A Phase 3, Randomized, Open-Label, Active-Controlled Study to Evaluate the Efficacy and Safety of Roxadustat in the Maintenance Treatment of Anemia in End Stage Renal Disease Subjects on Stable Dialysis

ISN/Protocol 1517-CL-0613

ClinicalTrials.gov Identifier: NCT02278341

Date of SAP: 10 Sep 2018

Sponsor: Astellas Pharma Europe B.V. (APEB)

Sylviusweg 62,

2333 BE Leiden

The Netherlands

STATISTICAL ANALYSIS PLAN

Final Version 3.0, dated 10 September 2018

A Phase 3, Randomized, Open-Label, Active-Controlled Study to Evaluate the Efficacy and Safety of Roxadustat in the Maintenance Treatment of Anemia in End Stage Renal Disease Subjects on Stable Dialysis

ISN: 1517-CL-0613
EudraCT number: 2013-001497-16

Sponsor name
Astellas Pharma Europe B.V. (APEB)
Sylviusweg 62,
2333 BE Leiden
The Netherlands

Table of Contents

I. LIST OF ABBREVIATIONS AND KEY TERMS

List of Abbreviations

Abbreviations	Description of abbreviations
ADaM	Analysis Data Model
AE	Adverse Event
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
AnS	Anemia Subscale
anti-HCV Ab	Anti-hepatitis C Virus Antibody
APEB	Astellas Pharma Europe B.V.
Apo	Apolipoproteins
ASC	Analysis Set Classifications
ASP1517	=FG-4592(roxadustat); codename(name) of investigational product
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
BL	Baseline
BL Hb	Baseline Hemoglobin (please refer to key definitions for information)
BMI	Body Mass Index
BP	Blood Pressure
BUN	Blood Urea Nitrogen
CAPD	Continuous ambulatory peritoneal dialysis
CBC	Complete Blood Count
CHr	Reticulocyte Hemoglobin Content
CKD	Chronic Kidney Disease
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Form
CRO	Contract Research Organization
CRP	C Reactive Protein
CSE	Composite Safety Endpoint
CSR	Clinical Study Report
DBP	Diastolic Blood Pressure
DILI	Drug-induced Liver Injury
dL	Deciliter
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
eCRF	Electronic CRF
EH	Excessive Hematopoiesis
EOS	End of Study
EOT	End of Treatment
EQ-5D 5L	European Quality of Life 5 Domain 5 Level
ESA	Erythropoiesis Stimulating Agent

Abbreviations	Description of abbreviations
ESRD	End Stage Renal Disease
EU	European Union
EudraCT	Clinical trial database regulated by European Community
EWB	Emotional Well being
FACT-An	Functional Assessment of Cancer Therapy-Anemia
FACT-G	Functional Assessment of Cancer Therapy-General
FDA	Food and Drug Administration
FAS	Full Analysis Set
FG-4592	=ASP1517; codename of investigational product (roxadustat)
FWB	Functional Well-being
g	gram
DS	Data Science
GGT	Gamma Glutamyl Transferase
Hb	Hemoglobin
HbA1c	Hemoglobin A1c; Glycated hemoglobin
HBsAG	Hepatitis B Surface Antigen
HCG	Human chorionic gonadotropin
HD	Hemodialysis
HDF	Hemodiafiltration
HDL	High-density Lipoprotein
HEENT	Head, Eyes, Ears, Neck and Throat
HIV	Human Immunodeficiency Virus
HIF	Hypoxia-inducible Factor
HRQoL	Health-Related Quality of Life
hs-CRP	High Sensitivity C-reactive protein
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
ICH E3	Guidance for Industry Structure and Content of Clinical Study Reports
ICH E9	Statistical Principles for Clinical Trials
ICU	Intensive Care Unit
IEC	Independent Ethics Committee
IF	Information Fraction
INR	International Normalized Ratio
IRS	Interactive Response System
ITT	Intention to treat
ISN	International Study Number
IV	Intravenous(ly)
Kg	Kilogram
LA-CRF	Liver Abnormality Case Report Form
LDL	Low-density Lipoprotein
LFT	Liver Function Tests
LLN	Lower Limit of Normal

Abbreviations	Description of abbreviations
LSO	Last Subject Out
MAP	Mean Arterial Pressure
MAR	Missing At Random
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
MI	Multiple Imputation
mL	Milliliters
mmHg	Millimeters of Mercury
MMRM	Mixed Model of Repeated Measures
MNAR	Missing Not At Random
NCI-CTCAE	National Cancer Institute - Common Terminology Criteria for Adverse Events
O	Optional
PCS	Physical Component Score
PD	Protocol Deviation
PD	Peritoneal Dialysis
PEY	Patient-Exposure-Year
PF	Physical Functioning
PGIC	Patients' Global Impression of Change
PK	Pharmacokinetic
PKAS	Pharmacokinetic Analysis Set
PMM	Pattern Mixture Model
PPS	Per Protocol Set
PT	Preferred Term
PWB	Physical Well-being
QoL	Quality of Life
QRS	QRS interval
QTc	QT Interval corrected for heart rate
QTcB	QTc calculated according to Bazett's formula
QTcF	QTc calculated according to Fridericia's formula
RBC	Red Blood Cell
RR	Respiratory Rate
RR Interval	Interval between successive Rs of the ECG
SAE	Serious AE
SAF	Safety Analysis Set
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SBP	Systolic Blood Pressure
SC	Subcutaneous(ly)
SDTM	Study Data Tabulation Model
SI	International System of Units
SF-36	Short Form 36
SF-36 PCS	SF-36 Physical Component Score

Abbreviations	Description of abbreviations
SF-36 PF	SF-36 Physical Functioning
SF-36 MCS	SF-36 Mental Component Score
SF-36 VT	SF-36 Vitality
SmPC	Summary of Product Characteristics
SMQ	Standardized MedDRA Query
SOC	System Organ Class
SWB	Social Well-being
TEAE	Treatment-Emergent Adverse Event
TIBC	Total Iron-Binding Capacity
TIW	Thrice Weekly
TLF	Tables, Listings and Figures
TSAT	Transferrin Saturation
ULN	Upper Limit of Normal
VAS	Visual Analogue Scale
VT	Vitality
WBC	White Blood Cell
WHO-DRL	World Health Organization Drug Reference List
Wk(s)	Week(s)
µg	Microgram

List of Key Terms

Terms	Definition of terms
Adverse Event	An adverse event is as any untoward medical occurrence in a subject administered study drug or has undergone study procedures and which does not necessarily have a causal relationship with this treatment. AE collection starts after obtaining signed informed consent and continues until the End of Study (EOS) visit. SAEs and cardiovascular and thromboembolic AEs will be collected during the post study follow-up period. AEs will not be collected during the period between first screen where subject has failed screening and first rescreening visit.
Baseline	1) Observed values/findings which are regarded as calibrated zero status in the present study; 2) Time when 'Baseline' is observed.
Baseline Hemoglobin **(**Hb**)** value	Baseline Hb is defined as the mean of four central laboratory Hb values: four latest Hb values prior or on the same date as first study drug intake (pre-dose).
Discontinuation	The act of concluding participation in either the study treatment or the study, prior to completion of all protocol-required elements, in a trial by an enrolled subject. Four categories of discontinuation are distinguished: a) dropout: Active discontinuation by a subject (also a noun referring to such a discontinued subject); b) investigator-initiated discontinuation (e.g., for cause); c) loss to follow-up: cessation of participation without notice or action by the subject; d) Sponsor-initiated discontinuation. Note that subject discontinuation does not necessarily imply exclusion of subject data from analysis. "Termination" has a history of synonymous use, but is now considered non-standard.
Efficacy Emergent Period	Defined as the evaluation period from the Analysis date of first dose intake up to EOT Visit.
Endpoint	Event or outcome that can be measured objectively to determine whether the intervention being studied is beneficial. Primary and secondary variables supporting objectives of the study are called endpoints.
Enroll	To register or enter into a clinical trial; transitive and intransitive. Informed consent precedes enrollment, which precedes randomization.
Hb Response	Mean Hb during weeks 28 to 36 within the target range of 10.0 to 12.0 g/dL.
Intervention	The drug, device, therapy or process under investigation in a clinical trial which has an effect on outcome of interest in a study: e.g., health-related quality of life, efficacy, safety, pharmacoeconomics.
Investigational period	Period of time where major interests of protocol objectives are observed, and where the test drug or comparative drug (sometimes without randomization) is usually given to a subject, and continues until the last assessment after completing administration of the test drug or comparative drug.

Terms	Definition of terms
Post study follow-up	Period of time from EOS visit to projected week 108 or until consent withdrawn. This period is only applicable to subjects who prematurely discontinued treatment. These subjects will be followed up at a 6-monthly frequency for vital status, SAEs and cardiovascular and thromboembolic AEs.
Pre-investigational period	Period of time before entering the investigational period, from the time of starting a subject enrolling into study until just before the test drug or comparative drug is given to a subject
Randomization	The process of assigning trial subjects to treatment or control groups using an element of chance to determine assignments in order to reduce bias. After randomization subjects receive either roxadustat or ESA (epoetin alfa or darbepoetin alfa) from day 1 until End Of Treatment (EOT).
Roxadustat	International Nonproprietary Name (INN) of ASP1517/FG-4592 investigational product
Re-screening	Process of repeating screening, *See also screening and screen failure* If subject fails screening they may be re-screened once if deemed appropriate; all screening procedures will be repeated. Renal ultrasound only to be repeated if not within 12 weeks prior to randomization.
Safety Emergent Period	Defined as the evaluation period from the Analysis date of first drug intake up to 28 days after the end of treatment taking into account the different dosing frequencies of the study treatments
Screening	A process of active consideration of potential subjects for enrollment in a trial.
Screening Hb value	Mean of subject's three last Hb values collected during the screening period and prior to the day of randomization.
Screen /Re-screen failure	Potential subject who did not meet one or more criteria required for participation in a trial during screening or re-screening.
Serious Adverse Event	An adverse event is considered "serious" if, in the view of either the investigator or Sponsor, it results in any of the following outcomes: results in death, is life threatening, results in persistent or significant disability/incapacity or substantial disruption of the ability to conduct normal life functions, results in congenital anomaly, or birth defect, requires in-subject hospitalization or leads to prolongation of hospitalization, or a medically important event.
Study period	Period of time from first subject screened to end of the last scheduled visit of the last subject randomized
Subject	An individual who participates in a clinical trial
Time to event	Time from a defined starting point (analysis date of first dose intake) to the time of occurrence of the event of interest.
Time to censoring	Time from a defined starting point (analysis date of first dose intake) to the time of end of observation period in case the event did not occur.

Terms	Definition of terms
Time to event analysis	Time to event analyses are statistical methods, such as survival analysis, that take into account 2 types of timing: the time to occurrence of an event (if an event occurred) and the time to censoring (if an event did not occur during the time we observed the subject). For time to censoring, we only know the total number of days in which the event didn't occur until the subject ceased to be followed (censored).
Treatment period	It is the period of time - between first dose of the test drug and EOT visit - where major interests of protocol objectives are observed, and where roxadustat (study drug) or ESA (comparative drug) is given to a subject.
Variable	Any quantity that varies; any attribute, phenomenon or event that can have different qualitative or quantitative values

1 INTRODUCTION

This Statistical Analysis Plan (SAP) contains a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and includes detailed procedures for executing the statistical analysis of the primary and secondary endpoints and other data.

The SAP is finalized and signed prior to accumulation of a substantial amount of data (open-label study) to ensure lack of bias. For operational efficiency an earlier time is usually targeted and wherever possible, the SAP should be developed in parallel with protocol finalization. If needed, revisions to the approved SAP may be made prior to database hard lock. Revisions will be version controlled.

This statistical analysis is coordinated by the responsible biostatistician of APEB. Any changes from the analyses planned in the SAP will be justified in the Clinical Study Report (CSR).

All details of the Pharmacokinetics Analysis Set (PKAS) will be described in a separate analysis plan, and a separate PKAS modeling report will be written.

This SAP is based on protocol version 2.2, dated 15 April 2016 and is an amendment version of SAP versions 1.0 and 2.0.

Prior to database hard lock, a final review of data and TLFs meeting will be held to allow a review of the clinical trial data and to verify the data that will be used for analysis set classification. If required, consequences for the statistical analysis will be discussed and documented. A meeting to determine analysis set classifications may also be held prior to database hard lock.

2 FLOW CHART AND VISIT SCHEDULE

Flow Chart

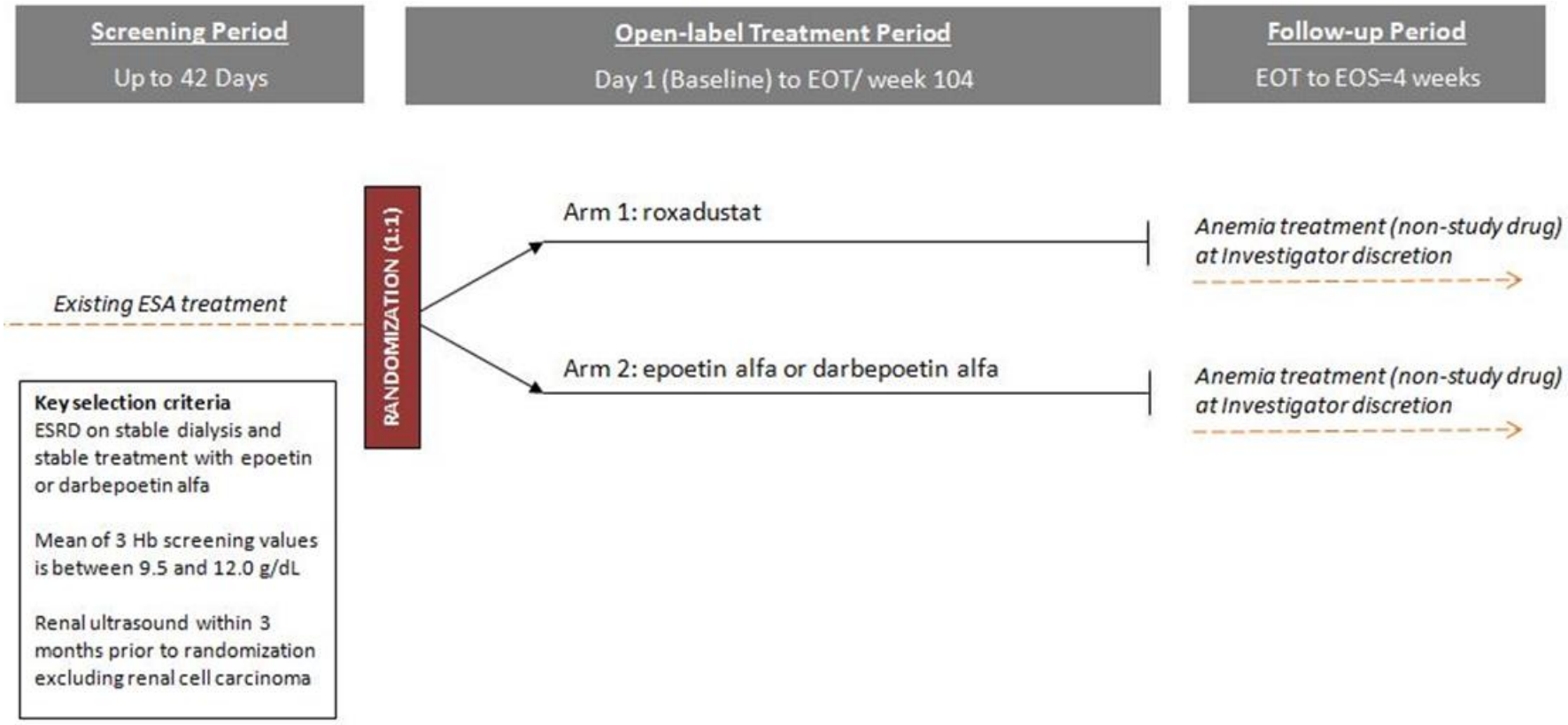


Table 1 Schedule of Assessments

Study Period:	**Screening Period**			**Treatment Period**					**Follow-up Period**		**Unscheduled Visits**	Post study Follow up[c]
Visit/Week:	**2-6 Weeks[a]**			**Day 1[b]**	Weekly (Wks 1 to 8) ± 2 days	Every 2 Weeks (Wks 10 to 36) ± 2 days	Every 4 Weeks (Wks 40 to 100) ± 3 days	EOT[c] (Week 104) ± 3 days	EOT + 2 wks ± 3 days	EOS[c] (EOT + 4 wks) ± 3 days		Every 6 months until projected Wk 108
	S1	**S2**	**S3**									
Dialysis procedure	◄										►	
ESA pre-treatment	◄											
Written informed consent	X											
Randomization				X								
Eligibility Criteria	X			X								
Demographics	X											
Height, weight[d]	X	X	X	X		Wks 12, 24, 36	Wks 52, 76	X		X	O	
Medical History	X											
Physical Examination	X			X		Wks 12[e], 24[e], 36[e]	Wks, 52[e], 76[e]	X		X[e]	O	
Blood pressure[f], heart rate[f], respiratory rate[g]	X	X	X	X	X	X	X	X		X	X	
CBC with WBC diff, red cell indices and platelets	X			X	Wks 1, 2, 3, 4, 8	Wks 12, 20, 28, 36	Wk 44 + every other 8 wks	X		X	O	
Hemoglobin[h]		X	X		X	X	X		X		X	
HemoCue Assessment				X	X	X	X				X	
Reticulocyte count and Hb in reticulocytes (CHr)	X			X	Wks 1, 2, 4, 6, 8	Wks 12, 16, 20, 28, 36	Wk 44 + every other 8 wks	X		X	O	
12-lead ECG	X			X		Wks 12, 24, 36	52, 76	X			O	
Serum Chemistry[i]	X			X	Wks 4, 8	Wks 12, 20, 28, 36	Wk 44 + every other 8 wks	X	X	X	O	
Renal Ultrasound [J]	X										O	
Liver Function Tests (LFTs)[k]					Wks 2, 6	Wk 16					O	

Table continued on next page

Study Period:	**Screening Period**			**Treatment Period**					**Follow-up Period**		**Unscheduled Visits**	Post study Follow up[c]
Visit/Week:	**2-6 Weeks[a]**			**Day 1[b]**	Weekly (Wks 1 to 8) ± 2 days	Every 2 Weeks (Wks 10 to 36) ± 2 days	Every 4 Weeks (Wks 40 to 100) ± 3 days	EOT[c] (Week 104) ± 3 days	EOT + 2 wks ± 3 days	EOS[c] (EOT + 4 wks) ± 3 days		Every 6 months until projected Wk 108
	S1	**S2**	**S3**									
Serum Lipid Panel[l]	X			X	Wks 4, 8	Wks 12, 20, 28, 36	Wks 44, 52, 68, 84	X		X	O	
Serum iron, ferritin, TIBC, TSAT[m]	X			X	Wks 4, 8	Wks 12, 20, 28, 36	Wk 44 + every other 8 wks	X		X	O	
HbA1c	X			X		Wk 12, 28, 36	Wks 44, 52, 60, 84	X		X	O	
Vitamin B_{12}, folate	X										O	
HIV, HBsAg, anti-HCV Antibody	X										O	
Serum Pregnancy Test (HCG)[n]	X					Wks 12, 24, 36	Wks 48, 60, 72, 84, 96	X			O	
high sensitivity C-reactive protein (hs-CRP), hepcidin				X	Wk 4	Wks 12, 20, 36	Wk 52	X		X		
Archival Serum Samples for Biomarkers				X	Wk 4	Wks 12, 20	Wks 52, 76	X		X		
Blood Sample for PK[o]					Wks 2 to 8							
QoL Questionnaires[p]				X	Wk 8	Wks 12, 28, 36	Wks 52, 76	X				
Dialysate[q]					Wks 2 to 8							
Optional genotyping[r]				O								
Study Treatment: roxadustat dispensing OR ESA administration[s]				⟵	———	———	⟶				O	
Dose Adjustment Review[t]					X	X	X				O	
AE and Concomitant Medication Recording	⟵	———	⟶									
Procedure and non-drug Therapy Recording	⟵	———	⟶									
Vital status, SAEs, cardiovascular and thromboembolic AEs												X

Footnotes appear on next page

S1/S2/S3=Screening visit 1, 2 and 3; EOT = End of Treatment; EOS = End of Study; Wk(s) = Week(s); X = mandatory test/assessment; O = optional test/assessment. In HD subjects all lab sampling is to be performed prior to or at initiation of the dialysis session.

Note: Instructions for roxadustat-treated subjects moving from Protocol v1.0 to Protocol v2.0 in terms of dosing are provided in Protocol Appendix 12.3.

[a] Due to the requirement of 4-day separation between the screening Hb values, the screening period will last a minimum of 2 weeks approximately, but is allowed to be a maximum of 6 weeks. Sites are recommended to schedule the three screening visits in the shortest time span possible.

[b] Randomization and administration of the first dose of study drug (roxadustat or ESA) are to occur on Day 1, which should correspond to a day when the subjects' next dose of epoetin or darbepoetin alfa would have been administered. In HD subjects only, post-dialysis weight and post-dialysis vital signs do not have to be completed prior to randomization and first study drug administration.

[c] In case of premature treatment discontinuation, the subject will complete the EOT visit and EOS visit. Thereafter subjects who have taken at least one dose of study drug will continue to be followed up at a 6-monthly frequency for vital status, SAEs and cardiovascular and thromboembolic AEs until their projected date of completion of the Follow-up Period (i.e. projected week 108 date) or until consent withdrawn. No additional study visits will be required during the Post study Follow-up.

[d] Height at first Screening visit only; Weight is post-dialysis weight in HD subjects. For day 1 visit the post-dialysis weight on the third screening visit is to be recorded in the Interactive Response System (IRS) for HD subjects. Weight for PD subjects is the weight measured at site minus the abdominal fluid based on last filling. For drug dispensing purposes (at visits after randomization), weight from the previous dialysis session may be used.

[e] Targeted physical examination (e.g., respiratory and cardiovascular).

[f] Blood pressure and heart rate measured singly during the screening period, and in triplicate at all other visits; in HD subjects both pre- and post-dialysis.

[g] Respiratory rate measured singly during screening period and all other visits; in HD subjects both pre- and post-dialysis.

[h] Separate Hemoglobin should be collected at all the visits where Complete Blood Count (CBC) is not collected; i.e. Hemoglobin at Weeks 5, 6, 7, 10 until the end of the study. It is recommended that the Hb assessments be performed on the same day of the patient's dialysis schedule throughout the study.

[i] Serum chemistry includes LFTs.

[j] Renal ultrasound examination can be performed at any time during the screening period. Not required if results of a previous renal ultrasound (or other renal imaging modality such as CT scan or MRI providing a conclusive report on the kidney) within 12 weeks prior to randomization is available and ruling out renal cell carcinoma.

[k] LFTs to be collected at further visits where Serum Chemistry is not collected.

[l] Fasting, whenever possible.

[m] TSAT = [FeSat (Ferro Saturation = % Iron Saturation) measured by central lab and derived from iron and TIBC].

[n] Human chorionic gonadotropin (HCG). To be collected from female subjects of childbearing potential only.

[o] Only for subjects randomized to roxadustat at 6 time points over 1 to 3 visits. At each PK visit, an additional sample will be collected for albumin and alpha-acidglycoprotein determination.

[p] Includes Short Form 36 (SF-36), Functional Assessment of Cancer Therapy-Anemia (FACT-An), European Quality of Life 5 Domain 5 Level (EQ-5D 5L) and Patients' Global Impression of Change (PGIC). PGIC will not be performed on Day 1. In HD subjects, the quality of life questionnaires are to be completed by the subject prior to any other assessment or at the latest at the start of dialysis session. It is important that the physician does not discuss the status of the subject prior to completion of the questionnaires. In PD subjects the questionnaires are to be completed prior to any study assessment.

Footnotes continued on next page

[q] Only for continuous ambulatory peritoneal dialysis (CAPD) subjects randomized to roxadustat. Sample to be withdrawn from the drainage bag after draining complete volume at the site and at the same visit as the PK blood samples (see footnote o).

[r] Optional assessment and only for subjects randomized to roxadustat. If a separate (optional) informed consent form is signed by the subject, a sample can be collected after randomization, preferably on Day 1.

[s] Dosing of epoetin alfa and darbepoetin alfa according to approved SmPC (Eprex® United Kingdom SmPC; Aranesp® European Union SmPC). It is recommended that roxadustat is administered any time after completion of dialysis (if dosing is scheduled on a dialysis day).

[t] Subjects randomized to roxadustat: dose adjustment review from week 4 onward, and every 4 weeks thereafter until EOT (deviation from the 4-week period is allowed anytime during the study in case of excessive hematopoiesis or Hb ≥13.0 g/dL). If next dose adjustment interval falls on a non-visit study week, the dose adjustment review should be performed at the next scheduled study visit. ESA dose adjustments should follow the SmPC.

3 STUDY OBJECTIVES AND DESIGN

3.1 Study Objectives

3.1.1 Primary Objective

The primary objective of this study is to evaluate the efficacy of roxadustat compared to ESA (epoetin alfa and darbepoetin alfa) in the maintenance treatment of anemia in End Stage Renal Disease (ESRD) subjects on stable dialysis.

3.1.2 Secondary Objectives

The secondary objectives of this study are to:

- Evaluate the safety of roxadustat compared to ESA (epoetin alfa and darbepoetin alfa) in the maintenance treatment of anemia in ESRD subjects on stable dialysis.
- Evaluate the effects on health-related quality of life (HRQoL) of roxadustat compared to ESA (epoetin alfa and darbepoetin alfa) in the maintenance treatment of anemia in ESRD subjects on stable dialysis.

3.2 Study Design

3.2.1 General

This is a phase 3, multi-center, randomized, open-label, active-controlled study. This study is planned to recruit approximately 750 subjects from approximately 150 study centers, mainly located in Europe.

The study is planned to provide key efficacy and safety data for the approval of roxadustat in the treatment of anemia associated with ESRD on stable dialysis. Study FGCL-4592-064 with similar design is conducted by FibroGen Inc, in study centers across North America, Latin America and Asia Pacific.

3.2.2 Study Population

The study population consists of adult subjects with ESRD who are on stable Hemodialysis (HD) or Peritoneal dialysis (PD) and on stable treatment with epoetin (i.e., epoetin alfa, beta, theta, zeta, delta or omega) or darbepoetin alfa for anemia. Subjects on polyethylene glycol-epoetin beta (Mircera®) are not to be included. Subjects on hemodiafiltration (HDF) are also allowed to participate in the study and will follow the same study procedures and requirements as HD subjects. Prior to randomization a subject must be on the same mode of dialysis for ≥4 months and on epoetin or darbepoetin alfa treatment for ≥ 8 weeks with stable average weekly doses (defined as ≤30% change) during the last 4 weeks prior to randomization. Anemia is defined as the mean Hb of the three most recent screening measurements ≥ 9.5 g/dL and ≤ 12.0 g/dL with an absolute difference ≤ 1.3 g/ dL between the highest and the lowest value. Subjects must be iron replete at baseline; inclusion is permitted if ferritin ≥ 100 ng/mL (≥220 pmol/L) and TSAT ≥ 20%. Anemia of non-renal origin is to be excluded. Washout periods of at least 4 weeks for any iron-chelating agents, and at least 8 weeks for any RBC

transfusion prior to randomization have been mandated in order to exclude a potential impact of these recent anemia treatments on the assessment of efficacy.

3.2.3 Description of Study

Subjects assigned to the roxadustat treatment arm will receive roxadustat orally as a combination of tablets of different strengths. Subjects assigned to the ESA treatment arm will be administered epoetin alfa if they were on epoetin (i.e., epoetin alfa, beta, theta, zeta, delta or omega) prior to randomization or darbepoetin alfa in case they were on darbepoetin alfa prior to randomization. Study treatment administration is implemented in an open-label manner. Subjects on polyethylene glycol-epoetin beta (Mircera®) are not to be included.

The study will consist of three study periods:

- Screening Period: up to 6 weeks
- Treatment Period: 104 weeks
- Follow-up Period: 4 weeks

During the course of the study, visits and assessments will be performed as defined in the Schedule of Assessments.

