22 July 2014

Submission of comments on 'Guideline on the investigation of subgroups in confirmatory clinical trials' (EMA/CHMP/539146/2013)

Comments from:

Name of organisation or individual
German Region of the International Biometric Society (IBS-DR) and German Society for Medical Informatics, Biometry and Epidemiology (GMDS)

Please note that these comments and the identity of the sender will be published unless a specific justified objection is received.

When completed, this form should be sent to the European Medicines Agency electronically, in Word format (not PDF).

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8416
E-mail info@ema.europa.eu **Website** www.ema.europa.eu

1. General comments

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	In some parts of the text (399-402, 472-475) an increase in sample size is recommended to ensure an estimate of effect that can be made with reasonable precision so that the applicant is able to substantiate therapeutic efficacy in important subgroups. If a study is advised to be adequately powered for subgroups, this seems to be not far from advice to perform separate studies. In contrast to the text passages mentioned above, the guideline explicitly states in 397-399 that for factors under point 2, it is not required that a formal proof of efficacy is available in order to conclude on effects across the breadth of the trial population. The guideline should clarify this and should give an explicit statement in the executive summary about the expected precision of effect estimates in key subgroups. The guideline should take into account potential implications of a sample size increase with regard to organization and financing of clinical trials.	
	It seems that there are in fact 3 levels of subgroup analyses (SGAs), line 385ff: Factors defined in number 1: confirmatory SGAs Factors defined in number 2: “exploratory key” SGAs Factors defined in number 3: “truly exploratory” SGAs	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	The discrimination of factors in number 2 and 3 is a crucial point. The relevance of “truly exploratory” SGAs for factors defined in number 3 (homogeneity is plausible) is unclear. A clarification with better terminology to distinguish between “exploratory key” and “truly exploratory” SGAs would be very useful. Scientific advice at the planning stage will be necessary for the decision which factors are considered key factors for exploratory subgroup analyses. The role of SGAs defined by factors for which homogeneity of treatment effects can be assumed should be clarified.	
	The guideline states its aim as maximizing a priori discussion and minimizing a posteriori discussion of the importance of subgroups in order to focus on the important subgroups, to reduce the risk of performing abundant analyses, and thus to reduce the risk of erroneous conclusions (see e.g. 71-75, 425-429, 449-450). This idea is highly appreciated and would really lead to a better interpretation of clinical trials’ results. However, throughout the guideline one gets the impression that in fact the sponsor should investigate subgroups defined by almost all baseline variables in all possible variations (see e.g. 278-279, 287-293, 311-314, 422-424, 613-616). This contradicts the declared	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	intention and, therefore, the guideline should try to be more focused on which subgroups really should be investigated and which not.	
	Sponsors can plan investigations only for expected heterogeneity. In this case sponsors may decide either to plan separate studies in more homogenous patient populations or to deliberately accept heterogeneity. In the latter case the sentence in 56-57 ”The more heterogeneous the study population, the greater the importance of subgroup analyses “ applies. However, unexpected heterogeneity will sometimes emerge (usually post-hoc). If this happens, it is much important to consider the credibility (as outlined in the respective section) of the unexpected findings. Therefore we would suggest to differentiate between unexpected and expected heterogeneity.	
	The discussion on statistical interactions should differentiate between quantitative and qualitative interactions (sometimes also called removable and non-removable interactions due to the fact that the removable interaction may only be present on a certain scale and is thus sensitive to scale transformations). Qualitative interactions are a) probably less influenced by data-driven decisions, b) may result in difficulties for a general claim and c) are more relevant. At present,	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	qualitative interaction is mentioned only once on page 13-line 510.	
	For clarity, we recommend a glossary with a definition of the most important terms. This glossary should detail (among others) the meaning of the term “factor”, since “factor” in e.g. a factorial design or some statistical software packages is limited to categorical variables only whereas here it is also applied to e.g. continuous variables.	

