HiLo: A pragmatic trial of higher vs lower serum phosphate targets in patients undergoing hemodialysis

NCT04095039

Statistical Analysis Plan

Version 1.0, January 31, 2024

HiLo

A pragmatic trial of higher vs lower serum phosphate targets in patients undergoing hemodialysis

Statistical Analysis Plan
Version 1.0

Contents

LIST OF ABBREVIATIONS

Abbreviation	Explanation
DCRI	Duke Clinical Research Institute
EHR	Electronic health record
ESRD	End-stage renal disease
HR	Hazard ratio
IRB	Institutional review board
ITT	Intent-to-treat
PCR	Protein catabolic rate
PPY	Per-person year
PRO	Patient-reported outcomes
PTH	Parathyroid hormone
SAP	Statistical Analysis Plan
SD	Standard deviation
TSAT	Transferrin saturation

AMENDMENT HISTORY

Date	Brief description of change
09 07 2022	Final draft version of SAP
01/31/2024	Defined variable used in variance formula in Section 4.5.1; rejection of null hypothesis in secondary analysis clarified in Section 4.5.2; added Summarizing Statistical Output section.

1. STUDY DETAILS

1.1 Study Objectives

The primary objective of HiLo will be to test the primary and secondary study hypotheses:

Primary hypothesis: Compared to the current standard approach of targeting serum phosphate levels of <5.5 mg/dl, less stringent control of serum phosphate to target levels ≥6.5 mg/dl will improve time to all-cause mortality and number of hospitalizations. This hypothesis will be tested using a hierarchical endpoint that prioritizes improvements in all-cause mortality over hospitalizations.

Secondary hypotheses: The trial will test the secondary hypotheses that less stringent control of serum phosphate will result in increased time to death; lower per-person year (PPY) hospitalization rates; and increased serum albumin and protein catabolic rate (PCR), as markers of diet and nutrition.

1.2 Study Design

HiLo is a pragmatic, multicenter, open-label trial that will compare the effects on hospitalization, mortality, diet, and nutrition of two different phosphate management strategies in 4400 patients being treated with maintenance hemodialysis at 100-150 facilities. HiLo will randomize participants to either liberal control of serum phosphate, targeting ≥6.5 mg/dl (Hi treatment arm, which is the new intervention to be tested), or strict control of serum phosphate, targeting <5.5 mg/dl (Lo treatment arm, which is the current standard of care).

1.3 Number of Patients

Sample size calculations

A simulation was performed in order to determine whether a sample size of 3800 provides sufficient power to reject the null hypothesis of no difference between the treatment arms in time to all cause death and number of hospitalizations.

The simulation population was created under the following parameters:

- 1:1 randomization
- For each participant, administrative censoring occurs uniformly randomly between 2 to 4 years
- Annual loss to follow up rate of 5% (for example, due to kidney transplantation, change in dialysis modality, withdrawal due to participant choice, etc.)
- Annual mortality rate of 15% in the Lo treatment arm vs. 12.8% in the Hi treatment arm.
- An average of 35% of the study population are not susceptible to hospitalization and have no chance of experiencing a hospitalization during the course of the study (zero-inflated distribution of hospitalization)
- The remaining susceptible population experiences an average a 2 hospitalizations per year in the Lo treatment arm vs. an average of 1.89 in the Hi treatment arm.

The above population was simulated across 1000 iterations. The simulated population and outcomes were analyzed using the methods described in protocol section 4.5.1. Power was estimated as the number of iterations with a p-value less than 0.05 divided by the total number of iterations.

The estimated power from the simulation indicate that a sample size of 3800 patients will have more than 95% power.

1.4 Data Sources

The primary source of data for the HiLo study will be electronic health record data received from the dialysis provider partners of the study. Dialysis providers routinely collect data on their patients in their electronic health record in the process of care, and they will share the data necessary to perform the analyses specified in this document.

Among the study population that are Medicare beneficiaries, cause-specific hospitalization data will be collected from the Centers for Medicare & Medicaid Services (CMS) Virtual Research Data Center.

