Statistical Analysis Plan

Albumin levels, inflammation and nutrition in chronic hemodialysis patients treated with medium cut-off or high flux membranes: a cohort study

Protocol number: RCS2022-002, Date: February 28, 2023, version 1.0

Change control

Version	Date	Change	Observation
1.0	February 28, 2023	Not Applicable	Original version

Prepared by: Ricardo Sanchez Pedraza

Signatures of the Statistical Analysis Plan

With our signatures we certify the approval of the statistical analysis plan version 1.0 dated February 28, 2023, for the study entitled: Albumin levels, inflammation and nutrition in chronic hemodialysis patients treated with medium cut or high flux membranes: a cohort study.

MSc Biostatistician (Author)

Name: **Ricardo Sanchez Pedraza**

Signature: R. Sanchez

February 28, 2023

PhD in Economics (Author)

Name: **Giancarlo Buitrago**

Signature: ______________________

February 28, 2023

PhD Clinical Epidemiologist (Author)

Name: **Henry Oliveros**

Signature: ______________________

February 28, 2023

Principal Investigator

Name: **Dr. Juan Carlos Mario Castillo**

Signature: Juan Carlos Castillo. S.

February 28, 2023

Table of Contents

1. Introduction :

Many of the current synthetic HD membranes are less effective than the glomerular membrane, in the sense that they do not adequately remove higher molecular weight toxins (4). In previous decades, technological innovation in membranes focused on improving the biocompatibility profile. However, the development of bioengineering has allowed the development of membranes with a higher clearance of uremic toxins (5). Several molecules (greater than 200 kDa) have been identified as toxic molecules, such as kappa (κ) and lambda (λ) free light chains (FLCs) , all in the medium-high molecular weight range, which are above the removal capacity of classical high-flux membranes, given that they have a molecular radius greater than that of the membrane pores (6).

The concept of offering greater removal of protein-bound uremic toxins and medium molecular weight molecules makes HD based on this type of medium cut-off membranes, generically called expanded hemodialysis (MCO), increase (expand) the clearance capacities of the membrane, constituting an important therapeutic step with which positive results have been obtained in some efficacy studies, aimed at evaluating inflammation, and the clearance of medium molecules (8).

Although current renal replacement therapy systems have reduced patient morbidity and mortality and significantly improved their quality of life, they have not achieved an efficiency equivalent to that of the kidney. New membranes for dialysis systems have recently been developed (medium cut -off membranes (COMs) that allow the filtration of medium molecular weight molecules and that, in some studies, have shown promising results. However, for this new technology, there is still little information from longitudinal studies with sufficient follow-up time on clinical outcomes of importance for the patient and health systems, such as survival, frequency of cardiovascular events or hospitalization.

This study aims to establish the role of serum albumin levels, as well as that of markers of inflammation and malnutrition on outcomes such as mortality, non-fatal cardiovascular events, hospitalization and hospital days in patients on hemodialysis, treated with medium cut-off (MCO) or high flux (HD AF) membranes.

2. Objectives and outcomes

2.1. General objective

To establish the effect of serum albumin levels, as well as inflammation and malnutrition markers on outcomes such as mortality/survival, non-fatal cardiovascular events, hospitalization and hospital days in hemodialysis patients treated with medium cut-off (MCO) or high flux (HF) membranes.

2.2. Specific objectives

1. To describe the demographic and clinical characteristics of both cohorts (MCO and HD AF).
2. To estimate and compare the days of hospital stay and the incidence of mortality and hospitalization in the studied cohorts.
3. To compare the cumulative hospitalization time during participation in the study, between patients receiving MCO and those receiving HD AF.
4. To compare the time to death event between patients receiving MCO and those receiving HD AF.
5. To compare the frequency of non-fatal cardiovascular events in both cohorts.
6. To estimate the association between serum albumin levels and the outcomes mortality, non-fatal cardiovascular events, hospitalization events and days of hospital stay, controlling for the effect of third variables such as PEW, MIS, hs CRP, type of dialyzer.

