Submission of comments on Prognostic Covariate Adjustment (PROCOVA™) with responses

Comments from:

Name of organisation or individual
1. **La Roche Ltd. - F.Hoffmann**
2. **Pfizer**
3. **MSD**
4. **International Society for Clinical Biostatistics, ISCB - Statistics in Regulatory Affairs Subcommittee**
5. **Junfeng Wang - PhD Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University** (the comments only represent the opinions from the submitter himself, not on behalf of his organisation)
6. **EuropaBio**
7. **Novartis**
8. **Kelly Van Lancker, Ph.D. (Johns Hopkins Bloomberg School of Public Health) and Prof. Stijn Vansteelandt (Ghent University)**
9. **EULAR (Margreet Kloppenburg)**

Please note that these comments and the identity of the sender are public unless a specific justified objection is received.

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000 An agency of the European Union

1. General comments

Stakeholder number	General comment	Outcome (if applicable)
1	We appreciate the applicant for raising this important topic to the public, as well as the EMA for the thoughtful and thorough response. The draft qualification opinion comprehensively covers all the critical considerations from the statistical and analytical perspectives. The methodology being discussed is potentially useful and could help run efficient clinical trials. We agree with the EMA that PROCOVA is a special case of ANCOVA, which was invented by Fisher in the 1930s and its properties have been thoroughly studied at great depth since then, and its benefits widely recognised. The concept of using a risk score predictive of outcome as a covariate is also not novel, for example in the analysis of studies of Diffuse Large B-cell Lymphoma (DLBCL), the International Prognostic Index (IPI) is commonly used as a covariate. Against this background, we suggest to introduce in the executive summary of the document a statement (for example, between LINE 60 and 61) to make its use more generalizable, such as: "The topics and considerations discussed in this qualification opinion are applicable to similar products providing predictive covariate adjustment in various disease areas."	The comment is acknowledged and agreed. We agree that prognostic modelling and risk score derivation is in principle also applicable to different indications and we appreciate the example. Nevertheless, it is not our policy to amend the background information document from the Applicant. No action taken.
1	Rigorous and trustworthy covariate adjustment methods have been extensively studied and are useful for increasing the power of clinical trials. In this digital age, big data efforts are widespread across healthcare in general and the drug development industry in particular. Considerable resources have	The comment is acknowledged and partly agreed. The updated text touches therapeutic areas, where we state that we cannot say however that the approach will be equally useful for all areas.

Stakeholder number	General comment	Outcome (if applicable)
	been spent on data curation with the explicit goal of improving clinical trial design and analysis. Using big data for optimal covariate adjustment, e.g. via selection of adjustment factors or development of prognostic scores for standard statistical adjustment, is a straightforward and obvious application. In particular, the three step process described in the briefing document has already been used in industry for other trials. In addition to providing this qualification opinion and in order to ensure consistency, we believe it would be beneficial to update the existing EMA guidance "Guideline on adjustment for baseline covariates in clinical trials (EMA/CHMP/295050/2013)" to include thorough and comprehensive considerations on covariate adjustment (see draft FDA guidance "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products" as an example).	We agree that the procedure described to formalise the steps involved has been used before. The recommendation to update the baseline covariates guideline is well taken and appropriate. Updates of a guideline is however out of scope of this qualification opinion and is a different project. EMA will consider scientific progress along the general review of GL and will update consequently. No action taken.
2	The Applicant sufficiently demonstrates that PROCOVA improves traditional ANCOVA methods by constructing the optimal adjustment covariate. The Applicant shows that that PROCOVA is an appropriate method that produces unbiased treatment effect estimates, reduces the variance of treatment effect estimates, and controls the type-I error rate. This in contrast to other statistical approaches outlined in p. 6 that may increase power, but do not control the type-I error rate, and thus PROCOVA is an improvement over previous methods. It is valuable that the Applicant adequately proves that the choice or performance of the prognostic model does not lead to biased estimates, and that this methodology creates robust results.	The comment is acknowledged and agreed. The intention of this qualification procedure is not to single out a specific method for statistical modelling. A statement to express this is added to the qualification opinion.

Stakeholder number	General comment	Outcome (if applicable)
	The Applicant sufficiently demonstrates that trial power will be increased with adjustment for the PROCOVA score, leading to a smaller minimum sample size needed to achieve a desired level of power. This will reduce randomized control trial (RCT) timelines and costs, creating more efficient and ethical trials with smaller control groups. In general, it is a good idea using trial subjects' predicted outcomes on placebo (prognostic scores) to reduce variance of treatment response estimated and thus reduce sample size. However, its relative merits compared to other strategies such as Bayesian methods using historical data as priors or various propensity score methods to enhance control arms should be studied.	
3	Comment (Prognostic scores): Practical difficulties in calibrating the prognostic scores across sponsors: If an adjustment is driven by modeling arbitrary external data, then it means different sponsors can build different models and lead to different conclusions, even by borrowing information from the identical external data to analyze the same trial data. Proposed change (if any): The proposal lacks solutions on enforcing common standards to ensure the inherent reproducibility of the prognostic modeling procedure.	The comment is acknowledged and partly agreed. This observation pertains to all types of model-based analysis that involves external data. Even without external data, sponsors could choose different statistical models and come to different conclusions. The approach to prognostic modelling is out of scope of the qualification and no approach to prognostic modelling is qualified. No action taken.
3	Comment (Prognostic scores): Risk of accidental or intentional p-hacking:	The comment is acknowledged but not fully shared.

Stakeholder number	General comment	Outcome (if applicable)
	Since the prognostic score model is subjectively built and selected without constraints, there is a risk where multiple models are built but one model is selected only due to better adjustment results in the actual analysis.	We disagree in case 'actual analysis' pertains to the estimating the treatment effect from the completed phase 2/3 study as in Step 3 of the procedure. The importance of pre-specification including the prognostic score is emphasized and even mentioned as potential drawback in the answer to question 6. For prognostic model development, preference for external validation is expressed in the opinion to avoid too optimistic estimation of a correlation coefficient. The approach to prognostic modelling is out of scope of the qualification and no approach to prognostic modelling is qualified. No action taken.
3	Comment (Model generalizability): It is at least questionable if the scores built from the external data can be generalized to the trial population, not to mention how to properly evaluate it.	The comment and stated problem is acknowledged. The problem of external validity is addressed in the answer to question 5 and opinion statement. No action taken.
3	Comment (Low interpretability models can be problematic for clinical applications): The prognostic score from certain types of models built from the first stage can be hard to interpret and justify clinically. If so, this is not helpful for addressing unmet medical needs, demonstrating the treatment superiority, and negotiating reimbursement policies.	The comment is acknowledged and agreed. Potential advantage of interpretable models is mentioned in the answer to question 6.

