen-adjuvant). No proposals are included to address these aspects. • In order to have a more general approach we should talk more about “groups” rather than “dose” because this is a term is specific for pharma trials and not for biologicals. Vaccines trials for instance are basic dose range studies (2 to 3 levels tested in	Development of vaccines was declared as outside scope.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	an ANOVA model) and often decisions are not based on pre-specified criteria but rather an overall view. Therefore this document provides a more scientific approach which is a very good point. • We suggest that the MCP-Mod is therefore not proposed as a standard for Phase II dose finding studies with vaccines. Other Comments • It may be worth considering the need for further documentation before MCP-Mod adoption to ensure users fully understand all of the nuances. • One aspect that isn't clear is how MCP-Mod fits some models – e.g. log-linear. When a placebo dose is present does MCP-Mod fit an offset for zero dose? How is this selected? • A series of R libraries (available R programs to apply these methods) are proposed, however we are not aware of a clear position from the Agency regarding the validation of the R package. Is SAS also a possibility? • Assumptions are made when modelling is used, however, nothing is provided in the draft document about evaluation if assumptions are not fulfilled or other alternatives when this occurs. • Phase II studies are presented here like isolated studies, however it would be useful to say something about the so-called Phase II/III trials at	The Qualification Opinion does not seek to be a text book for researchers. The submission dossier and the references contained therein provide a large body of additional information. An offset is included; see Table 6-1 in the submission dossier. Validation of software packages was out of scope. In principle, SAS is a possibility (also other packages) but we are not aware of a publicly available code as of now. The Qualification Opinion does not seek to be a text book for researchers. It is clear that technical competence, understanding and experience will be important for implementing new methodology. The applicant suggests at the design stage to run clinical trial scenario evaluations to understand the operating performance in situations where "assumptions" are not fulfilled. At the analysis stage, it is suggested to run sensitivity analyses to check the model assumptions,

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	some point.	etc. Use of MCP-Mod in trials with a confirmatory component were not presented or discussed in the scope of this procedure.
5	The goal of phase II studies, and indeed phase III studies, should be to determine the dose response relationships for multiple efficacy and safety endpoints. Only with precisely quantified relationships can drug companies and regulators determine a suitable range of doses to consider for approval. For an overview on how I see drug development, you might wish to view: http://www.youtube.com/watch?v=Iv5lR8fmm8A The current work covers a very important topic. Phase II dose response studies should be designed and analysed around dose response modelling. It is woeful that in 2013 we are discussing whether dose response modelling should be employed for dose response studies. Of course they must be. Are there idiots out there who would disagree?! The "current practice" of multiple pairwise comparisons to placebo is truly terrible. The document comments that "...that current practice is repeatedly sub-optimal and inefficient." This sentence is "too polite". To design studies to determine the dose response without consideration of dose response modelling is wholly unscientific and unethical. Put simply, no biostatistician (or ethics board) should ever sign off a phase II study unless the protocol explains the dose response models that will be employed, and the level of precision expected (based on a set of apriori estimates of the shape of the dose response). An experiment is only ethical and credible if it generates meaningful and useful data.	The comment is welcome.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	Thus clearly using dose/exposure response modelling is sensible, and hence the work presented should be fully applauded and supported. I have spent much time considering the design and analysis of clinical studies, and Phase II study designs in particular. For a non-technical introduction to the design of phase II studies, I would encourage people to view: http://www.youtube.com/watch?v=hLJAcu5yQCI For a more technical presentation, see: http://www.youtube.com/watch?v=8hKGejj-344 Given this background, I will now comment on my main criticisms of the document. a) The work frequently refers to targeting "a dose". For example, "target dose estimation" line 35, "select optimal dose" line 108 and "selecting a target dose" line 127 etc. It is misguided to think in terms of "target dose" and MED, or to focus solely on the precision of such metrics. One can think in terms of "which dose provided the best balance of efficacy and safety" in the target patient population, but this is still too simplistic, as it ignores the heterogeneity in patients which should (nearly) always lead to multiple dose levels being approved. A "target dose" or MED may be some criteria determined by the drug company relating to the desired efficacy, but this is quite pointless without having a clear and precise estimation of the safety profile across the dose range. That is, one cannot, and should not, try to	a) It is endorsed that exploratory trials are very unlikely to be definitive and hence that a target dose range may be sought and this has been amended in the Opinion.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	disassociate the efficacy and safety profiles when discussed the "target dose", or, as I prefer**, "target dose range".** Since at the end of phase II we will still only have limited precision on our dose responses for our efficacy and safety endpoints, phase III needs to be designed to improve this precision, and provide precision for AEs we may not have previously considered. **Thus we must have multiple doses in phase III as well.