Screening Period

During the Screening Period of up to 42 days, eligibility assessments will be performed. Subjects will continue epoetin or darbepoetin alfa treatment per local standard of care.

Subjects meeting all of the inclusion criteria and none of the exclusion criteria will be randomized at Day 1 (baseline), which marks the end of Screening Period and start of the Treatment Period.

Treatment Period

Subjects are randomized to one of two treatment arms as illustrated in Table 2

Table 2 Treatment Arms, Dosing Frequencies and Patient Numbers

Treatment Arms	Study Treatment	Dosing Frequency	N
1	roxadustat	TIW*	375
2	ESA**	Dosing per SmPC***	375

*TIW = three times weekly; ** ESA = epoetin alfa or darbepoetin alfa

***SmPC = Summary of Product Characteristics

- Treatment arm 1: subjects will be switched from epoetin or darbepoetin alfa treatment to roxadustat treatment (see Table 3 for doses).
- Treatment arm 2: subjects will continue ESA treatment, i.e. epoetin alfa if pre-treated with any epoetin (i.e. epoetin alfa, beta, theta, zeta, delta or omega) and darbepoetin alfa if pre-treated with darbepoetin alfa, at approximately the same average weekly dose the subject was on prior to randomization. It is not allowed to switch from epoetin alfa to darbepoetin alfa or vice versa, during the treatment period.

The initial study drug (roxadustat) dose is based on the dosing scheme shown in Table 3 and is determined by the subject's average prescribed weekly dose of epoetin or darbepoetin alfa within four weeks prior to randomization.

Table 3 Initial Dose of Roxadustat

epoetin* (IU/week)	**darbepoetin alfa*** (µg/week)	**roxadustat** (mg/dose) TIW
<8,000	<40	100
8,000 to 16,000	40-80	150**
>16,000	>80	200***

* Average prescribed weekly dose in the last four weeks prior to randomization.

** If the initial dose of 150 mg exceeds the maximum dose of 3.0 mg/kg, then 100 mg is to be used as the starting dose.

*** If the initial dose of 200 mg exceeds the maximum dose of 3.0 mg/kg, then 150 mg is to be used as the starting dose.

The randomization will result in an overall 1:1 ratio of subjects receiving roxadustat administered orally, or epoetin alfa or darbepoetin alfa administered by SC or IV injection.

Randomization to treatment arms will occur through an Interactive Response System (IRS).

Randomization and administration of the first dose of study treatment (roxadustat or ESA) is to occur on Day 1 which should correspond to a day when their next dose of epoetin or darbepoetin alfa would have been administered. The first dose of study treatment should be taken after all study assessments have been completed; in HD subjects only, study assessments should also be performed prior to or at initiation of dialysis (except for post-dialysis weight and post-dialysis vital signs, which should be measured post dialysis).

During the Treatment Period, subjects will attend weekly study visits from Day 1 to Week 8, followed by every other week study visits from Week 10 to 36. Following Week 36, study visits will occur every four weeks until the End of Treatment (EOT).

During the Treatment Period, the aim is to maintain the Hb levels between 10.0 g/dL and 12.0 g/dL. For both treatment arms dose adjustments are to be made based on Hb values using HemoCue®, a point-of-care device.

All subjects receiving roxadustat will initiate and continue roxadustat dosing at a dosing frequency of TIW throughout the entire period and dose adjustment is to follow pre-specified dose adjustment rules (see protocol, Section 5.1.2.1, Table 4). Intake should occur at approximately the same time of day and, if possible, dosing days should remain consistent throughout the study. For HD subjects it is recommended that roxadustat is administered any time after completion of dialysis if dosing is scheduled on a dialysis day to avoid potential bias on certain study assessments. Specific rules apply on the dialysis day(s) at which PK sampling is scheduled (see protocol, Section 5.6).

Subjects will receive study treatment (roxadustat or ESA) for 104 weeks.

Follow-up Period

After the end of the Treatment Period subjects will proceed to the 4-week Follow-up Period. The choice of anemia treatment during the follow-up period is at the discretion of the investigator and according to local standards.

Post Study Follow-up *(only for subjects prematurely discontinued from treatment)*

Subjects who have prematurely discontinued study treatment will complete the EOT and EOS visits. Thereafter, these subjects who have taken at least one dose of study treatment, will continue to be followed up at a 6-monthly frequency for vital status, SAEs and cardiovascular and thromboembolic AEs until their projected date of completion of the Follow-up Period (i.e. projected week 108) or until consent withdrawn. Data collection during this Post study Follow-up occur through phone calls or information collected from source data of subjects' regular, non-study, visits to the clinic. No additional study visits will be required from these subjects during the Post study Follow-up. The last Post study Follow-up data collection should occur around the projected week 108.

3.2.4 Comparative drugs

Epoetin alfa (Eprex®) or darbepoetin alfa (Aranesp®) will be administered IV or SC according to the Package Insert or Summary of Product Characteristics (United Kingdom [UK] SmPC for Eprex® and European Union [EU] SmPC for Aranesp®). Copies of these SmPCs will be distributed as part of the study materials.

For subjects randomized to ESA epoetin alfa or darbepoetin alfa, the starting dose, frequency and mode of application (IV or SC) are at the investigator's discretion provided the average weekly dose remains approximately the same as prior to randomization.

Epoetin alfa and darbepoetin alfa will be provided as a solution for injection in pre-filled syringes in the following strengths: 10, 20, 30, 40, 60 and 100 µg for darbepoetin alfa and 1000, 2000, 3000, 4000, 6000 and 8000 IU for epoetin alfa.

Dose adjustment and administration rules, as per respective SmPCs, are summarized in the protocol, Section 5.1.

3.2.5 Interim analysis

No interim analysis will be performed.

3.3 Randomization

A randomized design has been chosen in order to ensure a balanced allocation of study subjects to the Treatment Arms and to minimize bias in therapeutic management and assessment of outcomes.

An open-label design was chosen for several reasons. Epoetin alfa and darbepoetin alfa will be administered by SC or IV injection, whereas roxadustat is administered orally, and the investigational and comparator drugs have a different requirement for iron supplementation.

Randomization and treatment assignments will be performed via Interactive Response System (IRS) prepared on behalf of the Sponsor (under the responsibility of the Data Science (DS) Department of APEB). Specific procedures for randomization through the IRS are contained in the study procedures manual.

A total of approximately 750 subjects will be randomized to one of the two Treatment Arms in a 1:1 ratio as follows:

- Treatment Arm 1: roxadustat
- Treatment Arm 2: ESA (epoetin alfa or darbepoetin alfa)

See also Table 2 for the Treatment Arms description.

Randomization will be stratified by the following five factors:

- Previous ESA treatment (epoetin versus darbepoetin alfa)
- Region (region A versus region B)*
- History of cardiovascular, cerebrovascular or thromboembolic diseases (Yes versus No)
- Average prescribed weekly ESA dose four weeks prior to randomization (≤200 IU/kg epoetin or ≤ 1 µg/kg darbepoetin alfa versus >200 IU/kg epoetin or > 1 µg/kg darbepoetin alfa)
- Screening Hb value (≤11.0 g/dL versus >11.0 g/dL)

** The assignment to region (see Section 6.5.2) will be determined based on health care system comparability.*

4 SAMPLE SIZE

The study is sufficiently powered for both regionally-based primary efficacy endpoints. Approximately 750 subjects will be randomized to receive roxadustat or ESA in an open-label fashion in a 1:1 ratio.

EU (EMA)

Results from simulations of the EU primary efficacy endpoint in stable dialysis subjects suggest that dosing levels of ESA prior to randomization may play a role in Hb response. Therefore, the EU primary analysis of this study will be tested both in the overall population and in the subset (subgroup) of patients defined as subjects with an average prescribed weekly epoetin or darbepoetin dose within the last four weeks prior to randomization ≤ 200 IU/kg or ≤ 1 µg/kg respectively, following a parametric chain procedure [Millen et al, 2011; Spiessens et al, 2010]. The information fraction will be calculated at the time of database hard lock; it is defined as the number of subjects in the subset with respect to the number of subjects in the total study population.

Assuming that the Per Protocol Set (PPS) analysis will consist of 80% of the randomized subjects, 750 randomized subjects will lead to approximately 600 subjects in the PPS. The overall one-sided significance level (alpha) is fixed at 0.025 and the non-inferiority margin for the EU primary endpoint has been fixed to -0.75 g/dL. The overall alpha will be equally

allocated to each of the two test populations. If we assume an information fraction of 0.80, this rule will lead to a significance level of 0.0174.

In this setting, three hundred (300) subjects for the roxadustat treatment group and 300 subjects for the ESA treatment group will provide 97% power to statistically demonstrate non-inferiority of roxadustat versus ESA in the EU primary endpoint in both the total study population and the planned subset population analysis assuming a difference (roxadustat minus ESA) of -0.25 g/dL in the Hb change from baseline and an SD of 1.5 g/dL. In the case that these assumptions do not hold across the overall population, the planned procedure can still allow for a successful conclusion for the subset population.

US (FDA)

With 750 randomized subjects (All Randomized), the study will provide at least 99% power to demonstrate statistical non-inferiority of roxadustat versus ESA in the primary endpoint for US submission. The overall one-sided significance level (alpha) is fixed at 0.025 and the non-inferiority margin for the US primary endpoint has been fixed to -0.75 g/dL.

5 ANALYSIS SETS

In accordance with International Conference on Harmonization (ICH) recommendations in guidelines E3 and E9, the analysis sets below will be used for the analyses.

Detailed criteria for analysis sets will be laid out in Classification Specifications (CS) and the allocation of subjects to analysis sets will be determined prior to database hard lock.

5.1 All Randomized

Criterion for exclusion from All Randomized is defined as follows:

- Not randomized

The selection of subjects for the All Randomized will be confirmed in the Analysis Set Classification (ASC) meeting.

The All Randomized will be used for summaries, selected primary and secondary analyses of efficacy endpoints, as well as selected demographic and baseline characteristics.

The All Randomized is regarded as the main analysis set for the US primary endpoint testing non-inferiority.

The All Randomized Set will exclude the 2 randomized subjects from site [redacted], which has been terminated prematurely due to GCP violations including data integrity issues. Consequently, these subjects will not be included in any of the analyses.

5.2 Full Analysis Set (FAS)

The Full Analysis Set (FAS) consists of all randomized subjects (All Randomized) who received at least one dose of study drug and have at least one non-missing post-dose Hb assessment. Subjects will be assigned to their planned treatment provided by the IRS.

Criteria for FAS exclusion is defined as follows:

- No study drug taken, or
- No post-dose Hb value

The selection of subjects for the FAS will be confirmed in the Analysis Set Classification (ASC) meeting.

The FAS will be used for summaries, selected primary and secondary analyses of efficacy endpoints, as well as selected demographic and baseline characteristics.

5.3 Per Protocol Set (PPS)

The Per-Protocol Set (PPS) includes all FAS subjects who do not meet any of the reasons to exclude a complete subject from PPS, listed in Table 4. This PPS will be used for the primary EU endpoint analysis, for selected secondary analyses and for all disposition, demography and baseline characteristics.

Table 4 Criteria for excluding a subject from PPS

Number	Reasons for exclusion from PPS
1	Subject who receives less than 12 weeks of study treatment.
2	Subject without a valid corresponding Hb. 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.
3	Prescribed study drug compliance during treatment < 75% during the first 36 weeks analysis period.
4	Violation of inclusion or exclusion criteria which may affect the assessment of the efficacy of the study drug.
5	Administration of wrong randomization study drug for more than one week during the reference period (first 36 weeks) or until EOT, whatever comes first.
6	Administration of prohibited concomitant medication affecting efficacy listed in Appendix 12.1 of the protocol during the reference period (first 36 weeks) or until EOT, whatever comes first.
7	Administration of rescue therapy significantly deviating from the protocol during the reference period (first 36 weeks) or until EOT, whatever comes first.

The selection of subjects for the PPS will be confirmed in the ASC meeting. More details will be provided in the ASC specifications.

5.4 Safety Analysis Set (SAF)

The safety analysis set consists of all randomized subjects (All Randomized) who received at least one dose of study drug. Subjects will be assigned to their actual treatment received during the trial.

The SAF will be used to describe demographic and baseline characteristics and all safety and tolerability related variables.

5.5 Pharmacokinetics Analysis Set (PKAS)

The PKAS will be defined in a separate analysis plan. Results of the population PK analysis will not be reported in the Clinical Study Report but in a separate population PK report.

5.6 Pharmacodynamic Analysis Set (PDAS)

Not applicable in this study.

6 ANALYSIS VARIABLES

6.1 Efficacy Endpoints

The **Efficacy Emergent Period** will be defined as the evaluation period from the Analysis date of first dose intake up to EOT Visit. This period will be used as reference period for the time to event analyses related to efficacy endpoints, unless specified otherwise. More details on the derivation of the date of End of Efficacy Emergent Period are provided in Section 7.11.6

6.1.1 Primary Efficacy Endpoint

There are two separate regionally based primary efficacy endpoints in this study, depending upon whether the data are being filed to support submission to the EU EMA or to Ex- EU health authorities, such as the US FDA.

6.1.1.1 Primary Efficacy Endpoint for EU (EMA)

The primary efficacy endpoint for EU is Hb change from baseline (BL) to the average Hb of weeks 28 to 36, without having received rescue therapy (i.e., RBC transfusion for all subjects and ESA for roxadustat subjects) within 6 weeks prior to and during this 8-week evaluation period.

For the average in weeks 28-36, all available Hb values obtained from the central laboratory will be used (i.e., both scheduled and unscheduled Hb values).

For the analyses that require an Hb value by visit, the last available Hb value in the visit window will be kept.

In case a subject does not have any available Hb value during this week 28-36 period (because they died or withdrew from the study for example), imputation rules will be applied (refer to Section 7.11.1 for imputation rules).

In case a subject received rescue therapy within 6 weeks prior to week 28-36 period, i.e., in weeks 22-36, Hb value will be set to missing for 6 weeks after they stopped rescue and will be imputed similarly.

Baseline Hb is defined as the mean of four central laboratory Hb values: four latest Hb values prior or on the same date as first study drug intake (pre-dose).

6.1.1.2 Primary Efficacy Endpoint for US (FDA)

The primary efficacy endpoint for US is Hb change from BL to the average Hb of weeks 28 to 52 regardless of rescue therapy.

For the average in weeks 28-52, all available Hb values obtained from the central laboratory will be used (i.e., both scheduled and unscheduled Hb values).

For the analyses that require an Hb value by visit, the last available Hb value in the visit window will be kept.

In case a subject does not have any available Hb value during this week 28-52 period (because they died or withdrew from the study for example), imputation rules will be applied (refer to Section 7.11.1 for imputation rules).

Baseline Hb is defined as in Section 6.1.1.1

6.1.2 Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints in this study are listed in Table 5

Table 5 Key Secondary Efficacy Endpoints

Number	Endpoint
1	Hb response defined as mean Hb during weeks 28 to 36 within the target range of 10.0 to 12.0 g/dL without having received rescue therapy within 6 weeks prior to and during this 8-week evaluation period.
2	Change from BL in Low Density Lipoprotein (LDL) cholesterol to the average LDL cholesterol of weeks 12 to 28.
3	Mean monthly IV iron use (mg) during day 1 to week 36 (monthly defined as a period of 4 weeks).
4	Change from BL in SF-36 Physical Functioning (PF) sub-score to the average PF sub-score of weeks 12 to 28.
5	Change from BL in SF-36 Vitality (VT) sub-score to the average VT sub-score of weeks 12 to 28.
6	Change from BL in mean arterial pressure (MAP) to the average MAP value of weeks 20 to 28.
7	Time to first occurrence of increase in blood pressure: An increase from BL of ≥ 20 mm Hg systolic blood pressure (SBP) and SBP ≥ 170 mmHg or an increase from baseline of ≥15 mmHg diastolic blood pressure (DBP) and DBP ≥ 100 mmHg during weeks 1 to 36.

6.1.2.1 Hb response defined as mean Hb during weeks 28 to 36 within the target range of 10.0 to 12.0 g/dL without having received rescue therapy within 6 weeks prior to and during this 8-week evaluation period.

The average Hb value from weeks 28 to 36 weeks will be calculated as defined below. The Hb response will then be defined as a binary variable, Hb response (Yes/No), where Hb response="Yes" is defined as:

- Average Hb ≥10.0 g/dL, and
- Average Hb ≤ 12.0 g/dL.

For this calculation, all available Hb values in weeks 28-36 will be used, i.e. both scheduled and unscheduled Hb values from the central laboratory.

In case a subject does not have any available Hb value in weeks 28-36 (if they died or withdrew from the study for example), imputation rules will be used (refer to Section 7.11.1 for imputation rules).

In case a subject requires rescue therapy within 6 weeks prior to weeks 28-36, in weeks 22 to 36, values will be set to missing for 6 weeks after they stopped rescue and will be imputed similarly.

6.1.2.2 Change from BL in Low Density Lipoprotein (LDL) cholesterol to the average LDL cholesterol of weeks 12 to 28.

The analysis will be done on all values (fasted and non-fasted) of Day 1 and up to week 28.

All available LDL values will be used (regardless the fasting status), i.e. both scheduled and unscheduled LDL values. LDL values in analysis visit windows at weeks 12, 20 and 28 will be selected for the calculation of the average LDL cholesterol of weeks 12-28 (see Table 34 and Section 7.11.4 for the analysis windows definition and the differentiation between the MMRM and ANCOVA analyses).

For missing LDL imputation rules, refer to Section 7.11.1. Baseline LDL is defined as the LDL value on Day 1. If this value is missing, the latest value prior to first study drug administration will be used.

This analysis will also be repeated for fasted values only, as a sensitivity analysis.

6.1.2.3 Mean monthly IV iron use (mg) during day 1 to week 36

The Mean Monthly (defined as a period of 4 weeks or 28 days) IV iron use during day 1 to week 36, per subject (in mg) is defined by the following formula:

$$\frac{Total\ IV\ iron\ use\ (mg)\ from\ \ Analysis\ date\ of\ \ first\ dose\ intake\ to\ Min\ (Analysis\ date\ of\ week\ 36\ visit, End\ of\ Efficacy\ Emergent\ Period)}{(Min(Analysis\ date\ of\ Week\ 36\ \ visit, End\ of\ Efficacy\ Emergent\ Period) - Day\ 1\ date + 1)\ /\ 28}$$

Analysis visits will be used as indicated in Table 32, Section 7.11.4.

The use of IV iron is recorded in the *Concomitant Medication* form of the eCRF, where Route is INTRAVENOUS. All medications are coded with WHO-DD. Records selected will be those coded as IRON PREPARATIONS.

Having received IV iron is a binary variable (Yes/No), where "Yes" is defined as having at least one record selected during the Efficacy Emergent Period. Subjects using IV iron pre-treatment with an end date of Day 1 will not be included. Subjects without a relevant concomitant medication record will be assumed that they used no IV iron, thus set to 0 mg.

Time to first IV iron during day 1 to week 36 will be derived similarly to time to first occurrence of an increase in blood pressure, in Section 6.1.2.7, with first event date corresponding to the Date of first dose of IV iron during the day 1 to week 36 period. The time to censoring for a subject without the event of interest is calculated as:

(Min (Week 36 analysis date, end of efficacy emergent period) – Analysis date of first dose intake + 1) / 365.25

6.1.2.4 Change from BL in SF-36 Physical Functioning (PF) sub-score to the average PF sub-score of weeks 12 to 28.

For details on the SF-36 PF subscore, refer to Section 6.1.3.12.1

All available SF-36 PF values will be used i.e. both scheduled and unscheduled SF-36 PF values. SF-36 PF values in analysis visit windows at weeks 12 and 28 will be selected for the calculation of the average PF sub-score of weeks 12 to 28 (see Table 33 and Section 7.11.4 for the analysis windows definition and the differentiation between the MMRM and ANCOVA analyses).

For missing SF-36 PF, refer to Section 7.11.1 for the imputation rules.

Baseline assessment is the assessment from Day 1 visit.

6.1.2.5 Change from BL in SF-36 Vitality (VT) sub-score to the average VT sub-score of weeks 12 to 28.

For details on the calculation of the SF-36 VT scale subscore, see Section 6.1.3.12.1.

Similar rules as for Section 6.1.2.4 will be used for the calculation of the SF-36 VT scale sub-score.

6.1.2.6 Change from BL in pre-dialysis mean arterial pressure (MAP) to the average MAP value of weeks 20 to 28.

Blood pressure will be measured singly for the three visits during the screening period and in triplicate with a 2-minute interval for all other visits during the study. During the study, for systolic blood pressure (SBP) and diastolic blood pressure (DBP), the average will be calculated for each visit using the three readings. If less than three readings are available, all non-missing readings will be used in the calculation of the average.

For this analysis, only pre-dialysis blood pressure values will be considered.

Pre-dialysis MAP will be derived for each visit from the above averaged pre-dialysis SBP and the pre-dialysis DBP using the following equation:

$$MAP = (2/3) * DBP + (1/3) * SBP$$

All available pre-dialysis MAP values will be used, i.e. both scheduled and unscheduled MAP values. Pre-dialysis MAP values in analysis visit windows at weeks 20, 22, 24, 26 and 28 will be selected for the calculation of the average pre-dialysis MAP during weeks 20-28 (see Table 32 and Section 7.11.4 for the analysis windows definition and the differentiation between the MMRM and ANCOVA analyses).

For missing data imputation rules, refer to Section 7.11.1. Baseline assessment is the assessment on Day 1 (average of the three readings). If the baseline assessment is missing, then the latest available value prior to first drug administration will be used.

6.1.2.7 Occurrence and time to first occurrence of an increase in pre-dialysis blood pressure during weeks 1 to 36.

For this analysis, only pre-dialysis blood pressure values will be considered.

Occurrence of an increase in pre-dialysis blood pressure is a binary variable (Yes/No), defined as:

- Pre-dialysis systolic blood pressure (SBP) increase from BL ≥20 mmHg AND SBP ≥170 mmHg, or
- Pre-dialysis diastolic blood pressure (DBP) increase from BL ≥15 mmHg AND DBP ≥100 mmHg.

at any visit from week 1 to week 36. The date of occurrence is defined as the first date where SBP criterion or DBP criterion is met, whichever occurs first.

At each visit, pre-dialysis SBP and DBP are calculated as the average from the three readings, assessed before dialysis. If less than three readings are available, the non-missing readings will be used in the calculation of the average.

Baseline assessment is the assessment from Day 1 (pre-dialysis). If this value is missing, then the latest available value from the screening period will be used.

Only events starting during the Safety Emergent Period and up to week 36 analysis date will be taken into account.

The time to occurrence of an increase in blood pressure for a subject with the event of interest will be calculated (in years) as:

(First event date – Analysis date of first dose intake + 1) / 365.25

Where 'Analysis date of first dose intake' is defined in Section 6.5.4, and where 'First event date' is the first date of occurrence of increase in blood pressure.

The time to censoring for a subject without the event of interest is calculated as:

(Min (Week 36 analysis date, Date of last vital signs assessment during the week 1-36 period) – Analysis date of first dose intake + 1) / 365.25

Refer to Section 6.2 and 7.11.5 for more details regarding the Safety Emergent Period.

6.1.3 Additional Secondary Efficacy Endpoints

The additional secondary efficacy endpoints are listed in Table 6

Table 6 Additional Secondary Efficacy Endpoints

Number	Endpoint
	Hb Maintenance
1	Hb response during weeks 28 and 36 defined as mean Hb – regardless of use of rescue therapy within the target range of 10.0 to 12.0 g/dL. Additional definitions to be included for sensitivity analyses are mean Hb >=10.0 g/dL during Week 28 and 36, during Week 28 and 52 and mean Hb between 10.0 and 12.0 g/dL during Week 28-52.
2	Hb change from BL to each post-dosing time point.
3	Hb level averaged over weeks 28 to 36, 44 to 52, and 96 to 104 without use of rescue therapy within 6 weeks prior to and during these 8-week evaluation periods.
4	Hb change from BL to the average Hb value of weeks 28 to 36, 44 to 52, and 96 to 104 regardless of the use of rescue therapy.
5	Proportion of Hb values >=10 g/dL and within 10.0 to 12.0 g/dL in weeks 28 to 36, 44 to 52, and 96 to 104 without use of rescue therapy within 6 weeks prior to and during these 8-week evaluation periods.
	Hospitalizations
6	Occurrence (number) of hospitalizations (HD days are not counted as hospitalizations, even when performed overnight), number of days of hospitalization per patient-year exposure and time to first hospitalization.
	Rescue Therapy Use
7	Occurrence and time to first use of rescue therapy [composite of RBC transfusions (all patients) and rescue ESA (roxadustat treated subjects only)], occurrence and time to first use of RBC transfusions, number of RBC packs per subject, volume of RBC transfused per subject.
	Iron Use
8	Occurrence of IV iron supplementation. Mean monthly IV iron (mg) per subject during weeks 37-52 and weeks 53-104 (monthly defined as a period of 4 weeks). Use of oral iron.
	Change in Cholesterol Levels, Apolipoproteins
9	Change from BL to each post-dosing study visit in Total cholesterol, LDL/High-density Lipoprotein (HDL) ratio, Non-HDL cholesterol, Apolipoproteins A1 and B, ApoB/ApoA1 ratio.
10	Occurrence of mean LDL cholesterol <100 mg/dL calculated over weeks 12 to 28.
	Blood Pressure Effect
11	Occurrence of achieved antihypertensive treatment goal in CKD subjects (SBP< 140 mmHg and DBP< 90 mmHg) based on the mean SBP and mean DBP calculated over weeks 12 to 28.
	HRQoL
12	Change from BL to the average value of weeks 12 to 28 (SF-36 Physical Component Score (PCS), Anemia Subscale ("Additional Concerns") of Functional Assessment of Cancer Therapy (FACT-An) Score, Total FACT-An Score, EQ-5D 5L VAS Score).
13	Patient Global Impression of Change (PGIC).
	Hepcidin, Iron status, HbA1c
14	Changes from BL to each study visit (when measured) in Serum hepcidin, Serum ferritin, TSAT and HbA1c level.

6.1.3.1 Hb response regardless of use of rescue therapy.

Hb response is defined as the mean Hb during weeks 28 to 36 within the target range of 10.0 to 12.0 g/dL.The same rules as defined in Section 6.1.2.1 apply for this endpoint, except that is defined regardless of use of rescue therapy (values in the rescue period will be used too).

Additional endpoints to be used are mean Hb ≥ 10.0 g/dL over weeks 28 to 36 and Hb response (defined as mean Hb values between 10.0 to 12.0 g/dL and also defined as mean Hb values >=10.0 g/dL) during weeks 28 to 52.

6.1.3.2 Hb change from BL to each post-dosing time point.

All scheduled and unscheduled hemoglobin values that belong to each window will be taken into account using one value per analysis window, as defined in Table 32 and Section 7.11.4

Baseline Hb is defined in Section 6.1.1.1

At each visit, Hb will also be categorized into the following categories: <10 g/dL, 10-12 g/dL and > 12 g/dL.

6.1.3.3 Hb level averaged over weeks 28 to 36, 44 to 52, and 96 to 104 without use of rescue therapy within 6 weeks prior to and during these 8-week evaluation periods.

All scheduled and unscheduled hemoglobin values that belong to each period will be taken into account for calculating the average using the analysis windows (defined in Table 32, Section 7.11.4).

In addition, the averages over weeks 28-36, 44-52 and 96-104 will be categorized into the following categories:

- <10 g/dL,
- >=10 g/dL,
- 10-12 g/dL,
- >12 g/dL.

In case a subject does not have any available Hb value within this evaluation period, or in case a subject requires rescue therapy within 6 weeks prior to and during this 8-week evaluation period, refer to Section 7.11.1 for imputation rules.

6.1.3.4 Hb change from BL to the average Hb value of weeks 28 to 36, 44 to 52, and 96 to 104 regardless of the use of rescue therapy.

The same rules as defined in Section 6.1.1, apply for these periods except that it is regardless of rescue therapy.

