<Date of submission>

Submission of comments on 'Adjustment for Baseline Covariates' (EMA/295050/2013)

Comments from:

Name of organisation or individual
European Federation of Statisticians in the Pharmaceutical Industry (EFSPI)

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)*
	We recommend that the scope of this document (lines 119-120) is clearly identified in the title – We suggest changing the title to 'Guideline on the adjustment for baseline covariates for clinical trials'. Refer to this scope in the executive summary	
	We have a concern with the document as it is currently written relating to the stratification of studies by centre and the advice around inclusion of centre as a covariate in the analysis which we feel might be inappropriate and unworkable. The specific issues are as follows: Lines 191-192 of Section 4.2.2 state 'The primary analysis should reflect the restriction on the randomisation implied by the stratification. For this reason, stratification variables – regardless of their prognostic value – should usually be included as covariates in the primary analysis.' This is common advice, but is not well considered. Almost all clinical trials are randomized using permuted block designs that are a 'restriction on the randomization', and a form of stratification. The advice given means nearly all clinical trials must include randomization block as a factor in their analyses. This is unfeasible and never done. Provided variation with the randomization blocks is less than or equal to that between blocks, it has been shown that ignoring the blocks in the analyses produces somewhat conservative results. This is the reason the stratification by randomization blocks is routinely and appropriately ignored in the subsequent analyses. The preceding point has a very practical implication. Section 4.1.2 lines 159-162 state 'Most multicentre trials are stratified by centre (or investigator) either for practical reasons or because centre (or investigator) is expected to be confounded with other known or unknown prognostic factors. When multicentre trials are not stratified by centre, then the reason for doing so should be explained and justified in the protocol.' While not a directive to always include centre as a stratification factor, the instruction to justify not including centre effectively makes stratification by centre the default advice. Section 4.2.3 lines 201-205 then state	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	'Adjusting for many small centres might be possible but raises analytical problems for which there is no best solution. Analyses either ignoring centres used in the randomisation or adjusting for a large number of small centres might lead to unreliable estimates of the treatment effect and *P*-values that may be either too large or too small. Furthermore, pooling small centres to form one centre of size comparable to that of other centres has little or no scientific justification.' The default is to include centre in stratification, but then adjusting for centre, or not adjusting for centre, may be inappropriate. Hence, this unworkable advice. When numerous small centres are included in the design primarily for administrative reasons (e.g., drug dispensing), without prior evidence of high between-centre variability in outcome, excluding centre from the analysis is appropriate. The rationale for this decision is exactly the same rationale as is applied to exclude the randomization blocks from the analysis. As currently written, the document strongly discourages this simple approach. Nor is it acceptable to permit an analysis excluding the large number of centres but then routinely require a sensitivity analysis based on unstable estimation methods that adjust for the numerous centres.	
	All the text relates directly to parallel group studies. Cross over studies are sometimes carried out in late phase, especially for equivalence studies in respiratory disease indications. Either the scope should specifically exclude crossover studies or a section should be added addressing the extra issues and interpreting the comments elsewhere in the document. For instance the bullet "Variables measured after randomisation and so potentially affected by the treatment should not normally be included as covariates in the primary analysis." would presumably need "randomisation" replaced by "treatment initiation". The important difference is the possible inclusion of period level baseline covariates, often the outcome measured prior to start of treatment in each period. These are measured at the end of a washout period and before treatment starts for that period. Important topics that ought to be covered in any such extended guidance include: 1) Carryover is more likely to impact any baseline covariate than an outcome measured at the end of the later period, especially when the length of period is much longer than the washout. 2) When subject is treated as a random effect the potential introduction of cross-level bias requires the use of both period-level and subject-level versions of the baseline covariate (Kenward, M.G. & Roger, J.H. The use of baseline covariates in cross-over studies. *Biostat (2010) 11 (1): 1-17*).	
	Consideration should be given further guidance in this document for methods that should be used in	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	the circumstance of a prognostic (or potentially predictive) continuous quantitative baseline covariate. If stratification is justified then information will have been lost and cut-points used to create a categorical factor may be imperfect.	
	Outcome studies with risk enrichment for baseline covariates. Risk factors (e.g. high age and prior MI) used to increase the event rate in event-driven studies are clearly judged to be related to the (composite) endpoint of interest. It may be worth commenting on whether such variables should be incorporated in the primary efficacy analysis or not.	
	It is specified line 114-115 that "A question that is often encountered is whether the adjusted or unadjusted analysis should be declared as primary in the protocol. This guidance document addresses that critical issue" However, the hierarchy (primary/sensitivity analysis) between adjusted and unadjusted analysis depending on the criteria analysed, model used (and more especially non-linear model …) is not so clear in the guidance and should be clarified.	

