EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

26 February 2015
EMA/40143/2014
CHMP Biostatistics Working Party (BSWP)

Overview of comments received on 'Guideline on adjustment for baseline covariates' (EMA/CHMP/295050/2013)

Interested parties (organisations or individuals) that commented on the draft document as released for consultation.

Stakeholder no.	Name of organisation or individual
1	Alisa J. Stephens, University of Pennsylvania, USA; Eric J. Tchetgen Tchetgen, Harvard University, USA; Victor De Gruttola, Harvard University, USA.
2	European Federation of Statisticians in the Pharmaceutical Industry (EFSPI)
3	Merck Sharp & Dohme (MSD)
4	F. Hoffmann – la Roche Ltd.
5	The European Federation of Pharmaceutical Industries and Associations (EFPIA)
6	Teva Pharmaceutical Ltd
7	Ablynx (Heidi Wouters en Katrien Verschueren)
8	IDeAl (Integrated Design and AnaLysis of small population group trials) FP7 Consortium
9	SciencePharma (Poland)

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

1. General comments – overview

Stakeholder no.	General comment (if any)	Outcome (if applicable)
1	1. Introduction The proposed EMA guidelines highlight important issues surrounding covariate adjustment in the analysis and interpretation of results from randomized trials. As the guidelines state, proper covariate adjustment can enhance precision in the estimation of treatment effects; however, doing so in practice raises several important issues, including interpretation of the estimated effects, evaluation of the impact of model misspecification, and the consequences of inclusion of postrandomization covariates. In the following commentary, we discuss these issues and some recent methodological innovations that allay some concerns regarding validity of results. Our main message is that covariate adjustment is often underutilized or improperly conducted. 2. Methods to enhance robustness Intuitively, it seems reasonable that accounting for variability in outcomes by conditioning on their baseline correlates can improve the efficiency of analyses. For nonlinear models, such as logistic regression, however, it has been shown that covariate adjustment by including covariates in the regression in view can lead to loss of efficiency [1]. Additionally, for nonlinear models, a limitation to adjustment is its impact on interpretation; whereas unadjusted analyses yield estimated marginal treatment effects, adjusted analyses using standard regression yield estimated subgroup-specific effects, where subgroups are defined by the covariates included in the adjustment. Marginal and conditional effects do not generally agree, which can make interpretation challenging. A third critique of covariate adjustment for nonlinear models, also as noted in the draft	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	EMA guidelines, is the treatment effect bias that may be caused by model misspecification of the covariates included in adjustment. Augmented estimating equations introduced by [2,3] and recently summarized in [4] are a modern approach to incorporate covariates into the analysis of data from randomized trials that retains the marginal interpretation while exploiting the association between baseline covariates and outcomes to gain efficiency. These methods yield unbiased marginal treatment effect estimates even under misspecification of covariate forms in regression models. When the goal is hypothesis testing, robust standard errors may be used to construct unbiased variance estimates that preserve type I error when used in tests of treatment effects obtained through misspecified generalized linear models [5,6]. Unlike the augmentation method, hypothesis testing with robust standard errors may be performed using standard software by appealing to variance estimation options. 3. Adaptive covariate selection The current draft of the EMA guidelines takes a strong stance against adaptive covariate selection in which the data are used to determine the covariates most associated with the outcome. Although we agree that the goal of covariate adjustment is to obtain an 'unbiased estimate of the true difference between the treatments', this goal is advanced by using appropriate methods that increase precision in estimates, thereby providing greater confidence that the estimated effect is reasonably close to the true value. Large variability can result in particular data realizations that yield estimates far from the center of their distributions, even when they are unbiased. While for late phase trials (Phase III) most moderate to strong correlates of outcomes may have already been identified, the sample sizes of such trials may not permit associations with baseline covariates to be statistically confirmed. Furthermore, adaptive techniques for	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	selection of covariates as well as appropriate functional form for their incorporation in models (e.g. SuperLearner, LASSO, SCAD) may also yield covariate transformations that explain greater variability in outcomes than approaches like discretization into clinically meaningful categories or leaving covariates untransformed. Model selection may also reduce concerns that model misspecification will yield biased estimated treatment effects. Arguments against model selection in randomized trial analysis often arise from concerns over variance inflation and subsequent loss of type I error when variability of the selection process is not appropriately accounted for in analysis. In large samples, it has been suggested that augmented methods may be used with model selection to flexibly incorporate baseline covariates while preserving type I error [4] although this may only perform well if the number of candidate models is not too large relative to sample size. In trials with relatively small samples, it has been established recently, that model selection may be coupled with randomization inference to prevent type I error inflation [7]. To reduce concerns of selecting covariates expressly to produce a desirable treatment effect several strategies may be taken: 1) model building may be completed on a pooled dataset with treatment assignment information removed or 2) as suggested in [8], data may be split by treatment assignment, and the two data sets given to separate analytical teams, each of which independently builds predictive models for outcomes by treatment arm, which in turn are incorporated into an augmentation estimator. This estimator can accommodate separate covariate-adjustment models in determining a single marginal treatment effect. Both strategies maintain blinding of treatment assignments to the analysts that select covariates for adjustment. Given that outcome values may provide information about treatment assignment, the second	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	strategy has distinct advantages. A third strategy for objective covariate adjustment makes use of inverse probability weighting. Although the distribution of the treatment assignment is known in randomized studies, it may be shown that inversely weighting by the predicted probability of treatment as a function of baseline covariates maintains unbiasedness and improves efficiency over unadjusted techniques. These ideas were described in the work of [2] and may be found more recently in [9]. This approach may be taken without consideration of outcomes, solely by model building for the treatment assignment. To maximize efficiency using inverse probability weighting, analysts may consider strategy 1) to find correlates of the outcome which are then entered into a treatment assignment model that is then used to construct inverse probability weights. In summary, several methods are available to perform adaptive covariate adjustment in analyses of data from randomized studies that decouple covariate selection from treatment effect estimation and thereby preserve type I error. 4. Conclusion We support more frequent use of appropriate flexible covariate-adjusted analyses in randomized trials than the EMA guidelines. It is clear that naive adjustment can yield biased estimates and invalid tests. Nonetheless, recently developed methods that permit flexible selection yield robust, unbiased, and interpretable estimates and preserve the type I error control of tests. Such approaches may in fact be preferable to selection of a set of prespecified covariates. In the past flexible covariate adjustment may have been deemed exploratory and incapable of providing reliable conclusions regarding treatment efficacy. Perhaps contrary to popular belief, flexible selection may actually provide more interpretable results, particularly	The GL focus on confirmatory trials where important covariates should already be known.