Stakeholder number	General comment	Outcome (if applicable)
		The approach to prognostic modelling is out of scope of the qualification and no approach to prognostic modelling is qualified. No action taken.
3	Comment (Model robustness to adversarial examples): Research has shown that there are ways to construct models that can change the predicted outcomes of any input with only a slight perturbation, while hiding the mechanism to be undetected by any computationally-bounded observers (https://doi.org/10.48550/arXiv.2204.06974). This poses a realistic challenge on certifying the authenticity and robustness of the predictive models described in the first stage.	The comment and stated problem is acknowledged. The approach to prognostic modelling is out of scope of the qualification and no approach to prognostic modelling is qualified. No action taken.
3	Comment ((Stratification): By the time a phase 3 trial is planned with continuous responses, adjusting for the baseline response and any known prognostic factor (possibly used for stratification) delivers an estimated treatment effect with notably more precision than an unstratified analysis. It is unrealistic to expect additional gains in precision, if any, by adding an additional 'composite' covariate X established using machine learning tools applied to prior data, especially since X will likely include the baseline response and stratification factor(s).	The comment is acknowledged and partly agreed. The issue is addressed in the answer to question 5 and the procedure qualified involves as first step that an attainable advantage over using ANCOVA with individual covariate adjustment should be justified. No action taken.
3	Comment (randomization): Simulations conducted were restricted to 1:1 randomization. Proposed change (if any):	The comment is acknowledged and agreed. The issue of unequal group sizes is addressed in the answer to question 4.

Stakeholder number	General comment	Outcome (if applicable)
	Results for imbalanced (e.g., 2:1) randomization (see link) suggest that, unless a robust/sandwich variance estimator is used (NOT typically done in practice), there can be issues with type 1 error and/or bias inflation.	No action taken.
3	Comment (interactions): In most cases there will be true treatment by covariate interactions. Given this, excluding interaction terms can needlessly take away the benefits of covariate-adjustment. Proposed change (if any): Include interaction terms.	The comment is acknowledged and partly agreed. According to the Guideline on adjustment for baseline covariates in clinical trials (EMA/CHMP/295050/2013) the primary model should not include treatment-by- covariate interactions. If substantial interactions are expected a priori, this should be considered for trial design. Models with treatment-by-covariate interactions are out of scope of this qualification procedure. No action taken.
3	Comment (Composite covariate): PROCOVA involves developing a 'composite' covariate X with prior data for covariate adjustment in a future RCT. The authors talk about using real world data, genomics information, etc., to build X. Even if a suitable X can be developed, its use in the future RCT is a huge challenge because it requires that all the inputs of X are measured and available for use in the RCT; this is impractical/unrealistic.	The comment is acknowledged and partly agreed. The answer to question 1 states that the variables used by the prognostic model must be measured at baseline for subjects in the historical data set and the new clinical trial. It is not agreed that this is an unrealistic scenario. No action taken.
3	Comment (Composite covariate):	The comment is acknowledged and agreed.

Stakeholder number	General comment	Outcome (if applicable)
	Use of a composite X (or any) covariate will not change the estimand with continuous endpoints but that is not the case with time-to-event endpoints, so any suggestion that PROCOVA can be used in a similar manner and with similar 'advantages' for survival analysis is problematic. Proposed change (if any):	The opinion it is stated approaches with non-linear models for analysis and direct comparisons to such models are out of scope of this qualification procedure. No action taken.
3	Comment (Composite covariate): What seems to be different is that the covariate is not identified as a single variable that would be assessed for each trial subject at baseline such as gender or age, but rather is a composite score that can be a complex function of baseline attributes. The availability of diverse large databases means that the information about factors that can influence the control group response can in principle be obtained from prior randomized trials, observational or real world data bases, biomarker studies, genomic studies, etc. The development of nonlinear regression methods such as random forests, various machine learning algorithms such as deep learning and neural networks, and sophisticated AI techniques means only that there are many more ways to construct potential predictor scores to use as covariates.	The comment is acknowledged and agreed. Various options for prognostic model development exist. The approach to prognostic modelling is out of scope of the qualification and no approach to prognostic modelling is qualified. No action taken.
3	Comment (Composite covariate): If a composite score is constructed as some function of patient characteristics that are observable at baseline, then the rule for constructing this score must be fixed. For example, if the score is a weighted sum of characteristics, then the weights must be fixed, and not changed to reflect some differences between the observation values that were used to obtain the weights and the observations in the control group of the current study. This seems especially	The comment is acknowledged and agreed. The need for pre-specification is addressed in the answers to questions 2 and 6. The prognostic score must be pre-specified including a scale factor, and weights used within the score cannot be adjusted to possible differences between the training setting and the actual trial setting.

Stakeholder number	General comment	Outcome (if applicable)
	important if part of the analysis of the data from the completed trial is an evaluation of treatment by covariate interaction. Proposed change (if any): Rule for constructing composite score must be fixed.	No action taken.
3	Comment: The key issue seems to be how one should construct a prognostic score, not how the score should be used for evaluating treatments or whether there is anything remarkable about how the actual data analysis should proceed. Proposed change (if any): One should use conventional data mining principles, i.e., use a training set to develop the predictor and a validation set to confirm that it makes sense. All of this can be done before the trial starts.	The comment is acknowledged and partly agreed. Preference for external validation is expressed in the opinion and the prognostic model needs to be pre-specified. No action taken.
3	Comment: In general, RCT Control populations can be quite dynamic across trials and even how baseline prognostic factors are measured can be subjective (such as disease status per investigator assessments). However, this approach generally controls alpha-levels under model misspecification and the unbounded number of ways that the historical data could NOT be representative of the actual trial (selection biases).	The comment is acknowledged and agreed. The problem of external validity is addressed in the answer to question 5 and opinion statement. No action taken.
3	Comment (Stratification): While there could be some scenarios where this type of adjustment can gain efficiency – it seems that basic (pre-specified) covariate adjustment based on	The comment is acknowledged and agreed

Stakeholder number	General comment	Outcome (if applicable)
	actual trial data could achieve relatively the same level of efficiency but without the "selection bias" issues of incorporating historical data. Usually randomization is stratified based on established prognostic factors, which may go a long way in achieving some of the potential efficiency gains.	The procedure qualified involves as first step that an attainable advantage over using ANCOVA with single covariate adjustment should be justified. See e.g. the answer to question 5. No action taken.
3	Comment: There could be more done with covariate adjustment. Proposed change (if any): To lean towards the methods being developed in the causal-inference area or even just standard modeling … but that the adjustments are based on the actual trial data.	The comment is acknowledged. A statement was added that the qualification does not intend to single out a specific method for statistical modelling.
3	Comment (International harmonization): Recent statistical methods that received FDA/EMA qualification or fit-for-purpose designation include MCP-Mod and BOIN. If we regard those as the standard, the proposed method is below the bar in terms of innovation.	The comment is acknowledged and partly agreed. Level of innovation is not the only criterion for qualification and the opinion acknowledges that the proposed procedure can be considered an acceptable formal presentation of approaches that were used in clinical trial settings before when prognostic covariates were included in analysis models. We amended our qualification opinion to express that we do not qualify a method as 'the' best, but as an 'acceptable' one.
3	Comment:	The comment is acknowledged and partly agreed.

