Statistical Analysis Plan

Study Code	D5740C00002
Edition Number	4.0
Date	28 September 2018

A Phase 3, Multicenter, Randomized, Open-Label, Active-Controlled Study of the Safety and Efficacy of Roxadustat in the Treatment of Anemia in Dialysis Patients

A Phase 3, Multicenter, Randomized, Open-Label, Active-Controlled Study of the Safety and Efficacy of Roxadustat in the Treatment of Anemia in Dialysis Patients

Study Statistician PPD _____ 28 Sep 2018
Date

A Phase 3, Multicenter, Randomized, Open-Label, Active-Controlled Study of the Safety and Efficacy of Roxadustat in the Treatment of Anemia in Dialysis Patients

Global Product Statistician

PPD

2018-09-29
Date

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
ALT	Alanine aminotransferase
AST	Aspartate transaminase
ANCOVA	Analysis of Covariance
BIW	Twice weekly
BP	Blood pressure
CHr	Reticulocyte hemoglobin content
CKD	Chronic kidney disease
CKD-DD	Chronic kidney disease with subject on dialysis
CKD-ND	Chronic kidney disease with subject not on dialysis
CMH	Cochran-Mantel-Haenszel
CRF	Case report form
CRP	C-reactive protein
CS	Clinically significant
CSE	Composite safety endpoint
CSP	Clinical study protocol
CSR	Clinical study report
DSMB	Data and safety monitoring board
ECG	Electrocardiogram
ECRF	Electronic case report form
EOS	End of study
EOT	End of treatment
EQ-5D-5L	EuroQol Health Utility Index, 5 dimensions 5 levels
ESA	Erythropoiesis-stimulating agent
FAS	Full Analysis Set
FDA	Food and Drug Administration
Hb	Hemoglobin
HDL	High-density lipoprotein
HR	Hazard ratio
HRQoL	Health Related Quality of Life

Abbreviation or special term	Explanation
ITT	Intention To Treat
IV	Intravenous
KM	Kaplan-Meier
LDL	Low-density lipoprotein
LTFU	Lost to follow up
MACE	Major Adverse Cardiovascular Event
MAP	Mean Arterial Pressure
MAR	Missing At Random
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
MCMC	Markov Chain Monte Carlo
MCV	Mean Corpuscular Volume
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model of Repeated Measures
MNAR	Missing Not At Random
N (or n)	Sample size
NCS	Not clinically significant
PCS	Potentially clinically significant
PEY	Patient-exposure-year
PMM	Pattern Mixture Model
PPS	Per-protocol set
PSAP	Pooled Statistical Analysis Plan
RBC	Red blood cell
SAE	Serious adverse event
SAP	Statistical Analysis Plan
TEAE	Treatment Emergent Adverse Event
TESAE	Treatment Emergent Serious Adverse Event
TIBC	Total iron binding capacity
TIW	Three times weekly
TSAT	Transferrin saturation
US	United States
VAS	Visual analogue scale

Abbreviation or special term	Explanation
WBC	White blood cell

SAP AMENDMENT HISTORY

Date	Brief description of change
28 September 2018	The primary safety objective of this study is to contribute to the safety data in the pool safety analysis. Thus, analyses of CV safety will be conducted in accordance with the pooled safety analysis plan (PSAP).. These analyses will be made by study, as described in the PSAP, then pooled adopting meta-analysis techniques. The primary efficacy endpoint has been specified as the primary efficacy endpoint for the FDA. The first secondary efficacy endpoint for the FDA has been specified as the primary endpoint for the EU health authorities. Another change in this SAP is the ordering of the secondary efficacy endpoints to align with the order of the other phase 3 studies in the DD program. The censoring criteria at a primary analysis censoring date has been omitted to align with the censoring rules in the PSAP. A full list of the major changes are listed in Section 6.
25 January 2018	The most important change in edition 3.0 is how the analysis of the primary safety composite endpoint MACE is conducted. There has been a strategic change in how to address CV safety in the project. No hypothesis testing, and no formal NI-margin for the safety endpoints will be adopted for this SAP. In alignment with the pooled SAP (PSAP) the focus will be on estimation and in providing a complete and transparent pattern of results for CV safety through the adoption of the "Totality of Evidence" approach. For full details regarding the "Totality of Evidence" approach, please refer to the PSAP. Other main changes are the addition of a primary efficacy endpoint, and addition of further secondary endpoints, as well as ordering of the secondary endpoints, done to harmonize with the other pivotal studies in this indication. A comprehensive list of changes from the previous edition of the SAP is available in Section 6.

Date	Brief description of change
08 July 2016	Changes from SAP Edition 1.0: Following comments from the US Food and Drug Administration (FDA) on the CSP on March 13, 2015, it was decided that the analysis method for the first secondary endpoint regarding Hb change to be a multiple imputation ANCOVA. Furthermore, additional changes have been made to align with the CSP amendment, edition 6.0. For all major changes from SAP Edition 1.0, see Section 6.

1. STUDY DETAILS

1.1 Study objectives

This study is part of a study program for chronic kidney disease dialysis dependent (CKD-DD) subjects. The other phase 3 studies in the study program are the FibroGen FG-4592-063 and FG-4592-064 studies and the Astellas 1517-CL-613 study. There is a separate pooled statistical analysis plan (PSAP) for the statistical considerations concerning the overall program. The primary objective of this study is to evaluate the efficacy of roxadustat and to contribute CV safety data to the pooled safety analysis evaluating the safety of roxadustat for the treatment of anemia in CKD subjects on dialysis.

The objectives of the current study are to evaluate the safety and efficacy of roxadustat compared to epoetin alfa for the treatment of anemia in subjects receiving dialysis. Subjects on hemodialysis (HD) or peritoneal dialysis (PD) who have been treated with an erythropoietin analogue or have an indication for treatment with an erythropoietin analogue will be evaluated for eligibility and randomized at a 1:1 ratio to treatment with roxadustat (with discontinuation of prior erythropoietin analogue therapy) or to an active-control group treated with epoetin alfa. This study is also known as “ROCKIES”.

1.1.1 Primary efficacy objective

The primary efficacy objective is to evaluate the efficacy of roxadustat as compared to epoetin alfa based on Hb response during the study.

1.1.2 Primary safety objective

The primary safety objective is to contribute adjudicated CV safety data to pooled safety analyses across the phase 3 program per pooled SAP.

1.1.3 Secondary efficacy objectives

The secondary efficacy objectives are to evaluate:

- The efficacy of roxadustat as compared to epoetin alfa based on Hb response and level during the study.
- The efficacy of roxadustat based on Hb response in inflamed subjects.
- The effect of roxadustat on low-density lipoprotein (LDL) cholesterol as compared to epoetin alfa.
- The need for IV iron use in subjects treated with roxadustat as compared to epoetin alfa.
- The need for RBC transfusion as rescue therapy in subjects treated with roxadustat as compared to epoetin alfa.

1.1.4 Secondary safety objectives

The secondary safety objective is to evaluate the safety and tolerability of roxadustat as compared to epoetin alfa..

1.2 Study design

This is a Phase 3, multicenter, randomized, open-label, active-controlled study to evaluate the safety and efficacy of roxadustat compared to epoetin alfa for the treatment of anemia in dialysis subjects. Subjects on hemodialysis (HD) or peritoneal dialysis (PD) who have been treated with an erythropoietin analogue or have an indication for treatment with an erythropoietin analogue will be evaluated for eligibility and randomized at a 1:1 ratio to treatment with roxadustat (with discontinuation of prior erythropoietin analogue therapy) or to an active control group treated with epoetin alfa.

The study periods are as follows:

- **Screening Period:** Up to 6 weeks.
- **Treatment Period:** Subjects will be randomized (1:1) to open-label treatment with either roxadustat or epoetin alfa. Treatment duration is variable for individual subjects (estimated treatment up to 4 years). A common closeout will occur when the target number of MACE events has been accrued.
- **Post-Treatment Follow-Up Period:** 4 weeks from the end of treatment (EOT) visit to the end of study (EOS) visit. Subjects who discontinue study medication prematurely will be followed up for CV events, Hb measurements, vital status and hospitalizations until the end of the study (EOS), according to ITT principles, unless consent to participate is withdrawn.

1.2.1 Dosing

Roxadustat

Subjects currently treated with an erythropoietin analogue who are randomized to the roxadustat group will discontinue prior erythropoietin analogue therapy and initiate treatment with roxadustat at a starting dose according to Section 7 in the Clinical Study Protocol (CSP). Moreover, the dose is subsequently adjusted to achieve and maintain Hb levels between 10 and 12 g/dL. Subjects not currently treated with erythropoietin analogue will initiate roxadustat with dose selection based on body weight. Roxadustat will be administered orally three times a week (TIW) throughout the Treatment Period unless dose frequency reduction is required based on Hb levels.