2.3. Exploratory objective

1. Describe the use and type of nutritional supplements in the evaluated cohorts.

2.4. Primary Outcomes

- Incidence of hospitalization events (per patient-year)

- Length of hospitalization (hospital days per patient-year)
- Mortality
- 4-year survival

2.5. Secondary Outcomes

- Major non-fatal cardiovascular events

2.6. Exploratory Outcomes

- Use and type of oral nutritional supplements .

3. Statistical analysis:

3.1. General information

3.1.1. Descriptive component:

This study includes demographic variables, medical history, hemodialysis treatment, laboratory parameters, medication, nutritional variables, variables related to hospitalization, variables related to mortality, and end-of-follow-up variables. The descriptive component will be carried out taking into account the continuous, ordinal , or nominal nature of the above variables, taking into account the following procedures:

1. Continuous variables: Means or medians will be used as numerical methods to describe variables, along with their corresponding dispersion measures: standard deviation (SD) or interquartile range (IQR). Box plots will be used as graphical methods to evaluate the shape of the distribution, as well as to assess the

presence of symmetry and extreme values. Normal quantile graphs will be used as a tool to assess the normality of the variables.

2. Ordinal variables: Medians (IQR) or absolute and relative frequency measures will be used as summary measures in the case of variables with six or fewer modalities. Bar graphs will be used as summary graphic methods.
3. Nominal and count variables: Absolute frequencies and percentages will be used for their description using numerical methods. The summary graphic component will be carried out using bar graphs.

For the outcome variables (time between cohort entry and death, time between cohort entry and hospitalization, presence of non-fatal cardiovascular events, and days of hospitalization) their estimators will be calculated together with their 95% confidence intervals.

Each of these analyses will be performed for the entire sample, as well as for each of the exposure groups (MCO and HD AF).

3.1.2. Analytical component:

3.1.2.1. Bivariate analysis

1. Since there is no random assignment to treatments, these analyses will be performed to compare baseline characteristics between the two groups and to assess differences between variables with repeated measures, over time. The variables that will be incorporated into these analyses will be: Analysis are Albumin, PEW, MIS, hs CRP, dialyzer type. Longitudinal analysis of covariance will be used considering that in this approach the outcome variable, which is measured during different time periods, is adjusted by the baseline value.

3.1.2.2. Multivariate analysis

Considering that the results of analyses that include effect measures may be compromised by potential confounding bias given the non-experimental design of the study, it has been proposed that statistical models incorporate propensity score analyses using inverse probability of treatment weighting (IPTW) methods. These analyses will have the following steps:

These analyses will have the following steps [1]:

1. Data preparation: This stage includes the selection of covariates, and the assessment of data balance and the pattern of missing data. Not only the variables that are associated with the type of dialysis and with each of the outcomes used in this research will be selected, but also those that are associated only with the outcome, since it has been reported that this decision can increase the power to evaluate the treatment effect [2]. In the event that the handling of missing data is required, multiple imputation methods with chained equations will be carried out, carrying out the imputation separately in each exposure group [3].
2. Propensity score (PS) estimation : Given the dichotomous nature of the exposure, these scores will be estimated using logistic regression methods [4].
3. Application of the PS method: As mentioned before, the IPTW method will be applied, considering that it has advantages such as sample size optimization (it does not eliminate observations that cannot be matched) and versatility for its use with different multivariate analysis methods [5].

[1] Austin PC, Stuart EA. (2015) Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 10;34(28):3661-79.

[2] Cuong, N. V. (2013). Which covariates should be controlled in propensity score matching? Evidence from a simulation study. Statistica Neerlandica, 67(2), 169–180.

[3] Puma, MJ, Olsen, R.B., Bell, S.H., & Price, C. (2009). What to do when data are missing in group randomized controlled trials. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

[4] Setoguchi, S., Schneeweiss, S., Brookhart, MA, Glynn, RJ, & Cook, E.F. (2008). Evaluating uses of data mining techniques in propensity score estimation: A simulation study. Pharmacoepidemiology and Drug Safety, 17(6), 546–555.