Stakeholder number	General comment	Outcome (if applicable)
	Step 1 of the method cannot be generalized and must be handled on a case-by-case basis. it is difficult to anticipate how the scientific community can benefit from the designation.	It is agreed that model development can only be handled on a case-by-case basis and the approach to prognostic modelling is out of scope of the qualification. However, prognostic models could help understanding disease characteristics or even mechanistic properties. No action taken.
3	Comment (Type I error control and historical data): "Type 1 error rate can be controlled with historical data" – that would be a major breakthrough. However, this seems to be assumed from the property of ANCOVA. The problem is, adding a "score" in the regression model could change the interpretation of the treatment effect parameter unless causal assumptions are made, as is the case with propensity score. Such things are not discussed in the proposal.	The comment is acknowledged and agreed. Approaches to borrowing from historical data are briefly discussed in the background information submitted by the Applicant. The background information document from the Applicant is not to be amended. No action taken.
3	Comment (Type I error control and historical data): Even with proper causal assumptions, the interpretation of the treatment effect parameter will change in non-linear models (e.g., binary outcome). This is discussed in causal inference literature. Therefore, the proposed method cannot be generated beyond the continuous outcome where a linear model is employed.	The comment is acknowledged and agreed. Approaches with non-linear models for analysis and direct comparisons to such models are out of scope of this qualification procedure. No action taken.
3	Comment (Type I error control and historical data):	The comment is acknowledged and partly agreed.

Stakeholder number	General comment	Outcome (if applicable)
	The difference in patient population between historical and current trials is concerning.	The problem of external validity is addressed in the answer to question 5 and opinion statement. No action taken.
3	Comment (Type I error control and historical data): There are two primary ways that historical data can be useful for trial design and data evaluation: 1. Provide a plausible estimate of residual variance so as to avoid underpowering a trial because of an overly optimistic guesstimate of residual variability 2. Adjust for covariates to reduce noise (what a conventional covariance analysis does) The proposed approach addresses the second of these.	The comment is acknowledged and partly agreed. The approach to sample size estimation differs from a standard approach for ANCOVA and Step 2, accounting for the prognostic score while estimating the sample size required for a prospective study, is discussed in the opinion. No action taken.
4	Generally, this is a great job. Having gone through the handbook and the additional references, my major concerns and questions/issues/advices about PROCOVA™ have been addressed. Well done. PROCOVA appears to be a product of a particular company, UNLEARN. There is a blurring between the EMA and this company, e.g., EMA should not be saying (line 9) "Our proposed statistical methodology …".	The comment is acknowledged and appreciated. We agree that a clearer distinction between the background information and statements from the Applicant and EMA/CHMP is necessary. The format of the document was revised to clearly indicate those sections that were included in the Applicant's submission, and those authored by EMA.
4	EMA should not be advocating a specific approach over others. The document reads like marketing.	The comment is acknowledged and agreed.

Stakeholder number	General comment	Outcome (if applicable)
		The intention of this qualification procedure is not to single out a specific method for statistical modelling, but to qualify acceptability of a method. A statement to express that it is not the intention of this qualification procedure to single out a specific method for statistical modelling was added to the qualification opinion.
4	The proposal is very simple and therefore not novel from a methodological point of view. We would be surprised if it hasn't been done before.	The comment is acknowledged and agreed. We agree and discussed this in the first part of the answer to question 4. No action taken.
4	A number of qualifications of the proposed methodology need to be made more forcefully 1. It assumes that the prognostic index in the trial is the same as (or proportional to) that in the historical data. 2. It is likely to lead to improvements over adjustment for multiple covariates only in the case of small sample sizes (where adding multiple parameters to the model can be problematic). 3. Why use predictions from historical data rather than overfit to the sample data? This is the approach that the classic paper by Tsiatis et al. used and they did have some optimality results. (https://onlinelibrary.wiley.com/doi/10.1002/sim.3113) 4. It seems practically far fetched to suggest that you could get a good prognostic score by using e.g., control arms from old trials. Work on sample size for development and validation by Richard Riley, Maarten van	The comments are acknowledged and partly agreed. (1) Differences between historical data and future trial have to be taken into account and problem of external validity is addressed in the answer to question 5 and opinion statement. (2) Justification of an advantage over ANCOVA with adjustment for (multiple) covariates is necessary as part of the procedure. We do not qualify to which extent of sample sizes there is a practical advantage, but in principle it applies for finite samples.

Stakeholder number	General comment	Outcome (if applicable)
	Smeden & others shows that you need far larger sample sizes that you could hope to get from a control arm of a trial. 5. They claim that PROCOVA 'produces unbiased estimates for treatment effects' (p3 lines 43–44) but of course it could be biased in small samples as with regular ANCOVA. 6. The term 'optimal' seems to be used in an exaggerated way. They use the Pearson correlation and have expressions for variance, power and sample size that depend on R^2. But R^2 depends on distribution of prognostic score among those recruited, e.g., this is not considered in table 3. Unless eligibility criteria are the same in the trial as in the historical data, you need to know the distribution (or perhaps just variance) of the prognostic score in the trial data, and using this to calculate sample size is hard, just as with any covariates.	(3) We agree that other methods could be used without leveraging historical data. It is not the intention to single out a specific method with this qualification procedure. (4) We agree that further work for derivation of a prognostic model is necessary and this should build on existing literature. (5) Approaches in the proposed procedure for small sample sizes are discussed in the answer to question 4. (6) We agree that the term optimal should not suggest that the proposed method would be superior to other approaches to include covariates or historical information in model based statistical modelling. (7) We agree that potential differences for between historical population and the future trial population have to be taken into account. Measures to avoid too optimistic estimation of a correlation coefficient and preference for external validation for the progostic model is expressed in the opinion. A statement to express that it is not the intention of this qualification procedure to single out a specific method for statistical modelling was added to the qualification opinion.
5	Only the point estimate of correlation coefficient ρ was used in power and minimum sample size calculation (Line 606), thus sample size of the out-of-	The comment is acknowledged and agreed. We agree that it could be valuable to include

