NCT # NCT02052310

**FibroGen, Inc.**

**Protocol Number: FGCL-4592-063**

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

Protocol Amendment 4

# STATISTICAL ANALYSIS PLAN

**Version:** **V1.0**
**Release Date:** **14Oct2018**

**Approvals**

I have reviewed and accept the information in this document to be a true and accurate representation of the Statistical Analysis Plan for Study FGCL-4592-063.

**Initiator:**

Signature: [redacted] Date: [redacted]

**Reviewed by:**

Signature: [redacted] Date: [redacted]

Signature: [redacted] Date: [redacted]

Signature: [redacted] Date: [redacted]

Signature: [redacted] Date: [redacted]

Signature: [redacted] Date: [redacted]

**Signature Significance**

The following significance is lent to the signatures on the *Approvals* page of this document.

| Signature | Significance |
|---|---|
| **Author** | By signing, the author is attesting that the content of the document is complete and accurate. |
| **Reviewer** | By signing, the reviewer is attesting that the document's approach and contents are compliant with the study protocol, all appropriate, regulatory requirements, and other significant guidelines. This individual(s) has reviewed the document for accuracy and completeness. |

# Table of Contents

## LIST OF ABBREVIATIONS

| | |
|---|---|
| **AE** | Adverse Event |
| **ANCOVA** | Analysis of Covariance |
| **ANOVA** | Analysis of Variance |
| **ATC** | Anatomical Therapeutic Class |
| **CRF** | Case Report Form |
| **CRP** | C-Reactive Protein |
| **CPK** | Creatine Phosphokinase |
| **DB** | Double-Blind |
| **ECG** | Electrocardiogram |
| **EDC** | Electronic Data Capture |
| **FDA** | US Food and Drug Administration |
| **GCP** | Good Clinical Practice |
| **hs-CRP** | High Sensitivity C-Reactive Protein |
| **ICH** | International Conference on Harmonization |
| **ICH E8** | General Considerations for Clinical Trials |
| **ICH E9** | Statistical Principles for Clinical Trials |
| **IDMC** | Independent Data Monitoring Committee |
| **FAS** | Full Analysis Set |
| **LOCF** | Last Observation Carried Forward |
| **MedDRA** | Medical Dictionary for Regulatory Activities |
| **MMRM** | Mixed model of repeated measures |
| **MNAR** | Missing Not At Random |
| **OL** | Open-Label |
| **OC** | Observed Case |
| **PCS** | Potentially Clinically Significant |
| **PD** | Pharmacodynamics |

| | |
|---|---|
| **PK** | Pharmacokinetics |
| **PMN** | Pattern Mixture Model |
| **PPS** | Per Protocol Set |
| **QOL** | Quality of Life |
| **SAE** | Serious Adverse Event |
| **SAP** | Statistical Analysis Plan |
| **SmPC** | Summary of Product Characteristics |
| **SOC** | System Organ Class (used in MedDRA dictionary) |
| **TEAE** | Treatment Emergent Adverse Event |
| **TLF** | Tables, Listings, and Figures |
| **USPI** | United States Package Insert |
| **WC** | Worst Case |
| **WHO** | World Health Organization |

# 1 INTRODUCTION

This statistical analysis plan (SAP) provides a more technical and detailed elaboration of the statistical analyses of efficacy and safety as outlined and/or specified in the final study protocol. Specifications of tables, figures, and data listings are contained in a separate document (Table shells).

A separate pooled-analysis SAP for pre-specified analysis based on adjudicated composite safety data will complement this study specific SAP. This SAP is based on the Amendment 4 of the protocol dated September 20, 2017.

# 2 STUDY OBJECTIVES

## 2.1 PRIMARY OBJECTIVE

Evaluate the efficacy and safety of roxadustat in the treatment of anemia in incident dialysis subjects compared to active control (Epoetin alfa).

## 2.2 SECONDARY OBJECTIVES

- Evaluate the utilization of intravenous (IV) iron with roxadustat compared to active control
- Evaluate the effect of roxadustat on serum lipid parameters compared to active control
- Evaluate the effect of roxadustat on blood pressure (BP) compared to active control
- Evaluate time to achieve hemoglobin (Hb) response compared to active control

# 3 STUDY DESIGN

This is a Phase 3, multicenter, randomized, open-label, and active-controlled study to evaluate the efficacy and safety of roxadustat in incident dialysis subjects with anemia.

A total of up to 1,200 subjects are planned to be randomized to receive roxadustat or epoetin alfa (active control) in a 1:1 ratio, respectively.

Randomization stratification factors include geographical region (US vs. Ex-US), screening Hb values (≤8 g/dL vs. >8 g/dL), and cardiovascular/cerebrovascular/thromboembolic medical history (yes vs. no).

In addition, the change of protocol were incorporated via amendments on 20 Oct 2014, 24 Nov 2015, 12 Aug 2016, and current one on 20 September 2017.

Subjects randomized to roxadustat will have doses administered thrice weekly (TIW) throughout the Treatment Period using an initial tiered, weight-based, dosing scheme (see Table 1 below), followed by dose titration every 4 weeks.

**Table 1. Initial Study Drug Dosing**

| **Study Drug** (Dose Frequency) | **Low Weight/Dose** | **Median Weight/Dose** | **High Weight/Dose** |
|---|---|---|---|
| Roxadustat (TIW) – Original Protocol | 45 to 60 kg/70 mg | >60 to 90 kg/100 mg | >90 to 160kg/150 mg |
| Roxadustat (TIW) – Amended Protocol | ≤ 70 kg/70 mg | | >70 to 160 kg/100 mg |
| Epoetin alfa HD (TIW) | IV dosing according to epoetin alfa USPI or SmPC | | |
| Epoetin alfa PD (TIW) | Epoetin alfa should be administered according to epoetin alfa USPI or SmPC or local standard care. | | |
| | Abbreviations: HD = hemodialysis; IV = intravenous; PD = peritoneal dialysis; SmPC = summary of product characteristics; TIW = three times a week; USPI = United States Package Insert; wt = weight.<br>Note: Weight in HD subjects = subject's dry weight. | | |

The study periods are as follows:

- Screening Period: Up to 6 weeks
- Treatment Period: Treatment duration is variable for individual subjects with maximum treatment duration of up to approximately 3 years after the last subject is randomized. The minimum treatment duration may be less than 52 weeks.
- Post-Treatment Follow-Up Period: 4 weeks

**Dose Adjustments**

***Roxadustat arm***

Dose adjustments will occur in two separate study dosing phases: the Correction Phase and the Maintenance Phase. Each of these phases will follow unique dose adjustment rules according to Appendix 2 in the amended protocol of 20 October 2014. All subjects in the roxadustat arm will be dosed orally TIW during the Treatment Period. The maximum roxadustat dose is 3.0 mg/kg per dose or 400 mg, whichever is lower.

All dose adjustments as well as assessments of predefined out of range hemoglobin elevations are based on Hb values using a point-of-care device such as HemoCue® or CritLine®. In the event that the central lab Hb value of the site visit is significantly different and the dose adjustment decision based on the HemoCue® or CritLine® value is being reconsidered, the Medical Monitor should be contacted, if possible.
The determination of Hb response and transition from the Correction to the Maintenance Phase of the study is based on the central laboratory Hb value.

***Active Control arm***

For subjects receiving Active Control, dose adjustment rules are implemented according to the epoetin alfa US Package Insert (USPI) or Summary of Product Characteristics (SmPC).

### Correction Phase of Dosing

***Roxadustat arm***

The aim of the Correction Phase is to increase Hb levels from baseline to the desired Hb level defined as a responder in Section 4.1 by using the dose adjustment algorithm in appendix 2 of the amended protocol of 20 September 2017. This phase is variable in length for each subject.

### Maintenance Phase of Dosing for Roxadustat Arm

The aim of the Maintenance Phase is to maintain Hb levels after the initial correction by using the dose adjustment algorithm in appendix 2 of the amended protocol of 20 September 2017.

### Dosing and Dose Adjustment for Active Control Arm (Epoetin alfa)

Subjects on epoetin alfa should maintain their Hb levels within the target range accepted by their health authorities, specifically:

- Hb 10 to 11 g/dL in the United States
- Hb 10 to 12 g/dL in countries outside the United States

Subjects receiving HD on epoetin alfa will be dosed IV TIW, with starting doses and dose adjustment rules according to the epoetin alfa USPI or SmPC.

Subjects receiving PD on epoetin alfa will be dosed according to the epoetin alfa USPI or SmPC, or local standard of care.

For countries using prefilled syringes, the initial epoetin alfa dose and dose adjustments should be approximated to the closest calculated weekly dose.

# 4 STUDY ENDPOINTS

## 4.1 PRIMARY EFFICACY ENDPOINT

There are 2 separate primary endpoints which are defined for US (FDA) submission and Ex-US submission respectively.

US (FDA) submission.

The primary efficacy endpoint for the US submission is the Mean Hb change from baseline (using central laboratory values) to the average level during the Evaluation Period, defined as Week 28 until Week 52. This analysis will be based on the intent-to-treat (ITT) population. Hemoglobin values under the influence of rescue therapy (see definition below) will not be censored for the primary analysis.

Ex-US submission

The primary efficacy endpoint for the Ex-US submission is defined as the proportion of subjects who achieve a Hb response at two consecutive visits at least 5 days apart during the first 24 weeks of treatment, without rescue therapy (see definition below) within 6 weeks prior to the Hb response in the Per Protocol Set (PPS). A Hb response is defined, using central laboratory values, as:

- Hb ≥11.0 g/dL and a Hb increase from baseline by ≥1.0 g/dL in subjects whose baseline Hb >8.0 g/dL, or
- Increase in Hb ≥2.0 g/dL in subjects whose baseline Hb ≤8.0 g/dL.

Rescue therapy for roxadustat treated subjects is defined as recombinant erythropoietin or analogue (ESA) or RBC transfusion, and rescue therapy for Epoetin alfa treated subjects is defined as RBC transfusion. All endpoints using Hb are based on the central lab data.

Baseline Hb is defined as the mean of at least the last 3 available central laboratory Hb values prior to first dose of study medication: three last screening Hb values plus the pre-dose Hb value collected on day 1. In subjects with missing Day 1 Hb value, the mean of three last screening laboratory Hb values will be considered as baseline Hb value.

Hb values from the central laboratory will be sorted by visit date. If two consecutive on-treatment Hb values from the central laboratory meet the Hb response criteria defined by primary efficacy endpoint for the Ex-US submission, the subject will qualify for Hb response as long as there is no rescue therapy 6 weeks before the date of the first of these 2 Hb measurements.

The classification of Hb response shall take into account all consecutive Hb values regardless of whether they are obtained at scheduled or unscheduled visits.

RBC transfusion is collected in the Blood Transfusions form of the eCRF. The use of ESA is recorded in the related forms of the eCRF and coded into ATC (Code: B03XA01) and generic name.

## 4.2 ALTERNATIVE DEFINITIONS OF THE PRIMARY EFFICACY ENDPOINT

### 4.2.1 Hb Response Regardless Use of Rescue Medication (for Sensitivity Analysis)

An alternative definition of the Hb response, for sensitivity analysis purposes, is defined as:

- Hb ≥11.0 g/dL and a Hb increase from baseline by ≥1.0 g/dL in subjects whose baseline Hb >8.0 g/dL, or
- Increase in Hb ≥2.0 g/dL in subjects whose baseline Hb ≤8.0 g/dL.

at two consecutive visits with at least 5 days apart within the first 24 weeks of treatment regardless use of rescue therapy. Subjects who discontinue study medication before Hb response, as defined above, will be considered as non-responder.

## 4.3 SECONDARY EFFICACY ENDPOINTS

The secondary efficacy endpoints are for both US and Ex-US submissions unless specified otherwise:

US (FDA submission): The proportion of subjects who achieve a Hb response at two consecutive visits with at least 5 days apart during the first 24 weeks of treatment, without rescue therapy within 6 weeks prior to the Hb response. This analysis will be based on the ITT Population.

A Hb response is defined, using central laboratory values, as

- Hb ≥ 11.0 g/dL and a Hb increase from baseline by ≥ 1.0 g/dL in subjects whose baseline Hb > 8.0 g/dL, or
- Increase in Hb ≥ 2.0 g/dL in subjects whose baseline Hb ≤ 8.0 g/dL

Ex-US submission: Mean Hb change from baseline to the average level during the Evaluation Period, defined as Week 28 until Week 36 without rescue therapy within 6 weeks prior to and during the evaluation period. This analysis will be based on the PPS population.