Epoetin alfa

Initial dose selection of epoetin alfa for subjects treated with an erythropoietin analogue will be determined using a conversion table based on the subject's average prescribed

erythropoietin analogue dose during the preceding 4-8 weeks prior to enrollment in the study. Initial dosing of epoetin alfa for subjects not currently receiving an erythropoietin at study entry will be 50 IU/kg TIW with subsequent dose adjustments to achieve an Hb level of 11 g/dL and maintain an Hb level between 10 and 12 g/dL, consistent with approved prescribing information or the Summary of Product Characteristics for epoetin alfa.

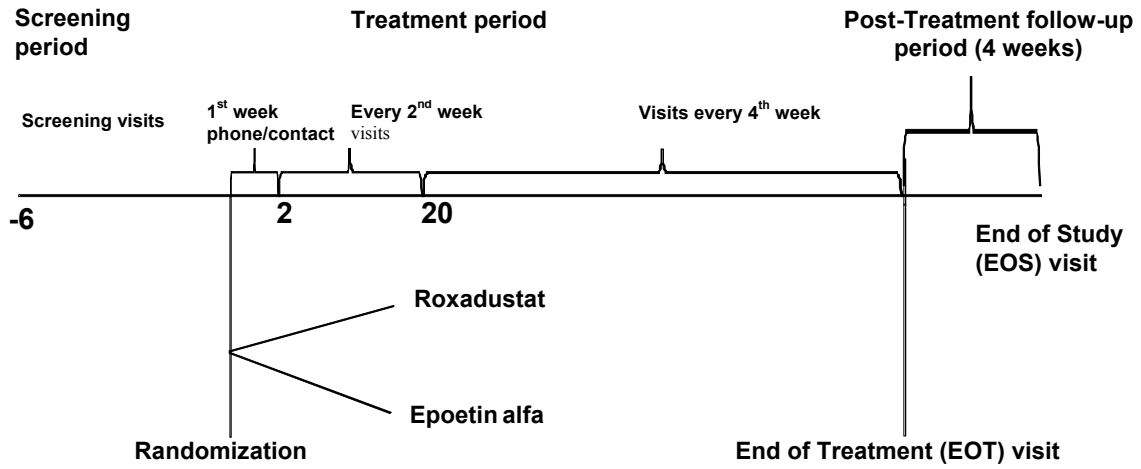
Dosing for both treatment groups

Rescue therapy guidelines are provided to optimize standardization of the use of rescue therapy by investigators and to ensure safety of the individual study subjects. In the event of excessive erythropoiesis or excessive Hb levels equal or greater than 13 g/dL, the dose will be adjusted or put on hold at any time. Excessive erythropoiesis is defined as an Hb increase by >2.0 g/dL within a 4 week period.

1.2.2 Scheduled visits during treatment

During the treatment period, subjects will be contacted by telephone at week 1, and will attend study visits every two weeks from weeks 2 to 20. After week 20, study visits will occur every four weeks until the end of treatment period. A study end date will be defined based on when the planned number of events are estimated to be accrued; the EOT visit will occur as soon as possible after that date. An EOS visit will be performed 4 weeks after the EOT.

Figure 1 Study Flowchart



1.2.3 Stratification variables

The randomization in this study will be stratified by country. The stratification variables for the other studies in the program will be used in the analyses for this study as covariates, with addition of incident vs stable dialysis. The stratification variables are:

1. Baseline Hb (≤ 10.5 g/dL vs > 10.5 g/dL)

2. cardiovascular/cerebrovascular/thromboembolic medical history (Yes vs. No)

3. geographical region (US vs. Ex-US)

4. incident vs. stable dialysis (dialysis duration ≤4 months vs >4 months from the randomization date)

Baseline Hb will be included in the analyses as a continuous covariate, hence, not as a dichotomous factor, unless specified otherwise. Throughout this document, the variable cardiovascular/cerebrovascular/thromboembolic medical history will be shortened as CV history, and the variable incident vs stable dialysis will be shortened as dialysis duration

CV history at baseline will be defined for subjects with history of any of the following diseases:

- Myocardial infarction
- Percutaneous coronary intervention
- Coronary artery bypass
- Cardiac failure congestive
- Ischaemic stroke
- Haemorrhagic stroke
- Cerebrovascular accident

1.3 Number of subjects

Primary efficacy variable: With at least 600 subjects, the study will provide at least 99% power to demonstrate non-inferiority of roxadustat versus epoetin alfa for the primary efficacy endpoint (i.e., Hb change from baseline to the average level during the evaluation period defined as Week 28 until Week 52). This assumes a difference (roxadustat minus epoetin alfa) of -0.30 g/dL, a non-inferiority margin for this difference of -0.75 g/dL and a standard deviation of 1.25 g/dL.

To contribute adjudicated CV events for the pooled CV analyses across the phase 3 program: approximately 2000 subjects will be randomized in a 1:1 ratio to either roxadustat or active control, i.e. epoetin alfa. The sample size for this study is driven by the overall requirement of adjudicated CV events for the phase 3 program in dialysis-treated CKD subjects (which consists of four studies in total targeting 611 subjects with MACE events). The three other studies in the study program are FG-4592-064, 1517-CL-0613, FG-4592-063.

,.

2. ANALYSIS SETS

2.1 Definition of analysis sets

2.1.1 Intention To Treat Analysis Set (ITT)

All subjects who have been randomized to study treatment will be included irrespective of their protocol adherence and continued participation in the study. Subjects will be analyzed according to their randomized study medication irrespective of intake of study medication.

2.1.2 Per Protocol Set (PPS)

All randomized subjects without important protocol deviations and who have received at least 8 weeks of study treatment and who have valid corresponding Hb measurements will be included in the PPS. A valid corresponding Hb is defined as an Hb value from the central laboratory that is measured at least 2 weeks after the first dose and was either before the last study drug intake or at maximum three days after the last drug intake. Subjects will be analysed according to their randomized study medication irrespective of intake of study medication. Subjects with an important protocol deviation will be included in the PPS up to the time point when the violation was met. For criteria for PPS exclusion, see Table 1 in Section 2.2. Further details of important protocol deviations are available in a Protocol Deviation Plan. Subjects will be censored at the earliest of date of violating an important protocol deviation, the EOS visit, or at last intake of study drug.

2.1.3 Safety Analysis Set

All subjects who received at least one dose of randomized study drug will be included in the Safety Analysis Set. Throughout the safety results sections, erroneously treated subjects will be accounted for in the actual treatment group. If a subject has received both treatments, only the initial period will be utilized. Subjects will be censored at either, 7, 3 or 0 days after last intake of study drug. Consequently, there will be three versions of the Safety Analysis Set, and they will be referred to as On-treatment+7 (OT+7), OT+3 and OT+0 respectively.

2.1.4 Full Analysis Set (FAS)

The FAS consists of all patients in the ITT analysis set who received at least one dose of study drug and have baseline and at least one post-dose Hb assessment. If actual study medication received differs from the randomized treatment arm, the randomized treatment arm will be used for analysis for the FAS. This analysis set is primarily used for EX-US submissions.

2.1.5 Subjects who will not be included in any analysis sets

Subjects or sites identified prior to database lock (breaking of Sponsor blindness, open-label study) with major Good Clinical Practice violations and where the integrity of the data is strongly questioned through thorough independent investigations will be excluded from all analyses and all analysis sets. This includes but are not limited to subjects who have been identified to be part of a potential fraud investigation, subjects who have not signed an informed consent, subjects randomized in error (e.g. a subject considered to be a screen fail but by mistake randomized in the IWRS due to a technical error). Further, subjects being

randomized more than once will only contribute to the analysis one time. These patients will be analyzed according to their first assigned randomization number and treatment code. All AE's reported for the subjects will be assigned to the subject´s first randomization number. All subjects excluded from all analysis sets will be properly documented.

2.2 Violations and deviations

The important protocol deviations are defined in Table 1. Protocol deviations will be presented in a data listing.

Table 1 Criteria for Assessing Important Protocol Deviations

Number	Important Protocol Deviations	Level of Deviation[1]
1	Study drug compliance <75% where drug compliance is measured by comparing dispensed and returned drug (see Section 3.4).	Subject
2	Administration of wrong type of study drug (i.e. the one not randomized to) cumulatively more than 1 week	Visit
3	Administration of prohibited concomitant medication or non-drug therapy as defined in the protocol.	Visit
4	Administration of rescue therapy deviating from the protocol	Visit
6	Violation of inclusion or exclusion criteria. The key inclusion criteria are numbers 3-8. The key exclusion criteria are numbers 1-6, 9-16, 18 and 23. The full inclusion and exclusion criteria is available in the CSP.	Subject

3. PRIMARY AND SECONDARY VARIABLES

3.1 Efficacy variables

3.1.1 Primary efficacy endpoint

US FDA: The primary efficacy endpoint for the US is the mean change from baseline in Hb averaged over week 28 to week 52. A multiple imputation approach with analysis of

[1] Subject-level deviations refer to important protocol deviations that will cause subjects to be excluded from the Per Protocol set, and therefore all their collected data from analyses based on this population.