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	when the precise nature of the specification of covariates and of the models that incorporate them is not known. For example, the paper by [10] reported the results of a study that compared the effect of monotherapy with two different drugs on mortality among patients with HIV infection; they found 100 deaths in the didanosine group and 88 in the zalcitabine group, for a relative risk of 0.78 (P = 0.09). After adjustment for baseline CD4 count, Karnofsky score, and presence of AIDS, the adjusted relative risk was 0.63 (P = 0.003). Although the paper states that the variables used in the adjustment were pre-specified, it does not provide information regarding which (if any) other variables were considered, how they were included in models, and whether any model selection was done. As a result, these findings are difficult to interpret without more explicit acknowledgement of how they were produced. Application of flexible models to these data do not require pre-specification of variables (transformed or otherwise) or model and could, even now, provide an assessment of the validity of adjusted analyses reported in the paper. With recent methodological advances the benefits of more aggressive covariate adjustment can be attained without compromising the validity of results. 5. References [1] D. Robinson and N. P. Jewell. Some surprising results about covariate adjustment in logistic regression models. International Statistical Review, 59(2):227-240, 1991. [2] J. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. Birkhäuser, 1992. [3] J. Robins and A. Rotnitzky. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association, 90:106-121, 1995.	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	[4] M. Zhang, A. A. Tsiatis, and M. Davidian. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. Biometrics, 64:707-715, 2008. [5] M. H. Gail, W. Y. Tan, and S. Piantadosi. Tests for no treatment effects in randomized clinical trials. Biometrika, pages 57-64, 1988. [6] M. Rosenblum and M.J. van der Laan. Using regression models to analyse randomized trials: Asymptotically valid hypothesis tests despite incorrectly specified models. Biometrics, 65(3):937-945, 2009. [7] A. J. Stephens, E. J. Tchetgen Tchetgen, and V. De Gruttola. Flexible covariate-adjusted exact tests of randomized treatment effects with application to a trial of HIV education. Annals of Applied Statistics, in press, 2013. [8] A. A. Tsiatis, M. Davidian, M. Zhang, and X. Lu. Covariate adjustment for two-sample treatment comparisons for randomized clinical trials: A principled yet flexible approach. Statistics in Medicine, 27:4658-4677, 2008. [9] C. Shen, X. Li, and L. Li. Inverse probability weighting for covariate adjustment in randomized studies. Statistics in Medicine, in press, 2013. [10] D. I. Abrams, A. I. Goldman, C. Launer, J. A. Korvick, J. D. Neaton, L. Crane, M. Grodesky, S. Wakeeld, K. Muth, S. Kornegay, D. L. Cohn, A. Harris, R. Luskin-Hawk, N. Markowitz, J. Sampson, M. Thompson, L. Deyton, and The Terry Beirn Community Programs for Clinical Research on AIDS. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. The New England Journal of Medicine, 330(10):657-662, 1994.	
2	We recommend that the scope of this document (lines 119-120) is	Title modified

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	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.	
2	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	**Text modified**

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	'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 '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	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	centres but then routinely require a sensitivity analysis based on unstable estimation methods that adjust for the numerous centres.	
2	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).	The GL states that a covariate that may be affected by treatment should 'not normally' be included in the primary analysis. This does not prohibit - provided a reasonable justification and assurance of type I error control & lack of bias is given - the use of covariates measured following randomisation (e.g. baseline values in cross-over trials).
2	Consideration should be given further guidance in this document for methods that should be used in the circumstance of a prognostic (or potentially predictive) continuous quantitative baseline covariate. If stratification is justified then information will have been lost and cut-	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	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.	
4	Throughout the document the word 'stratification' is sometimes referring to randomization, and other times analysis. It would be helpful to be clear throughout which is being referenced. (e.g. "Stratified randomization" vs "stratified analysis").	
5	We welcome the publication of the draft guideline on adjustment of baseline covariates (EMA/295959/2013) and the opportunity to comment on the document. The revision leads to an improved standard of regulatory assessment of confirmatory trials and improves the planning of confirmatory trials by sponsors. The guideline acknowledges that sensitivity analyses are important (for example, to understand the impact of additional covariates not included in the primary analysis). Even an appropriately stratified	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	trial may produce important baseline imbalances by chance. Sensitivity analyses in general should be leveraged to support the conclusions of the primary analyses, but protocols should not be written so as to suggest that all sensitivity analyses should produce similar results. Differences may be observed and these should be appropriately discussed. The guideline stresses that adjusted and unadjusted analyses need to be compared with the expectation that the difference is small. The difficulty with this view is that if covariates are relevant there may be a non-trivial difference between the adjusted analysis and unadjusted analysis, and results will be sensitive to including or omitting covariates. This should not automatically lead to the conclusion that the analysis is not reliable. Another aspect that needs to be further addressed is the regulator's interpretation of a covariate adjusted analysis. For example, does a covariate adjusted analysis provide estimates for the effect of treatment in an individual (for a given set of covariates) or for the effect in a patient population with a similar distribution of covariates as those in the study?	
5	It is stated several times that the primary model should not include treatment by covariate interactions. Instead the trial should be designed to allow separate effect estimates in subgroups. The latter can be achieved by including an interaction term in the model for simple categorical covariates. In the case of continuous covariates, an interaction term would prevent categorisation of the covariate which can affect the analysis as well. It will be helpful if this aspect could be addressed in the final guideline.	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
3	In general factors used for stratification are included as covariates in the model. When the stratification variable is a categorisation of a continuous baseline covariate, general guidance on how the variable can be included in the analysis would be useful. Further guidance on methods that could be used in the circumstance of a prognostic (or potentially predictive) continuous quantitative baseline covariate would also be useful. The dictum that stratification variables should be included regardless of their prognostic value (192-194) seems to be at odds with the assertions that the number of covariates should be limited (257-261). More guidance on how to balance these two positions would be helpful. In some cases transformation of covariates may be required. This could impact the analysis and interpretation. Further details on suitability and interpretation of baseline covariates in non-linear models and in particular Cox regression models will be helpful (appreciating these models are more complex in terms of their structural form and link functions). Some guidance on the applicability for adjusting covariates in cross-over trials would be beneficial. For 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 is worth commenting on whether such variables should be incorporated in the primary efficacy analysis or not.	**Not in the scope of the GL** **Use minimal number of stratification factors** **Not in the scope of the GL**