- The time to achieve the first hemoglobin (Hb) response defined by the primary endpoint for Ex-US or the first secondary efficacy endpoint for the US submission.
- Proportion of patient exposure time (months) with Hb>= 10 g/dL during Weeks 28-52, similarly analysis for the period between Weeks 28 - 36
- Mean change from baseline in Low-density lipoprotein (LDL) cholesterol averaged over Weeks 12-24
- Mean change from baseline in Hb levels between Weeks 18 to 24 in patients whose baseline hs-CRP> ULN
- Average monthly IV iron use per subject during weeks 28-52
- Time to first RBC transfusion during the treatment
- Mean change in mean arterial pressure (MAP) averaged over Weeks 8-12

- Time to first exacerbation of hypertension (defined as [systolic BP≥170 mmHg AND systolic BP increase from BL ≥20 mmHg] or [diastolic BP ≥110 mmHg AND diastolic BP increase from BL ≥15 mmHg]) during weeks 28 to52

## 4.4 ADDITIONAL EVALUATION OF EFFICACY

The additional efficacy evaluations in this study are:

- **Hb Correction and Maintenance:**
  - Hemoglobin maintenance: Mean change from baseline in Hb averaged over 8 weeks of treatment at Weeks 28 to 36, without rescue therapy within 6 weeks prior to and during this 8-week evaluation period.
  - Hemoglobin long-term Maintenance: Mean change in Hb averaged over 8 weeks of treatment at Weeks 44-52 without rescue therapy within 6 weeks prior to and during this 8-week evaluation period.
  - Mean change from baseline in Hb averaged over the 96 to 104 weeks of treatment, without rescue therapy within 6 weeks prior to and during this 8-week evaluation period.
  - Change from baseline in Hb at each of the selected post-dosing time points.
  - Proportion of subjects with Hb >=10 g/dL averaged over Weeks 28-36, without rescue therapy within 6 weeks prior to and during this 8-week evaluation period.
  - Proportion of subjects with mean Hb < 9 g/dL, 9-<10, 10-<11, 11-<12, 12-<13, >=13 g/dL during Weeks 28-36
  - Proportion of patient exposure (PEY) with Hb < 9 g/dL, 9-<10, 10-<11, 11-<12, 12-<13, >=13 g/dL during weeks 28 to 52
- **Hospitalizations:**
  - Time to first hospitalization (% of subjects) up to Week 52.
  - Time to first hospitalization or skilled nursing facility (% of subjects) up to Week 52
  - Number of days of hospitalizations per patient-exposure year (PEY).
  - Number of days of hospital or skilled nursing facility per PEY
  - Number of medical-facility free days (hospital, skilled nursing facility, emergency room, or overnight observation) per PEY
  - Number of days on treatment out of hospital and skilled nursing facility up to Week 52, 7 days after Last Dose.
- **Missed dialysis sessions up to Week 52**
  - Occurrence (number) of missed dialysis sessions
  - Proportion of subjects with missed dialysis sessions
  - Number of days of missed dialysis sessions per patient-exposure year (PEY)

- **Rescue Therapy Use up to Week 52:**
  - Proportion of subjects who receive RBC transfusions.
  - Number of RBC packs per patient-month exposure to study medication.
  - Roxadustat subjects: The proportion of subjects requiring ESA rescue therapy (ATC code: B03XA01 )
- **Changes in Cholesterol Levels:**
  - Change at each of the protocol specified treatment time points in:
    - total cholesterol,
    - low-density lipoprotein (LDL)/high-density lipoprotein (HDL) ratio,
    - non-HDL cholesterol.
  - Proportion of subjects achieving LDL target of <100 mg/dL averaged over Weeks 12-24 of treatment.
- **Blood Pressure Effect:**
  - Time to an exacerbation of hypertension over weeks 28-52, meeting at least 1 of the following criteria: Increase in blood pressure: An increase from baseline of ≥ 20 mm Hg systolic BP and sBP >170 mmHg or an increase from baseline of ≥ 15 mm Hg diastolic BP and dBP>100 mmHg.
  - Proportion of subjects achieving blood pressure treatment goal in ESRD subjects (pre dialysis systolic BP <140 mmHg and diastolic BP<90 mmHg) averaged over Weeks 12-28.
- **Health Related Quality of Life (HRQoL) and EQ-5D-5L Benefits of Anemia Therapy:**
  Mean change averaged over Weeks 12, 36 and 52 of treatment including those listed below.
  - Vitality Sub-score of SF-36:
    - In FAS subjects with baseline Vitality Sub-score below 50.
    - In all FAS subjects.
  - Physical Functioning Sub-scores of SF-36:
    - In FAS subjects with baseline *Physical Functioning Sub-scores* below 40.
    - In all FAS subjects.
  - Anemia Subscale ("Additional Concerns") of Functional Assessment of Cancer Therapy-Anemia (FACT-An) Scores:
    - In FAS subjects with baseline subscale scores below 55 (generally associated with fatigue).
    - In all FAS subjects.

- Total FACT-An Scores:
  - In FAS subjects with baseline FACT-An scores below 135
  - In all FAS subjects.
- EQ-5D-5L Scores and other QoL Measures/Other Component scores of SF-36: In all FAS subjects.

- **Hepcidin, Iron, CHr and HbA1c:**
  - Change from baseline in serum hepcidin at each of the selected time points (e.g., Weeks 4, 12, 20, 44 and every 8 weeks onwards)
  - Change in serum iron from baseline to Week 28
  - Change in TSAT from baseline to Week 28
  - Change from baseline in serum ferritin at each of the selected time points, total and sub-grouped by baseline values of, >=400 ng/mL, 400 to 100 ng/mL, and <100 ng/mL.
  - Change from baseline in TSAT at each of the selected time points, total and sub-grouped by baseline values of >= 40%, 40% to 20%, and <20%.
  - Serum iron at each of the time points tested
  - Change from baseline in CHr at each of the selected time points (e.g., Weeks 4. 8, 12, 20, 28, 36 and every 8 weeks onwards)
  - Proportion of patients with CHr > LLN at each timepoint tested: Weeks 4, 8, 12, 20, 28, 36, and every 8 weeks onwards)
  - Change from baseline in HbA1c level at each of the selected time points in subjects without history of diabetes, in subjects with history of diabetes, and all subjects.
  - Changes from BL to each study visit (when measured) in fasting blood glucose,

## 4.5 SAFETY ASSESSMENTS

Study-specific safety will be assessed by evaluating the following:

- Occurrence of treatment emergent adverse events (TEAEs), treatment emergent serious adverse events (TESAEs) and clinically significant laboratory values
- Changes from Baseline in vital signs, electrocardiogram (ECG) findings, and clinical laboratory values.

- Safety interpretation will also be made based on analyses of composite endpoints derived from pre-specified and adjudicated events pooled across multiple studies in the roxadustat Phase 3 program. The members of an independent adjudication committee blinded to treatment assignment will adjudicate the following events in multiple phase 3 studies. Death from any cause, MI, stroke, heart failure requiring hospitalization, unstable angina requiring hospitalization, hypertensive emergency, deep venous thrombosis, pulmonary embolism, and vascular access thrombosis.

- Safety analyses based on these adjudicated events will be pooled across multiple studies. The analyses of the adjudicated events will be detailed in the pooled SAPs.

# 5 GENERAL STATISTICAL CONSIDERATIONS

## 5.1 SAMPLE SIZE DETERMINATION

The sample size calculation is based on the primary endpoints for the US (FDA) submission and Ex-US submission.

During the course of this study, which is being conducted in parallel with other Phase 3 studies, approximately 1000 subjects were to be enrolled to contribute to safety evaluation of roxadustat in comparison to epoetin alfa in CKD patients on dialysis less than 6 months. The final number of patients to be enrolled in this study will be based on the enrollment rate of other roxadustat phase 3 studies on dialysis patients, in order to optimize program timeline to generate sufficient adjudicated safety data across dialysis studies.

With at least 600 subjects, the study will provide at least 99% power to demonstrate statistical non-inferiority of roxadustat versus ESA in the primary endpoint for US (FDA) submission (i.e., specifically, Hb change from baseline to the average level during the evaluation period defined as Week 28 until Week 52). This assumes a difference (roxadustat minus ESA) of -0.30 g/dL, a non-inferiority margin for this difference of -0.75 g/dL (see Appendix 6 for the NI margin justification) and a standard deviation of 1.25 g/dL. This endpoint will be analyzed using the ITT population for the US (FDA) submission.

The study will provide at least 99% power to demonstrate statistical non-inferiority of roxadustat versus ESA in the primary endpoint outside of the United States (i.e., specifically, the proportion of subjects who achieve a Hb response at two consecutive visits during the first 24 weeks of treatment, without rescue therapy within 6 weeks prior to the Hb response). This assumes an 80% responder rate for both roxadustat and epoetin alfa, in order to support the primary efficacy analysis (i.e., a non-inferiority comparison in responder rate between roxadustat and epoetin alfa) and assuming a non-inferiority margin of -15% for this difference (roxadustat minus epoetin alfa).

Appendix 6 has description on the justification that the non-inferiority margin that were used.

## 5.2 ANALYSIS POPULATIONS

### 5.2.1 Intent-to-treat (ITT) Population

The ITT population will consist of all randomized subjects. If treatment received differs from the randomized treatment, the randomized treatment assignment will be used.

### 5.2.2 Safety Population

The Safety Population will consist of all randomized/enrolled subjects who received at least one dose of study medication. If treatment received differs from the randomized treatment, the actual treatment will be used for the safety analysis.

### 5.2.3 Full Analysis Set (FAS)

The FAS population will consist of all randomized/enrolled subjects who received at least one dose of study drug and have at least one post-dose Hb assessment. If treatment received differs from the randomized treatment, the randomized treatment assignment will be used for efficacy analysis.

### 5.2.4 Per Protocol Set (PPS)

The PPS population will consist of all subjects in the FAS population who received at least 8 weeks of treatment, have at least one valid post-dose Hb assessment and are without major protocol violations.

### 5.2.5 Major Protocol Deviations

Major protocol deviations of interest may include, but are not limited to the criteria in Table 2.

A subset of pre-specified major protocol deviations will exclude some patients in the PPS analyses. These will be identified while data are collected prior to database lock. Considerations will be given according to the following table.

Table 2. Criteria for Assessing Major Protocol Deviations

| Number | Major Protocol Deviation |
|---|---|
| 1 | Violation of key* inclusion or exclusion criteria which may affect the assessment of the efficacy of the study drug |
| 2 | Administration of wrong randomization study drug for more than 4 week before week 24; any duration from week 24 to week 52 |
| 3 | Study drug compliance < 75% (up to Week 52) |
| 4 | Administration of prohibited concomitant medication that may impact evaluation of efficacy of the study drug* |
| 5 | Significant noncompliance with study procedures that may impact evaluation of efficacy of the study drug will be evaluated case by base* |

*Subject to Medical Monitor's decision

The number and percentage of major protocol deviations will be categorized and summarized by treatment group as deemed appropriate.

## 5.3 METHODOLOGY AND CONVENTIONS

Safety and efficacy data will be summarized and presented by treatment group and time point in summary tables. Continuous variables will be presented by descriptive statistics: n, mean, standard deviation, median, minimum, and maximum. Categorical variables will be tabulated by frequency count and percentage.

Lab results obtained from the central laboratory, rather than local laboratories, will be used for all efficacy and safety analyses. Local laboratory values, if collected in the CRF's, will be listed only in data listing.

When the actual treatment received by a subject is different from the randomized treatment assigned, the subject will be analyzed per the randomized treatment for the efficacy parameters; while they will be analyzed per actual treatment that was taken for the safety parameters for US submission.

Unless otherwise stated, all statistical tests will be two-sided hypothesis tests performed at the 5% level of significance for main effects and all confidence intervals will be two-sided 95% confidence intervals.

The secondary endpoints will be tested sequentially using the fixed sequence approach for multiplicity adjustments at an alfa level of 0.05. There will be no adjustments for multiple comparisons for other tests.

All analyses will be performed using SAS® Version 9.1.3 or higher.

## 5.4 ADDITIONAL DATA HANDLING RULES AND PRESENTATION SPECIFICATIONS

The following general guidelines will apply to all statistical analyses and data presentations:

- Baseline is defined as the last available value obtained prior to the first dose of study drug, unless otherwise specified in this SAP.
- Hb baseline is defined as the mean of the at least last three available values obtained prior to the first dose.
- Baselines for reticulocyte count, reticulocyte hemoglobin content (CHr), hepcidin, serum iron parameters (transferring, TIBC, TSAT, Ferritin, sTfR, and iron), lipids, blood pressures and heart rate are defined as the mean of values obtained within 6 weeks prior to the first dose.
- Randomization stratification factors and enrollment protocol version derived from actual data (not the ones from the randomization system) will be used in all applicable analysis models.

  The stratification factors to be used in efficacy analyses are:

  1. Us vs Ex-US Screening Hb values (≤ 8 g/dL vs. > 8 g/dL) (other than Hb related endpoints)
  2. Cardiovascular/cerebrovascular/thromboembolic medical history (Yes vs. No)
  3. Geographic Region
  4. The original protocol and amended protocols

- Unscheduled visits within an allowable window will be grouped into the closest scheduled visits based on the visit window specified in Appendix 1. For subjects who have more than one measurement at a certain scheduled visit, the last measurements

will be used, with the exception of CPK, WBC, liver function tests (i.e., ALT, AST, GGT, ALP, and total bilirubin), in which the maximum measurement will be used.
- By default, US conventional units will be used for laboratory value presentations. A set of lab summary tables in SI units will also be provided based on TLF index.
- Age is calculated as of date that the informed consent form was signed.
  - age = INTCK('YEAR', Birth date, date of Informed Consent, 'C') where INTCK is a SAS function.
- Duration of treatment or days in treatment is calculated as: last dose date – first dose date +1
- Body weight, height and temperatures will converted using the following formula:
  - kg = lb/2.2
  - cm = 2.54 x in
  - $C^o = (5/9) \times (F^o - 32)$
- The mean, standard deviation and median will be presented with adding one more decimal to raw data with rounding off. The minimum and maximum will be presented with the same number of decimals as in the raw data.
- 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.
- Any p-values will be rounded to four decimal places and will be presented as '<.0001' if they are less than 0.0001 after rounding.
- All tables and listings will have a header showing "FibroGen, Inc.", the protocol number, and Page x of y. Footer will indicate the program file path/name, run date and run time.
- 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.
- Decimal points will be presented as follows: N will be presented without decimal, minimum and maximum in same precision as in the database, mean and median in one more decimal than minimum and maximum, and SD in one more decimal than mean and median.
- Tables and figures will use derived analysis visit. Listings will use nominal visits, show the flag to indicate analysis visit to be used. Namely, both Nominal visit and analysis visit will be presented in the listing.
- Additional data handling conventions are detailed in Appendix 1.

# 6 SUBJECT ACCOUNTABILITY AND DISPOSITION

The following subject data will be summarized and presented by treatment group (roxadustat and active control) if applicable:

- Number and percentage of subjects screened and randomized (using the screened subjects population)

- Number and percentage of subjects randomized at each center, and for all centers combined, by treatment group (using the ITT population)

- Number and percentage of subjects in each analysis set, by treatment group (using the ITT population)

- Number and percentage of subjects excluded from the Per Protocol analysis set by reason for exclusion, treatment group (using the ITT population)

- Kaplan-Meier plots will be generated for premature treatment discontinuation by randomization arm showing 2 curves (one curve per treatment group).