6.1.3.5 Categorical analysis of Hb values

The following endpoints will be analyzed:

- Proportion of Hb values >=10 g/dL and within 10.0-12.0 g/dL by time intervals;

- Percentage of time with Hb values falling in each Hb interval (<10.0 g/dL, >=10 g/dL, within 10.0-12.0 g/dL, > 12.0 g/dL, > 13.0 g/dL and > 14.0 g/dL) during the Efficacy Emergent Period;
- Potential Excessive Hematopoiesis (EH).

Proportion of Hb values

The following proportions in percentage for each subject will be defined:

- Number of Hb values >= 10.0 g/dL / Total number of Hb values*100
- Number of Hb values within 10.0-12.0 g/dL / Total number of Hb values*100

in weeks 28 to 36, 44 to 52 and 96 to 104 without use of rescue therapy within 6 weeks prior to and during this 8 week evaluation period. All scheduled and unscheduled hemoglobin values that belong to each period will be taken into account using the analysis windows defined in Table 32, Section 7.11.4. The following time periods will be defined in addition: 56 to 64, 68 to 76 and 80 to 92.

Percentage of time

The percentage of time each patient has an Hb value <10.0 g/dL, >= 10.0 g/dL, within 10.0-12.0 g/dL, > 12.0 g/dL, > 13.0 g/dL or > 14.0 g/dL will be calculated (as a percentage of the total of the length of time between the first and last Hb assessment during the evaluated period). The percentage of time will be calculated via linear interpolation. That is, if the change in Hb category (for instance from <= 12.0 g/dL to > 12.0 g/dL) occurs between two visits V0 and V1, the day of change will be calculated by:

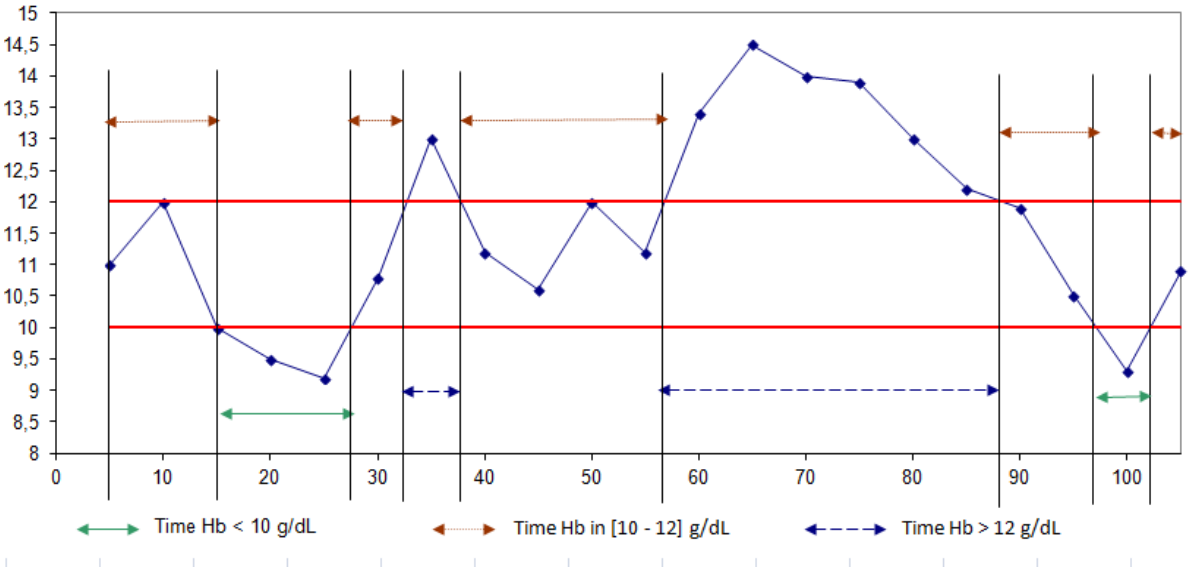
$$x = x_0 + (y - y_0)\frac{(x_1 - x_0)}{(y_1 - y_0)}$$

Where x1 and x0 are the dates when Hb was measured at V0 and V1 respectively,

y_0 and y_1 are the Hb value at the respective visits V0 and V1 and y is the level of the Hb boundary (i.e 12.0, 13.0 or 14.0g/dL).

Figure 1 shows visually how the linear interpolation will calculate the total number of days that a subject is in each Hb category for an example subject:

Figure 1 Example of Linear Extrapolation



In case that several Hb values are on the same day the average of these values will be used to represent the Hb of that day in the above formula. This calculation will provide the day that the change in Hb value occurs. The number of days that the Hb value has been in each category will be determined and the percentage calculated based on the length of time between the first and last Hb assessment during the evaluated period, i.e.:

Date of Last Hb assessment during the evaluated period – Date of first assessment during the evaluated period.

No imputation will be performed if no Hb value is available in relevant time windows.

In case a subject requires rescue therapy within 6 weeks prior to and during these 8-week evaluation periods, refer to Section 7.11.1 for imputation rules.

Potential Excessive Hematopoiesis (EH), regardless use of rescue therapy, based on Hb central laboratory

The presence of potential EH will be defined as:

- Hb increase by >2.0 g/dL between any 2 visits within 4 weeks of treatment

Time to first occurrence of potential EH regardless the use of rescue therapy during the treatment period will be defined in weeks as:

(First event date – Analysis date of first dose + 1) / 7

where 'First event date' is defined as first date of occurrence of the criterion met during the treatment period.

For a subject without potential EH, the time to censoring will be calculated (in weeks) as:

(Min(Date of last hemoglobin assessment during the Efficacy Emergent Period, End Date of the Efficacy Emergent Period) – Analysis date of first dose + 1) / 7

Refer to Section 7.11.6 for the definition of the Efficacy Emergent Period.

6.1.3.6 Occurrence (number) of hospitalizations, number of days of hospitalization per year and time to first hospitalization.

The occurrence and the number of non-HD hospitalizations per subject during the Efficacy Emergent Period will be calculated. HD days are not counted as hospitalizations, even when performed overnight.

The number of days of hospitalization per year treated will be calculated as:

[Sum of the durations of all non-HD hospitalizations in days (Date of discharge – Date of admission + 1)] / (duration of efficacy emergent period in days / 365.25).

In case of missing dates, the hospitalization duration will be imputed by the average duration per stay derived from the subjects with non-missing duration within the same treatment group. Duration of treatment exposure is defined in Section 6.5.4.

If the date of admission of a hospitalization record is the same as the date of discharge of the previous record, for example because two records are created to illustrate that the subject is moved from one hospital to another hospital or from a standard care to the intensive care unit (ICU), then the two records will be combined in one record.

Hospitalizations will also be described by reason for admission (admission for anemia or other reasons).

Time to first hospitalization in years will be defined in years as:

(First event date during the Efficacy Emergent Period – Analysis date of First dose intake + 1)/365.25

With ‘First event date’ defined as ‘Date of first Admission and ‘Analysis Date of first dose intake’ defined in Section 6.5.4.

For a subject without hospitalization, the time to censoring will be calculated as:

(Date of End of Efficacy Emergent Period – Analysis Date of first dose intake + 1) / 365.25

With date of End of Efficacy Emergent Period defined in Section 7.11.6

6.1.3.7 Occurrence and time to first use of rescue therapy, occurrence and time to first use of RBC transfusions, number of RBC packs per subject, volume of RBC transfused per subject during the treatment period.

Rescue therapy is defined as RBC transfusion for all subjects and ESA (SC or IV) for roxadustat-treated subjects.

RBC transfusion is collected and coded into the ATC and WHO-DRL dictionaries. The following WHO-DRL codes will be classified as RBC transfusion: '01186901001'. ESA rescue treatment is recorded on the dosing CRF.

Only rescue medication that started during the study treatment and up to end of efficacy emergent period will be taken into account and considered as use of rescue medication.

For a subject with use of rescue therapy, the time to first use of rescue therapy will be calculated (in years) as:

(First event date – Analysis Date of first dose intake + 1) / 365.25

With 'First event date' defined as 'Date of first dose of rescue medication' during the Efficacy Emergent Period and 'Analysis date of first dose intake' defined in Section 6.5.4

For a subject without use of rescue therapy, the Time to censoring is calculated as:

(Date of End of Efficacy Emergent Period – Analysis date of first dose intake + 1) / 365.25

With Date of End of Efficacy Emergent Period defined in Section 7.11.6

The volume of blood transfused and the total number of RBC units/packs (for each subject, the sum of blood volume and units/packs transfused) during the Efficacy Emergent Period will be calculated.

For RBC transfusions, when the number of units is not given but the volume transfused is given, the number of units will be estimated by volume transfused/250 mL (for transfusion of packed cell units) or volume transfused/500 mL (for transfusion of full blood).

For subjects with use of RBC transfusion, the time to first use of RBC transfusion is calculated as:

(First event date – Analysis date of first dose intake + 1) / 365.25

With 'First event date' defined as 'Date of first RBC transfusion' during the Efficacy Emergent Period and 'Analysis Date of first dose intake' is defined in Section 6.5.4

For a subject without use of RBC transfusion, the Time to censoring is calculated as:

(Date of End of Efficacy Emergent Period – Analysis date of first dose intake + 1) / 365.25

With Date of End of Efficacy Emergent Period defined in Section 7.11.6

The ESA-weeks are defined as the duration of ESA exposure. Each period when ESA [ATC code = B03XA] was taken will be summed by subjects as follows:

sum (each period (min(Analysis Date of last dose, End date of ESA therapy) + X days – Start date of ESA therapy +1)/7);

X will be defined as the duration of the effect of ESA following the last ESA administration for each ESA therapy use period, based on the following rules:

If Frequency = 1 PER WEEK then X=7

If Frequency = 2 PER WEEK then X=3

If Frequency = 3 PER WEEK then X=2

If Frequency = 1 PER MONTH then X=28

If Frequency = ONCE then X=0

If Frequency = BIM (Bi-monthly) then X=14

If Frequency = QD (daily) then X=1

If Frequency = QM (monthly) then X=28

If Frequency =QOD (every other day) then X=2

If Frequency = TID (three times daily) then X=1

If Frequency = 4 TIMES PER WEEK then X=2

If Frequency = EVERY 3 WEEKS then X=21

Additional frequency may be considered depending on the data.

6.1.3.8 Occurrence of iron supplementation.

The use of IV iron is collected in the Concomitant Medication form of the eCRF. All medications are coded with WHO-DD. Records selected will be those coded as IRON PREPARATIONS, and where route is INTRAVENOUS.

Having received IV Iron is a binary variable (Yes/No), where "Yes" is defined as having at least one record selected during the Efficacy Emergent Period. Time to first IV Iron will be derived similarly to time to first use of rescue therapy, in Section 6.1.3.7, with first event date corresponding to the Date of first dose of IV Iron during the efficacy emergent period. Censoring rules will be the same as for use of rescue therapy.

The Efficacy Emergent Period will be divided in periods of 28 days and for each of these periods, the monthly mean of IV iron will be calculated. The total amount of IV iron used will be calculated for the overall Efficacy Emergent Period. Only use of IV Iron that was ongoing or started during the Efficacy Emergent Period will be taken into account.

The Mean Monthly IV iron use per subject (in mg) for weeks 37 to 52 and 53-104 will be calculated similarly as in Section 6.1.2.3. The occurrence of oral iron (coded as IRON PREPARATIONS and route ORAL) use will be also derived by the relevant periods (week 1-36, week 37-52, week 53-104) as a yes or no use of oral iron. This will be combined with IV iron use for these periods.

6.1.3.9 Change from BL to each post-dosing study visit in Total cholesterol, LDL/High-density Lipoprotein (HDL) ratio, Non-HDL cholesterol, Triglycerides, Apolipoproteins A1 and B, ApoB/ApoA1 ratio.

For each sample, the following will be calculated:

- LDL/HDL ratio (LDL Cholesterol divided by HDL Cholesterol)
- Non-HDL cholesterol (Total Cholesterol minus HDL Cholesterol)

Change from baseline to each post-dosing study visit will be calculated for the following lipid parameters:

- Total cholesterol
- LDL cholesterol
- HDL cholesterol
- Low-density lipoprotein (LDL) / high-density lipoprotein (HDL) ratio
- Non-HDL cholesterol
- Triglycerides
- Apolipoproteins A1 and B (ApoA1 and ApoB)
- ApoB/ApoA1 ratio.

All available data will be summarized descriptively for all parameters above, for fasted values and regardless of fasting status.

No imputation will be performed in case of a missing value. If several values are available in the same window, one value will be used, as defined in Table 34 and Section 7.11.4

Baseline assessment is the assessment from Day 1 visit. If this value is missing, then the latest screening period value will be used as baseline.

6.1.3.10 Occurrence of mean LDL cholesterol <100 mg/dL calculated over weeks 12 to 28.

The evaluation period is defined as the average of all available LDL cholesterol values in weeks 12-28 (visit at 12, 20 and 28 weeks, as defined in Table 34 and Section 7.11.4). The occurrence of mean LDL cholesterol <100mg/dL (2.59 mmol/L) over weeks 12 to 28 will then be defined as a binary variable (Yes/No), where ”Yes” is defined as mean LDL cholesterol <100mg/dL (2.59 mmol/L) over weeks 12 to 28.

No imputation will be performed in case of a missing value.

This endpoint will be reported on fasting values and regardless of fasting status.

6.1.3.11 Occurrence of achieved antihypertensive treatment goal in CKD subjects (SBP< 140 mmHg and DBP< 90 mmHg) based on the mean SBP and mean DBP calculated over weeks 12 to 28.

Occurrence of achieved antihypertensive treatment goal (SBP< 140 mmHg and DBP< 90 mmHg) based on the mean SBP and mean DBP will be calculated over an evaluation period defined as the average of all available values in weeks 12 to 28, similarly as in Section 6.1.2.6. The analysis windows is defined in Table 32, Section 7.11.4

Occurrence of achieved antihypertensive treatment goal will then be defined as a binary variable (Yes/No), where “Yes” is defined as SBP< 140 mmHg and DBP< 90 mmHg.

No imputation will be performed in case of a missing value.

6.1.3.12 Change from BL to the average value of weeks 12 to 28 in Quality of Life scores.

All study subjects will be required to complete Quality of Life (QoL) questionnaires as indicated in the schedule of assessments:

- SF-36
- FACT-An
- EQ-5D 5L

In HD subjects, questionnaires will be completed by the subject preferably prior to, or at the start of dialysis session. In Peritoneal Dialysis subjects, it is done preferably prior to any study assessment.

In addition to week 12-28, change from BL to the average value of weeks 36 to 52 will be calculated.

The next sections provide further details on how to derive these instruments, some derivations will be provided by an external vendor (QualityMetric).

6.1.3.12.1 Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36)

The Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) is a multi-purpose, short-form health survey with 36 questions (see Appendix 1). It yields an 8-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures. It is a generic measure, as opposed to one that targets a specific age, disease, or treatment group. Accordingly, the SF-36 has proven useful in surveys of general and specific populations, comparing the relative burden of diseases, and in differentiating the health benefits produced by a wide range of different treatments.

The SF-36 contains 36 items that measure eight dimensions or scales: (1) physical functioning (PF); (2) role limitations due to physical health problems (RP); (3) bodily pain (BP); (4) social functioning (SF); (5) general health perceptions (GH); (6) role limitations due to emotional problems (RE); (7) vitality, energy or fatigue (VT); and (8) mental health (MH) (see Appendix 1). In addition, two summary measures, defined as the Physical Component Score (SF-36 PCS) and Mental Component Score (SF-36 MCS) will be provided.

Scoring of each dimension and the summary measure will be performed by QualyMetric using QualityMetric Health Outcomes(tm) Scoring Software 4.5.

Change from baseline to the average value in weeks 12-28 will be calculated for the Physical Component Scores of SF-36 (SF-36 PCS), following the same rules as defined in Section 6.1.2.4.

For missing SF-36 PCS values, refer to Section 7.11.1 imputation rules. Baseline always refers to the assessment at day 1 to be performed prior to first study drug administration.

The number and percent of subjects with a change from baseline of <3 / >=3 points and of <5 / >=5 points will be calculated for each visit for the following: Vitality Score (SF-36 VT), Physical Functioning score (SF-36 PF) and Physical Component score (SF-36 PCS).

In addition, the eight dimensions and the two summary measures and their associated change from baseline will be reported by visit.

The normalized values will be used for analysis.

6.1.3.12.2 Functional Assessment of Cancer Therapy –Anemia (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, and emotional (EWB) – 6 items.

The 'additional concerns' section contains 20 items: 13 fatigue specific items plus 7 additional items related to anemia were developed for use in conjunction with the FACT-G (Cella 1997). The 13 fatigue items plus the seven additional items related to anemia comprise the Anemia Subscale (AnS). Administration of the FACT-G plus the Anemia Subscale (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. A final higher score indicates better QoL (see Appendix 2).

Each individual item is scored from 0 (Not at all) to 4 (Very much), and then the total score is obtained by summation of the resulted scores.

If there are missing items, subscale scores can be standardized. This is done by multiplying the sum of the subscale by the number of items in the subscale, then dividing by the number of items actually answered. This can be done on the scoring guide or by using the formula below:

Prorated subscale score = [Sum of item scores] x [N of items in subscale] / [N of items answered]

When there are missing data, standardizing by subscale in this way is acceptable (Webster 2003) as long as more than 50% of the items were answered (e.g., a minimum of 4 of 7 items, 4 of 6 items, etc). The total score is then calculated as the sum of the un-weighted subscale scores. The FACT scale is considered to be an acceptable indicator of a subject's quality of life as long as overall item response rate is greater than 80% (e.g., at least 22 of 27 FACT-G items completed). This is not to be confused with individual subscale item response rate, which allows a subscale score to be prorated for missing items if greater than 50% of items are answered. In addition, a total score should only be calculated if ALL of the component subscales have available scores.

The FACT-An instrument will be scored according to Appendix 10.2. The following 8 scores will be calculated:

- PWB subscale score
- SWB subscale score

- EWB subscale score
- FWB subscale score
- AnS subscale score
- FACT-An TOI score
- FACT-G total score
- FACT-An total score
- Fatigue subscale score

Change from baseline to the average value in weeks 12-28 will be reported for the two scores (Anemia subscale 'Additional concerns' of FACT-An score and Total FACT-AN score). In addition, the score and change from baseline will be reported for each visit for all six scores.

Change from baseline to the average value in weeks 12-28 will be calculated for the FACT-An subscore following the same rules as defined in Section 6.1.2.4

No imputation will be performed in case of a missing value.

Baseline always refers to the assessment at day 1 to be performed prior to first study drug administration.

6.1.3.12.3 EQ-5D 5L

The EQ-5D 5L is an international standardized non-disease specific (i.e. generic) instrument for describing and valuing health status, and a multi-dimensional measure of health-related QoL (see Appendix 10.3).

It includes two main components: (1) a VAS scale rating perception of overall health and (2) 5 qualitative domains: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. For more details and description of the questionnaire, refer to Appendix 10.3

Change from baseline to the average value in weeks 12-28 will be calculated for the EQ 5D 5L VAS, following the same rules as defined in Section 6.1.2.4

Frequency distributions will be described for each visit for:

- EQ-5D 5L Mobility Score
- EQ-5D 5L Self-Care Score
- EQ-5D 5L Usual Activities Score
- EQ-5D 5L Pain/Discomfort Score
- EQ-5D 5L Anxiety/Depression Score

The evaluation period for the VAS score is defined as the average of available EQ-5D 5L VAS scores of weeks 12-28 (visit at W12 and W28).

No imputation will be performed in case of a missing item. Baseline always refers to the assessment at day 1 to be performed prior to first study drug administration.

6.1.3.13 Patients' Global Impression of Change (PGIC)

The Patients' Global Impression of Change (PGIC) is a subject-rated instrument that measures change in subjects' overall status since the start of the study on a 7-point qualitative scale ranging from 1 (very much improved) to 7 (very much worse) (see Appendix 10.4).

Data will be reported qualitatively by assessment as follows:

- Reported subject status,
- Combined Categories as binary:
  - Very Much Improved + Much Improved (yes/no)
  - Very Much Improved + Much Improved + Minimally Improved (yes/no)

No imputations will be performed in case of a missing item, except that the score at the last post-baseline assessment will be derived too.

6.1.3.14 Changes from BL to each study visit (when measured) in Serum hepcidin, Serum ferritin, TSAT and HbA1c level.

Changes from baseline to each study visit (see analysis windows in Table 32) will be calculated for these parameters:

- Serum hepcidin,
- Serum ferritin,
- TSAT,
- Serum Iron,
- Total Iron Binding Capacity (TIBC).
- HbA1c level (Subjects with and without diabetes in history, and overall),

For all variables listed above, baseline assessment is the assessment from Day 1 visit. If this value is missing, then the screening period value, if collected for that parameter, will be used.

6.1.4 Other exploratory variables: hs-CRP (High Sensitivity C-Reactive Protein) and Post-dialysis BP

The variable hs-CRP will be collected from the central laboratory on the following visits: Day 1, weeks 4, 12, 20, 36, 52, EOT and EOS. Absolute values and changes from baseline to each study visit will be calculated. Baseline assessment is the assessment from Day 1 visit. If Day 1 assessment is missing, change from baseline will not be reported. Analysis windows are defined in Table 32

Post-dialysis BP will be used to perform additional exploratory analyses for BP-related endpoints: 1) MAP and 2) Time to first occurrence of increase in blood pressure.

6.2 Safety Variables

Safety will be assessed by evaluation of the following variables:

- Treatment-emergent adverse events (TEAEs; frequency, severity, seriousness, and relationship to study drug), and pre-specified adjudicated cardiovascular and cerebrovascular events (will be reported separately),

- Vital signs (systolic and diastolic blood pressure, pulse, respiratory rate and weight),
- Clinical laboratory variables (hematology, biochemistry including liver enzymes and total Bilirubin),
- Physical examination,
- 12-lead electrocardiogram (ECG),
- Vascular Access Thrombosis.

The **Safety Emergent Period** will be defined as the evaluation period from the Analysis date of first drug intake up to 28 days after the Analysis Date of Last Dose taking into account the dosing frequency. Refer to Section 7.11.5 for more details on the derivation of the date of End of the Safety Emergent Period. This period will also be used to identify the minimum or maximum values collected on-treatment, defined as values collected from Day 2 up to the end of the Safety Emergent Period.

6.2.1 Adverse Events

6.2.1.1 Treatment emergent adverse event (TEAE)

TEAE is defined as an adverse event observed after starting administration of the test drug/comparative drug. If the adverse event occurs on Day 1 and the onset check box is marked “Onset after first dose of study drug” or the onset check box is left blank, then the adverse event will be considered treatment emergent. If the adverse event occurs on Day 1 and the onset check box is marked “Onset before first dose of study drug”, then the adverse event will not be considered treatment emergent. If a subject experiences an event both during the pre-investigational period and during the investigational period, the event will be considered as TEAE only if it has worsened in severity (i.e., it is reported with a new start date). All adverse events collected during the Safety Emergent Period will be counted as TEAE.

A drug-related TEAE is defined as any TEAE with at least possible relationship to study treatment as assessed by the investigator or with missing assessment of the causal relationship.

Severity of AEs will be graded according to National Cancer Institute - Common Terminology Criteria for Adverse Events (NCI-CTCAE) Version 4.0.

For AE onset date imputation rules, refer to Section 7.11.2

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

6.2.1.2 Standardized MedDRA Queries

No standardized MedDRA Queries (SMQs) will be performed.

6.2.1.3 Time to occurrence of a TEAE (by type of AE group)

TEAEs are classified into a number of groups depending on the following factors:

- Serious TEAEs
- Death during the Safety Emergent Period

- All deaths during study period and post study safety follow up period
- Related Serious TEAEs
- TEAEs Leading to permanent Discontinuation of study drug
- TEAEs NCI CTC(AE) Grade 3 or Higher
- MedDRA System Organ Class (SOC)

The time to occurrence for a subject with a TEAE for a given type (e.g., serious AE leading to discontinuation) will be calculated (in years) as:

(First TEAE date of the given type during the Safety Emergent Period – Analysis date of first dose intake + 1) / 365.25

With 'Analysis date of first dose' defined in Section 6.5.4. All adverse events collected during the Safety Emergent Period will be counted as TEAE, irrespective of use of rescue therapy.

Subjects who have not experienced a TEAE for that given type will be censored; the subject will be censored and the time to censoring for these subjects will be calculated (in years) as:

(Date of End of Safety Emergent Period – Analysis date of first dose intake +1) / 365.25

With Date of End of Safety Emergent Period defined in Section 7.11.5

For any deaths during the 24-month period (including those occurring during the Post-Study Follow Up Period), time to occurrence for a subject who died during study / post study follow up) will be calculated (in years) as:

[End date for corresponding Fatal AE – Analysis date of first dose intake + 1] / 365.25

With 'End date for corresponding Fatal AE' which cover both study and post-study FU occurring up to Day 760 (i.e., month 24 + one month follow up + 3 days window as per protocol).

Subjects who have not died; the subject will be censored and the time to censoring for these subjects will be calculated (in years) as:

(Minimum between [Day 760 and Max (Date of last known date when subject alive (End of Post-study FU), Date of last contact (Post study FU visit/call), Date of last Study evaluation (EOS form))] – Analysis date of first dose intake +1) / 365.25

6.2.1.4 Definition of incidence rate

The incidence rate (per 100 subject years at risk) will be calculated as follows:

$$\frac{\text{Number of subjects with event}}{\text{Total cumulative time at risk (years)}} \times 100$$

Where Total cumulative time at risk is the sum of individual time at risk defined as either time to occurrence of the event or time to censoring for subjects with no event.

Time to occurrence of the event and time to censoring are defined in Section 6.2.1.3

6.2.1.5 Definition of event rate

The event rate (number of events per 100 subject years) will be calculated as follows:

$$\frac{\text{Number of treatment - emergent events}}{\text{Total patient years (years)}} \times 100$$

Where Total patient years is the sum of individual time during the safety emerging period expressed in years.

6.2.1.6 Adverse events up to 7 days after last dose

Additional analyses restricted to any treatment-emergent adverse event starting up to Analysis Date of Last Dose (adjusted for dosing frequency) + 7 days will be conducted.

Time to occurrence of event will be calculated (in years) as:

(First event date of the given type occurring from Day 1 (post dose) up to Analysis date of Last Dose + X + 7 days) – Analysis date of first dose intake + 1) / 365.25, where X days is defined depending on the dosing frequency.

With 'Analysis date of first dose intake' Analysis Date of Last Dose defined in Section 6.5.4 and 7.11.5. Subjects who have not experienced an AE for that given type will be censored and the time to censoring for these subjects will be calculated (in years) as:

Minimum [Analysis Date of Last Dose + X + 7 days, *Max* (EOS, Date of Death)] – Analysis date of first dose intake +1) / 365.25.

6.2.2 Vital Signs

The following endpoints will be assessed, pre and post dialysis:

- Systolic blood pressure (SBP)
- Diastolic blood pressure (DBP)
- Pulse

In addition, the following assessments will be done:

- Respiratory rate
- Weight

Single measurements for blood pressure (BP) will be taken at three visits during the screening period. Measurements will be taken in triplicate with 2-minute intervals for all other visits. An average will be calculated from the three readings, the average in the eCRF system will not be used. If less than three readings are available, the non-missing readings will be used in the calculation of the average.

The position and date for the assessment will also be recorded. Change from baseline will be calculated as the measurement taken at the specific visit minus the measurement at baseline visit.

For all vital parameters, the minimum and the maximum post-baseline value during the Safety Emergent Period will be defined. For this calculation, only values from day 2 up to the date of End of the Safety Emergent Period (see Section 7.11.5) will be used.

Baseline assessment is the assessment from day 1 visit. If day 1 assessment is missing, screening period assessment will be used in the analysis. For missing visits, the last observation will be carried forward.

For vital signs, if there are pre-dialysis and post-dialysis assessments, the pre-dialysis assessment will be used for classification of the analysis visit (see Section 7.11.4) and the definition of potentially clinical vital signs criteria will use pre-dialysis and post-dialysis assessments separately.