[5] Austin PC, Stuart EA. (2017) The Performance of Inverse Probability of Treatment Weighting and Full Matching on the Propensity Score in the Presence of Model Misspecification When Estimating the Effect of Treatment on Survival Outcomes. Stat Methods Med Res, 26(4), 1654-1670

4. Evaluation of the balance of covariates: It will be carried out with graphical methods (QQ plots for continuous variables and bar graphs for nominal variables), and with descriptive numerical methods (standardized mean differences and variance ratios) and inferential methods (t tests and rank sum tests) [6].
5. Estimation of treatment effect: Treatment effects, according to the type of outcome, will be estimated using generalized linear models with weighting of the observations in the sample [7].
6. Sensitivity analysis: This analysis will be performed to determine the magnitude of the bias that may be produced by the omission of covariates. For this analysis, the methodology proposed by Carnegie, Harada, Dorie and Hill will be used [8].

In accordance with the specific objectives set out in the protocol, the following types of analysis are proposed:

1. For the objectives related to the comparison of time to hospitalization and time to death event between both cohorts, the respective incidence rates for each group and the survival functions will be calculated using Kaplan-Meier estimators. Comparisons between the survival functions of each of the two groups will be made with log- rank , Tarone-Ware or Peto-Peto Prentice tests, depending on the censoring pattern. Additionally, Cox models will be used to estimate the strength of association between type of treatment and time to event using the HRs metric . In these models, weighted samples will be used with the IPTW method and a group of time-dependent covariates will be incorporated (serum albumin, PEW, MIS, hs CRP, type of dialyzer).
2. To meet the objectives related to counting variables (days of hospitalization and non-fatal cardiovascular events), Poisson regression models or negative binomial

[6] Austin P.C. (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 10;28(25):3083-107

[7] Austin P.C. (2016) Variance Estimation When Using Inverse Probability of Treatment Weighting (IPTW) With Survival Analysis. Stat Med. 35 (30), 5642-5655

[8] Carnegie, N.B., Harada, M., Dorie, V., & Hill, J. (2016). treatSens: Sensitivity analysis for causal inference. Retrieved from https://cran.r-project.org/web/packages/treatSens/index.html

regression models will be used in case of overdispersion in the outcome variable. Within these models, the covariates that will be candidates for inclusion in the models are:

   a. For the outcome days of hospitalization: Serum albumin, PEW, MIS, hs CRP, dialyzer type

3. For non-fatal cardiovascular events outcome: Serum albumin, PEW, MIS, hs CRP, dialyzer type
4. For the objective related to the estimation of the association between serum albumin levels and the outcomes of mortality, non-fatal cardiovascular events, hospitalization events and days of hospital stay, the use of linear mixed models for longitudinal data is proposed. This strategy allows modeling the association between repeated measurements of independent variables and the proposed outcomes, in the presence of other variables that also vary over time. These models involve the combination of fixed effects and random effects models, and allow the inclusion of variables that change over time. To identify potential confounding variables, a directed acyclic causal diagram (DAG) will be performed. Additionally , G methods will be used for longitudinal data, which allow estimating the causal effect of an exposure that varies over time (in this case albumin levels) in the presence of confounding variables that also vary over time (hs CRP, MIS, PEW, FRR, type of dialyzer) that are in turn affected by the exposure. These methods may be more robust than traditional regression-based methods. The parametric G- formula and the weighted inverse probability of treatment estimation with marginal structural models [9]will be used [10].[11]

[9] James M. Robins, Sander Greenland, Fu-Chang Hu. Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome, J Am Stat Assoc. 1999; 94:447, 687-700.

[10] Miguel A Hernán, Babette Brumback, James M Robins. Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments. J Am Stat Assoc. 2001; 96:454, 440-448.