- All subjects who prematurely discontinued during the treatment period will be listed by discontinuation reason for the randomized population.

# 7 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographic parameters and important baseline and disease characteristics will be summarized by treatment group for the ITT, Safety, FAS and PPS populations. These include but may not be limited to age, age group (18 to 64, 65 to 74, ≥75), sex, ethnicity, race, region, weight, body-mass index (BMI), Hemoglobin, baseline Hb categories, iron repletion status at baseline, Ferritin, Ferritin group (<100 vs. >=100 ng/mL), TSAT and TSAT group (<20% vs. >=20%), iron deplete (ferritin 〉=100 and TSAT 〉=20%) vs. not, cHepcidin< eligibility threshold or not, baseline C-reactive protein (CRP) group (CRP ≤ULN vs. CRP > ULN), cardiovascular or cerebrovascular or thromboembolic medical history (yes vs. no), primary reason for CKD/ESRD as one of the baseline characteristics (DM and HPT vs. all others).

In addition, 25%-75% values of Hb and platelets will be presented. Frequency distributions (number and percentage of subjects) will be presented for categorical variables.

The baseline characteristics, iron indices, and iron IV given between ESA naïve and those treated with ESA then washed out will be summarized and presented in a table.

Descriptive statistics of baseline values for other parameters will be presented in their change from baseline tables.
A summary table for the patient population for enrolled patients before and after the protocol amendments will be presented.

# 8 MEDICAL HISTORY

Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA). Medical History of interest including Chronic Kidney Disease (CKD) History inclusive of CKD Cause, Cardiovascular Disease, Cerebrovascular Disease, Thrombosis History, Hypertension History, Diabetes History, and Anemia History will be summarized by system organ class, preferred term and treatment group for the Safety Population.

# 9 STUDY MEDICATION

## 9.1 EXTENT OF EXPOSURE

Exposure to study medication will be summarized by treatment group in terms of treatment duration in weeks, which is calculated using the following formula: (the date of last medication taken - the date of first dose taken +1)/7.

Total weekly study drug exposure is defined as the total prescribed dose (in mg and mg/kg for Roxadustat and IU and IU/kg for EPO) of study drug administered within the week (windowed by 7-day period from Day 1).

Duration of exposure, weekly exposure and total study drug exposure will be tabulated by treatment group for the safety population.

Per administration amount and administration frequency will also be tabulated by treatment group for the safety population and PPS population.

Patient-Exposure-Year (PEY) is defined as (Last Dose Date – First Dose Date + 1)/365.25.

### 9.1.1 Dosing Changes

Dosing changes for both treatment groups are collected in the Study Drug Administration/HemoCue/Dose Adjustment Form in the eCRF. Two types of dosing changes will be calculated.

A dose-per-intake change is the change in the number of milligrams on the intake day (for example from 200 mg TIW to 250 mg TIW). A weekly-dose change is the change in the prescribed weekly dose, calculated as the dose-per-intake times the weekly frequency.

For example a change from 200 TIW to 250 TIW is a change of 600 mg to 750 mg per week which is considered as a change in dose-per-intake and a change in weekly-dose as well.
For each subject the total number of dose-per-intake changes and the weekly-dose changes will be calculated.

### 9.1.2 Duration of Exposure

Exposure time will be categorized according to the following categories by treatment groups (roxadustat and ESA):

- Less than 2 weeks
- At least 2 weeks, less than 4 weeks
- At least 4 weeks, less than 26 weeks
- At least 26 weeks, less than 52 weeks
- At least 52 weeks, less than 78 weeks
- At least 78 weeks, less than 104 weeks
- At least 104 weeks, less than 130 weeks
- At least 130 weeks, less 156 weeks
- More than 156 weeks
- Unknown

## 9.2 TREATMENT COMPLIANCE

Study medication dosing compliance for a specified period is defined as the total dose (mg actually taken by a patient during that period divided by the prescribed dose expected to be taken during the same period multiplied by 100. An overall per-dose compliance measure can be calculated by (# of actual dose administrations)/ (Total # of expected dose administrations) *100 during the subject's treatment period. Descriptive statistics for study medication compliance will be presented by treatment group for the entire treatment period of the study.

Compliance will be summarized as follows:

- Descriptive statistics will be summarized by the 2 treatment groups for entire treatment period.
- Percent compliance will be categorized according to the following categories for entire treatment period for safety population by the 2 treatment groups:
    - less than 50% (significant drug noncompliance)
    - at least 50%, less than 75% (moderate drug noncompliance)
    - greater or equal 75% (drug compliance)
    - unknown

## 10 PRIOR AND CONCOMITANT MEDICATIONS

The World Health Organization Drug Dictionary (WHO Drug) will be used to classify prior and concomitant medications by therapeutic class and generic name based on ATC code level 3. Prior medication is defined as any medication taken and stopped prior to the first dose of the study medication. Concomitant medication is defined as any medication taken between the day of first dose of the study medication and the day of last study medication date + 28 days.

Medication start and end dates and times will be compared with the start date of study drug and classified as per Table 3.

In case of partial or missing dates, comparisons will be made based on the level of detail available. For example, if start date of study drug is 04Jan2013, and a medication has a start date of Jan2013 but missing day, the medication will be classified as concomitant.

**Table 3: Classification of prior and concomitant medications**

| **End date** / **Start date** | **Before start of study drug administration** | **On or after start of study drug administration** | **Missing** |
|---|---|---|---|
| **Before start of study drug administration** | Prior | Concomitant | Concomitant |
| **On or after start of study drug administration** | – | Treatment Emergent Concomitant | Treatment Emergent Concomitant |
| **Missing** | Prior | Concomitant | Concomitant |

Both prior and concomitant medication usage will be summarized by the number and proportion of subjects in each treatment group. Subjects will only be counted one time in each unique ATC Class and generic name if multiple drugs are used by a subject.

Detailed analyses may be performed on prior and concomitant medications of special interests such as oral iron, blood pressure medications and lipids medications.

# 11 EFFICACY ANALYSES

Efficacy analysis will be conducted on the ITT and FAS for US (FDA) submission and analysis for non-inferiority on the PPS population and analysis for superiority on FAS population for EU regulatory submission.

The **Efficacy Emergent Period** is defined as the evaluation period from the Analysis date of first dose intake up to 7 days after the Last Dose of study drug or EOT Visit, whichever occurs first. This period will be used as reference period for the time to event analyses related to efficacy endpoints, unless specified otherwise.

## 11.1 ANALYSIS OF PRIMARY ENDPOINT

### 11.1.1 Primary Endpoint

There are 2 primary efficacy endpoints: one for the US submission and one for the Ex-US submission.

The primary efficacy endpoint for the US submission is the Mean Hb change from baseline (using central laboratory values) to the average level during the Evaluation Period, defined as Week 28 until Week 52. This analysis will be based on the intent-to-treat (ITT) population. Hemoglobin values under the influence of rescue therapy will not be censored for the primary analysis.

The primary efficacy endpoint for the Ex-US submission is defined as the proportion of subjects who achieve a Hb response at two consecutive visits during the first 24 weeks of treatment, without rescue therapy within 6 weeks prior to the Hb response. This analysis will be based on the PPS population.

- A Hb response is defined, using central laboratory values, as: Hb ≥11.0 g/dL and a Hb increase from baseline by ≥1.0 g/dL in subjects whose baseline Hb >8.0 g/dL, or
- Increase in Hb ≥2.0 g/dL in subjects whose baseline Hb ≤8.0 g/dL.

Rescue therapy for roxadustat treated subjects is defined as ESA rescue or RBC transfusion, and rescue therapy for Epoetin alfa treated subjects is defined as RBC transfusion.

### 11.1.2 Primary Analysis

Efficacy analysis for superiority will be conducted on the ITT population for US (FDA) submission and on the FAS population for Ex-US submission.

Efficacy analysis for non-inferiority will be conducted on the ITT population for US (FDA) submission and on the PPS population for Ex-US submission.

**US (FDA) Submission**: The primary efficacy endpoint for US (FDA) submission is defined as the mean Hb change from baseline to the average level during the Evaluation Period, defined as Week 28 until Week 52. The analysis will be based on the ITT

population. Hb values under the influence of rescue therapy will not be censored for the primary efficacy analysis.
The primary hypothesis to be tested for the primary efficacy analysis is:

> $H_0$: Hb mean change from baseline to the average level from Week 28 to Week 52 in the roxadustat arm ≤ Hb mean change from baseline in the epoetin alfa arm minus 0.75 g/dL
> Versus:
> $H_1$: Hb mean change from baseline to the average level of Week 28 to Week 52 in the roxadustat arm > Hb mean change from baseline in the epoetin alfa arm minus 0.75 g/dL

A multiple imputation analysis of covariance (MI ANCOVA) model will be used. The model will contain terms for treatment group, baseline Hb measurement, and stratification factors except Screening Hb values (≤ 8 g/dL vs. > 8 g/dL). The primary efficacy analysis will be based on the estimated difference between the two treatments overall mean effects throughout the evaluation period based on the pooled ANCOVA model.

This null hypothesis will be rejected if the two-sided 95% CI for the difference between the two treatment groups using MI ANCOVA model lies entirely above -0.75 g/dL.

The following steps will be used to conduct the primary analysis of the primary endpoint:

1. Generate 200 datasets, using seed 162345 for the U.S., where only 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 baseline hemoglobin, and the available non missing hemoglobin for each scheduled Week by treatment group.

   The MCMC statement in the SAS PROC MI procedure with monotone option will be used. As a result, each dataset will only have 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. As a result, 200 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.

3. Analyze each imputed complete dataset using the ANCOVA model using the mean of all observed or imputed Hb values within the evaluation period (Week 28 – 52). The model will contain terms for baseline Hb measurement as a covariate, treatment group and stratification factors except Screening Hb values (≤ 8 g/dL vs. > 8 g/dL) as fixed effects.

Sample SAS code:

```
PROC MIXED data=xx;
  class treatment categorical covariates;
  model change_Week36 = treatment covariates / solution;
  lsmeans treatment / diff cl;
  ods output Diffs=lsdiffs LSMeans=lsm solutionF=Parms;
  by _Imputation_;
run;
```

4. Combine estimates from the results for each of the 200 ANCOVA model using SAS PROC MIANALYZE.

```
PROC MIANALYZE parms(classvar=full)=lsdiffs;
  class treatment categorical covariates;
  modeleffects treatment;
  ods output ParameterEstimates=MIAN_lsdiffs;
run;
```

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 CFB in hemoglobin for the treatment group minus the least-squares mean CFB in hemoglobin for the active comparator group) and the corresponding p-values during the evaluation period.

The analysis will be repeated with Hb values under the influence of rescue therapy censored Namely, set Hb to missing for Rescue Medication and 6 weeks afterwards (or 8 weeks if during the evaluation period) .

**Ex–US Submission**: The primary efficacy endpoint for Ex-US submission is defined as the proportion of subjects who achieve an Hb response at two consecutive visits during the first 24 weeks of treatment, without rescue therapy within 6 weeks prior to the Hb response.

- A Hb response is defined, using central laboratory values, as: Hb ≥ 11.0 g/dL and a Hb increase from baseline by ≥ 1.0 g/dL in subjects whose baseline Hb > 8.0 g/dL, or
- Increase in Hb ≥ 2.0 g/dL in subjects whose baseline Hb ≤ 8.0 g/dL

Rescue therapy for roxadustat treated subjects is defined as ESA rescue or RBC transfusion, and rescue therapy for epoetin alfa treated subjects is defined as RBC transfusion.
The hypothesis to be tested for the primary efficacy analysis is:

$H_0$: Hb response rate for subjects in the roxadustat arm - Hb response rate for subjects in the epoetin alfa arm ≤ -15%
Versus
$H_1$: Hb response rate for subjects in the roxadustat arm - Hb response rate for subjects in the epoetin alfa arm > -15%

A two-sided 95% CI for the difference of 2 responder rates (roxadustat minus epoetin alfa) based on the Miettinen & Nurminen approach adjusting for treatment and stratification factors will be calculated and this null hypothesis will be rejected if the lower bound of the 95% CI is greater than -15%. Subjects who dropped out from the study without data for the assessment will be treated as non-responder.

### 11.1.3 Sensitivity Analyses of Primary Endpoint

The following secondary analyses will be performed on the U.S. primary efficacy endpoint as sensitivity analyses to examine the potential impact of missing data on the estimates. These analyses will be performed using ITT Population. The results of the analyses will be summarized in Table 4. These sensitivity analyses are further detailed in each Section.

**Table 4: U.S. Primary Endpoint Analysis Results**

| Analysis ( ITT Population) | Trt Diff and 95% CI (roxadustat vs EPO) | Std. Err. | Degree of Freedom* | t-statistics | p-value |
|---|---|---|---|---|---|
| ANCOVA with Multiple Imputations (Primary) | .xx (.xx, .xx) | .xx | xxx | .xxx | .xxxx |
| ANCOVA-MI with Hb censored**for rescue therapy | .xx (.xx, .xx) | .xx | xxx | .xxx | .xxxx |
| MMRM | .xx (.xx, .xx) | .xx | xxx | .xxx | .xxxx |
| PMM-Last Mean Carried Forward | .xx (.xx, .xx) | .xx | xxx | .xxx | .xxxx |
| PMM –Baseline Carried Forward (Roxadustat only) | .xx (.xx, .xx) | .xx | xxx | .xxx | .xxxx |
| PMM –Baseline Carried Forward (Both Groups) | .xx (.xx, .xx) | .xx | xxx | .xxx | .xxxx |

* Maximum of Degree of Freedom from each individual ANCOVA. The actual Degrees of Freedom in PMM will be calculated using $df = (m-1)\left(1+\frac{m\,\overline{U}}{(m+1)B}\right)^2$. where B is between-imputation variance and $\overline{U} = \frac{1}{m}\sum_{j=1}^{m} U_j$. with U the standard error associated with estimates, Rubin (1987).