2. Specific comments on text

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
56-58		Comment: The text refers to a need for the estimated overall effect to be “broadly applicable.” The true issue of concern is not whether there is heterogeneity of effect but whether the effect in each of the subpopulations is sufficient for the conclusion of efficacy and adequate benefit-risk for the proposed indication. Proposed change (if any): The more heterogeneous the study population, the greater the importance of subgroup analyses to ~~check that the estimated overall effect is broadly applicable and~~ support assessment of risk-benefit across the breadth of the proposed indication.	
104		Comment: Please explain the abbreviation SmPC. Proposed change (if any): "… of the summary of products characteristic (SmPC) …"	
145-146		Comment: We think the statement “Whilst a number of the consideration outlined in this document will apply to the former …” is somewhat misleading. Most considerations outlined in this document refer to a more exploratory investigation of subgroups rather than a confirmatory testing strategy.	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		Proposed change (if any): Delete this statement or reformulate the corresponding sentence.	
174-177		Comment: "This is predominately a consideration of whether information on subgroups would be useful to the prescriber but, depending on the circumstance, criteria outlined in Section 6 may also be useful for a determination of whether the evidence generated may be considered reliable for presentation." The meaning of this sentence is unclear. Proposed change (if any): Please give a more distinct and clear wording.	
194		Comment: The definition of a subgroup should always be based upon factors that are identifiable prior to randomisation. Proposed change (if any): "These factors and the categorisation of patients **should** be identifiable prior to randomisation …"	
198		Comment: See comment on line 194 above Proposed change (if any): "Post-baseline covariates which may be affected by treatment received will **in general** not be appropriate …"	
201-208		Comment: Please add that treating continuous factors on a continuous scale by means of appropriate interaction terms has	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		statistical advantages compared to categorisation. Proposed change (if any): Please add: "… for decision making in clinical practice. **However, treating continuous factors on a continuous scale by means of appropriate interaction terms has statistical advantages compared to categorisation (see section 4.3).**"	
210		Comment: We do not think that subgroups defined by multiple factors may be of interest only on occasion. Proposed change (if any): Delete "on occasion".	
218-219		Comment: "The risk score itself may serve as a factor by which subgroups of patients may be defined in addition to a categorical factor against which response to treatment may be modelled." The meaning of this sentence is unclear. We understand that subgroups of patients can be defined by categorization of a continuous risk score. What does "in addition to a categorical factor" mean? Proposed change (if any): Please rephrase for clarification.	
220-223		Comment: "For factors where categorisation depends on a biological measure there is a risk of misclassification, in particular due to measurement or diagnostic error. Information will be needed to quantify the influence of this risk on the	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		classification of patients into subgroups and on the inferences that can reliably be made therefrom." What kind of information is expected here? Proposed change (if any): Please specify (or give an example) what kind of information on potential misclassification is expected.	
266-279		Comment: An alternative term for "interaction" is given by "effect modification". It could be helpful for some readers to make this clear. Proposed change (if any): Please add that "interaction" and "effect modification" is the same.	
278-279		Comment: "If categorising a continuous covariate, sensitivity analyses using different cut-offs should routinely be performed." It has to be noted that this further increases the risk of a type I error (as in the related situation of investigating the prognostic effect of a continuous variables by using maximally selected test statistics). If an established or biologically meaningful cut-off point exists for a continuous factor, this should be used, and it is questionable if further cut-off points must be investigated "routinely". Proposed change (if any): The sentence should be changed taking into account this comment. At least the word "routinely"	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		should be deleted.	
307		Comment: Please be specific that this statement relates to the absolute scale. Proposed change (if any): add: "... severe disease at baseline, **considering the absolute rather than relative scale**"	
310		Comment: Please be specific that this statement relates to the absolute scale. Proposed change (if any): Change "(larger)" to "(**absolute** larger)"	
311		Comment: Please be specific that this statement relates to the absolute scale. Proposed change (if any): Change "(smaller)" to "(**absolute** smaller)"	