Visit-level deviations refer to important protocol deviations that will cause only some data for subjects to be excluded from analyses based on the Per Protocol set, while the subjects remain in the Per Protocol set given that they did not meet any subject-level deviations. Data to be excluded from the Per Protocol analyses could be either data from a certain date, at which the deviation was met for the first time, onwards to the end of the study, or data during a period defined by the start and end dates of the deviation.

covariance (ANCOVA) will be applied as a method to handle missing data. Details of the multiple imputation ANCOVA are provided in Sections 4.2.4 and 4.2.5.

Hb results obtained from the central laboratory will be used for all Hb efficacy analyses. Baseline Hb is defined as the mean of the three last central laboratory Hb values from the screening and randomization visits.

Hb values under the influence of a rescue therapy will not be censored.

EU health authorities: The primary efficacy endpoint for the EU health authorities is the mean change from baseline in Hb averaged over week 28 to week 36, without having received rescue therapy within 6 weeks prior to and during this 8-week evaluation period.

3.1.2 Hb related secondary efficacy variables

The Hb related secondary efficacy variables are:

- The EU primary endpoint as specified in Section 3.1.1 is the first secondary efficacy endpoint in the analysis for FDA
- Mean change in Hb from baseline to the subjects mean level between week 28 to week 52 in subjects with baseline high-sensitivity C-reactive protein (hsCRP) greater than the Upper Limit Normal (ULN)
- Proportion of total time of Hb ≥10 g/dL from week 28 to week 52. The proportion of total time will be computed as follows: For each subject, the recorded Hb values will first be linearly interpolated between measurements. The time this interpolated curve is ≥10 g/dL will be computed and subsequently divided by the time between the measurements at week 28 and week 52. Subjects without any Hb measurements from week 28 will not be considered for this variable.
- Proportion of total time of Hb within the interval of 10-12 g/dL from week 28 to week 52. The proportion of total time will be computed as follows: For each subject, the recorded Hb values will first be linearly interpolated between measurements. The time this interpolated curve is within 10-12 g/dL will be computed and subsequently divided by the time between the measurement at week 28 and week 52. Patients without any Hb measurements from week 28 will not be considered for this variable.

3.1.3 Lipid related secondary efficacy variables

To evaluate the roxadustat effect on lipids, the following variable will be evaluated

- Mean change from baseline in LDL cholesterol to week 24.

3.1.4 Rescue therapy related secondary efficacy variables

The need for rescue therapy will be evaluated as

- Time-to-first (and proportion of subjects who received) administration of red blood cell (RBC) transfusion as rescue therapy (rescue therapy guidelines are specified in the CSP, Section 7.7.4).

For analyses based on the Safety Analysis set, time to the event will be calculated as the number of days plus one between the day of first dose of study drug and date of the first occurrence of the event, or if no event has occurred before censoring, the date of censoring (see Section 4.1.1). For analyses based on the ITT analysis set, FAS and PPS, time will be calculated as the number of days plus one between the day of randomization and date of the first occurrence of the event, or if no event has occurred before censoring, the date of censoring.

3.1.5 IV iron related secondary efficacy variables

The use of IV iron will be investigated with the variable:

- Average monthly IV iron use per subject during, week 36 to EOS (monthly defined as a period of 4 weeks).

3.1.6 Exploratory variables

The exploratory variables are:

3.1.6.1 Hb related exploratory variables

Mean change from baseline in Hb, utilizing all Hb values from week 28 until the EOT visit. An imputation for mean change in baseline will only be applied for subjects with no Hb values from week 28 to the EOT visit.

- Time to achieving target Hb for anemic (Hb<10 g/dL at baseline) subjects not receiving ESA <= 4 weeks prior to randomization and were ESA-naïve or near ESA-naïve (no ESA use <=4 weeks prior to randomization). Time will be computed analogously as time to first rescue therapy. Target Hb is achieved when Hb level is within 10-12 g/dL at two consecutive measurements. This will be repeated for subjects with no ESA use <= 4 weeks prior to randomization.

- Proportion of ESA-naïve anemic patients achieved Hb response by Week 24 in the subset of patients who were anemic (Hb< 10 g/dL at baseline) and were ESA-naïve or near ESA-naïve (no ESA use <=4 weeks prior to randomization).

  Hb response (Yes/No), where Yes is defined as:

  - Hb ≥ 11.0 g/dL and Hb increase from baseline by ≥ 1.0 g/dL, for subjects with baseline Hb > 8.0 g/dL; or

- Hb increase from baseline by ≥ 2.0 g/dL, for subjects with baseline Hb ≤ 8.0 g/dL

  at two consecutive visits [dates] (with available data) separated at least 5 days during the first 24 weeks of treatment without having received rescue therapy (RBC transfusion, ESA, or IV iron) prior to Hb response. The first date of the two consecutive visits will be used as the date of response. The second date of the two consecutive visits will be used when evaluating the presence or absence of rescue therapy.

- Percent of total patient exposure time with achieved Hb level <9, 9-<10, 10-<11, 11 - <12, 12-<13, >=13 g/dL during treatment.

3.1.6.2 RBC transfusion related exploratory variables

- Number of rescue therapy treatments given, RBC transfusion per patient exposure year (PEY).

- Proportion of subjects with RBC transfusions during week 28 to week 52.

3.1.6.3 Quality of life related exploratory efficacy variables

European Quality of Life Questionnaire in Five Dimensions, Five Levels (EQ-5D-5L)

The EQ- 5D-5L consists of the EQ-5D-5L descriptive system and the EQ visual analogue scale (VAS). The EQ-5D-5L descriptive system comprises 5 dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, extreme problems. The visual analogue scale records the respondent's self-rated health status on a graduated (0–100) scale, where the endpoints are labelled ‘Best imaginable health state’ and ‘Worst imaginable health state’ with higher scores for higher HRQoL. The EQ-5D-5L variables are:

- Change from baseline in EQ-5D-5L index value

- Change from baseline in EQ VAS

and will be computed according to the EQ-5D-5L documentation.

Results from the EQ-5D-5L questionnaire will also be summarized descriptively.

3.1.6.4 Hospitalization related exploratory variables

- Number of hospitalization(s) and number of days of hospitalizations per PEY.

- Number of days spent in Intensive Care Unit (ICU) per PEY.

- Proportion of subjects who are re-admitted to hospital within 30 days per PEY.

- Proportion of subjects who are re-admitted to hospital within 30 days due to heart failure per PEY following a preceding hospitalization due to heart failure.

- Proportion and number of on treatment days of hospitalization-free survival.

- Proportion and number of on treatment days of hospitalization-free, emergency room- free, and skilled nursing facility-free survival. Number of days spent in a Skilled Nursing Facility that follow hospitalizations per PEY, and the total number of days covering both hospitalizations and subsequent days in Skilled Nursing Facility.

3.1.6.5 Other exploratory variables

- Variables concerning serum iron profiles: Iron, TIBC, Ferritin and TSAT. Mean value at each time-point tested, and change from baseline.

- Variables concerning lipids: total cholesterol, high-density lipoprotein (HDL) and triglyceride: mean values and change from baseline at each timepoint tested; percent of subjects who achieved target LDL level <100 mg/dL, at each timepoint ; subgroup analyses on patients on statins and those not on statins

- Variables concerning heart rate and blood pressure: heart rate, systolic blood pressure (SBP), diastolic blood pressure (DBP) and mean arterial pressure (MAP).

- Variables concerning concomitant medication: Usage of statins and types of statins, initiation of ESA therapy post study drug discontinuation

- Change from baseline in hepcidin to week 24.

3.2 Safety assessments

Safety will be further assessed by evaluating the following:

- Occurrence of treatment emergent adverse events (TEAEs) and treatment emergent serious adverse events (TESAEs)

- Changes from baseline in vital signs and physical examinations.

- Mean change from baseline in clinical laboratory values.

- Occurrence of clinically significant changes from baseline in vital signs and ECG values.

3.3 Adjudicated CV Events Analyses for Safety Assessments

The CV events being analysed in the PSAP will be adjudicated by the Independent Event Review Committee (IERC) according to the IERC charter. The same adjudication committee will be used for all the phase 3 studies (FG-4592-064, 1517-CL-0613, FG-4592-063, and D5740C00002) in the CKD-NDD program. Analyses of these adjudicated events are described in a separate pooled analysis plan.

3.4 Treatment compliance

Subjects will be asked to return all unused study medication and empty packages to the clinic at each visit. The amount of dispensed and returned study medication will be recorded in the eCRF. The percentage treatment compliance will be calculated as:

((Overall amount of dose actually taken)/(Overall amount of dose to be taken))*100%
Subjects taking ≥75% and ≤125% of planned study medication are considered to be compliant.

Compliance will be summarized as follows:

- Descriptive statistics will be summarized by the two treatment groups
- Percent compliance will be categorized according to the following three categories:
  - <50% (significant drug non-compliance)
  - ≥50%, <75%, >125% (moderate drug non-compliance)
  - ≥75%, ≤125% (drug compliance)

4. ANALYSIS METHODS

Statistical analyses will be performed by IQVIA™ using SAS® Version 9.4 or higher and, where appropriate, additional validated software.