[11] Sally Picciotto, Miguel A. Hernán, John H. Page, Jessica G. Young, James M. Robins. Structural Nested Cumulative Failure Time Models to Estimate the Effects of Interventions. J Am Stat Assoc. 2012;107:499, 886-900.

3.1.5. Handling missing data

Considering that in the present study data from a standardized clinical registry of proven quality in previous research will be used, a low amount of missing data is expected in the study variables. However, it is noted that the amount or pattern of missing data may affect the quality of the statistical estimates. Missing data and the mechanism by which they occur will be evaluated: MCAR (Missing Data). completely at random), MAR (Missing at random) and MNAR (Missing not at random), according to this evaluation, the appropriate data imputation procedures will be defined for each case

3.1.5 . Sample size:

Based on the availability of the assembled cohorts of 1098 patients, 564 on MCO in the Theranova® dialyzer cohort and 534 on conventional HD in the AF dialyzer cohort. This provides 90% power considering an IRR 0.82 for all-cause hospitalization and a significance level of 0.05 (26).

3.2. Detailed analysis proposal

3.2.1. Descriptive component:

The following tables show how the results of the descriptive component of the analysis will be reported, according to the different variables considered in the study:

Table 1Demographic variables

Variable	HDx	HF-HD	Total
Sex			
Male	n(%)	n(%)	n(%)
Female	n(%)	n(%)	n(%)
Age	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Ethnicity			
Indigenous	n(%)	n(%)	n(%)
African American	n(%)	n(%)	n(%)
Mestizo	n(%)	n(%)	n(%)
City			
1	n(%)	n(%)	n(%)
2	n(%)	n(%)	n(%)
3	n(%)	n(%)	n(%)
4	n(%)	n(%)	n(%)
5	n(%)	n(%)	n(%)
6	n(%)	n(%)	n(%)
7	n(%)	n(%)	n(%)
Insurer			
1	n(%)	n(%)	n(%)
2	n(%)	n(%)	n(%)
3	n(%)	n(%)	n(%)
4	n(%)	n(%)	n(%)
5	n(%)	n(%)	n(%)
Kidney clinic			
1	n(%)	n(%)	n(%)
2	n(%)	n(%)	n(%)
3	n(%)	n(%)	n(%)
4	n(%)	n(%)	n(%)
5	n(%)	n(%)	n(%)
6	n(%)	n(%)	n(%)
7	n(%)	n(%)	n(%)
8	n(%)	n(%)	n(%)
9	n(%)	n(%)	n(%)
10	n(%)	n(%)	n(%)
11	n(%)	n(%)	n(%)
12	n(%)	n(%)	n(%)

Table 2Medical history variables

Variable	**HDx**	**HF-HD**	**Total**
Seniority in dialysis	Median (IQR)	Median (IQR)	Median (IQR)
Etiology of CKD			
1= Hypertension	n(%)	n(%)	n(%)
2= Diabetes mellitus	n(%)	n(%)	n(%)
3= Autoimmune glomerular	n(%)	n(%)	n(%)
4= Obstructive	n(%)	n(%)	n(%)
5= Disease polycystic	n(%)	n(%)	n(%)
6= Unknown	n(%)	n(%)	n(%)
7= Other	n(%)	n(%)	n(%)
Dx Hypertension	n(%)	n(%)	n(%)
Dx Diabetes	n(%)	n(%)	n(%)
Dx Cardiovascular Disease	n(%)	n(%)	n(%)
Charlson index	Median (IQR)	Median (IQR)	Median (IQR)
Karnofsky	Median (IQR)	Median (IQR)	Median (IQR)
Urine (diuresis ml/day)			
< 100 ml/day	n(%)	n(%)	n(%)
100 to 400 ml/day	n(%)	n(%)	n(%)
>400 ml/day	n(%)	n(%)	n(%)