2. ANALYSIS POPULATIONS

2.1 Definition of Analysis Populations

2.1.1 Intention-To-Treat (ITT) Population

The ITT population will include all patients that have consented to participation in the study. The treatment arms in analyses in the ITT population will be the participant's assigned phosphate target arm, regardless of the participant's actual phosphate levels throughout the course of the trial. The ITT population will be used in the primary analysis of efficacy and all major secondary analyses.

2.1.2 In Range (IR) Population

The aim of the IR population is to include the patients that have reached the targeted phosphate levels for the trial. Table 1 in section 4.4.1 below presents the "in range" phosphate range for each treatment arm. Phosphate levels will be measured on a monthly basis for each patient in the trial. Patients that are in range for 75% of their monthly phosphate measures will be included in the IR population.

2.1.3 In Range 2 (IR2) Population

The IR2 population will be defined similarly to the IR population, except it will include patients that are in range for 50% of their monthly phosphate measures.

2.2 Protocol Deviations

Because this is a pragmatic trial intended to evaluate effectiveness of the phosphate targets under real-world conditions, protocol deviations will be limited to: 1) enrollment of individuals who do not meet eligibility criteria or provide informed consent, and 2) breaches of participant confidentiality. Deviations will be reported by the DCRI to the IRB.

3. STUDY ENDPOINTS

3.1 Primary Endpoint

The primary outcome for HiLo is the hierarchical composite of: 1) time to all cause death followed by 2) number of all cause hospitalizations. The use of this endpoint will allow for simultaneous testing of both mortality and hospitalizations while prioritizing mortality as the more important health outcome to consider. Time zero for the time to all cause death portion of the endpoint will be the date of the first dialysis session after the patient has consented to participate in HiLo.

3.2 Secondary Endpoints

Time to all cause death and PPY hospitalization rates will be investigated separately as secondary endpoints. The patient level PPY hospitalization rate will be calculated by dividing the patient's number of hospitalizations during the follow-up period by the total length of the patient's follow-up time in years.

Other secondary outcomes will be the change from baseline in the following clinical measures: serum albumin and PCR. The baseline measure for these will be the latest available measure before consent.

3.3 Additional Endpoints

The percent of hospitalization days during the follow-up period will be investigated as an additional endpoint. At the individual patient level, this endpoint will be calculated as the total number of days spent in the hospital divided by the total number of days of follow-up for that patient.

If data are available, an endpoint filtering out COVID-19 related death and hospitalizations will be created. This endpoint will be identical to the primary endpoint, however, deaths and hospitalizations that are attributed to COVID-19 will not be counted as events as they would be in the primary endpoint.

4. ANALYSIS METHODS

4.1 General Principles

This analysis plan is intended to support the primary and main secondary manuscripts planned for the HiLo study. Analysis plans for other manuscripts will be provided in separate documents.

In addition to specific analyses and presentations that are detailed in the following sections, results will be summarized for continuous measures using descriptive statistics, including the number of patients, mean, standard deviation, median and range as appropriate. For categorical variables, counts and percentage per treatment group will be presented. Unless otherwise specified, all statistical tests will have a two-sided significance level of α=0.05.

4.1.1 Control of Type I Error

To account for repeated significance testing of the accumulating data, the group sequential method of Lan and DeMets will be used as a guide for interpreting interim analyses. Monitoring boundaries for the primary endpoint will be based on a two-sided symmetric O'Brien-Fleming type spending function with an overall two-sided significance level of α=0.05. The O'Brien-Fleming approach requires large critical values early in the study but relaxes (i.e., decreases) the critical value as the trial progresses. A single planned interim analysis using alpha spending will be targeted to occur after the first 2000 to 2400 subjects have been enrolled with at least 24 months of median follow up time accrued. Additional interim analyses are not currently planned but may be performed upon request by the Data Safety and Monitoring Board (DSMB).

Assuming that the interim analyses take place as currently planned with just a single analysis after ~50% of subjects have been enrolled with at least 24 months of median follow up time accrued, the critical p-value for the test of the null hypothesis of no difference at the first interim analysis will be 0.001 and a result of $|z| < 3$ or $p > 0.001$ would suggest that the trial should continue. In order to maintain the overall 5% significance level of the trial, the critical p-value for declaring there to be a significant difference in the primary endpoint at the final analysis will be decreased from 0.05 to 0.049. However, the critical p-values for the interim analysis and final analysis can be adjusted appropriately if additional interim analyses are requested by the DSMB.