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)*
046-048		Add a sentence also in the summary that number of stratification variables should be limited to the most relevant to avoid empty cells. We feel that this is important enough to be briefly mentioned already in the summary.	
046-48 (and 191-194)		Consideration should be given to whether quantitative variables are continuous or categorised when used as covariates	
047-048		Comment: In addition to including stratification factors as covariates in the primary model, sometimes stratified or conditional analysis by stratification factors (eg, stratified Cox regression model or stratified logistic regression) may be conducted. Proposed change (if any): Please provide guidance on which method is recommended and/or if both are appropriate.	
047-48		Comment: When stratification is carried out for administrative reasons rather than to control variation, there is no need to include the stratification variable in the analysis model. Indeed with randomization carried out within many small centres, such additional covariate will increase rather than decrease precision while having no impact on bias. It has recently become common practice to exclude centre from the analysis model and include something more useful such as country or region with fewer and more appropriate levels. This should be reflected in the guidance.	

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): "The factors that are the basis of stratification should normally be included as covariates in the primary model." to become "The factors that are the basis of stratification should normally be included as covariates in the primary model, except where stratification was carried out purely for administrative reasons."	
053-054		Comment: As specified in the §4.4.2, in case of strong baseline imbalance in a variable, some sensitivity analyses including this variable as covariate should be provided to assess the robustness of the primary analysis. Proposed change (if any): To be consistent with §4.4.2, adding of this recommendation.	
64		We suggest the use of the word 'categorised' instead of 'dichotomised'	
070-72		Comment: It is not clear how the presentation of the treatment effects in the subgroups enables an assessment of validity of the model assumptions. Proposed change (if any): Suggest remove 'of the validity' so that the statement reads '..... an assessment of the model assumptions'	
082-083		Comment: Missing data in covariate is an important topic and is solely included in the executive summary. Proposed change (if any): Expand the discussion on this topic in 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)*
		specific section	
098		Comment: "primary variable" should be clarified. Proposed change (if any): specify this as the "primary outcome variable".	
133-134		Comment: It is stated that randomisation is expected to balance treatment groups amongst covariate levels, where as in fact we don't expect perfect balance, it is just that randomisation means there is no a-priori reason why one treatment group should be favoured by an imbalance compared to another. Suggest more nuanced wording. Proposed change (if any): Replace lines 133-134 with "The use of randomisation means that none of the treatment groups is any more likely than any other group to receive a more favourable allocation with respect to a given baseline covariate. However randomisation cannot guarantee perfect balance and it is not unusual to observe some imbalances post-hoc even if they may be purely due to chance."	
134-135		Comment: The statement ""Such imbalances are of particular concern if they favour the experimental group"" is conservative. Nevertheless, there is also a concern if imbalances favour the control since the estimation of the treatment effect is biased. Indeed, as mentioned lines 273-274, "the aim of a RCT is (…) to provide an unbiased estimate of the true difference between treatments".	

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): To remove "Such imbalances are of particular concern if they favour the experimental group".	
156-157		Comment: The downside of a large number of covariates is explained in section 4.1.1, and section 4.2.2 details the expectation on including all stratification factors as covariates. Could it be made clear whether there is a place for an important covariate in the primary analysis which has not been stratified for and that is not a baseline to the primary outcome? This is never stated as such, but one would infer that this may be acceptable in certain cases if the importance of the covariate is justifiable, but there may be concerns over including too many covariates. Proposed change (if any): At line 157: "As such it may be justifiable to include covariates in the primary analysis which have not been used as factors for stratifying the randomisation."	
159-162		Comment: "Most clinical trials are stratified by centre (or investigator)." Is this still the case? Our experience suggests centre is rarely used as a stratification variable; rather region is more commonly used. Proposed change (if any): Adjust the wording accordingly and refer to other stratification variables that may be used instead of centre.	
168-176		Advice on inclusion of the variables used in the dynamic allocation scheme in the analysis is desirable. Contrary to stratified randomization where balance is sought for each combination of level of the stratification variables, in dynamic allocation models 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)*
		allocation is performed simultaneously for the different factors. Due to this aspect, even in small trials, it may be possible to allocate according to several factors but the recommendation of inclusion of these variables in the model and the restriction on the number of covariates requires additional details. Especially in the case of small trials, it is desirable to ensure at least some balance with regard to some known prognostic factors, even if they are not included in the primary analysis (due to small number of subjects within each combination of levels of the covariates). We therefore consider that further details in the guideline will be beneficial to the readers. Proposed change (if any): Recommendations on dynamic allocation should be clarified	
181-183		Comment: It is stated that adjustment for covariates generally improves efficiency. Whilst this is true to a point in terms of a reduction in variance, the more covariates that are included or the more that are included with less evidence of prognostic effects, the more chance there is of accidental confounding with treatment. Worth pointing not to use more covariates than are needed. Proposed change (if any): The points about number of covariates are made in section 4.3.2, but maybe there is the chance to introduce that idea here and to explicitly say at the end of 4.2.1 that "Covariates with little expected association to the primary outcome variable should not be included".	
187-189		Comment: This section seems to conflict with the guidance in section 4.2.6 relating to the inclusion of baseline as a covariate in the analysis.	