4.1 General principles

All study data will be listed by treatment group, centre, and subject number. Throughout subject data listings, figures and tables, treatment groups will be labelled as “Roxadustat” and “Epoetin alfa”.

Study Day will be listed in all data listings whenever an assessment date is presented. Study Day is a relative number, relative to the date of first dose of study drug. Week is also a relative number, relative to the number of weeks from the first dose of study drug. Study Days 1-7 is defined as Week 0, Days 8-14 as Week 1, etc. Clinical events and other variables reported after the EOT visit for a subject will not be included in the primary efficacy and safety

analysis. If collected, these events will be included in tables. Events that are recorded as beginning prior to the date and time of randomization will not be included in a listing.

In addition to specific analyses and presentations that are detailed in the following sections, results will be summarised using descriptive statistics, including the number of subjects (n), mean, standard deviation (SD), median and range (i.e. minimum and maximum) as appropriate. For categorical variables counts and percentage n (%) per treatment group will be presented. Summaries of continuous variables will be based on non-missing observations. For time to event data, the number and percentage of subjects recording the event will be summarised. Ninety-five percent confidence intervals will also be included where appropriate, as a measure of precision. Demographic characteristics, qualifying risk factors and other specific medical and surgical history will be summarised for the ITT analysis set using descriptive statistics. Mean Hb values over time will be graphically displayed and grouped by treatment.

Dates will be presented in the format YYYY-MM-DD.

When the last dose date is missing, it will be imputed as the earliest date of last drug dispense date + number of days of drug dispensed, date of death, date of EOT visit or date of EOS visit.

4.1.1 Censoring

For analyses based on OT+7, subjects will be censored at 7 days after last intake of study drug, or the EOS visit, whichever is earliest.

For analyses based on OT+3, subjects will be censored at 3 days after last intake of study drug, or the EOS visit, whichever is earliest.

For analyses based on OT+0, subjects will be censored at last intake of study drug, or the EOS visit, whichever is earliest.

For analyses based on the ITT analysis set and FAS, subjects will be censored at their individual EOS visit, regardless of if they have discontinued study drug or not. Complete endpoint information will be pursued with every effort for all subjects, unless they exercise their right to withdraw consent. Subjects who withdraw consent will be censored at date of withdrawal of consent, and subjects who are lost to follow up (LTFU) will be censored at last available contact.

Subjects who withdraw consent and for whom only vital status (known to be alive at study closure, or date of death) may be obtained from public records, the occurrence of all components of the primary endpoint cannot be assessed, and will thus be censored at date of consent withdrawn for all analyses. However, the determination of all-cause death will utilize all publicly known mortality data, even that extending beyond date of consent withdrawal. The vital status information will be included in the analysis of all-cause death as a single endpoint, and in sensitivity analysis and tabulations. Similarly, complete information on the endpoint may not be obtained for subjects who are LTFU. Any such subject will be censored in the analysis at the last contact where all elements of the endpoint were assessed. A subject

will not be recorded as LTFU until the end of the study, after every allowable effort to get in contact has been made. Hence, it is anticipated that the number of subjects LTFU will be limited.

4.1.2 Premature permanent discontinuation of study medication

Premature discontinuation from study medication is not the same as withdrawal from the study. As described in Section 3.9 in the CSP there are several options for continuing the study.

It is expected that complete information on the safety composite endpoint events, and as much as possible of the remaining eCRF data, will be obtained for all subjects who prematurely discontinue study medication, unless they refuse any form of follow-up and withdraw consent or are LTFU.

4.1.3 Choice of the non-inferiority margins

Non-inferiority for the Hb efficacy endpoint will be established if the lower limit of the two-sided 95% CI for the difference between the means of the primary endpoint (roxadustat minus control epoetin alfa) is ≥-0.75 g/dL, see Section 0. In other words, a non-inferiority margin of -0.75 g/dL will be used in the Hb efficacy assessments.

Support for the use of a non-inferiority margin of -1.0 g/dL in the peginesatide submission for the dialysis studies was initially provided by estimating the magnitude of the effect of erythropoietin analogue therapy. This estimate was based on summary information from the darbepoetin alfa development program (Aranesp 2001, Omontys 2012), with publicly available summary data. Based on these data, the estimated treatment effect of erythropoietin analogue therapy in the dialysis population was approximately 2.0 g/dL. Based on these data, the estimated treatment effect of erythropoietin analogue therapy in the dialysis population with a 10-12 g/dL target range (at least 2.0 g/dL and was considered to be appreciably larger than 1.0 g/dL, thus supporting the choice of a non-inferiority margin of -1.0 g/dL.

Therefore, the choice of -0.75 g/dL as non-inferiority is more conservative than prior programs, and consequently, an appropriate margin to use.

The comparison between the study drugs for the time-to-first instance of receiving RBC transfusion, or erythropoietin analogue (for roxadustat subjects only) as rescue therapy is also based on a NI evaluation. For this purpose, in the lack of support for the choice from the literature, a 1.8 margin has been selected. NI will thus be claimed if the upper bound of the 2-sided 95% CI for the hazard ratio (roxadustat/epoetin alfa) is less than or equal 1.8.

The analysis of the proportion of total time of Hb within the interval of 11±1 g/dL from week 28 until week 52 between the treatment groups also relies on a non-inferiority evaluation. For this purpose a NI margin of -0.15 has been adopted, i.e. the 2-sided 95% CI around the difference between roxadustat and epoetin alfa has to exceed -0.15. No support from the literature has been found for this selection.

4.2 Analysis methods

4.2.1 Demography

The following will be reported on subjects who are randomised: sex, age, race and ethnic group, baseline Hb value, geographical region, incident vs stable dialysis (dialysis duration ≤4 months vs >4 months from the randomization date), dialysis type, cardiovascular/cerebrovascular/thromboembolic medical history, congestive heart failure history, coronary artery disease history and cerebrovascular history, other relevant medical and surgical history, concomitant medication, weight, height, BMI, tobacco use, CKD diagnosis, diabetes history and baseline blood pressure. Continuous and categorical demographic variables will be presented as described in Section 4.1. The following continuous variables will also be presented as range-based categories:

- baseline Hb value (≤10.5g/dL vs >10.5g/dL),
- age (≥18 - < 50, ≥ 50 - < 65, ≥ 65 - < 75, ≥ 75 years),
- BMI (<30, ≥30 kg/m2), and
- weight (<70, >=70 - <100, >=100 kg)).

4.2.2 Confirmatory analysis for the efficacy endpoints

To address the issue of multiple testing while maintaining the overall type-I error, adopting a 5% two-sided significance level, a closed testing sequence will be used for the efficacy endpoints. First, the primary efficacy endpoint analysis according to Section 0 will be performed. If successful, the testing will continue with the secondary efficacy endpoints in the order as specified in Section 4.3.4. Confirmatory statistical hypothesis testing will continue until the first statistically non-significant treatment difference is observed. However, treatment comparisons following and including the first non-significant comparison will be examined in an exploratory manner.

All analyses other than part of this confirmatory analysis will be interpreted descriptively. Consequently, no adjustments for multiplicity will be necessary for such analyses. Ninety-five percent confidence intervals will be calculated, where appropriate, as measures of study precision. P-values may be calculated but are to be regarded as descriptive.

4.2.3 Time to event analysis

For time to event variables, treatments will be compared using a Cox proportional hazards model. Unless specified otherwise baseline Hb as a continuous variable will be used as a covariate, and treatment group, CV history, dialysis duration and geographic region as fixed effects for all analyses. The Efron method will be used for ties. The p-values (calculated using the Wald test), hazard ratio (HR) and 95% confidence intervals for the HR will be reported. Summary tables of these analyses will also include the number of subjects with an event and Kaplan-Meier estimates of the event rates per treatment group estimated at a time point determined on the basis of the available follow-up. Kaplan-Meier estimates of the cumulative proportion of subjects with events will be estimated and plotted, with the number of subjects at risk indicated below the plot at specific time points.

4.2.4 Difference in proportions analysis

For analysis of difference in proportions the approach by Miettinen & Nurminen 1985 will be used and a 2-sided 95% confidence interval for the difference of two proportions (roxadustat vs epoetin alfa) will be computed, adjusting for stratification factors. The model will include the terms of baseline Hb, treatment group, cardiovascular history, geographic region and dialysis duration. The stratified statistics will be based on the standard normal statistic proposed by Gart and Nam 1990.

4.2.5 Analysis of Covariance (ANCOVA)

When using ANCOVA in analysis of change from baseline for a continuous variable, the mean value of all change from baseline values available within the pre-specified timeframe will be used as the dependent variable. Unless specified otherwise, baseline Hb will be used as a covariate and treatment group, cardiovascular history, geographic region and dialysis duration as fixed effects for all analyses. Any further details will be given case-by-case for each endpoint (see Section 4.3). The least squares mean estimates of change from baseline for each treatment group and their difference, and associated 95% CI will be provided.

