Response to EMA Qualification Advice List of Issues Dated 2 December 2021

Unlearn.AI, Inc
75 Hawthorne Street, Suite 560
San Francisco, CA 94105

29 December 2021

INTRODUCTION

Unlearn appreciates the thoughtful questions and robust scientific discussions with the EMA related to our submission for PROCOVA™, a statistical methodology for analysis of continuous outcomes in Randomized Controlled Trials. We are pleased that the SAWP has confirmed the intention to proceed with a qualification opinion (QO).

Unlearn acknowledges the EMA's concern that it is vital for trial statisticians who intend to use PROCOVA™, to understand the impact of potentially over-optimistic assumptions. We agree that trial developers should be able to make informed decisions in choosing one of three paths – no adjustment, ANCOVA with one or more pre-specified covariates, or PROCOVA™ – as well as to understand and to be able to effectively address a number of other uncertainties described in the EMA's List of Issues.

Here, we provide responses to the questions posed in Qualification Advice List of Issues dated 2 December 2021. Additionally, Unlearn has created a document entitled PROCOVA™ Handbook for the Trial Statistician (hereafter referred to as The Handbook, and included with this submission). The Handbook represents a significant expansion on the "Checklist for the Practitioner" which was included at the end of the "FINAL EMA Discussion Meeting Minutes_PROCOVA Qualification_27Sep2021" submitted by Unlearn on 1 October 2021.

Unlearn acknowledges that the PROCOVA™ approach involves certain planning and application steps that are different from the conventional approach. Therefore, The Handbook provides a brief overview of the three steps of PROCOVA™ followed by a detailed guidance for the practical application of PROCOVA™, encompassing further elaboration on all the aspects raised in the Qualification Advice List of Issues.

ISSUES ON STATISTICAL METHODOLOGY

The Applicant is asked to provide instructions for the practical use of the PROCOVA approach, with respect to:

ISSUE 1

The attainable advantage over using ANCOVA with "conventional" covariate adjustment and the decision to choose one of the three paths: no adjustment, ANCOVA with one or more pre-specified covariates, or PROCOVA. Instructions should be given for the choice of the deflation factor λ, and the conduct of sensitivity analyses taking into account a potential over-optimism of the prognostic model, the external validity of the historical validation data sets and the fact that the correlation of the prognostic score with the outcome may be smaller under experimental treatment.

SPONSOR RESPONSE

The elements raised as part of Issue 1 are now addressed in the instructions for Step 1 and Step 2 of PROCOVA™ provided in The Handbook. Since they are addressed in The Handbook in a slightly different order, we have bolded individual elements of Issue 1 for ease of navigating the response.

Step 1 of PROCOVA™ is to "Validate the prognostic score (obtained from a prognostic model) for use in the Target Trial; *collaborate with the Model Developers*". The following information on Step 1 is included in The Handbook introduction:

> "The purpose of Step 1 is to validate the prognostic score generated by a prognostic model for use in a particular planned trial which we will call the Target Trial. This validation involves estimating the Pearson correlation coefficient R between the prognostic score and the actual outcomes obtained from a separate dataset which was *not* used to train the prognostic model, and which contains data from subjects whose baseline characteristics are similar to those in the Target Trial. The activities involved in Step 1 require a close collaboration between the Target Trial Statistician and the Model Developers."

For Step 1, The Handbook provides definitions and instructions on how to establish **the external validity of the historical validation data sets.** This section of The Handbook is reproduced below:

> "**1a.** Confirm that the Pearson correlation coefficient R between the prognostic score (computed by a prognostic model) and the outcome was obtained using an out-of-sample validation dataset, i.e., a dataset not used to train the prognostic model. When such out-of-sample validation dataset is not available, PROCOVA™ is not recommended.
>
> A prognostic model is defined as a mathematical function of a subject's baseline covariates that predicts the subject's expected outcome if he or she were to receive a control treatment (e.g., placebo) in the Target Trial. A subject's prognostic score is the output of the prognostic model for a given subject.
>
> **1b.** Confirm that this out-of-sample validation dataset is similar to the population of the Target Trial, i.e., contains data from subjects meeting the main inclusion criteria of the Target Trial. Such criteria should include intended indication and baseline severity/stage of disease, as well as other baseline characteristics known or strongly suspected to be correlated with the outcome in a particular disease area, such as age, time since onset of symptoms, or known biomarkers.
>
> For instance, if the Target Trial will be conducted in subjects over age 65 who have severe disease, the out-of-sample validation dataset should not include subjects with mild or moderate disease or aged 65 or younger.
>
> **1c.** Determine if the correlation R obtained using the out-of-sample validation dataset is at least 90% of the R provided by the Model Developers and obtained using an in-sample dataset (defined as a historical dataset that was used to train the prognostic model). If it is less than 90% of the in-sample R, factor lambda can be used to keep the estimates conservative, see Step 2a below."