Vital signs values are potentially clinically significant (PCS) if they meet both the observed value criteria and the change from baseline criteria listed in Table 7 (10 combined criteria).

Table 7 Potentially Clinically Significant (PCS) Vital signs Criteria

Vital Sign Parameter	Flag	Criteria	
		Observed Values	Change from Baseline
Respiratory Rate (breaths per min)	High	≥ 20	Increase of ≥ 5
	Low	≤ 10	Decrease of ≥ 5
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 (beats per min)	High	≥ 120	Increase of ≥ 20
	Low	≤ 50	Decrease of ≥ 20
Weight (kg)	High	-	Increase of ≥ 10%
	Low	-	Decrease of ≥ 10%

Potentially Clinically Significant Vital Signs Criteria will be calculated at each study visit and on-treatment (at any moment during the Safety Emergent Period) using the worst value among all available measurements.

Time to occurrence of PCS Vital signs:

For each potentially clinically significant vital signs criterion (i.e 10 combined criteria), the time to occurrence of a PCS at any moment during the Safety Emergent Period (see Section 6.2) will be calculated (in years) as:

(First occurrence date – Analysis date of first dose intake + 1) / 365.25

With ‘First occurrence date‘ defined as the first date when both criteria (i.e., on observed and change from baseline) are met and Analysis date of first dose intake is defined in Section 6.5.4

Subjects without abnormality will be censored and time to censoring for these subjects will be calculated (in years) as:

(Date of last vital signs assessment where the parameter analyzed is non-missing during the Safety Emergent Period – Analysis date of first dose intake +1) / 365.25

6.2.3 Clinical laboratory variables

6.2.3.1 Potentially Clinically Significant (PCS) Laboratory Criteria

Laboratory test values are potentially clinically significant (PCS) if they meet either the low or high PCS criteria listed in Table 8 below.

Table 8 Potentially Clinically Significant (PCS) Laboratory Criteria

Potentially Clinically Significant Laboratory Criteria			
Laboratory Parameter	**Unit**	**Low PCS Criteria**	**High PCS Criteria**
Alanine Aminotransferase (ALT)	U/L	no lower limit	> 3X ULN > 5X ULN[#] > 8X ULN[#] > 10X ULN[#] > 20X ULN[#]
Aspartate Aminotransferase (AST)	U/L	no lower limit	> 3X ULN > 5X ULN[#] > 8X ULN[#] > 10X ULN[#] > 20X ULN[#]
Alkaline Phosphatase (ALP)	U/L	no lower limit	> 1.5 X ULN[#] > 3 X ULN[#]
ALT or AST > 8× ULN	U/L for ALT or AST	no lower limit	ALT or AST > 8X ULN
Total Bilirubin	µmol/L	no lower limit	> 1.5 X ULN > 2 X ULN[#]
Moderate Liver Abnormality[*]	U/L for ALT and AST, µmol/L for Total Bilirubin	no lower limit	ALT and/or AST > 3X ULN **or** Total Bilirubin > 2X ULN
Severe Liver Abnormality[*]	U/L for ALT and AST, µmol/L for Total Bilirubin	no lower limit	ALT and/or AST > 3X ULN **and** Total Bilirubin > 2X ULN
Gamma Glutamine Transaminase (GGT)	U/L	no lower limit	> 3X ULN
Calcium	mmo1/L	< 0.8 X LLN	>1.2 X ULN
Creatinine	µmo1/L		>1.5 X Baseline[$]
Potassium	mmo1/L	< 0.75 X LLN	>1.2 X ULN
Sodium	mmo1/L	< 0.9 X LNL	>1.1 X ULN
Total Protein	g/L	< 0.9 X LNL	>1.1 X ULN
Blood Urea Nitrogen (BUN)	mmo1/L		>1.5 X Baseline[$]
Neutrophils	10^6/L	≤1000	
Platelet Count	10^9/L	≤100	≥700
White Blood Cell Count	10^9/L	≤2.5	≥15
Lipase	U/L	no lower limit	> 3X ULN **or** > 2 X Baseline[$]
LLN: Lower limit of normal, value provided by the laboratory ULN: Upper limit of normal, value provided by the laboratory			

Footnotes appear on next page

* a subject's ALT and Total Bilirubin laboratory draw date or AST and Total Bilirubin laboratory draw date must occur on the same blood sample in order to be counted.

[#] Additional criteria required for summary of Liver Function Tests only (see Section 7.5.2.1).

[$] Criteria of > nX baseline only applicable for post-baseline assessments.

Potentially Clinically Significant Laboratory Criteria will be calculated at each study visit and on-treatment (see Section 6.2) using the worst value among all available measurements, except for Moderate and Severe Liver Abnormalities which will be calculated on-treatment only.

Time to occurrence of an abnormality (for selected criteria)

Time to occurrence of an abnormality, will be derived only for the following PCS criteria:

- Alanine Aminotransferase (ALT) > 3X ULN
- Aspartate Aminotransferase (AST) > 3X ULN
- Total Bilirubin > 1.5 X ULN

For each above potentially clinically significant laboratory criteria, the time to occurrence of an abnormality for a subject with an abnormality at any moment during the Safety Emergent Period (see Section 6.2) will be calculated (in years) as defined in Section 6.2.2

Time to censoring will be defined (in years) as:

(Date of last laboratory assessment where parameter analyzed is non-missing during the Safety Emergent Period – Analysis date of first dose intake +1) / 365.25

6.2.3.2 Laboratory assessments

For all laboratory parameters, the minimum and the maximum values on treatment (see Section 6.2) will be defined.

In addition, each laboratory result will be classified as low (L), normal (N), or high (H) at each visit according to the laboratory supplied reference ranges.

Change from baseline will be calculated as the measurement taken at the specific visit minus the measurement at baseline visit.

Baseline is defined as the assessment on Day 1 (prior to dosing), except for Hb (see Section 6.1.1.1). If Day 1 is missing, the screening or unscheduled assessment (prior to dosing) that is closest to Day 1 will be used.

Screening is defined as the screening or unscheduled assessment (during screening) that is closest to Day 1.

For the lipid panel and glucose parameter, two baseline values will be defined based on fasting status: regardless of fasting and fasted.

6.2.4 Physical Examination

A comprehensive physical examination will be conducted during the screening period and at the EOT visit and recorded in the source documents. This examination will include general appearance and the following body regions and systems: head, eyes, ears, neck and throat

(HEENT), lungs, heart, chest and back, abdomen, genitourinary, extremities, skin and any other, if deemed necessary.

A targeted examination (e.g., respiratory and cardiovascular) will be conducted and recorded in the source documents.

Only the date of the physical examination will be recorded in the eCRF. There will be no table or listing for the physical examination. Any clinically relevant adverse change will be recorded as an AE in the eCRF.

6.2.5 12-lead Electrocardiogram (ECG)

The 12-lead ECG measurements will be performed on all subjects at specific times. A single ECG measurement will be taken with the subject in the supine position, after the subject has been lying quietly for 5 minutes. Clinically significant abnormalities will be reported as an AE.

The visit, ECG date, Pulse, RR Interval, PR interval, QRS Interval, QT Interval, overall interpretation and relevant comments will be recorded in the eCRF.

Baseline assessment is the assessment from day 1 visit. If day 1 assessment is missing, screening period assessment will be used in the analysis.

QTc interval will be calculated using both

- Bazett ($QTcB = QT/(RR\ Interval)^{1/2}$) and
- Fridericia ($QTcF = QT/(RR\ Interval)^{1/3}$) corrections,

where QT is in msec and RR Interval in seconds; and if RR Interval is not available, it will be replaced with 60/HR.

For all ECG parameters, the maximum post-baseline value on treatment will be defined (see Section 6.2).

The QTc interval will be categorized at baseline and at each treatment visit using the range criteria defined in Table 9 below. This will be done for both Bazett and Fridericia correction formulas. The QTc interval changes from baseline will be categorized at each visit using the criteria defined in Table 9 below:

ECG values are potentially clinically significant (PCS) if they meet or exceed the upper limit values listed in Table 9 below.

Table 9 QTc Interval Classification

ECG Parameter	Classification
QTc interval (msec)	> 450 msec, > 480 msec, > 500 msec
QTc interval change (msec)	> 30 msec and > 60 msec

QRS and PR interval will also be categorized at baseline and at each treatment visit as defined in Table 10 below:

Table 10 Other ECG Parameters Classification

ECG Parameter	Classification
QRS (msec)	≥ 150 msec
PR (msec)	≥ 250 msec

Time to occurrence of PCS ECG:

For the two QTc criteria (QTc > 500 msec; change from baseline in QTc > 60 msec), the time to occurrence (in years) for a subject with occurrence of the PCS at any moment during Safety Emergent Period (defined in Section 6.2) will be calculated (in years) as defined in Section 6.2.2

Time to censoring will be defined as:

(Date of last ECG assessment where parameter analyzed is non-missing during the Safety Emergent Period – Analysis date of first dose intake +1) / 365.25

6.2.6 Vascular Access Thrombosis (VAT)

For each AE marked as VAT, additional information regarding the event will be listed.

Time to first VAT will be derived similarly to time to first occurrence of a TEAE, in Section 6.2.1.3, with first event date corresponding to the Date of first VAT. Censoring rules will be the same as without an event of interest.

6.3 Pharmacokinetic Variables

All details of the population PK analysis will be described in a separate analysis plan.

6.4 Pharmacodynamic Variables

Not applicable.

6.5 Other Variables

6.5.1 Eligibility criteria

Eligibility at screening will be recorded as a yes/no variable for each criterion. The date of the informed consent for the subjects will also be documented.

6.5.2 Demographic and Baseline Characteristic Variables

Demographic characteristics will be recorded at screening (sex, the day, month and year of birth, age, race, height and weight).

Collection of date of birth depends on local regulations. Day of birth will be recorded in the eCRF as the first of the month when the day is not allowed to be collected. In cases where only year of birth is allowed to be collected, day and month will be recorded in the eCRF as the first of January. Age will be recalculated in SDTM and ADaM datasets.

Age at baseline is calculated as Age= $(D_{First}-D_B)/365.25$, with D_B as the Date of Birth and D_{First} as the Date of First Dose intake. If the Date of First Dose intake is not available, the Date of Informed Consent will be used.

Based on recalculated age, three categories will be defined:

- < 65 years,
- 65 - 74 years,
- ≥ 75 years

Each subject's body mass index (BMI) will be calculated as:

$$BMI\ (Kg/m^2) = Weight\ (Kg)/\ [Height\ (m)]^2,$$

in which the height will be converted from cm. into m. by dividing by 100.

Tobacco history and use will be recorded at screening.

The average maximum quantity of tobacco per week will be calculated using the average maximum quantity and frequency filled in the CRF. If the frequency is "Day", the average maximum quantity of tobacco per week will be determined as follows:

average maximum quantity per day x 7

If the frequency is "/Month", the average maximum quantity of tobacco per week will be determined as follows:

average maximum quantity per month / 4.3482

Screening Hb will be defined as the mean of the three latest central laboratory Hb values prior to the day of randomization.

Baseline Hb is defined in Section 6.1.1.1

Based on the mean screening Hb value, two categories will be defined:

- ≤ 11.0 g/dL
- > 11.0 g/dL

History of cardiovascular, cerebrovascular or thromboembolic diseases at baseline will be defined for subjects with any of the diseases that have been reported in the Cardiovascular Disease History eCRF. History will be categorized as:

- Yes
- No

Countries and Regions

Subjects will be enrolled from the following 17 countries:

Region A (Western Europe)

- Belgium
- France

- Germany
- Italy
- Portugal
- Spain
- UK

Region B (Central and Eastern Europe)

- Bulgaria
- Croatia
- Hungary
- Poland
- Romania
- Russia
- Serbia
- Czech Republic
- Slovakia
- Georgia

Randomization will be stratified by region using two categories:

- Region A: Western Europe
- Region B: Central and Eastern Europe.

Time to Treatment Discontinuation

Time to Treatment Discontinuation in years is defined as:

Time to Treatment Discontinuation (years) = ('Date of Treatment Discontinuation'-'Analysis date of first dose intake'+1)/365.25

In case a subject completed the treatment period, time to censoring will be calculated as:

(EoT visit – Analysis date of first dose intake + 1) / 365.25

Time in years from Diagnosis of Anemia

Time from diagnosis of anemia in years is defined as:

Time from Diagnosis of Anemia (years) = ('Analysis date of first dose intake '-'Date of Diagnosis')/365.25

In case of partial dates, imputation rules apply and are detailed in Section 7.11.2

Time in years from Diagnosis of CKD

Time from diagnosis of CKD in years is defined as:

Time from Diagnosis of CKD (years) = ('Analysis date of First Dose intake '-'Date of Diagnosis of CKD')/365.25

In case of partial dates, imputation rules apply and are detailed in Section 7.11.2

Time from Diagnosis of Targeted Medical History

Onset date and the start date of analysis for the study drug are collected and the time from diagnosis of targeted medical history in years is defined as

Time from Diagnosis of Targeted Medical History (years) =

('Analysis date of first dose intake'– 'Onset Date')/365.25

In case of partial dates, imputation rules apply and are detailed in Section 7.11.2

This will be calculated for each patient who was diagnosed with the targeted medical history: hypertension, diabetes mellitus type 1, type 2 and combined, dyslipidemia and vascular access.

Iron Repletion at Screening

Subjects will be classified in one of the following groups according to the TSAT and ferritin levels collected at Screening (prior to first drug intake):

- Ferritin <100 ng/mL and TSAT <20%
- Ferritin < 100 ng/mL and TSAT >=20%
- Ferritin >=100 ng/mL and TSAT <20%
- Ferritin >=100 ng/mL and TSAT >=20%
- Other

Iron Repletion at Baseline

Subjects will be classified in one of the following groups according to the TSAT and ferritin levels collected on Day 1:

- Ferritin <100 ng/mL and TSAT <20%
- Ferritin < 100 ng/mL and TSAT >=20%
- Ferritin >=100 ng/mL and TSAT <20%
- Ferritin >=100 ng/mL and TSAT >=20%
- Other

If no Day 1 assessment is available, Iron Repletion at Screening will be used.

Average weekly ESA dose before randomization.

All subjects randomized will be classified into two groups:

- Subjects with lower average weekly ESA dose prior to randomization (≤200 IU/kg epoetin or ≤ 1 µg/kg darbepoetin alfa), versus
- Subjects with higher average weekly ESA dose prior to randomization (>200 IU/kg epoetin or > 1 µg/kg darbepoetin alfa).

It will be derived using data in 'Treatment History of Anemia' eCRF. The methodology to derive the dose is described in Section 6.5.4

Dialysis type

Dialysis type is recorded in the eCRF *Cumulative* form: *Dialysis method*. All subjects will be classified into two groups (if two types of dialysis are reported, the latest one prior to first dosing will be used):

- Hemodialysis subjects (including hemodiafiltration), versus
- Peritoneal dialysis subjects (continuous ambulatory and automated)

Previous ESA treatment.

In the form *Screening Procedures - Treatment history for Anemia and ESA Treatment history for Anemia,* if the variable 'Medication type'='ESA' and 'Specify type of ESA'='Darbepoetin Alfa' then the variable 'Previous ESA Treatment will be coded as 'Darbepoetin Alfa'.

For all other values, as there are several epoetin types, it will take the value 'Epoetin'.

6.5.3 Previous and concomitant medication

Previous medication is defined as a medication with at least one dose taken before the date of first dose of study drug.

Concomitant medication is defined as a medication with at least one dose taken between the date of first dose (inclusive) and end of safety emergent period.

Previous and concomitant drug use will be recorded, including non-prescription medication, complementary and alternative medications. Handling of missing date information for prior or concomitant medications is given in Section 7.11.2

If the medication start date and end date are both missing, the medication will be counted as previous and concomitant.

If the medication start date is missing and the end date is prior the date of first drug administration, the medication will be counted as previous medication.

If the medication start date is missing and the end date is after the date of first drug administration, the medication will be counted as previous and concomitant medication.

Renal transplant status (Yes or No), and the date of renal transplant (if applicable) will be derived. Similarly, ESA use after study treatment (Yes or No), and the start date of ESA use after study treatment (if applicable) will be derived.

6.5.4 Variables related to study drugs

Randomization/Treatment Arms

Table 11 below presents the groups to which subjects are randomized.

Table 11 Treatment arms

Randomization Arm Code (ARMCD)	Randomization Arm (ARM)
A	Roxadustat - TIW
B	ESA (Darbepoetin Alfa or Epoetin Alfa)

Analysis date of First Dose Intake

Date of first study drug dose intake is collected in the Day 1 visit in the Randomization eCRF. In case of a missing/partial date, the earliest available date will be used. It will be on the same day than the randomization date and before the next dose date.

Analysis Date of Last Dose

Date of Last Study Drug Dose is collected at the End of Treatment visit in the End of Treatment eCRF. When this date is not known, all efforts should be made to obtain the date from the investigator. If it is still missing after all efforts have been made, then the visit date of the End of Treatment Visit will be used as the Analysis Date of Last Dose. If subject is lost to follow-up and none of these dates are available, then the date of the last available assessment during the study will be used.

Duration of exposure in days

For each subject, the Length of Time on treatment will be calculated in days, using the following formula:

('Analysis Date of Last Dose' - 'Analysis Date first dose intake') + 1

Amount of Prescribed (planned) Medication

Roxadustat

The number of milligrams prescribed at each visit (including unscheduled visits) is captured in the *Changes in Dosing* eCRF. The investigator reported dose and frequency will be used when available. For first administration, the IRS reported dose and frequency will be used.

The prescribed weekly dose at each visit (including unscheduled visits) will be calculated as prescribed dose x 3 as the prescribed frequency is TIW for roxadustat.

Each visit will have an associated start and end date as follows:

- Each visit (including unscheduled) will have an associated start date,
- Each visit (including unscheduled) will have an associated end date. This date will be the date of the next consecutive visit [including unscheduled visits and EOT] minus 1 day.

Time periods of interest are defined below (monthly is defined as a period of 4 weeks or 28 days) as follows:

Time Period	Analysis Start Day	Analysis End Day*
Week 4 (Month 1)	Day 1	Day 28
Week 8 (Month 2)	Day 29	Day 56
Week 12 (Month 3)	Day 57	Day 84
Week 16 (Month 4)	Day 85	Day 112
Week 20 (Month 5)	Day 113	Day 140
Etc….		
Week 104 (Month 26)	Day 701	Day 728
Treatment Period	Day 1	Day of EOT
Day 1 - Week 28	Day 1	End Day of Month 7 (day 196)
Day 1 – Week 36	Day 1	End Day of Month 9 (day 252)

*or EOT whichever is first

This will allow us to calculate the amount of prescribed medication in a time period as the sum of the daily prescribed amount within the time windows defined above.

In addition, amount of prescribed medication in mg/kg will be calculated. To convert a dose given in mg into a dose in mg/kg, the body weight recorded at day 1 will be used.

Darbepoetin alfa and epoetin alfa

The same methodology will be used for darbepoetin alfa and epoetin alfa. Note that prescribed frequency for darbepoetin alfa is once weekly, once every other week, or once every four weeks, and the prescribed frequency for epoetin alfa is: once weekly, twice weekly, three times weekly, or once every other week. The unit for darbepoetin alfa is microgram (µg), the unit for epoetin alfa is IU.

In addition, amount of prescribed medication in µg/kg and IU/kg will be calculated. To convert a dose given in µg into a dose in µg/kg or in IU into a dose in IU/kg, the body weight recorded at day 1 will be used.

Amount of Consumed Medication

Roxadustat

The consumed Investigational Product Medication is captured in the *Study Drug Roxadustat Accountability* eCRF which includes the following kit information:

- Kit strength, kit treatment, kit dispensed, date of kit dispensed, kit strength total number of tablets dispensed
- Kit returned, returned date, total number of tablets returned

For each kit, the following will be calculated:

- Start Day of Exposure for each kit: study day kit dispensed
- End Day of Exposure for each kit: study day kit returned – 1 Day
- Amount dispensed for each kit: kit strength x number of tablets dispensed
- Amount returned for each kit: kit strength x number of tablets returned

- Amount consumed for each kit: amount dispensed – amount returned
- Daily consumed dose for each kit: amount consumed/(end day of exposure-start day of exposure +1)

The same methodology as described for Amount Prescribed (planned) Medication will apply to calculate amount of consumed medication for each time period by summing up the different daily consumed amount of the different kits on a given day (subjects will be dispensed more than one kit on a given visit).

In addition, amount of consumed medication in mg/kg will be calculated.

Darbepoetin alfa and epoetin alfa

The same methodology will be used for darbepoetin alfa and epoetin alfa. Note that instead of tablets dispensed/returned, there will be syringes dispensed/administered. Administration date is not captured, therefore the start and end day of exposure will be defined as follows:

- Start Day of Exposure for each kit: study day kit dispensed
- End Day of Exposure for each kit: for kits administered, the next consecutive date that the ESA was dispensed – 1 Day

In addition, amount of consumed medication in µg/kg (darbepoetin alfa) and IU/kg (epoetin alfa) will be calculated.

Compliance

Compliance will be calculated for the time periods defined in Amount of Prescribed (planned) Medication. Compliance in % will be calculated for each time period (not reported cumulatively) as:

$$\frac{\text{Amount consumed during time period}}{\text{Amount prescribed during time period}} \times 100$$

The following compliance categories will be defined:

- less than 50% (significant drug noncompliance)
- at least 50%, less than 75% (moderate drug noncompliance)
- at least 75% (acceptable compliance)
- unknown

Dose Changes for roxadustat, darbepoetin alfa and epoetin alfa

Dosing changes are collected in the Study Drug Roxadustat - Dosing Decisions eCRF for roxadustat, in the Study Drug Darbepoetin - Dosing Decisions eCRF for darbepoetin alfa and in the Study Drug Epoetin alfa- Dosing Decisions eCRF for epoetin alfa.

For example, for roxadustat, a change from 200 TIW to 250 TIW is a change of 600 mg to 750 mg per week.

The same applies for darbepoetin alfa and epoetin alfa, but not weekly as it could be administered weekly, bi-weekly, monthly, or other frequencies.

For each subject the total number of dose changes, dose increases, dose decreases, and dose holds will be calculated.

7 STATISTICAL METHODOLOGY

7.1 General Considerations

- All statistical comparisons will be made using two sided tests at the α=0.05 significance level unless specifically stated otherwise. Null hypotheses for superiority testing will be of no treatment difference and corresponding alternative hypothesis will be two-sided. Null hypotheses for non-inferiority testing will be of inferiority of roxadustat treatment and will be one-sided at the α=0.025.
- All data processing, summarization, and analyses will be performed using SAS® Version 9.3 (SAS Enterprise Guide 4.3) or higher. Specifications for tables, data listings and figures (TLFs) formats can be found in the TLF Specifications for this study.
- All data will be summarized by treatment arm (roxadustat and ESA), type of ESA treatment (darbepoetin alfa and epoetin alfa) and for the total, unless specified otherwise. For all results split by type of ESA treatment, previous ESA treatment before randomization will be separated in two categories (darbepoetin alfa or epoetin), unless specified otherwise.
- For continuous variables that are recorded as "< X" or "> X", the value of "X" will be used in the calculation of summary statistics. The original values will be used for the listings.
- All percentages will be rounded to one decimal place and lined up by the decimal place. The percentage will be suppressed when the count is zero.
- For continuous variables, descriptive statistics will include the number of subjects (n), mean, standard deviation, median, minimum and maximum. Frequencies and percentages will be displayed for categorical data. Percentages by categories will be based on the number of subjects with no missing data, i.e. will add up to 100%. Number of missing values will be shown in the frequency tables
- All data included in summary tables, inferential analyses or figures will also be listed.
- Listings will be done on all randomized subjects and all assessments (all collected data in the eCRF will be listed except the physical examination data).
- Pre-dialysis vital signs including blood pressure measurement will be used for analysis, unless otherwise specified.
- For the SAF population analyses, the actual treatment will be used. For all other analysis populations, the planned treatment will be used.

For the definition of subgroups of interest please refer to Section 7.8.

7.2 Study Population

For this section, unless specified otherwise, PPS refers to the analysis set which excludes from the FAS all the subjects violating criteria at subject level from Table 4. See Section 5.3 and Classification specifications for more details.

7.2.1 Disposition of Subjects

The following subject data will be summarized and presented:

- Number and percentage of subjects with informed consent, who discontinued before randomization and randomized (overall only),
- Number and percentage of subjects randomized in each analysis set, by treatment arm, type of ESA treatment and overall,
- Number and percentage of subjects who completed and discontinued treatment, by primary reason for treatment discontinuation, by treatment arm and type of ESA treatment for randomized subjects, SAF, FAS and PPS;
- Number and percentage of subjects who completed and discontinued the study, by primary reason for study discontinuation, by treatment arm and type of ESA treatment, for randomized subjects, SAF, FAS and PPS;
- Number and percentage of subjects who completed the post study safety follow-up period, for all subjects who prematurely discontinued the study;
- Number and percentage of subjects for post study follow-up status, for all randomized subjects within 2 years; and
- Number and percentage of subjects excluded from PPS by reason for exclusion defined in Section 5.3, by treatment arm and type of ESA treatment, for the FAS.

The following data will be presented graphically by treatment arm and type of ESA treatment for the SAF and PPS:

- Treatment discontinuation by reason using bar chart;
- Treatment discontinuation by time interval and reason using bar chart; and
- Time to treatment discontinuation using a Kaplan-Meier plot.

In addition, the following graphs will be done to explore missing patterns and support efficacy analyses, by treatment arm, for the All Randomized and PPS:

- Treatment discontinuation for lack of efficacy, using a cumulative incidence plot;
- Treatment discontinuation for adverse event, using a cumulative incidence plot;
- Treatment discontinuation for withdrawal by subject, using a cumulative incidence plot.

Time intervals will be analyzed using the following categories:

- Less than 2 weeks
- At least 2 weeks, less than 4 weeks
- At least 4 weeks, less than 12 weeks
- At least 12 weeks, less than 24 weeks
- At least 24 weeks, less than 36 weeks
- At least 36 weeks, less than 52 weeks
- At least 52 weeks, less than 78 weeks
- At least 78 weeks, less than 104 weeks
- 104 weeks or more
- Unknown.

In addition, the randomization stratification strata from both sources (CRF and IRS) will be reported by treatment arm. Discrepancy between stratification from CRF and IRS will be summarized and the total number of patients with discrepancy overall and for each stratification factor will be provided.

Subjects who had a screen failure will be listed.

All other available data will also be listed.

All data collected during the Post-Study Follow up period will be listed by visit (i.e., type of contact, subject status and occurrence of overnight hospitalizations).

7.2.2 Protocol Deviations

Protocol deviations, as defined in the study protocol (Section 8.1.6: Protocol Deviations) will be assessed for all randomized subjects. The number and percentage of subjects meeting any criteria will be summarized for each criterion and overall, by treatment arm, type of ESA treatment and overall, as well as by study site. Subjects deviating from a criterion more than once will be counted once for the corresponding criterion. Subjects who have more than one protocol deviation will be counted once in the overall summary. A data listing will be provided by site and subject.

The protocol deviation criteria will be uniquely identified in the summary table and listing. The unique identifiers will be as follows:

- PD1 - Entered into the study even though they did not satisfy entry criteria,
- PD2 - Developed withdrawal criteria during the study and was not withdrawn,
    - PD2_1 Withdrawal of consent
    - PD2_2 Pregnancy
    - PD2_3 Significant non compliance
    - PD2_4 ESA rescue therapy (roxadustat subjects)
    - PD2_5 Organ transplant
- PD3 – Received wrong treatment or incorrect dose
    - PD3_1- Received wrong treatment kit
    - PD3_2- Received incorrect dose
    - PD3_3- Study drug non compliance
- PD4 – Received excluded concomitant treatment
    - PD4_1- Received prohibited medication
    - PD4_2- Received restricted medication

7.2.3 Demographic and Other Baseline Characteristics

Demographic and other baseline/screening characteristics will be summarized by descriptive statistics and frequency tabulations.

Number and percentage of subjects randomized in each country and site will be presented for the SAF, FAS, All Randomized and PPS.

