Statistical Analysis Plan

Study Code	D5740C00001
Edition Number	3.0
Date	28 September 2018

A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety and Efficacy of Roxadustat for the Treatment of Anemia in Chronic Kidney Disease Patients not on Dialysis

A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety and Efficacy of Roxadustat for the Treatment of Anemia in Chronic Kidney Disease Patients not on Dialysis

Study Statistician PPD _____ 28sep2018
Date

A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Safety and Efficacy of Roxadustat for the Treatment of Anemia in Chronic Kidney Disease Patients not on Dialysis

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 model
BIW	Twice weekly
BP	Blood pressure
CHr	Reticulocyte hemoglobin content
CKD	Chronic kidney disease
CKD-DD	Chronic kidney disease with subject on dialysis
CKD-NDD	Chronic kidney disease with subject not on dialysis
CMH	Cochran-Mantel-Haenszel
CRF	Case report form
hsCRP	High-sensitivity C-reactive protein
CS	Clinically significant
CSE	Composite Safety Endpoint
CSR	Clinical Study Report
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
eGFR	Estimated glomerular filtration rate
EOS	End of study
EQ-5D-5L	EuroQol Health Utility Index, 5 dimensions 5 levels
FACT-An	Functional Assessment of Cancer Therapy-Anemia
EOS	End of study
EOT	End of treatment
ESA	Erythropoiesis-stimulating agent
ESRD	End-stage renal disease
FACT-An	Functional Assessment of Cancer Therapy - Anemia
FAS	Full Analysis Set

Abbreviation or special term	Explanation
FDA	Food and Drug Administration
Hb	Hemoglobin
HDL	High-density lipoprotein
HR	Hazard ratio
HRQoL	Health Related Quality of Life
ICU	Intensive Care Unit
IPCW	Inverse Probability of Censoring Weights
ITT	Intention to Treat
IV	Intravenous
LDL	Low-density lipoprotein
LOCF	Last observation carried forward
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
PF	Physical Functioning subscale component of SF-36
PGIC	Patients' Global Impression of Change
PPS	Per-protocol Set
PMM	Pattern Mixture Model

Abbreviation or special term	Explanation
PSAP	Pooled Statistical Analysis Plan
QW	Once weekly
RBC	Red blood cell
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SF-36	Short Form 36
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
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 adjudicated CV safety data to the pool safety analysis across the study program to ensure sufficient data. 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 choice of secondary objectives and variables, as well as the ordering of the secondary efficacy endpoints in the hierarchical testing, to align with the order of the other phase 3 studies in the NDD 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 is listed in Section 6.
25 January 2018	An important change in edition 2.0 is on the analysis of adjudicate safety data which will be conducted according to the pooled statistical analysis (PSAP) on studies across the phase 3 program. Another important change is the addition of a primary efficacy endpoint, and addition of further secondary endpoints, as well as ordering of the secondary endpoints, done to test efficacy of roxadustat and 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, Changes of Analysis from Protocol.

1. STUDY DETAILS

1.1 Study objectives

This study is part of the study program for subjects with Chronic Kidney Disease who are not dependent on dialysis (CKD-NDD). The other studies in the study program are FG-4592-060 and 1517-CL-0608 which are also placebo-controlled. 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 provide MACE events required for the pooled safety analysis evaluating the safety of roxadustat for the treatment of anemia in CKD patients not on dialysis. This study is also known as "OLYMPUS".

The objectives of the current study are to evaluate the efficacy and safety of roxadustat compared to placebo for the treatment of anemia in subjects with Stage 3, 4 or 5 CKD who are not on dialysis.

1.1.1 Primary efficacy objective

The primary efficacy objective is to evaluate the efficacy of roxadustat compared to placebo for the treatment of anemia in CKD subjects not on dialysis.

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 placebo based on Hb response and level during the study
- The efficacy of roxadustat compared to placebo based on Hb response in inflamed subjects
- The effect of roxadustat on Low-density lipoprotein (LDL) cholesterol as compared to placebo.
- The need for rescue therapy in subjects treated with roxadustat as compared to placebo.
- The effect of roxadustat on anemia symptoms and health-related quality of life (HRQoL) based on comparison with placebo.

- The effect on the CKD progression of roxadustat as compared to placebo.

1.1.4 Secondary safety objectives

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

1.2 Study design

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled study in anemic subjects (Hb <10 g/dL) with Stage 3, 4, or 5 CKD who are not on dialysis.

This study will consist of three study periods as follows:

- **Screening Period**: Up to 6 weeks
- **Treatment Period:** Subjects will be randomized (1:1) to double-blind treatment with roxadustat or placebo. 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. 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

1.2.1.1 Investigational product, dosage and mode of administration

The initial study drug dose (per dose occasion) is 70 mg orally administered three times a week (TIW). Moreover, the dose is subsequently adjusted to achieve and maintain Hb levels between 10 and 12 g/dL. Study drug will be dosed TIW throughout the study treatment period unless downward dose adjustment requires a change to twice or once weekly.