4.2.6 Multiple imputation ANCOVA

For the primary efficacy analysis, a multiple imputation ANCOVA method (O'Kelly & Ratitch, 2014) will be used. It will be conducted with the following steps:

1. 200 datasets will be generated, using seed number 326154, where non-monotone missing Hb data will be imputed, meaning intermediate visits that subjects skip, but return for evaluations at subsequent visits. The data points are imputed assuming MAR, using the MCMC imputation model baseline Hb, CV history, geographic region and dialysis duration, and the available non-missing Hb for each scheduled week are used as covariates, by treatment group. The MCMC statement in the SAS PROC MI procedure with monotone option will be used. As a result, each dataset will only have a monotone missing data pattern.

2. For each dataset from step 1, the missing monotone data points will be imputed, which is when a subject misses one visit, and all subsequent visits. As a result, 200 imputed complete datasets will be generated.

    - Missing data at Week 2 will be imputed using the regression imputation model with baseline Hb and Hb from Week 2, CV history, geographic region and dialysis duration as terms in the model, by treatment group This will be performed with the SAS PROC MI procedure with the REGRESSION option in the MONOTONE statement.

    - Repeat for all other scheduled weeks sequentially. Subjects whose missing data were imputed for previous weeks will contribute to the imputation for the current week.

    - The regression imputation model includes an intercept and the slopes of the Hb from previous weeks.

3. Fit an ANCOVA model on each of the 200 datasets where the average of the imputed and observed Hb values between weeks 28 to 52 for each subject is taken as the dependent variable and baseline Hb, treatment group, CV history, geographic region and dialysis duration as covariates.

4. Combine the results of all 200 ANCOVA models using Rubin's rules (Rubin, 1987) with the SAS PROC MI ANALYZE procedure.

The least squares mean estimates of change from baseline for each treatment group and their difference, together with their associated 95% CI and p-value will be reported.

The exploratory efficacy endpoint of Hb change from baseline to the average between week 28 to the EOT visit will be conducted in the following steps:

1. 200 datasets will be generated, using seed number 326154, assuming a return-to-baseline values, with imputed values being sampled from a posterior Bayesian distribution of baseline Hb for all treatment groups combined, using a regression imputation model with stratification variables as predictor variables.

2. The imputations will be obtained using SAS PROC MI as follows. The input dataset will contain a set of temporary records to be used for imputation model estimation, where a new record is created for each subject, assigning their baseline Hb value to a variable representing mean Hb values from week 28 to EOT. This variable will be used as a dependent variable in a regression imputation model, thus estimating a distribution of baseline values. The MONOTONE REG statement will be used to estimate the imputation model, and the MNAR statement with MODEL option will be used to specify the subset of temporary records as described above from which the imputation model will be estimated. Once the imputation is complete, the temporary records will be removed prior to analyzing the imputed data, and for subjects whose mean Hb values from week 28 to EOT were imputed, a change from baseline will be calculated as the imputed value minus baseline Hb value.

3. Fit an ANCOVA model on each of the resulting 200 datasets where the mean change in Hb from baseline from weeks 28 to EOT for each patient is taken as the dependent variable and baseline Hb, treatment group, CV history, geographic region and dialysis duration as covariates.
4. Combine the results of all 200 ANCOVA models using Rubin's rules (Rubin, 1987) with the SAS PROC MIANALYZE.

5. Non-inferiority between roxadustat compared to epoetin alfa will be declared, and this test successful, if the lower bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa exceeds -0.75 g/dL.

4.2.7 Mixed Model of Repeated Measures (MMRM)

As one of the sensitivity analyses, the mixed model of repeated measures (MMRM) will be used. Longitudinal models with correlated errors, otherwise widely known as MMRMs, have been increasingly used for the analysis of clinical trials with missing data. A longitudinal model is often used even though the primary objective is to estimate a treatment effect and test a null hypothesis of no treatment effect at a single specific time-point (typically at the end of double-blind period). The advantage of using an MMRM analysis in this context (compared to ANCOVA at the primary time-point) is that longitudinal models include all randomized subjects regardless of whether they completed the study (provided data for the primary time-point) or not. Model estimation and inference is done without performing any imputation of the missing data for subjects who discontinued early, yet partial data available for these subjects is fully utilized and contributes to the estimation of effects and to the variance-covariance structure of the longitudinal model.

The MMRM can contain terms for baseline measurement, treatment arm, visit, treatment by visit interaction, and the stratification variables. Details will be given case-by-case for each endpoint. The least squares mean estimates of change from baseline for each treatment group and their difference, and associated 95% CI will be provided. Due to the large amount of visits to include in the model, the unstructured covariance pattern model will be selected first. If the algorithm for unstructured covariance pattern does not converge, then the heterogeneous Toeplitz structure will be used instead. If this second model also does not converge, then the (homogeneous) Toeplitz structure will be selected, thereafter the compound symmetry and finally the first order autoregressive covariance structure will be used to achieve convergence.

4.2.8 Pattern Mixture Models

To address the possibility of the Hb data being missing not at random (MNAR), Pattern Mixture Models (PMM) will be implemented as sensitivity analyses. Pattern Mixture Models (PMM) provide a general and flexible framework for sensitivity analyses that allows formulating assumptions regarding missing data in a transparent and clinically interpretable manner. A variety of PMMs with different types of MNAR assumptions will be implemented.

4.2.8.1 PMM –Last Mean Carried Forward

A pattern-mixture model using a last mean carried forward multiple imputation method (Carpenter et al, 2013) will be used as another sensitivity analysis to explore the robustness of the ANCOVA results for the primary efficacy variables. Using this method, missing data after ending Week will be imputed based on the last non-missing mean from its own treatment group.

The steps to implement this sensitivity method is the same as for the multiple imputation ANCOVA method described in Section 4.2.6 with the exception of step 2, where the monotone datasets are imputed. The procedure for that is as follows. Parameters below refer to the parameters of the multivariate normal distribution for baseline and post baseline Hb measurement.

1. Create posterior distribution of parameters: Separately for each treatment group, take all subjects observed data and assuming MAR to fit a multivariate normal distribution with unstructured mean (i.e. a separate mean for each of the baseline plus post-baseline scheduled weeks and unstructured variance covariance matrix using a Bayesian approach with an improper prior for the mean and an uninformative Jereys' prior for the variance-covariance matrix (Schafer, 1997, p. 155).

2. Draw parameters: Separately for each treatment group, draw variance-covariance matrix from the posterior distribution for the parameters using seed 453628. The mean Vector would be set to the marginal mean for their randomized treatment arm at their last non-missing measurement.

3. Build joint distribution of missing data and observed data: For each subject with missing data, using the draws for the parameter to build the joint distribution of their observed and missing data.

4. Construct conditional distribution of missing data give observed data: For each subject with missing data, use their joint distribution in previous step to construct their conditional distribution of missing given observed outcome data. Sample their missing data from this conditional distribution, to create a "completed" data set, using seed 732545.

Repeat the above steps for 200 times and resulting in 200 fully imputed data sets. Then fit an ANCOVA model for each imputation data set, and combine the resulting parameter estimates and standard errors using Rubin's rules (Rubin, 1987) for final inference.

4.2.8.2 PMM –Baseline Carried Forward (roxadustat only and both groups)

The analysis is similar to PMM – Last Mean Carried Forward, with a different assumption in imputing the missing data. The imputation data will be generated similarly as last mean carried forward method described above but instead of using post-baseline observed data, only baseline data will be used. The similar analyses will be conducted in two scenarios.

- The baseline carried forward imputation will be performed for the roxadustat treatment group only, while for active control group, the imputation data will be generated using the last mean carried forward described above.

- The baseline carried forward imputation will be performed for both treatment groups.

Similarly, the Rubin's method will be then used to combine the estimates and the differences between the least square mean differences between the two treatment groups from each of the ANCOVA analysis.

4.3 Statistical Analyses

4.3.1 CV safety endpoints analyses

The CV safety evaluation strategy is to conduct pooled analyses of adjudicated data across the study program to ensure that the overall number of events is high enough to provide adequate power. Thus, all analyses of CV safety will be conducted in accordance with the PSAP.

4.3.2 Primary efficacy endpoint analysis for US

Mean change in Hb from baseline to the subjects mean level from week 28 to week 52 will analyzed with multiple imputation ANCOVA as described in Section 4.2.5 and 4.2.6. The model will contain terms for the baseline Hb measurement, treatment arm, CV history, geographic region and dialysis duration. Non-inferiority of roxadustat compared to epoetin alfa will be declared, and this test as successful, if the lower bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa exceeds -0.75 g/dL. The ITT analysis set will be used.

4.3.3 Primary efficacy endpoint analysis for EU (First Secondary endpoint for FDA)

Mean change in Hb from baseline to the subjects mean level from week 28 to week 36, without having received rescue therapy within 6 weeks prior to and during this 8-week evaluation period and will be analysed using MMRM. The model will contain terms for the baseline Hb measurement, treatment arm, visit, visit by treatment, CV history, geographic region and dialysis duration. Data up to visit of Week 52 will be included in the model.