Descriptive statistics for age, weight, body mass index (BMI) and height at screening will be presented. Frequency tabulations for sex and race will be presented. Descriptive statistics and frequency tabulations will also be presented for the subgroup variables presented in Section 7.8. Additionally, demographic and other baseline characteristics will be presented for the following variables:

- Average screening Hb value as continuous and categorical (≤ 11.0 g/dL versus >11.0 g/dL),
- Baseline Hb as continuous,
- Iron repletion at baseline, as a categorical value (see Section 6.5.2).

Demographic and baseline characteristics summaries above will be done for the All Randomized, SAF, FAS and PPS. This table will be repeated for subjects randomized after amendment 2.0 (or any further amendment) is implemented.

Selected demographics collected at screening will be done on Screen Failure subjects.

All Medical History will be analyzed using the SAF, as described below:

Medical History including anemia, chronic kidney disease (CKD), cardiovascular disease and targeted medical history are coded in MedDRA, they will be summarized by System Organ Class (SOC) and Preferred Term (PT).

Anemia history, CKD history, targeted medical history (diabetes type I and II, hypertension and dyslipidemia), cardiovascular disease history, tobacco history and family history of cardiovascular disease will be summarized.

The number and proportion of subjects with each typical symptom for CKD will be described, as well as the number and proportion of subjects with each typical symptom for anemia.

Demographic and baseline data will also be listed.

7.2.4 Previous and Concomitant Medications

Previous and concomitant medications are coded with WHO-DD, and will be summarized by therapeutic subgroup (ATC 2nd level), chemical subgroup (ATC 4th level) and preferred WHO name for the SAF.

Subjects taking the same medication multiple times will be counted once per medication.

Treatment history for anemia will be summarized separately.

Missing dates' imputation rules are detailed in Section 7.11.2

7.3 Study Drugs

Roxadustat (=Investigational Product Medicine) is for oral administration and supplied as red coated oval tablets of 20, 50 and 100 mg.

Darbepoetin alfa (=Comparative Medicine) is administered by SC or IV injection, and supplied as a solution for injection in a pre-filled syringe of 10, 20, 30, 40, 60 and 100 µg.

Epoetin alfa (=Comparative Medicine) is administered by SC or IV injection, and supplied as a solution for injection in a pre-filled syringe of 1000, 2000, 3000, 4000, 6000 and 8000 IU.

7.3.1 Exposure

The following information on drug exposure will be presented by treatment arm and type of ESA treatment, for the SAF.

Exposure related variables are defined in Section 6.5.4

Descriptive statistics will be produced for:

- The average weekly dose (roxadustat, darbepoetin alfa and epoetin alfa) the subject was exposed to during the treatment period (in mg, µg or IU and in mg/kg, µg/kg or IU/kg), by month, during the first 28, 36, and 52 weeks; and overall.
- Number and percentage of subjects with dose changes, dose increases, dose decreases or dose hold.
- Number of times for dose changes, dose increases, decreases or dose hold.

Duration of exposure will be summarized in two ways:

- Descriptive statistics and frequency tabulations will be presented;
- Exposure time will be categorized according to the following categories:
    - Less than 2 weeks
    - At least 2 weeks, less than 4 weeks
    - At least 4 weeks, less than 12 weeks
    - At least 12 weeks, less than 24 weeks
    - At least 24 weeks, less than 36 weeks
    - At least 36 weeks, less than 52 weeks
    - At least 52 weeks, less than 78 weeks
    - At least 78 weeks, less than 104 weeks
    - 104 weeks or more
    - Unknown.

Counts and percentages of subjects in each of these categories will be summarized.

For roxadustat, box-plots of average weekly dose and average weekly dose/kg by month will be produced. Separate box-plots for darbepoetin alfa will be produced, using a different unit (µg instead of mg), and separate box-plots will be produced for epoetin alfa too, using IU instead of mg. In the box-plots, the prescribed amount will be used.

Study drug medication will also be listed showing for each subject and each visit the dispensed kit numbers and the actual medication in each kit. For instance, if a subject randomized to ESA, at one visit was mistakenly dispensed a kit containing roxadustat then the listing will show that roxadustat was given to the subject on the intended visit.

7.3.2 Treatment Compliance

Overall compliance with the dosing schedule will be examined for subjects in the SAF whose total study drug count and first and last days of treatment are known.

Percent overall compliance will be summarized in two ways, during the first 28 weeks, 36 weeks, 52 weeks and overall:

- Descriptive statistics will be presented,
- Percent compliance categories will be categorized according to the categories defined in Section 6.5.4

Counts and percentages of subjects in each of these categories will be summarized.

Results will be displayed by treatment arm and type of ESA treatment.

7.4 Analysis of Efficacy

For all continuous efficacy variables, in addition to inferential analyses, descriptive statistics will be produced for the actual values and for the changes from baseline (BL) by visit.

Similarly, for all categorical efficacy variables, frequencies and proportions will be produced by analysis visit.

Quantitative endpoints with repeated measures over time will be analyzed using MMRM and ANCOVA with MI, as the preferred methods for imputation of missing data.

For this section, unless specified otherwise, PPS refers to the analysis set which excludes from the FAS all the subjects violating criteria at subject level from Table 4. See Section 5.3 and Classification specifications for more details.

Unless otherwise specified, the efficacy analyses will be performed for All subjects, and by Type of ESA treatment (Darbepoetin Subjects and Epoetin Subjects).

For efficacy analyses by treatment arm (Roxadustat versus ESA), the covariates list will include the variable Previous ESA treatment (epoetin versus darbepoetin alfa). But for efficacy analyses by Type of ESA treatment (Subgroup Darbepoetin and Subgroup Epoetin), because the Type of ESA treatment is based on the Previous ESA treatment, then this variable Previous ESA treatment will not be included in the covariates list.

For the efficacy analyses by subgroup, the subgroup analysis by Previous ESA treatment will not be presented in tables, as this analysis is already performed by Type of ESA treatment (Subgroup Darbepoetin and Subgroup Epoetin).

Analysis visits windows are detailed in Section 7.11.4.

Missing data imputation rules are detailed in Section 7.11.1

7.4.1 Analysis of Primary Endpoint(s)

There are two separate regionally based primary efficacy endpoints in this study, depending upon whether the data are being filed to support submission to the US FDA or to Ex-US health authorities, such as the European Medicines Agency (EMA).

7.4.1.1 EU (EMA) Primary Endpoint

7.4.1.1.1 Primary Analysis of the EU (EMA) Primary Endpoint

The primary EU (EMA) efficacy endpoint will be analyzed using the PPS.

The change from baseline to the average Hb of weeks 28 to 36, without having received rescue therapy within 6 weeks prior to and during this 8-week period, will be computed from the Mixed Model of Repeated Measures method (MMRM), adjusting for stratification factors (covariates defined below), comparing roxadustat to ESA (pooled darbepoetin alfa and epoetin alfa). After fitting the data, a computation statement will be added to the model to calculate the average Hb values estimate during the period under consideration.

Difference of least square means (roxadustat minus ESA) and its 2-sided 100*(1-α*2)% confidence interval will be estimated for the change from baseline to the average of weeks 28 to 36. The significance level α is fixed by the parametric chain procedure explained below.

The primary analysis of the EU (EMA) primary efficacy endpoint will be tested both in the overall population and in the subset of patients defined as subjects with an average prescribed weekly epoetin or darbepoetin dose within the last four weeks prior to randomization ≤ 200 IU/kg or ≤ 1 µg/kg respectively, following a parametric chain procedure.

The two null hypotheses to be tested are:

- H_{A0}: Hb change from baseline to the average of weeks 28 to 36 in the roxadustat arm ≤ Hb change from baseline to the average of weeks 28 to 36 in the ESA arm minus 0.75 g/dL in the total study population
- H_{B0}: Hb change from baseline to the average of weeks 28 to 36 in the roxadustat arm ≤ Hb change from baseline to the average of weeks 28 to 36 in the ESA arm minus 0.75 g/dL in the subset population

The alternative hypotheses H_{A1} and H_{B1} are defined as the negation of the null hypotheses H_{A0} and H_{B0} respectively. The information fraction (subjects in the subset population/total study population) will be calculated at the time of database hardlock.

The overall one-sided significance level (alpha) is fixed at 0.025. The overall alpha will be equally allocated to each of the two null hypotheses defined above. In case of an information fraction of 0.80, this rule will lead to a significance level of 0.0174.

The following table (Table 12) lists a selection of one-sided significance levels (alpha) to be considered should the IF varies from the protocol.

Table 12 Information Fraction (IF) and One Sided Significance Level (alpha) for the Primary Endpoint

Information Fraction (IF)	alpha (1-sided)
0.5	0.0147
0.6	0.0154
0.7	0.0162
0.8	0.0174
0.9	0.0191
0.97	0.0216

Following the parametric chain procedure, if at least one of the two null hypotheses is rejected at the significance level of 0.0174, the other one can be tested at a significance level of 0.025 while still controlling the overall type I error. The study will be successful for the EU primary efficacy endpoint if either of the null hypotheses is rejected, i.e., if the lower limit of the two-sided 100*(1-α*2)% CI for the difference in least square means from the MMRM model is greater than -0.75g/dL.

If both null hypotheses are rejected, secondary endpoints will be assessed on the overall population. Otherwise, secondary endpoints will not be formally tested.

The covariates will be:

- Previous ESA treatment (epoetin versus darbepoetin alfa);
- Region (region A versus region B);
- History of cardiovascular, cerebrovascular or thromboembolic diseases (Yes versus No)[#];
- Average prescribed weekly ESA dose prior to randomization (≤200 IU/kg epoetin or ≤ 1 μg/kg darbepoetin alfa versus >200 IU/kg epoetin or > 1 μg/kg darbepoetin alfa)[*];
- Baseline Hb value (Continuous).

[#]: This covariate is derived according to a list of cardiovascular disease preferred terms received from the study physician after softlock.

[*]: For the analyses, the average prescribed weekly ESA dose prior to randomization will be taken from the screening procedure form, in the eCRF. Upon review of data, this covariate was excluded from analysis as number of subjects in the high dose group was too low for the variable to be meaningful.

For these covariates, the IRS value used at randomization will be compared to the eCRF value.

In case of stratification errors during the randomization process, the analysis will use the correct covariates as recorded on the eCRF for the subject.

MMRM model:

An MMRM model will be run for the purpose of implicit imputation of missing data by using all the available information from the observed data via the within-patient correlation

structure. The analysis will be based on the estimated difference between the two treatment arms overall mean effects throughout the evaluation period (weeks 28 to 36) based on this MMRM model.

The model will contain treatment arm, region, CV History, previous ESA treatment, visits and visit by treatment as categorical variables. It will also contain baseline Hb and baseline Hb by visit as continuous variable.

The unstructured covariance pattern model will be used. If the algorithm for unstructured covariance pattern does not converge, then heterogeneous Toeplitz structure will be used. If this later model does not converge either, then the (homogeneous) Toeplitz structure will be tried. Finally, if none of them converge, first order autoregressive (AR (1)) as a covariance structure will be used to achieve convergence.

A similar model as the general example below will be used:

$$c_{ikjn} = intercept + \beta_n M_{baseline,ikjn} + \tau_i + \alpha_n + (\alpha\tau)_{in} + \gamma_k + \varepsilon_{ikjn}$$

where

- c_{ikjn} is each analysis visit change from baseline of subject j in treatment arm i, and stratum k at time n,
- β_n is the slope of c_{ikjn} at visit n as a function of the baseline Hb,
- $M_{baseline,ikjn}$ is the baseline measurement of subject j in treatment arm i and stratum k at time n,
- τ_i , is the mean effect of treatment arm i,
- α_n is the mean effect at time n,
- $(\alpha\tau)_{in}$ is the interaction term between treatment arm i and time n,
- γ_k is the mean effect of stratum k,
- ε_{ikjn} is the residual at time n for subject j in treatment arm i and stratum k.

The SAS procedure will be similar to the following:

```
proc mixed;
 class subject_id treatment previous_ESA_treatment region cv_history
      visit;
 model change = treatment previous_ESA_treatment region cv_history
      baseline_Hb visit treatment*visit;
 repeated visit /subject = subject id type=un;
 lsmeans visit*treatment / cl alpha = 0.05;
 estimate 'Roxadustat v.s. ESA averaged at weeks 28-36'
      treatment 1 -1
      treatment*visit 0 0 0 0 .. 0.2 0.2 0.2 0.2 0.2 -0.2 -0.2 -0.2
 -0.2  / cl;
 where visit in ('Week1','Week2',...,'Week8','Week10',...,'Week36');
run;
```

One analysis Hb value for each visit will be used, as defined in analysis windows in Section 7.11.4 and Table 32.

Subjects receiving rescue therapy treatment within 6 weeks prior to and during this 8-week period will be set to missing and imputed similarly in the model (see Section 7.11.1 for imputation details).

In addition, MMRM least square means and 2-sided 95% confidence intervals will be calculated for each visit for the difference in treatment arms. The results will be presented as least square means and 95% confidence intervals for treatment differences, and also present the nominal p-values for non-inferiority and superiority. MMRM least square means and their 95% confidence intervals will be plotted versus time.

Model checking:

Residual plots will be produced showing the following:

- Model residuals in the y-axis and model predictions in the x-axis,
- Model residuals in the y-axis and baseline Hb in the x-axis.

Different dot styles will be used for the two treatment arms. Solid black symbols will be used for roxadustat and non-solid red symbols will be used for ESA.

Residual plots will be done for the MMRM analysis. The plot will be repeated by visit at Weeks 28, 30, 32, 34 and 36.

In addition, an empirical cumulative distribution function of the residuals will be plotted for the MMRM analysis.

Descriptive analyses

In addition to the inferential analysis, central laboratory and HemoCue hemoglobin Hb values and their associated change from baseline, will be reported descriptively by visit. For central lab Hb values, the average of weeks 28-36 will also be reported.

The following data will be presented graphically, by treatment arm:

- Hb results using mean values (+/- 95% CI) plot
- Hb change from baseline results using mean values (+/- 95% CI) plot.

7.4.1.1.2 Secondary Analyses (sensitivity) of the EU (EMA) Primary Endpoint

The following analyses will be done as sensitivity analyses of the EU (EMA) primary endpoint:

- The primary analysis will be repeated on the FAS and All Randomized;
- The primary endpoint will be analyzed similarly as for the primary analysis, but regardless of rescue therapy, on the PPS and FAS.
- The primary endpoint will be analyzed using an ANCOVA and Multiple Imputation (MI) of data, and including the same covariates, on the PPS.
- The primary endpoint will be analyzed using different Pattern Mixture Models (PMM) with different assumptions for missing patterns, including the same covariates, on the PPS.

All sensitivity analyses will be done on the total population.

The analysis using MMRM will be similar to the one provided in Section 7.4.1.1.1

For all sensitivity analyses, no hypothesis testing will be done, only confidence intervals presented and graphically represented in a forest plot.

Table 13 summarizes all sensitivity analyses to be performed with the EU (EMA) primary endpoint.

Table 13 Primary and Sensitivity Analyses for the EU (EMA) Primary Endpoint

Code	Set	Population	Endpoint	Model	Covariates*
Primary	PPS	Total	Change to the Average Hb in weeks 28-36, without rescue therapy	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb by visit, BL Hb
S1	FAS	Total	Change to the Average Hb in weeks 28-36, without rescue therapy	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb by visit, BL Hb
S2	All Randomized	Total	Change to the Average Hb in weeks 28-36, without rescue therapy	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb by visit, BL Hb
S3	PPS	Total	Change to the Average Hb in weeks 28-36, regardless of rescue therapy	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb by visit, BL Hb
S4	FAS	Total	Change to the Average Hb in weeks 28-36, regardless of rescue therapy	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb by visit, BL Hb
S5	PPS	Total	Change to the Average Hb in weeks 28-36, without rescue therapy	ANCOVA with MI	Region, History of CV, Previous ESA treatment, BL Hb
S6	PPS	Total	Change to the Average Hb in weeks 28-36, without rescue therapy	PMM (Last mean carried forward)	Region, History of CV, Previous ESA treatment, BL Hb
S7	PPS	Total	Change to the Average Hb in weeks 28-36, without rescue therapy	PMM – Last Mean Carried Forward for Roxadustat and Randomized arm MAR for ESAs	Region, History of CV, Previous ESA treatment, BL Hb

ANCOVA with MI:

The sensitivity analysis using ANCOVA with MI will be similar to the primary analysis of the US endpoint, provided in Section 7.4.1.2.1

Pattern Mixture Models (PMM)

Pattern Mixture models will be used as an alternative to impute missing values, using different assumptions for missing patterns.

PMMs provide a general and flexible framework for sensitivity analyses that allows formulating assumptions regarding missing data in a transparent and clinically interpretable manner. This is expected to address the possibility of the data being missing not at random. All factors mentioned in the primary analysis will be included in the PMM (refer to Section 7.4.1.1.1).

The following aspects of data missingness, may affect the estimates.

- Timing and extent of missingness
- Assumed underlying mechanism for data missingness

A. Timing and Extent of Missing Data

To assess the potential effect of data missingness on the estimate of treatment effect, subjects will be classified as full data or missing data cases. Patterns of missingness will be based on non-missing hemoglobin before the end of the evaluation period.

- Full data cases are defined as subjects with non-missing hemoglobin for all scheduled weeks of the Treatment period.
- Missing data cases are defined as subjects with a missing hemoglobin on at least one scheduled Week of the treatment period. The missing data cases are further grouped into intermittent missing and monotone missing cases.
    - Intermittent missing hemoglobin cases are defined as subjects with a missing hemoglobin for at least one scheduled week of but not on consecutive scheduled weeks up to end of the evaluation period.
    - Monotone missing hemoglobin cases are defined as subjects who have consecutive scheduled Weeks with missing hemoglobin up to the end of evaluation period. A subject who is a Monotone missing case could have intermittent missing hemoglobin prior to the ending Week.

Subjects will be grouped as follows:

- Full data cases
- Intermittent missing data cases
- Monotone missing data cases

Should the incidence of monotone missing data cases and intermittent missing data cases be relatively small, then those cases will be combined so that the groups are full data cases and missing data cases. The summary of missing patterns in the first 36 scheduled visits will be presented in a table/graph.

B. Assumptions on Missing Data Mechanism

In addition to the extent of data missingness, the mechanism under which missing data occur may affect the estimate of the parameter of interest.

The potential impact of missing efficacy endpoints on the estimates of treatment effects will be assessed using alternative statistical models with different underlying assumptions on the missing data mechanism for Missing Not AT Random (MNAR) (Little and Rubin, 1987).

C. 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 MMRM and ANCOVA with MI 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.

D. PMM –Mixed (Last Mean Carried Forward for Roxadustat and Randomized Arm MAR for Active Comparator)

This method is a combination of PMM-Last Mean Carried forward for Roxadustat and PMM-Randomized Arm Missing At Random (MAR) for ESA group. The imputation data will be generated based on last mean carried forward method described above for the Roxadustat treatment group, while for the ESA group, the imputation data will be generated using the Randomized Arm MAR group described below.

The Randomized Arm MAR is similar to PMM-Last Mean Carried Forward except that the joint distribution of the patient's observed and missing data is multivariate normal with mean and covariance matrix from their randomized treatment group.

The ANCOVA model will then be performed for each combined imputed complete data set. 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 analyses.

Model checking:

Residual plots will be produced showing the following:

- Model residuals in the y-axis and model predictions in the x-axis,
- Model residuals in the y-axis and baseline Hb in the x-axis.

Different dot styles will be used for the two treatment arms. Solid black symbols will be used for roxadustat and non-solid red symbols will be used for ESA.

Residual plots will be done for the ANCOVA analysis. In addition, an empirical cumulative distribution function of the residuals will be plotted for the same ANCOVA analysis.

If a relevant baseline variable is identified for which a clinically important imbalance exists at baseline between treatment groups, additional sensitivity analyses of the primary endpoint

may be performed using an ANCOVA model adjusting for this baseline variable. This will allow us to assess the impact of these imbalances on the treatment comparisons.

7.4.1.1.3 Additional Analyses of the EU (EMA) Primary Endpoint

In addition to the main estimates cited in Section 7.4.1.1.1, supportive estimates for each ESA treatment (darbepoetin alfa and epoetin alfa), will be presented for the primary analysis, on the PPS. Differences between treatments change from BL (roxadustat randomized subjects on epoetin before randomization minus epoetin alfa randomized subjects, and similarly for darbepoetin alfa) will be presented with indicative p-values, to support the analysis.

The analysis of the primary endpoint will be repeated by subgroup of interest. For definitions of subgroups of interest, see Section 7.8

The subgroups with a potential impact on Hb will be analyzed to check if there is an imbalance between the groups. If one of the subgroups is the same as a stratification factor, the factor will be omitted from the model.

Table 14 Additional Analyses of the EU (EMA) Primary Endpoint, by Subgroup

Code	Set	Population	Endpoint	Model	Covariates*
A1	PPS	Total	Change to the Average Hb in weeks 28-36, without rescue therapy by Subgroup	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb

Subgroup analyses will be done by producing separate summaries similar to those produced for the primary analysis. In addition, forest plots will be generated per each subgroup showing subgroup factors on the y-axis and change from baseline and their 95% confidence interval on the x-axis.

The potential existence of subgroup by treatment interaction will be visually inspected.

7.4.1.2 US (FDA) Primary Endpoint

7.4.1.2.1 Primary Analysis of the US (FDA) Primary Endpoint

The US (FDA) primary efficacy endpoint will be analyzed using the All Randomized.

The change from baseline to the average Hb of weeks 28 to 52 regardless of rescue therapy will be computed from an ANCOVA model with MI, adjusting for covariates defined in Section 7.4.1.1.1, comparing treatments (roxadustat vs. ESA).

The null hypothesis to be tested for the US primary efficacy analysis is:

- H_0: Hb change from baseline in the roxadustat arm $\leq$ Hb change from baseline in the ESA arm minus 0.75 g/dL

The alternative hypothesis H_1 is defined as the negation of the null hypothesis H_0. The overall one-sided significance level (alpha) is fixed as 0.025. The study will be successful for the US primary endpoint if the null hypothesis is rejected.

ANCOVA model:

The following model will be used:

$$c_{ikj} = intercept + \beta M_{baseline,ikj} + \tau_i + \gamma_k + \varepsilon_{ikj}$$

where

- c_{ikj} is the change from baseline to the average Hb values in weeks 28-52 of subject j in treatment arm i and stratum k,
- β is the slope of c_{ikj} as a linear function of the baseline Hb,
- $M_{baseline,ikj}$ is the baseline Hb of subject j in treatment arm i and stratum k,
- τ_i is the mean effect of treatment arm i
- $\gamma_{k,}$ is the mean effect of stratum k
- ε_{ikj} is the residual for subject j in treatment arm i and stratum k.

The SAS procedure will be similar to the following:

```
proc mixed;
   class treatment region CV_history previous_ESA_treatment;
   model change = treatment previous_ESA_treatment region cv_history
baseline_Hb;
run;
```

ANCOVA with MI model:

The MI ANCOVA model will be used to compare the roxadustat and ESA groups in a fixed sequence procedure:

The following steps will be used to conduct the primary analysis.

1. Generate 1000 datasets, using seed 253056, where intermittent missing hemoglobin data will be imputed for each treatment relying on non-missing data from all subjects within each treatment group using the Monte Carlo Markov Chain (MCMC) imputation model with treatment, baseline hemoglobin, randomization stratification factors and the available non missing hemoglobin for each scheduled Week.
   The MCMC statement in the SAS PROC MI procedure with monotone option will be used. As a result, each dataset will only have missing ending data, or a monotone missing data pattern.
2. For each dataset from step 1, missing ending data (hemoglobin up through end of evaluation period) will be imputed using seed 35286853. As a result, 1000 imputed complete datasets will be generated.

- Missing data at Week 1 will be imputed using the regression imputation model with baseline stratification factor, baseline and hemoglobin from Week 1, using the SAS PROC MI procedure with the REGRESSION option in the MONOTONE statement.
- The SAS PROC MI procedure will use data separately from each treatment subjects to impute the missing data for a specific Week (i.e. only those that need the imputation for the Week). Since subjects from the different treatment groups for that Week are excluded from the step, they will not contribute to the imputation for the Week.

- Repeat for all other scheduled Weeks sequentially (Week 2 to the end of evaluation period). 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 hemoglobin from previous Weeks and the stratification factors.

3. Analyze each imputed dataset using the ANCOVA using the mean of all observed or imputed Hb values within the evaluation period. The model will contain terms for baseline Hb measurement as a covariate and treatment arm and the other randomization stratification factors except screening Hb (≤11 vs >11) as fixed effects.
4. Combine estimates from the results of each of the 1000 ANCOVA runs using SAS PROC MIANALYZE.

Report the results of the least-squares mean estimates of the change from baseline in hemoglobin during the evaluation period, the estimates of treatment effect (e.g., least-squares mean change form baseline in hemoglobin for the treatment group minus the least-squares mean change from baseline in hemoglobin for the ESA group) and the corresponding p-values and 95% CIs during the evaluation period. Summary statistics for the imputed data by MI will be presented too.

All available Hb values will be used for the calculation of the average in weeks 28 to 52, as defined in analysis windows in Section 7.11.4.

A forest plot will be generated showing subgroup factors on the y-axis and differences in mean changes from BL to the average in week 28-52 and their 95% confidence interval on the x-axis.

If a relevant baseline variable is identified for which a clinically important imbalance exists at baseline between treatment groups, additional sensitivity analyses of the primary endpoint may be performed using an ANCOVA model adjusting for this baseline variable. This will allow us to assess the impact of these imbalances on the treatment comparisons.

Descriptive analyses

In addition to the inferential analysis, central laboratory and HemoCue hemoglobin Hb values and their associated change from baseline, will be reported descriptively by visit. For central lab Hb values, the average of weeks 28-52 will also be reported.

The following data will be presented graphically, by treatment arm:

- Hb results using mean values (+/- 95% CI) plot
- Hb change from baseline results using mean values (+/- 95% CI) plot.

7.4.1.2.2 Secondary Analyses (sensitivity) of the US (FDA) Primary Endpoint

Table 15 summarizes all sensitivity analyses to be performed with the US (FDA) primary endpoint. No analysis will be performed on the subset population with lower ESA doses, only the total population.

Table 15 Primary and Sensitivity Analysis for the US (FDA) Primary Endpoint

Code	Set	Endpoint	Method	Covariates
Primary	All Randomized	Change to the Average Hb in weeks 28-52, regardless rescue therapy	ANCOVA with MI	Region, History of CV, Previous ESA treatment, BL Hb
S1	FAS	Change to the Average Hb in weeks 28-52, regardless rescue therapy	ANCOVA with MI	Region, History of CV, Previous ESA treatment, BL Hb
S2	PPS	Change to the Average Hb in weeks 28-52, regardless rescue therapy	ANCOVA with MI	Region, History of CV, Previous ESA treatment, BL Hb
S3	All Randomized	Change to the Average Hb in weeks 28-52, regardless rescue therapy	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb by visit, BL Hb
S4	FAS	Change to the Average Hb in weeks 28-52, regardless rescue therapy	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb by visit, BL Hb
S5	FAS	Change to the Average Hb in weeks 28-52, without rescue therapy	ANCOVA with MI	Region, History of CV, Previous ESA treatment, BL Hb
S6	All Randomized	Change to the Average Hb in weeks 28-52, regardless rescue therapy	PMM (Last mean carried forward)	Region, History of CV, Previous ESA treatment, BL Hb
S7	All Randomized	Change to the Average Hb in weeks 28-52, regardless rescue therapy	PMM – Last Mean Carried Forward for Roxadustat and Randomized arm MAR for ESA	Region, History of CV, Previous ESA treatment, BL Hb

For all sensitivity analyses, no hypothesis testing will be done, only confidence intervals presented and graphically represented in a forest plot.

MMRM:

The analysis using MMRM model will be similar to the one provided in Section 7.4.1.1.1

One analysis Hb value for each visit will be used, as defined in analysis windows in Section 7.11.4 and Table 32.