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): Suggest that this section is reworded to indicate that the justification for the association between covariate and primary outcome variable is not required in the case of baseline.	
191-194		Comment: It is stated that stratification factors need to be adjusted for in the primary analysis. What are the consequences of not adjusting for stratification factors in the analysis? Proposed change (if any): Suggest adding explanation.	
195		Comment: Section 4.1.2 refers to stratifying by variables other than centre, e.g. region, when this is appropriate. Should there be an acknowledgement of stratifying by region in section 4.2.3? Proposed change (if any): Also refer to stratifying by other related variables other than centre in section 4.2.3.	
195-208		Comment: Analytical problems due to adjustment for many small centres are discussed. However, no reference is made to Random effect model with centre as random variable. Proposed change (if any): References to "Fixed effect model" and to "Random effect model" to be added	
196-200		Comment: Assume that a study, for practical reasons, have been stratified by centre. There is also one or a couple of baseline covariates known to be associated with the efficacy variable. If it is not feasible to adjust for both centre and the prognostic covariate(s), recommend clarifying in these situations which takes priority. Proposed change (if any):	

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)*
		Section 4.2.3 could mention country or region so as to be consistent with section 4.1.2.	
201-205		Comment: We agree with this. The arbitrary pooling of smaller centres is an older practice that was commonly conducted to solve the sparse centre problem. However, there was often no rationale to think that the pooled centres have anything in common other than the sparse data they contributed. But sometimes, pooling centres within country could be reasonable in a multi-national trial because of similar background conditions in the countries, eg. Similar medical practice. Proposed change (if any): Adding a sentence under what conditions a pooling of centres might be considered.	
216		Comment: Section 2.4.5: It is legitimate to use covariates measured on-treatment (after randomization) in any imputation model used to handle missing data. Proposed change (if any): Add sentence "However, post-randomization covariates, including the outcome variable itself measured at previous visits, should be considered for use in any multiple imputation models to handle missing data, either as primary or as sensitivity.".	
222-224		Comment: As stated, the adjusted treatment effect may be biased. Need to clarify what these suggested exploratory covariate or subgroup analyses are intended for. Proposed change (if any): Clarify the purpose of the suggested exploratory covariate or subgroup analyses.	

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)*
225		Comment: There is no mention of the issue of adjusting for baseline value in an analysis of percentage change from baseline resulting in a possible over adjustment. Proposed change (if any):	
234		Comment: Section 4.3: Specification of primary analysis. It has become common for primary analyses to handle the problem of early withdrawal by fitting some form of repeated measures model. The guidance should reflect this by commenting on the importance of fitting an interaction between baseline covariates and visit (time). Proposed change (if any): Proposed change (if any): Add "If a longitudinal analysis is used, for example a Gaussian multivariate linear model, then the full baseline outcome by visit interaction must be included, to avoid unrealistic constraints on the implied covariance structure of the outcomes. Also full baseline covariate by visit interaction should be included for other baseline covariates except where the impact of that covariate is likely to remain constant across visits. Severity of disease would often require a full interaction while centre or gender might simply be included as main effects."	
235, 312		Comment: Header and body text not separated. Proposed change (if any): Separate body text from header by adding hard return after header.	
237 -240		It is clearly stated that the inclusion of covariates in the primary	

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)*
		efficacy analysis has to be pre-specified. Where the state of knowledge changes it should be sufficient to document these changes in the SAP provided the SAP is signed off before database lock.	
243-244		This is too sweeping a statement. In 4.3.2 it is conceded that some models are particularly stable against even a large number of covariates. In a simple randomised experiment, the treatment variable should be independent of all baseline covariates, and even multicollinearity between different predictors, while looking ugly, does not impact the treatment effect estimate. The phrase "fewer, well-chosen" suggests that it is better to err on the side of parsimony. However, it is well known, for instance in the logistic model, that it is the omission of important covariates rather than the inclusion of ancillary covariates that may bias the treatment effect. A preference for sparse models is generally prudent, but not "in all cases".	
245-246		Comment: Examples of such non-linear models should be given (Are they, for example, logistic regression models, Poisson regression models?). For such models, a brief explanation (no more than a couple of sentences) should be presented on why the adjusted parameters and the unadjusted parameters have different interpretations. It would not be obvious to many (if not most) readers. Proposed change (if any): Suggest including additional detail as described above.	
245-248		Comment: We agree that the interpretation may be different and the hierarchy between the adjusted and unadjusted analyses may depend on the context.	