Step 2 of PROCOVA™ is to "Estimate sample size and plan the Target Trial taking the prognostic score into account; *collaborate with the Model Developers*". The following information on Step 2 is included in The Handbook introduction:

> "The purpose of Step 2 is to estimate the sample size and plan the Target Trial using PROCOVA™ for the primary analysis. In Step 2, the R (as defined above) can be used to calculate sample size reduction and/or power increase compared to a traditional design. To keep estimates conservative, as is common for sample size

estimation, one can use lambda (the deflation factor for the correlation coefficient R) and gamma (the inflation factor for the standard deviation).

In addition, the expected variances attainable with PROCOVA™, ANCOVA with conventional covariate adjustment, and no adjustment, should be compared to enable the selection of the optimal procedure that will result in the greatest reduction in variance. If PROCOVA™ is chosen, *R* can be used to calculate the new sample size for the Target Trial and the associated statistical power. The decisions and actions involved in Steps 2 also require a close collaboration between the Target Trial Statistician and the Model Developers.

The Target Trial protocol must pre-specify all design and analysis choices including those related to the application of PROCOVA™ as the primary analysis. The protocol must also indicate whether adjustment for additional covariates in the regression is part of the primary analysis or is included as a sensitivity analysis. These decisions must be also pre-specified in the Statistical Analysis Plan (SAP) finalized in advance of database lock."

For Step 2, The Handbook provides guidance **for the choice of the deflation factor λ, and for the conduct of sensitivity analyses taking into account a potential over-optimism of the prognostic model and the fact that the correlation of the prognostic score with the outcome may be smaller under experimental treatment** (Step 2a)**.**

The relevant text of The Handbook is included below:

"2a. Gather the standard inputs needed to compute a sample size for a given power (i.e., the target effect size, the standard deviation of the outcome, the proportion of subjects to be randomized to the intervention, the expected dropout rate, and the alpha level).

To keep sample size estimates conservative, PROCOVA™ makes explicit use of two factors designed to help avoid undue optimism, i.e., lambda (the deflation factor for the correlation coefficient R) and gamma (the inflation factor for the standard deviation). Note that lambda is specific to PROCOVA™ because the use of R is specific to PROCOVA™. Gamma, however, is relevant to any sample size calculation (as is standard deviation).

To obtain a conservative estimate of the correlation coefficient R, choose an appropriate value of lambda (the deflation factor for R) using the following rules-of-thumb:

- Choose lambda ~ 0.95 if *similar correlation coefficients R* were obtained by the Model Developers using the in-sample dataset and *two or more* out-of-sample datasets that matched the Target Trial, see Step 1c above.
- Choose lambda ~ 0.90 if *similar correlation coefficients R* were obtained by the Model Developers using the in-sample dataset and *a single* out-of-sample dataset that matched the Target Trial, see Step 1c above. Reduce lambda further (or consider requesting another out-of-sample dataset assessment) if the out-of-sample dataset performance is less than 90% of the in-sample dataset performance.

- In collaboration with the Model Developers, identify sensitivity analyses to be conducted with reduced lambda (by approximately 0.05 for each) if any of the following conditions apply to the Target Trial:
  - Significant differences in the standard of care (SOC) exist between the Target Trial and the out-of-sample validation dataset, e.g., due to a rapid and broad adoption of a new therapy for a component of the disease etiology. Such event may alter the likely outcome of the Target Trial *vs* the original out-of-sample validation dataset which did not contain data from subjects on the new SOC. [Note that in practice, SOCs rarely undergo such major changes in a short amount of time].
  - Significant differences in data completeness exist between the Target Trial and the out-of-sample validation dataset. The model generates prognostic scores for all Target Trial participants, regardless of missing data; however, the correlation coefficient R may be lower if one or more important variables are expected to be missing frequently (or with a different pattern of missingness) in the Target Trial compared to the out-of-sample validation dataset, and if the missing variable(s) are known or suspected to be highly prognostic.
  - The prognostic score includes a potentially predictive biomarker (which identifies the likely responders to treatment) rather than a prognostic biomarker (which is associated with a particular clinical outcome regardless of treatment). This could result in a weaker correlation between the prognostic score and the expected outcome in the active treatment arm compared to the control arm, and thus a lower lambda should be considered for the treatment arm *vs* the control arm.