** In the seminar paper from 1997 ("Learning versus confirming in clinical drug development") Lewis Sheiner actually wrote about phase III studies that "a larger number of toxicity outcomes may be observed, but this is because the analysis of a confirmatory trial for toxicity is actually a learning analysis". We have now moved on from 1997, and safety is as important, if not more important, than efficacy. Thus both phase II and phase III should be designed with the precise and accurate dose/exposure response modelling central to the design and analysis for both efficacy and safety endpoints. A well designed (and analysed) "learning study" is much more valuable than any "confirming study". b) The absence of the sigmoidal Emax model considered. **Without question, the sigmoidal Emax model is the best (a priori) model for dose/exposure response modelling**. See: http://www.youtube.com/watch?v=IzaiahFy5U4 Thus the sigmoidal Emax model should, at a minimum, be clearly among the list of potential models. The simple Emax model used is too simple (Hill coefficient fixed to 1), whilst the logistic model (in the current	b) There is no dispute that a sigmoidal Emax model could be included (Table 6.1 of submission dossiers), though it is understood that the sigmoid Emax model requires at least 4 active doses in order to have a chance to be fitted reliably (fewer doses risks non-convergence).

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	parameterisation) is ugly (effect at dose 0 not equal to 0) and quite useless in drug development (drugs do not have this "shape" around the ED50). A better parameterisation of the logistic model is one which looks the same shape, but using log(dose) on the x axis. This, however, is simply a reparameterisation of the sigmoidal Emax model. It is important to point out that the sigmoidal Emax model can cover all of the monotonically increasing functions shown. Thus please add the sigmoidal Emax model to the list of potential models! c) The MCP part (contrasts) looks curious to me. In short, if we plan to consider multiple potential models, we can clearly fit each candidate model and rank them (based on multiple statistical criteria including the value of the likelihood relative to the model complexity/number of parameters, residual plots, predictive check performance etc.). **However the goal of this first step should be to simply rule out clearly wrong models**. The remaining models can be used to generate the predicted dose response (incorporating parameter uncertainty) under each model, as we can then see if important decisions are sensitive or not to model choice. Thus the presentation of multiple models is both important and scientific. However we need to think here, as using a generic tool like MCP-MOD which, a priori, sees a linear model and the sigmoidal Emax as 'equivalent' candidate models is misguided. We must use our knowledge of the endpoints and drugs to avoid doing predictions based on silly models. However the principle of presenting multiple models is a sound one, but we would clearly disagree on what weight to put on each after the event. Perhaps more importantly though, at the design stage, the construction of, for example, D optimal designs	c) The goal of the first step (MCP step) is to provide a robust signal detection assessment and to control the probability of an incorrect decision for dose response under a flat model (= continue developing an inefficacious drug). It is fully agreed that best possible understanding of pharmacology is critical and that will inform selection of candidate models across the dose range under study, but scenarios indeed exist where there is considerable model uncertainty across a dose range under study.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	across a mix of sensible models (e.g. the sigmoidal Emax) and silly (linear) should be avoided. We need to be smart in our choice of candidate models at the design stage. d) It is important for biostatisticians to recognise the importance of **pharmacokinetics** in many therapeutic areas. Thus a simple summary measure of PK exposure to the parent compound (like AUC) should replace dose. This is important at the analysis stage, but much more important at the design stage, since then dose selection in not about 'optimal doses' but rather 'optimise doses to generate optimal exposure distributions for the subsequent exposure response analyses'. Here the lowest dose and highest dose become critical, with intermediate doses only needed to (simplistically speaking) 'fill in the gaps' in the exposure distributions between the top and lowest dose. Modelling in terms of exposure will also help to 'rule out' some of the more curious models they may wish to entertain (e.g. quadratic/umbrella) since on the exposure scale (where doses yield overlapping exposure distributions) these may yield unphysiological predictions. e) It is important we move on from simple models to more informative models. Thus when we collect longitudinal data in a dose response study, it is still very common to focus on LOCF analysis. We should be designing the study for a **longitudinal dose response model**, and presenting the results across multiple models, from the most basic (e.g. LOCF) to the more advanced longitudinal and/or PK/PD models. f) The authors mention that a minimum of 3 doses with greater than a 10 fold range of doses be considered.	d) This is agreed, though the scope of this procedure was dose-response modelling. Exposure-response modelling is highlighted as further work that would be of interest to pursue. e) MCP-Mod can also be used for longitudinal dose response modelling; see Sec 3.1.2.3 of the submission dossier for a detailed exampled. f) This is endorsed. MCP-Mod can also be used for response-

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	Drug companies consistently underestimate the width of the dose range they should investigate, and it is not uncommon to see all doses doing well (or to see no effect, and then the company revisit the MAD study, and then restart Phase II at higher doses). Whilst it may be great to see efficacy across all doses, clearly the company would then need to repeat the study at lower dose levels. To avoid this, it is essential that we think in terms of **adaptive randomisation** (place new patients at doses which are most informative (not the 'best dose'), using a very large dose range. Let the accruing data be unblinded to a select cohort of individuals who can (optimally) refine the randomisation schedule such that the final data is as informative as possible for a selected set of efficacy and safety endpoints. It is simply unethical to continue to randomise patients to doses which add very little to our knowledge/understanding, when other dose levels would have been informative and useful. **That is, it is unethical to <u>not</u> look at the data.**	adaptive dose finding; see Sec 3.1.2.1 of the submission dossier for a related discussion.
