**Renal Outcome with** **Empagliflozin in non-diabetic Chronic Kidney Disease Patients: A Randomized Control Trial**

**Statistical Analysis Plan**

**(SAP)**

Contents

[1 INTRODUCTION 3](#_Toc87363710)

[2 DATA SOURCE AND COLLECTION PROTOCOL 3](#_Toc87363711)

[3 ANALYSIS OBJECTIVES 5](#_Toc87363712)

[4 SAMPLE SIZE AND POWER CONSIDERATIONS 6](#_Toc87363713)

[5 ENDPOINTS AND COVARIATES 8](#_Toc87363714)

[6 HANDLING OF MISSING VALUES AND OTHER DATA CONVENTIONS 9](#_Toc87363715)

[7 STATISTICAL ANALYSES 9](#_Toc87363716)

[8 REFERENCES 11](#_Toc87363717)

# 1 INTRODUCTION

In the context of CKD etiologies, randomized controlled trials (RCTs) play a fundamental role in establishing evidence of the drug effects. Therefore, the statistical analysis of data in such RCTs has a core significance. This section addresses every detail relevant to collecting the data, the collection protocol, analyzing the data, and using the most relevant statistical methods to accomplish the goal of this study.

# 2 DATA SOURCE AND COLLECTION PROTOCOL

Inclusion criteria are the following three:

* Non-diabetic CKD patients, stage 3&4.
* Age ≥ 18 years.
* Both male & females

The following are the five exclusion criteria:

* Major surgery during the last 6 months
* Pregnant or nursing mothers
* Patients on transplantation list
* Patients on dialysis
* Patients who are allergic to any drug
* Patients with malignancy

Patients meeting the inclusion criteria will be included in the study after they have signed the informed consent form. Patients will come to the hospital for five visits: baseline, 1 month, 2 months, 3 months, and final assessment (Figure 1). During the first visit baseline measurements will be recorded, and the three following visits will allow for the 30-day, 60-day, and 90-day measurements respectively.

Those patients who will return on visit 2, will be randomized into one of the treatment groups in a 1:1 ratio. A randomization list will be generated by IRT system.

Baseline investigations will be done at this time. The patients will then be asked to return at 1, 2 and at 3rd month. In between this time, telephonic follow up will be maintained with patients for their safety and in case any adverse events start to emerge.

Table 1 presents the thirty-five variables that will be collected from the subjects throughout the 90-day period. At the completion of the study, all the participants will be asked for a follow up after 30 day.

Figure 1. The five visits of the study

Table 1. The thirty-five variables of the study

|  |  |  |
| --- | --- | --- |
| Variable | Data Type | Description |
| Age | Continuous |  |
| Gender | Binary |  |
| Ethnicity | Categorical |  |
| TotalCholesterol | Continuous | Baseline Total Cholesterol  |
| HDL | Continuous | Baseline High density lipoprotein cholesterol |
| LDL | Continuous | Baseline Low density lipoprotein cholesterol |
| Triglycerides | Continuous | Baseline Triglycerides |
| TotalCholesterol30 | Continuous | Total Cholesterol after 30 days |
| HDL30 | Continuous | High density lipoprotein cholesterol after 30 days |
| LDL30 | Continuous | Low density lipoprotein cholesterol after 30 days |
| Triglycerides30 | Continuous | Triglycerides after 30 days |
| TotalCholesterol60 | Continuous | Total Cholesterol after 60 days |
| HDL60 | Continuous | High density lipoprotein cholesterol after 60 days |
| LDL60 | Continuous | Low density lipoprotein cholesterol after 60 days |
| Triglycerides60 | Continuous | Triglycerides after 60 days |
| TotalCholesterol90 | Continuous | Total Cholesterol after 90 days |
| HDL90 | Continuous | High density lipoprotein cholesterol after 90 days |
| LDL90 | Continuous | Low density lipoprotein cholesterol after 90 days |
| Triglycerides90 | Continuous | Triglycerides after 90 days |
| Proteinuria | Continuous | Baseline urinary protein excretion greater than 500mg per day  |
| Proteinuria30 | Continuous | Urinary protein excretion greater than 500mg per day, after 30 days  |
| Proteinuria60 | Continuous | Urinary protein excretion greater than 500mg per day, after 60 days  |
| Proteinuria90 | Continuous | Urinary protein excretion greater than 500mg per day, after 90 |
| EGFR | Continuous | Baseline Estimated Glomerular filtration rate |
| EGFR30 | Continuous | Estimated Glomerular filtration rate, after 30 days |
| EGFR60 | Continuous | Estimated Glomerular filtration rate, after 60 days |
| EGFR90 | Continuous | Estimated Glomerular filtration rate, after 90 days |
| BMI | Continuous | Baseline body mass index |
| BMI30 | Continuous | Body mass index after 30 days |
| BMI60 | Continuous | Body mass index after 60 days |
| BMI90 | Continuous | Body mass index after 90 days |
| UTI | Integer | Baseline frequency of urinary tract infection |
| UTI30 | Integer | Frequency of urinary tract infection after 30 days |
| UTI60 | Integer | Frequency of urinary tract infection after 60 days |
| UTI90 | Integer | Frequency of urinary tract infection after 90 days |

# 3 ANALYSIS OBJECTIVES

This two arm clinical trial follows a pre-post design with a continuous outcome (eGFR change). The aim of the analysis is to measure the difference in post-treatment scores between the groups. The corresponding point estimators of treatment effect are all unbiased.

