PROCOVA™ Handbook for the Target Trial Statistician

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

29 December 2021

PROCOVA™ Handbook for the Target Trial Statistician.

PROCOVA™ is a statistical methodology that leverages historical data (from control arms of clinical trials and from observational studies) along with prognostic modeling to decrease the uncertainty in treatment effect estimates from Phase 2 and 3 Randomized Controlled Trials measuring continuous responses, in the large-sample setting. As a special case of ANCOVA, PROCOVA™ preserves the control of type I error rate when used with a prognostic score regardless of how this prognostic score was derived, e.g., using one or more baseline covariates, mechanistic models, linear statistical models, or machine-learning-based methods. The latter are particularly useful since they can use large databases to train non-linear, highly prognostic models, and generate prognostic scores which are highly correlated with future outcomes. This document provides step-by-step guidance for the practical application of PROCOVA™ following a brief overview of the methodology below.

PROCOVA™ Step 1.

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.

PROCOVA™ Step 2.

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.

PROCOVA™ Step 3.

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.

Step	Item	Recommended Actions for Application of PROCOVA™
1.		**Validate the prognostic score (obtained from a prognostic model) for use in the Target Trial; *collaborate with the Model Developers*.**
	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.
2.		**Estimate sample size and plan the Target Trial taking the prognostic score into account; *collaborate with the Model Developers*.**
	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.

To obtain a conservative estimate of the standard deviation, choose an appropriate value of gamma. A value of 1 may be used if consistent standard deviations are seen across out-of-sample validation datasets. If there is a wide range of standard deviations across the literature and out-of-sample validation datasets, choose a middle value and select a gamma greater than 1 to ensure conservative approach. Multiply the standard deviation by gamma and use the resulting value when estimating power with or without PROCOVA™.

With all the parameters now defined, compute the power with PROCOVA™ over a range of sample sizes to obtain a power curve. In collaboration with the

Model Developers, choose the minimum sample size to achieve the desired power using PROCOVA™.

In general, the PROCOVA™ formula does not require a 1:1 randomization and allows for the lambda and gamma values to be different between the control and treatment groups. Here we provide an example of a simplified formula where a common gamma and a common lambda are used for two equally sized treatment groups. We also assume that the outcomes have a common variance (σ^2); that n represents the trial sample size without dropouts; and that d represents the proportion of subjects who drop out.

- Estimate the standard error of the treatment effect (ν) as the square root of $[1/(n*(1-d))] * [(2\gamma\sigma)^2*(1-(\lambda R)^2]$.
- Compute the power as $\Phi(\Phi^{-1}(\alpha/2) + \beta/\nu) + \Phi(\Phi^{-1}(\alpha/2) - \beta/\nu)$ where Φ is the cumulative distribution function of the standard normal distribution, α is the type I error, and β is the target treatment effect.
- For a range of n (e.g., 100-1000) compute the power and plot the corresponding curve.

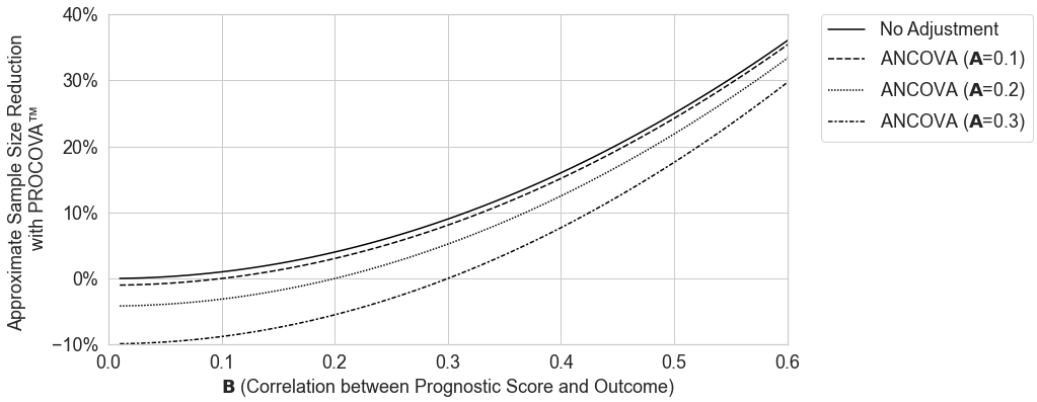
If the Target Trial utilizes co-primary endpoints, repeat the bulleted steps above for each endpoint. Select the sample size sufficient to simultaneously address the null hypothesis for both endpoints. Note that repeating the process for multiple endpoints requires a prognostic score for each of them, and that, in turn, necessitates either a machine learning model that predicts multivariate results or multiple individual models for each endpoint of interest.

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.



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.

Pre-specify in the Target Trial Protocol all design and analysis choices described above as they relate to the application of PROCOVA™, including whether adjustment for additional covariates in the regression is part of the primary analysis or is included as a sensitivity analysis. In addition, these decisions must be pre-specified in the SAP finalized in advance of database lock.

3. Analyze trial results using a linear model while adjusting for the prognostic score.

As soon as *clean* baseline data on all subjects randomized into the Target Trial are available, provide them to the Model Developers for the generation of trial-specific prognostic scores. The Model Developers *must remain blinded* to the randomization code *until after* the final trial-specific prognostic score is

	delivered to the Target Trial Statistician and applied in treatment effect estimation as described below.
	Estimate the treatment effect with PROCOVA™ using a linear model while adjusting for the prognostic score. If specified in the protocol and SAP, this may also involve adjusting for additional covariates including stratification factors in the case of trials with stratified randomization. In all cases, estimate the variance for the treatment effect using a heteroskedasticity-consistent covariance (HCC) matrix. As there is more than one method for obtaining heteroskedasticity-consistent estimates, the specific option (e.g., the HCC parameter to specify in the analysis code) should be detailed in the SAP. For this primary analysis, assess the null hypothesis by computing a two-sided p-value based on a t-distribution using the covariate-adjusted treatment effect and the heteroskedasticity-consistent standard error.
	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.