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.

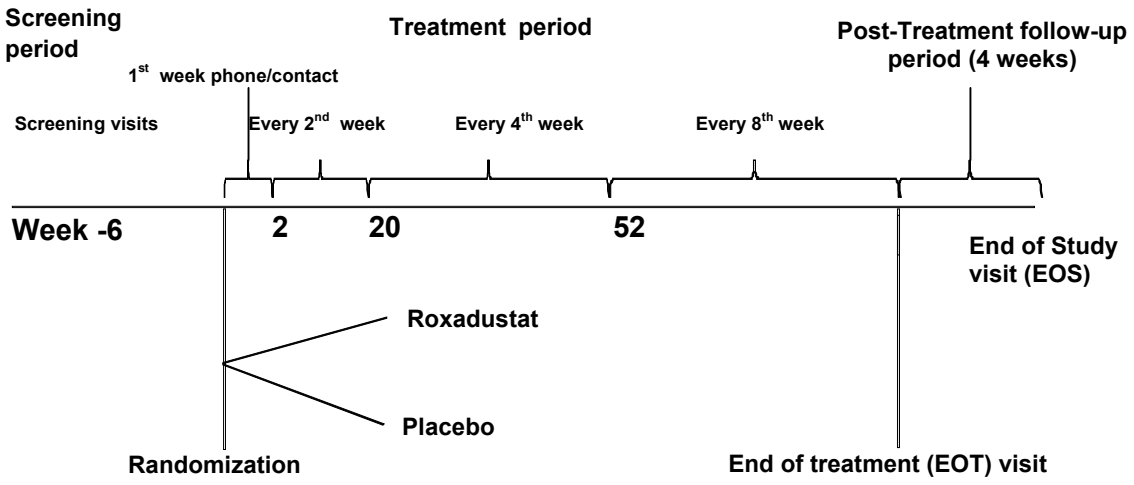
In the event of excessive erythropoiesis or excessive Hb levels ≥13 g/dL, the dose will be adjusted or put on hold at any time. Excessive erythropoiesis is defined as an Hb increase by >2.0 g/dL within a 4-week period.

1.2.2 Scheduled visits during treatment

During the screening period, eligibility will be confirmed at a minimum of 2 screening visits. If eligible, the subject will be randomized at the randomization visit. 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 week 52, then every 8 weeks until the end of the treatment period. A study end date will be defined based on when the planned number of events are estimated to be accrued, the end of treatment (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 flow chart



1.2.3 Stratification variables

The randomization in this study will only be stratified by country. The stratification variables for the other two studies in the program will be used in the analyses for this study as covariates. The variables are:

1. Baseline eGFR ($\leq$ 30 mL/min/1.73 m^2 vs > 30 mL/min/1.73 m^2)
2. Baseline Hb ($\leq$ 8 g/dL vs > 8 g/dL)
3. Cardiovascular/cerebrovascular/thromboembolic medical history (Yes vs. No)
4. geographical region (US vs Ex-US)

Baseline Hb and baseline eGFR will be included in the analyses as continuous covariates, not as dichotomous factors, unless specified otherwise. Throughout this document, the variable cardiovascular/cerebrovascular/thromboembolic medical history will be shortened as CV history.

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 endpoint: A sample size of 450 subjects will have > 99% power to detect a 0.75 g/dL difference in mean Hb values between the two treatment groups, assuming that the common standard deviation is 1.2 g/dL, using an analysis of variance (ANOVA) test with a 0.05 two-sided significance level.

To contribute an adequate number of adjudicated CV events for the pooled CV analyses across the phase 3 program, approximately 2600 subjects are planned to be randomized in this study in a 1:1 ratio to either roxadustat or placebo. This sample size is driven by the overall requirement of adjudicated CV events for the overall study program in CKD-NDD (which consists of 3 studies in total targeting 465 subjects with MACE events), all evaluating the cardiovascular safety of roxadustat compared to placebo. The two other placebo-controlled studies in the study program are FGCL-4592-060 and 1517-CL-0608.

All event numbers in this section refer to events accrued in the Safety Analysis Set.

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 analysed 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. 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 2.1.4. Further details of important protocol deviations are available in a Protocol Deviation Plan. Subjects will be censored at the earliest of date of an important protocol deviation, the EOS visit, or 28 days after 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 28 days after last intake of study drug. The Safety Analysis Set may also be referred to as On-treatment+28 (OT+28).