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	It would be helpful to clarify the distinction between design considerations and analysis considerations in this guideline, and to discuss the link between the two, at the outset. For example, stratifying by site may be an example of an exception to the rule that stratification variables should be included in the analysis model; it is meaningful and may be feasible to stratify randomisation by site, but it may not make sense or be infeasible to include site in the analysis. We welcome the fact that the guideline acknowledges circumstances in which dynamic allocation may be appropriate and useful. 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 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.	
5	(Editorial) The guideline does not cover non-randomized trials, observational studies etc. We recommend amending the title for the guideline to “Guideline on adjustment for baseline covariates in clinical trials”.	Title modified
6	We would like to know the view of the Agency on the use of propensity models, taking into account many covariates, as a	Not in the scope of the GL

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	sensitivity analysis?	
8	We welcome this statement. It yields a very good guideline for the adjustment on covariates. We especially appreciate the claim for justification of every covariate that is included in the study, as well as keeping the total amount of covariates as low as possible by elimination of dependent covariates.	Not clear which statement meant
8	We welcome that baseline covariates can be accounted for in two stages of a clinical trial, the randomization and/or the analysis. In particular, we appreciate the explanation, that stratified randomization is the typical approach for handling baseline covariables in the randomization process.	
8	We regard it as dangerous, though, to use oversimplified models for the primary analysis. From Senn [2005, 2012] it is known, that all relevant observed baseline covariates must be included in the primary analysis of the study. The credibility of the trial is not compromised by many covariates if their relevance is explained in the study protocol.	The GL focus on confirmatory trials where important covariates should already be known.
8	We recommend the use of a suitable randomization procedure to diminish the increased effects of (selection) bias that might arise due to many strata.	
8	We support the claim, that post-hoc testing for baseline-covariates should be avoided in randomized clinical trials if randomization and blinding are properly conducted in the study.	
8	The agency should carefully elaborate on the way the term “imbalance” is used throughout the whole document for covariables as the meaning of the term “imbalance” is twofold and its interpretation is related to the related main objective: If the interest of covariates is as a main effect, that is to say to adjust the treatment effect for them, it is the degree of imbalance of the covariate between treatment arms that adversely affects power.	

Stakeholder no.	General comment (if any)	Outcome (if applicable)
	If the interest is in the covariate as in interaction, then in addition one needs (for a categorical covariate) that each category is well-represented. However, this requirement adversely affects recruitment time and is usually impractical.	
8	The need for randomization and blinding to avoid bias in clinical trials cannot be overstated. In particular, knowing important covariate measurement, may lead to strong selection bias in trials, where the person who recruits the patients is not blinded to previous treatment allocations. We would therefore recommend to further stress this point in the statement.	

2. Specific comments on text

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
40-43	5	Comments: Adjusting for covariates may not only increase efficiency, but also avoiding conditional bias resulting from chance covariate imbalance. This could be clarified in the text. Proposed change (if any): “Although baseline adjustment is not always necessary, […] generally improves the efficiency of the analysis and avoids conditional bias from chance covariate imbalance.”	Accepted.
41	5	Comments: The definition for baseline covariates could be provided here. There are good definitions on the next pages (line 97-99).	Not accepted. Not considered necessary in the executive summary
46-48	2	Comments: 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.	Not accepted: The key message is the restriction of the number of stratification variables (for whatever reason)
46-48 191-194	2	Comments: Consideration should be given to whether quantitative variables are continuous or categorised when used as covariates.	Not accepted This depends on the situation and cannot be solved generally
46-48	7	Comments: Here is also mentioned that you should correct for stratification factor by including the factor into the model. Is the stratification using STRATA in SAS then a kind of subgroup analysis? Is this allowed when looking at this paragraph because this procedure	Accepted: Wording modified