4.1.2 Censoring Scheme

Primary Censoring Scheme: for time-to-event endpoints, patients will be censored at the earliest of: 1) end of study date, and 2) date of dialysis status change.

End of study date is defined as the earliest of:

- date of withdrawal from study
- date of lost to follow-up
- date of study completion

Date of dialysis status change is the earliest of the following:

- date of transfer to another dialysis facility
- date of kidney transplantation
- date of transfer to peritoneal dialysis
- date of withdrawal from dialysis

4.2 Study Conduct

Major protocol deviations will be identified for all patients who are randomized (see Section 2.2)

4.3 Study Population

4.3.1 Patient Disposition

The number of patients included in each study population will be summarized by treatment group. The number and percentage of patients who completed, patients who had a change in status, and patients who withdrew from the study or were lost to follow-up will be presented for each treatment group and overall for the ITT population.

4.3.2 Demographics and Other Baseline Characteristics

Demographics and baseline characteristics will be summarized by treatment group for the ITT population. The following demographics and baseline characteristics will be summarized:

- Age
- Sex
- Race
- Ethnicity
- Height
- Weight
- Dialysis vintage
- Co-morbid illnesses noted on admission to the dialysis facility (using ICD-9/10 codes)
- Cause of end-stage renal disease
- Medication use, specifically including phosphate binders and activated vitamin D

4.4 Application of Study Treatment

4.4.1 Treatment Adherence

Serum phosphate will be monitored throughout the trial. Participants are anticipated to have phosphate levels measured at the site at least once per month. The participant monthly average phosphate level and the site level average of the patient average will be monitored for each site to ensure that the site assignment to a phosphate target has the intended effect of changing the patients' phosphate levels. The patient monthly average will be the average of all

phosphate measures in a calendar month. The target ranges for patients and site average for each arm are provided in Table 1. Sites will be provided a listing of patient IDs of patients falling outside of the "In range" range will be provided monthly.

Table 1. Phosphate target ranges for Hi and Lo arms.

Target phosphate (mg/dL)	Hi	Lo
In range	>6.5	<5.5
Near range	5.5 – 6.5	5.5 – 6.5
Out of range	<5.5	>6.5

4.4.2 Concomitant Medication

Partner clinics will provide all available data on medications on a monthly basis. These data should include phosphate binders, activated vitamin D, calcimimetics, and home medications.

4.5 Analysis of Primary and Secondary Endpoints

4.5.1 Analysis of the Primary Endpoint

The hierarchical endpoint of time to all-cause death followed by number of hospitalizations will be assigned a rank score using the method proposed by Finkelstein and Schoenfeld.[1]

First, each patient in the trial is assigned a rank score by comparing their outcomes with the outcomes of every other patient in the trial. The comparison occurs first on mortality: the patient that survived longer is the "winner" and the other patient that died first is the "loser". The winner in the comparison has their rank score increased by 1 while the loser has their rank score reduced by 1.

In order to make this comparison appropriately, at least one of the two patients being compared had to have an observed death prior to the other patient being censored. If both patients are censored or if censoring occurred before the observed death of the comparator patient, then it cannot be determined who died first. In cases where the comparison cannot be made on death, then the patients are then compared on hospitalizations.

The comparison on hospitalizations is done through the time of the earliest censoring time of the two patients being compared. At the comparison time, the patient with fewer hospitalizations is considered the winner and the patient with more hospitalizations is considered the loser. The rank score of the winners and losers is adjusted in the same way as described above. If both patients have the same number of hospitalizations at the comparison

time (including zero), then the rank score of both patients is unchanged based on the comparison.

Once a patient has been compared to all other patients in the trial, their final rank score describes their overall outcomes in comparison to the outcomes of the complete population. A higher rank score indicates the patient had better outcomes than most of the study population, while increasingly negative scores indicate worse outcomes.