** Set Hb to missing for Rescue Medication and 6 weeks afterwards (or 8 weeks if during the evaluation period)

#### 11.1.3.1 MMRM Model

A mixed model of repeated measures (MMRM) will be used as one sensitivity analysis for the primary endpoint for US Submission (change from baseline in Hb from Week 28 to Week 52).

The model will contain terms for treatment arm, baseline measurement, visit (up to Week 52), visit by treatment interaction, and stratification factors except Screening Hb values (≤ 8 g/dL vs. > 8 g/dL). The primary efficacy analysis (for US [FDA]) will be based on the estimated difference between the two treatments overall mean effects throughout the evaluation period based on the MMRM model. Hb values under the influence of a rescue therapy will not be censored in the primary analysis.

Due to the large amount of visits to include in the model, data up to Week 52, the unstructured covariance pattern model will be applied first. If the algorithm for unstructured covariance pattern does not converge or the likelihood ratio test is not statistically significant then heterogeneous Toeplitz structure will be used. If this second model does not converge either then the (homogeneous) Toeplitz structure will be tried and finally compound symmetry as a covariance structure to achieve convergence. If none of them converge, first order autoregressive (AR (1)) as a covariance structure will be used to achieve convergence.

#### 11.1.3.2 Pattern Mixture Model (PMM)

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 (MNAR). All factors mentioned in the primary analysis will be included in the PMM.

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

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

##### 11.1.3.2.1 Timing and Extent of Missing Data

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

- Full data cases are defined as subjects without any missing hemoglobin values for all scheduled assessments in the evaluation period.

- Missing data cases are defined as subjects with a missing hemoglobin on at least one scheduled Visit Week of the evaluation 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 pattern in first 52-week scheduled visit will be presented by treatment group in a table.

#### 11.1.3.2.2 Assumptions on Missing Data Mechanism

In addition to the extent of missing data, 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 (missing not at random(MNAR)) (Little and Rubin, 1987).

A pattern-mixture model using a treatment-based multiple imputation method (Ratitch et al, 2011) will be used for a sensitivity analysis to explore the robustness of the ANCOVA results for the primary efficacy results.

#### 11.1.3.2.3 PMM –Last Mean Carried Forward

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

The steps to implement this sensitivity analysis are as follows. Parameters below refer to the parameters of the multivariate normal distribution for baseline and post baseline Hb measurement.

1. Create posterior distribution of parameters: Separately for each treatment group, take all patients observed data and assuming MAR to fit a multivariate normal distribution with unstructured mean (i.e. a separate mean for each of the baseline plus post-baseline

scheduled weeks and unstructured variance covariance matrix using a Bayesian approach with an improper prior for the mean and an uninformative Jereys' prior for the variance-covariance matrix (Schafer, 1997, p. 155).

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

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

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

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

#### 11.1.3.2.4 PMM –Baseline Carried Forward (Roxadustat Only and Both Groups)

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

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

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

Similarly, the Rubin's method will be then used to combine the estimates and the differences between the least square mean differences between the two treatment groups from each of the ANCOVA analysis. This sensitivity analysis will be performed for U.S. primary endpoints only.

### 11.1.3.3 Logistic Regression

The primary analysis for Ex-US endpoint will also be repeated using logistic regression for PPS population. In addition to the stratification factors except Screening Hb values (≤ 8 g/dL vs. > 8 g/dL) and treatment group, and baseline Hb will be included as continuous covariates. The odds ratio for roxadustat versus placebo and its 95% confidence interval will be provided.

This model will also be repeated using logistic regression and adjusting by sex and age, in addition to the factors mentioned above.

#### 11.1.3.4 Subgroup Analysis

The primary analysis of the US primary endpoint may be repeated separately by sex, age group, geographic region, baseline Hb categories, baseline CRP, baseline iron status, and cardiovascular/cerebrovascular/thromboembolic medical history on ITT Population. The subgroup analysis of the Ex-US primary endpoint may be performed on PPS for non-inferiority and FAS for superiority.

## 11.2 ANALYSIS OF SECONDARY ENDPOINTS

The secondary endpoints are for both U.S. and Ex-U.S. submission unless otherwise specified. Once the primary hypothesis has been rejected for the primary endpoint, the secondary endpoints in the order that was specified in section 4.3 will be tested using a fixed sequence testing procedure, in order to maintain the overall two-sided type I error of 0.05. If p-value from a test is < 0.05, 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 5 Key Secondary Endpoints Fixed Sequence Testing Procedure**

| Test | Variable | Comparison |
|---|---|---|
| 1 | Proportion of subjects who achieve an Hb response at two consecutive visits during the first 24 weeks of treatment, (U.S.). or the mean Hb change from baseline to the average level during the Evaluation Period, defined as Week 28 until Week 52 for Ex-U.S.. | Non-inferiority of roxadustat versus EPO.<br>The non-inferiority margin for the difference of responder analysis between groups is 0.15<br>The non-inferiority margin for the difference of average of Change in Hb values between groups is - 0.75 g/dL. |
| 2 | LDL cholesterol change from BL to the average of weeks 12 to 24. | Superiority of roxadustat versus EPO |
| 3 | Mean change from baseline in Hb levels between Weeks 18 to 24 in patients whose baseline CRP> ULN; | The non-inferiority with margin of -0.75 of roxadustat versus EPO |
| 4 | Monthly IV iron (mg) use per subject during weeks 28 to 52. | Superiority of roxadustat versus EPO |
| 45 | Time to first RBC transfusion during the treatment. | Non-inferiority of roxadustat versus EPO with Non-inferiority margin of 1.8 |
| 5 | Change from BL to the average MAP of weeks 8 to 12 | Superiority of roxadustat versus EPO |
| 6 | Time to first exacerbation of hypertension (defined as [systolic BP≥170 mmHg AND systolic BP increase from BL ≥20 mmHg] or [diastolic BP ≥110 mmHg AND diastolic BP increase from BL ≥15 mmHg]) during weeks 28 to 52 | Superiority of roxadustat versus EPO & Non-inferiority with margin of 1.8 |

### 11.2.1 Primary Analysis of Secondary Endpoints

The primary analysis for the secondary endpoints will be based on the PPS for the non-inferiority tests and FAS for the superiority tests for Ex-U.S. and FAS for U.S.

#### 11.2.1.1 Hb

The first secondary efficacy endpoint for the US submission and the Ex-US submission will be the primary efficacy endpoint for the Ex-US submission and the US submission, respectively. The Ex-US secondary efficacy endpoint will use MMRM model described in 11.1.3.1. The US first secondary efficacy endpoints will use the same statistical analysis methodology for the primary efficacy endpoint for Ex-US.

#### 11.2.1.2 Mean change from baseline in LDL cholesterol averaged over Weeks 12-24

The mean change from baseline in LDL cholesterol averaged over Weeks 12-24 will be compared between the 2 treatment groups using the MMRM model with baseline LDL cholesterol as a covariate, treatment group, visit, interaction of visit and treatment group, and the above-mentioned 4 stratification factors as fixed effects. The same strategy as that used in MMRM for Hb will be used to choose variance covariance structure. Data up to visit of Week 52 will be included in the model. The estimates for difference of LDL Cholesterol averaged over Weeks 12 to 24 between the two treatment groups will be generated from an estimate statement from Visit Week 12 to 24. Superiority will be declared if the upper bound of the 2-sided 95% confidence interval of the difference between roxadustat and Epoetin alfa (roxadustat - epoetin alfa) is less than 0.

#### 11.2.1.3 Mean change from baseline in Hb levels between Weeks 18 to 24 in patients whose baseline CRP> ULN

Change from baseline in Hemoglobin from baseline to the average level during the Week 18 to 24 will be analyzed using the ANCOVA MI as the primary endpoints. Both Non-inferiority of roxadustat vs. Epoetin and superiority will be tested. The non-inferiority margin is fixed as a difference of -0.75.

#### 11.2.1.4 Average monthly IV iron use during 28 to 52Weeks

The average monthly IV iron use during the treatment period will be calculated for monthly intervals. The treatment period will be divided in periods of 28 days and for each of these periods the monthly mean of IV iron will be used using the following formula:

Monthly iron use for each subject = Total IV iron in mg / [(last visit date – first drug date +1)/ 28]

The average monthly iron use will be compared between the 2 treatment groups using an ANCOVA model adjust for baseline iron replete, treatment group and above-mentioned stratification factors as fixed effects. Superiority will be declared if the lower bound of the 2-

sided 95% confidence interval of the difference between Epoetin alfa and roxadustat exceeds 0.

#### 11.2.1.5 Mean change in mean arterial pressure (MAP) averaged over Weeks 8-12

Mean Arterial Pressure (MAP) will be calculated for each subject using the following formula: MAP = (2/3) * DBP + (1/3) * SBP.

Mean change from baseline in MAP will be analyzed and compared between the 2 treatment groups using an MMRM model with baseline MAP as a covariate, treatment group, visit, interaction of visit and treatment group, and above-mentioned stratification factors as fixed effects. The same strategy as that used in MMRM for Hb will be used to choose variance covariance structure. Data up to visit of Week 52 will be included in the model. The estimates for difference of MAP averaged over Weeks 8 to 12 between the two treatment groups will be generated from an estimate statement from Visit Week 8 to 12. Non-inferiority margin for the difference between groups is 1mmHg. Superiority will be declared if the upper bound of the 2-sided 95% confidence interval of the difference between roxadustat and ESA (roxadustat – ESA) is below 0.

#### 11.2.1.6 Time to first exacerbation of hypertension over Week 28 to 52

Subject with exacerbation of hypertension is defined as meeting the following criterion:

Increase in blood pressure: An increase from baseline of ≥ 20 mm Hg systolic BP and sBP ≥170 mmHg or an increase from baseline of ≥ 15 mm Hg diastolic BP and dBP≥100 mmHg. Increases from baseline in blood pressure are considered as confirmed by taking the mean of triplicates.
An exacerbation of hypertension is defined as an increase from baseline of ≥ 20 mm Hg systolic BP and sBP >170 mmHg or an increase from baseline of ≥ 15 mm Hg diastolic BP and dBP >100 mmHg).

Time to an exacerbation of hypertension in blood pressure will be analyzed and compared between the 2 treatment groups using the Cox Proportional Hazards model adjusting for baseline stratification factors. Both Non-inferiority and Superiority of roxadustat vs. ESA will be tested. The Non-inferiority margin for the difference between groups is 1.3 and the superiority margin for the difference between groups is 1. Subjects will be censored at the time of the last available blood pressure if an increase in blood pressure does not occur.

### 11.3 ADDITIONAL EFFICACY ANALYSES

The additional efficacy endpoints stated in section 4.3 will be analyzed using both the FAS.

#### 11.3.1 Hb Correction and Maintenance

#### 11.3.1.1 Time to achieve the first hemoglobin (Hb) response

Hemoglobin response is defined as the primary endpoint for the Ex-US submission and the first secondary endpoint for the US submission. It will be analyzed and compared between the 2 treatment groups using the Cox Proportional Hazards model adjusting for the above-

mentioned stratification factors. Non-inferiority of roxadustat vs. ESA will be tested. Non-inferiority will be declared if the upper bound of the 2-sided 95% confidence interval of the hazard ratio is less than 1.3. Subjects will be non-responder if the subject drop out the study early.

#### 11.3.1.2 Hemoglobin Maintenance

Hemoglobin maintenance at Weeks 28-36 will be assessed by the mean change in Hb averaged over 8 weeks of treatment at Weeks 28-36 without rescue therapy within 6 weeks prior to and during this 8-week evaluation period.

The Hb values will be used on the central laboratory values. The Weeks will be defined using the visit window defined in Appendix 1.

The mean change in Hb will be analyzed using a MMRM model described in 11.1.3.1.

#### 11.3.1.3 Hemoglobin Long-Term Maintenance

Hemoglobin long-term maintenance will be assessed by the mean change in Hb averaged over 8 weeks of treatment at Weeks 44-52 without rescue therapy within 6 weeks prior to and during this 8-week evaluation period. The mean change in Hb will be analyzed using the MMRM model described in 11.1.3.1

The mean change in Hb averaged over 8 weeks of treatment at Weeks 96-104 will be analyzed using the MI analyses as the one for the primary efficacy endpoint of US-submission.

Hemoglobin long-term maintenance will also be assessed by % patients with average Hb level >= 10 g/dL during weeks 28-36; 44-52; 68-76, 96-104, etc.

To evaluate Hb maintenance by other dosing frequencies, the following summary will be provided for the subgroups of subjects treated on BIW or QW (including any frequency <QW) for longer than 8 weeks (i.e., >=56 days):

- Average Hb values over time for every 4-8 weeks after the initiation of BIW or QW
- Average weekly dose over time for every 4-8 weeks after the initiation of BIW or QW

#### 11.3.1.4 Hb Correction

Hb correction will be assessed using the following endpoints:

- Change in Hb at each of the selected post-dosing time points (see Table 4). Mean change in Hb at each of post-doing time points will be presented for the 2 treatment groups. The mean change will be compared between the 2 treatment groups using the MMRM model described in 11.1.3.1.

- Change in Hb averaged over 8 weeks of treatment at Weeks 28-36, without rescue therapy within 6 weeks prior to and during this 8-week evaluation period, in subjects who have reached an Hb ≥11 g/dL prior to Week 28. The same MMRM model will be

used to compare the mean change from baseline in Hb within this subgroup of subjects.