Stakeholder number	General comment	Outcome (if applicable)
	sample validation dataset (or the confidence interval of ρ) was not taken into account in application of PROCOVA. Especially when historical trial data was used for out-of-sample validation, the (relatively small) sample size can be an issue. This was also not covered by the deflation factor lambda. Although the applicant/sponsor mentioned they "selected a value that was close to 1 because our cross validation and external datasets which gave us a high degree of confidence in our correlation estimates", this consideration/criterion was not explicitly stated in Step 2a about how to determine deflation factor lambda. Either the sample size of validation dataset or the confidence of ρ needs to be added to Step 2a.	information on the sample size of the validation data set or the confidence interval of the correlation coefficient into account for Step 2. However, Step 2 does not prescribe a quantitative way to determine a deflation factor. It is not intended to amend the background information document from the Applicant. No action taken
5	In Step 2a, the applicant/sponsor only mentioned the situation when "Significant differences in the standard of care (SOC) exist between the Target Trial and the out-of sample validation dataset". However, when the control group received placebo, and "observational and nature history studies, and real-world sources" (Line 126) were used as validation data, it is worth noting that, there was no placebo given to patients in real-world (receiving no treatment ≠ receiving placebo). This consideration should also be added to Step 2a.	The comment is acknowledged and agreed. We agree that it could be valuable to take the type of control or natural history information into account for Step 2. However, Step 2 does not prescribe a quantitative way to determine a deflation factor. It is not intended to amend the background information document from the Applicant.
5	In Step 2a, the methodological quality (or risk of bias) in prognostic model development and validation, and the applicability of the prognostic model were not incorporated. Well established tools, such as PROBAST (Prediction model Risk Of Bias ASsessment Tool), can be used to give a more structured guideline in determining deflation factor lambda.	The comment is acknowledged and agreed. We agree that it would be valuable to take established tools and guidelines into account for derivation of the prognostic model in Step 1. It could also be valuable to use information on potential limitations of the prognostic model into account for Step 2. It is not intended to amend the

Stakeholder number	General comment	Outcome (if applicable)
		background information document from the Applicant. No action taken.
5	The planned sample size of a new clinical trial needs to ensure some diversity in patients included, thus there should be a floor of sample size requirement. Thus when applying PROCOVA, a cap should be applied to ρ (e.g. ρ_{max}=0.6 or even 0.4), to ensure the planned sample size will not be too small.	The comment is acknowledged and partly agreed. We agree that it will be essential that the sample size is not too small. While a lower limit based on the correlation could be of value, this may not be only guided by this criterion and it may be avoided to prescribe a quantitative value. A broader context should be applied and the last paragraph in the answer to question 5 addresses this. No action taken.
6	In the draft opinion (and other available documents including those provided by the applicant), PROCOVA is defined as a three-step process, first step being "Training and evaluating a prognostic model to predict control outcomes." However, in their Briefing Book, the applicant reiterates that PROCOVA is independent from the method used to produce the prognostic model. This alone is a contradiction and makes it unclear whether PROCOVA is a methodology for utilising pre-existing validated prognostic models and therefore consists only of Steps 2 and 3, or is it a methodology that comprises all three steps. Draft Opinion (lines 1276 and onwards) touches on this issue and points out dependence of usability of methodology of Steps 2 and 3 on the successful completion of Step 1, but it fails to conclude that steps 2 and 3 are not independent from Step 1. The information provided on Training and Evaluation of Prognostic Model (Step 1) is minimal, yet this is the most contentious issue surrounding the use of historical controls in clinical trials. There is no information as to how would	The comment is acknowledged and partly agreed. As pointed out in the comment, discussion on the importance of Step 1 is included in the qualification opinion. Independence of the derivation step for the prognostic model from the final application to a future trial and treatment effect estimation is essential to ensure the properties of the method. While understanding covariates in the prognostic model would be valuable, this is not a pre-requisite for application of a prognostic model. The opinion contains an important discussion on qualitative steps to address the external validity of a prognostic model, including expressing preference for external validation.

Stakeholder number	General comment	Outcome (if applicable)
	the biases in selection of historical data be addressed, what is considered a "large" data set, how is the validation data set selected etc. The entire approach in regards to the Prognostic Model seems to be based on the assumption that all covariates can be fully understood and that they are mostly limited to the inclusion and exclusion criteria. In reality, numerous factors other than those defined in the patient selection criteria can influence the outcome of a trial, and the unique reason for having a control group within the trial is to account for these unknown covariates. Nothing in the proposed PROCOVA methodology addresses these concerns. The draft opinion should be clear on whether PROCOVA comprises all three steps or only the last two. It should also be clearer on the severe practical limitations of the methodology in view of the absence of instructions for the Step 1.	Statements were added to the qualification opinion that all three steps are a necessary part of the procedure and that it is acknowledged that work on prognostic model is necessary for application of the method.
7	Comment: We welcome the CHMP's thoughtful suggestions and proposed considerations for implementing an idea that has been sporadically used for many years. Covariate adjustment based on a prognostic risk score, derived from data external to the study, even in the setting of non-linear models, is not a new idea with previous RCTs e.g., using the Framingham risk score as a covariate when analyzing major adverse cardiovascular events (Dahlöf et al. 2002) or the Rudolph Risk Score in the analysis of delirium after cardiac surgery (Hakim et al. 2012). However, there was little specific advice around potential challenges when doing so until now and the thoughtful comments in the draft opinion help address this gap and provide several useful suggestions. In the past, the used prognostic covariates were mostly developed for risk assessment in clinical practice and not with the analysis of RCTs in mind, but the idea to specifically develop a prediction model on external data to use for covariate adjustment has been also proposed e.g., by Branders et al. (2017) and Branders et al. (2021). It would be worth	The comment is acknowledged and agreed. We agree that it would be valuable to take existing literature into account for derivation of the prognostic model in Step 1 and we appreciate the references provided. It could also be valuable to use information on potential limitations of the prognostic model into account for Step 2. It is noted that the qualification opinion states that a formalised procedure for prognostic model development cannot be qualified. It is not intended to amend the background information document from the Applicant.

Stakeholder number	General comment	Outcome (if applicable)
	emphasizing that the same best practices for developing clinical prediction models would apply when using a predicted outcome covariate, even if a model for such predictions only needs to generalize to the target trial of interest and even if to some extent the risk of a mis-specified prediction model may be primarily a risk for the trial sponsor. Dahlöf, Björn, et al. "Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol." The Lancet 359.9311 (2002): 995-1003. Hakim, Sameh M., Ahmed I. Othman, and Dina O. Naoum. "Early treatment with risperidone for subsyndromal delirium after on-pump cardiac surgery in the elderly: a randomized trial." The Journal of the American Society of Anesthesiologists 116.5 (2012): 987-997. Branders Samuel, Pereira Alvaro, Bernard Guillaume, Ernst Marie, Albert Adelin. Leveraging Historical Data for High-Dimensional Regression Adjustment, a Composite Covariate Approach 2021. Preprint posted online under https://arxiv.org/abs/2103.14421 March 26, 2021. Branders Samuel, Pereira Alvaro, Clermont Frederic, Gossuin Chantal, Demolle Dominique. Bayesian Modeling Of The Placebo Response In Neuropathic Pain Scientific Poster presented on June 1, 2017 at the Promoting Statistical Insight Conference, London (UK). 2017.	The importance of external validity is mentioned in the opinion. A statement is added to the opinion that current existing literature should be taken into account.
7	Comment regarding description and outline of the document. The draft document did not have a preamble, or an outline and it was unclear where the qualification statements started or ended and who authored the different sections of the document. For example, the executive summary in page 1 refers to "our approach". Thus, it was unclear whether pages 1-20 refer to the applicant's executive summary, position, and full application or whether these pages or part of these sections were a summary by the EMA of the application. Without a clear outline and attribution of authorship, all	The comment is acknowledged and agreed. We agree that a clearer distinction between the background information and statements from the Applicant and EMA/CHMP is necessary. The format of the document was revised.