# 4 SAMPLE SIZE AND POWER CONSIDERATIONS

This study aims at assessing whether Empagliflozin added with Standard therapy slows down the progression of CKD compared to standard therapy alone in non-diabetic CKD patients. It is estimated that approximately 115 patients per arm will provide a power of 95% to detect a significant effect in the progression of CKD with Empagliflozin and Standard therapy, assuming type I error (alpha) of 5%, reference survival exponential hazard rate of 7.5%, and hazard ratio of 61% (Table 3, Figure 2). It is estimated that approximately 115 patients per arm will enable a power of 95% to detect a beneficial effect of 0.95. Table 2 presents the range of sample size per arm from a power of 80% to 95% (Figure 3). According to Lachin (1988), a randomized clinical trial with a total sample size of more than 200 patients allows for a near perfect balance.

Table 2. Log-Rank Test for Two Survival Curves

|  |  |
| --- | --- |
| Method | Lakatos normal approximation |
| Form of Survival Curve 1 | Exponential |
| Form of Survival Curve 2 | Exponential |
| Number of Sides | 2 |
| Follow-up Time | 30 |
| Total Time | 90 |
| Alpha | 0.05 |
| Reference Survival Exponential Hazard | 0.075 |
| Hazard Ratio | 0.61 |
| Number of Time Sub-Intervals | 12 |
| Group 1 Loss Exponential Hazard | 0 |
| Group 2 Loss Exponential Hazard | 0 |

Table 3. Estimated number of patients per arm, per power.

|  |  |  |  |
| --- | --- | --- | --- |
| Index | Nominal Power | Actual Power | N per Group |
| 1 | 0.80 | 0.805 | 70 |
| 2 | 0.85 | 0.854 | 80 |
| 3 | 0.90 | 0.901 | 93 |
| 4 | 0.95 | 0.951 | 115 |



Figure 2. Sample size per group versus hazard ratio.



Figure 3. Range of sample size per group per power.

# 5 ENDPOINTS AND COVARIATES

This study has three main outcomes: change in eGFR (using the CKD-EPI creatinine equation) from baseline, change in proteinuria from baseline, and change in lipid profile from baseline. These are continuous variables. Table 4 presents the variables contributing to the outcomes and Table 5 presents the eleven covariates of the study.

Table 4. The variables that contribute to the three outcomes of the study

|  |  |  |  |
| --- | --- | --- | --- |
| Outcome | Variable | Data Type | Description |
| Change in Lipid Profile | TotalCholesterol | Continuous | Baseline Total Cholesterol  |
| HDL | Continuous | Baseline High density lipoprotein cholesterol |
| LDL | Continuous | Baseline Low density lipoprotein cholesterol |
| Triglycerides | Continuous | Baseline Triglycerides |
| TotalCholesterol30 | Continuous | Total Cholesterol after 30 days |
| HDL30 | Continuous | High density lipoprotein cholesterol after 30 days |
| LDL30 | Continuous | Low density lipoprotein cholesterol after 30 days |
| Triglycerides30 | Continuous | Triglycerides after 30 days |
| TotalCholesterol60 | Continuous | Total Cholesterol after 60 days |
| HDL60 | Continuous | High density lipoprotein cholesterol after 60 days |
| LDL60 | Continuous | Low density lipoprotein cholesterol after 60 days |
| Triglycerides60 | Continuous | Triglycerides after 60 days |
| TotalCholesterol90 | Continuous | Total Cholesterol after 90 days |
| HDL90 | Continuous | High density lipoprotein cholesterol after 90 days |
| LDL90 | Continuous | Low density lipoprotein cholesterol after 90 days |
| Triglycerides90 | Continuous | Triglycerides after 90 days |
| Change in Proteinuria | Proteinuria | Continuous | Baseline urinary protein excretion greater than 500mg per day  |
| Proteinuria30 | Continuous | Urinary protein excretion greater than 500mg per day, after 30 days  |
| Proteinuria60 | Continuous | Urinary protein excretion greater than 500mg per day, after 60 days  |
| Proteinuria90 | Continuous | Urinary protein excretion greater than 500mg per day, after 90 |
| Change in eGFR | EGFR | Continuous | Baseline Estimated Glomerular filtration rate |
| EGFR30 | Continuous | Estimated Glomerular filtration rate, after 30 days |
| EGFR60 | Continuous | Estimated Glomerular filtration rate, after 60 days |
| EGFR90 | Continuous | Estimated Glomerular filtration rate, after 90 days |