- Proportion of subjects with Hb within 10-12 g/dL in U.S. and 10-13 g/dL in Ex-U.S. averaged over Weeks 28-36, without rescue therapy within 6 weeks prior to and during this 8-week evaluation period. The proportion of subjects meeting the above-mentioned criteria will be compared between the 2 treatment groups using logistic regression model adjusting for baseline Hb value as covariate with treatment group and baseline stratification factors except screening Hb values (≤ 8 g/dL vs. > 8 g/dL) as fixed effects.
    - Mean change in Hb from baseline to Week 28-36, from baseline to Week 28 to 52 in the subgroup of patients with baseline CRP>ULN.
    - Time to and minimum effective dose evaluation: Time to Hb ≥11.0 g/dL and an increase by >=1 g/dL from baseline and the study drug dose will also be evaluated, with subgroup analysis by starting dose in the original protocol vs starting dose in the amended protocol.

Table 9 presents the analysis visits assigned for hemoglobin samples collection corresponding to the range of treatment days (window) during which an actual visit may have occurred. All scheduled and unscheduled hemoglobin samples that belong to each window will be taken into account.

### 11.3.2 Hospitalizations (including skilled nursing facility)

Hospitalizations are collected in the Hospitalization Records Form of the eCRF. For each subject, an entry will be recorded for each hospitalization. The days of hospitalization will be calculated as the sum of all hospitalizations durations in days (Date of discharge – Date of Admission + 1). In case of missing dates, the hospitalization duration will be assumed to be 5 days.

Only hospitalizations with admission dates that occur during the treatment period and up to 7 days after the last study medication date will be taken into account.

The following hospitalization-related data will be analyzed and compared between the 2 treatment groups:

- Time to first hospitalization up to Week 52- Proportion of subjects hospitalized. Time to first hospitalization or skilled nursing facility up to Week 52.
- Number of days of hospitalizations
- Number of days of medical-facility
- Number of days of hospitalizations per patient-exposure year (PEY).
- Number of days in hospital or skilled nursing facility per patient-year exposure (PEY)
- Number of days of medical-facility per subject-exposure year (PEY).

- Number of days on treatment out of hospital and skilled nursing facility up to Week 52, 7 days after Last Dose. The days will be compared between the 2 treatment groups using ANCOVA model with baseline Hb and the baseline stratification factors.

Note: Elective procedures may be excluded from analyses by using only hospitalization due to AE.

The time to hospitalization event will be compared between the two treatment groups using cox model including baseline stratification factors. The proportion of subjects hospitalized will be compared between the 2 treatment groups using logistic regression model and stratified by the baseline stratification factors. The mean number of days of hospitalization will be analyzed and compared between the 2 treatment groups using the Mantel-Haenszel mean score test adjusting for stratification factors.

### 11.3.3 Missed Dialysis Sessions

Missed dialysis sessions are collected in the Dialysis Modality Form of the eCRF. For each subject, an entry will be recorded for missing dialysis. The number of missing dialysis will be summarized by treatment group.
The following missing dialysis-related data will be analyzed and compared between the 2 treatment groups:

- Occurrence (number) of missed dialysis sessions. The number of missed dialysis sessions for each subject will be collected. The mean number of missed dialysis sessions will be analyzed and compared between the 2 treatment groups using Mantel-Haenszel mean score test adjusting for stratification factors.

- Proportion of subjects with missed dialysis sessions. The proportion of subjects missed dialysis sessions will be compared between the 2 treatment groups using CMH model and stratified by the baseline stratification factors.

- Number of days of missed dialysis sessions per patient-exposure year (PEY). The number of days of missed dialysis sessions for each subject will be calculated (# of missed dialysis/weekly prescribed dialysis frequency* 7), where weekly prescribed dialysis is TIW in this study. The mean number of days of missed dialysis sessions will be analyzed and compared between the 2 treatment groups using the Mantel-Haenszel mean score test adjusting for stratification factors.

### 11.3.4 Rescue Therapy Use

#### 11.3.4.1 Blood Transfusion

For a subject receiving RBC transfusion, the Time at Risk (time up to first RBC transfusion) will be calculated (in years) as:

(First RBCtransfusion date – First dose date of study medication + 1) / 365.25

For a subject not receiving transfusion, the Time at Risk (time until they get censored) is calculated as:

(Date of last study medication – – First dose date of study medication + 1) / 365.25

The blood transfusion form of the eCRF in the cumulative visit will be used to derive the number of RBC packs. The number of RBC units is collected in this form. For transfusions where the number of units is not given but the volume transfused is given, the number of units will be estimated by dividing the volume transfused by 250 mL (for transfusion of packed cells) or by dividing the volume transfused by 500 mL (for transfusion of whole blood).

The total number of RBC units/packs during the treatment period is calculated for each subject by the sum of the transfused units between the Analysis Date of First Dose and up to the Analysis Date of Last Dose. The following 2 endpoints will be analyzed:

- Proportion of subjects who receive RBC transfusions. The proportion of subjects who received RBC transfusion will be compared between the 2 treatments using logistic regression model adjusting for baseline stratification factors.

- Number of RBC packs per patient-month exposure to study medication. The mean number of RBC packs will be compared between the 2 treatment groups using ANCOVA model with baseline hemoglobin as covariate and treatment and stratification factors as fixed effect.

#### 11.3.4.2 ESA Usage as Rescue Therapy for roxadustat-treated Subjects

For roxadustat-treated subjects, ESA, as rescue therapy will be recorded in the ESA log of the eCRF. The total number of ESA-week dose per subject will be calculated.

For each entry that meets the criteria below the ESA-week will be calculated as follows:

- If drug is epoetin alfa, epoetin beta, or an epoetin biosimilar (ATC code: B03XA01), then ESA-Weeks = (stop date – start date + 1) / 7;

- If drug is darbepoetin SQ or IV dose (ATC code: B03XA02), then ESA-Weeks = 2 x (stop date – start date + 1) / 7;

- If drug is Mircera IV or SQ dose (ATC code: B03XA03), then ESA-Weeks = 4 x (stop date – start date + 1) / 7;

- If drug is peginesatide dose IV or SQ (ATC cod: B03XA04), then ESA-Weeks = 4 x (stop date – start date + 1) / 7.

The total number of ESA-week during the treatment period is calculated for each subject by the sum of the ESA-week between the analysis date of First dose and the analysis date of Last dose.

#### 11.3.4.3 Average Monthly IV iron use in 2nd and 3rd year of treatment

The same analysis of 11.2.1.4 will be perform for the above endpoint.

### 11.3.5 Changes in Cholesterol Levels

- Change at each of the selected treatment time points in:
  - total cholesterol,

- low-density lipoprotein/high-density lipoprotein ratio,
- non-HDL cholesterol.

The mean change in these 3 endpoints at each post-dosing time point will be analyzed and compared between the 2 treatment groups using the MMRM model with baseline value as a covariate, treatment, visit and interaction of treatment and visit and stratification factors as fixed effects. The same strategy as that used in MMRM in Section 11.1.3.1 will be used to choose variance covariance. Data up to visit of Week 52 will be included in the analyses.

- Proportion of subjects achieving LDL target of <100 mg/dL averaged over Weeks 12-24 of treatment. The proportion of subjects achieving LDL target will be compared between the 2 treatment groups using logistic regression model adjusting for baseline LDL value and baseline stratification factors as fixed effects.

### 11.3.6 Blood Pressure Effect

- Proportion of subjects achieving blood pressure treatment goal in ESRD subjects (pre dialysis systolic BP <140 mmHg systolic and diastolic BP<90 mmHg) averaged over Weeks 12-28. The proportion of subjects achieving blood pressure goal will be compared between the 2 treatment groups using logistic regression model adjusting for baseline pre dialysis systolic and pre dialysis diastox1lic blood pressures as covariates and baseline stratification factors as fixed effects.
- Mean change in mean arterial pressure (MAP) averaged over Weeks 20-28

The same analysis as 11.2.1.5 will be performed for the above endpoint.

### 11.3.7 Vascular Access Thrombosis

- Time to a treatment-emergent AE of vascular access thrombosis
- Proportion subjects with a treatment-emergent AE of vascular access thrombosis

### 11.3.8 Health Related Quality of Life (HRQoL) and EQ-5D-5L Benefits of Anemia Therapy

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 2). 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: (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).

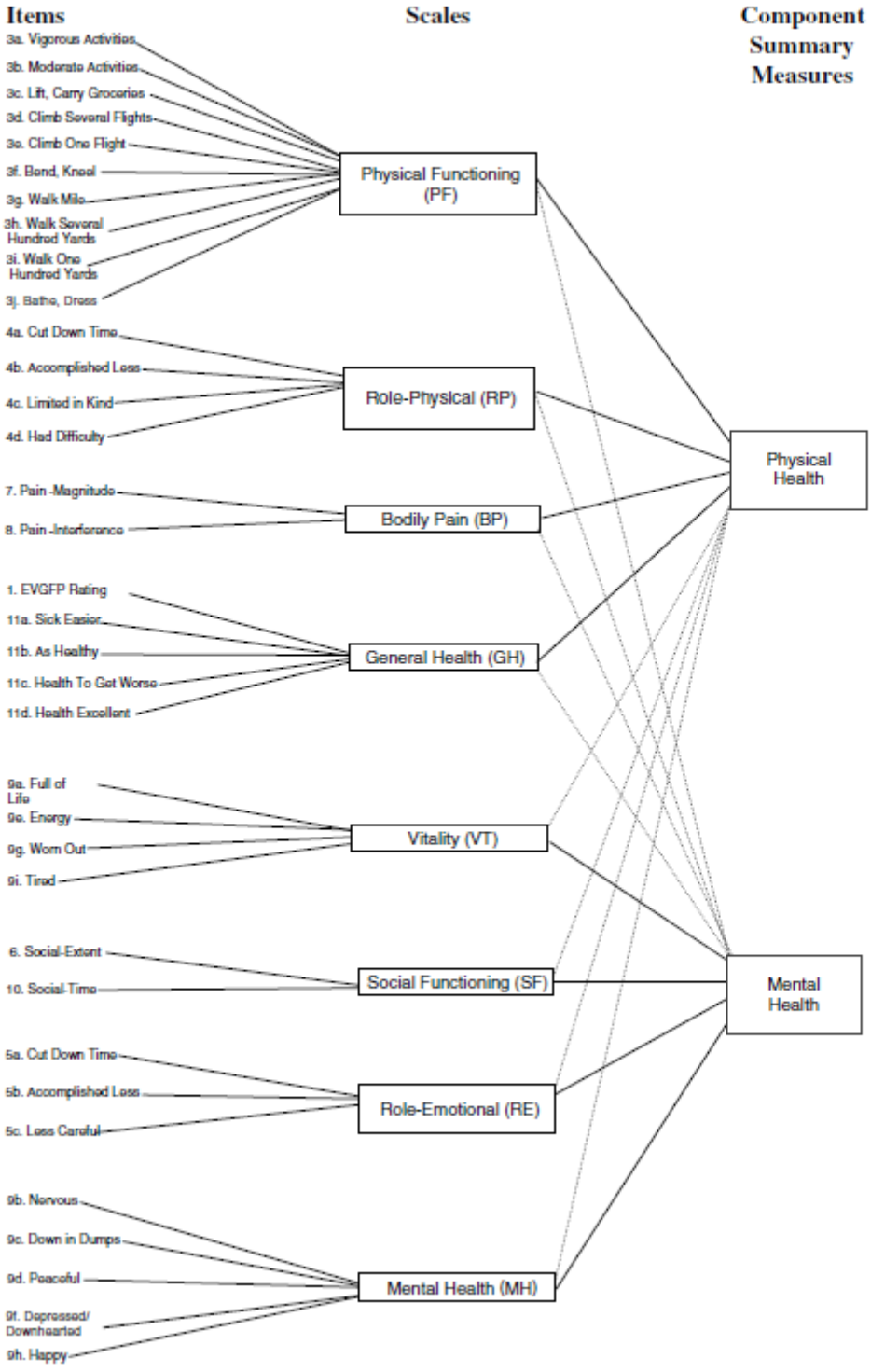
**Table 6 Transformation of SF-36 items to a scale from 0 to 100**

| Question | Self-perceived health | |
|---|---|---|
| | 0 (Poor) | 100 (Better) |
| 1, 2, 6, 8, 9a, 9d, 9e, 9h, 11b, 11d | 5 | 1 |
| 3a to 3j | 1 | 3 |
| 4a to 4d, 5a to 5c, 9b, 9c, 9f, 9g, 9i, 10, 11a, 11c | 1 | 5 |
| 7 | 6 | 1 |
| Programming note: Once transformed – the items should each span the range 0 to 100. | | |

Item scores for each dimension are coded, summed, and transformed to a scale from 0 to 100, with higher scores indicating better self-perceived health (See detail in 4). The transformed items are then averaged to give the subscales. The subscales are averaged to give the composite scores. The composite scores are averaged to give the overall score. The reliability and validity of the SF-36 is well documented in a variety of different patient groups, including patients with vascular diseases.

For each of the 8 dimensions, if less than 50% of the items which constitute that dimension are missing, the dimension score will be calculated by the mean of the available non-missing items. The physical health composite summary and mental health composite summary will only be calculated if a score has been calculated for all 4 dimensions that constitute the composite summary.

**Figure 1. SF-36 Model**



The mean change from baseline at weeks 12, 36, and 52 weeks will be computed for each treatment group. A paired t-test will be used to assess within treatment effect. Between treatment difference will be assessed using the MMRM model with baseline sub-score as a covariate, and treatment group, visit, interaction of treatment and visit, and stratification factors as fixed effects. The same strategy as that used in MMRM in Section 11.1.3.1 will be used to choose variance covariance. Data up to visit of Week 52 will be included in the analyses. Non-inferiority of roxadustat vs. ESA will be tested. The non-inferiority margin is fixed as a difference of 2 points.