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		is not the same as a simple inclusion of the covariate? Proposed change (if any): Add more information on these different ways to correct for stratification factors and also an advice on what should be done in different situations.	
47	5	Comments: Medical practice can differ across regions, and "country" can be used as a stratification factor. This is not solely for "admin" reason.	Accepted Wording modified
47-48	5	Comments: The guideline states that stratification variables should usually be included as covariates in the primary analysis (47-48). We are wondering whether “Included as covariates” refers only to statistical models with treatment and covariates as main effects or whether it is meant as a term for general models including a stratified model with treatment as only main effect and random strata as stratification variables. A clarification concerning stratified models would be useful. With one categorical stratification factor there is no major difference expected between the stratified and adjusted analysis (included as main effect). In case of an adjusted Cox model there is the additional assumption of proportional hazards which has to be fulfilled by the covariate (as indicated in line 66 of the draft guideline). Thus, the stratified approach offers adjusting for potential prognostic factors without necessity to check additional model assumptions.	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		When there is more than one factor the stratified and the adjusted model differ as the stratified model accounts for the combination of all levels of the factors (i.e. interaction). In contrast the inclusion of the two variables as covariates (main effects) in the model does not. This would require including an interaction term for both covariates. Thus, the adjusted analysis without interaction term does not reflect the nature of the stratified randomisation. Proposed change (if any): Please provide guidance on which method of stratified analysis and analyses including stratification factors as covariates is recommended and/or if both are appropriate. Amend the existing to text as “The factors that are the basis of stratification should normally be included as covariates or as stratification variables in the primary model.”	
47-48	2	Comments: In addition to including stratification factors as covariates in the primary model, sometimes stratified or conditional analysis by stratification factors (e.g. 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.	Accepted
47-48	2	Comments: When stratification is carried out for administrative reasons rather than to control variation, there is no need to include the stratification variable in	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		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. 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.”	
53-54	2	Comments: 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.	Accepted
57-59	5	Comments: There is no explanation why a continuous baseline should be included as a covariate mainly in ordinary linear models and not in non-linear models. We propose to include a wider class of models into this proposal or to provide some explanation of why this	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		advice cannot be generalized beyond linear models.	
57-59	7	Comments: It is mentioned that in either case you should add the baseline covariate, even if you use a change from baseline? Is this not an overcorrection? Proposed change (if any): Can this be more motivated why the correction is still necessary even when you take the change from baseline as an outcome. What is most correct? Is it not better to concentrate on raw outcome when baseline is included? Why?	Not accepted No overcorrection
64	5	Comments: "...dichotomising a continuous scale..." In some cases it is appropriate to classify a continuous variable into more than two groups. Proposed change (if any): "...categorising a continuous scale".	Accepted
54	2	Comments: We suggest the use of the word 'categorised' instead of 'dichotomised'.	Accepted
70-72	2	Comments: 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'.	Accepted
70-72 63-65 283-288	8	Comments: We have some doubts regarding the statement, where "appropriate categorization of covariables" or simple functional forms for the relationship are mentioned. A loss of information	Not accepted. According to the text, categorization should only be done if appropriate.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		results from categorization and possible erroneous relationships may result in biased treatment estimates. See lines 277-283. Proposed change (if any): We recommend that categorization or linearization of continuous covariables should not be done, apart from the case where well established clinical categorizations are used, meaning that the relationship of the categories to the treatment estimate are established. For exploratory analysis categorized analysis – if in agreement with the results of the primary analysis – may be helpful in interpretation of the data in relevant subgroups.	
77	4	Comments: A loss of efficiency will occur by following this guidance in situations where the covariate modulates the treatment but not the control. For example, predictive biomarkers being tested with new treatments in phase II (Mackey and Bengtsson 2013, Contemporary Clinical Trials.)	Not accepted. The GL is main concerned with confirmatory clinical trial (ususall not phase II trials)
82-83	5	Comments: The reader is referred to the guideline on missing data for direction on dealing with missing baseline covariates. However, this guideline primarily addresses the case of missing response variables. If a baseline variable is missing, options are to impute or delete the observation, or, if the variable of interest is the baseline value of the response variable, one could fit the baseline observation as part of the response vector in a repeated measures analysis, thus accounting for the missingness as part of the response.	Partly accepted. The GL is not intended to provide guidance on missing values, the aim is to mention the problem of missing covariate values. The text has been sligthly modified to point out that the GL on missing values provide examples to deal with missing values (but other approaches are also possible).

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): More explicit commentary on handling missing baseline covariates in general would be beneficial; relevant literature citations would help.	
82-83	2	Comments: 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 specific section.	Accepted
83	4	Comments: Please refer to the specific guidance document.	Not accepted. The pertinent GL is mentioned in section 3 (Legal basis and relevant Guidelines)
88-89 168-176	8	Comments: The term “dynamic allocation” could be misleading, because there exist procedures without randomisation element, e.g. minimisation method, Pocock-Simon range method with p=1. In our opinion only “stratified randomization methods” or “baseline adaptive randomization methods” with a true random element should be recommended. Methods without random elements should be avoided. Proposed change (if any): We recommend to use the term “appropriate random allocation”.	Not accepted. Line 171 already states ‘deterministic schemes should be avoided’
98	5	Comments: “primary variable” should be clarified. Proposed change (if any): specify this as the “primary outcome of the study”.	Accepted
98	2	Comments: “primary variable” should be clarified. Proposed change (if any): specify this as the “primary	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		outcome variable".	
103	5	Comments: Change to "primary outcome measure is also considered as a covariate".	Accepted
108-111	5	Comments: The paragraph mentions mainly regression methods. Covariate adjustment can also be implemented through stratification (e.g., for the logrank test). Therefore, we propose addition of stratification as an adjustment method.	Accepted
109-111	6	Comments: We suggest adding the reference to Poisson regression for count data.	Not accepted The list provides only examples and is not intended as a complete enumeration of all possibilities
114-115	5	Comments: "... A question that is often encountered is whether the adjusted or unadjusted analysis should be declared as primary in the protocol..." The terms "adjusted analysis" and "unadjusted analysis" are ambiguous and further clarification will be helpful. Alternatively, an appendix to include definition of terminologies will be useful.	Not accepted. The text considered to sufficiently clear
119	4	Comments: The recommendations in this guidance document appear to be exclusively meant for confirmatory clinical trials rather than just "mostly concerned". Proposed change (if any): It would help to state that many of the recommendations discourage methods that while not acceptable for a primary analysis of a Phase III trial are in fact entirely reasonable approaches in	Not accepted While the emphasis of the GL is with confirmatory clinical trials, the principles laid down in this GL are applicable to all clinical trials.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		early phase trials when one is learning about the effect of the drug.	
119-120	6	Comments: Should it be mentioned in the executive summary that the guideline is mostly concerned with confirmatory trials?	Not accepted While the emphasis of the GL is with confirmatory clinical trials, the principles laid down in this GL are applicable to all clinical trials.
133-134	2, 5	Comments: It is stated that randomisation is expected to balance treatment groups amongst covariate levels, whereas 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."	Accepted Inserted that balance is only to be expected on average.
134-135	2	Comments: 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".	Not accepted The conservative approach is favoured