Stakeholder number	General comment	Outcome (if applicable)
	statements in the document may appear to either represent the EMA opinions or applicant's opinions agreed upon by the EMA. Proposed change: we suggest that the qualification document includes a preamble or an outline describing the structure of different sections of the document with information on page number and authorship for each section.	
8	As researchers interested in covariate adjustment to improve randomized trial efficiency, we appreciate the development of new statistical methodology intended to improve the efficiency of Phase 2 and 3 clinical trials by using patient covariates. We are therefore pleased to provide comments to the European Medicine Agency on the DRAFT Qualification opinion for Prognostic Covariate Adjustment. Overall, we agree with the claim that PROCOVA preserves the Type I error of the test of the null hypothesis of no treatment effect, even in the presence of model misspecification. Even so, as we argue below, we disagree with the optimality claims made - or at least, we find those claims ambiguous and potentially misleading - and find the empirical evidence provided to be unsatisfactory. First, the recent FDA guidance for industry on 'Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products' recommends a - now well developed and validated - framework that improves upon PROCOVA (see e.g., Tsiatis et al, 2008; Ge et al, 2011; Díaz et al, 2019). Across all unbiased estimators of the treatment effect, it delivers the most efficient one when the prediction model for the outcome is correctly specified, and more generally retains efficiency within a large class of treatment effect estimators that includes the unadjusted estimator (Rubin and van der Laan, 2008). This is the case, no matter what outcome is considered, no matter whether it obeys a linear or non-linear model, and no matter whether there is heterogeneity in treatment effect. In contrast, PROCOVA is limited to	The comment is acknowledged and partly agreed. We agree that a clearer distinction between the background information and statements from the Applicant from statements by EMA/CHMP is necessary. We agree that other methods could be used with or without leveraging historical data. The qualification procedure discusses the proposal of a special case of ANCOVA with a specific limited context of use. It is not the intention to single out a specific method with this qualification procedure. We appreciate the references provided. We do not agree that lack of evidence for a gain in efficiency (and potentially validity) when using historical data in statistical modelling as opposed to using only trial data would per se support the preference for using only trial data and recommendation against the use of PROCOVA. We agree that further work comparing PROCOVA to methods that do not use historical data (e.g.by

Stakeholder number	General comment	Outcome (if applicable)
	continuous outcomes and attains efficiency only when in truth the expected difference in outcome between treated and untreated individuals is the same in all covariate strata. This is a severe limitation since treatment effect heterogeneity is common, and since there is no easy way of extending PROCOVA to ordinal, binary or time-to-event endpoints, for which covariate adjustment is much more underutilized. The more general methods recommended in the FDA guidance share the simplicity of PROCOVA, can easily incorporate machine learning based predictions (either trained on the trial data or historical data), and are available in software (see e.g. Williams, Rosenblum and Diaz, 2021). Second, while these more general methods do not typically make use of historical data, this is only because the underlying theory shows that there can be no efficiency benefit from using historical data on the control arm, as opposed to the trial data themselves, without invoking untestable assumptions; interestingly, this is so even when the historical data has a larger sample size than the trial! It is therefore a pity that the EMA Draft does not show an empirical comparison of PROCOVA with versus without the use of historical data (i.e., in the latter case, basing machine learning predictions on data for the control arm of the trial itself, as opposed to historical data). While in principle a small finite-sample benefit could be seen from the use of historical data, in practice any benefit will likely be offset by the fact that the historical trial population will generally differ from that in the considered trial, which in turn would result in a loss of efficiency. Although the empirical study considered a case where the patients in the historical trial have a different covariate distribution, it did not consider the more worrying setting where the mean of Y given X differs between the historical and trial population; this is especially likely to occur in disease areas with fast developments of treatments. This also has consequences for the proposed sample size	using machine learning predictions on data for the control arm of the trial itself) would be an interesting and valuable task. In general, comparisons of properties of different methods for statistical modelling that use adjustment for covariates may be of value in a specific trial setting considering the sample size of a future trial. The qualification opinion states that establishing external validity of historical data is of paramount importance when applying a prognostic model in a future clinical trial. Steps to safeguard that the sample size estimation is appropriate for clinical trial purposes are discussed in the answers to the questions and opinion statement. The following actions were taken: The format of the document was revised to better distinguish between the background document by the Applicant and EMA/CHMP statements. A statement to the qualification opinion was added that the qualification does not intend to single out a specific method for statistical modelling, but qualify acceptable methods. A statement was added that the derivation of a suitable prognostic model in Step 1 is part of the procedure.

Stakeholder number	General comment	Outcome (if applicable)
	estimation approach: if the prognostic value of covariates differs between both trials, then this will lead to an under- or overpowered trial. In conclusion, we recommend against the use of PROCOVA, despite its validity. Instead, we recommend the use of the methods advocated in the FDA guidance for industry on 'Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products'. These methods are more generally optimal (also in the presence of treatment effect heterogeneity, or non-linear models) without being more difficult to apply. Furthermore, evidence is lacking that the use of historical data (as opposed to trial data) leads to greater validity and efficiency. In particular, the methods proposed in the FDA guidance have been shown to remain valid when the selection of baseline covariates is based on the trial data itself (as opposed to historical data), so long as one pre-specifies how the selection will be done (i.e., based on what covariate set and algorithm) (Williams, Rosenblum and Diaz, 2021). Furthermore, the use of historical data implies a loss in efficiency when - as is likely - those data obey a different prediction model. Finally, we found many claims in the EMA draft on PROCOVA to be ambiguous or over-worded. Given the nature of this 'general comments' section, we have omitted those specific concerns, but are pleased to provide specific comments if desired. Díaz, I., Colantuoni, E., Hanley, D. F., & Rosenblum, M. (2019). Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards. *Lifetime data analysis*, *25*(3), 439-468. Ge, M., Durham, L. K., Meyer, R. D., Xie, W., & Thomas, N. (2011). Covariate-adjusted difference in proportions from clinical trials using logistic	

Stakeholder number	General comment	Outcome (if applicable)
	regression and weighted risk differences. *Drug information journal: DIJ/Drug Information Association*, *45*(4), 481-493. Rubin, D. B., & van der Laan, M. J. (2008). Empirical efficiency maximization: improved locally efficient covariate adjustment in randomized experiments and survival analysis. *The International Journal of Biostatistics*, *4*(1). Tsiatis, A. A., Davidian, M., Zhang, M., & Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in medicine*, *27*(23), 4658-4677. Williams, N., Rosenblum, M., & Díaz, I. (2021). Optimizing Precision and Power by Machine Learning in Randomized Trials, with an Application to COVID-19. *arXiv preprint arXiv:2109.04294*.	
9	This method to use historical data to perform a Prognostic Covariate Adjustment is a great opportunity to make use of all the data from trials done to facilitate new trials in the future and enable more patients from the start to receive an actual medicine. No further comments.	The comment is acknowledged and appreciated. We agree that the proposed procedure is an option to leverage historical data for clinical trials. No action taken.