Non-inferiority of roxadustat compared to epoetin alfa will be declared, and this test as successful, if the lower bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa exceeds -0.75 g/dL. The PPS will be used for non-inferiority. In addition to the comparison based on the PPS population, to address the formal test for non-inferiority, results will also be provided based on the FAS population, to allow also for a potential superiority comparison. The latter test is not part of the formal testing sequence.

4.3.4 Secondary efficacy endpoints analyses

Secondary efficacy endpoints will be tested using a fixed sequence approach to adjust for multiple testing. If the p-value from a test is less than 0.05, the test will be declared as successful and the analysis will continue to the next comparison in the sequence. Formal statistical hypothesis testing will be stopped as soon as a test is accompanied by a p-value ≥0.05. The PPS will be used for the first secondary endpoint for non-inferiority, OT+3 analysis set will be used for the secondary endpoints related to RBC transfusion as rescue therapy, and the ITT analysis data set will be used for all the remaining secondary endpoints.

1. The EU primary endpoint for non-inferiority is the first secondary efficacy endpoint for FDA (see above).

2. Mean change from baseline in LDL cholesterol to week 24 will be analysed using ANCOVA. Baseline Hb and baseline LDL will be used as covariates and treatment groups, CV history, geographic region and dialysis duration as fixed effects. Superiority will be declared if the upper bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa falls below 0.

3. Mean change in Hb from baseline to the subjects mean level from week 28 to week 52 in subjects with baseline hsCRP greater than the Upper Limit Normal (ULN) will be analysed analogously as the primary efficacy endpoint. Superiority of roxadustat compared to epoetin alfa will be declared, and this test as successful, if the lower bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa exceeds 0 g/dL.

4. Proportion of total time of interpolated Hb values ≥ 10 g/dL from week 28 until week 52 will be estimated for each subject and used as dependent variable. The difference between roxadustat and epoetin alfa will be compared using ANCOVA. Baseline Hb will be used as a covariate and the treatment groups, CV history, geographic region and dialysis duration as fixed effects. Non-inferiority between the groups will be declared, and this test as successful, if the lower bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa exceeds -0.15.

5. Proportion of total time of interpolated Hb values within the interval 10-12 g/dL from week 28 until week 52 will be estimated for each subject and used as dependent variable. The difference between roxadustat and epoetin alfa will be compared using ANCOVA. Baseline Hb will be used as a covariate and the treatment groups, CV history, geographic region and dialysis duration as fixed effects. Non-inferiority between the groups will be declared, and this test as successful, if the lower bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa exceeds -0.15.

6. The average monthly IV iron use during Week 36 to EOS will be compared between the two treatment groups using a Wilcoxon Rank Sum test. Superiority will be declared if the p-value is less than 0.05.

7. Time-to-first (and proportion of subjects who received) RBC transfusion as rescue therapy, will be analysed using Cox proportional hazard model. The baseline Hb, geographic region, dialysis duration and CV history will be included as covariates. Non-inferiority will be claimed, and this test successful, if the upper bound of the 2-sided 95% CI for the hazard ratio (roxadustat/epoetin alfa) is less than or equal to 1.8.

4.3.5 Exploratory endpoint analysis

The baseline value for each exploratory variable is defined as the last measurement of the variable prior to randomization, including the measurement from the randomization visit, unless stated otherwise.

The analysis set to be used for all exploratory analyses will be ITT analysis set, unless specified otherwise. The variables will be analysed as follows:

4.3.5.1 Hb related exploratory endpoint analysis

- Mean change from baseline in Hb to the subjects mean level from week 28 to EOT will be analyzed using multiple imputation with ANCOVA as described in Section 4.2.6. The mean change from baseline will only be imputed for subjects with no measurements from week 28 to EOT. Non-inferiority of roxadustat compared to epoetin alfa will be declared, and this test as successful, if the lower bound of the 2-sided 95% confidence interval of the difference between roxadustat and epoetin alfa exceeds -0.75 g/dL.

- Mean change in Hb from baseline to the subjects mean level from week 28 to week 36 in subjects with baseline hsCRP greater than the Upper Limit Normal (ULN) will be analysed analogously as the primary efficacy endpoint. PPS will be used.

- Proportion of total time of interpolated Hb values ≥10 g/dL from week 28 until week 36 will be estimated for each subject and used as dependent variable. The difference between roxadustat and epoetin alfa will be compared using ANCOVA with treatment group, geographic region and CV history as fixed factors and baseline Hb and baseline eGFR as covariates. PPS will be used.

- Time to achieving target Hb for anemic (Hb<10 g/dL at baseline) subjects who were ESA-naïve or near ESA-naïve (no ESA use <=4 weeks prior to randomization). This will be analyzed analogously as Time to first rescue therapy (composite). Target Hb is achieved when Hb level is within 10-12 g/dL at two consecutive measurements.

- Estimation of median time (in weeks) to achieve target Hb for anemic (Hb<10 g/dL who were ESA-naïve or near ESA-naïve (no ESA use <=4 weeks prior to randomization), based on the definition of two consecutive Hb levels within 10-12 g/dl, by treatment arm.

- Proportion of ESA-naïve anemic patients achieved Hb response by Week 24 in the subset of patients who were anemic (Hb< 10 g/dL at baseline) and were ESA-naïve or near ESA-naïve (no ESA use <=4 weeks prior to randomization).

  Hb response (Yes/No), where Yes is defined as:

  - Hb ≥ 11.0 g/dL and Hb increase from baseline by ≥ 1.0 g/dL, for subjects with baseline Hb > 8.0 g/dL; or
  - Hb increase from baseline by ≥ 2.0 g/dL, for subjects with baseline Hb ≤ 8.0 g/dL

  at two consecutive visits [dates] (with available data) separated at least 5 days during the first 24 weeks of treatment without having received rescue therapy

(RBC transfusion, ESA, or IV iron) prior to Hb response. The proportion of responders in the primary efficacy variable will be compared using Miettinen & Nurminen model, adjusting for the region, history of CV, baseline Hb (<=8, >8 g/dL) and dialysis duration, comparing roxadustat to epoetin alfa.

4.3.5.2 Rescue therapy related exploratory endpoint analysis

- Time-to-first instance rescue therapy (composite) of receiving RBC transfusions, or erythropoietin analogue as rescue therapy will be analysed analogously as Time to first RBC transfusion in Section 4.3.4. The OT+3 will be used.
- Number of rescue therapy treatments given; RBC transfusion or erythropoietin analogue per PEY will be reported descriptively, together and separately. The OT+3 will be used.
- Proportion of subjects receiving RBC transfusion during week 28 to week 52 will be analysed using Miettinen & Nurminen model adjusting for the region, history of CV and baseline Hb (<=8, >8 g/dL) and dialysis duration, comparing roxadustat to epoetin alfa. The OT+3 will be used.

4.3.5.3 Quality of life related exploratory endpoint analysis

- Mean change in EQ-5D-5L index value from baseline to average EQ-5D-5L index value of weeks 28-52 will be analyzed using MMRM with treatment, visit, treatment-by-visit interaction, and baseline covariates, including the baseline index value, baseline Hb, CV history, geographic region and dialysis duration, as fixed effects and subject as a random effect.
- Change in EQ-5D-5L index value from baseline at weeks 12, 28 and 52 will be analyzed using the same MMRM as specified in Section 4.3.4 for this variable.
- Shift tables of EQ-5D-5L levels 1-5 by dimension and treatment arm.
- Mean change in EQ-5D-5L VAS value from baseline to average EQ-5D-5L VAS value of weeks 28-52 will be analyzed using MMRM with treatment, visit, treatment-by-visit interaction, and baseline covariates, including the baseline VAS value, baseline Hb, CV history, geographic region and dialysis duration, as fixed effects and subject as a random effect.
- Change in EQ-5D-5L VAS value from baseline at weeks 12, 28 and 52 will be analyzed using the same MMRM as specified above for this variable.
- EQ index value and VAS mean values (+SD) and median values (+ 25^{th} & 75^{th} percentiles) at baseline and each visit per treatment arm.

4.3.5.4 Hospitalization related exploratory endpoint analysis

All endpoints in this subsection will use OT+7.

- Proportion of subjects with hospitalizations and number of days of hospitalizations per PEY will be reported descriptively.
- Number of days spent in ICU per PEY for each treatment arm will be reported descriptively.
- Proportion of subjects who are re-admitted to hospital within 30 days per patient-exposure year for each arm will be reported descriptively.
- Proportion of subjects who are re-admitted to hospital within 30 days due to heart failure preceding a hospitalization due to heart failure per PEY for each arm will be reported descriptively.
- Proportion of subjects by number of days spent in a Skilled Nursing Facility that follow hospitalizations per PEY will be reported descriptively.
- Proportion and number of days of hospitalization-free survival on treatment will be reported descriptively.
- Proportion and number of days of hospitalization-free, emergency room- free, and skilled nursing facility-free survival on treatment will be reported descriptively. Proportion of subjects with days spent in a Skilled Nursing Facility that follow hospitalization and number of days spent in a Skilled Nursing Facility that follow hospitalizations per PEY will be reported descriptively. The total number of days covering both hospitalizations and subsequent days in Skilled Nursing Facility will also be reported.