6	EFPIA welcome the opportunity to comment on the qualification opinion and wish to congratulate the Applicant for the work delivered which has been successfully received by the EU regulators. It will certainly help future development since now in the public domain. We are broadly supportive of the proposed qualification opinion that MCP-Mod is an efficient statistical methodology for model-based design and analysis of	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	Phase II dose-finding studies under model uncertainty. However, MCP-Mod presents limitations including: • The MCP-Mod method relies entirely on use of empirical models. Mechanistic drug-disease models are increasingly being used to understand diseases and the impact of drugs on disease, and due to the complexity of these models they could not be used in the formal model comparison methods used by MCP-Mod. • The opinion states that this method has not yet been compared to other modelling techniques. Modelling exposure response would be a better way to understand variability in the data. • Instead of the focused towards traditional endpoints (i.e. change from baseline at end of study) but to consider open up for use of longitudinal data, which is essential for improved phase IIb efficiency • The qualification procedure does not address applicability or lack of applicability to specific therapeutic areas. For example, selection of dose levels in oncology is more likely to be based on a maximum tolerability approach. It would be useful to have comment about applicability to oncology. • The MCP-Mod method chooses the most appropriate empirical model based on statistical criteria indicating which model best fits the data. Choosing a model which best fits observational data implies that the data are considered an accurate representation of the truth, but in many cases observational data are affected by trial design and execution artefacts. Choosing a model which best fits erroneous observational data can lead to erroneous conclusions. It would be more appropriate to choose models which reflect the expected nature of the relationship between dose and response based on understanding the disease	In Sec 2.5 of the submission dossier it is stated that "Exposure-response analyses or PK-PD models are possible (if appropriate models are available) but not the purpose of this request per se" MCP-Mod has been compared to other modelling techniques (see Sec 3.2.3.2 of the submission dossier), but these comparisons are per se not the purpose of this qualification opinion. Exposure-response modelling is highlighted as work to be pursued in the future. MCP-Mod can also be used for longitudinal dose response modelling; see Sec 3.1.2.3 of the submission dossier for a detailed exampled. The principles of MCP-Mod are valid regardless of the specific therapeutic areas. However, the context of use may limit the practical value of MCP-Mod, and MTD in oncology trials appears to be a good example of this. In respect of the comments below, it is fully agreed that best possible understanding of pharmacology and disease is critical and that will inform selection of candidate models across the dose range under study, but scenarios indeed exist where there is considerable model uncertainty across a dose range under study.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	and pharmacology of the drug. • At the design stage the MCP-Mod method makes use of multiple empirical models to simulate potential relationships between dose and response. In most cases there is significant information available from earlier studies that would allow a team to select only one model. Even if such information was not available, it would be more appropriate to select the model which represents the expected dose-response relationship (which in almost all cases is an Emax model) and then incorporate uncertainty in the model parameters. Incorporation of uncertainty in model parameters in a single model that reflects known pharmacology is a more useful representation of uncertainty than using multiple empirical models. • In many cases, inverted U dose response shapes ("umbrella") are the result of experimental or observational artefacts, for example during titration to effect studies and when using endpoints that are composites of both efficacy and safety. For this reason, "umbrella" models should be used with caution when describing dose-response data. • The MCP-Mod method does not appear to make use of patient characteristics (covariates) which can have a significant impact on understanding dose-response. Having established that context, the use of modelling approaches to understand dose-response and support dose selection for studies is already known to be superior over traditional pairwise comparisons and has been standard practice in the drug development industry for many years. MCP-Mod should not be promoted through the qualification opinion as the gold-standard for approaching Phase II dose-finding studies:	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	other methods may be more appropriate depending on the situation. **Therefore, it should be made clear in the opinion that it does not preclude other statistical methodology, although not qualified by the CHMP, from being used for model-based design and analysis of Phase II dose finding studies under model uncertainty, as appropriate (see proposal below on line 184-185).**	This is clear in the Opinion.