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 Hb 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 unblinding 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, and 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 1-7, 10-17, 19 and 24. 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 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 whether subjects achieved Hb response as described below.

[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 patient-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.

- 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 [according to 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. Subjects who have discontinued study medication or received rescue therapy before a Hb response could be achieved will be considered as a non-responder.

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 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 within the interval of ≥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 measurements at week 28 and week 52. Subjects 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 in LDL cholesterol from baseline to week 24.

3.1.4 Rescue therapy related secondary efficacy variables

The need for rescue therapy will evaluated as

- Time-to-first (and proportion of subjects who received) instance of receiving intravenous (IV) iron, red blood cell (RBC) transfusions, or erythropoietin analogue (composite) as rescue therapy (rescue therapy guidelines are specified in the clinical study protocol (CSP), Section 7.7.4)
- Time-to-first (and proportion of subjects who received) administration of RBC transfusion as rescue therapy

For analyses based on the OT+28, 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.

Furthermore, the other components of the above composite will be analyzed separately as exploratory endpoints:

- Time-to-first (and proportion of subjects who received) administration of erythropoietin analogue as rescue therapy
- Time-to-first (and proportion of subjects who received) administration of IV iron as rescue therapy

3.1.5 36-Item Short Form Health Survey

The SF-36 is a QoL instrument designed to assess generic health concepts relevant across age, disease, and treatment groups. It is aimed at both adults and adolescents aged eighteen years and older. The SF-36 consists of eight domains of health status: Physical functioning (10 items), Role-physical (4 items), Bodily pain (2 items), General health (5 items), Vitality (4 items), Social functioning (2 items), Role emotional (3 items) and Mental health (5 items). Two component scores, the Physical Component Summary and the Mental Component Summary can also be calculated. For both the SF-36 domain scores and summary scores, higher scores indicate better health status. The SF-36 has a recall period of the ‘past four weeks’.

- Change from baseline in SF-36 Vitality (VT) sub-score
- Change from baseline in SF-36 Physical Functioning (PF) sub-score

The SF-36 scores will be computed according to its documentation.

3.1.6 CKD progression related secondary efficacy variables

To evaluate the effect of roxadustat on the CKD progression prior to initiation of dialysis or kidney transplant, the following variable will be analyzed:

- Annual rate of eGFR change, calculated as the slope of eGFR values to prior to initiation of dialysis/kidney transplant

Baseline eGFR is defined as the mean of all available central lab values prior to randomization, including the value from the randomization visit.

3.1.7 Exploratory variables

The exploratory variables are:

3.1.7.1 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 11 g/dL. 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. Time will be computed analogously as time to first rescue therapy.

- Proportion of time on study with Hb <7, 7 to <8, 8 to < 9, 9 to <10, 10 -<11, 11-<12, 12-<13, >13g/dL; proportion of time on study with Hb < 10 g/dL after week 12; proportion of time on study with Hb < 10 g/dL after week 24;

3.1.7.2 Rescue therapy related exploratory variables

- Number of rescue therapy treatments given, IV iron, RBC transfusion or erythropoietin analogue, per patient exposure year (PEY).

3.1.7.3 CKD progression related exploratory variables

All eGFR values obtained after the initiation of dialysis or renal transplant will be censored.

- Time to a renal composite endpoint which is defined by the components:
    - Doubling of creatinine or
    - Initiating chronic dialysis treatment or
    - All-cause mortality,

for subjects with eGFR >15 ml/min/1.73m2 at baseline. A doubling of creatinine from baseline is confirmed with two consecutive measurements separated at least 21 days apart. The date of the first measurement will be used as the date when the event was reached. Time will be computed analogously as time to first rescue therapy.

- Time to initiation of chronic dialysis. Time will be computed analogously as time to first rescue therapy. This will be analysed in all patients and in subgroups with eGFR >15, >20, >30 mL/min/1.73 m^2.

- Time to an eGFR decrease of 30%, 40% and 50% from baseline, prior to initiation of dialysis or renal transplant, and will be analysed using the Cox model analogously as Time to-first rescue therapy (composite) in Section 4.3.4, this will be analysed in all patients and in subgroups with eGFR >15, >20, >30 mL/min/1.73 m^2. A eGFR decrease of 30%, 40% or 50% from baseline is confirmed with two consecutive measurements separated at least 21 days apart. The date of the first measurement will be used as the date when the event was reached. Time will be computed analogously as time to first rescue therapy.