4.3.5.5 Other exploratory endpoint analysis

- The average monthly IV iron use during Week 0 to 36 and week 28 to 36 will be compared between the two treatment groups analogously as the secondary efficacy endpoint of IV iron.
- Average monthly IV iron usage per PEY per arm will be reported descriptively.
- Mean change in heart rate from baseline throughout week 28 to the EOT visit. For each subject, the change from baseline to the mean level across all heart rate values from week 28 until the EOT visit will be used as the dependent variable. An ANCOVA approach will be used with baseline heart rate, baseline Hb as covariates and the treatment groups, CV history, geographic region and dialysis duration as fixed effects.

- Change in blood pressure (DBP, SBP and MAP) from baseline throughout week 28 to the EOT visit. Analysis will be conducted using ANCOVA analogously as the analysis of change in heart rate.

- Iron parameters: Serum iron, TIBC, Ferritin and TSAT level at each testing time-point and mean change from baseline throughout week 28 to the EOT visit. For each of the serum profiles, analysis will be conducted using ANCOVA analogously as the analysis of change in heart rate.

- Change in variables concerning lipids: Total cholesterol, LDL, HDL and triglyceride. For each of the lipids, analysis will be conducted using ANCOVA analogously as the analysis of change in heart rate, from week 24 to the EOT visit. Percent of subjects who achieved target LDL level <100 mg/dL will also be compared at all available time points

- The usage of statins, types of statins and statin dose levels will be reported descriptively.

- Subject initiation of ESA therapy post study drug discontinuation will be reported descriptively.

- Change from baseline in hepcidin to week 24. Analysis will be conducted using ANCOVA analogously as the analysis of change in heart rate.

4.3.6 Sensitivity analysis of efficacy endpoints

- The analysis of the primary efficacy endpoint for US will be repeated but will exclude Hb values 6 weeks after the use of rescue therapy. ITT analysis set will be used.

- The analyses of primary efficacy endpoint and the secondary efficacy endpoints will be repeated using the OT+7.

- The analysis of the US primary efficacy endpoint and the Hb related secondary efficacy endpoints will be repeated using the PPS.

- The secondary endpoint of RBC transfusion as rescue therapy will be repeated using the ITT analysis set.

- Change in Hb from baseline using MMRM. Mean change from baseline across all Hb values from week 28 to week 52 will be analysed using baseline Hb as a covariate and treatment group, visit, visit by treatment interaction, CV history, geographic region and dialysis duration as fixed effects. ITT analysis set will be used.

- Proportion of total time of interpolated Hb within the interval of 10-12 g/dL from week 28 to the EOT visit. The difference between roxadustat and epoetin alfa will be compared using an ANCOVA model with baseline Hb as a covariate, and CV history, geographic region and dialysis duration, as fixed effects. ITT analysis set will be used.

- Change in Hb from baseline using PMM – Last Mean Carried Forward, as specified in Section 4.2.8.1. ITT analysis set will be used.

- Change in Hb from baseline using PMM – PMM – Baseline Mean Carried Forward (ANCOVA), as specified in Section 4.2.8.2. ITT analysis set will be used.

- Change in Hb from baseline using PMM – Baseline Mean Carried Forward for roxadustat subjects, MAR assumption for epoetin alfa group (ANCOVA), as specified in Section 4.2.8.2. ITT analysis set will be used.

4.3.7 Subgroup analyses

Subgroup analysis will be performed for both the primary efficacy endpoints of Hb, with the ITT analysis set for the primary endpoint for US and the PPS for the primary endpoint for EU.

- Age: <65 and ≥65; <75 and ≥75 years

- Gender: Male vs Female

- Race: White, Black or African American, Asian, Native Hawaiian or other Pacific Islander, American Indian or Alaska native, other

- Weight: <70 kg vs ≥70 kg; and <100 kg vs ≥ 100 kg

- Weight by gender-specific median (4 groups)

- Body mass index (BMI): <30 and ≥30 kg/m^2

- Geographical region: US vs Ex-US

- Geographical region:

  – North America

  – South America

  – Asia and Australia

  – Europe

- Peritoneal dialysis vs. Hemodialysis

- Cardiovascular/cerebrovascular/thromboembolic history: Yes or No
- Baseline Hb value: ≤10.5g/dL and >10.5g/dL
- Incident vs stable dialysis: dialysis duration ≤4 months vs >4 months from the randomization date
- Diabetes history: Yes vs No
- Epoetin alfa dose prior to randomization: ≤ 12,500 IU/week and >12,500 IU/week
- Baseline hsCRP (≤ULN vs >ULN).

4.3.8 Safety assessment analysis

The safety analysis will be performed using the OT+7. Safety variables include adverse events (AE), laboratory variables, vital signs, ECG variables and physical examinations. For each safety variable, the last assessment made on the screening visits or the randomization visit will be used as the baseline for all analyses, unless specified otherwise.

4.3.8.1 Adverse events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 20.0 higher.

An AE (classified by preferred term) started during the treatment period will be considered a treatment-emergent adverse event (TEAE) if it was not present prior to the first dose of study medication. An AE that starts more than 7 days after the last dose of study medication will not be counted as a TEAE.

The number, percentage and percentage per PEY of subjects reporting TEAEs in each treatment group will be tabulated by system organ class and preferred term; by system organ class; by preferred term and by system organ class, preferred term, and relationship to study medication. If more than one event occurs with the same preferred term for the same subject, the subject will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to the study medication. In addition to reporting TEAEs by number of subjects, the table by system organ class and preferred term will also be reported by patient years and event rates. Thus, allowing for potential systematic differences in mean exposure between the treatment groups. The event rate for a particular AE will be derived as the number of subjects with the AE, divided by total number of days at risk for the AE across all subjects in given group, multiplied by 365.25 multiplied by 100.

The distribution of TEAEs by severity and relationship to study medication will be summarized by treatment group.

The incidence of common (≥5% of subjects in any treatment group) TEAEs, common treatment-emergent serious AEs (TESAE), and AEs leading to discontinuation of study medication will be summarized by preferred term and treatment group, sorted in decreasing overall (across treatments) frequency. Moreover, TEAEs leading to hospitalization by system organ class and preferred term will be presented. In addition, related deaths and fatal SAEs (i.e., events that caused death) will be summarized separately by treatment group, system organ class and preferred term. TEAEs with outcome of deaths and TESAEs will also be presented for the ITT analysis set.

TEAEs and TESAEs by system organ class and preferred term will also be reported with OT+3 and OT+0.

Listings will be presented of subjects with serious adverse events (SAEs), subjects with adverse events leading to discontinuation, and subjects who died.

4.3.8.2 Laboratory variables

Descriptive statistics for laboratory values and mean percent changes from baseline at each assessment time point will be presented by treatment group for the following laboratory variables collected in the study:

- Hematology: Hemoglobin, hematocrit, RBC count, MCV, MCH, MCHC, WBC count, WBC differential, platelet counts and Reticulocyte count
- Chemistry: Alkaline phosphatase, ALT, AST, total bilirubin, LDH, total protein, albumin, fasting glucose, phosphate, uric acid, BUN, creatinine, sodium, and potassium.
- Serum iron, ferritin, TIBC, TSAT
- CHr
- Hepcidin and hsCRP

The laboratory values will be presented in SI units, except for Hb, ALT, AST, ALP and Gamma Glutamyl Transferase, which will be presented in conventional units

4.3.8.3 Vital signs

Blood pressure baselines are defined as the average of all measurements from the screening visits and randomization visit. For subjects on hemodialysis, vital signs should be recorded pre-dialysis. For subjects on peritoneal dialysis, vital signs may be recorded at any time during the visit.

Descriptive statistics for vital signs (e.g., systolic and diastolic blood pressure and MAP) and their changes from baseline at each visit and at the end of study will be presented by treatment group.

Vital sign values are potentially clinically significant (PCS) if they meet both the observed value criteria and the change from baseline criteria listed in Table 2 below. The number and percentage of subjects with post-baseline PCS values will be tabulated by treatment group.

Table 2 Criteria for Potentially Clinically Significant Vital Signs

Vital Sign Parameter	Flag	Criteria*	
		Observed Value	Change
Systolic Blood Pressure (mmHg)	High	≥170	Increase of ≥20
	Low	≤90	Decrease of ≥20
Diastolic Blood Pressure (mmHg)	High	≥110	Increase of ≥15
	Low	≤45	Decrease of ≥15
Pulse Rate (bpm)	High	≥120	Increase of ≥20
	Low	≤50	Decrease of ≥20

* A post-baseline pre-dialysis or post-dialysis value is considered as a PCS value if it meets both criteria for observed value and change from pre-dialysis or post-dialysis baseline

4.3.8.4 Electrocardiogram

QTc interval will be calculated using both Bazett ($QTcB = QT/(RR)^{1/2}$) and Fridericia ($QTcF = QT/(RR)^{1/3}$) corrections; and if RR is not available, it will be replaced with 60/HR in the correction formula.