2. Specific comments on text

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
Lines 124-134 Lines 390-411 Lines 629-637	2	Comment: To train highly accurate prognostic models, large longitudinal datasets are typically required. In the empirical application, data from ~6,900 subjects were utilized. Sometimes fit-for-purpose databases with high quality consists of limited number of subjects. It would be helpful if the Applicant can elaborate or/and demonstrate the impact of the number of subjects on the proposed PROCOVA and then under what situation the PROVOVA may not be appropriate. Proposed change (if any): N/A	The comment is acknowledged and agreed. We agree that further work on prognostic modelling in Step 1 is necessary. Prognostic model development can only be handled on a case-by-case basis and the approach to prognostic modelling is out of scope of the qualification. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 124-134 Lines 390-411 Lines 512-517	2	Comment: The applicant indicates that PROCOVA borrows information from a historical data. Sometimes concurrent data from prospective studies can be leveraged. Please comment/elaborate if PROCOVA can accommodate both history data as well as concurrent data.	The comment is acknowledged and partly agreed. We agree that concurrent data could be leveraged in principle, but the prognostic score has to be pre-specified with all properties (scale factor, weights) before use in analysis of a confirmatory trial. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 618 - 728	2	Comment: The Applicant outlines empirical application of PROCOVA through using historical controls in completed Alzheimer disease clinical trials.	The comment is acknowledged and partly agreed.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		Please elaborate if other non-RCT historical patient data, such as natural history studies, can also be leveraged as the historical data to train the PROCOVA prognostic model, or would additional mathematical, simulation, and empirical demonstrations be needed to substantiate this? Comment: As the relationship between prognostic score and outcome (either in the historical data or in the trial data) can be nonlinear and the Pearson correlation is only measuring a linear association between prognostic score and observed outcome in the historical data, I would suggest using Spearman rank correlation in any prospective sample size estimation. Proposed change (if any): Replace Pearson correlation with Spearman rank correlation in any prospective sample size estimation.	Data from different sources could be used for derivation of a prognostic model and further work on prognostic modelling in Step 1 on a case-by-case basis is necessary We also agree that further work on details on the application of the prognostic model in Step 2 would be desirable. It is not intended to amend the background information document from the Applicant. Additional general comments on Step 1 were included in the opinion statement. No specific recommendation on methodology is given in the qualification opinion.
	2	Comment: In the step 3 estimating the treatment effect from the completed study using a linear model while adjusting for the control outcomes predicted by the prognostic model, it is unclear how the prognostic score in the linear regression.	The comment is acknowledged and partly agreed. Non-linear models for analysis in Step 3 are out of scope of the qualification procedure. Nevertheless, further work could be done to incorporate specific approaches.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		Proposed change (if any): As the relationship between prognostic score and outcome can be nonlinear and the relationship between prognostic score and the outcome is not of our primary interests, smoothing techniques should be recommended as regular practice.	No action taken.
(Type I error control and historical data) Lines 464-465 Lines 594-595	3	Comment: Regardless whether the regression model used is a linear or nonlinear model, or whether the outcome is continuous, categorical, or time to event, Type 1 error is preserved because the covariate adjustment is defined independently of the trial outcomes so that randomization guarantees balance and unbiasedness (at least in sufficiently large trials). In fact, the document pretty much says this (lines 464-465). Consequently, there is no point of the theoretical discussion to establish that Type 1 error is conserved. Also, since covariance analysis can increase power if the response (at least in the control group) is related to the covariate (lines 594-595), the point of the theoretical discussion of power is not clear, either. Of course, if the covariate turns out not to be related to (control group) response, then adjusting using the covariate will not increase power and, in fact, may decrease power. This is well known for common garden variety covariance analyses and there is no	The comment is acknowledged. The described properties pertain to linear models as special case of ANCOVA. Approaches with non-linear models for analysis and direct comparisons to such models are out of scope of this qualification procedure. No action taken.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		reason to believe it is not true as well when more exotic covariates are considered. Proposed change (if any):	
P4 line 124	4	TM & IW: Comment: 'A number of recent technological developments have led to substantial improvements in the ability to train highly accurate prognostic models'. This is a real stretch and reviews of prognostic models suggest the quantity is high but quality uniformly low e.g. Wynants et al. (https://www.bmj.com/content/369/bmj.m1328) Proposed change (if any):	The comment is acknowledged and partly agreed. We agree that the quality of prognostic models varies. Nevertheless we see potential for application of new methodology to development of prognostic models. No procedure for development of a prognostic model can be qualified as stated in the opinion. The background information document from the Applicant is not to be amended. No action taken.
Line 606	4	Comment: The mathematical results compare PROCOVA with no covariate adjustment, not with other forms of covariate adjustment; and the result for power is clearly wrong since it does not respect that power is bounded at 100% Proposed change (if any):	The comment is acknowledged and partly agreed. An advantage over ANCOVA with covariate adjustment has to be justified as part of Step 2. This could involve comparisons. The background information document from the Applicant is not to be amended. No action taken.
Line 507–8	4	Comment: Theorem 2 is misleading. It should say 'If the treatment effect is constant, then the optimal covariate to adjust for in ANCOVA is ~~a~~	The comment is acknowledged and partly agreed.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		~~prediction of~~ the potential control outcome'. The difference is shown e.g., by table 1. Proposed change (if any):	We agree that the intended statement and wording for a situation where the true outcome is not known may be clarified. The background information document from the Applicant is not to be amended. No action taken.
Tables 1 & 2	4	Comment: should report empirical SE rather than MSE (though these will be about the same), given that their estimators are unbiased.	The comment is acknowledged and partly agreed. We agree that additional information on results may be provided. The background information document from the Applicant is not to be amended. No action taken.
Line 131 and Line 525	5	Comment: "large, analysable databases containing high-dimensional outcomes, ...". Did the applicant mean "high-dimensional covariates/predictors", which can allow for develop more complicated and more accurate prediction models? Proposed change (if any): "large, analysable databases containing high-dimensional covariates/predictors, ..."	The comment is acknowledged. We assume that the description could be more precise. It could still apply to both, covariates and outcomes. It is not intended to amend the background information document from the Applicant.
Line 4	7	Comment: The method is proposed for general control outcomes not only placebo Proposed change: Rephrase to say "...predicted outcomes on control"	The comment is acknowledged and agreed. We agree that the wording could be more precise. However, it is not our policy to amend the