  After selecting lambda, and for each sensitivity analysis, multiply the correlation coefficient R by lambda and use the resulting value when estimating power with PROCOVA™. If different lambdas were selected for the treatment and control groups, complete this step separately for each group."

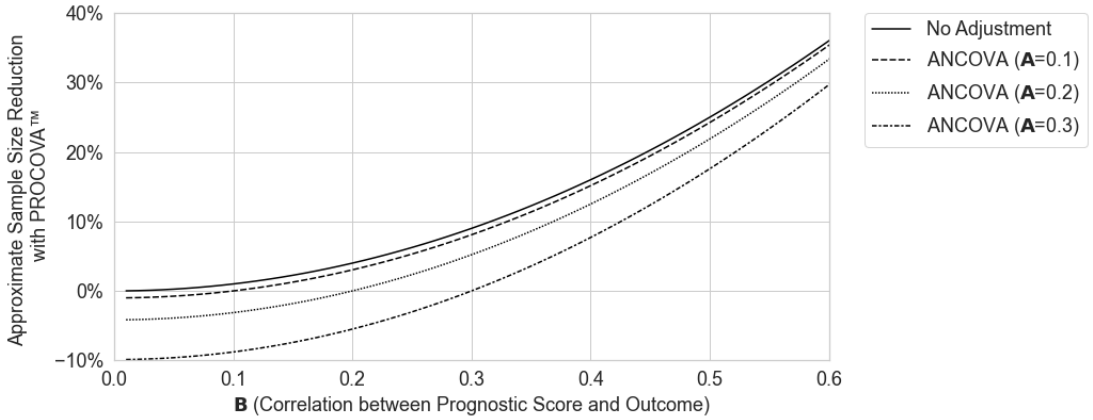
Later in Step 2, The Handbook provides guidance to help the Target Trial Statistician make an informed **choice among the three paths - no adjustment, ANCOVA with one or more pre-specified covariates, or PROCOVA™**, including the figure below and step-by-step instructions on how to use it.

The relevant section of The Handbook is reproduced below:

> **"2b.** Compare the expected impact of PROCOVA™ (obtained in Step 2a above) to that of ANOVA or traditional ANCOVA adjusting for baseline covariate(s), in order to choose the optimal approach.
>
> To estimate the potential sample size reduction attainable from PROCOVA™ and ANCOVA/ANOVA, use the figure below where:
>
> - The quantity **B**, plotted along the X-axis, is a *conservative estimate of R* in Step 2a (i.e., the correlation R multiplied by lambda).
> - The quantity **A** represents the correlation between a single baseline variable, or a linear combination of a small set of variables, and the outcome. The quantity **A**

may be estimated using an out-of-sample dataset. For a single covariate, multiplication by lambda may not be necessary, but, for multiple covariates, consider choosing the same lambda as used to estimate **B** in the bullet above.

- The incremental sample size reduction with PROCOVA™ *vs* ANCOVA/ANOVA is a function of **A** and **B**, and can be approximated as 100% - (1-$\mathbf{B}^2$) / (1-$\mathbf{A}^2$). This formula was used to generate the curves in the figure below.
- Find the point on the graph below corresponding to quantities **A** and **B** associated with the Target Trial. The corresponding value along the Y-axis is the expected incremental sample size reduction attainable in the Target Trial with PROCOVA™ over and above the sample size reduction achievable with ANCOVA or, when **A**=0, with ANOVA ("No Adjustment" in the graph below).
- If the original sample size was estimated with ANOVA as the primary analysis, then ANOVA (**A**=0) should be the reference for determining the incremental power gains attainable with PROCOVA™. If the original sample size was estimated with ANCOVA as the primary analysis, and the power gains associated with ANCOVA were factored into the sample size calculation, the Y-axis represents the incremental sample size reduction attainable *vs* ANCOVA."