3.1.7.4 Quality of life related exploratory efficacy variables

Evaluate the effect of roxadustat on anaemia symptoms and disease-specific health-related quality of life (HRQoL) and overall health status: The variables are:

- Changes in anaemia symptoms as measured by the Functional Assessment of Cancer Therapy-Anemia (FACT-An).
- Changes in self-reported health status as measured by the EuroQol Health Utility Index, 5 dimensions, 5 levels (EQ-5D-5L) and
- Patients' Global Impression of Change (PGIC).
- Change from baseline in SF-36 Vitality (VT) sub-score in subjects with baseline VT subscore in subjects with with baseline VT subscore below 50
- Change from baseline in SF-36 Physical Functioning (PF) sub-score in subjects with with baseline PF subscore below 35

The FACT-An, EQ-5D-5L, and PGIC scores will be computed according to their respective documentation.

Patient Global Impression of Change Scale (PGIC)

The PGIC is a subject-rated instrument that measures change in subjects' overall status on a 7-point scale ranging from 1 (very much improved) to 7 (very muchworse), when compared to the start of the study treatment.

FACT-An

The Functional Assessment of Cancer Therapy- General (FACT-G) Version 4 contains 27 items that cover four dimensions of well-being: physical (PWB)—7 items, functional (FWB)—7 items, social/family (SWB)—7 items each, and emotional (EWB)—6 items. A subscale of 13 fatigue specific items (the Fatigue Subscale) plus seven additional items related to anemia were developed for use in conjunction with the FACT-G. The 13 fatigue items plus the seven additional items related to anemia comprise the Anemia Subscale (AnS). Administration of the FACT-G plus the AnS is referred to as the FACT-An. The FACT-An has a recall period of the 'past seven days'. Respondents are asked to provide responses, (i.e., 'Not at all', 'A little bit', 'Somewhat', 'Quite a bit' and 'Very much'), to a list of statements which are either positively or negatively phrased. For all FACT-An scales, a higher score indicates better QoL.

- Changes in the score of the FACT-An in total score and anemia subscale.

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

3.1.7.5 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 and number of on-treatment days hospitalization-free
- Proportion and number of on-treatment days hospitalization-free, emergency room-free, and skilled nursing facility-free.
- 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.

- 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.

3.1.7.6 Other exploratory variables

- Variables concerning lipids: total cholesterol, high-density lipoprotein (HDL) and triglycerides. Lipid levels at all available timepoints and change from baseline to testing timepoints. Percent of subjects who achieved LDL target of <100 mg/dL will also be compared at all available time points.

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

- Change from baseline in hepcidin to week 24.

- Initiation of ESA therapy post study drug discontinuation

- Usage of statins and types of statins

3.2 Safety assessment

Safety will be assessed by evaluating the following:

- Occurrence and event rates 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, and proportion of values above, within or below normal ranges

- Occurrence of clinically significant changes from baseline in electrocardiogram (ECG) values.

3.3 Adjudicated CV Events Analyses for Safety Assessments

The CV events to be analyzed according to 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-4592-060, 1517-CL-608 and D5740C00001) 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 analysis 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 “Placebo”.

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 the Safety Analysis Set, subjects will be censored at the 28 days after 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.

In analyses of eGFR and creatinine, all eGFR and creatinine measurements after initiation of dialysis or renal transplant will be censored.

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.

4.2 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, baseline eGFR, geographical region, 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, baseline iron replete 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 <=8, >8 - <=9, >9 g/dL
- baseline eGFR; <10, 10 to <15, 15 to <30, 30 to <45, 45 to <60, ≥60 mL/min/1.73m^2,
- 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 4.3.2 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 as described in detail below. Unless specified otherwise baseline Hb, baseline eGFR as continuous variables will be used as covariates, and treatment group, CV history 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 based on 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.

4.2.3.1 Cox regression adopting Inverse of Probability of Censoring Weights

A separate document will describe the IPCW specification to be used to adjust for potential asymmetric dropout due to informative censoring.

4.2.4 Analysis of Covariance (ANCOVA)

When using ANCOVA for the 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, baseline eGFR will be used as covariates, and treatment group, cardiovascular history and geographic region 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.

4.2.5 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 with baseline Hb, baseline eGFR, CV history and geographic region, and the available non-missing Hb for each scheduled week are used as covariates, and 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, baseline eGFR, CV history and geographic region 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 observed and imputed Hb values between weeks 28 to 52 for each subject is taken as the dependent variable and baseline Hb, baseline eGFR, treatment group, CV history and geographic region 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.

4.2.6 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 are 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 at the end of the double blind period and to the variance-covariance structure of the longitudinal model.

The MMRM can contain terms for baseline measurements, 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 number 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.

4.2.7 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. 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.

4.2.7.1 PMM – Last Mean Carried Forward

A PMM 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 variable. 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 method is the same as for multiple imputation ANCOVA described in Section 4.2.5 with the exception of step 2, where the monotone datasets are imputed. Parameter below refers to the parameter of the multi-normal distribution for baseline and post baseline Hb measurement.