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
			background information document from the Applicant. No action taken.
Line 5 (first mention)	7	Comment: The model predicting the prognostic score is called differently throughout the document. Here it is called “predictive model”, elsewhere it is called “prognostic model”. Proposed change: Use a consistent terminology/keyword throughout the document, for example we suggest using “prognostic model”	The comment is acknowledged and agreed. We agree that the wording could be more uniform. It is not intended to amend the background information document from the Applicant. To at least be consistent in the opinion text, we used the term 'prognostic model' for the model predicting the prognostic score.
Lines 5 to 8	7	Comment: We disagree that the proposed approach is better than all other available approaches that either leverage historical data or covariate adjustment. For example, historical data can be leveraged in multiple efficient ways including augmenting controls and to inform prior distributions of outcomes. The proposed approach uses historical data to derive an optimal way of adjustment, and the data is not used directly in the analysis. Further, it is by no means certain that the proposed approach would be more efficient than more traditional ANCOVA covariate adjustment performed directly using the RCT data only. Proposed change: Remove the part of the sentence that states “better than other available approaches”	The comment is acknowledged and agreed in principle. We agree that there should not be the suggestion that the proposed method would be the only and best option in all situations. However, it is not intended to amend the background information document from the Applicant. A statement to the qualification opinion was added that the qualification does not intend to single out a specific method for statistical modelling.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
Line 9	7	Comment: We disagree with the use of terminology "prognostic covariate adjustment". There is a standard terminology used for decades in drug development (for e.g., in paper by Chen CH and George SL in 1985 https://doi.org/10.1002/sim.4780040107 and the paper by Senn S. in 1989 https://doi.org/10.1002/sim.4780080410), which has a much more general form than proposed in this document. Thus, using this name and acronym may lead to confusion. Proposed change: use a different name for the adjustment	The comment is acknowledged and partly agreed. We agree that the wording could be more precise. It is not intended to amend the background information document from the Applicant. As above, we used adapted the opinion, using less terms, while maintaining for legibility Applicant terminology.
Lines 13-18, lines 44-45	7	Comment: Please clarify whether this is an EMA recommendation. For instance, does this statement imply that the EMA recommends PROCOVA in these situations and that other types of covariate adjustment are no longer recommended? Proposed change: propose to replace "...is recommended for use..." in line 13 by "..may be used...". Also replace the summary statement in line 45 by "Discuss with health authority implications of this methodology on general recommendations on pre-specification"	The comment is acknowledged and agreed. We agree that a clearer distinction between the background information and statements from the Applicant and EMA/CHMP is necessary. It is not intended to amend the background information document from the Applicant. The format of the document was revised.
Line 18	7	Comment: Clarify that imputation refers to imputing baseline covariates and the prognostic score, not the outcome.	The comment is acknowledged and partly agreed.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		Proposed change: Replace statement by "(and missing baseline data or prognostic value imputation scheme should be prespecified)"	We agree that the wording could be more precise. It is not intended to amend the background information document from the Applicant. No action taken.
Line 23	7	Comment: While we agree that expertise in machine learning is useful, building prognostic factors and identifying suitable historical data to derive them requires a broader expertise. Proposed change: replace "machine learning experts" with "experts in building prognostic clinical models"	The comment is acknowledged and partly agreed. We agree that the wording could be more precise. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 28, 147	7	Comment: We disagree with the statement that the methodology is "easy to explain, interpret". In our practical experience, deriving a prognostic score from historical data using machine learning methods, which can be black-box and use a large number of potential baseline characteristics, is neither easy to explain nor easy to interpret to clinicians, patients, or researchers involved in design and analysis of randomized clinical trials. Proposed change: We suggest dropping the qualifiers "easy to explain, interpret"	The comment is acknowledged and agreed. We agree that models for prognostic scores may not be easy to understand and explainability needs to be addressed. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 79-81, 82-88, 147-149	7	Comment: We disagree with the use of the terminology "historical borrowing" in relation to the proposed approach. The term "historical borrowing" is typically reserved for directly using historical data in the analysis, potentially after down-weighting when	The comment is acknowledged and partly agreed. We agree that the wording could be more precise. It is not intended to amend the background information document from the Applicant.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		using Bayesian approaches (e.g., Schmidli et al (2013) https://doi.org/10.1177/0962280211432512 In the proposed approach, historical data are only used to derive the functional form adjustment that will be used in the new trial. Thus, the proposed approach does not use historical data in the final outcome model/analysis, else type 1 error control would be impossible. Proposed change: We suggest deleting the paragraph in lines 82-88 and mention of this method as doing historical borrowing in other parts of the qualification package.	No action taken.
Line 104-105	7	Comment: We disagree that this approach is easy to pre-specify or that it is generic. For example, how can one pre-specify the prognostic functional form of adjustment (which may be very complex and thus not possible in text form in protocol or SAP)? How can one allow for independent replication of the trial analysis, if the prognostic covariate is not easy to explain/write down? How can pre-specification be logistically implemented and proven? Proposed change: Rewrite lines 104-105 as a recommendation to pre-specify the process with details to be discussed with health authorities along the lines "We suggest for users to pre-specify the process of identifying prognostic function and then using it in a clinical trial."	The comment is acknowledged and agreed. We agree that models for prognostic scores may not be easy to understand and explainability needs to be addressed. The necessity to pre-specify the statistical analysis for a future trial in Step 3 is emphasized. It is noted that no procedure for Step 1 is qualified, but pre-specification would be valuable. It is not intended to amend the background information document from the Applicant. No action taken.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
Line 133	7	Comment: We suggest using more cautious language regarding improvement of machine learning over general linear models as major improvements of the predictions of clinical outcomes have not yet materialized. There are also countless recent publications demonstrating that "flexible modelling" is not a good solution for scenarios with limited (historical) data. To see a few examples of this work, we refer to Ennis et al (1998) https://doi/10.1002/(SICI)1097-0258(19981115)17:21<2501::AID-SIM938>3.0.CO;2-M, van der Ploeg et al (2014) https://doi.org/10.1186/1471-2288-14-137 and the more recent systematic review Christodoulou et al (2019) https://doi.org/10.1016/j.jclinepi.2019.02.004) Proposed change: We suggest replacing "...which can substantially reduce variance/confidence intervals,..." with "...which might reduce variance/confidence intervals..."	The comment is acknowledged and partly agreed. We agree that the quality of prognostic models varies. Nevertheless, we see potential for application of new methodology to development of prognostic models. No procedure for development of a prognostic model can be qualified as stated in the opinion. The background information document from the Applicant is not to be amended. No action taken.
Lines 144-145 and lines 175-176	7	Comment: We disagree with the general statement that the method is effective and safe without caveats, specifically about the choice of historical data and its adequacy. Also, we disagree that the existence of historical data is sufficient for the methodology to be applicable. Proposed change: remove the wording "effective and safe" from the statement in lines 144-145. Provide	The comment is acknowledged and partly agreed. We agree that Step 1 is an important part of the proposed method. No procedure for development of a prognostic model can be qualified as stated in the opinion. The background information document from the Applicant is not to be amended.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		general statements about the need for evaluation of adequacy of historical data, most specifically when discussing scope of application throughout the document, including lines 175-176.	No action taken.
Lines 165-174	7	Comment: The advantages listed of the proposed approach are not unique as the same advantages apply to standard covariate adjustment for prognostic indices. Proposed change: remove the comparative statement "over other approaches" in line 165.	The comment is acknowledged and partly agreed. When applying the method an advantage over standard covariate adjustment has to be justified in Step 2. The background information document from the Applicant is not to be amended. No action taken.
Lines 81, 207 and 1137	7	Comment: The terminology matching to historical data may be misunderstood to represent matching methods used in observational data analysis to adjust for confounding (e.g., exact matching or propensity score matching). Proposed change: replace in lines 81 and 207"...suitably matched historical population" by "...suitable/adequate historical population". Replace "...how to match the validation dataset to the trial population..." with "how to select the validation dataset to be comparable to the trial population"	The comment is acknowledged and partly agreed. We agree that the wording could be more precise. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 236-237	7	Comment: It is unclear why deep learning models are specifically highlighted here. Best of class prediction model may change with time. We were also surprised not to find any reference (in step 1) to existing best	The comment is acknowledged and agreed. We agree that it would be valuable to take established tools and guidelines into account for derivation of the prognostic model in Step 1. It