Table 3. 3

Variable	**HDx**	**HF-HD**	**Total**
Duration of weekly sessions, hours	Median (IQR)	Median (IQR)	Median (IQR)
Dialyzer	n(%)	n(%)	n(%)
QD	Median (IQR)	Median (IQR)	Median (IQR)
QB	Median (IQR)	Median (IQR)	Median (IQR)
PHEW	Median (IQR)	Median (IQR)	Median (IQR)
Type of access			
1= Temporary catheter	n(%)	n(%)	n(%)
2= Catheter tunneled	n(%)	n(%)	n(%)
3= Native FAV	n(%)	n(%)	n(%)
4= AVF graft	n(%)	n(%)	n(%)
Anticoagulant			
1= Heparin	n(%)	n(%)	n(%)
2= Heparin LMW	n(%)	n(%)	n(%)
3= Not used	n(%)	n(%)	n(%)
Dose anticoagulant	Median (IQR)	Median (IQR)	Median (IQR)
TA_Systolic pre	Median (IQR)	Median (IQR)	Median (IQR)
TA_Pre- diastolic	Median (IQR)	Median (IQR)	Median (IQR)
Post Systolic BP	Median (IQR)	Median (IQR)	Median (IQR)
TA_Post diastolic	Median (IQR)	Median (IQR)	Median (IQR)

Table 4Laboratory parameter variables

Variable	HDx	HF-HD	Total
Albumin	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
PTHi	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Phosphorus	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Calcium	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Potassium	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Hemoglobin	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Platelets	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Lymphocytes	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Iron	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Ferritin	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
SATT	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
TIBC	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
PCRHS	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
BUN pre	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
BUNpost	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Kt/V	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Cholesterol	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
HDL	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
LDL	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Triglycerides	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)

Table 5Medication variables

Variable	HDx	HF-HD	HF-HD
Nutritional supplements yes or no	n(%)	n(%)	n(%)
Type of supplement			
1=Full formula	n(%)	n(%)	n(%)
2=Specialized complete formula	n(%)	n(%)	n(%)
1=Modular formula	n(%)	n(%)	n(%)

*Variables measured every 6 months

*Table 6variables **

Variable	HDx	HF-HD	HF-HD
BMI	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Categorized BMI			
1= Underweight (<18.5)	n(%)	n(%)	n(%)
2 = Normal (18.5 – 24.9)	n(%)	n(%)	n(%)
3 = Overweight (25.0 – 29.9)	n(%)	n(%)	n(%)
4= Obesity (>=30)	n(%)	n(%)	n(%)
PEW	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
PEW categorized			
0=No wear (0-2)	n(%)	n(%)	n(%)
1=With wear (>=3)	n(%)	n(%)	n(%)
Malnutrition-Inflammation Scale			
Hydration status			
0= No signs and symptoms	n(%)	n(%)	n(%)
1=Edema below the knees	n(%)	n(%)	n(%)
2=Edema above knees	n(%)	n(%)	n(%)

Karnofsky	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)
Nutritional assessment			
1	n(%)	n(%)	n(%)
2	n(%)	n(%)	n(%)
3	n(%)	n(%)	n(%)

*Variables measured every 6 months

Table 7Hospitalization variables

Variable	**HDx**	**HF-HD**	**HF-HD**
Hospitalization yes or no	n(%)	n(%)	n(%)
Diagnosis in hospitalization			
1=Cat 1	n(%)	n(%)	n(%)
2=Cat2	n(%)	n(%)	n(%)
3=Cat3	n(%)	n(%)	n(%)
Major cardiovascular event yes or no	n(%)	n(%)	n(%)
Fatal cardiovascular event yes or no	n(%)	n(%)	n(%)
COVID in hospitalization	n(%)	n(%)	n(%)
Days of hospital stay	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)	Mean (SD), Median (IQR)