1. Create posterior distribution of parameter: Separately for each treatment arm, 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) using a Bayesian approach with an improper prior. Moreover, fit an unstructured variance covariance matrix using an uninformative Jeffrey's' prior (Schafer, 1997, p. 155).

2. Draw parameters: Separately for each treatment arm, draw variance-covariance matrix from the posterior distribution for the parameters. The mean vector is 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 given observed data: For each subject with missing data, use their joint distribution in previous step to construct their conditional distribution of missing data given observed data. Sample the missing data from this conditional distribution, to create a "completed" data set.

Repeat the above steps 200 times, which will result 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 for final inference.

4.2.7.2 PMM – Baseline Mean Carried Forward (roxadustat group only and both groups)

The analysis is the same as PMM – Last Mean Carried Forward except 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 the placebo 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.

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.

4.2.7.3 PMM – Jump to Control

A PMM using jump to control multiple imputation method (Carpenter et al, 2013) will also be used as another sensitivity analysis similar to the PMM-Last Mean Carried Forward with the exception of step 3. The joint distribution of the subjects observed and missing data are considered multivariate normal with mean and covariance matrix from following the placebo treatment arm.

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 analysed with multiple imputation ANCOVA as described in Section 4.2.4 and 4.2.5. The model will contain terms for the treatment, baseline Hb measurement, baseline eGFR, geographic region and CV history. Superiority of roxadustat compared to placebo 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 placebo exceeds 0 g/dL. The ITT analysis set will be used.

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

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. The proportion of responders in the primary efficacy variable will be compared using a Cochran–Mantel–Haenszel (CMH) test adjusting for the region, history of CV, baseline Hb (<=8, >8 g/dL) and baseline eGFR (<=30, >30 mL/min/1.73 m2), comparing roxadustat to placebo. The FAS will be used.

4.3.4 Secondary Efficacy Endpoints:

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 FAS will be used on the first secondary endpoint, OT+28 analysis set will be used for the two secondary endpoints related to rescue therapy, and the ITT analysis data set will be used for all the remaining secondary endpoints."

1. Hb response as noted for EU Primary Endpoint.

2. 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 for US. Superiority of roxadustat compared to placebo 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 placebo exceeds 0 g/dL.

3. 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 placebo will be compared using ANCOVA with treatment group, geographic region and CV history as fixed factors and baseline Hb and baseline eGFR as covariates. Superiority between the groups will be declared, and this test as successful, if the lower bound of the 2-sided confidence interval of the difference between roxadustat and placebo exceeds 0.

4. 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 placebo will be compared using ANCOVA with treatment group, geographic region and CV history as fixed factors and baseline Hb and baseline eGFR as covariates. Superiority between the groups will be declared, and this test as successful, if the lower bound of the 2-sided confidence interval of the difference between roxadustat and placebo exceeds 0.

5. Mean change from baseline in LDL cholesterol to week 24 will be analysed using ANCOVA. Baseline Hb, baseline eGFR and baseline LDL will be used as covariates and treatment arm, CV history and geographic region as fixed effects. Superiority will

be declared if the upper bound of the 2-sided 95% confidence interval of the difference between roxadustat and placebo exceeds 0.

6. Time-to-first rescue therapy (composite) of any of IV iron, RBC transfusion or erythropoietin analogue (and proportion of subjects who received) as rescue therapy, will be analyzed using the Cox proportional hazard model with OT+28. Baseline Hb, baseline eGFR, geographic region, and CV history will be included as covariates. Superiority will be claimed, and this test successful, if the upper limit of the 2-sided 95% CI for the hazard ratio is less than or equal 1.0.

7. Time-to-first RBC transfusion (and proportion of subjects who received) as rescue therapy, will be analyzed using the Cox proportional hazard model with OT+28. Baseline Hb, baseline eGFR, geographic region, and CV history will be included as covariates. Superiority will be claimed, and this test successful, if the upper limit of the 2-sided 95% CI for the hazard ratio is less than or equal 1.0.

8. Mean change in SF-36 Vitality (VT) sub-score from baseline to average VT sub-score of weeks 12-28 will be analyzed using MMRM with treatment, visit, treatment-by-visit interaction, and baseline covariates, including the baseline score, baseline Hb, baseline eGFR, and geographic region and CV history, as fixed effects and subject as a random effect. Superiority between roxadustat and placebo will be declared, and this test successful, if the lower bound of the 2-sided confidence interval for the difference between roxadustat and placebo exceeds 0.

9. Annual rate of eGFR change prior to initiation of Dialysis/Transplant will be estimated with MMRM using all post-baseline eGFR values prior to initiation of dialysis/kidney transplant. Baseline eGF, baseline Hb and geographic region, CV history, treatment group and post-baseline eGFR measurement time will be treated as fixed effects, and subject and time as random effects, i.e. random intercept and slope.