7	The MCP- Mod. Is a good MODEL. Dose response and dose selection is a bio- assay problem. Sources of variation have to be identified at the design stage otherwise there will be errors in the analysis and interpretation becomes difficult	This is endorsed.
8	The comments concern the qualification of the procedure in question due to scientific issues and its regulatory implications. I. Lack of Scientific Progress For most drugs where monotone dose-response relationships are determined on grounds of biology and pharmacology, there is no need to consider the set of models used in the procedure. Thomas (2006) demonstrates that "the basis functions (Bretz et al., 2005) can be closely matched by the expanded Emax model, so the use of the single expanded model does not practically restrict their choice of contrasts.'' See Figure 2 of Thomas (2006). Thus, the modelling	I. It is agreed that much of the theory underpinning the proposed method is not novel, yet the use of this type of approach in regulatory submissions remains rare and hence, the fact that these sub-optimal approaches persist makes this a relevant topic for a CHMP Opinion

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	approach is essentially identical to previous work by Sheiner, Beal and Sambol (1989). The work by Thomas (2006) could have four implications. a) Sheiner et al. (1989) point out that with the sigmoid Emax model, only a single test against the null hypothesis Emax ≤ 0 is necessary. This can be achieved with a likelihood test or a triple trend test developed by Capizzi, Survill, Heyse and Malani (1992). There is no need for the MCP component of the procedure. b) In reference to CHMP Qualification Question 2 of June 11, 2013, there is no need to perform goodness-of-fit with different model selection criteria. Rather, the goal should be to fit the fully parameterized sigmoid Emax model to practical applications where the maximum likelihood estimate method could fail for practical applications (Kirby, Brain and Jones, 2011). Fitting data to a small set of selected models disguises this clear and present issue. c) Sheiner et al. (1989) demonstrate the difficulty in estimating Emin and ED50 of the sigmoid Emax model parameters (see first two lines of Table I and problems 12 and 13). With a more sophisticated PK/PD	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	modelling and simulation approach, Lockwood et al. (2003) demonstrate the difficulty with estimating the "minimum effective dose" (MED) of a certain effect size and conclude that "identification of the selected dose-response feature with any real precision from the trial design paradigm is borderline." Furthermore, "The marginal precision raises the question as to what is the best dose to study to ensure a clinical outcome of at least a one-point change in pain score, given the dosing options available." A literature review was provided to the PhRMA working group on Feb. 3, 2006. The lack of awareness of Sheiner et al. (1989) was acknowledged by a key member of the working group. Based on a PhRMA working group report, it appears that the FDA Office of Translational Science also suggested the work by Sheiner and Lockwood in May 2007. The work by Sheiner et al. (1989) was confirmed by comprehensive simulation studies of Sheiner, Hashimoto and Beal (1991). The fundamental issue with the MED or "target-dose" is that it is not an intrinsic characteristic of the dose-response relationship, since when the Emax is below a target effect, MED simply does not exist. In a 2006 Joint Statistical Meeting (JSM) presentation, it is shown that	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	the probability of a non-estimable “target-dose” can be as high as 26% based on simulations performed with a real case clinical trial application in neuropathic pain for which the target effect size is already at the plateau of the dose-response curve. The issue of MED was raised and extensively discussed in the PhRMA working group in June 2006. The probabilities of non-existing target doses were produced through simulation studies. However, neither the results nor the issue were included in the PhRMA working group report (Bornkamp et al., 2007). Without this critical information, the reported results of the PhRMA working group on estimation of MED with the modelling approach do not provide any additional scientific knowledge that was not already known prior to 2003. In addition, there is no acknowledgement of the prior work by Sheiner et al. (1989) and Lockwood et al. (2003) which already demonstrated difficulties of estimation in the report. The critical results on non-estimable “target-dose” were not reported in the request for CHMP Qualification. With regards to the CHMP Qualification Question 3, it is fairly straightforward to conclude that the approach in question requires unnecessary steps in analysis and trial planning; without providing the critical results on non-estimable “target-dose”; there lacks the additional scientific knowledge on estimability of the “target-	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	dose''. d) In reference to