Throughout Steps 1 and 2, particular attention is paid in The Handbook to the **external validity** of the datasets used to validate the prognostic score. Instructions are provided on how to match the validation dataset to the Target Trial population (see above, Section **1b**); how to account for the potential changes in the SOC (see above, Section **2a**, third bullet), and how to address the potential differences in the missing data or patterns of missing data between the validation dataset and the Target Trial (see above, Section **2a**, third bullet).

ISSUE 2

The implementation of stratified randomisation, having in mind that strong prognostic factors would usually be considered in design and analysis of a clinical trial.

SPONSOR RESPONSE

Instructions for combining PROCOVA™ with **stratified randomization** are provided in the last subsection of Step 2. The corresponding text of The Handbook is copied below:

> **"2c.** When making a final determination whether to use PROCOVA™, consider both ANCOVA and ANOVA (no adjustment) as alternatives. Also consider additional features of the trial that may be of relevance (see below).
>
> If PROCOVA™ is chosen to be used, specify if individual baseline covariates will be also included in the primary analysis. When PROCOVA™ is applied to trials utilizing **stratified randomization**, the strata should be included as covariates in the primary analysis (note, however, that the prognostic score is not designed to be used for stratification)."

Additional instructions are provided in the final section of The Handbook, i.e., Step 3, "Analyze trial results using a linear model while adjusting for the prognostic score", described in The Handbook introduction as follows:

> "Step 3 takes place after the completion of the Target Trial designed using the new estimate of sample size and/or power obtained in Step 2, and after database lock. The purpose of Step 3 is to estimate the treatment effect using a linear model while adjusting for the prognostic score. Finally, a null hypothesis is assessed by computing a two-sided p-value based on a t-distribution."

The following guidance for Step 3 is included in the corresponding section of The Handbook:

> "Applying PROCOVA™ using a linear model while adjusting for the prognostic score and any additional pre-specified baseline covariates, produces an unbiased estimate of the treatment effect, however, it does not produce an unbiased estimate of a subgroup effect. Therefore, to gain precision from PROCOVA™ when assessing the treatment effect for individual subgroups, adjust for a prognostic score on the subset of subjects in that particular subgroup or **strata**. These treatment effect estimates should also be used when evaluating treatment-by-subgroup interactions.
>
> Do not evaluate subgroup effects or treatment-by-subgroup interactions using the same linear model that was used for primary analysis of the treatment effect since doing so may introduce collinearities and undermine the accuracy of subgroup-specific treatment effect estimates."

ISSUE 3

The consequence of including covariates in the analysis model that are potentially correlated with the prognostic score and the opportunity to understand the importance of the co-variate for decision making (i.e., if there is interaction with the treatment effect).

SPONSOR RESPONSE

This issue arises in the case of stratification factors, addressed above, but also for prognostic covariates in general. Individual covariates (stratification factors or otherwise) which are strongly prognostic, are likely to be correlated with the prognostic score, and their effect on the outcome may be fully captured by the prognostic score.

The guidance for applying PROCOVA™ while also adjusting for covariates that are potentially correlated with the prognostic score are provided in the last section of The Handbook, Step 3, "Analyze trial results using a linear model while adjusting for the prognostic score". The description of Step 3 and the guidance for prognostic covariates are

already included in our response to Issue 2 above, and the instructions are also reproduced below:

> "Applying PROCOVA™ using a linear model while adjusting for the prognostic score and any additional pre-specified baseline covariates, produces an unbiased estimate of the treatment effect, however, it does not produce an unbiased estimate of a subgroup effect. Therefore, to gain precision from PROCOVA™ when assessing the treatment effect for individual subgroups, adjust for a prognostic score on the subset of subjects in that **particular subgroup** or strata. These treatment effect estimates should also be used when evaluating **treatment-by-subgroup interactions**. Do not evaluate subgroup effects or treatment-by-subgroup interactions using the same linear model that was used for primary analysis of the treatment effect since doing so may introduce collinearities and undermine the accuracy of subgroup-specific treatment effect estimates."

ISSUE 4

The interpretation (and possible impact on study planning) of subgroup analyses based on covariates included in the prognostic score that may be predictive of treatment effect.