To test the hypothesis of a difference between the treatment arms, the Finkelstein and Schoenfeld method will be employed. The test statistic will be

$$T = \sum_{i=1}^{N} U_i D_i$$

Where i is the participant indicator, N is the size of the analysis population, D_i is 1 in the treatment arm and 0 otherwise, and U_i is the value of the score described above.

Under the null hypothesis, the mean of T is zero and the variance is

$$V = \frac{p(N-p)}{N(N-1)} \sum_{i=1}^{N} U_i^2.$$

Here, p is the number of participants in the treatment arm. The value of $T/V^{1/2}$ will be compared to the standard normal distribution to determine the test p-value. A p-value of less than 0.05 will indicate that there is a difference between the treatment groups.

4.5.2 Analyses of the Secondary Endpoints

The hypothesis of the main secondary endpoint of time to all-cause death will be tested in the ITT population using a Cox Proportional Hazards model including phosphate target group as an explanatory factor. The hazard ratio and 95% CI for the high phosphate target versus the low phosphate target will be calculated from the estimated model parameters. If the confidence interval does not include 1, then the null hypothesis will be rejected.

The all-cause PPY hospitalization rates will be analyzed under the following hypotheses:

$$H_0: \mu_H = \mu_L \ \textit{vs.}\ H_1: \mu_H \neq \mu_L$$

Where μ_H and μ_L are the all-cause PPY hospitalization rates in the Hi and Lo treatment arms respectively. A two-sided 95% confidence interval for the difference in the cause-specific PPY hospitalization rates between the study arms will be estimated under normality assumptions. If confidence interval does not include 0, then the null hypothesis will be rejected.

For the continuous longitudinal outcomes (change from baseline in serum albumin and PCR), mixed effect models will be created to test the null hypothesis that the change of each measure

will not differ by treatment group. The phosphate target group and the baseline measure will be included as fixed effect covariates and study site will be included as a random effect covariate in the model.

Two covariance structures for this model will be considered: the compound symmetry structure (CS) and the auto-regressive(1) (AR1) structure. Each covariance structure will be compared using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The structure that provides better results on the AIC and the BIC will be chosen for the final adjustment model.

4.5.3 Additional Analyses

An adjusted analysis of the main secondary endpoints will be performed. The following adjustment variables will be included in the models:

- Age
- Sex
- Race
- Calculated BMI
- Cause of end-stage renal disease
- Vascular access type

The covariance structure for the model will be selected in the same manner as for the serum albumin and PCR analyses described in section 4.5.2.

An adjusted analysis of time to all-cause death will be performed by repeating the analysis of the endpoint described in section 4.5.2, but including the same adjustment variables listed for the adjusted analysis of the primary endpoint in the Cox model.

Similarly, the analyses for the serum albumin and PCR will be repeated including the same adjustment variables in their respective models.

The ratio of hospitalized days will be tested with the following hypotheses:

$$H_0: \mu_H = \mu_L \textit{ vs. } H_1: \mu_H \neq \mu_L$$

Where μ_H and μ_L are the mean ratio of hospitalized days in the Hi and Lo treatment arms respectively. If the two-sided 95% confidence interval for the difference in the mean ratios does not include 0, then the null hypothesis will be rejected.

Under a previous HiLo protocol, cluster randomization at the site level occurred but otherwise participants were enrolled and participated in the study in a similar manner to the current protocol. If it is determined to be appropriate, the primary outcomes of the participants enrolled under the cluster randomization will be analyzed separately. Furthermore, the results of this analysis will be combined with the primary analysis detailed in this SAP using meta-analysis methods. Details of both the analysis of the cluster randomized data and the meta-analysis will be provided in a future version of this document.

Additional analyses concerning COVID-19

Additional analyses are planned to investigate and control for the potential effect of the COVID-19 outbreak on the other planned analyses. If the data are available to construct an endpoint as described in section 3.3, then the primary and secondary analyses will be repeated using only non-COVID-19 related deaths and hospitalizations as endpoint events. Patients experiencing a death attributed to COVID-19 will be censored at the time of death.