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): To remove “Such imbalances are of particular concern if they favour the experimental group”.	
141-143	5	Comments: Clarify the following: stratifying randomisation simply reduces the chance of covariate imbalances between treatment groups; it’s the inclusion of stratification variables (or other relevant covariates) in the analysis model that improves the efficiency of the estimation. Proposed change (if any): Reword the bullet point as follows: “Stratifying randomisation reduces the chance of covariate imbalances between treatment groups, and the inclusion of stratification variables in the analysis model may improve the efficiency of the estimation of the treatment effect, …”	Partly accepted. Wording modified
143	5	Comments: Please consider if the wording "complementary" is clear enough. Our concern is that this wording could be interpreted as if stratification without adjusting for the stratification in the analysis is suggested.	Not accepted. ‘Complementary’ considered clear enough in the given context
146	5	Comments: This phrase is contradicting with the 1st statement of this bullet point. We don't expect the consistency between strata for a given stratification factor. Suggest change this phrase to "examine the differences and similarity across these subgroups".	Not accepted Wording considered sufficiently clear
147-148, 309-	8	Proposed change (if any): Please add references of the	Not accepted.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
310		relevant regulatory documents. (E.g., it is not clear which guidelines on subgroup analysis are meant.)	A draft GL on subgroup analysis has just been published for consultation.
150-157	4	Comments: The section on Dynamic Allocation (4.1.3) should follow directly from this paragraph as a strategy to address balancing margins for several covariates. We are pleased to see that dynamic allocation is now presented as an option for ensuring balance across many covariates and no longer discouraged by EMA. We agree with the authors that deterministic schemes should be avoided.	Accepted
151	5	Comments: Change “many” to "multiple".	Not accepted
156-157	2,5	Comments: 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.”	Partly accepted Wording in 4.2.2 modified
158-162	4	Comments: We are also surprised by the statement	Partly accepted.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		that one needs to justify not stratifying randomization by center. We have seen it as the norm not to stratify by site. Open-label trials with randomization stratified by site introduce predictability of treatment assignment for final patient in each block of a permuted block design.	Wording modified.
159-162	2	Comments: “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): Proposed change (if any): Adjust the wording accordingly and refer to other stratification variables that may be used instead of centre.	Partly accepted Wording modified
159-167	5	Comments: With the common use of IV/WRS systems, it is no longer true that most multicentre trials are stratified by centre (or investigator). We therefore propose to modify this section and delete the text "When multicentre trials are not stratified by centre, then the reason for doing so should be explained and justified in the protocol."	Partly accepted Wording modified
160	5	Comments: This could be stated as "because many known or unknown variability among centers that can impact on the outcome of clinical trials".	Partly accepted Wording modified
163	5	Comments: To stratify the randomization by centre does not require all centres to be small. It will be more	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		appropriate to use the word many centres instead. Proposed change (if any): replace “each” with “many”: “When the number of patients within each many centres is expected to be very small […] by centre.	
168-176	2	Comments: 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 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.	Not accepted. Dynamic randomisation is not considered to be the randomisation method of choice
170-172	6	Comments: The use of deterministic dynamic randomization schemes might influence the p-value and bias. In line 170 it is said that deterministic schemes should be avoided. In light of that, we would	Not accepted It is clearly stated that type I error control and possible bias should be addressed by the Sponsor

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		like to have clarification what is meant by the possible implication on p-value and bias and what is expected from the Sponsor.	
171-175	3	Comments: The sentence “Deterministic schemes should be avoided and possible implications of dynamic allocation methods on the analysis e.g. with regard to bias and type I error control should be carefully considered, taking into account that for some situations (e.g. planned unbalanced treatment allocation) it is has been shown that these methods might impact the validity of conventional statistical methods” implies that it had been shown that for dynamic allocation with planned unbalanced allocation the validity of conventional statistical methods might be impacted. This statement is likely based on the paper by Proschan M, Brittain E, and Kammerman L. “Minimize the use of minimization with unequal allocation” (Biometrics 2011; 67: 1135 – 41), where the authors show that with certain versions of unequal allocation minimization as well as unequal allocation expansion of the biased coin randomization (a non-dynamic allocation procedure), the unconditional randomization distribution of a test statistics is shifted away from 0. This shift in the randomization distribution causes low power of the randomization test and problems when interpreting study results. However, as demonstrated in Kuznetsova OM and Tymofyeyev Y. “Preserving the allocation ratio at every allocation with biased coin randomization and	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		minimization in studies with unequal allocation". Stat Med 2012; 31 (8): 701 – 23, the shift in unconditional randomization distribution is not peculiar to minimization or dynamic allocation procedures in general. Instead, it is common to all unequal allocation procedures for which the allocation ratio varies from allocation to allocation – as was the case in the examples considered by Proschan M, Brittain E, and Kammerman L. To avoid this problem, equal allocation procedures, dynamic or non-dynamic ones (including such common procedures as biased coin allocation, urn models, maximal procedure) should be expanded to unequal allocation in a way that preserves the allocation ratio at every allocation. In this case, the shift in the randomization distribution converges to 0 as the sample size increases and is typically negligible in studies of moderate size (for dynamic as well as non-dynamic allocation procedures). Additionally, Proschan et al. considered examples of unequal allocation where variations in allocation probabilities were confounded with a temporal trend so that one treatment had a higher probability to be assigned at the positions where patients were healthier. As a result, the Type I error of the Z-test was inflated. This type I error inflation is also a direct result of the variations in the allocation ratio and would not happen with unequal allocation procedures that preserve the allocation ratio at every allocation. Overall, variations in the allocation ratio from allocation to allocation cause the same problems for both non-	