HRQoL benefit will also be assessed using SF-36 vitality and physical functioning subscales. Mean change in following endpoints at above-mentioned time points will be analyzed and compared between the 2 treatment groups using MMRM using the baseline value as a covariate, treatment group, visit, the interaction of treatment and visit, adjusting for the stratification factors. The same strategy as that used in MMRM for Hb will be used to choose variance covariance structure. Other than below endpoints for exploratory analyses, other QoL variables may be performed

- Vitality Subscale of SF-36: In FAS subjects with baseline Vitality Sub-score below 50.
- Physical Component Scores of SF-36:
    - In FAS subjects with baseline *physical component scores* below 40.
    - In all FAS subjects.
- Anemia Subscale ("Additional Concerns") of Functional Assessment of Cancer Therapy-Anemia (FACT-An) Scores:
    - In FAS subjects with baseline subscale scores below 55 (generally associated with fatigue).
    - In all FAS subjects.
- Total FACT-An Scores:
    - In FAS subjects with baseline FACT-An scores below 135
    - In all FAS subjects.
- EQ-5D-5L Scores: In all FAS subjects.

### 11.3.9 Hepcidin, Iron, and HbA1c

The mean change from baseline in the following endpoints will be analyzed using the MMRM model with baseline value as a covariate and treatment group and stratification factors as fixed effects with the same strategy as that used in MMRM in Section 11.1.3.1 will be used to choose variance covariance and data up to visit of Week 52 will be included in the analyses:

- Change from baseline in serum hepcidin at each of the selected time points (e.g., Weeks 4, 12, 20, 44 and every 8 weeks onwards)

- Change from baseline in CHr at each of the selected time points (e.g., Weeks 4, 8, 12, 20, 28, and every 8 weeks onwards)

- Change from baseline in serum ferritin at each of the selected time points, total and sub-grouped by baseline values of <100 ng/mL, 100 to <400 ng/mL and >=400 ng/mL.
- Change from baseline in TSAT at each of the selected time points, total and sub-grouped by baseline values of <20%, 20% to <40%, and >=40%.
- Serum iron at each of the time points tested
- CHr at each timepoint tested (Weeks 4, 8, 12, 20, 28, 36, every 8 weeks onwards)
- Proportion of patients with CHr > ULN at each timepoint tested: Weeks 4, 8, 12, 20, 28, 36, every 8 weeks onwards)
- Change in HbA1c level at each of the selected time points in subjects without history of diabetes, in subjects with history of diabetes.

# 12 SAFETY ANALYSES

The safety analysis will be performed using the Safety Population. Safety parameters include adverse events, laboratory parameters, vital signs, ECG parameters, and physical examinations.

Safety interpretation will also be made based on analyses of composite endpoints derived from adjudicated events pooled across multiple studies in the roxadustat Phase 3 program per pooled statistical analysis plan (PSAP) on dialysis studies in the roxadustat program.

For each safety parameter, the last assessment made prior to the first dose of study medication will be used as the BL for all analyses of that safety parameter.

## 12.1 ADVERSE EVENTS

Adverse events will be coded using the latest MedDRA version.

### 12.1.1 Proportion of Subjects with TEAE

An AE (classified by preferred term) started during the treatment period will be considered a treatment-emergent adverse event (TEAE) if it was not present prior to the first dose of study medication, or it was present prior to the first dose of study medication but increased in severity during the treatment period up to 7 days after last dose of study drug or until the administration of another anemia drug (other than the randomized treatment). An AE that occurs more than 7 days after the last dose of study medication or after the administration of another anemia drug (other than the randomized treatment) will not be counted as a TEAE.

The number and percentage of subjects reporting TEAEs in each treatment group will be tabulated separately by system organ class and preferred term; by system organ class, preferred term, and severity; and by system organ class, preferred term, and relationship to study medication. If more than one event occurs with the same preferred term for the same patient, the patient will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to the study medication.

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

The incidence of common (≥ *5%* of subjects in any treatment group) TEAEs, treatment-emergent serious AEs (TESAE), and AEs leading to discontinuation of study medication will be summarized by preferred term and treatment group, sorted in decreasing overall (across treatments) frequency. In addition, the incidence of death and fatal SAEs (i.e., events that caused death) and incidence rate per PEY will be summarized separately by treatment group and preferred term.

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

Temporal profile of TEAEs of special interest may also be plotted by treatment group showing the subjects in the y-axis and time to these TEAEs in the x-axis (Appendix 8).

## 12.2 CLINICAL LABORATORY PARAMETERS

Descriptive statistics for laboratory values (in US conventional and SI units) and changes from baseline at each assessment time point will be presented by treatment group for the following laboratory parameters collected in the study including but are not limited to the following:

- Hematology: Hemoglobin, hematocrit, RBC count, MCV, MCH, MCHC, WBC count, WBC differential, platelet counts and Reticulocyte count;
- Chemistry: CPK, ALP, ALT, AST, total bilirubin, LDH, total protein, albumin, glucose, phosphate, uric acid, BUN, creatinine, sodium, and potassium; HbA1C;
- Serum iron, ferritin, TIBC, TSAT
- CHr
- Hepcidin
- CRP

Laboratory tests values are clinically significant (CS) if they meet either the low or high CS criteria. The number and percentage of subjects with post-baseline CS values will be tabulated by treatment group. The percentages are to be calculated relative to the number of subjects with available non-CS baseline values and at least one post-baseline assessment. The numerator is the total number of subjects with at least one post-baseline CS value. In addition, shift tables will be presented by treatment group and time point. The following 3 data listings will be presented by subject:

- A listing of lab values for all lab tests at all collected time points.
- A listing of subjects with post-baseline CS values will be provided including the baseline and post-baseline values.
- A listing of all AEs for subjects with CS laboratory values will also be provided.

## 12.3 VITAL SIGNS

Blood Pressures and Heart Rate baselines are defined as the mean of values obtained from the last 6 weeks of screening including Day 1 prior to the first dose. For subjects on hemodialysis, pre-dialysis vital signs will be used. For subjects on peritoneal dialysis, vital signs may be recorded at any time during the visit.

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

Vital sign values are potentially clinically significant (PCS) if they meet both the observed value criteria and the change from baseline criteria listed in Table 6 below. The number and percentage of subjects with post-baseline PCS values will be tabulated by treatment group. The percentages are to be calculated relative to the number of subjects with baseline and at least one post-baseline assessment. The numerator is the total number of subjects with at least one post-baseline PCS vital sign value. Shift tables may be presented. A supportive listing of subjects with post-baseline PCS values will be provided including the patient ID, study center, baseline, and post-baseline values. A listing of all AEs for subjects with PCS vital signs will also be provided.

Table 7. Criteria for Potentially Clinically Significant Vital Signs

| Vital Sign Parameter | Flag | Criteria* | |
|---|---|---|---|
| | | Observed Value | Change from Baseline |
| Systolic Blood Pressure (mmHg) | High | ≥ 170 | Increase of ≥ 20 |
| | Low | ≤ 90 | Decrease of ≥ 20 |
| Diastolic Blood Pressure (mmHg) | High | ≥ 100 | Increase of ≥ 15 |
| | Low | ≤ 50 | Decrease of ≥ 15 |
| Pulse Rate (bpm) | High | ≥ 120 | Increase of ≥ 20 |
| | Low | ≤ 50 | Decrease of ≥ 20 |
| Weight (kg) | High | - | Increase of ≥ 10% |
| | Low | - | Decrease of ≥ 10% |

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

Additional analyses include but are not limited to

- Subgroup analyses of patients without any change in BP meds during treatment period
- Proportion of subjects meeting NKF BP target: within sBP 120-140 mmHg/dBP 70-90 mmHg at baseline, during treatment, and 4 weeks post treatment

## 12.4 ELECTROCARDIOGRAM (ECG)

Descriptive statistics for ECG parameters (e.g., Heart Rate, PR interval, QRS interval, QT interval, and QTc interval) at baseline and changes from baseline at each assessment time point will be presented by treatment group. QTc interval will be calculated using both Bazett ($QTcB = QT/(RR)^{1/2}$) and Fridericia ($QTcF = QT/(RR)^{1/3}$) corrections; and if RR is not available, it will be replaced with 60/HR in the correction formula.

A plot for each parameter of mean (+/- 95% CI) versus visit will be produced by treatment group (roxadustat vs. ESA).

ECG parameters values are potentially clinically significant (PCS) if they meet or exceed the upper limit values listed in Table 8 below. The number and percentage of subjects with post-baseline PCS values will be tabulated by treatment group. The percentages are to be calculated relative to the number of subjects with available non-PCS baseline and at least one post-baseline assessment. The numerator is the total number of subjects with at least one post-baseline PCS ECG value. Shift tables may be presented. A listing for all subjects with post-baseline PCS value will be provided including the patient ID, study center, baseline, and post-baseline PCS values.

In addition, a listing of all TEAEs for subjects with PCS ECG values and a listing of subjects with post-baseline significant ECG abnormalities as reported by the investigators will also be provided.

Table 8. Criteria for Potentially Clinically Significant ECG

| ECG Parameter | Unit | Higher Limit |
|---|---|---|
| QRS interval | Msec | ≥ 150 |
| PR interval | Msec | ≥ 250 |
| QTc interval | Msec | > 500;<br>Change from baseline > 60 |

## 12.5 OTHER SAFETY ANALYSES

A separate meta-analysis SAP for pre-specified, adjudicated composite safety endpoints to assess Cardiovascular, cerebrovascular and thrombo-embolic Events will be developed to complement this study specific SAP.

# 13 ADDITIONAL AND SUBGROUP ANALYSES

The analysis of the primary endpoints of US and EU, and Hb change from baseline to the average level during Week 18 to 24 may be performed separately by sex, age group, baseline iron replete status, baseline CRP group (<= ULN vs > ULN) , and baseline stratification factors.

## 14 INTERIM ANALYSIS

Safety data and dosing decisions will be monitored on an ongoing basis. Additional ongoing review of safety data will be conducted by an independent DSMB (see protocol Section 3.7).

## 15 REFERENCES

ICH Harmonized Tripartite Guideline E3. Structure and Content of Clinical Study Reports, November 1995. (www.ich.org; Guidelines; "Efficacy" Topics)

ICH Harmonized Tripartite Guideline E9. Statistical Principles for Clinical Trials, February 1998. (www.ich.org; Guidelines; "Efficacy" Topics)

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

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

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

Haybittle, J. L. (1971), "Repeated Assessment of Results in Clinical Trials of Cancer Treatment," *British Journal of Radiology*, 44, 793-797.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., and Smith, P. G. (1976), "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient: I. Introduction and Design," *British Journal of Cancer*, 34, 585-612.

FDA Guidance for Industry Non-Inferiority Clinical Trials, 2010 (http://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf).

Little RJA, Rubin D.1987. Statistical Analysis with Missing Data. John Wiley and Sons.

Ratitch B, Kelly M 2011. Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures, PharmaSUG2011 – Paper SP4.

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

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

# 16 APPENDIX

## 16.1 APPENDIX 1: SCHEDULE OF ASSESSMENTS

| **Study Period:** | **Screening** | | | **Treatment** | | | | | **Follow-up** |
|---|---|---|---|---|---|---|---|---|---|
| **Visit / Week:** | **Up to 6 Weeks** [a] | | | **Day 1 (Wk 0)** | **Weekly (Wks 1 to 4)** ± 2 days | **Every 2 Weeks (Wks 6 to 24)** ± 2 days | **Every 4 Weeks (Wks 28 to EOT)** ± 3 days [b] | **EOT or ET**[l] ± 3 days | **EOS (4 wks post EOT or ET)** ± 7 days |
| | **1** | **2** | **3** | | | | | | |
| Written informed consent | X | | | | | | | | |
| Eligibility criteria | X | | | X | | | | | |
| Demographics and medical history | X | | | | | | | | |
| Physical examination | X | | | X | | Wks 12 [c], 24 [c] | Wks 36 [c], Q12wk [c, d] | X | X [c] |
| Height, weight | X | | | X [e] | | | Wk 24 and every 24 wks [e] | | |
| Blood pressure, heart rate, respiratory rate, temperature [f] | X | X | X | X | X | X | X | X | X |
| Hemoglobin | | X | X | | | X [g] | X [g] | | |
| CBC with WBC differential | X | | | X | X | Wks 8. 12, 20 | Wk 28, Q8wk | X | X |
| Serum chemistry | X | | | X | Wk 2 | Wks 8,12, 20 | Wk 28, Q8wk | X | X |
| LFTs, CPK | | | | | Wk 2 | Wks 6, 16 | | | |
| Lipid panel (whenever fasting possible) | X | | | X | Wk 4 | Wks 8, 12, 24 | Wks 32, 40, 48, 60, Q24wk | X | X |
| Serum iron, ferritin, TIBC, TSAT | X | | | X | Wk 4 | Wks 8, 12, 20 | Wk 28, Q8wk | X | X |
| CHr | X | | | X | Wk 4 | Wks 8, 12, 20 | Wk 28, Q8wk | X | X |
| HbA1c | X | | | X | | Wk 12 | Wks 28, 44, 60 Q16wk [d] | X | X |
| Vitamin $B_{12}$, folate | X | | | | | | | | |
| HIV ELISA, HBsAg, anti-HCV Ab | X | | | | | | | | |
| Serum hCG pregnancy test | X [h] | | | | | Wks 12, 24 | Wk 36, then Q12wks | X | X |
| Reticulocyte count | | | | X | Wks 1, 2 | Wks 8, 20 | Wk 44, Q24wk | X | X |
| Special laboratory analytes (hepcidin, hs-CRP) | | | | X | Wk 4 | Wks 12, 20 | Wk 44, Q24wk | X | X |
| Optional archival serum/plasma samples | | | | X | Wk 4 | Wks 12, 20 | Wk 44, Q24wk [d] | X | X |
| HemoCue® assessment | | | | X | X | X | X | | |
| Quality-of-life questionnaires | | | | X | | Wk 12 | Wks 36, 52 [d] | X | |
| 12-lead ECG | | | | X | | | Wk 24 and every 24 wks | X | |
| Renal ultrasound [i] | | X | | | | | | | |
| Dose adjustment [J] | | | | | X | X | X | | |
| Adverse event recording | X | X | X | X | X | X | X | X | X |
| Concomitant medication recording | X | X | X | X | X | X | X | X | X |
| Procedure and nondrug therapy recording | X | X | X | X | X | X | X | X | X |
| Study drug dispensing [k] | | | | X | X [l] | X | X | | |