CHMP Qualification Question 4, any existing software for fitting sigmoid Emax models can be used. For example, the software package NONMEM developed by Stuart L. Beal and Lewis B. Sheiner in the late 1970s at UCSF for population PK/PD modeling has become the "gold standard" for both the pharmaceutical industry and academia. Liao and Liu (2009) propose a 5-parameter model used in bioassays that includes the sigmoid Emax model as a special case. Their model permits asymmetry around ED50 to better reflect the underlying biological processes. This model can be applied to model dose-response when deviations from the sigmoid Emax model occur. At the fundamental level, the statistical models for dose-response studies need to reflect vast work in the biological sciences (e.g., bioassay and clinical pharmacology). For example, with a growing number of cases where hormesis is justified on biological ground, a biphasic dose-response model, which includes the sigmoid Emax modal as a special case, can be used for design and analysis. Again, there is no need to consider seemingly different and ad-hoc models. Regulatory policies should promote fruitful research to solve real problems.	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	II. Alternative Approaches It is all too common to discover that an approved drug has a high safety risk prompting regulatory agencies to withdraw marketing approval or require black-box safety warning restrictions. Occasionally, regulatory agencies may approve a drug with questionable efficacy and require black-box safety warning restrictions to limit the drug's usage only to patients for whom other drugs have failed. Because early dose-response trials are often designed with short or intermediate treatment and follow-up, the results of early trials cannot reliably predict the phase 3 dose with optimal long-term benefit-to-risk profile. Rather, the early trials should be designed and analysed in such a way that identifies a reliable lower effective bound for effective doses, which, along with the highest safe dose, forms the full dose range for carrying forward to dose-response phase 3 clinical development programs. While the importance of phase 3 dose-response trials has not been broadly incorporated into clinical development programs by the pharmaceutical industry, the consequence of limited phase 3 dose-response information has been well- recognized by regulatory agencies. Temple (2004) states that ''The impression that dose-finding is largely completed in Phase II is a terrible error. Phase II studies almost never can detect small differences in effect, and cannot give useful	II. This description of alternative approaches is welcome, though the current Qualification procedure is limited in scope to MCP-Mod. There is no intention to discourage alternative approaches. Indeed it is hoped and anticipated that Qualification of MCP-Mod affirms the interest of drug regulators in any and all robust strategies for the exploring and understanding of dose-response, exposure-response and mechanistic modelling. Discussion of alternative approaches would be welcome.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	information on safety except for the most common events.'' Temple (2004) further suggests the need to ``study a full range of doses in phase 3 to establish dose response for both favorable and unfavorable effects and to locate less than fully effective dose that may still be useful.'' The need for phase 3 dose response trials is also clearly elaborated by Hemmings (p. 30, and pp. 46-47, 2006), who details various clinical issues and regulatory ramifications of selecting one dose for phase 3 trials. In particular, Hemmings (p. 47, 2006) states that ``Where data on outcome are required for submission, it is considered that the continuation of dose-finding into phase III would usually be highly beneficial, using phase II trials with a surrogate only to narrow the potential dose range rather than to select a single dose for the phase III study.'' In reference to CHMP Qualification Question 3, there are at least three different approaches based on the traditional step-down framework that meet the regulatory principles. a) Step-down trend test approach Quan and Capizzi (1999), following Tukey, Ciminera and Heyse (1985), apply a triple trend test in a step-down fashion to identify a no-statistical-significance-of-trend (NOSTASOT) dose for a two-way dose-response design.	