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): Recognition that that the hierarchy of between adjusted and unadjusted analyses depends on context.	
246-248		Comment: Difference in treatment effect in non-linear models, even if the covariates are perfectly balanced, has important implications for non-inferiority – as exclusion of important covariates could be a means of falsely showing non-inferiority Proposed change (if any): The relevance to non-inferiority trials is important	
248		Comment: "precisely" explaining the effect size can still be incorrect. Proposed change (if any): Replace "precisely" with "accurately"	
249-250		Comment: Suggest adding "in confirmatory analyses" following "should be avoided". This type of analyses may be useful for hypothesis-generating purposes. Proposed change (if any): "Methods that select covariates by choosing those that are most strongly associated with primary outcome (...) should be avoided in confirmatory analyses."	
257-259		Comment: These lines could be backed up by justifications rather than just saying "it is safer". Proposed change (if any): "Although the addition of covariates can in general reduce variance, a large number of covariates may increase the chance of confounding with treatment or of the model failing to	

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)*
		converge.”	
268-270		Comment: For categorical covariates with many levels, combining categories is suggested. The point could also be made here that a continuous version could be used if the variable was originally quantitative. Proposed change (if any): Add “.or continuous covariates used where possible if measures are of a quantitative nature”. This does however speak to the separate point made about stratifying by an originally continuous covariate.	
285-287		Comment: A linear relationship is mentioned. Could be clarified that this is linear on whatever scale the analysis is being carried out on as we may already be working to a multiplicative scale, i.e. after taking into account the link function in a GLM or any transformations used. Proposed change (if any): “… based on a linear relationship between covariate and outcome (on whichever additive scale is to be used),…”	
318-319		Comment: The reason that testing for baseline imbalances is inappropriate should be reiterated here. Proposed change (if any): Add again “as any observed imbalances will be a random phenomenon”.	
320-321		*Comment*: Lines 318-319 state that statistical testing is inappropriate and we agree. It is then inconsistent to refer to such test in lines 320-321.	

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 delete the last part of the sentence, i.e. ", irrespective of whether a statistical test... treatment groups."	
322-325		Comment: Need clarification what "the process of allocation ... has not been random" refers to because section 2 states non-random trials are out of the scope. Also it is not clear what appropriate actions may take. Proposed change (if any): Add clarification what is meant if the process of allocation has not been random, or remove if it refers to something outside the scope of the guideline.	
338-330		For clarity: "no interaction terms with treatment". Interaction terms among covariates are rarely employed but there is no reason to rule them out.	
347-349		Comment: If the observed interaction is particularly large, the interpretation of the overall results may become impossible. So, only the results at each level of the covariate could be interpreted. Proposed change (if any): Add that "in case of particularly large observed interaction, the interpretation of the results may only be done within each level of the covariate."	
351-354		Comment: A simple analysis of variance or covariance model is stated to have its model assumptions generally hold under quite "weak conditions." What is meant by "weak conditions"? A simple analysis of variance or covariance model is a special type of generalized linear model. Yet it is stated that mis-specification of a generalized linear	

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)*
		model could lead to incorrect estimates of the treatment effect. More clarity is needed here on this point and also on why mis-specification of a non-linear model could lead to incorrect estimates of the treatment effect. What is different about non-linear models? Don't some non-linear models belong to generalized linear models (e.g., logistic regression, Poisson regression)? Do differences in interpretation between certain models relate to marginal effects versus individual effects? Proposed change (if any): Suggest rewording.	
369-370		Comment: It could be stated that the covariate responsible for discrepancies between analyses should be discussed. Proposed change (if any):	
370-372		Comment: We suggest adding caveat that the results between adjusted and unadjusted may be different but explainable, e.g. by imbalance in influential covariate between treatment groups. Proposed change (if any): Suggest rewording to 'If the conclusions from the primary analysis and the sensitivity analyses are very different in terms of clinical and statistical significance, and that the difference cannot be explained by (for example) imbalance between treatment groups in the covariates, then the results of the trial could become inconclusive'.	
373-376		Comment: It should be explained a bit why adjusted and unadjusted treatment effects from "generalized linear models or non-linear models" may not have the same interpretation? Why might adjusted and unadjusted treatment effects be different also for generalized	

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)*
		linear models (of which simple analysis of variance or covariance models are members of)? Proposed change (if any): Suggest rewording	

Please add more rows if needed.