Table 5. The eleven covariates of the study

|  |  |  |
| --- | --- | --- |
| Variable | Data Type | Description |
| Age | Continuous |  |
| Gender | Binary |  |
| Ethnicity | Categorical |  |
| BMI | Continuous | Baseline body mass index |
| BMI30 | Continuous | Body mass index after 30 days |
| BMI60 | Continuous | Body mass index after 60 days |
| BMI90 | Continuous | Body mass index after 90 days |
| UTI | Integer | Baseline frequency of urinary tract infection |
| UTI30 | Integer | Frequency of urinary tract infection after 30 days |
| UTI60 | Integer | Frequency of urinary tract infection after 60 days |
| UTI90 | Integer | Frequency of urinary tract infection after 90 days |

# 6 HANDLING OF MISSING VALUES AND OTHER DATA CONVENTIONS

Baseline missing values or drop-out values that could happen in the following visits are expected and will be handled accordingly, to ensure balance between the two arms. Geiker et al. (2016) suggests considering the baseline outcome as a response in the following visits, applying an imputation technique, or a combination of both methods.

# 7 STATISTICAL ANALYSES

7.1 EXPLORATORY DATA ANALYSIS

Once the clinical trial has been started, baseline data on the subjects will be collected and re-measured again after 30, 60, and 90 days. This data will be summarized to calculate the mean, median, interquartile range (IQR) and standard deviation for continuous variables. Categorical variables will be summarized by means of counts and percentages per category.

Furthermore, normality will be investigated for continuous variables since it is an essential assumption for the proposed analysis of variance and generalized linear modelling techniques.

7.2 DATA TRANSFORMATION

If the continuous variables are not normally distributed, they will be transformed using the natural logarithm function. This ensures that the normality assumption is not violated when conducting the statistical methods.

7.3 CORRELATION ANALYSIS

Pairwise correlation coefficient will be calculated among the variables to explore between-participant variation. Moreover, graphs will be drafted to visualize the relationships among the variables.

7.4 MULTIVARIATE ANALYSIS OF COVARIANCE

For this type of clinical trial, multivariate analysis of covariance (MANCOVA) main effect model is the most efficient approach with smallest variance estimator. MANCOVA is more meaningful conceptually because pre-score is the baseline covariates, and because there are multiple outcomes. MANCOVA tests the overall null hypothesis that all groups have the same means on the various dependent variables.

The following are few of the main assumptions of MANCOVA. SAS code tests them before conducting MANCOVA.

1. Normal Distribution: The dependent variable should be normally distributed within groups.
2. Linearity: MACNOVA assumes that there are linear relationships among all pairs of dependent variables, all pairs of covariates, and all dependent variable-covariate pairs in each cell. Therefore, when the relationship deviates from linearity, the power of the analysis will be compromised.
3. Homogeneity of Variances: Homogeneity of variances assumes that the dependent variables exhibit equal levels of variance across the range of predictor variables. Homoscedasticity can be examined graphically or by means of a number of statistical tests (F test or Levene’s test).
4. Homogeneity of Variances and Covariances: In multivariate designs, with multiple dependent measures, the homogeneity of variances assumption described earlier also applies. However, since there are multiple dependent variables, it is also required that their intercorrelations (covariances) are homogeneous across the cells of the design. There are various specific tests of this assumption (e.g. Box's M test).

7.5 DATA INTEGRITY, TRIAL VALIDITY, AND SOFTWARE USED

SPSS will be used to collect the raw data. SAS will be used to manage the data and conduct the statistical analyses. These two applications are reliable and thus provide a high degree of data integrity and trial validity.

# 8 REFERENCES

1. Anderson, F. ANOVA by Example: Hands on Approach using SAS. US: KDP; 2017.
2. Anderson, F. Clinical Trials Statistics by Example: Hands on approach using R. US: CreateSpace Independent Publishing Platform; 2017.
3. Lachin, J.M. Properties of simple randomization in clinical trials. Controlled Clinical Trials. [Online]1988;9(4): 312–26. Available from: https://www.sciencedirect.com/science/article/abs/pii/0197245688900463?via%3Dihub [Accessed 9 November 2021]
4. Geiker NRW, Ritz C, Pedersen SD, Larsen TM, Hill JO, Astrup A. A weight-loss program adapted to the menstrual cycle increases weight loss in healthy, overweight, premenopausal women: a 6-mo randomized controlled trial. Am J Clin Nutr. 2016;104:15–20. https://doi.org/10.3945/ajcn.115.126565.