Pattern Mixture Model (PMM)

Pattern Mixture models will be used as an alternative to impute missing values, using different assumptions for missing pattern. The PMM analysis method is described in Section 7.4.1.1.2, the analyses will be similar:

- PMM –Last Mean Carried Forward
- PMM – Last Mean Carried Forward for Roxadustat and Randomized arm MAR for ESA

Model checking:

Residual plots will be produced showing the following:

- Model residuals in the y-axis and model predictions in the x-axis,
- Model residuals in the y-axis and baseline Hb in the x-axis.

Different dot styles will be used for the two treatment arms. Solid black symbols will be used for roxadustat and non-solid red symbols will be used for ESA.

Residual plots will be done for the ANCOVA analysis. In addition, an empirical cumulative distribution function of the residuals will be plotted for the same ANCOVA analysis.

If a relevant baseline variable is identified for which a clinically important imbalance exists at baseline between treatment groups, additional sensitivity analyses of the primary endpoint may be performed using an ANCOVA model adjusting for this baseline variable. This will allow us to assess the impact of these imbalances on the treatment comparisons.

7.4.1.2.3 Additional Analyses of the US (FDA) Primary Endpoint

In addition to the main estimates cited in Section 7.4.1.2.1, supportive estimates for each ESA treatment (darbepoetin alfa and epoetin alfa), will be presented for the primary analysis, on the All Randomized.

The analysis of the US (FDA) primary endpoint will be repeated by subgroup of interest. For definitions of subgroups of interest, see Section 7.8

The subgroups with a potential impact on Hb will be analyzed to check if there is an imbalance between the groups. If one of the subgroups is the same as a stratification factor, the factor will be omitted from the model.

Table 16 Additional Analyses of the US (FDA) Primary Endpoint, by Subgroup.

Code	Set	Endpoint	Model	Covariates
A1	All Randomized	Change to the Average Hb in weeks 28-52 by Subgroup	ANCOVA with MI	Region, History of CV, Previous ESA treatment, BL Hb

Subgroup analyses will be done by producing separate summaries similar to those produced for the primary analysis. In addition, forest plots will be generated per each subgroup showing subgroup factors on the y-axis and change from baseline and their 95% confidence interval on the x-axis.

The potential existence of subgroup by treatment interaction will be visually inspected.

7.4.2 Analysis of Key Secondary Endpoints

Once the null hypothesis has been rejected for the EU (EMA) primary efficacy endpoint for both the total population and the subset population, the key secondary endpoints will be tested for the total population, as described below.

The EU (EMA) primary analysis set for the analysis of the key secondary endpoints will be the PPS for the non-inferiority tests and the FAS for the superiority tests.

All inferential analyses will evaluate the difference between the treatment arms: roxadustat versus ESA (i.e. pooled darbepoetin alfa and epoetin alfa).

The key secondary endpoints will be tested using a fixed sequence testing procedure, as depicted in Table 17 in order to maintain the overall one-sided type I error rate for the set of key secondary endpoints at 0.025. If the null hypothesis is rejected for a test, the claim of superiority (or non-inferiority) will be considered successful and the test will progress to the next comparison in sequence as follows:

Table 17 Key Secondary Endpoints Fixed Sequence Testing Procedure

Test	Analysis sets	Endpoint	Comparison
1	PPS	Hb response defined as mean Hb during weeks 28 to 36 within the target range of 10.0 to 12.0 g/dL without having received rescue therapy within 6 weeks prior to and during this 8-week evaluation period.	Non-inferiority of roxadustat versus ESA (the non-inferiority margin for the difference between groups is -15%
2	FAS	Change from BL in Low Density Lipoprotein (LDL) cholesterol to the average LDL cholesterol of weeks 12 to 28.	Superiority of roxadustat versus ESA
3	FAS	Mean monthly IV iron use (mg) during day 1 to week 36 (monthly defined as a period of 4 weeks).	Superiority of roxadustat versus ESA
4	PPS	Change from BL in SF-36 Physical Functioning (PF) sub-score to the average PF sub-score of weeks 12 to 28.	Non-inferiority of roxadustat versus ESA (the non-inferiority margin is fixed as -3)
5	PPS	Change from BL in SF-36 Vitality (VT) sub-score to the average VT sub-score of weeks 12 to 28.	Non-inferiority of roxadustat versus ESA (the non-inferiority margin is fixed as -3)
6	PPS	Change from BL in mean arterial pressure (MAP) to the average MAP value of weeks 20 to 28.	Non-inferiority of roxadustat versus ESA (the non-inferiority margin for the difference between groups is 1 mmHg)
7	PPS	Time to first occurrence of an increase in blood pressure: An increase from BL of ≥ 20 mm Hg systolic blood pressure (SBP) and SBP ≥ 170 mmHg or an increase from baseline of ≥15 mmHg diastolic blood pressure (DBP) and DBP ≥ 100 mmHg during weeks 1 to 36.	Non-inferiority of roxadustat versus ESA (the non-inferiority margin for the difference between groups is fixed as a hazard ratio of 1.3).
8	FAS	Change from BL in mean arterial pressure (MAP) to the average MAP value of weeks 20 to 28.	Superiority of roxadustat versus ESA
9	FAS	Time to first occurrence of an increase in blood pressure: An increase from BL of ≥ 20 mm Hg systolic blood pressure (SBP) and SBP ≥ 170 mmHg or an increase from baseline of ≥15 mmHg diastolic blood pressure (DBP) and DBP ≥ 100 mmHg during weeks 1 to 36).	Superiority of roxadustat versus ESA

Details of the analysis for each of these secondary endpoints are given below.

Descriptive statistics and frequency tabulations will be reported.

7.4.2.1 Hb response defined as mean Hb during weeks 28 to 36 within the target range of 10.0 to 12.0 g/dL without having received rescue therapy within 6 weeks prior to and during this 8-week evaluation period.

The difference in the proportion of responders between the treatment arms will be analyzed using a Miettinen & Nurminen (MN) approach and adjusting for covariates defined below:

The covariates will be:

- Previous ESA treatment (epoetin versus darbepoetin alfa);
- Region (region A versus region B);
- History of cardiovascular, cerebrovascular or thromboembolic diseases (Yes versus No);
- Average prescribed weekly ESA dose prior to randomization (≤200 IU/kg epoetin or ≤ 1 µg/kg darbepoetin alfa versus >200 IU/kg epoetin or > 1 µg/kg darbepoetin alfa) [*];
- Baseline Hb value (≤11.0 g/dL versus >11.0 g/dL).

[*]: The average prescribed weekly ESA dose prior to randomization will not be used in the analysis due to the number of subjects in the high dose group was too low for the variable to be meaningful.

The SAS code will be similar to the one below:

```
proc freq;
  tables  ESA_trt*region*CV_history* Hb_category*arm*response/
riskdiff (cl=MN common);
run;
```

The stratified MN method has recently been implemented in SAS version 9.4. The current SAP states use of SAS version 9.3 or higher (see Section 7.1). Therefore, depending on the availability of SAS version 9.4 at the time of analysis, the stratified MN statistics could be replaced by the alternative methodology based on the standard normal statistic proposed by Gart and Nam (see reference in Section 9).

The SAS code will be similar to the one below:

```
proc genmod;
  class arm ESA_trt region CV_history;
  model response=ESA_trt region CV_history Hb_continuous arm
/dist=bin link=identity;
  lsmeans arm / diff cl;
run;
```

This endpoint will be analyzed using the PPS. See Section 7.11.1 for imputation rules, in case of missing values in the evaluation period or in case of rescue therapy within 6 weeks prior to and during this 8-week evaluation period.

The point estimates for the proportion of responders in the roxadustat group and the ESA group and the difference in proportions between these rates will be calculated (roxadustat –

ESA). The MN method will be used to calculate the two-sided 95% CI for the difference in rates, adjusting for covariates. If the resulting lower bound of the two-sided 95% CI between roxadustat and ESA (roxadustat minus ESA) is > -15%, non-inferiority will be concluded.

Two sensitivity analyses (not part of the sequence) will be performed, one on All Randomized and the other one on the PPS regardless the use of rescue therapy.

In addition, a 95% confidence interval for the proportion of roxadustat and ESA responders based on the exact method of Clopper-Pearson will be calculated and presented, for all analyses.

The average Hb value of weeks 28-36 will also be reported.

A forest plot will be generated showing subgroup factors on the y-axis and differences in proportions and their 95% confidence interval on the x-axis.

Table 18 Primary and Sensitivity Analysis for the Hb Response

Code	Set	Rescue Therapy	Endpoint	Method	Covariates
Primary	PPS	Without rescue therapy	Hb response	Miettinen & Nurminen (MN) approach	Region, History of CV, Previous ESA treatment, BL Hb (categorical)
S1	FAS	Without rescue therapy	Hb response	Miettinen & Nurminen (MN) approach	Region, History of CV, Previous ESA treatment, BL Hb (categorical)
S2	PPS	Regardless rescue therapy	Hb response	Miettinen & Nurminen (MN) approach	Region, History of CV, Previous ESA treatment, BL Hb (categorical)

7.4.2.2 Change from BL in Low Density Lipoprotein (LDL) cholesterol to the average LDL cholesterol of weeks 12 to 28.

LDL change from baseline to the average LDL value in weeks 12-28 will be compared by treatment arms using a MMRM model as in Section 7.4.1.1.1 (with the addition of LDL at baseline as a continuous covariate).

The analysis will be similar to the primary analysis provided in Section 7.4.1.1.1, except that it will be a superiority test instead of a non-inferiority test. This superiority test will be considered successful if the upper bound of the two-sided 95% confidence interval of the difference between treatment arms (roxadustat minus ESA) is lower than 0.

The analysis will be done on the FAS. An additional analysis (not part of the sequence) will be done on All Randomized.

For missing LDL imputation rules, refer to Section 7.11.1.

Theses analyses will be done on all values (regardless of fasting status).

In addition to the inferential analysis, LDL cholesterol and LDL cholesterol change from baseline will be reported descriptively by visit for fasted values. The average of weeks 12-28 will also be reported.

Residual plots will be produced showing the following:

- Model residuals in the y-axis and model predictions in the x-axis,
- Model residuals in the y-axis and baseline LDL-cholesterol in the x-axis.

Table 19 Primary and Sensitivity Analysis for Change from BL in LDL to the Average LDL of weeks 12 to 28

Code	Set	Endpoint	Method	Covariates
Primary	FAS	Change from BL in Low Density Lipoprotein (LDL) cholesterol to the average LDL cholesterol of weeks 12 to 28.	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL LDL
S1	All Randomized	Change from BL in Low Density Lipoprotein (LDL) cholesterol to the average LDL cholesterol of weeks 12 to 28.	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL LDL

7.4.2.3 Mean monthly IV iron use (mg) during Day 1 to Week 36 (monthly defined as a period of 4 weeks).

The average monthly IV iron use (mg) during Day 1 to Week 36 will be compared by treatment arm using an ANCOVA model, as in Section 7.4.1.2.1 (except that it will be done on the total average of weeks 1 to 36 and that there will be no imputation). Subjects on study with no medication record of IV Iron will be assumed that they received no IV Iron. For those subjects (on study), monthly IV iron use will be set to zero mg.

The analysis will be done on the FAS. This superiority test will be considered successful if the upper bound of the two-sided 95% confidence interval of the difference between treatment arms (roxadustat minus ESA) is lower than 0.

An additional analysis (not part of the sequence) will be performed on All Randomized.

Furthermore, the mean monthly IV iron use will be reported for Day 1 to Week 36, weeks 37 to 52, and weeks 53 to 104. Number and percentage of subjects took IV iron and/or Oral Iron at any time and during Day 1 to week 36, weeks 37 to 52, and weeks 53 to 104 will be summarized.

Time to first IV iron during day 1 to week 36 will be analyzed similarly as in Section 7.4.2.7

Table 20 Primary and Sensitivity Analysis for Mean Monthly IV Iron use (mg) During Day 1 to Week 36

Code	Set	Endpoint	Method	Covariates
Primary	FAS	Mean monthly IV iron use (mg) during Day 1 to Week 36	ANCOVA	Region, History of CV, Previous ESA treatment, BL Hb
S1	All Randomized	Mean monthly IV iron use (mg) during Day 1 to Week 36	ANCOVA	Region, History of CV, Previous ESA treatment, BL Hb
S2	FAS	Occurrence and time to first use of IV iron in day 1 to week 36	Cox regression + Kaplan Meier	Stratified on Region, History of CV, Previous ESA treatment, and adjusted on BL Hb as continuous covariate

7.4.2.4 Change from BL in SF-36 Physical Functioning (PF) sub-score to the average PF sub-score of weeks 12 to 28.

Change from baseline in PF subscore of SF-36 to the average of weeks 12–28 will be compared by treatment arm, for the PPS. It will be done using a MMRM method using region, history of CV, previous ESA treatment as categorical covariates and baseline SF-36 PF subscore, baseline Hb as continuous covariates. Non-inferiority can be concluded if the lower bound of the two-sided 95% CI of the difference between the two treatment arms (roxadustat minus ESA) is greater than -3 points.

The analysis will be similar to the one provided in Section 7.4.1.1.1

Residual plots will be produced showing the following:

- Model residuals in the y-axis and model predictions in the x-axis,
- Model residuals in the y-axis and baseline SF-36 PF subscore in the x-axis.

An additional analysis (not part of the sequence) will be done on All Randomized.

The analysis will be repeated in the subsets of subjects with baseline PF subscore below 40 and equal or above 40.

In addition to the inferential analyses, SF-36 PF and SF-36 PF change from baseline will be reported descriptively by visit, using all available data for PPS and FAS. The average of weeks 12-28 will also be reported.

Table 21 Primary and Sensitivity Analysis for Change from BL in SF-36 PF Sub-score to the Average PF Sub-score of weeks 12 to 28

Code	Set	Endpoint	Method	Covariates
Primary	PPS	Change from BL in SF-36 PF sub-score to the average PF sub-score of weeks 12 to 28	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL SF-36 PF subscore, BL Hb
S1	FAS	Change from BL in SF-36 PF sub-score to the average PF sub-score of weeks 12 to 28	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL SF-36 PF subscore, BL Hb

7.4.2.5 Change from BL in SF-36 VT subscore to the average in weeks 12–28

Change from baseline in VT subscore of SF-36 to the average of weeks 12–28 will be compared by treatment arm, for the PPS. It will be done using a MMRM method using region, history of CV, previous ESA treatment as categorical covariates and baseline SF-36 VT subscore, baseline Hb as continuous covariates. Non-inferiority can be concluded if the lower bound of the two-sided 95% CI of the difference between the two treatment arms (roxadustat minus ESA) is greater than -3 points.

The analysis will be similar to the one provided in Section 7.4.1.1.1

Residual plots will be produced showing the following:

- Model residuals in the y-axis and model predictions in the x-axis,
- Model residuals in the y-axis and baseline SF-36 VT subscore in the x-axis.

An additional analysis (not part of the sequence) will be done on All Randomized.

The analysis will be repeated in the subsets of subjects with baseline vitality subscore below 50 and equal or above to 50.

In addition to the inferential analyses, SF-36 VT and SF-36 VT change from baseline will be reported descriptively by visit, using all available data for PPS and FAS. The average of weeks 12-28 will also be reported.

Table 22 Primary and Sensitivity Analysis for Change from BL in SF-36 VT Sub-score to the Average VT Sub-score of weeks 12 to 28

Code	Set	Endpoint	Method	Covariates
Primary	PPS	Change from BL in SF-36 VT sub-score to the average VT sub-score of weeks 12 to 28	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL SF-36 VT subscore, BL Hb
S1	FAS	Change from BL in SF-36 VT sub-score to the average VT sub-score of weeks 12 to 28	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL SF-36 VT subscore, BL Hb

7.4.2.6 Change from BL in pre-dialysis mean arterial pressure (MAP) to the average MAP value of weeks 20 to 28.

The blood pressure effect is shown by use of the change from baseline in MAP. MAP values are pre-dialysis MAP values.

MAP change from baseline to the average MAP value in weeks 20-28 will be analyzed using a MMRM model as in Section 7.4.1.1.1 (with the addition of MAP at baseline as continuous covariate).

For missing MAP imputation rules, refer to Section 7.11.1

Non-inferiority (by the absence of a rise in MAP) can be concluded if the upper bound of the two-sided 95% CI of the difference between the two treatment arms (roxadustat minus ESA) is lower than 1 mm Hg. This will be calculated on the PPS.

As per Table 17, once the non-inferiority is concluded, superiority will be checked for this variable on the FAS as part of the fixed sequence testing procedure. Superiority can be concluded if the upper bound of the two-sided 95% CI of the difference between roxadustat and ESA (roxadustat minus ESA) is lower than 0 mmHg.

A sensitivity analysis will be done on All Randomized.

In addition to the inferential analyses, MAP and MAP change from baseline will be reported descriptively by visit, using all available data. The average of weeks 20-28 will also be reported.

Residual plots will be produced showing the following:

- Model residuals in the y-axis and model predictions in the x-axis,
- Model residuals in the y-axis and baseline MAP-cholesterol in the x-axis.

Table 23 Primary and sensitivity analysis for change from BL in MAP to the average MAP value of weeks 20 to 28

Code	Set	Endpoint	Method	Covariates
Primary	PPS	Change from BL in MAP to the average MAP value of weeks 20 to 28	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL MAP
Primary	FAS	Change from BL in MAP to the average MAP value of weeks 20 to 28	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL MAP
S1	All Randomized	Change from BL in MAP to the average MAP value of weeks 20 to 28	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL MAP

7.4.2.7 Time to first occurrence of an increase in pre-dialysis blood pressure: An increase from BL of ≥ 20 mm Hg systolic blood pressure (SBP) and SBP ≥ 170 mmHg or an increase from baseline of ≥15 mmHg diastolic blood pressure (DBP) and DBP ≥ 100 mmHg during weeks 1 to 36.

Time to first occurrence of an increase in blood pressure during the first 36 weeks, including time to censoring, defined in Section 6.1.2.7, will be used in a Cox Proportional Hazards regression analysis. It will compare treatment arms, using covariates defined in Section 7.4.1.1.1, and provide hazard ratio and their 95% confidence intervals.

Blood pressures are pre-dialysis blood pressures.

Non-inferiority will be declared if the upper bound of the 2-sided 95% confidence interval of the hazard ratio (roxadustat as relative to ESA), calculated on the PPS, is lower than 1.3. As per Table 17, once the null hypothesis is rejected, superiority will be checked for this variable on the FAS, as part of the fixed sequence testing procedure. Superiority will be concluded if the upper bound of the two-sided 95% CI of the hazard ratio of the two treatment arms (roxadustat as relative to ESA) is lower than 1.

A sensitivity analysis will be done on All Randomized.

The Cox Model can be written:

$$\lambda_j(t;\underline{x}) = \lambda_{0j}(t)\exp(\alpha z)$$

where

j - indicator for stratum
z – treatment indicator – which is either roxadustat or ESA

The SAS procedure for the Cox regression will be similar to the following:

```
proc phreg;
  model time_to_event*cens_var(1) = treatment Hb_base / rl;
  strata previous_ESA_treatment region cv_history;
where visit in
('Day1','Week1','Week2',...,'Week8','Week10',...,'Week36');
run;
```

In addition, the cumulative incidence curve of subjects with an increase in blood pressure from baseline will be plotted by treatment arm and type of ESA treatment.

The cumulative incidence will be calculated as one minus the Kaplan-Meier estimate of the survival function. Two types of analyses can be performed: modeling the cause specific hazard or modeling the hazard of the sub-distribution. The first approach has been chosen in this SAP, because competing risks are assumed not to exist, since the main interest is to study the treatments effect and this method provides results to be generalized across datasets with different competing risks. One minus the Kaplan-Meier estimate can be interpreted as the probability that an event of interest occurs to a subject by time t, in the absence of any competing risk.

In addition to the cumulative incidence plot, the cumulative incidence at 0.25, 0.5 and 0.75 year with the 95% confidence interval will be reported using Greenwood's formula.

Model checking

The proportional hazards assumption will be checked graphically using a log-cumulative hazard plot against log-survival time. This plot will give approximately parallel lines if the proportional hazards assumption between treatment arms subgroups holds.

The proportional hazards assumption will also be tested by fitting a Cox Proportional Hazards regression, stratified for region, history of CV, previous ESA treatment, and adjusted for baseline Hb (continuous) with one dependent variable: treatment (0 for ESA, 1 for roxadustat).

Table 24 Primary and Sensitivity Analysis for the time to First Occurrence of an Increase in Blood Pressure in the First 36 weeks

Code	Set	Endpoint	Method	Strata/Covariates
Primary	PPS	Occurrence and time to first occurrence of increase in blood pressure in week 1 to 36	Cox regression + Kaplan Meier	Stratified on Region, History of CV, Previous ESA treatment, and adjusted on BL Hb as continuous covariate
Primary	FAS	Occurrence and time to occurrence of increase in blood pressure in week 1 to 36	Cox regression+ Kaplan Meier	Stratified on Region, History of CV, Previous ESA treatment, and adjusted on BL Hb as continuous covariate
S1	All Randomized	Occurrence and time to occurrence of increase in blood pressure in week 1 to 36	Cox regression+ Kaplan Meier	Stratified on Region, History of CV, Previous ESA treatment, and adjusted on BL Hb as continuous covariate

In addition, the number and percentage of subjects with an increase from baseline in blood pressure during weeks 1 to 36 will be reported by treatment arm and type of ESA treatment, on the PPS. They will also be reported for the total duration of the study. For subjects who have experienced more than one increase in blood pressure, only their first event following study treatment will be used in the analysis.

In addition, the incidence rate (per 100 subject years at risk) will also be calculated as follows:

$$\frac{\text{Number of subjects with event}}{\text{Total cumulative time at risk (years)}} \times 100$$

Where Total cumulative time at risk is the sum of individual time at risk defined as either time to occurrence of the event or time to censoring for subjects with no event.

7.4.2.8 Additional Analyses of the Key Secondary Endpoints

In addition to the main estimates cited in Sections 7.4.2.1 to 7.4.2.7, descriptive supportive estimates for each type of ESA treatment (darbepoetin alfa and epoetin alfa) will be presented for the primary analysis of all key secondary endpoints.

Each of these key secondary endpoints will also be analyzed by the subgroups of interest defined in Section 7.8, using only the primary analysis method.

Forest plots will be produced where all subgroup factors will appear in the y-axis and the appropriate statistic comparing roxadustat to ESA and the 95% confidence interval will appear in the x-axis.

Table 25 Additional Analyses of the Key Secondary Endpoints

Code	Analysis sets	Endpoint	Method	Covariates
A1	PPS	Hb response defined as mean Hb during weeks 28 to 36 within the target range of 10.0 to 12.0 g/dL without having received rescue therapy within 6 weeks prior to and during this 8-week evaluation period by Subgroup	MN method	Region, History of CV, Previous ESA treatment, BL Hb (categorical)
A2	FAS	Change from BL in Low Density Lipoprotein (LDL) cholesterol to the average LDL cholesterol of weeks 12 to 28 by Subgroup.	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL LDL
A3	FAS	Mean monthly IV iron use (mg) during day 1 to week 36 (monthly defined as a period of 4 weeks) by Subgroup.	ANCOVA	Region, History of CV, Previous ESA treatment, BL Hb
A4	PPS	Change from BL in SF-36 Physical Functioning (PF) sub-score to the average PF sub-score of weeks 12 to 28 by Subgroup.	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL SF-36 PF subscore, BL Hb
A5	PPS	Change from BL in SF-36 Vitality (VT) sub-score to the average VT sub-score of weeks 12 to 28 by Subgroup.	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL SF-36 VT subscore, BL Hb
A6	PPS	Change from BL in mean arterial pressure (MAP) to the average MAP value of weeks 20 to 28 by Subgroup.	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL MAP

Table continued on next page

Code	Analysis sets	Endpoint	Method	Covariates
A7	PPS	Time to an increase in blood pressure: An increase from BL of ≥ 20 mm Hg systolic blood pressure (SBP) and SBP ≥ 170 mmHg or an increase from baseline of ≥15 mmHg diastolic blood pressure (DBP) and DBP ≥ 100 mmHg during weeks 1 to 36 by Subgroup.	Cox regression	Stratified on Region, History of CV, Previous ESA treatment, and adjusted on BL Hb as continuous covariate
A8	FAS	Change from BL in mean arterial pressure (MAP) to the average MAP value of weeks 20 to 28 by Subgroup.	MMRM	Region, History of CV, Previous ESA treatment, Visits and Visit by Treatment, BL Hb, BL MAP
A9	FAS	Time to an increase in blood pressure: An increase from BL of ≥ 20 mm Hg systolic blood pressure (SBP) and SBP ≥ 170 mmHg or an increase from baseline of ≥15 mmHg diastolic blood pressure (DBP) and DBP ≥ 100 mmHg during weeks 1 to 36 by Subgroup.	Cox regression	Stratified on Region, History of CV, Previous ESA treatment, and adjusted on BL Hb as continuous covariate

7.4.3 Analysis of Additional Secondary Efficacy Endpoints

All the analyses below will be superiority tests. These inferential analyses will be performed on the FAS and present the treatment effect of Roxadustat versus ESA.

The analyses will be performed on the total population.

Descriptive statistics will be presented by treatment arm and type of ESA treatment, when applicable.

7.4.3.1 Hb response regardless of use of rescue therapy.

This analysis is identical to the one defined in Section 7.4.2.1, except that it is regardless of rescue therapy and a superiority test will be performed on the FAS.

The two-sided 95% CI for the difference in rates will be calculated similarly. If the resulting lower bound of the two-sided 95% CI between roxadustat and ESA is > 0, superiority will be concluded.

Hb responses (defined as mean Hb values between 10.0 to 12.0 g/dL and also defined as mean Hb values >=10.0 g/dL) during weeks 28 to 36, and during weeks 28 to 52 will be analyzed similarly.

For Hb response (mean Hb between 10.0 to 12.0 g/dL) during weeks 28 to 52 will be analyzed using logistic regressions too, the odds ratio (roxadustat vs. ESA) and their 95% confidence intervals will be produced, if convergence achieved.

The SAS procedure will be similar to the following:

```
proc logistic;
class treatment previous_ESA_treatment region CV_history;
model response = treatment previous_ESA_treatment region cv_history
baseline_Hb;
run;
```

7.4.3.2 Hb change from BL to each post-dosing time point.

Hb value and change from BL Hb to each post dosing time point will be described by treatment arm on the FAS.

The analysis will be similar to the one provided in Section 7.4.1.1.1

7.4.3.3 Hb level averaged over weeks 28 to 36, 44 to 52, and 96 to 104 without use of rescue therapy within 6 weeks prior to and during these 8-week evaluation periods.

Averaged Hb over weeks 28-36, 44 to 52 and 96 to 104 will be described by treatment arm on the FAS.

The analysis will be similar to the one provided in Section 7.4.1.1.1

In addition, the number and proportion of subjects with average Hb over weeks 28-36 and over weeks 44-52, 72-80 and 96-104 and by visit within the <10 g/dL, >=10 g/dL, 10-12 g/dL and >12 g/dL categories will be reported.

7.4.3.4 Hb change from BL to the average Hb value of weeks 28 to 36, 44 to 52, and 96 to 104 regardless of the use of rescue therapy.

The analysis will be done similarly as in Section 7.4.1.1.1, on the FAS except that it will be for weeks 28-36, 44-52 and 96-104.

7.4.3.5 Categorical analysis for Hb values

For the analyses below, results will be displayed by treatment arm and type of ESA treatment.

Proportion of Hb values:

The proportion of Hb values >=10 g/dL and within 10-12 g/dL is a quantitative variable in the 0 – 100 range for each subject. Descriptive statistics for this variable will be presented. This variable will be presented in weeks 28-36, in weeks 44-52 and in weeks 96 -104, on the FAS.

Percentage of time during the Efficacy Emergent Period:

Descriptive statistics for the percentage of time with Hb values falling in each interval (<10.0 g/dL, >=10 g/dL, within 10.0-12.0 g/dL, > 12.0 g/dL, > 13.0 g/dL and > 14.0 g/dL) between the first and last Hb assessment during the Efficacy Emergent Period will be presented.