**Footnotes on following page**

Abbreviations: Ab = antibody; BP = blood pressure; CBC = complete blood count; CHr = reticulocyte hemoglobin content; ELISA = enzyme-linked immunosorbent assay; EOT = End of Treatment; EOS = End of Study; ET = early termination; Hb = hemoglobin; HbA1c = glycated hemoglobin A1c; HBsAg = hepatitis B surface antigen; hCG = human chorionic gonadotropin; HCV = hepatitis C virus; HD = hemodialysis; HIV = human immunodeficiency virus; HR = heart rate; HRQoL = health-related quality of life questionnaire; HS-CRP = high-sensitivity C reactive protein; ICF = informed consent form; LFTs = liver function tests; PE = physical examination; RBC = red blood cell; RR = respiratory rate; SmPC = summary of product characteristics; TIBC = total iron binding capacity; TSAT = transferrin saturation; TX = treatment; WBC = white blood cells; Wk(s) = week(s); X = mandatory test/assessment.

a Screening Hb values must be obtained at least 4 days apart.

b Treatment duration is variable for each subject. All subjects will remain in the study treatment until up to 3 years s after the last subject randomized.

c Targeted PE only (e.g., respiratory and cardiovascular).

d If the indicated assessments fall on a study treatment visit that is within two weeks of the planned EOT visit then these specified assessments can be postponed until the EOT visit.

e Weight only (HD subjects: use dry weight).

f Perform HR and BP at all week visits. Additional respiratory rate and temperature are measured at Day 1/Week 0 and EOT/ET.

g Dedicated Hb sample for central lab should be collected during the visits where CBC is not collected.

h Collect from female subjects of child bearing potential only.

i A renal ultrasound examination will be performed during screening if no record of a renal imaging modality exists within 12 weeks prior to randomization.

j Roxadustat subjects: Dose adjustments will be permitted from Week 4 onward, and every 4 weeks thereafter (except in extenuating circumstances) to correct and maintain subjects to a target Hb range. Please refer to dose adjustment rules as stated in **Error! Reference source not found.**.

Epoetin alfa subjects: Dose adjustments for HD subjects receiving epoetin alfa will follow the country specific product labeling for dosing and dose adjustments (e.g., PI; SmPC). For PD subjects, local standard of care may be followed for dosing and dose adjustments.

k All assessments should be done prior to first study drug administration.

l Dispense every other week

m Subjects who prematurely discontinue study treatment will complete the ET and EOS visits, and – unless consent is withdrawn – will continue to be followed for CV events of interest, vital status and hospitalizations until study closure. These subjects will be asked to return for study visits every 3 to 6 months, or be available via telephone.

## 16.2 APPENDIX 2: DATA HANGLING CONVENTIONS

### 16.2.1 VISIT TIME WINDOW

Table 9 below presents the visits assigned for efficacy and safety analyses corresponding to the range of treatment days (window) during which an actual visit may have occurred.

**Table 9. Analysis Visit Windows**

| Derived Visit | Scheduled Visit Day [a] | Window |
|---|---|---|
| Baseline, Week 0 | Day 1 | Days ≤ 1 |
| Week 1 | Day 7*(Week #)+1 | [Day 2, 10] |
| Weeks 2-3 | Day 7*(Week #)+1 | Days [Scheduled Day ± 3 ] |
| Week 4 | Day 7*(Week #)+1 | [Scheduled Day -3, Scheduled Day +6] |
| Weeks 4-22 | Day 7*(Week #)+1 | [Scheduled Day -7, Scheduled Day +6] |
| Week 24 | Day 7*(Week #)+1 | [Scheduled Day -7, Scheduled Day +13] |
| Week 24 to xx | Day 7*(Week#)+1 | [Scheduled Day -14, Scheduled Day +13] |
| ET | Earlier Termination, Match to a closest scheduled visit in protocol if patient had not been off drug for more than 7 days. | |
| EoT | Last assessment between Day 2 and EOT visit day, match to a closest scheduled visit in protocol if patient had not been off drug for more than 7 days. | |
| EoS (FU-4Wk) | Final visit for the Study 15 – 31 days after the last dose (excluding long term follow-up for early termination) | |

[a]: Relative to the first study medication date. For example, Day 1 = the first dose date of study medication.

Analysis Visit windows, as depicted in Table 10 below, will be used for the quality of life efficacy study assessments:

**Table 10: 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 12 | Day 7 * (Week #) + 1 | [Target Day -14, Target Day + 27] | Week 12 |
| 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 |
| EOT Visit | | Last assessment between Day 2 and EOT visit day, remapped to the closest next scheduled visit for HRQoL collection. | Week 12, 36, 52 and > 1 year |

[a]: Relative to Day 1 (first dose date of study medication)

**Table 11. Analysis Visit Windows for Lipid Panel**

| Derived Visit | Scheduled Visit Day [a] | Window |
|---|---|---|
| Baseline, Week 0 | Day 1 | Days 1 |
| Week 4 | Day 7*(Week #)+1 | [Day 2, 42] |
| Week 8 | Day 7*(Week #)+1 | [Scheduled Day-14, Scheduled Day +13] |
| Weeks 12 | Day 7*(Week #)+1 | [Scheduled Day -14, Scheduled Day +27] |
| Week 24 | Day 7*(Week #)+1 | [Scheduled Day -28, Scheduled Day +27] |
| Weeks 32-40 | Day 7*(Week #)+1 | [Scheduled Day -28, Scheduled Day +27] |
| Week 48 | Day 7*(Week #)+1 | [Scheduled Day -28, Scheduled Day +41] |
| Week 60 | Day 7*(Week #)+1 | [Scheduled Day -42, Scheduled Day +83] |
| Week 84 to xx | Day 7*(Week#)+1 | [Scheduled Day -84, Scheduled Day 83] |
| EoT | Last assessment between Day 2 and EOT visit day, match to a closest scheduled visit in protocol if patient had not been off drug for more than 7 days | |

[a]: Relative to Day 1 (first dose date of study medication)
Visit Day is calculated by (visit date - date of first study medication + 1). If a patient has ≥ 2 actual visits within the same window, the last visit with non-missing value will be used for analysis.

### 16.2.2 Repeated or Unscheduled Assessments of Safety Parameters

If a patient has repeated assessments prior to the start of study medication, then the results from the final assessment made prior to the start of study medication will be used as baseline. If end of study assessments are repeated or unscheduled, the last post-baseline assessment will be used as the end of study assessment for generating summary statistics. However, all post-baseline assessments will be used for PCS value determination and all assessments will be presented in the data listings.

### 16.2.3 Missing Date of Study Medication

When the last date of study medication during the study treatment phase is missing, all efforts should be made to obtain the date from the investigator. If it is still missing after all efforts, then the last dispensation visit date during the treatment period will be used in the calculation of treatment duration.

### 16.2.4 Missing Severity Assessment for Adverse Events

If severity is missing for an AE started prior to the first study medication, then a severity of "Mild" will be assigned. If the severity is missing for an AE started on or after the first study medication dosing, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summary, while the actual missing values will be presented in data listings.

### 16.2.5 Missing Relationship to Study Drug for Adverse Events

If the relationship to the study medication is missing for an AE started after baseline, a causality of "Related" will be assigned. The imputed values for relationship to study medication will be used for incidence summary, while the actual values will be presented in data listings.

### 16.2.6 Missing Date Information for Adverse Events

The following imputation rules only apply to the case where the start date is incomplete (i.e., partial missing) for adverse events.

**Incomplete Start Date**

*Missing day and month*

- If the year is same as the year of first day on study medication, then the day and month of the start date of study medication will be assigned to the missing fields.
- If the year is not the same as the year of first day on study medication, then January 1 will be assigned to the missing fields.

*Missing month only*

Treat day as missing and replace both month and day according to the above procedure.

*Missing day only*

If the month and year are same as the year and month of first day on study medication, then the start date of study medication will be assigned to the missing day.

If the month and year are not the same as the year and month of first day on study medication, then the first day of the month will be assigned to the missing day.

- **Incomplete Stop Date**

If the stop date is complete and the imputed start date as above is after the stop date, the start date will be imputed by the stop date.

If needed, the following imputation rules apply to the case where the end date is incomplete (i.e., partially missing) for adverse events. Other partial end date will not be imputed.

Missing day and month, or Missing month only

December 31 will be assigned to the missing fields.

Missing day only

The last day of the month will be assigned to the missing day.

**Table 16.6-1 Imputation of the Analysis Adverse Event Start Date**

| Reported Date | Date of First Drug Intake | Analysis Date (Derived) |
|---|---|---|
| --/MM/YYYY | DD/MM/YYYY | |
| --/02/2008 | 14/02/2008 | 14/02/2008* |
| --/02/2008 | 14/02/2007 | 01/02/2008 |
| --/02/2008 | 14/02/2009 | 01/02/2008 |
| --/--/YYYY | DD/MM/YYYY | |
| --/--/2008 | 14/02/2008 | 14/02/2008 |
| --/--/2008 | 14/02/2007 | 01/01/2008 |
| --/--/2008 | 14/02/2009 | 01/01/2008 |
| DD/--/----<br>--/MM/----<br>--/--/---- | | No imputation |

**Table 16.6-2 Imputation of the Analysis Adverse Event Stop Date**

| Reported Date | Analysis Date (Derived) * |
|---|---|
| --/MM/YYYY | 31/MM/YYYY or<br>30/MM/YYYY or<br>29/MM/YYYY or<br>28/MM/YYYY |
| --/--/YYYY | 31/12/YYYY |
| DD/--/----, or<br>--/MM/----, or<br>--/--/---- | No imputation |

*Death has to be taken into consideration when calculating this.

### 16.2.7 Missing Date Information for Prior or Concomitant Medications

For prior or concomitant medications, including rescue medications, incomplete (i.e., partial missing) start date and/or stop date will be imputed. When the start date and the stop date are both incomplete for a patient, impute the start date first.

- **Incomplete Start Date**

The following rules will be applied to impute the missing numerical fields. If the stop date is complete and the imputed start date is after the stop date, then the start date will be imputed using the stop date.

*Missing day and month*

If the year of the incomplete start date is the same as the year of the first dose date of study medication, then the day and month of the first dose date will be assigned to the missing fields.

If the year of the incomplete start date is not the same as the year of the first dose date of study medication, then January 1 will be assigned to the missing fields.

*Missing month only*

Treat day as missing and replace both month and day according to the above procedure.

*Missing day only*

If the month and year of the incomplete start date are the same as the month and year of the first dose date of study medication, then the day of the first dose date will be assigned to the missing day.

If the month and year of the incomplete start date are the same as the first dose date of study medication, then the first day of the month will be assigned to the missing day.

- **Incomplete Stop Date**

The following rules will be applied to impute the missing numerical fields, if needed. If the last dose date of study medication is missing, impute it with the last visit date. If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be equal to the start date.

*Missing day and month*

If the year of the incomplete stop date is the same as the year of the last dose date of study medication, then the day and month of the last dose date will be assigned to the missing fields.

If the year of the incomplete stop date is not the same as the year of the last dose date of study medication, then December 31 will be assigned to the missing fields.

*Missing month only*

Treat day as missing and replace both month and day according to the above procedure.

*Missing day only*

If the month and year of the incomplete stop date are the same as the month and year of the last dose date of study medication, then the day of the last dose date will be assigned to the missing day.

If the month and year of the incomplete stop date are not the same as the month and year of the last dose date of double-blind study medication, then the first day of the month will be assigned to the missing day.

### 16.2.8 Missing Date Imputation for last dose date

Imputed last dose date = earliest date of (last drug dispense date + number of days of drug dispensed, date of death, date of EOT/EOS visit, and other dates as appropriate).

### 16.2.9 Character Values of Clinical Laboratory Parameters

If the reported value of a clinical laboratory parameter cannot be used in a statistical summary table due to, for example, that a character string is reported for a parameter of the numerical type, coded value needs to be appropriately determined and used in the statistical analyses. However, the actual values as reported in the database will be presented in data listings.