Box plots for each variable versus visit will be produced by treatment group (roxadustat vs. epoetin alfa).

ECG values are PCS if they meet or exceed the upper limit values listed in Table 3 below. The number and percentage of subjects with post-baseline PCS values will be tabulated by treatment group. The percentages are to be calculated relative to the number of subjects with available baseline and at least one post-baseline assessment. The numerator is the total number of subjects with at least one post-baseline PCS ECG value.

Table 3 Criteria for Potentially Clinically Significant ECG

ECG Parameter	Unit	High Limit
QRS interval	Msec	≥150
PR interval	Msec	≥250
QTc interval	Msec	>500; Change from baseline >30 and >60

4.3.8.5 Physical examination

Incidence of physical examination abnormalities will be summarized for the randomization visit and the EOT visit by treatment group. Shift tables of baseline vs last observation will be provided.

4.3.9 Population PK analysis

A population PK analysis of data collected in the CKD-dialysis dependent program will be performed as outlined in a separate population PK analysis plan.

5. INTERIM ANALYSES

No interim analysis specific to this study will be conducted.

6. CHANGES OF ANALYSIS FROM PROTOCOL

There are no changes of analysis from protocol version 8.0, 19 September 2018.

6.1 Changes of analysis from previous edition of the SAP

Table 4 Major changes of analysis from SAP Edition 3.0

SAP Section	Description of change	Rationale
1.1	The objectives of the study have been split to efficacy objectives and safety objectives.	For clarification.
1.1.4	A secondary objective to evaluate the efficacy of roxadustat based on Hb response in inflamed subjects has been added A secondary objective to evaluate the effect of roxadustat on LDL cholesterol has been added. The secondary objective of the effect on self-reported health status has been removed.	To harmonize with the secondary objectives of the other phase III studies in the program.
1.3	The description of the determination of the sample size related to CV safety has been shortened	Reference to a more detailed description of the requirements for sample size to address CV safety for this

		indication is made to the pooled statistical analysis plan.
2.1.1	FAS is renamed as ITT analysis set.	To align with the definition and terminology adopted in the other phase 3 trials in the study program.
2.1.2	An additional criteria to PPS has been added, which requires subjects to be on study drug for at least 8 weeks.	To align with the definitions adopted in the other phase 3 trials in the study program, and for clarification.
2.1.3	Added additional safety analysis sets; OT+7, OT+3 and OT+0. Removed OT+28.	To align with safety analysis sets in the PSAP.
	Full analysis set (FAS) is newly defined in a new section, Section 2.1.4.	To align with the definitions and terminology adopted in the other phase 3 trials in the study program, this analysis set will be required for the EU submission.
	Have added a subsection 2.1.5 that describes how subjects who will not be included in any analysis sets will be handled.	Not included in previous editions of the SAP.
2.2	Changed the level of deviation for the important protocol deviation of compliance to subject level from visit level.	To simplify the derivation of compliance and harmonize with the other phase III studies in the program
3	The structure of this section and its subsections are rearranged. The subsection on primary efficacy variables is split into two parts, one for US FDA and the other for EU health authority. The subsections on primary and secondary safety variables is renamed as “Adjudicated CV events Analyses	To harmonize with the primary variables of the other phase III studies in the program

	for Safety Assessments" and its description is replaced by new texts on the pooling of the adjudicated composite safety endpoints from all the phase 3 studies of the program.	
Subsections of Section 3 related to secondary and exploratory efficacy variables	The primary efficacy variable designated for EU health authority is added as the first Hb-related secondary efficacy variable designated for US FDA. Changed the timing of the hsCRP variable to the average level between week 28 to week 52. Added a variable for proportion of total time of interpolated Hb values ≥10 g/dL from week 28 until week 52 to the secondary variables. A corresponding analysis has been added to Section 4 as a secondary efficacy analysis. Downgraded the secondary variables of mean change from baseline in Hb, averaged over week 28 to EOT visit, Hb response, FACT-An variable, PGIC variable and EQ-5D-5L variable to exploratory variables and their corresponding analyses to exploratory in Section 4. Added a new section for lipid related secondary efficacy variables. A corresponding analysis has been added as a secondary efficacy analysis in Section 4. Added exploratory endpoint of proportion of subjects with RBC transfusion within weeks 28-52 in	To harmonize with the secondary variables of the other phase III studies in the program To investigate the added secondary objective to evaluate LDL cholesterol.

	Section 3 and the corresponding exploratory analysis in Section 4. Added exploratory endpoint of proportion of anemic ESA-naïve subjects in Section 3 and the corresponding exploratory analysis in Section 4. Added exploratory variables related to hospitalization-free, emergency room-free, and skilled nursing facility-free survival in Section 3 and their corresponding exploratory analysis in Section 4. Added exploratory variables for measuring the proportion of time subjects were on different Hb levels in Section 3 and their corresponding exploratory analysis in Section 4. Added exploratory variables that repeat the secondary endpoints number 3, 4 and 6 on the time period week 28 to week 36.	Exploratory variables and analyses of interest
3.4	Changed the derivation of compliance	To harmonize with the definition with the other phase 3 studies in the program
4.1	A method to impute the last dose date, if missing, has been added.	To handle missing last dose dates.
4.1.1	The criteria to censor at PACD has been removed. Subjects will be censored at the EOS instead of EOT for FAS.	To harmonize with the censoring rules of the PSAP.
4.1.4	Deleted Section “Investigation of informative censoring”.	Not applicable since the CV analyses will not be performed for the individual CSR.

4.2.6	Decreased the number of multiple imputations to 200 from 1000.	To reduce the computational runtime.
Subsections of Section 4 related to statistical analyses on CV Safety	Removed the statistical analyses of the adjudicated CV events from this SAP. Deleted section "Model checking".	All analyses of CV safety will be conducted in accordance with the PSAP, and will not be done for the individual CSR. Not applicable since the CV analyses will not be performed for the individual CSR.
Subsections of Section 4 related to statistical analysis on efficacy	The section on primary efficacy endpoint analysis is split into two sections, one for US FDA and the other for EU health authority. The primary efficacy endpoint designated for EU health authority is added as the first secondary efficacy endpoint designated for US FDA Changed the ordering of the secondary efficacy endpoints. The PPS will be used for the first secondary endpoint, OT+3 analysis set will be used for the secondary endpoints related to RBC transfusion as rescue therapy, and the ITT analysis data set will be used for all the remaining secondary endpoints. Changed the analysis model of IV iron to Wilcoxon Rank Sum test. Analyses of the exploratory efficacy endpoints newly added in Section 3 are specified accordingly.	To harmonize with the primary variables of the other phase III studies in the program To harmonize with the secondary endpoints of the other phase III studies in the program. A decision based on balancing clinical importance of different endpoints together with the likelihood for success. To better the fit the distribution of the IV iron data. Analyses of interest

	The analysis data sets of individual sensitivity analyses are updated, and new sensitivity analyses are added.	To harmonize with the sensitivity analyses of the other phase III studies in the program
4.3.10.1	The sentence "Finally, TESAEs that occurred during the 4-week period preceding an excessive erythropoiesis event will be presented by system organ class, preferred term and treatment group" is removed. Change the safety analyses to be on OT+7 instead of OT+28. Added analyses on key safety variables for OT+7, OT+3 and OT+0.	To align with safety analysis sets in the PSAP and the safety analysis in the other phase 3 studies in the program. Analyses of interest.
4.3.10.3	Vital sign baseline definition has changed from using the last assessment prior to the first dose to the average of all measurements from the screening visits and randomization visit.	Deemed to be a more meaningful definition of the baseline from a clinical perspective.

7. REFERENCES

Aranesp 2001
Darbepoetin alfa Product Approval Information Summary. Available at **http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/TherapeuticBiologicApplications/ucm080442.htm. Accessed on 22MAY2014**.

Carpenter et al. 2013
Carpenter JR, Roger JH and Kenward MG, *Analysis of longitudinal trials with protocol deviation: - A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation.* Journal of Biopharmaceutical Statistics, issue 6 (November/December) in volume 23 (2013). 1352-137

Gart & Nam 1990
Gart J., and Nam J. Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables. Biometrics

1990, 46: 637-643.

Miettinen & Nurminen 1985
Miettinen O, Nurminen M. *Comparative analysis of two rates.* Statistics in Medicine; 4(2), 213-26, 1985

O'Kelly & Ratitch, 2014
O'Kelly M, Ratitch B, 2014. *Clinical Trials with Missing Data – A Guide for Practitioners*. West Sussex: John Wiley & Sons

Omontys 2012
Omontys Product Approval Information Summary. Available at http://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/202799_Omontys_Orig1s000TOC.cfm. Accessed on 22MAY2014.

Rubin, 1987
Rubin, D. B. *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Schafer, 1997
Schafer, J. L. *Analysis of incomplete multivariate data*. London: Chapman and Hall.

8. APPENDIX

Not applicable.