The number and percentage of subjects, and the number of events, will be reported by for the following potential EH events of interest, based on Hb values from the central laboratory:

- Hb increase by >2.0 g/dL between any two visits within 4 weeks of treatment

Time to first occurrence of potential EH will also be analyzed similarly as in Section 7.4.2.7 (except that all data during the Efficacy Emergent Period will be taken into account for the analysis).

Subjects with at least one Hb value > 13g/dL during the treatment period will be summarized.

7.4.3.6 Occurrence (number) of hospitalizations, number of days of hospitalization per year and time to first hospitalization during the study.

The number and percentage of subjects with hospitalization will be reported. Descriptive statistics and frequency tabulations by treatment arm and type of ESA treatment of the total duration of hospitalization (days), the average duration of each hospitalization (days), the number of hospitalizations, number of days of hospitalization per year and reason for hospitalization will also be reported.

Time to first hospitalization will also be analyzed similarly as in Section 7.4.2.7 (except that all data during the Efficacy Emergent Period will be taken into account for the analysis).

Superiority will be declared if the upper bound of the two-sided 95% confidence interval of the hazard ratio of the two treatment arms is below 1.

The analysis will be done on the FAS.

7.4.3.7 Occurrence and time to first use of rescue therapy, occurrence and time to first use of RBC transfusions, number of RBC packs per subject, volume of RBC transfused per subject.

The number and percentage of subjects with rescue therapy will be reported by treatment arm and type of ESA treatment. Time to first use of rescue therapy will also be analyzed similarly as in Section 7.4.2.7 (except that all data during the Efficacy Emergent Period will be taken into account for the analysis).

Mean monthly number of RBC packs and volume of RBC transfused during the Efficacy Emergent Period will be compared by treatment arm using an ANCOVA model as in Section 7.4.1.2.1. Subjects with no medication record of RBC will be assumed that they received no RBC and therefore number of packs and volume will be set to 0.

The analysis will be done on the FAS.

Superiority will be declared if the upper bound of the two-sided 95% confidence interval of the hazard ratio of the two treatment arms is below 1.

In addition, the number and percentage of subjects with RBC transfusion will be reported. Descriptive statistics by treatment arm and type of ESA treatment for number of RBC packs and volume of blood transfused will be reported. Time to first use of RBC transfusion will be analyzed similarly as in Section 7.4.2.7 (except that all data during the Efficacy Emergent Period will be taken into account for the analysis).

The number and percentage of roxadustat subjects with ESA use as rescue therapy, and the number of ESA-weeks will be reported descriptively for the roxadustat treatment arm only.

7.4.3.8 Occurrence of iron supplementation.

The number and percentage of subjects who received IV Iron and the total amount of IV Iron used during the study will be reported descriptively by treatment arm and type of ESA treatment.

The incidence and cumulative incidence of subjects receiving IV Iron will be calculated similarly as in Section 7.4.2.7 (except that all data during the Efficacy Emergent Period will be taken into account for the analysis).

Mean monthly IV iron (mg) per subject during weeks 37-52 and weeks 53-104 will be compared by treatment arm and analyzed similarly as in Section 7.4.2.3

The analysis will be done on the FAS.

7.4.3.9 Change from BL to each post-dosing study visit in Total cholesterol, LDL/High-density Lipoprotein (HDL) ratio, Non-HDL cholesterol, Triglycerides, Apolipoproteins A1 and B, ApoB/ApoA1 ratio.

Descriptive statistics (value, change from baseline) by visit, treatment arm and type of ESA treatment will be presented regardless of fasting status, including averages in week 12-28. Descriptive statistics will also be reported with fasted values.

7.4.3.10 Occurrence of mean LDL cholesterol <100 mg/dL calculated over weeks 12 to 28.

The number and percentage of subjects with mean LDL cholesterol less than 100 mg/dL (2.59 mmol/L) (regardless fasting status) on average in weeks 12-28 will be reported by treatment arm and type of ESA treatment. Descriptive statistics and frequency tabulations will also be reported on fasted values, and regardless of fasting status.

7.4.3.11 Occurrence of achieved antihypertensive treatment goal in CKD subjects (SBP< 140 mmHg and DBP< 90 mmHg) based on the mean SBP and mean DBP calculated over weeks 12 to 28.

The number and percentage of subjects with achieved antihypertensive treatment goal over an evaluation period defined as the average of available values in weeks 12-28 will be reported by treatment arm and type of ESA treatment.

7.4.3.12 Health related Quality of Life Questionnaires Change from BL to the average value of weeks 12 to 28

The following questionnaires will be analyzed:

- SF-36 (Physical Component Score (PCS));
- FACT-An (Anemia Subscale (“Additional Concerns”), Total FACT-An Score);
- EQ-5D 5L (VAS Score);
- Patients’ Global Impression of Change (PGIC).

SF-36 and FACT-An

Descriptive statistics and frequency tabulations (value, change from baseline) will be presented for SF-36 (subscale and component scores) and FACT-An (total and subscale scores) by visit, treatment arm and type of ESA treatment. Mean values will also be plotted over time and by treatment arm.

In addition, for the average value in weeks 12-28, an inferential analysis, similar to the one defined in Section 7.4.1.1.1 will be performed for the following endpoints:

- Physical Component Scores of SF-36 (SF-36 PCS)
- Anemia Subscale ("Additional Concerns") of FACT-An Scores
- Total FACT-An Scores
- Fatigue

For SF-36, at each visit, the frequency and proportion of subjects with a change from baseline <3/>= 3 points and <5/>=5 points will be reported for Physical Functioning, Vitality and Physical Component scores.

In addition to this, change from BL to the average value of weeks 36 to 52 will be summarized similarly as for change from BL to the average value of week 12 to 28.

EQ-5D 5L

For the EQ-5D 5L VAS score, change from baseline to each visit and to the average of weeks 12-28 and 36-52 will be described by treatment arm.

For the 5 EQ-5D 5L qualitative domains, the number and percentage of subjects in each response level value will be reported by visit, treatment arm and type of ESA treatment.

7.4.3.13 Patients' Global Impression of Change (PGIC)

Patients' Global Impression of Change will be summarized descriptively by visit, treatment arm and type of ESA treatment.

7.4.3.14 Changes from BL to each study visit (when measured) in Serum hepcidin, Serum ferritin, TSAT and HbA1c level.

Changes from baseline to each study visit will be calculated for these parameters:

- Serum hepcidin
- Serum ferritin
- TSAT
- Serum Iron
- TIBC
- HbA1c level (Subjects with diabetes, Subjects without diabetes, and overall)

Descriptive statistics and frequency tabulations will be presented for these parameters and for the change from baseline by visit, treatment arm and type of ESA treatment. Mean values will also be plotted versus visit by treatment arm.

7.4.4 Analysis of Exploratory Variables: hs-CRP (High Sensitivity C-Reactive Protein) and Post-dialysis BP.

For each visit, descriptive statistics with the absolute values and change from baseline for hs-CRP will be displayed by treatment arm and type of ESA treatment.

Genotyping will be shipped to a delegated CRO and analyzed under the responsibility of Bioanalysis-Europe of Astellas Pharma Europe B.V. A separate report will be provided.

The BP-related key secondary endpoints will be analyzed using the post-dialysis BP assessments: MAP and time to first occurrence of increase in blood pressure, similarly with the analyses in Sections 7.4.2.6 and 7.4.2.7, excluding the analysis using All randomized subjects and subgroup analysis.

7.5 Analysis of Safety

Safety analyses will be performed using the Safety Analysis Set (SAF).

Missing dates' imputation rules for AE onset date and stop date are detailed in Section 7.11.2

For each safety parameter, the last non-missing assessment prior to the first dose of study drug will be used as the baseline for all analyses, unless specified otherwise.

For the safety analyses by subgroup, the subgroup analysis by Previous ESA treatment will not be presented in tables, as this analysis is already performed by Type of ESA treatment (Subgroup Darbepoetin and Subgroup Epoetin). In forest plots, results for this subgroup factor Previous ESA treatment will be included.

7.5.1 Adverse Events

All adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 20.0. They will be summarized by System Organ Class (SOC) and Preferred Term (PT).

7.5.1.1 Overview

An overview table will include the following details, by treatment arm and type of ESA treatment:

- Number and percentage of subjects with TEAEs,
- Number and percentage of subjects with causally drug related TEAEs,
- Number and percentage of subjects with serious TEAEs,
- Number and percentage of subjects with serious drug related TEAEs,
- Number and percentage of subjects with TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with causally drug related TEAEs leading to permanent discontinuation of study drug,
- Number and percentage of subjects with NCI CTC Grade 3 or higher TEAEs,
- Number and percentage of subjects with TEAEs leading to death,
- Number and percentage of subjects with drug related TEAEs leading to death and
- Number of deaths occurring during the Safety Emergent Period and overall.

7.5.1.2 Proportion of subjects with TEAEs by SOC/PT

The number and percentage of subjects with TEAEs, as classified by SOC and PT will be summarized by treatment arm and type of ESA treatment. Summaries will be provided for:

- TEAEs,
- TEAEs (by PT only),
- Drug related TEAEs,
- TEAEs with NCI CTC Grade 3 or higher,
- TEAEs by severity,
- Drug related TEAEs by severity,
- Serious TEAEs,
- Drug related serious TEAEs,
- TEAEs leading to permanent discontinuation of study drug,
- Drug related TEAEs leading to permanent discontinuation of study drug,
- TEAEs leading to death,
- Drug-related TEAEs leading to death,
- TEAEs excluding serious adverse events that equal to or exceed a threshold of 5.0% in any treatment arm (roxadustat or ESA),
- Common TEAEs that equal to or exceed a threshold of 5.0% in any treatment arm (roxadustat or ESA).
- TEAEs by relationship to the study drug,
- AEs occurring during the post study follow up period,
- Serious AEs during the post-study follow up period.

For the summaries by severity or relationship to the study drug, in the subject count, if a subject has multiple TEAEs with the same SOC or PT, but with different severity or relationship, then the subject will be counted only once with the worst severity and highest degree of relationship, however, if any of the severity or relationship values are missing then the subject will be counted only once with missing severity or relationship. In the adverse event count, the adverse events will be presented in each category they were classified to. Drug related TEAEs will be presented in a similar way by severity only.

Maximum severity or relationship will be defined as the worst severity and highest degree of relationship on-treatment (see Section 6.2).

No queries will be done for SMQs.

7.5.1.3 Number of Events/100 Patient Exposure Year (PEY)

The number of events and number of events/Patient Exposure Year (PEY) during the Safety Emergent Period with TEAEs, as classified by SOC and PT will be calculated by treatment arm. Summaries will be provided for:

- TEAEs
- TEAE NCI CTC Grades 3 or higher
- Serious TEAEs,

7.5.1.4 Incidence Rates and Cumulative Incidence

Since the percentage of adverse events might be different between treatment arms due to the difference in the discontinuation rates, in addition to the frequency tables above, the incidence rate (per 100 subject years at risk) and the cumulative incidence at 0.5 year, 1 year, 1.5 years & 2 years with the 95% confidence interval, using Greenwood's formula, will be reported by treatment arm and type of ESA treatment.

It will be done for each of the following event types of special interest:

- Serious TEAEs,
- Deaths occurring during the Safety Emergent Period,
- All deaths during study period and post study safety follow up period,
- Related serious TEAEs,
- TEAEs leading to discontinuation of study drug, and
- TEAE NCI CTC Grades 3 or higher.

Hazard ratios of each of the TEAE categories of special interest above will be computed by using Cox Proportional Hazards regression stratified on region, history of CV, previous ESA treatment and adjusted on baseline Hb as continuous covariate. Hazard ratios (roxadustat as relative to ESA) and their 95% CI will be calculated for each TEAE category.

A dot-and-forest plot will be produced showing each of the above TEAE category on the y-axis. The incidence rates by treatment arm, the stratified Cox hazard ratios (roxadustat as relative to ESA) and their 95% CI will be shown on the x-axis.

In addition, cumulative incidence plots for subjects experiencing each of the TEAE categories above will be produced by treatment arm and type of ESA treatment.

Incidence rate (per 100 subject years at risk) for most common PT (>=5% in the total population) by SOC and the cumulative incidences of TEAE by SOC will also be produced for TEAEs in each treatment arm: Cumulative Incidence plots of subjects experiencing each SOC will be produced. A dot-and-forest plot will be produced showing each SOC in the y-axis and the incidence rates, the Cox hazard ratio and its 95% CI in the x-axis. The SOC will be sorted by the hazard ratio.

7.5.1.5 Sensitivity/Subgroup Analyses

- *Subgroups of interest:*

The number and percentage of subjects reporting TEAEs in each treatment arm will be tabulated by SOC and PT for the subgroups of interest defined in Section 7.8

7.5.1.6 Adverse events up to 7 days after last dose

Additional summaries or analyses with treatment emergent adverse events to be considered restricted to any event starting up to Analysis Date of Last Dose + 7 days considering the last dosing frequency:

- Overview summary table of subjects with AEs
- Number and percentage of subjects with AEs leading to death, as classified by PT

- Incidence rate for serious AEs (i.e., incidence rate, hazard ratios, and cumulative incidence plots as described in Section 7.5.1.4 will be provided)
- Incidence rate for AEs NCI CTC Grades 3 or higher (i.e., incidence rate, hazard ratios, and cumulative incidence plots as described in Section 7.5.1.4 will be provided)
- Number and percentage of subjects with AEs by SOC and PT.

Pre-specified adjudicated cardiovascular and thrombo-embolic events will be analyzed in meta-analyses across multiple phase 3 studies and compared between treatment groups. The statistical method for analysis of the composite safety endpoint (CSE) pooling studies will be detailed in a Meta-Analysis Statistical Analysis Plan. The results will be presented in a separate report.

7.5.2 Clinical Laboratory Evaluation

Descriptive statistics for laboratory values (in SI units) and changes from baseline at each assessment time point and for the maximum and minimum on-treatment (i.e., during Safety Emergent Period) value will be presented by treatment arm and treatment for the quantitative laboratory parameters.

Maximum and minimum on-treatment values will be determined using all the original values and not the derived windows.

Shift tables and number and percentage of subjects with shift to low and shift to high will be reported by treatment arm and type of ESA treatment for the quantitative laboratory parameters.

Box plots of quantitative laboratory values (in SI units) versus visit will be produced by treatment arm (two arms in one page) and type of ESA treatment (darbepoetin alfa and epoetin alfa in another page).

Summary by visit for qualitative laboratory parameters will be provided by treatment arm.

All clinical laboratory data will also be listed.

Potentially clinically significant (PCS) laboratory abnormalities

For each potentially clinically significant (PCS) criterion defined in Section 6.2.3.1, the percentage of subjects with abnormalities by visit and at any moment during the Safety Emergent Period and who did not meet the criteria at baseline will be reported by treatment arm and type of ESA treatment.

Incidence rate (per 100 subject years at risk) and the cumulative incidence at .5 year, 1 year, 1.5 years & 2 years with the 95% confidence interval using Greenwood's formula of PCS abnormalities will also be reported, using only subjects who did not meet the criteria at baseline, by treatment arm and type of ESA treatment. Risk of PCS abnormalities will be compared using the same Cox model as used on Section 7.5.1. Hazard ratio and its 95% will be calculated for the frequency of roxadustat as relative to ESA. A dot-and-forest plot will be produced showing the PCS abnormalities above in the y-axis and the incidence rates, the Cox hazard ratio and their 95% CI in the x-axis.

In addition, cumulative incidence plots for subjects experiencing each PCS abnormality will be produced by treatment arm (two arms in one page) and type of ESA treatment (darbepoetin alfa and epoetin alfa in another page).

7.5.2.1 Liver function tests

Descriptive summary of PCS values in Liver Enzymes and Total Bilirubin will be provided as per Astellas standard TLB_005.

In addition, a matrix scatter plot of Liver Enzymes and Bilirubin will be plotted showing the maximum ALT, AST, ALP and total bilirubin during the Safety Emergent Period crossed against each other. Different dots will be used for roxadustat and ESA.

Individual displays of Liver Enzymes and Bilirubin parameters, listed in Section 6.2.3.1, will be reported for all subjects with either ALT or AST > 3 x ULN or total bilirubin > 2 x ULN during the Safety Emergent Period.

For subjects who require further liver function investigations, additional information will be collected and listed.

7.5.3 Vital Signs

Descriptive and changes from baseline for vital signs (systolic blood pressure, diastolic blood pressure, respiratory rate, weight and pulse) at each assessment time point and for the maximum and minimum on-treatment (i.e during Safety Emergent Period) value will be presented by treatment arm and type of ESA treatment.

Maximum on-treatment value will be determined using all the original values and not the derived windows.

A plot for each parameter of mean (+/- 95% CI) versus visit will be produced by treatment arm (two arms in one page) and type of ESA treatment (darbepoetin alfa and epoetin alfa in another page) for pre-dialysis and post-dialysis values.

PCS Vital signs criteria (10 Combined) will be analyzed in the same way as explained in Section 7.5.2 for PCS laboratory abnormalities.

All vital signs data will also be listed.

7.5.4 Electrocardiograms (ECGs)

Descriptive and changes from baseline for ECG parameters (Pulse, PR Interval, RR Interval, QRS interval, QT interval, and QTc interval) at each assessment time point and for the maximum - on-treatment (i.e during Safety Emergent Period) value will be presented by treatment arm and type of ESA treatment.

Maximum on-treatment value will be determined using all the original values and not the derived windows.

The number and percentage of subjects with post-baseline PCS values (see Table 9 and Table 10) will be tabulated by treatment group and type of ESA treatment. The percentages are to be calculated relative to the number of subjects with available baseline and at least one

post-baseline assessment. The numerator will be total number of subjects with at least one post-baseline PCS ECG value. Shift tables may be presented.

The following Potentially Clinically Significant (PCS) QTc Criteria (both QTcB and QTcF) :

- QTc > 500 msec
- Change from baseline in QTc > 60 msec

will be analyzed in the same way as explained in Section 7.5.2 for PCS laboratory abnormalities.

A plot for each parameter of mean (+/- 95% CI) versus visit will be produced by treatment arm (two arms in one page) and type of ESA treatment (darbepoetin alfa and epoetin alfa in another page).

In addition, ECG parameters will be reported according to Astellas standards TEG_003 and TEG_004.

All ECG data will also be listed.

7.5.5 Vascular Access Thrombosis (VAT)

Time to first VAT will be analyzed similarly as in Section 7.5.1.4 (Time to event analysis for TEAEs).

7.5.6 Pregnancies

A listing of pregnancy test results will be provided.

7.6 Analysis of PK

The statistical methods for PK data will be described in a separate analysis plan. Results of the population PK analysis will not be reported in the Clinical Study Report but in a separate population PK report.

Plasma concentration data of roxadustat will be listed. In addition, concentration of roxadustat in dialysates will also be listed.

7.7 Analysis of PD

Not Applicable.

7.8 Subgroups of Interest

Selected efficacy and safety endpoints will be summarized for the subgroups defined on the basis of the categorized variables listed below in Table 26

Table 26 Subgroups of Interest

Grouping variables	Subgroups
Age group	< 65 years 65 - 74 years ≥ 75 years
Sex	Female Male
Previous ESA treatment	Darbepoetin alfa Epoetin
Region	Western Europe Central and Eastern Europe
Baseline Hb	≤11 g/dL >11 g/dL
Dialysis type*	Hemodialysis Peritoneal dialysis
History of cardiovascular, cerebrovascular or thromboembolic diseases [#]	Yes No
Baseline hs-CRP	<=ULN >ULN

*: Only for primary Hb analyses.

[#]: This subgroup is derived according to a list of cardiovascular disease preferred terms received from the study physician after softlock.

7.9 Other Analyses

No other analyses are planned.

7.10 Interim Analysis (and Early Discontinuation of the Clinical Study)

The study will have no interim analysis with statistical inference. Safety data and dosing decisions will be monitored on an ongoing basis. Ongoing review of safety data will be completed by an independent Data and Safety Monitoring Board (DSMB).

7.11 Handling of Missing Data, Outliers, Visit Windows, and Other Information

7.11.1 Missing Data

For relevant analyses without rescue therapy, for subjects who used rescue therapy, the reported Hb values after the initiation of rescue therapy will be set to missing (instead of the reported values) for 6 weeks from the start date of rescue therapy (or the end in case the duration of rescue therapy > 1 day).

For response variables, if no Hb value is available during the reference period, an imputation technique will be used, with the mean of all available values from Day 1 to minimum (End of Efficacy Emergent Period, End of the reference period) carried forward.

The following imputations will be performed for the continuous endpoints, unless specified otherwise:

- An MMRM model will be run for the purpose of implicit imputation of missing data by using all the available information from the observed data via the within-patient correlation structure for continuous endpoints with inferential analysis.
- An ANCOVA model with multiple imputations (MI) will also be run (refer to Section 7.4.1.2.1).
- PMM models with different missingness patterns assumptions will be run, in addition (refer to Section 7.4.1.1.2).

7.11.2 Missing Dates

As a general rule, the worst case scenario imputation rule is usually used. A start date is generally imputed to the first possible day, unless the available information in the partly missing date is equal to the one in the reference date. In this case, the substituted date is set to the reference date. An end date is generally imputed to the last possible day.

Completely missing dates will not be imputed.

Diagnosis of anemia, CKD and Targeted Medical History

The following rules will be applied to impute partially missing dates of diagnosis of Anemia, CKD and targeted medical history, as defined in Table 27 below.

Table 27 Definitions of the Analysis Date of Diagnosis of Anemia, CKD and Targeted Medical History

Reported Date (from the eCRF)	Analysis Date (Derived)
--/MM/YYYY	01/MM/YYYY
--/--/YYYY	01/01/YYYY
DD/--/----, or --/MM/----, or --/--/----	No imputation

Previous or Concomitant medication:

For previous or concomitant medications, including rescue medications, partially missing start dates and/or stop dates will be imputed as defined in Table 28 and Table 29 below:

Table 28 Definitions of the Previous or Concomitant Medication Analysis Start Date

Reported Date (from the eCRF)	Analysis Date (Derived)
--/MM/YYYY	01/MM/YYYY
--/--/YYYY	01/01/YYYY
DD/--/----, or --/MM/----, or --/--/----	No imputation

If the imputed start date is after the stop date, then the imputed start date will be one day prior to the stop date.

Table 29 Definitions of the Previous or Concomitant Medication Analysis Stop Date

Reported Date (from the eCRF)	Analysis Date (Derived)
--/MM/YYYY	31/MM/YYYY, or 30/MM/YYYY, or 29/MM/YYYY, or 28/MM/YYYY
--/--/YYYY	31/12/YYYY
DD/--/----, or --/MM/----, or --/--/----	No imputation

AE Onset date

For adverse events, partially missing start dates and/or stop dates will be imputed as defined in Table 30 and Table 31 below:

Table 30 Definitions of the Analysis Adverse Event Onset Date

Reported Date	Date of First Drug Intake	Analysis Date (Derived)
--/MM/YYYY --/02/2008 --/02/2008 --/02/2008	DD/MM/YYYY 14/02/2008 14/02/2007 14/02/2009	 14/02/2008* 01/02/2008 01/02/2008
--/--/YYYY --/--/2008 --/--/2008 --/--/2008	DD/MM/YYYY 14/02/2008 14/02/2007 14/02/2009	 14/02/2008 01/01/2008 01/01/2008
DD/--/---- --/MM/---- --/--/----		No imputation

* If the month and year is the same as the month and year of first drug intake, use date of the first drug intake.

Table 31 Definitions of the Analysis Adverse Event End Date

Reported Date	Analysis Date (Derived) *
--/MM/YYYY	31/MM/YYYY or 30/MM/YYYY or 29/MM/YYYY or 28/MM/YYYY
--/--/YYYY	31/12/YYYY
DD/--/----, or --/MM/----, or --/--/----	No imputation

*Death has to be taken into consideration when calculating this.

7.11.3 Outliers

As a general rule, all values, including outliers will be analyzed.

7.11.4 Visits Windows

The study protocol gives the overall study schedule and the permissible intervals for visits expressed as the number of days relative to the first study medication date (Day 1).

For all study assessments reported by visit, the value which assessment day is the latest collected within the corresponding analysis visit window will be used. If more than one value is collected that day, then the latest value will be used in the analysis.

Analysis Visit windows, as depicted in Table 32 below, will be used for the following study assessments reported by visit:

- Central laboratory parameters (except Lipid Panel),
- Vital Signs,
- ECG parameters.

Table 32 Analysis Visit Windows

CRF Visit	Target Day[a]	Analysis Visit Windows Actual Assessment Day	Analysis Visit
Screening		Day -42 to Day -1	Screening
Day 1	Day 1	Day 1	Baseline
Week 1	Day 7 * (Week #) + 1	Day 2 to Day 11	Week 1
Week 2 – 7	Day 7 * (Week #) + 1	[Target Day ± 3]	Week 2 - 7
Week 8	Day 7 * (Week #) + 1	[Target Day – 3, Target Day + 6]	Week 8
Week 10 - 34	Day 7 * (Week #) + 1	[Target Day – 7, Target Day +6]	Week 10 - 34
Week 36	Day 7 * (Week #) + 1	[Target Day – 7, Target Day +13]	Week 36
Week 40 – 100	Day 7 * (Week #) + 1	[Target Day – 14, Target Day + 13]	Week 40 - 100
Week 104/EOT (for completers)	Day 7 * (Week #) + 1	[Target Day – 14, Target Day + 28]	Week 104
EOT Visit (for premature discontinuations)	NA	NA	Analysis Visit corresponding to the actual visit window
EOT + 2 Weeks Visit	14 Days after EOT Visit date	[DAY of EOT + 1, DAY of EOT +20]	EOT + 2 Weeks
EOS Visit	28 Days after EOT Visit date	[DAY of EOT + 21, DAY of EOT + 31]	EOS
Unscheduled	NA	NA	Analysis Visit corresponding to the actual visit window

[a]: Relative to Day 1 (first dose date of study medication)

Analysis Visit windows, as depicted in Table 33 below, will be used for the quality of life efficacy study assessments:

Table 33 Analysis Visit Windows for QoL

CRF Visit	Target Day[a]	Analysis Visit Windows Actual Assessment Day	Analysis Visit
Day 1	Day 1	Day 1	Baseline
Week 8	Day 7 * (Week #) + 1	[Target Day – 14, Target Day + 13]	Week 8
Week 12	Day 7 * (Week #) + 1	[Target Day – 14, Target Day + 27]	Week 12
Week 28	Day 7 * (Week #) + 1	[Target Day – 28, Target Day + 27]	Week 28
Week 36	Day 7 * (Week #) + 1	[Target Day – 28, Target Day +27]	Week 36
Week 52	Day 7 * (Week #) + 1	[Target Day – 56, Target Day + 83]	Week 52
Week 76	Day 7 * (Week #) + 1	[Target Day – 84, Target Day + 97]	Week 76
Week 104/EOT (for completers)	Day 7 * (Week #) + 1	[Target Day – 98, Target Day + 28]	Week 104/EOT
EOT Visit (for premature discontinuations)	NA	NA	Analysis Visit corresponding to the actual visit window
Unscheduled	NA	NA	Analysis Visit corresponding to the actual visit window

[a]: Relative to Day 1 (first dose date of study medication)

Analysis Visit windows, as depicted in Table 34 below, will be used for the Lipid Panel, including LDL cholesterol efficacy study assessment:

Table 34 Analysis Visit Windows for Lipid Panel

CRF Visit	Target Day[a]	Analysis Visit Windows Actual Assessment Day	Analysis Visit
Day 1	Day 1	Day 1	Baseline
Week 4	Day 7 * (Week #) + 1	Day 2 to Day 42	Week 4
Week 8	Day 7 * (Week #) + 1	[Target Day – 14, Target Day + 13]	Week 8
Week 12	Day 7 * (Week #) + 1	[Target Day – 14, Target Day + 27]	Week 12
Week 20	Day 7 * (Week #) + 1	[Target Day – 28, Target Day + 27]	Week 20
Week 28	Day 7 * (Week #) + 1	[Target Day – 28, Target Day + 27]	Week 28
Week 36	Day 7 * (Week #) + 1	[Target Day – 28, Target Day + 27]	Week 36
Week 44	Day 7 * (Week #) + 1	[Target Day – 28, Target Day + 27]	Week 44
Week 52	Day 7 * (Week #) + 1	[Target Day – 28, Target Day + 55]	Week 52
Week 68	Day 7 * (Week #) + 1	[Target Day – 56, Target Day + 55]	Week 68
Week 84	Day 7 * (Week #) + 1	[Target Day – 56, Target Day + 69]	Week 84
Week 104/EOT (for completers)	Day 7 * (Week #) + 1	[Target Day – 70, Target Day + 28]	Week 104/EOT
EOT Visit (for premature discontinuations)	NA	NA	Analysis Visit corresponding to the actual visit window
EOS Visit	NA	Last assessment between Day 2 and study termination day	EOS
Unscheduled	NA	NA	Analysis Visit corresponding to the actual visit window

[a]: Relative to Day 1 (first dose date of study medication)

For the MMRM analyses, which require one value per visit, one analysis value for each visit will be used.