If the data for the analyses described in the previous paragraph are not available, there may be other options to control for the possible confounding from COVID-19. If it is known whether an enrolled patient was infected by COVID-19 in the past or whether a patient was infected during follow-up, then patient level COVID-19 status will be used as a covariate in adjustment models using mixed models as described at the beginning of this section. Two adjustment models using COVID-19 infection status will be fit to the data: 1) a model using infection status as the only adjustment covariate 2) a repeat of the adjustment model described at the beginning of this section including infection status as an additional covariate.

COVID-19 infection may lead to patients being removed from the study site for an extended time period (e.g. extended hospitalization, transfer to non-study dialysis clinic designated for infected patients, etc.) In these cases, it is unlikely that patients that should be receiving a higher phosphate target will continue on that target and will be treated so as to reduce phosphate to the standard of care. If the date of a patient's transfer away from their study site or dietitian and the date of return are available, then there will be an additional analysis repeating the primary analysis disregarding events that occur while the patient is not receiving dialysis at the study site. If a death occurs during a period away from the study site, then the patient will be censored at the date of transfer away from the study site. Additionally, hospitalizations will not be compared using the count of hospitalizations. Instead, the analysis will use hospitalization rate calculated as the number of hospitalizations occurring while the patient is at study site divided by the time that the patient received dialysis at the study site.

All primary, secondary, and additional analyses will be repeated in the IR population.

4.6 Other Safety Considerations

4.6.1 Adverse Events

Among patients undergoing hemodialysis, adverse events of moderate or higher severity are extremely common and usually result in hospitalizations. Since hospitalization is part of the primary outcome of HiLo, additional information on adverse events will not be collected.

4.6.2 Clinical Safety Laboratory Events

Laboratory measurements are collected either monthly or quarterly for all patients as part of routine clinical care. Predetermined laboratory safety events will be collected. Below are defined the laboratory events of interest in HiLo:

- Hypophosphatemia, defined as serum phosphate <2.0 mg/dL
- Hyperphosphatemia, defined as serum phosphate >7.5 mg/dL
- Hypercalcemia, defined as total uncorrected serum calcium >10.5 mg/dL
- Excessive iron supplementation, defined as serum ferritin >1000 ng/mL and transferrin saturation (TSAT) >50%
- Secondary hyperparathyroidism, defined as parathyroid hormone (PTH) levels greater than 9 times the upper limit of normal for the assay

A listing of laboratory safety events will be created, along with a table of counts of each type of event by treatment group.

4.7 Interim Analysis

An interim analysis is planned for this trial when 50% of planned patients have enrolled with at least 24 months of median follow-up time accrued.

The interim analysis will investigate for futility at two possible levels: futility of enrollment, futility of serum phosphate separation (inability to sustain a separation of ≥0.75 mg/dl between arms).

A detailed interim analysis plan will be provided in a separate document.

5. CONVENTIONS

5.1 Baseline Measurements

Unless specified otherwise, a baseline value is the last assessment taken prior to the first dialysis session after consent. When there is a missing baseline assessment, it will not be imputed, thus, patients will be excluded from any changes from baseline analysis for which they have a missing baseline value.

5.2 Missing Dates

If a date is missing just the date portion (i.e. year and month are reported) then the date will be imputed with “15”. If both the month and date portion are missing, then the date will be

imputed with “June 15”. If the imputed date with this method falls before the date of consent, then the date will be imputed with the consent date instead. Likewise, if the imputed date falls after the end of study date, then the date will be imputed with the end of study date.

If additional information is available that allows for a date that is missing both month and day to be deduced as occurring between two other dates, then the date will be imputed as the midpoint between the two known dates that the missing date occurs between. An example of this situation would be if a dialysis date is missing, but is known to have occurred between two other dialysis dates.

5.3 Other Missing Data

No imputation will be made for missing data other than the date imputation as described section 5.2.

6. CHANGES OF ANALYSIS FROM PROTOCOL

There are no planned changes in analysis from what is specified in the study protocol.

7. SUMMARIZING STATISTICAL OUTPUT

A separate document is provided that details planned statistical output. The planned statistical output may be updated during the course of the trial.

8. REFERENCES

1. Finkelstein DM, Schoenfeld DA. Combining Mortality and Longitudinal Measures in Clinical Trials. Statist Med 1999; 18:1341-1354.