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		dynamic and dynamic unequal allocation procedures (potential for selection bias and evaluation bias even in double-blind studies; accidental bias associated with the time trend; shift in re-randomization distribution) and should be avoided with both non-dynamic and dynamic procedures. Thus, we suggest to add a qualifier “unbalanced treatment allocation that does not preserve the allocation ratio at every allocation” in the sentence above. Proposed change (if any): We suggest to add underlined text to the sentence below: “Deterministic schemes should be avoided and possible implications of dynamic allocation methods on the analysis e.g. with regard to bias and type I error control should be carefully considered, taking into account that for some situations (e.g. planned unbalanced treatment allocation that does not preserve the allocation ratio at every allocation) it has been shown that these methods might impact the validity of conventional statistical methods.”	
175	6	Comments: We would like to have clarification what is meant by re-randomization. If this refers to sampling methods like the bootstrap, we suggest not to use re-randomization since in some indications patients are being randomized several times (for example in Chron’s disease patients are randomized at entrance to induction study and then re-randomized at entrance to maintenance study).	Not accepted. It is clear from the context that here ‘re-randomization methods’ relate to the analysis

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
175-176	8	Comments: The authors of this comment consider the term “re-randomisation” too vague. Proposed change (if any): Specify what is meant by re-randomisation, e.g. permutation testing.	Not accepted. The text refers to re-randomization methods in the analysis, several methods are possible, specific guidance is outside the scope of the GL
175-176	9	Comments: It is not clear whether the re-randomization should be performed during analysis of the results or prior to clinical study commencement. Moreover, there are no details of the methods of re-randomization given. This issue should be clarified and the methods of re-randomization described in details, especially as re-randomization is not addressed in ICH Topic E9 Statistical Principles for Clinical Trials (CPMP/ICH/363/96).	Not accepted The text refers to re-randomization methods in the analysis Furthermore, providing information on specific analysis methods is not in the scope of the GL
176	4	Comments: Different randomization techniques (permuted block versus dynamic) do different things: Blocks balance all cells while dynamic balances margins. Recommend a comment about the level of balance for dynamic randomization being reflected in the analysis.	Not accepted. This is already to be found in 4.2.2
180	5	Comments: “...The main reason to include a covariate in the analysis of a trial is the existence of strong or moderate...” The above statement is worded strongly. Proposed change (if any): “... the main reason to include a covariate in the analysis of a trial is evidence of the existence of strong or moderate...”.	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
181-183	2, 5	Comments: 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”.	Not accepted. Section 4.2.1 deals with the reasons to identify variables as covariates
187-189	2	Comments: 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. 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.	Not accepted. Text not considered sending conflicting messages
188	5	Comments: It may not be appropriate to evaluate known or expected associations with primary outcome variable from current trials unless they are open-label. Apparent associations may be confounded with treatment effects.	Accepted
191-196	4	Comments: This implies that a trial where	Not accepted.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		randomization is stratified by center must also always stratify the analysis by center. In trials with a large number of centers where many centers contribute only a few patients, many times the analysis is not stratified by center as this will result in considerable loss of power. This is not an uncommon occurrence in Phase II.	Not stratifying by centre in the situation mentioned in the comment is not excluded by the GL text.
191-194	2	Comments: 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.	Not accepted. There is no general answer.
191-194	5	Comments: Most randomisations are restricted by randomising in blocks, usually with a fixed block size. Please clarify whether you mean that the block should be included as a covariate in the model. To our knowledge, this is not the current standard.	Accepted
191-193 338-339	5	Comments: As the “primary analysis should reflect the restriction on the randomisation implied by the stratification” and “stratification variables ... should be included as covariates in the primary analysis” (191-193) the approach for two or more variables should be clarified. The statement in lines 338 and 339 could also be adjusted accordingly. Proposed change (if any): 191-193	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		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, either as main factors with interaction terms or as stratification variable. 338-339 The primary analysis should include only the covariates pre-specified in the protocol and no treatment interaction terms.	
193-194	5	Comments: Baseline covariates are observed/ measured before randomisation. The value of the covariate used for stratified randomisation might be identified later as being incorrect. Using the value of the covariate as used for the randomisation would follow the ITT principle. Using the correct value could be considered as a sensitivity analysis. Proposed change (if any): Any mismatch of covariates between randomisation and case report forms must be explained and justified. Sensitivity analyses should be performed.	Not accepted. Section 4.2.2 does not deal with a mismatch on an individual basis but with the situation that the stratification at randomisation differs from the stratification variables in the analysis.
195	2	Comments: 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	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		other related variables other than centre in section 4.2.3.	
195-208	2	Comments: 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.	Not accepted. The wording is quite general not excluding random effects models
195-208	5	Comments: This section discusses adjusting for centre when centre is a stratification factor in randomization. What is the recommendation with centre when it is not a stratification factor in randomization? Proposed change (if any): Add any recommendations for analysis when centre is not a stratification factor in randomization. Harmonize 4.2.3 and 4.1.2.	Partly accepted Wording modified
196	5	Comments: Whether centre can be adjusted for as a fixed effect or as a random effect is not discussed. Proposed change (if any): Consider referring to different modelling approaches, such as fixed effects or random effects to model centre effect as an acceptable strategy for dealing with small centres.	Not accepted. The wording is quite general not excluding random effects models
196-200	2	Comments: 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	Partly accepted Wording in 4.2.3 modified