10. Mean change in SF-36 Physical Functioning (PF) sub-score from baseline to average PF sub-score of weeks 12-28 will be analyzed using MMRM with treatment, visit, treatment-by-visit interaction, and baseline covariates, including the baseline score, baseline Hb, baseline eGFR, and geographic region and CV history, as fixed effects and subject as a random effect. Superiority between roxadustat and placebo will be declared, and this test successful, if the lower bound of the 2-sided confidence interval for the difference between roxadustat and placebo exceeds 0.

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 the ITT analysis set, unless specified otherwise. The variables will be analysed as follows:

4.3.5.1 Hb related exploratory endpoint analysis

- Mean change in Hb from baseline to the subjects mean level from week 28 until the EOT visit will be analysed with ANCOVA where the mean change in Hb from baseline of all available values from week 28 to the EOT visit for each subject will be the dependent variable. The mean change from baseline will be imputed for subjects with no measurements from week 28 to the EOT visit with a multiple imputation ANCOVA approach and a MAR assumption. The model will contain terms for the baseline Hb measurement, baseline eGFR, CV history and geographic region and treatment arm.

- Time to achieving target Hb will be analyzed analogously as Time to first rescue therapy (composite) in Section 4.3.3. Target Hb is achieved when Hb level is within 10-12 g/dL at two consecutive measurements.

- Proportion of time on study with Hb <7, 7 to <8, 8 to < 9, 9 to <10 g/dL, 10-12, 12-<13, >13 g/dL; proportion of time on study with Hb < 10 g/dL after week 12; proportion of time on study with Hb < 10 g/dL after week 24. These will be reported descriptively.

- Estimation of median time (in weeks) to achieve target Hb, based on the definition of two consecutive Hb levels within 10-12 g/dl, by study drug.

4.3.5.2 Rescue therapy related exploratory endpoint analysis

- Number of rescue therapy treatments given; RBC transfusion, IV iron or erythropoietin analogue per PEY will be reported descriptively, together and separately. OT+28 will be used.

4.3.5.3 CKD progression related exploratory endpoint analysis

- Time to the renal composite endpoint for subjects with baseline eGFR >15 will be analysed using the Cox proportional hazard model. Baseline Hb, baseline eGFR, geographic region, and CV history will be included as covariates.

- The secondary efficacy endpoint of annual rate of eGFR change will be repeated for subjects with baseline eGFR >30 and >20 mL/min/1.73 m^2.

- Time to initiation of dialysis will be analyzed using the Cox model analogously as analogously as Time to-first rescue therapy (composite) in Section 4.3.4. This will be analysed in all patients and in subgroups with eGFR>15, >20, >30 mL/min/1.73m^2

- Time to an eGFR decrease of 30%, 40% and 50% from baseline, prior to initiation of dialysis or renal transplant, will be analysed using the Cox model analogously as Time to-first rescue therapy (composite) in Section 4.3.4, this will be analysed in all patients and in subgroups with eGFR >15, >20, >30 mL/min/1.73 m^2.

4.3.5.4 Quality of life related exploratory endpoint analysis

- Mean change in FACT-An score from baseline to average FACT-An score of weeks 28-52 will be analyzed using MMRM with treatment, visit, treatment-by-visit interaction, and baseline covariates, including the baseline score, baseline Hb, baseline eGFR, and geographic region and CV history, as fixed effects and subject as a random effect.

- Mean change in FACT-An subscale score of anemia from baseline to average score of weeks 28-52 will be analyzed using MMRM with treatment, visit, treatment-by-visit interaction, and baseline covariates, including the baseline score, baseline Hb, baseline eGFR, CV history and geographic region, as fixed effects and subject as a random effect.

- Change in FACT-An score 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.

- Change in SF-36 PF and VT score 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.

- Change in PGIC 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.

- 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.

- 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, baseline eGFR, CV history and geographic region 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.

- Shift tables of EQ-5D-5L levels 1-5 by dimension and treatment arm.

- EQ index value and VAS mean values (+SD) and median values (+ 25^{th} & 75^{th} percentiles) at baseline and each visit per treatment arm. (descriptive)

- Mean change in PGIC score from baseline to average PGIC score of weeks 28-52 will be analyzed using MMRM with treatment, visit, treatment-by-visit interaction, and baseline covariates, including the baseline Hb, baseline eGFR, and geographic region and CV history, as fixed effects and subject as a random effect. Superiority between roxadustat and placebo will be declared, and this test successful, if the lower bound of the 2-sided confidence interval for the difference between roxadustat and placebo exceeds 0.

4.3.5.5 Hospitalization related exploratory endpoint analysis

All endpoints in this subsection will use OT+28.

