1. General comments

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	The title of the guidance is „Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics with Continuous Outcomes" and it is stated "continuous endpoints that are appropriate for analysis with normal-theory methods". However, the guidance only covers ANCOVA and no other methods for the analysis of continuous endpoints e.g. for longitudinal data like Mixed-Effect Models Repeated Measurements or Random Slope and Intercept Models commonly used in clinical trials is discussed.	
	A guidance for non-continuous settings would also be welcome. For example, would recently proposed adjusted estimators for such settings such as those in https://onlinelibrary.wiley.com/doi/full/10.1002/sim.6507 be acceptable? For non-continuous endpoints, there are models which lead to biased estimates (Cox proportional hazard model, logistic regression) if important prognostic factors are not included. Is there any recommendation for such models?	
	Could the agency comment on the risks of high leverage observations, typically outlier observations for the covariate, having a high influence on the fitted	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	relationships for covariates and the role of transforming variables as a mitigation strategy for this, for example taking logarithms of gene expression or many laboratory measures. This is related to the existing statements on selecting the appropriate functional form, but has a different emphasis.	
	Could the agency comment on the appropriateness of flexible regression methods e.g. splines if fully pre-specified. For example, would it be acceptable to say we will include the covariate age as a spline function with 4 degrees of freedom.	
	Depending on the countries or geographic regions, concomitant medications & so on, specified in the inclusion/exclusion criteria of a trial, the sample might not be fully representative of the target population. In such cases, the new guidance could assist by specifying the suitability of using baseline variables to adjust inferences so that they might extend to the target population.	
	ICH E9 says that "When the potential value of an adjustment is in doubt, it is often advisable to nominate the unadjusted analysis as the one for primary attention, the adjusted analysis being supportive." Can the agency	

Stakeholder number *(To be completed by the Agency)*	General comment (if any)	Outcome (if applicable) *(To be completed by the Agency)*
	comment on how they see the priority of an adjusted vs. an adjusted analysis as primary? Is there a preference how many covariates should be included (few vs. many at least as general statement) and is there any guidance on the choice, e.g. continuous vs. categorical covariates?	
	In relation to the above comment, should there be a clear recommendation that any analyses should adjust for any stratification factors in the study design?	
	The problem of missing data on covariates has not been addressed. Bearing in mind the guidance is not prescriptive, would it still be useful to include something in the recommendations to state that the planned analysis should address how missing covariate data will be handled since complete case analysis will lead to a reduction in power, and exclusion of covariates with missing data could cause deviation from the prespecified analysis?	

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)*
24		Using "population" to refer to both target populations and samples has led to much confusion in designing, analysing and interpreting clinical trials. More recently, it has complicated and clouded discussions about estimands, too. We suggest, therefore, replacing "the population studied" with "the sample studied." ICH E9, referred to elsewhere in this draft guidance, avoids this confusion by using "analysis set" to refer to the sample.	
25		Suggest replacing "prognostic" with "prognostic and predictive." The differences between these are explained & illustrated in this paper: (https://www.sciencedirect.com/science/article/pii/S1574789107001020). Similar comments apply to subsequent uses of "prognostic."	
26		In many fields, including pharmaceutical development, the interpretation of statistical tests has been hampered by conflating hypothesis testing and significance testing. These two types of testing have different purposes, formulations, and interpretations. "Power" is a concept of hypothesis tests, not applicable to significance testing [which concerns only the null (tested) hypothesis]. For more explanation, see www.perfendo.org/docs/BayesProbability/5.3_GoodmanAnnInt	