385ff		Comment: Three kinds of factors are defined. It is the 2nd kind of factors that might cause problems, because of their numbers and because the majority of subgroup analyses will belong to it, in particular because some biological reasons or previous experience will not be difficult to be found in the literature for each of the following: stratification factors, gender, age (including cut points for dichotomisation), region, BMI, genomic	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		factors, pharmacology, stage of disease, severity, location of disease, phenotype, concomitant medication, combination of these, and more. The potentially large number makes it very difficult to follow the advice of the guideline and to plan for consistency for each of them simultaneously. A trial could soon become impractical if the advice in line 399 would be followed: "It would, however, be prudent to design the trial accordingly such that a sufficient number of patients are recruited to the subgroup to ensure an estimate of effect that can be made with reasonable precision so that the applicant is able to substantiate therapeutic efficacy and a favourable risk-benefit in important subgroups". Proposed change (if any): Multiplicity issues in relation to subgroups defined by factors belonging to category 2 should be considered and discussed in more detail. Will the agency accept some selection process in order to maintain practicability of clinical trials? Please provide some ideas what "sufficient number of patients", "reasonable precision" and "substantiation of therapeutic effect" could mean.	
386-387		Comment: The text suggests that separate trials should be done (according to level of a particular factor) when it is suspected that response to treatment may differ according to level of a particular factor. While separate subpopulation-only trials may be done, this is not the only approach and hence the guideline should not restrict the use of other valuable designs.	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		Proposed change (if any): 1. For a particular factor there is strong reason to expect a heterogeneous response to treatment across the different levels of the factor. In this case ~~separate trials should usually be planned.~~ **it is relevant to consider design and analysis strategies to enable assessment of subpopulation effects.**	
399-402 and 407-408		Comment: 399-402 says that in the planning of studies it should be ensured that subgroups are large enough to give sufficiently precise estimates. 407-408 says that patients should be recruited in a way that the epidemiology of disease is reflected. This seems to be inconsistent. Proposed change (if any): Please clarify.	
418-424		Comment: It seems that there are in fact 3 levels of subgroup analyses (SGAs): Factors defined in number 1: confirmatory SGAs Factors defined in number 2: “exploratory key” SGAs Factors defined in number 3: “truly exploratory” SGAs The discrimination of factors in number 2 and 3 is a crucial point. The relevance of “truly exploratory” SGAs for factors defined in number 3 (homogeneity is plausible) is unclear. Proposed change (if any): A clarification with better terminology to distinguish between “exploratory key” and “truly exploratory”	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		SGAs would be very useful. Scientific advice at the planning stage will be necessary for the decision which factors are considered key factors for exploratory subgroup analyses. The role of SGAs defined by factors for which homogeneity of treatment effects can be assumed should be clarified.	
422-424 and 426-427 and 449-450		Comment: With statement 422-424, the guideline requires subgroup analyses also with regard to factors for which homogeneity of response to treatment is plausible (factors covered by definition number 3). As a consequence, subgroup analyses with respect to almost all baseline variables are demanded from the sponsor. Although the idea of stating in advance the “key” subgroups is appreciated, 422-424 seems to contradict the guideline’s aim of focussing on the important subgroups, reducing the risk that abundant analyses are performed (449-450), and thus reducing the risk of erroneous conclusions (426-427). Proposed change (if any): The guideline should try to be more focused on which subgroups really should be investigated and which not.	
472-475		Comment: According to the guideline, sample size increase (above the size determined for achieving the primary study aims for the total patient population) is justified for an investigation of substantial regional differences. The particular	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		role of region is not fully understood, because substantial differences may also be expected in other key characteristics such as age or gender. Also, if a study is advised to be adequately powered for subgroups, this seems to be not far from advice to perform separate studies. In addition sample size increase for subgroups will surely increase the financial burden for sponsors with the potential consequences that in the future mainly financial strong and large companies can afford performing such trials. Proposed change (if any): If a sample size increase is recommended, the guideline should take into account and discuss the above mentioned potential consequences.	
491		Comment: Please explain the abbreviation FAS. Proposed change (if any): "… switched from the Full Analysis Set (FAS) to …"	