- 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 and number of days of hospitalization-free days on treatment will be reported descriptively.

- Proportion of hospitalization-free, emergency room- free, and skilled nursing facility-free days 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.

4.3.5.6 Other exploratory endpoint analysis

- Change in heart rate from baseline throughout week 28 to the EOT visit. For each subject, the mean change from baseline across all heart rate values from week 28 to the EOT visit will be used as the dependent variable. An ANCOVA approach will be used with baseline Hb, baseline eGFR, baseline heart rate, treatment arm, CV history and geographic region as covariates.

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

- Change in the serum iron profiles: Ferritin, TIBC and TSAT from baseline throughout week 28 to 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 triglycerides. For each of the lipids, analysis will be conducted using ANCOVA analogously as the analysis of change in heart rate, from week 24 to EOT. Percent of subjects achieved target LDL of <100 at the various time points that lipids were measured. Subgroup analysis will be conducted in subjects on statin and/or other cholesterol-lowering drug

- Subjects 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 analysis of the primary efficacy endpoint and the secondary efficacy endpoints (with the exception of the rescue therapy secondary endpoints) will be repeated using OT+28.

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

- 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, baseline eGFR, visit and visit by treatment interaction, treatment arm, CV history and geographic region as covariates. 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 placebo will be compared using an ANCOVA model with baseline Hb and eGFR, CV history and geographic region as covariates. ITT analysis set will be used.

- Change in Hb from baseline using PMM – Last Mean Carried Forward, as specified in Section 4.2.7.1. ITT analysis set will be used.
- Change in Hb from baseline using PMM – Baseline Mean Carried Forward, as specified in Section 4.2.7.2. ITT analysis set will be used.
- Change in Hb from baseline using PMM – Jump to Control as specified in Section 4.2.7.3. ITT analysis set will be used.
- The analyses of the time to rescue therapy endpoints will be repeated using IPCW instead of the standard Cox regression.

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 FAS 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 vs ≥30 kg/m^2
- Geographical region: US vs Ex-US
- Geographical region:
  - North America
  - South America
  - Asia
  - Europe
- Cardiovascular/cerebrovascular/thromboembolic history: Yes or No
- Baseline Hb value: ≤8g/dL vs >8 g/dL and ≤9 vs >9 g/dL
- Baseline eGFR value: <30 vs ≥30 and <15 vs ≥15, <10 vs ≥10 mL/min/1.73m^2

- Diabetes history: Yes vs No
- Baseline hsCRP (≤ULN vs >ULN).
- Iron replete at baseline: (Ferritin>100 and TSAT>20%)

4.3.8 Safety assessment analysis

The safety analysis will be performed using the OT+28. Safety parameters include adverse events (AE), laboratory parameters, 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 or 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 28 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; by system organ class, preferred term, and relationship to study medication; and by system organ class and the baseline eGFR categories <10, 10 to <15, 15 to <30, 30 to <45, 45 to <60 mL/min/1.73m^2. 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.

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.

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	High	≥170	Increase of ≥20

Vital Sign Parameter	Flag	Criteria*	
		Observed Value	Change
(mmHg)	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 value is considered as a PCS value if it meets both criteria for observed value and change from 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. placebo).

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- non-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

Changes of analysis from protocol version 6.0, 31 August 2018.

The analysis of eGFR rate will not use log transformation of eGFR values.

6.1 Changes of analysis from previous edition of the SAP

Table 4 Major changes in the SAP from edition 2, 25 January 2018.

SAP Section of previous edition	Description of change	Rationale
1.1	The objectives of the study have been split to efficacy objectives and safety objectives.	For clarification.
1.1.2	The safety objectives have been revised.	Harmonizes D5740C00001 with the other phase 3 trials in the roxadustat CKD non-dialysis program, to serve as a pivotal study for confirming efficacy and safety, and facilitates assessment of pooled safety across the phase 3 trials.
1.1.3	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 (PGIC) is now	To harmonize with the secondary objectives of the other phase 3 studies in the program.