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		practice and guidance documents on developing prognostic modelling. Those include the following: • Probast, - https://pubmed.ncbi.nlm.nih.gov/30596875/ • Tripod - https://www.equator-network.org/reporting-guidelines/tripod-statement/ • Remark - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3362085/ • Path statement - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7750907/ Proposed change: Remove specific reference to deep learning or offer that as an example. In addition, add references to existing best practice guidelines on developing prognostic scores.	could also be valuable to use information on potential limitations of the prognostic model into account for Step 2. It is not intended to amend the background information document from the Applicant. No action taken.
Line 249	7	Comment: The procedure is the same as before (no formalization). The contribution of the applicant was to prove the statistical properties of this approach. Proposed change: remove the sentence "our approach is a formalization of what has previously been an ad-hoc procedure"	The comment is acknowledged and not agreed. The term appears applicable to describing a procedure with defined steps. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 325-329	7	Comment: We disagree that these are sole properties of PROCOVA, they generally apply to ANCOVA.	The comment is acknowledged and partly agreed.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		Proposed change: rephrase in line 325 "...primary benefits of PROCOVA..." to "...primary benefits of ANCOVA..."	We agree that the wording could be more precise. It is not intended to amend the background information document from the Applicant. No action taken.
Line 544	7	Comment: There is a typo in a formula relating to the scenario Proposed change: replace " $E[Y1-Y0|X] \neq \mu 1(X)-\mu 0(X))$" with "E[Y1-Y0|X] ≠const.)	The comment is acknowledged and agreed. We agree that the description should be precise. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 557-564	7	Comment: Some important simulation specifications are missing from this summary and results. Proposed change: Add all specifications used in the simulation in the body of the document, including correlation values (between prognostic and outcome) and discounting factor values. Add simulation results for low values of the correlation (for an R2 in the range of 0.1-0.3 as is commonly observed in clinical outcome data)	The comment is acknowledged and agreed. We agree that the description of the simulation could be more precise. It is not intended to amend the background information document from the Applicant. No action taken.
Line 601	7	Comment: Additional important limitation about sample size and power is needed. Proposed change: Add that if this methodology had been used also to reduce the sample size at the design stage, that the trial would not have been adequately powered.	The comment is acknowledged and party agreed. We agree that the aspect could be mentioned. It is not intended to amend the background information document from the Applicant. No action taken.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
Line 615	7	Comment: To our knowledge, there has been limited success of neural network models use with clinical data and the success was broadly within reach of standard methods such as random forest. Proposed change: Please support the statement with references to systematic reviews.	The comment is acknowledged and partly agreed. We agree that Step 1 is an important part of the proposed method. No procedure for development of a prognostic model can be qualified as stated in the opinion. The background information document from the Applicant is not to be amended. No action taken.
Line 668	7	Comment: Some fitted values are missing. Proposed change: Please add confidence interval estimates for the correlation for random forest and deep learning models in Table 3	The comment is acknowledged and agreed. We agree that the description of the simulation could be more informative. It is not intended to amend the background information document from the Applicant. No action taken.
Lines 1248-1250	7	Comment: The qualification statement of PROCOVA includes all phase 2 and 3 clinical trials. However, some phase 2 and phase 3 clinical trials are single arm or not randomized. The proposed use of PROCOVA and demonstrated properties of ANCOVA were for randomized clinical trials where the prognostic scores is not associated with treatment assignment and thus not a confounding factor. An association of the prognostic factor with treatment, possible in a non-randomized study, could potentially	The comment is acknowledged and agreed. We agree that the trial setting should be part of the qualification opinion statement. The qualification opinion statement was amended accordingly.

Line number(s) of the relevant text	Stakeholder number	Comment and rationale; proposed changes	Outcome
		lead to a biased treatment effect estimate with ANCOVA. Proposed change: We propose to include the word "randomized" in the qualification for use as in "CHMP qualifies PROCOVA as prognostic score adjustment and the proposed procedures as described in a handbook for trial statisticians could enable increases in power or precision of treatment effect estimates in controlled randomized clinical trials with continuous outcomes"	
Lines 935, 1049, 1112, 1127, 1134, 1226, 1249, 1273, 1288, 1290, 1311	7	Comment: The Q & A section and the qualification sections refer to the applicant's "handbook" with step-by-step guide for trial statisticians. This is an important document with some guidelines on checking assumptions and sensitivity analyses around them. However, there is no link or citation to the handbook and it is unclear whether this refers to part of pages 1-20 of the application in its current or earlier format or whether this will be a separate documentation in the qualification. Proposed change: We suggest that the qualification document provides a cited reference when references to this handbook are made.	The comment is acknowledged. We apologise that the reference and link to the handbook document was apparently not clear. The handbook will be part of the final documentation.