506-509		Comment: It would make sense to distinguish between quantitative and qualitative interaction. The existence of an irrelevant quantitative interaction is in general not very important. However, tests of qualitative interaction or relevant quantitative interaction are in general not performed in practice. We agree with the stated principle that absence of statistical significance should not be taken to imply equality or consistency. A possible solution would be the exclusion of a	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		relevant quantitative or qualitative interaction. If sponsors plan larger sample sizes for having power also for key subgroups it becomes more likely that tests of quantitative interactions will have smaller p-values even if the differences between subgroups in effect sizes are not relevant. Proposed change (if any): Please add some discussion on quantitative, qualitative, and relevant quantitative interaction.	
509		Comment: There is a typing error in “points estimates”. Proposed change (if any): Write “… point estimates …”	
523-525		Comment: It is really important to mention (as the guideline did) that unexpected heterogeneity without full consideration of other important factors (in particular the positive statement of credibility) would be a very weak basis for restricting the licence. Proposed change (if any):	
526-530		Comment: In general, methodological and statistical principles are discussed in this guideline without commenting on specific statistical methods or measures. In contrast, in this paragraph, I^2 and Q are discussed without giving a recommendation which measures should be presented.	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		Proposed change (if any): A more detailed discussion should be given or, alternatively, the paragraph should be skipped.	
547-551		Comment: A comparison of the CIs for a subgroup and the overall effect is hard to interpret because the corresponding estimates are not independent. The relevant comparison is given by the CIs of a considered subgroup and the corresponding complement. Proposed change (if any): Please add that the comparison of the CIs of a considered subgroup and the corresponding complement is easier to interpret than the comparison of the CIs for a subgroup and the overall effect.	
547-550		Comment: “For subgroups where the effect can also be estimated with reasonable precision (such that the width of the relevant confidence interval is up to approximately 2x or 3x as wide as for the overall effect) a flag for inconsistency would be an estimated effect that is outside the span of the CI for the overall effect … ” A 2 or 3 fold width of the CI will correspond to sub group sample sizes of 11-25% of the total sample size. With these small sample sizes it occurs in about 50% or more of randomly selected subgroups that at least one subgroup mean is outside the span of the 95% CI for the total population. Proposed change (if any): The guideline’s authors should	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		investigate the properties of their proposal and should reconsider their statement.	
613		Comment: Some people might have the understanding that subgroup analyses are analyses of subgroups. It is helpful to underline that the analysis of interaction terms is also a part of heterogeneity exploration. Proposed change (if any): Write "… include covariate-adjusted analyses, **analyses of appropriate interaction terms,** and subgroup analyses."	
613-616		Comment: The strategy described in section 5.1 and 5.2 already includes all possible subgroups (see comment on 422-424 above). Now, "for completeness", further exploratory analyses are requested. This contradicts the guideline's aim of focussing on the important subgroups, reducing the risk of performing abundant analyses, and thus reducing the risk of erroneous conclusions. Proposed change (if any): The guideline should try to be more focussed on which subgroups really should be investigated and which not.	
623-625		Comment: It would be helpful to describe the possible consequences of an incomplete pre-specification of key subgroups.	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		Proposed change (if any): Please add some examples of incomplete pre-specification of key subgroups.	
637-638		Comment: please avoid terms like “borderline significant” – even the term “significant” very likely makes no sense in the exploratory setting in which no significance level is usually defined. Proposed change (if any): Maybe just say some evidence (e.g. small nominal p-value)	
645-696		Comment: Annex 2 is not referenced in the description of scenario 2. Proposed change (if any): Give reference to annex 2.	
660-662		Comment: “3. Benefit in the all-randomised population is statistically and clinically persuasive, but risks and uncertainties are present in a subset of the population to the extent that a positive risk-benefit cannot be concluded in that subset.” This is described as reason for scenario 2, but it seems rather be related to scenario 1 than to scenario 2. Proposed change (if any): Delete 660-662.	
691-693		Comment: If a treatment recommendation is to be based on a	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		subgroup, it is not only mandated that benefit/risk is carefully inspected in that subgroup (by extrapolating safety data from the all-randomised population) but also that the safety data in that subgroup are considered. Proposed change (if any): Please clarify that (in addition to the safety data of the all-randomized population and possible extrapolations from there) the safety data in the considered subgroup should also be carefully inspected.	
720		Proposed change (if any): Please add: "... **of subgroup effects**"	

Please add more rows if needed.