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	They show that this step-down trend test is more powerful than a commonly used step-down pair-wise test. The triple trend test is also robust against various shapes of the dose-response relationship. The step-down trend test has been successfully applied for Phase II as well as phases 3 dose-response trials. This author was involved in a case where the step-down trend test was applied to the design and analysis of a phase 3 clinical program and was accepted by regulatory agencies world-wide for marketing approval. b) Adaptive step-down test with dose-response modelling approach A potential issue of relying only on the models of a triple trend test is the ability to identify an effect size of interest. For example, this can happen when there is a discrepancy between the shape of the dose-response curve and the three trend scales of Tukey, Ciminera and Heyse (1985). In a two-stage adaptive dose-response design by Liu and Pledger (2005), the sigmoid Emax model is used to fit the first stage data. The fitted model is used to derive an adaptive trend test statistic for the second stage which is then combined with the first period pair-wise test statistic. The combination trend test is applied in a step-down fashion to identify a NOSTASOT dose. By construction, their proposed adaptive design has the ability to identify an effect size	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	of interest with high probabilities over a broad range of models (see Tables 1 and 2 of Liu and Pledger, 2005). The approach hybridizes pair-wise multiple comparisons with adaptive trend test based on modelling. Prior to its publication in JASA in 2005, the work was presented at ENAR 2003, BASS X 2003 and DIA 2004. c) Lower effective bound approach It is known that a trend test is limited only to doses that are studied in a trial. For many applications, the feature that allows inference of treatment effect for doses not studied in the trial is also important. Given the knowledge of the poor performance in estimating MED as described by Lockwood et al. (2003), the lower effective bound approach using likelihood inference was also developed as a part of the research by author in the context of the PhRMA working group on dose-finding. A generalized triple trend test for broad families of distributions whose canonical link follows a nonlinear dose-response model (e.g., sigmoid Emax model) was proposed. For the triple trend test, a procedure for calculating the sample size was developed. The main goal of the research is to develop a likelihood inference on the effect of any given dose, which is not necessarily studied in the trial. The core method is a likelihood test with a set of superiority margins that can be calibrated	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	for an effect size of interest. The likelihood test leads to the lower effective bound, which, by construction, achieves a specified probability of coverage for the dose with an effect size of interest. The lower effective bound is then used to define the full dose range for phase 3 trials. As a result, the phase 3 trials are likely to contain critical information that allows assessment of dose(s) with optimal benefit-to-risk profiles. It is explained below that this clinical development strategy has a high probability of technical, regulatory and post-marketing success. Extensive simulation studies were conducted. The results reveal issues of MED as mentioned in Section I that were not previously identified by Lockwood et al. (2003). More importantly, it is shown that the lower effective bound approach yields a high probability of technical, regulatory and post-marking success (71%-75%), whereas the scheme of carrying over only the "target-dose" to phase 3 only has a low probability of success (46%-49%). The high probability of success of the lower effective bound approach is not at all surprising as the approach implements the regulatory principles described by Temple (2004) and Hemmings (p. 47, 2006). The low probability of success of the "target-dose" scheme is easily explained by the fact that the "target-dose" simply cannot be reliably identified when the "target effect" is already at the	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	plateau of the dose-response. This result was later confirmed through other real practical applications. For a case example where a drug was withdrawn from major markets due to post-marketing hepatotoxicity, the ``target-dose'' scheme has a probability of success of 37%, whereas the lower effective bound approach still holds the probability success at around 75%. The lower effective bound approach was presented at the 2006 JSM's special topic contributed session on dose-ranging organized by the PhRMA working group. A manuscript on the lower effective bound approach is under preparation. III. Conclusion The procedure in question for the CHMP Qualification does not represent meaningful scientific progress. The problem of MED estimation was known in 2003. Through the PhRMA working group research, it was revealed that the "target-dose" scheme suffers from additional identifiability problems for real practical applications. The low probability of success raises serious regulatory concerns on its application to clinical trials when in fact alternative approaches are available, well-understood, and can be easily implemented. Selected Reference	III. See I and II above. MCP-Mod is to be qualified as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. The MCP-Mod approach is efficient in the sense that it uses the available data better than traditional pairwise comparisons. This opinion does not preclude any other statistical methodology for model-based design and analysis of exploratory studies from being used.