For the ANCOVA analyses, which use the average, all available values in the analysis windows will be used for the calculation.

7.11.5 End of Safety Emergent Period

The end of Safety Emergent Period will be defined as:

- minimum between [(Analysis Date of last dose + 28 days + x), EOS visit, Date of death], with x corresponding to additional days based on the last dosing frequency.

The value of x days is defined to standardize to 3 times per week,:

if frequency = “3 TIMES PER WEEK” then x=0

if frequency = “2 TIMES PER WEEK” then x=1

if frequency = “1 TIME PER WEEK” then x=5

if frequency = “EVERY 2 WEEKS” then x=12

if frequency = “QM” then x =26 days.

With Analysis Date of Last Dose defined in Section 6.5.4

7.11.6 End of Efficacy Emergent Period

For all subjects, the end of Efficacy Emergent Period is defined as the treatment period up to the EOT Visit. Only exception for subjects who discontinued due to death, the last Hb assessment is taken as per convention the date of death is the EOT visit.

8 DOCUMENT REVISION HISTORY

Version	Date	Changes	Comment/rationale for change
1.0	04-Dec-2014	NA	Document finalized
2.0	22-Mar-2016	1) Reduction of the number of sensitivity analyses for the secondary endpoints. 2) Revise the ordering for the key secondary endpoints by moving up the QoL endpoints. 3) Implementation of the Time to event approach for the PPS and adjustment of the relevant sections. 4) Change from ITT to All Randomized as per Astellas standard. Definition remains unchanged. 5) Use of the Safety Emergent Period (i.e., last dose + 28 days) as evaluation period by default for the safety endpoints. 6) Use of the Efficacy Emergent Period (i.e last dose + 7 days) as evaluation period by default for most of the efficacy time to event endpoints. 7) Use of "Time to censoring" instead of "time at risk" for patient with no event for more clarity. 8) Clarification of time to censoring for events evaluated during the Safety Emergent Period/Efficacy Emergent Period. 9) Time to PCS vital signs added. 10) Time to PCS ECG focused to QTc instead of all ECG assessments. 11) Update regarding derivation of MAP values: average of the 3 measurements instead of 2 (now in line with FG SAP). 12) Implementation of the primary efficacy endpoint for the US	1) To limit the additional analyses. 2) Due to the importance and relevance of the QoL endpoints 3) PPS definition has been revised in order to limit the exclusion of data by using a time to event approach rather than creating one PPS set for each period of interest. 4) Astellas standards requirement. 5) Clarify a consistent approach for all the safety endpoints by defining the evaluated period up to 28 days after last dose, which matched the AE analyses. 6) Clarify a consistent approach for time to event efficacy endpoints by extending the evaluated period up to 7 days after last dose. 7) Clarification of the wording. Time to censoring is more appropriate. 8) Based on the definitions of Efficacy/Safety Emergent period, time to censoring for all time to events have been updated accordingly. 9) Time to PCS were already planned for lab and ECG but not for Vital signs. This was added to maintain a consistent approach. 10) Time to event for PCS ECG other than QTc was considered not of interest. 11) Using the average of all 3 measurements is more efficient and is consistent with the other studies 12) This specific endpoint has

Version	Date	Changes	Comment/rationale for change
		submission (change in Hb from baseline to the average level between week 28 and week 52) with the description of the statistical analyses including sensitivity analyses for missing data.	been added in protocol amendment 1.0.Sensitivity analyses for missing data were added following FDA feedback on the Fibrogen 064 study.
		13) Interim analysis changed from week 36 to week 52.	13) This has been changed in protocol amendment 1.0.
		14) Implementation of time to first hospitalization and time to first potential EH .	14) Due to harmonization with Fibrogen 064 study SAP and the expected difference in treatment durations
		15) Update regarding the analysis on cholesterol data: use of all data as main analysis and fasted only as additional..	15) Due to the assumption that most samples were not taken fasted and harmonization with Fibrogen 064 study SAP.
		16) Reduction in the number of analyses of subgroup of interests.	16) To limit the total number of analyses and the volume of tables
		17) Addition of the analysis of percentage of time when Hb > 12, 13 or 14.	17) The concept of percentage of time with high Hb values was implemented in order to account for differences in individual treatment durations.
		18) MMRM model added for number of RBC and volume of RBC during the treatment period.	18) Inferential analysis for RBC was missing in the v1.0.
		19) Removal of descriptive analysis on the categorical change form baseline in Total FACT-An Score.	19) Descriptive statistics for categorical change not necessary since analysis on absolute change already planned and considered sufficient.
		20) Addition of an analysis on a subset of subjects with baseline PF subscore <40 and ≥40, and baseline VT subscore <50 and ≥ 50.	20) Due to harmonization with Fibrogen 064 study SAP.
		21) Updated rule for complete missing start date and end date for concomitant medications. In that case, the medication will be considered as both previous and concomitant instead of only concomitant.	21) Previous rule considered too conservative since a missing year of start is usually for previous medications than concomitant. Decision was taken to consider such medications as both previous and concomitant.
		22) For the analysis of monthly IV Iron use, MMRM has been replaced by ANCOVA.	22) MMRM model for monthly IV Iron was not adapted due the data distribution. Since monthly average will be calculated, ANCOVA is more appropriate.
		23) Additional tables run on group of subjects enrolled after protocol	23) Added in order to assess the

Version	Date	Changes	Comment/rationale for change
		amendment implementation 24) Addition of response endpoint up to week 52, regardless of rescue therapy in the additional secondary endpoint 25) Addition of a week 36 to 52 endpoints, in the additional secondary endpoints for Quality of Life. 26) Removal of ANCOVA with LOCF for the primary endpoint sensitivity analysis, and addition of ANCOVA with MI and PMM.	impact of substantial protocol amendment 1.0. 24) Corresponding with the endpoint for the US 25) Clinically relevant to have an endpoint to measure long term effect on quality of life. 26) Sensitivity analyses for missing data were added following FDA feedback on the Fibrogen 064 study.
3.0	10-Sep-2018	1) Updating the safety and efficacy periods 2) Additional Hb response definition as the average Hb>=10. 3) Add MMRM analyses regardless of rescue therapy using FAS for primary EU endpoint 4) Removed analysis using all Hb values using All Randomized. 5) Added time to first IV iron during day 1 to week 36 6) Added the analysis of Time to first VAT 7) Details were added in All Randomized section regarding exclusion of subjects from site [redacted] 8) Additional analyses on adverse event were added considering events occurring up to last dose +7 days 9) The unique identifiers of protocol deviations are updated 10) Removed other baseline factors except baseline Hb and randomization stratification factors in ANCOVA with MI	1) Adjusting for the different dosing frequencies, especially for darbepoetin. 2) To include Hb values above 12.0 (e.g. 12.3) also as a good response. 3) Harmonization of Phase 3 studies. 4) Duplicate analysis 5) Majority of subjects did not use IV iron; therefore, time to first use proposed as alternative to the actual dose of IV iron. 6) Included VAT AE specific safety analysis. 7) Site [redacted] was terminated due to GCP violations and it is now clearly mentioned in the population section that those subjects are excluded from all analysis set. 8) Due to harmonization with Fibrogen 064 study SAP. 9) Updated according to latest protocol deviations 10) Per team discussion, other baseline factors will not be included for ANCOVA with MI

9 REFERENCES

Cella D. The functional assessment of Cancer Therapy-Anemia (FACT-An) Scale: A new tool for the assessment of outcome in cancer anemia and fatigue. Hematology Seminars 1997; 34: 13-19

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.

Ge M, Durham LK, Meyer RD, Xie W and Thomas N. Covariate-Adjusted Difference in Proportions from Clinical Trials Using Logistic Regression and Weighted Risk Differences. Drug Information Journal 2011; 45: 481

ICH Harmonized Tripartite Guideline E 3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E 9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

Miettinen, O. and Nurminen, M., Comparative Analysis of Two Rates*, Statistics in Medicine*, 4, 213–226, 1985

Millen BA, Dmitrienko A. Chain procedures: A class of flexible closed testing procedures with clinical trial applications. Statistics in Biopharmaceutical Research. 2011;3:14-30.

Spiessens B, Debois M. Adjusted significance levels for subgroup analyses in clinical trials. Contemporary Clinical Trials, Volume 31, Issue 6; November 2010.

Webster K, Cella D and Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Heath and Quality of Life Outcomes 2003; 1: 79.

10 APPENDICES

10.1 Appendix 1: SF-36 v2

Your Health and Well-Being

This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities. *Thank you for completing this survey!*

For each of the following questions, please tick the one box that best describes your answer.

1. **In general, would you say your health is:**

Excellent	Very good	Good	Fair	Poor
▼	▼	▼	▼	▼
☐1	☐2	☐3	☐4	☐5

2. **<u>Compared to one year ago</u>, how would you rate your health in general <u>now?</u>**

Much better now than one year ago	Somewhat better now than one year ago	About the same as one year ago	Somewhat worse now than one year ago	Much worse now than one year ago
▼	▼	▼	▼	▼
☐1	☐2	☐3	☐4	☐5

SF-36 v2

3. The following questions are about activities you might do during a typical day. Does <u>your health now limit you</u> in these activities? If so, how much?

	Yes, limited a lot	Yes, limited a little	No, not limited at all
a Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	☐ 1	☐ 2	☐ 3
b Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	☐ 1	☐ 2	☐ 3
c Lifting or carrying groceries	☐ 1	☐ 2	☐ 3
d Climbing several flights of stairs	☐ 1	☐ 2	☐ 3
e Climbing one flight of stairs	☐ 1	☐ 2	☐ 3
f Bending, kneeling, or stooping	☐ 1	☐ 2	☐ 3
g Walking more than a mile	☐ 1	☐ 2	☐ 3
h Walking several hundred yards	☐ 1	☐ 2	☐ 3
i Walking one hundred yards	☐ 1	☐ 2	☐ 3
j Bathing or dressing yourself	☐ 1	☐ 2	☐ 3

SF-36 v2

4. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a Cut down on the amount of time you spent on work or other activities	☐1	☐2	☐3	☐4	☐5
b Accomplished less than you would like	☐1	☐2	☐3	☐4	☐5
c Were limited in the kind of work or other activities	☐1	☐2	☐3	☐4	☐5
d Had difficulty performing the work or other activities (for example, it took extra effort)	☐1	☐2	☐3	☐4	☐5

5. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a Cut down on the amount of time you spent on work or other activities	☐1	☐2	☐3	☐4	☐5
b Accomplished less than you would like	☐1	☐2	☐3	☐4	☐5
c Did work or other activities less carefully than usual	☐1	☐2	☐3	☐4	☐5

SF-36 v2

6. **During the <u>past 4 weeks</u>, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?**

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

7. **How much <u>bodily</u> pain have you had during the <u>past 4 weeks</u>?**

None	Very mild	Mild	Moderate	Severe	Very severe
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

8. **During the <u>past 4 weeks</u>, how much did <u>pain</u> interfere with your normal work (including both work outside the home and housework)?**

Not at all	A little bit	Moderately	Quite a bit	Extremely
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

SF-36 v2

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks…

	All of the time	Most of the time	Some of the time	A little of the time	None of the time
a Did you feel full of life?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b Have you been very nervous?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c Have you felt so down in the dumps that nothing could cheer you up?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d Have you felt calm and peaceful?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
e Did you have a lot of energy?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
f Have you felt downhearted and low?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
g Did you feel worn out?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
h Have you been happy?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
i Did you feel tired?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

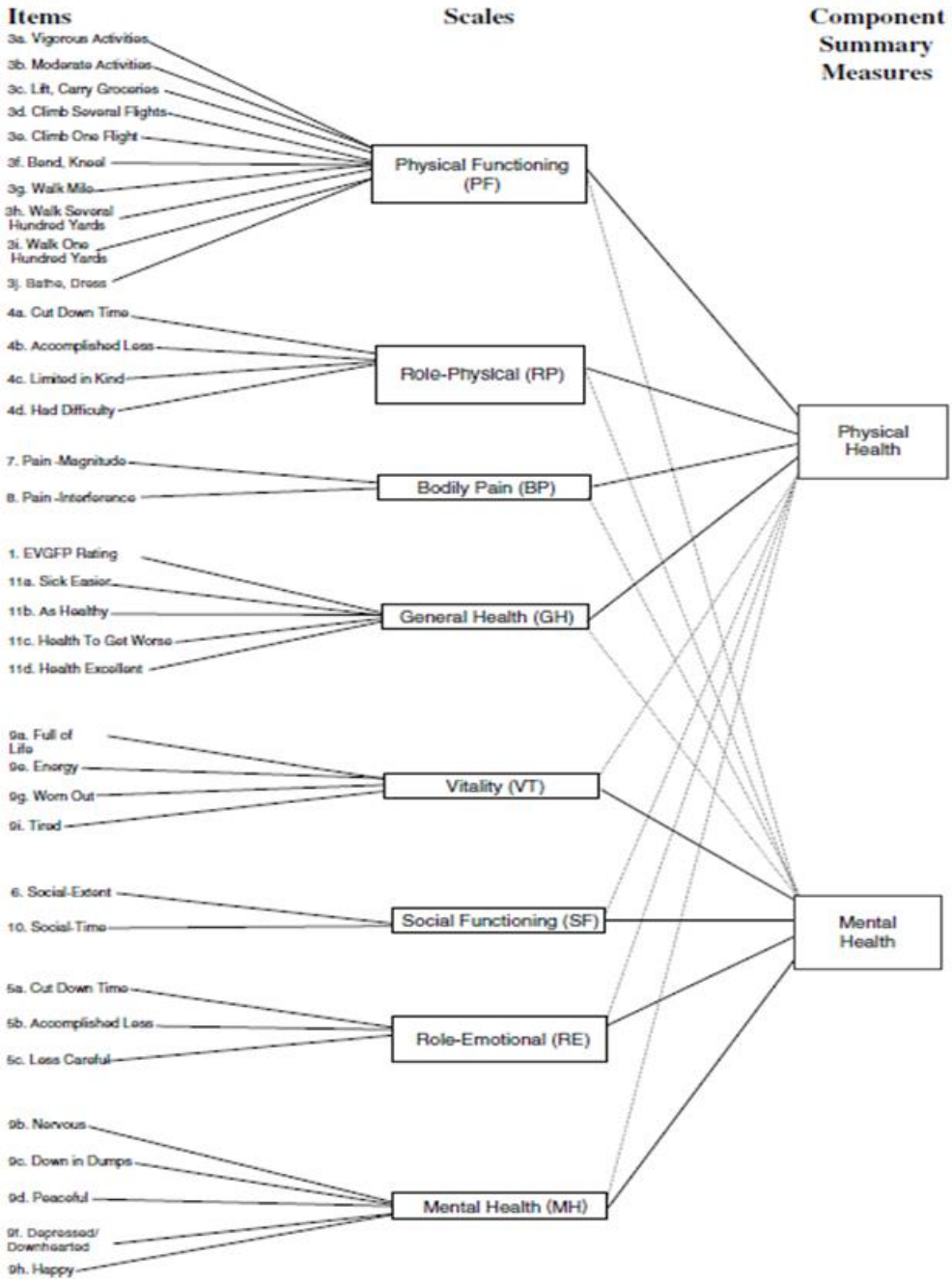
SF-36 v2

11. How TRUE or FALSE is <u>each</u> of the following statements for you?

	Definitely true	Mostly true	Don't know	Mostly false	Definitely false
a I seem to get ill more easily than other people	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b I am as healthy as anybody I know	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c I expect my health to get worse	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d My health is excellent	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Thank you for completing these questions!

SF-36 Model


Items
Scales
Component Summary Measures
3a. Vigorous Activities
3b. Moderate Activities
3c. Lift, Carry Groceries
3d. Climb Several Flights
3e. Climb One Flight
3f. Bend, Kneel
3g. Walk Mile
3h. Walk Several Hundred Yards
3i. Walk One Hundred Yards
3j. Bathe, Dress
Physical Functioning (PF)
4a. Cut Down Time
4b. Accomplished Less
4c. Limited in Kind
4d. Had Difficulty
Role-Physical (RP)
7. Pain -Magnitude
8. Pain -Interference
Bodily Pain (BP)
1. EVGFP Rating
11a. Sick Easier
11b. As Healthy
11c. Health To Get Worse
11d. Health Excellent
General Health (GH)
9a. Full of Life
9e. Energy
9g. Worn Out
9i. Tired
Vitality (VT)
6. Social-Extent
10. Social-Time
Social Functioning (SF)
5a. Cut Down Time
5b. Accomplished Less
5c. Less Careful
Role-Emotional (RE)
9b. Nervous
9c. Down in Dumps
9d. Peaceful
9f. Depressed/ Downhearted
9h. Happy
Mental Health (MH)
Physical Health
Mental Health

10.2 Appendix 2: FACT-An (Version 4)

10.2.1 FACT-An Questionnaire

Below is a list of statements that other people with your illness have said are important.
Please circle or mark one number per line to indicate your response as it applies to the past 7 days.

	PHYSICAL WELL-BEING	Not at all	A little bit	Some-what	Quite a bit	Very much
GP1	I have a lack of energy	0	1	2	3	4
GP2	I have nausea	0	1	2	3	4
GP3	Because of my physical condition, I have trouble meeting the needs of my family	0	1	2	3	4
GP4	I have pain	0	1	2	3	4
GP5	I am bothered by side effects of treatment	0	1	2	3	4
GP6	I feel ill	0	1	2	3	4
GP7	I am forced to spend time in bed	0	1	2	3	4

	SOCIAL/FAMILY WELL-BEING	Not at all	A little bit	Some-what	Quite a bit	Very much
GS1	I feel close to my friends	0	1	2	3	4
GS2	I get emotional support from my family	0	1	2	3	4
GS3	I get support from my friends	0	1	2	3	4
GS4	My family has accepted my illness	0	1	2	3	4
GS5	I am satisfied with family communication about my illness	0	1	2	3	4
GS6	I feel close to my partner (or the person who is my main support)	0	1	2	3	4
Q1	*Regardless of your current level of sexual activity, please answer the following question. If you prefer not to answer it, please mark this box ☐ and go to the next section.*					
GS7	I am satisfied with my sex life	0	1	2	3	4

FACT-An (Version 4)

Please circle or mark one number per line to indicate your response as it applies to the past 7 days.

	EMOTIONAL WELL-BEING	Not at all	A little bit	Some-what	Quite a bit	Very much
GE1	I feel sad	0	1	2	3	4
GE2	I am satisfied with how I am coping with my illness	0	1	2	3	4
GE3	I am losing hope in the fight against my illness	0	1	2	3	4
GE4	I feel nervous	0	1	2	3	4
GE5	I worry about dying	0	1	2	3	4
GE6	I worry that my condition will get worse	0	1	2	3	4

	FUNCTIONAL WELL-BEING	Not at all	A little bit	Some-what	Quite a bit	Very much
GF1	I am able to work (include work at home)	0	1	2	3	4
GF2	My work (include work at home) is fulfilling	0	1	2	3	4
GF3	I am able to enjoy life	0	1	2	3	4
GF4	I have accepted my illness	0	1	2	3	4
GF5	I am sleeping well	0	1	2	3	4
GF6	I am enjoying the things I usually do for fun	0	1	2	3	4
GF7	I am content with the quality of my life right now	0	1	2	3	4

FACT-An (Version 4)

Please circle or mark one number per line to indicate your response as it applies to the past 7 days.

	ADDITIONAL CONCERNS	Not at all	A little bit	Some-what	Quite a bit	Very much
HI7	I feel fatigued	0	1	2	3	4
HI12	I feel weak all over	0	1	2	3	4
An1	I feel listless ("washed out")	0	1	2	3	4
An2	I feel tired	0	1	2	3	4
An3	I have trouble starting things because I am tired	0	1	2	3	4
An4	I have trouble finishing things because I am tired	0	1	2	3	4
An5	I have energy	0	1	2	3	4
An6	I have trouble walking	0	1	2	3	4
An7	I am able to do my usual activities	0	1	2	3	4
An8	I need to sleep during the day	0	1	2	3	4
An9	I feel lightheaded (dizzy)	0	1	2	3	4
An10	I get headaches	0	1	2	3	4
B1	I have been short of breath	0	1	2	3	4
An11	I have pain in my chest	0	1	2	3	4
An12	I am too tired to eat	0	1	2	3	4
BL4	I am interested in sex	0	1	2	3	4
An13	I am motivated to do my usual activities	0	1	2	3	4
An14	I need help doing my usual activities	0	1	2	3	4
An15	I am frustrated by being too tired to do the things I want to do	0	1	2	3	4
An16	I have to limit my social activity because I am tired	0	1	2	3	4

10.2.2 FACT-An Scoring Guidelines

Instructions:* 1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.
4. Add subscale scores to derive total scores (TOI, FACT-G & FACT-An).
5. **The higher the score, the better the QOL.**

Subscale	Item Code	Reverse item?		Item response	Item Score
PHYSICAL WELL-BEING (PWB)	GP1	4	-	_____	=_____
	GP2	4	-	_____	=_____
	GP3	4	-	_____	=_____
	GP4	4	-	_____	=_____
Score range: 0-28	GP5	4	-	_____	=_____
	GP6	4	-	_____	=_____
	GP7	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____
=**PWB subscale score**

SOCIAL/FAMILY WELL-BEING (SWB)	GS1	0	+	_____	=_____
	GS2	0	+	_____	=_____
	GS3	0	+	_____	=_____
	GS4	0	+	_____	=_____
Score range: 0-28	GS5	0	+	_____	=_____
	GS6	0	+	_____	=_____
	GS7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____
=**SWB subscale score**

EMOTIONAL WELL-BEING (EWB)	GE1	4	-	_____	=_____
	GE2	0	+	_____	=_____
	GE3	4	-	_____	=_____
	GE4	4	-	_____	=_____
Score range: 0-24	GE5	4	-	_____	=_____
	GE6	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 6: _____
Divide by number of items answered: _____
=**EWB subscale score**

FUNCTIONAL WELL-BEING (FWB)	GF1	0	+	_____	=_____
	GF2	0	+	_____	=_____
	GF3	0	+	_____	=_____
	GF4	0	+	_____	=_____
	GF5	0	+	_____	=_____
Score range: 0-28	GF6	0	+	_____	=_____
	GF7	0	+	_____	=_____

Sum individual item scores: _____
Multiply by 7: _____
Divide by number of items answered: _____
=**FWB subscale score**

Subscale	**Item Code**	**Reverse item?**		**Item response**	**Item Score**
ANEMIA SUBSCALE (AnS)	HI7	4	-	_____	=_____
	HI12	4	-	_____	=_____
	An1	4	-	_____	=_____
	An2	4	-	_____	=_____
	An3	4	-	_____	=_____
Score range: 0-80	An4	4	-	_____	=_____
	An5	0	+	_____	=_____
	An6	4	-	_____	=_____
	An7	0	+	_____	=_____
	An8	4	-	_____	=_____
	An9	4	-	_____	=_____
	An10	4	-	_____	=_____
	B1	4	-	_____	=_____
	An11	4	-	_____	=_____
	An12	4	-	_____	=_____
	BL4	0	+	_____	=_____
	An13	0	+	_____	=_____
	An14	4	-	_____	=_____
	An15	4	-	_____	=_____
	An16	4	-	_____	=_____

Sum individual item scores:_____
Multiply by 20: _____
Divide by number of items answered: _____
=**An Subscale score**

To derive a FACT-An Trial Outcome Index (TOI):
Score range: 0-136

_____ + _____ + _____ =_____=**FACT-An TOI**
(**PWB** score) (**FWB** score) (**AnS** score)

To Derive a FACT-G total score:
Score range: 0-108

_____ + _____ + _____ + _____=_____=**FACT-G Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score)

To Derive a FACT-An total score:
Score range: 0-188

_____ + _____ + _____ + _____ + _____ =_____=**FACT-An Total score**
(**PWB** score) (**SWB** score) (**EWB** score) (**FWB** score) (**AnS** score)

*For guidelines on handling missing data and scoring options, please refer to the Administration and Scoring Guidelines in the manual or on-line at www facit.org.

FACIT-Fatigue Subscale Scoring Guidelines

Instructions:* 1. Record answers in "item response" column. If missing, mark with an X
2. Perform reversals as indicated, and sum individual items to obtain a score.
3. Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.
4. The higher the score, the better the QOL.

Subscale Score	Item Code	Reverse item?		Item response	Item
FATIGUE	HI7	4	-	_____	=_____
SUBSCALE	HI12	4	-	_____	=_____
	An1	4	-	_____	=_____
	An2	4	-	_____	=_____
Score range: 0-52	An3	4	-	_____	=_____
	An4	4	-	_____	=_____
	An5	0	+	_____	=_____
	An7	0	+	_____	=_____
	An8	4	-	_____	=_____
	An12	4	-	_____	=_____
	An14	4	-	_____	=_____
	An15	4	-	_____	=_____
	An16	4	-	_____	=_____

Sum individual item scores: _____
Multiply by 13: _____
Divide by number of items answered: _____
=**Fatigue Subscale score**

10.3 Appendix 3: EQ-5D 5L v2

EQ-5D 5L v2

10.4 Appendix 4: Patient Overall Impression of Change

Since the start of the study, my general state of health is:

tick one box only

- ☐ [1] Very Much Improved
- ☐ [2] Much Improved
- ☐ [3] Minimally Improved
- ☐ [4] No Change
- ☐ [5] Minimally Worse
- ☐ [6] Much Worse
- ☐ [7] Very Much Worse

10.5 Appendix 5: Medication WHO Drug Dictionary Codes

Name	Code
ESA except darbepoetin alfa	ATC level 4 = B03XA
Darbepoetin alfa	ATC level 4 = B03XA
IV Iron or Oral Iron	ATC Level 3 = “IRON PREPARATIONS” and Route = “INTRAVENOUS” or “ORAL”
RBC transfusion	ATC level 4 = B05AX
Any investigational drug	WHODD drug code = 99999701001
Hypoxia-inducible factor HIF-PHI	ATC level 4 = B03XA
Iron-chelating agents	ATC level 4 = B03AA, B03AD, B03AE, A12CX
Androgens	ATC level 3 = G03B and G03E
Dapsone	ATC level 4 = J04BA
Acetaminophen/paracetamol	ATC level 4 = N02BE

10.6 Appendix 6: Signatures

List of Key Contributors and Approvers

Key Contributors

The following contributed to or reviewed this Statistical Analysis Plan as relevant to their indicated discipline or role.

Primary author

[redacted]	[redacted]

Contributors and Reviewers

[redacted]	[redacted]
[redacted]	Data Science
[redacted]	Data Science
[redacted]	Clinical Science
[redacted]	Medical Science
[redacted]	Medical Science
[redacted]	Medical Science
[redacted]	[redacted]

Approver Signatories

(E-signatures are attached at end of document)

[redacted], was the study statistician for this study.

[redacted] was the Global Statistician Leader and biostatistics peer reviewer of this Statistical Analysis Plan

I approve the contents of this Statistical Analysis Plan:

[redacted]
Medical Science – Nephrology
Astellas Pharma Europe B.V.