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		priority. Proposed change (if any): Section 4.2.3 could mention country or region so as to be consistent with section 4.1.2.	
196-200	5	Comments: Assume that a study, for practical reasons, has been stratified by centre. There could also be 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), consider recommending in these situations which covariates takes priority.	Partly accepted Wording in 4.2.3 modified
201-205	2	Comments: 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, e.g. Similar medical practice. Proposed change (if any): Adding a sentence under what conditions a pooling of centres might be considered.	Accepted
211	5	Comments: "... a pronounced baseline imbalance is not expected a priori in a randomised trial: if the randomisation process has worked correctly, any	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		observed imbalance must always be a random phenomenon...” “... must always...” is rather strong. Proposed change (if any): “... process has worked correctly, any observed imbalance is likely to be a random phenomenon...”	
216	2	Comments: 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.”	Not accepted. The recent wording of the GL does not prohibit, if properly justified, the use of such variables
222-224	2	Comments: 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.	Not accepted.
222-224	5	Comments: 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 whether the	Not accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		suggested exploratory covariate or subgroup analyses are undertaken to evaluate the treatment effect.	
225	2	Comments: 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.	Partly accepted Wording modified
225-233	8	Comments: We welcome the comments on “Change from baseline”. However, in our opinon, a remark on stratification should be given and on reflecting baselines in the randomization process. Proposed change (if any): If baseline values were used as covariates, the measurement scale should be preserved. Consequently, a categorization is not recommended. Further, baseline value could be incorporated in the randomization procedure by using a covariate adaptive randomization procedures, where it is strongly recommended that methods without randomization element (e.g. minimization) are to be avoided.	Not accepted. Whether and how to include baseline into the randomisation should be decided on a case by case basis.
226-233	5	Comments: Change from baseline can be defined in several ways, for example as difference or percentage change (ratio). This may have implications on the adjustment for baseline covariates. It would be helpful if this could be further discussed.	Partly accepted Wording modified
234	2	Comments: Section 4.3: Specification of primary analysis. It has become common for primary analyses to handle	Not accepted. Outside the scope of the GL (better Missing value GL)

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		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): 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	2, 5	Comments: Header and body text not separated. Proposed change (if any): Separate body text from header by adding hard return after header.	Accepted
237 -240	2	Comments: It is clearly stated that the inclusion of covariates in the primary 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.	Accepted
240	5	Comments: Depending on the feasibility and timing for	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		making a protocol amendment, changes may have to be documented in a statistical analysis plan signed off before database unblinding. Including this as an alternative option where a protocol amendment is not able to be completed would be useful.	
243-244	2	Comments: 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”.	Partly accepted Wording modified
243-244 257-261	8	Comments: The authors along with the IDeAl consortium consider the statement that “In any cases, analyses including many covariates will always be less convincing than analyses with fewer, well-chosen, covariates.” misleading, as all relevant covariates must be included, even though they were many. (See further S. Senn, “Baseline Balance and Valid Statistical Analyses: Common Misunderstandings”, appeared in Applied Clinical Trials, 2005).	Not accepted In a confirmatory clinical trial there should be only a limited number of relevant covariables and parsimonious models are preferred.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): Avoid this implication.	
245-246	2	Comments: 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.	Not accepted Providing examples is outside the scope of the GL
245-248	2	Comments: We agree that the interpretation may be different and the hierarchy between the adjusted and unadjusted analyses may depend on the context. Proposed change (if any): Recognition that that the hierarchy of between adjusted and unadjusted analyses depends on context.	Not accepted. It is already stated that in such a situation the meaning of the adjusted effect sizes is explained.
246-248	2, 5	Comments: 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.	Not accepted. To a certain extent this is also true for superiority trials and already considered in the GL text by pointing out that in such a situation an accurate explanation of the estimated effect size is necessary.
248	2, 5	Comments: “precisely” explaining the effect size can still be incorrect.	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): Replace “precisely” with “accurately”.	
249-250	2, 5	Comments: 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.”	Accepted
257-259	2, 5	Comments: 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 converge.”	Not accepted The requested statement is provided with the next sentence in the GL: ‘Results based on such a model are more likely to be numerically stable, the assumptions underpinning the statistical model are easier to validate and generalisability of the results may be improved.’
263	5	Comments: Since only limited number of covariates can be included in the statistical analysis model, collinearity of covariates should also be considered in determining the analysis model.	Not accepted Already mentioned in the next sentence: ‘Potential covariates are often strongly correlated and so knowledge of the correlation can be a useful basis for eliminating some stratification variables at the planning stage’
268-270	2, 5	Comments: 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	Not accepted. This is not prohibited by the recent GL text

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		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.	
275	5	Comments: It will be useful to discuss, in addition to function forms of the covariates, how to handle covariates in the analysis. For example, some covariates are usually treated as random effect factor such as centre.	Not accepted Outside the scope of the GL
284-285	5	Comments: Delete either "often" or "most".	Accepted
285-287	2, 5	Comments: 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),…"	Accepted
289	5	Comments: Do you mean "subsequent"?	Accepted
294-297	5	Comments: Any regression model makes assumptions about the relationship between dependent and independent variables. However, it may not so much depend on distributional assumptions (e.g., on the error term). We understand that it is often difficult to obtain treatment effect estimators from these models.	Not accepted. It is already stated that treatment estimates should be provided