*Table 8Mortality variables**

	HDx	**HF-HD**	**Total**
Death yes or no	n(%)	n(%)	n(%)
Cause of death			
1=Cancer	n(%)	n(%)	n(%)
2= Cardiovascular	n(%)	n(%)	n(%)
3=Infectious	n(%)	n(%)	n(%)
4=Metabolic	n(%)	n(%)	n(%)
5=Respiratory	n(%)	n(%)	n(%)
6 = COVID	n(%)	n(%)	n(%)

7=Other	n(%)	n(%)	n(%)
8= Unknown	n(%)	n(%)	n(%)
Dialysis-related death yes or no	n(%)	n(%)	n(%)

*Variables measured every 6 months

*Table 9End-of-follow-up variables**

	HDx	**HF-HD**	**Total**
Reason for termination of follow-up			
1= Death	n(%)	n(%)	n(%)
2= Loss of tracking	n(%)	n(%)	n(%)
3= Change of health provider/insurer	n(%)	n(%)	n(%)
4= Recover kidney function	n(%)	n(%)	n(%)
5= Suspension of treatment	n(%)	n(%)	n(%)
6= Transplant	n(%)	n(%)	n(%)
7= Dialyzer change	n(%)	n(%)	n(%)
8= Changed to DP	n(%)	n(%)	n(%)
9= End of study	n(%)	n(%)	n(%)

*Variables measured every 6 months

4. Data extraction and preparation

After the measurement of variables and the collection of data are completed, the sponsor of the study must provide the analysis group with a file containing the results of the measurements of the variables considered in the project. It is recommended that this file be in a plain text format, preferably as comma-separated text (*. csv) or tab-separated text (*. txt). Additionally, the sponsor is required to provide a file with the variable dictionary that must contain the following information:

1. For continuous variables and outcome-related variables: The names of the variables considered in the database, the labels for each variable, and the values corresponding to the lower and upper ends of the range must be included.
2. For ordinal variables: Names of the variables included in the database, labels for each variable, labels for the categories of each variable, and the values corresponding to the lower and upper ends of the range.
3. For nominal variables: Names of the variables included in the database, labels for each variable, and labels for the categories of each variable.
4. Dates included in the database must be recorded consistently, using a single format for the entire database.
5. The database must contain an identification code for each patient to ensure the anonymity of the information provided by the participants.

After having received the database and the variable dictionary to their satisfaction, the personnel in charge of the analysis will sign a confidentiality agreement.
Subsequently, the group of data analysts will carry out the following activities:

1. Importing data into the Stata® program.
2. Creating a working copy of the original database.
3. Using the working copy, verify both the integrity of the database and the names and labels of variables and categories.
4. Using the working copy, quality check of the database: For each variable, a review of missing values and consistency of information will be carried out.

Any problems detected in the above activities will be reported to the sponsor of the study, so that the information in the original source of the data can be reviewed and the necessary corrections can be made to ensure the quality and validity of the database.
All stages related to the extraction and preparation of the database will be documented in a *.do file within the Stata® program.

5. Preliminary analysis and draft report

Once the statistical analyses have been completed, the results will be presented in tabular and graphical formats. These preliminary results will be delivered in an editable Word file, as a draft report. As additional material, *.do files will be delivered with documentation of the database recruitment and analysis processes. Based on the results of the review of the draft report, the sponsor of the study may propose additional analyses.

6. Final analysis and results report

Based on the recommendations resulting from the previous stage, it may be necessary to perform new analyses or to expand existing ones. In this case, the new results will be included in the final report of the statistical analysis plan. This last phase has the following products:

1. Delivery of the original database.
2. Delivery of the working databases, appropriately validated and labeled, in Stata format.
3. Delivery of database preparation documentation files and statistical analysis files in *.do format.
4. Delivery of a final report, in Word format, with the results of the analyses carried out.

7. Schedule

The following table presents the estimated duration for each of the processes related to statistical analysis:

Task	Timeline
Data extraction and preparation of databases	3 weeks
Preliminary analysis and draft report	2 weeks
Review by the sponsor	1 week
Final analysis and results report	3 weeks