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	Thomas, N. (2006). Hypothesis testing and Bayesian estimation using a sigmoid Emax model applied to sparse dose response designs. Journal of Biopharmaceutical Statistics 16, 657-677. Sheiner, L. B., Beal, S. L. and Sambol, N. C. (1989). Study designs for dose-ranging. Clin Pharmacol Ther. 46(1):63-77. Capizzi, T., Survill, T. T., Heyse, J. F., and Malani, H. (1992). An empirical and simulated comparison of some tests for detecting progressiveness of response with increasing doses of a compound. Biometrical Journal 34, 275-289. Kirby, S., Brain, P. and Jones, B. (2011). Fitting Emax models to clinical trial dose–response data. Pharmaceut. Statist. 10, 143–149 Lockwood PA, Cook JA, Ewy WE, Mandema JW (2003). The use of clinical trial simulation to support dose selection: application to development of a new treatment for chronic neuropathic pain. Pharm Res. 20(11):1752-9. Sheiner, L.B., Hashimoto, Y. and Beal, S.L. (1991). A simulation study comparing designs for dose ranging. Stat Med. 10(3):303-21. Liao, Z., and Liu, R. (2009). Re-parameterization of five-parameter logistic function. J. Chemometrics 23,	

Stakeholder number *(see cover page)*	General comment (if any)	Outcome (if applicable)
	248–253. Quan, H., Capizzi, T. (1999). Adjusted Regression Trend Test for a Multicenter Clinical Trial. Biometrics 55, 460-462. Liu, Q. and Pledger, W. G. (2005). Phase II and 3 combination designs to accelerate drug development. Journal of the American Statistical Association 100, 493-502. Temple (2004). The Critical Path Opportunities for Efficiency in Development. FDA Science Board Advisory Committee Meeting. Hemmings (2006). Philosophy and methodology of dose-finding – a regulatory perspective. Statistical Methods for Dose-Finding Experiments (ed. Sylvie Chevret). Wiley.	

2. Specific comments on text

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome
Lines 24-25 and Line 43 1st Bullet	6	**Comment:** On page 2, lines 24-25 states that "The…MCP-Mod…approach for dose response testing and estimation intended to enable more informative Phase II study designs…", whereas pg. 3 line 43 states that "in scope" are "Drug development stage: Phase II dose finding studies to support dose selection for Phase III". Clearly the intent is to encourage use of MCP-Mod for both the design and analysis of Phase II studies, but the bullet in line 43 could be interpreted to imply only analysis of Phase II data to support Phase 3 dose selection. This could be prevented by also identifying Phase 1 studies as "in scope". **Proposed change:** Consider editing line 43 as follows: "Drug development stage: ***Phase 1 studies to support design of Phase II studies:*** Phase II dose finding studies to support dose selection for Phase III.	It is clarified that the conclusions extend to exploratory trials investigating dose-response, traditionally this is done in Phase II trials.
Lines 49-51 2nd Bullet	6	**Comment:** "However, more broadly "dose" could be any univariate, continuous, quantitative measurement, as long as an ordering of the measurements is possible and the differences between measurements are interpretable. For example, sometimes it is possible to convert b.i.d. and o.d. regimen to a common univariate scale." • The ability to substitute concentration for dose should be explicitly stated as a possibility. This of course necessitates the ability in the Mod stage not only to	No change. Concentration-response relationships in MCP-Mod are outside the scope of this qualification. Line 52 will be amended as follows: "For example, sometimes it is possible to convert b.i.d. and o.d. regimen to a common univariate scale, by introducing additional parameter(s).

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome
		examine the concentration response relationship, but also to explore the dose concentration relationship. • Moreover it is also stated "sometimes it is possible to convert b.i.d. and o.d regimen to a common univariate scale"; it would be useful to expand on this example, i.e. how? By considering a common daily dose and assuming model parameters the same for both b.i.d and o.d.? Please clarify.	There is no “standard” way to deal with situations when there is dose and regimen. Converting o.d. dose and b.i.d. dose to total daily dose and performing a traditional dose-response analysis seems rarely adequate as a pre-specified analysis as it makes strong assumptions. In addition: Often it is of interest whether there is a difference between o.d. and b.i.d. regimen, this could no longer inferred from such a model. A slightly more complicated approach is the “regimen multiplier” approach that still makes strong assumptions, but it can sometimes be adequate. The idea is to model the dose-response curves for o.d. and b.i.d. in terms of total daily dose, but allow for one additional parameter in the bid curve. The model equation for the once daily regimen is then $\mu(x) = E_0 + E_1 f(x)$ And $\mu(x) = E_0 + E_1 f(rx)$ For the twice daily regimen, here f is a nonlinear (standardized) dose-response model, x is the total daily dose and r > 0 is the regimen multiplier, which hence adjusts the total daily dose for the bid regimen.