**Table 12. Example for Coding of Special Character Values for Clinical Laboratory Parameters**

| Lab Test | Possible Lab Results (in SI unit) | Coded Value for Analysis |
|---|---|---|
| Urinalysis: Ketones | = OR > 8.0, >=8.0, > 0 | Positive |
| | <= 0, Negative | Negative |
| Urinalysis: pH | > 8.0, >= 8.0 | 8.0 |
| | >= 8.5, | 8.5 |
| Urinalysis: Protein | = OR > 3.0, >=3.0, > 0 | Positive |
| | <= 0 | Negative |

## 16.3 APPENDIX 3: RANGES OF POTENTIALLY CLINICALLY SIGNIFICANT LAB VALUES

| Parameter | SI Unit | Lower Limit | Higher Limit |
|---|---|---|---|
| **CHEMISTRY** | | | |
| Alanine Aminotransferase (ALT) | U/L | | ≥3 * ULN |
| Alkaline Phosphatase | U/L | | ≥3 * ULN |
| Aspartate Aminotransferase (AST) | U/L | | ≥3 * ULN |
| GGT | U/L | | ≥3 * ULN |
| Calcium | mmol/L | <0.8*LLN | >1.2 * ULN |
| Creatinine | µmol/L | | > 1.5x Baseline value |
| Potassium | µmol/L | <0.75*LLN | >1.2 * UNL |
| Sodium | mmol/L | <0.9*LNL | >1.1 * UNL |
| Total Bilirubin | µmol/L | | >1.5 * UNL |
| Total Protein | µmol/L | <0.9*LNL | >1.1 * UNL |
| Urea (BUN) | mmol/L | | >1.5X Baseline value |
| HEMATOLOGY | | | |
| Neutrophils | $10^9$/L | ≤1 | |
| Platelet Count | $10^9$/L | ≤ 100 | ≥700 |
| White Blood Cell Count | $10^9$/L | ≤2.5 | ≥15 |
| LLN: Lower limit of normal, value provided by the laboratory<br>ULN: Upper limit of normal, value provided by the laboratory | | | |

## 16.4 APPENDIX 4: 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. **Compared to one year ago, how would you rate your health in general now?**

| 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 your health now limit you 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 past 4 weeks, 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 bodily pain have you had during the past 4 weeks?**

| None | Very mild | Mild | Moderate | Severe | Very severe |
|---|---|---|---|---|---|
| ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 | ☐ 6 |

8. **During the past 4 weeks, how much did pain 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!***

## 16.5 APPENDIX 5: FACT-AN (VERSION 4)

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 |

## 16.6 APPENDIX 6: JUSTIFICATION FOR THE NONINFERIORITY MARGIN OF 0.75 MG/DL

The **FDA draft Guidance for Industry: Non-Inferiority Clinical Trials** states that "the NI trial concept depends on how much is known about the size of the treatment effect the active comparator will have in the NI study compared to no treatment" and this effect "must be assumed, based on an analysis of past studies of the control".

Pursuant to this, we sought out studies that compared ESA therapies currently approved for use in the United States to placebo (or no therapy) in the treatment of anemia. Criteria to be considered relevant to estimate the potential treatment effect included:

1- Randomized trial design
2- Prospective follow up
3- Treatment groups which included epoetin-alfa (or other recombinant erythropoeitin derivatives) and either placebo or no treatment
4- Inclusion of treatment naïve adult patients with CKD or ESRD related anemia.
5- Post randomization monitoring of hemoglobin following randomization
6- Reporting either hemoglobin or hematocrit summary measures at baseline and during follow-up

Studies were identified through a search of the bibliographies peer-reviewed meta-analyses examining the effect of ESAs on outcomes.

1. Palmer et al. Meta-analysis: Erythropoiesis-Stimulating Agents in Patients with Chronic Kidney Disease. *Ann Intern Med.* 2010;153:23-33.
2. Phrommintikul A et al. "Mortality and target haemoglobin concentrations in anaemic patients with chronic kidney disease treated with erythropoietin: a meta-analysis." *The lancet* 369.9559 (2007): 381-388.
3. Koulouridis, Ioannis, et al. "Dose of erythropoiesis-stimulating agents and adverse outcomes in CKD: a metaregression analysis." *American Journal of Kidney Diseases* 61.1 (2013): 44-56.

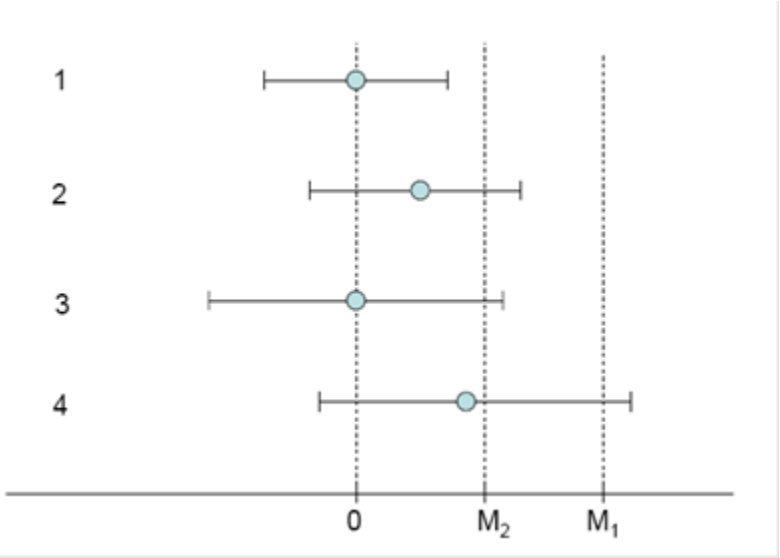
Identification of all appropriate trials was confirmed through a literature search (www.pubmed.gov) using combinations of the terms "ESA", "epoetin", "CKD", "ESRD", "placebo", and "anemia".

The approach taken for this NI margin is the fixed margin method described in the U.S. Food and Drug Administration's 2010 guidance "Non-Inferiority Clinical Trials." In this guidance, the fixed margin approach is summarized thus:
"Briefly, in the fixed margin method, the margin $M_1$ is based upon estimates of the effect of the active comparator in previously conducted studies, making any needed adjustments

for changes in trial circumstances. The NI margin is then pre-specified and it is usually chosen as a margin smaller than $M_1$ (i.e., $M_2$), because it is usually felt that for an important endpoint a reasonable fraction of the effect of the control should be preserved. The NI study is successful if the results of the NI study rule out inferiority of the test drug to the control by the NI margin or more. It is referred to as a fixed margin analysis because the past studies comparing the drug with placebo are used to derive a single fixed value for $M_1$, even though this value is based on results of placebo-controlled trials (one or multiple trials versus placebo) that have a point estimate and confidence interval for the comparison with placebo. The value typically chosen is the lower bound of the 95% CI (although this is potentially flexible) of a placebo-controlled trial or meta-analysis of trials. This value becomes the margin $M_1$, after any adjustments needed for concerns about constancy.

The fixed margin $M_1$, or $M_2$ if that is chosen as the NI margin, is then used as the value to be excluded for C-T in the NI study by ensuring that the upper bound of the 95% CI for C-T is $<M_1$ (or $M_2$). This 95% lower bound is, in one sense, a conservative estimate of the effect size shown in the historical experience. It is recognized, however, that although we use it as a "fixed" value, it is in fact a random variable, which cannot invariably be assumed to represent the active control effect in the NI study." This approach is shown schematically in Figure 2.

**Figure 2 Active Control – Test Drug differences (Point estimate, 95% CI)**




"1. C-T point estimate = 0 and upper bound of 95% CI < M2, indicating test drug is effective (NI demonstrated).

2. Point estimate of C-T favors C and upper bound of 95% CI < M1 but > M2, indicating effect > 0 but unacceptable loss of the control effect.

3. Point estimate of C-T is zero and upper bound of 95% CI < M1 but it is slightly greater than M2. Judgment could lead to conclusion of effectiveness.

4. C-T point estimate favors C and upper bound of 95% CI > M1, indicating there is no evidence of effectiveness for test drug."

Table 13 below displays the mean change in hemoglobin for recombinant human (EPO) based on 3 publications, [Canadian et al, 1990], [Bennett et al, 1991] and [Nissenson et al, 1995 ]. The 3 studies are all randomized, double-blinded, randomized, placebo controlled study to evaluate the effect EOP (epoetin alfa or beta) in anemia patient with renal End Stage Renal Disease with Peritoneal Dialysis or Hemodialysis.

Using the data in Table 13 below, the weighted mean of the point estimate of treatment effect in mean change from baseline for EPO (mean change in hemoglobin or hemoglobin equivalence in EPO group – mean change in placebo group) was calculated, along with its 95% confidence interval.

**Table 13 Historical studies with mean change and SD in hematocrit or hemoglobin available in End Stage Renal Disease (ESRD): EPO versus Placebo**

| Nissenson et al, 1995 (Erythropoietin in Peritoneal Dialysis) | Endpoint | EPO Mean (SD) | Placebo Mean (SD) | EPO-Placebo Mean (Std Err) |
|---|---|---|---|---|
| | | sample size = 78 | Sample size = 74 | |
| | Baseline in hematocrit | 23.8 (3.8) % | 23.8 (3.3) % | |
| | 12 week Follow up in hematocrit | 33.7 (4.8) % | 24.1 (3.8) % | |
| | HMG equivalent Mean Change* | 3.3(0.33) g/dL | 0.1( 0.17) g/dL | 3.2(0.042) g/dL |
| Bennett et al, 1991 (Epoetin Beta - Hemodialysis) | | sample size = 90 | sample size = 41 | |
| | Baseline | 7.1 (0.1*sqrt(90)) g/dL | 6.8 (0.2*sqrt(41)) g/dL | |
| | 12 week Follow up | 11.1 (0.2*sqrt(90)) g/dL | 7.6 (0.3*sqrt(41)) g/dL | |
| | HMG Mean Change* | 4.0 (0.95) g/dL | 0.8(0.64) g/dL | 3.2(0.141) g/dL |
| Canadian et al, 1990 (Erythropoietin - hemodialysis patients) | | sample size = 40 | sample size = 40 | |
| | Baseline | 7.1 (0.9) g/dL | 6.9 (1.0) g/dL | |
| | 6 month follow up | 10.2 (1.0) g/dL | 7.4(1.2) g/dL | |
| | HMG Mean Change* | 3.1(0.1) g/dL | 0.5 (0.2) g/dL | 2.6(0.035) g/dL |
| | | | | |

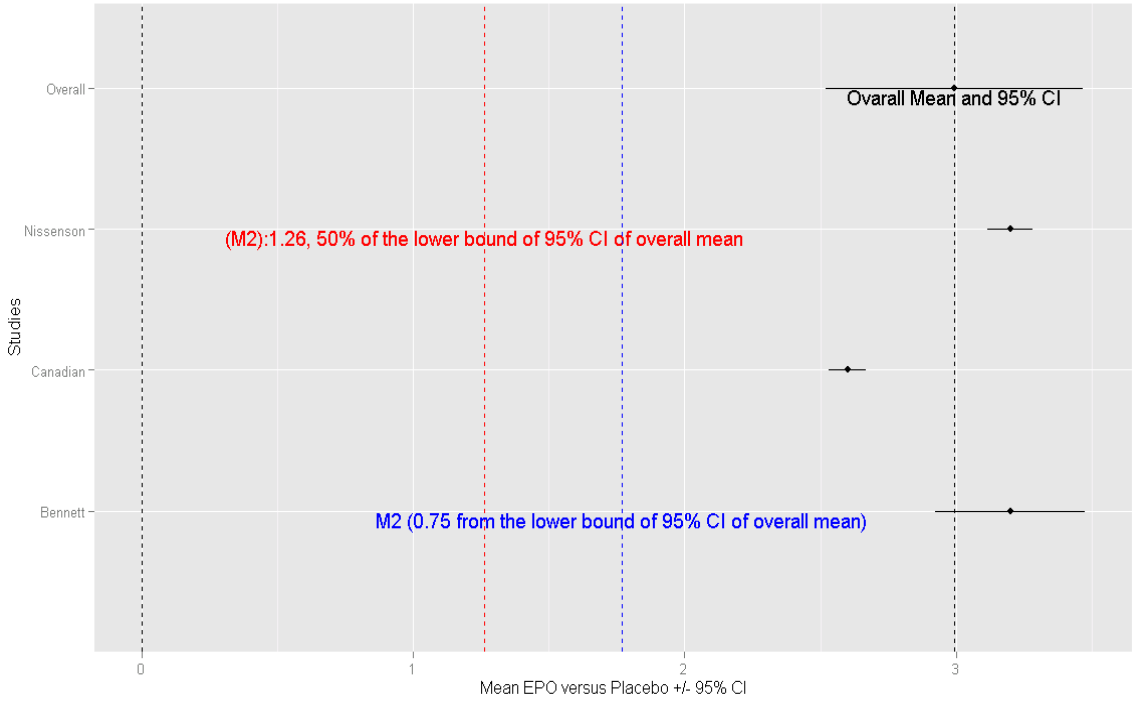
***Hematocrit in %** was converted to Hemoglobin in g/dL by dividing by 3. Due to the information are limited from the reference, the mean change in hemoglobin and the corresponding standard deviation were derived where the baseline and follow up hemoglobin were assumed to be independent which is very conservative.

The weighted mean point estimate for mean change from baseline in hemoglobin for EPO (mean change from baseline in hemoglobin for EPO group – mean change from baseline in hemoglobin for placebo group) was estimated to be 2.859 g/dL by treating the studies as fixed effect, with a 95% confidence interval of (2.806, 2.911) and 2.993 g/dL by treating the studies as random effect, with a 95% confidence interval of (2.520, 3.466) from meta-data analysis. The lower bound of this 95% confidence interval by treating the study as random effect, 2.520 is taken to be the NI margin $M_1$. To ensure that not more than 50% of the effect of EPO was lost, giving an NI margin ($M_2$) of 1.26. However, a very conservative $M_2$ margin was chosen, 0.75, which is 29.76% of $M_1$, thus ensuring preservation of 70.24% of $M_1$ in this study.

While for study with responder as the primary endpoint, Roth et al 1994 study results were used to calculate the NI margin. The study was undertaken to ascertain the effects of recombinant human erythropoietin (EPO) on renal function in chronic renal failure pre-dialysis patients. The study defined an increase to a hematocrit of at least 36% (Hemoglobin > = 12 g/dL) as correction or responder. There are total 43 subjects in EPO treatment group and 40 in the placebo group. Out of 43 subjects in EPO treatment group, 34 subjects were responder and none were responders from placebo group.

95% CI (66.9, 91.2) % was calculated for the estimated difference in responder rate, 79.1%. The lower bound of this 95% confidence interval 0.669 is taken to be the NI margin $M_1$. To ensure that not more than 50% of the effect of EPO was lost, giving an NI margin ($M_2$) of 0.335. Similarly, a very conservative $M_2$ margin was chosen, 0.15, which is 22.42% of $M_1$, thus ensuring preservation of 77.58% of $M_1$ in this study.

**Figure 2 Active Control – Placebo differences (Point estimate, 95% CI) in History Studies**



## 16.7 APPENDIX 7: TEMPEROAL PROFILE OF TEAES OF SPECIAL INTEREST