SPONSOR RESPONSE

Subgroup analyses based on covariates included in the prognostic score that may be predictive of treatment effect, are addressed in The Handbook guidance for Step 2 and Step 3 of PROCOVA™.

The instructions for Step 2, "Estimate sample size and plan the Target Trial taking the prognostic score into account", include the following text:

> **"2c.** When making a final determination whether to use PROCOVA™, consider both ANCOVA and ANOVA (no adjustment) as alternatives. Also consider additional features of the trial that may be of relevance (see below).
>
> If PROCOVA™ is chosen to be used, specify if individual baseline covariates will be also included in the primary analysis. When PROCOVA™ is applied to trials utilizing stratified randomization, the strata should be included as covariates in the primary analysis (note, however, that the prognostic score is not designed to be used for stratification).
>
> Also consider if the treatment effect is expected to differ between/among subgroups because a subgroup indicator is a **predictive biomarker** (which identifies the likely responders to treatment) rather than a prognostic biomarker (which is associated with a particular clinical outcome in the absence of therapy or with the application of a standard therapy). If that is the case, and if there is a subject subgroup for which precision of the treatment effect is especially important, additional power calculations are recommended to ensure sufficient power for both/all subgroups."

In addition, The Handbook instructions for Step 3, "Analyze trial results using a linear model while adjusting for the prognostic score", include the following text (already reproduced above in our responses to related Issues 2 and 3):

"Applying PROCOVA™ using a linear model while adjusting for the prognostic score and any additional pre-specified baseline covariates, produces an unbiased estimate of the treatment effect, however, it does not produce an unbiased estimate of a subgroup effect. Therefore, to gain precision from PROCOVA™ when assessing the treatment effect **for individual subgroups**, adjust for a prognostic score on the subset of subjects in that particular subgroup or strata. These treatment effect estimates should also be used when evaluating treatment-by-subgroup interactions.

Do not evaluate subgroup effects or treatment-by-subgroup interactions using the same linear model that was used for primary analysis of the treatment effect since doing so may introduce collinearities and undermine the accuracy of subgroup-specific treatment effect estimates."

ISSUE 5

The handling of incomplete data in covariates for prognostic score adjustment.

SPONSOR RESPONSE

This issue is addressed in The Handbook guidance for Step 2, "Estimate sample size and plan the Target Trial taking the prognostic score into account". The following text is included in instructions for Step 2a (third bullet), which deals with choosing an appropriate value of the deflation factor lambda:

- "In collaboration with the Model Developers, identify sensitivity analyses to be conducted with reduced lambda (by approximately 0.05 for each) if any of the following conditions apply to the Target Trial:
    - Significant differences in the standard of care (SOC) exist between the Target Trial and the out-of-sample validation dataset, e.g., due to a rapid and broad adoption of a new therapy for a component of the disease etiology. Such event may alter the likely outcome of the Target Trial *vs* the original out-of-sample validation dataset which did not contain data from subjects on the new SOC. [Note that in practice, SOCs rarely undergo such major changes in a short amount of time].
    - Significant differences in data completeness exist between the Target Trial and the out-of-sample validation dataset. The model generates prognostic scores for all Target Trial participants, regardless of missing data; however, the correlation coefficient R may be lower **if one or more important variables are expected to be missing frequently (or with a different pattern of missingness)** in the Target Trial compared to the out-of-sample validation dataset, and if the missing variable(s) are known or suspected to be highly prognostic.
    - The prognostic score includes a potentially predictive biomarker (which identifies the likely responders to treatment) rather than a prognostic biomarker (which is associated with a particular clinical outcome regardless of treatment). This could result in a weaker correlation between the prognostic score and the expected outcome in the active treatment arm compared to the

control arm, and thus a lower lambda should be considered for the treatment arm *vs* the control arm."

In conclusion, we would like to express our gratitude to the EMA for raising important scientific issues and for recommending that these be addressed in a "Handbook for the Trial Statistician". The Handbook provides detailed practical guidance for the application of PROCOVA™, including clarifications to all of the issues raised by the EMA, in order to help the Trial Statistician make informed decisions and avoid over-optimistic assumptions, enhancing the scientific and practical value of the novel statistical methodology to the practitioners.