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome
			In case of an Emax model $$f(x) = \frac{x}{x + ED_{50}}$$ this reduces simply to assuming different ED50 parameters between o.d. and b.i.d. regimen: $$\frac{rx}{rx + ED_{50}} = \frac{x}{x + ED_{50}/r} = \frac{x}{x + \widetilde{ED}_{50}}$$
Lines 53-56 4th Bullet	6	**Comment:** "Number of doses: For the Mod step, a minimum of four distinct doses (including placebo) is required, ideally distributed over the effective range. For the MCP step (e.g. for dose response signal testing or identifying the type of plausible dose response shapes), at least three distinct doses (including placebo) are needed." It is unclear why the number of distinct doses would be different for the MCP step versus the Mod step. Wouldn't these both draw data from the same experiment and therefore have the same number of distinct doses available? The choice of the exact numbers likely depends on considered scenarios; without specifics of these scenarios the numbers seem questionable. Please clarify.	No change. These doses represent a rule of thumb. Obviously the same experiment will inform both dose MCP and Mod step.
Lines 64-65	6	**Comment:** It is readily agreed that …. Current practice is repeatedly suboptimal and inefficient". Being sub-optimal may be not	No change. The document addresses a wide range of professionals involved in drug

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome
		too bad if the suboptimal solution is quite close to the optimal one. One needs to define the meaning of optimality and what the "distance" between optimal and suboptimal solutions is. We would suggest to avoid that type of generalities.	development and healthcare. More detailed evaluations are provided in the annexes.
Lines 118-120	6	**Comment:** The Opinion states that (in the MCP component) the choice of false positive error control at 5% is arbitrary and can be varied based on the certainty the Applicant wishes to have for decision-making. We all know that the standard choice of 5% is indeed arbitrary, and likely the Applicant uses this value for illustration purposes only. This highlights an important flexibility that sponsors should be able to retain, especially in the Phase II setting, and the wording should be retained in the final version.	Agreed. No Change.
Line 138	6	**Comment:** It is stated “the parameters investigated were relevant”. What does "relevant" mean in this context? Please clarify.	No change. The document addresses a wide range of professionals involved in drug development and healthcare. More detailed evaluations are provided in the annexes.
Lines 164-165	6	**Comment (typo):** “…but also important safety or tolerability variables, which will also influence dose selection for Phase II.” This should be dose selection for Phase 3. **Proposed change (if any):** “…but also important safety or tolerability variables, which will also influence dose selection for Phase **III**.”	Agreed. Typo will be corrected.
Lines 184-185	6	**Comment:** See general comment for the rational.	Agreed.

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome
		Proposed change (if any): The following sentence should be added: "As other methods may be more appropriate depending on the situation, this opinion does not preclude any other statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty from being used"	The following statement will be added: "This opinion does not preclude any other statistical methodology for model-based design and analysis of exploratory dose finding studies from being used."
Lines 186-188	6	**Comment:** The Opinion states that "…the anticipated benefits of **a modelling approach** are demonstrated by the simulations performed…" **Proposed change (if any):** Clarify whether the intent is to state that the simulations demonstrate that modelling approaches *in general* can be beneficial in the dose-finding setting, or whether the intent is to flag the fact that MCP-Mod is beneficial. The former interpretation, although rather broad and non-specific, would be appropriate.	The statement refers to MCP-Mod. The text will be updated.
Line 61	7	Comment:"uses available data" this does not define efficiency clearly Proposed change (if any): type I and type II can be used to define efficiency	No change.
Line 67	7	Comment: "the high failure rate in Phase III" is because not all sources of variation are accounted for Proposed change (if any):	No change.

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome
Line 78	7	Comment:" to quantify dose- response relationship" homogeneity of patients has to be ensured due to some medical criteria for drawing valid inferences. Proposed change (if any):	No change.
Line 143	7	Sample size "n" can be reduced if more sources of variation are identified, improve results and interpretations. Proposed change (if any):	No change.
Line 96	7	"external factors " will induce bias, cannot ensure randomisation!	No change.
Line 107	7	Evidence of drug effect is always confounded, not easy to isolate.	No change.
Line 152	7	"Robustness" does this mean that the model is not sensitive to the underlying distribution?	Robustness in this context means a good performance across all metrics and true dose response relationships. For more detailed information please refer to response to questions.
Line 160	7	"underlying scenario" is it the underlying distribution?	Underlying scenario = true dose-response relationship. The text was updated in the opinion.
Line 162	7	"power deterioration" which means type II error is large, more sources of variation have to identified.	No Change

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome
Line 165	7	“safety and tolerability” depends on an underlying variable Say- constitution or some measure of immunity EX: bone density or tissue health	No change