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)*
		Med99all.pdf. We suggest, therefore, replacing "significance" with "hypothesis," in this sentence and elsewhere.	
37-43		Here it would be a good place to expand the definition of ANCOVA to include MMRM and other methods, or alternatively just refer to the practice of covariate adjustment in models, rather than the ANCOVA model itself.	
45-47		An ANCOVA may also reduce bias in estimate of difference between treatment groups in case of larger baseline imbalance. Suggested wording change: "… use ANCOVA to adjust for differences between treatment groups in relevant baseline variables to improve the power of significance tests and to reduce bias and increase the precision of estimates of treatment effect."	
45-47		"Sponsors can use ANCOVA to adjust for differences between treatment groups in relevant baseline variables to improve the power of significance tests and the precision of estimates of treatment effect." It is not clear if the recommendation is to include relevant baseline variables in the model. The only reason for adjustment stated are (random) treatment group differences at baseline – which the sponsor becomes aware of only after database lock. It suggests that adjustment is only for	

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)*
		precision and potential baseline differences. However, in longitudinal models adjustment for covariates is also important for other reasons. Covariates which are associated with the continuous endpoint of interest and drop-out better account for the impact of missing data as the MAR assumption is more likely valid.	
60-65		The guidance mandates pre-specification of covariates and functional form. However, covariate adjustment that yield the desired gains in efficiency and that allow covariate relationships to be identified and exploited while circumventing the usual concerns have also been proposed (e.g. https://www.ncbi.nlm.nih.gov/pubmed/17960577). Would such approaches also be acceptable?	
60-65		In contrast to this guidance, the corresponding EMA guidance (https://lnkd.in/gvPVaKU) states: "Alternative analyses should always be presented to confirm that the conclusions of the study are not sensitive to the choice of covariates included or the choice of the relationship between covariates and outcome that has been assumed. Findings based on these sensitivity analyses should normally be considered exploratory but necessary to support the primary analysis ". Presumably, this difference in approach between regions is deliberate?	
60-65		There is a (relatively small?) risk of loss of study power should	

Line number(s) of the relevant text *(e.g. Lines 20-23)*	Stakeholder number *(To be completed by the Agency)*	Comment and rationale; proposed changes *(If changes to the wording are suggested, they should be highlighted using 'track changes')*	Outcome *(To be completed by the Agency)*
		the model adjust for a non-prognostic covariate ('The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies', Kahan, BC et al, Trials, 2014; 15:139) – does this deserve a mention?	
60-65		Would it be worth making readers aware that once a primary model is defined, then the primary evaluation of treatment effect will be based on that model regardless of results? Therefore, further model refinement and the consequent results would be a discussion matter with the agency, with the implication that complexity and consequence should be taken into consideration for the primary model. The context here is whether a specific covariate and/or interaction term should be in or out of the primary model, and the extent to which that choice is results-driven.	
68-73		Maybe give a reference to ICH E9 here: In most cases, however, subgroup or interaction analyses are exploratory and should be clearly identified as such; they should explore the uniformity of any treatment effects found overall. Suggested wording change: "Therefore, even though a primary analysis showing an overall treatment effect remains valid, differential effects in subgroups can also be important. In most cases subgroup or interaction analyses should be additional exploratory analyses, as also specified by ICH E9."	

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)*
68-73		Would it be fair to consider that, from a conceptual point of view, the inclusion of an interaction term means that there is a belief that the effect of the treatment may depend on the value of the covariate(s), therefore the concept of 'overall' effect becomes unclear. One might even say that any overall effect does not exist or that there are several overall effects depending on the weights applied to the effect associated with each value of the covariate(s). From a practical point of view, a paper by Chuang- Stein and Tong ("The impact of parametrization on the interpretation of the main effect in the presence of an interaction", Drug Inf. Journal 1996, 30: 421-424) showed how the overall treatment effect may change significantly just through changing the coding of the covariate(s). This may be especially relevant when the model includes at least one categorical factor.	
74-76		In our experience, it is more common practice that when the primary outcome measure is change from baseline, the analysis should always also adjust for baseline. Is the last sentence understated and, if so, perhaps it should be removed? Or, perhaps we should state: "As baseline value and change from baseline value are often highly correlated, it is generally recommended to adjust for the baseline value in an analysis of change from baseline."	