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): “Nonparametric regression models may be considered in case of uncertainties about the distribution of the data or the error terms in the model.”	
295-297	8	Comments: The authors along with the IDeAl consortium consider the randomization procedure should be reflected in the nonparametric regression as well. Proposed change (if any): However, in these cases, it is important that the randomization procedure is reflected in the model and appropriate estimates of the size of the treatment effect are still attainable and, not just the calculation of significance levels.	Accepted as already stated in the text
312	5	Comments: Insert line break after "General considerations".	Accepted
312-316	5	Comments: Recommend noting that except for trials only stratified by centre, results should also be presented by stratum when a stratified design, i.e. stratified randomization, has been used	Partly accepted. Modification in section 4.4.3
312-316	5	Comments: The wording “alternative methods may be equally valid” in case of ambiguously specified analyses and in case of difficulty to understand the adjustment for covariates. Whether other analyses may be equally valid does not depend on pre-specification. Therefore, please clarify the wording in this paragraph.	Accepted
318-319	2, 5	Comments: The reason that testing for baseline	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		imbalances is inappropriate should be reiterated here. Proposed change (if any): Add again “as any observed imbalances will be a random phenomenon”.	
320-321322-325	2, 5	Comments: 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. Proposed change (if any): Please delete the last part of the sentence, i.e. “, irrespective of whether a statistical test... treatment groups.”	Accepted
	2, 5	Comments: 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.	Accepted
332-333	7	Comments: If there is a strong imbalance and an adjustment is not sufficient. Is this then a reason to not adjust? Proposed change (if any): Should it not be stressed that the adjustment is necessary even if it is not sufficient?	Not accepted. As mentioned such imbalances would raise questions with regard to the reliability of the (study) results
338-330	2	Comments: For clarity: “no interaction terms with treatment”. Interaction terms among covariates are rarely employed but there is no reason to rule them out.	Accepted

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
338	5	Comments: Subgroup analysis can help to identify treatment by covariate interaction.	Accepted
340-343	7	Comments: Why can you not conclude that an interaction is relevant based on a significant test? Proposed change (if any): Is it possible to add an example that illustrates this?	Not accepted. As with any statistical test one has to differentiate between statistical significance and relevance
342-344/345	5	Comments: Lines 342 and 344/5 seem contradictory as written. Change “or” to “and” on line 344 to avoid this contradiction. Also, the term “interaction” in this context should be replaced by “treatment by covariate interaction”, to clarify intent. Proposed change (if any): If some treatment by covariate interactions turn out to be large from a clinical point of view or and significant from a statistical point of view, this provides evidence that the effect of treatment may vary across subgroups.	Not accepted There might be clinical relevant interactions that are not statistically significant. This situation would not be covered in case of substituting ‘or’ by ‘and’.
347-349	2	Comments: 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): 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.	Not accepted. The point is that with such interactions the interpretation on a study level might be impossible (this does not exclude the possibility for an interpretation on the various levels of that covariate)
351-352	5	Comments: “Model assumptions hold under quite weak	Not accepted.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		conditions" is not easy to understand since assumptions hold or do not hold. Please add clarification on the above statement. Is it meant that analysis of covariance is considered robust to deviations from normality, homoscedasticity etc.?	
351-354	2	Comments: 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 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.	Partly accepted. Text modified
351-359	5	Comments: If considered within scope of this guidance, recommend the covariance structure used also be commented upon since this section discusses the appropriateness of a pre-defined model?	Not accepted. Not in the scope of the GL
355	5	Comments: To what extent do we need to validate the	Not accepted.

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		model? (e.g. cox regression)	It is already stated that model assumptions should be checked
369-370	2	Comments: It could be stated that the covariate responsible for discrepancies between analyses should be discussed.	Not accepted The GL already asks for a discussion of strong differences in the conclusions from adjusted and unadjusted analyses
369-370	5	Comments: It could be stated that the covariate responsible for discrepancies between analyses should be discussed. Proposed change (if any): Conversely, if there are strong discrepancies between the conclusions drawn from adjusted and unadjusted analyses, these should be discussed and interpreted whenever possible and the particular covariate responsible for these discrepancies should be described.	Not accepted The GL already asks for a discussion of strong differences in the conclusions from adjusted and unadjusted analyses
370-372	2	Comments: 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'.	Accepted
373-376	2	Comments: It should be explained a bit why adjusted and unadjusted treatment effects from "generalized	Not accepted. Outside the scope of the GL

Line no.	Stakeholder no.	Comment and rationale; proposed changes	Outcome
		linear models or non-linear models" may not have the same interpretation? Why might adjusted and unadjusted treatment effects be different also for generalized linear models (of which simple analysis of variance or covariance models are members of)? Proposed change (if any): Suggest rewording.	
365-367	7	Comments: Estimates of the treatment effect should be the same when having the covariate included or not for linear models. This should ideally be the case but what if (baseline) covariates are confounding factors? Proposed change (if any): Is it possible to add guidelines with respect to confounding covariates?	Not accepted. Not in the scope of the GL. In confirmatory clinical trials (the focus of this GL) important covariates are considered to be known from previous trials, literature etc
373-375	9	Comments: Unlike in the case of ordinary linear model, there are no examples of sensitivity analyses for generalised linear models or non-linear models. The examples of sensitivity analyses for different statistical models should be presented to facilitate the choice of the analysis that fits best the model.	Not accepted. Not in the scope of the GL