	defined as an exploratory efficacy endpoint.	
1.3	The sample size determination text has been revised	Harmonization with the strategy for the analyses of CV safety for the study program as described in the PSAP.
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.
	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 3 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 for Safety Assessments" and its description is replaced by new texts on the pooling of	To harmonize with the primary variables of the other phase III studies in the program

	the adjudicated composite safety endpoints from all the phase 3 studies of the program.	
	Added a new section for lipid related secondary efficacy variables. A corresponding analysis has been added as a secondary efficacy analysis in Section 4.	To investigate the added secondary objective to evaluate LDL cholesterol.
3.3.4.1/ 3.3.4.3/ 3.3.4.4	Moved the sections of the secondary variables FACT-An, EQ-5D-5L and PGIC, section numbers 3.3.4.1, 3.3.4.3 and 3.3.4.4 respectively to exploratory variables and set their corresponding analysis as exploratory in Section 4.	These variables have been downgraded to exploratory variables.
3.3	The secondary efficacy variable of Hb response has been specified as the primary efficacy endpoint for EU. The primary efficacy variable designated for EU health authority is added as the first Hb-related secondary efficacy variable designated for US FDA. Downgraded the secondary variables of mean change from baseline in Hb, averaged over week 28 to EOT visit, FACT-An variable, PGIC variable and EQ-5D-5L variable to exploratory variables. Upgraded the exploratory variable of time-to-first (proportion of subjects receiving) RBC transfusion as rescue therapy to the secondary efficacy variable and set its corresponding analysis as secondary efficacy analysis in Section 4.	To harmonize with the secondary variables of the other phase 3 studies in the program
3.3.2	Changed the timing of the hsCRP variable to the average level between week 28 to week 52.	To harmonize with the primary efficacy variable To harmonize with the secondary

	Added a variable for proportion of total time of interpolated Hb values ≥10 g/dL from week 28 until week 52. A corresponding analysis has been added to Section 4 as a secondary efficacy analysis.	variables of the other phase 3 studies in the program
3.3.5	Set the annual rate of eGFR change as a secondary variable and its corresponding analysis in Section 4 as a secondary efficacy analysis. Set the renal composite endpoint as an exploratory variable and its corresponding analysis as an exploratory in Section 4. Added all-cause mortality as a component to the renal composite endpoint. Replaced the component of 40% eGFR decrease with doubling of creatinine.	To harmonize with the secondary variables of the other phase 3 studies in the program
3.3.6/ 4.3.6	Added analyses of proportion of time on study with different Hb categories over various timeframes Added further analyses of hospitalization-free, emergency room- free, skilled nursing facility-free survival and LDL as exploratory variables in Section 3 and exploratory analysis in Section 4. Added an analysis measuring the proportion of time subjects were on different Hb levels as exploratory variables in Section 3 and exploratory analysis in Section 4. Updated individual analyses of CKD progression related endpoints Removed analyses of proportion of subjects with at least one occurrence of excessive erythropoiesis, Hb≥12, Hb≥13 and Hb≥14.	Analyses of interest and also to harmonize with other phase 3 studies in the program.

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.
4.1.2	Removed the subsection and the criteria to censor at PACD. Subjects will be censored at the EOS instead of EOT for ITT and FAS. eGFR and creatinine values will also be censored at time of renal transplant.	To harmonize with the censoring rules of the PSAP. To harmonize with the censoring rules of the PSAP. To include another renal replacement therapy which influences eGFR and creatinine values.
4.1.4	Deleted Section “Investigation of informative censoring”.	Not applicable since the CV analyses will not be performed for the individual CSR.
4.2.5	Decreased the number of multiple imputations to 200 from 1000.	To reduce the computational runtime.
4.2.8	Referred the specification of the IPCW to a separate document.	To harmonize the analysis with the other phase III studies in the program.
4.3.1/ 4.3.2/ 4.3.3	Removed the statistical analyses of the adjudicated CV events from this SAP.	All analyses of CV safety will be conducted in accordance with the PSAP, and will not be done for the individual CSR
4.3.5	The ITT analysis set will be used for the primary efficacy endpoint for the US FDA. FAS will be used for the primary efficacy endpoint for the EU health authority.	To be in agreement with the other studies in the program.
4.3.2	The section on primary efficacy endpoint analysis is split into two sections, one for US FDA and the other for EU health authority.	To harmonize with the primary variables of the other phase 3 studies in the program

4.3.4	The analysis of Hb response (Y/N) during the first 24 weeks of treatment without having received any rescue therapy is moved to the primary efficacy endpoint analysis for EU health authority and the first secondary efficacy analysis for US FDA.	This efficacy endpoint is now the primary for EU health authority and the first secondary for US FDA.
4.3.5	Changed the ordering of the secondary efficacy endpoints. The primary efficacy endpoint designated for EU health authority is added as the first secondary efficacy endpoint designated for US FDA ITT analysis set is used for all the secondary efficacy endpoints except for the first secondary efficacy endpoint and the two rescue therapy related endpoints.	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.
4.3.7	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.
4.3.8	Added a subgroup of iron replete subjects	Subgroup if interest
4.3.9	Deleted section “Model checking”.	Not applicable since the CV analyses will not be performed for the individual CSR.
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

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.

Royston & Finkelstein 2000
Robins J, Finkelstein D. *Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests.* Biometrics 779-788 September 2000

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

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

8. APPENDIX

Not applicable.