AKB-6548-CI-0014 and AKB-6548-CI-0015

This Supplement Contains:

- Combined Statistical Analysis Plan (SAP) for study AKB-6548-CI-0014 (NCT02648347) and AKBˇ6548-0015(NCT02680574) (only 1 version)
- Combined SAP to monitor Major Adverse Cardiovascular Events (MACE) for study AKB-6548-CI-0014 (NCT02648347) and AKBˇ6548-0015(NCT02680574) (only 1 version)

Date: 29 Jan 2020

Statistical Analysis Plan

PHASE 3, RANDOMIZED, OPEN-LABEL, ACTIVE-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF ORAL VADADUSTAT FOR THE CORRECTION OF ANEMIA IN SUBJECTS WITH NON-DIALYSIS-DEPENDENT CHRONIC KIDNEY DISEASE (NDD-CKD)

(PRO$_2$TECT – CORRECTION, AKB-6548-CI-0014)

and

PHASE 3, RANDOMIZED, OPEN-LABEL, ACTIVE-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF ORAL VADADUSTAT FOR THE MAINTENANCE TREATMENT OF ANEMIA IN SUBJECTS WITH NON-DIALYSIS-DEPENDENT CHRONIC KIDNEY DISEASE (NDD-CKD)

(PRO$_2$TECT – CONVERSION, AKB-6548-CI-0015)

AKB-6548-CI-0014 and AKB-6548-CI-0015

Sponsored by:

Akebia Therapeutics, Inc.

Version: 1.0

Confidentiality Statement

1 TABLE OF CONTENTS

2 LIST OF TABLES

3 LIST OF APPENDIX TABLES

4 LIST OF APPENDIX FIGURES

5 LIST OF APPENDIX LISTINGS

6 LIST OF APPENDIX EXHIBITS

Abbreviations

Abbreviation	Definition
AE	adverse event
AESI	adverse events of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
CBC	complete blood count
CHF	congestive heart failure
CKD	chronic kidney disease
CRF	case report form
CRP	c-reactive protein
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DA	darbepoetin alfa
DBP	diastolic blood pressure
DD	dialysis dependent
ECG	electrocardiography
eCRF	electronic case report form
eGFR	estimated glomerular filtration rate
Epo	erythropoietin
EpoA	erythropoietin alfa
EOS	end of study
EOT	end of treatment
ESA	erythropoietin-stimulating agent
EU	European Union
FAS	Full Analysis Set
FCS	fully conditional specification
Hb	hemoglobin
HDL	high-density lipoprotein
HLGT	gigh-level group term
HLT	high-level term
HR	hazard ratio
IDMC	Independent Data Monitoring Committee
IV	intravenous

Abbreviation	Definition
IWR	interactive web response
LDL	low-density lipoprotein
LLD	lower limits of detection
MACE	major adverse cardiovascular event(s)
MAR	missing at random
MNAR	missing not at random
MMRM	mixed models for repeated measurements
MedDRA	Medical Dictionary for Regulatory Activities
NDD-CKD	non-dialysis-dependent chronic kidney disease
NEC	not elsewhere classfied
NYHA	New York Heart Association
PEP	primary efficacy period
PK	pharmacokinetic
PP	per protocol
PT	preferred term
RBC	red blood cell
ROW	Rest of World
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SC	subcutaneous
SD	standard deviation
SEP	secondary efficacy period
SMQ	Standardized MedDRA Query
SOC	system organ class
TEAE	treatment-emergent adverse event
TIBC	total iron-binding capacity
TSAT	transferrin saturation
uACR	urine albumin-to-creatinine ratio
ULN	upper limit of normal
US	United States
VEGF	vascular endothelial growth factor
WBC	White blood cells

1 INTRODUCTION

This statistical analysis plan (SAP) covers the two PRO$_2$TECT studies, Protocols AKB-6548-CI-0014 and AKB-6548-CI-0015, hereafter referred to as CI-0014 and CI-0015, respectively.

The SAP supports the Statistical Methods Section of the clinical study reports (CSRs). It elaborates upon the protocol-specified endpoint definitions and the formal statistical methods that will be used in analyzing both studies. Unless otherwise specified, all analyses will be conducted for each study separately.

Throughout this SAP, planned analyses of both studies are identical to each other except in those places that specifically identify where the plans diverge. Akebia does not intend to pool data together for formal efficacy and general safety analyses; however, analyses of major adverse cardiovascular events (MACE) will be pooled from both studies. A separate SAP will describe the planned methodology for the MACE analyses.

If the analyses described in the protocols differ from those in this SAP, the methods of the SAP prevail.

This SAP contains language and programming code that specifies the exact intent of each analysis. The sponsor will finalize and sign this document prior to locking the database and unblinding (see Section 2.5). Many of the analyses described in this SAP are quite complex. A blinded team of clinicians and statisticians will review the data carefully to develop conventions and analytic methods not anticipated in writing this SAP.

2 DESCRIPTION OF THE STUDIES

Both PRO$_2$TECT studies are Phase 3, randomized, open-label, active-controlled trials of the efficacy and safety of vadadustat compared to darbepoetin alfa (DA). CI-0014 addresses the correction of anemia and maintenance of hemoglobin (Hb) in subjects with anemia secondary to non-dialysis-dependent chronic kidney disease (NDD-CKD). CI-0015 addresses the maintenance treatment of anemia after conversion from current erythropoietin-stimulating agent (ESA) therapy in subjects with anemia secondary to NDD-CKD.

2.1 Randomization and Stratification

In each study, subjects who meet all inclusion and no exclusion criteria are randomized 1:1 to vadadustat or darbepoetin alfa. Randomization is stratified by the following three variables:

- Geographic region: United States (US); European Union (EU); Rest of World (ROW)
- New York Heart Association (NYHA) congestive heart failure (CHF) class: 0 or I; II or III

In addition, each study has the following strata defined by Hb levels at entry:

Study entry Hb level	*CI-0014*	*CI-0015*
	<9.5 g/dL; ≥9.5 g/dL	<10 g/dL; ≥ 10 g/dL

2.2 Study Periods

Following randomization, the studies have the following initial study periods:

Study CI-0014

- **Correction Period (Weeks 0 to 23):** initial period on study medication for correction of Hb
- Maintenance Period (Weeks 24 to 52):
  - Weeks 24 to 36: primary efficacy period (PEP)
  - Weeks 40 to 52: secondary efficacy period (SEP)

Study CI-0015

- Conversion and Maintenance Period (Weeks 0-52):
  - Weeks 0 to 23: conversion to study treatment for maintaining Hb
  - Weeks 24 to 36: PEP
  - Weeks 40 to 52: SEP

After Week 52, subjects in both studies will proceed to the following 2 periods:

- **Long-Term Treatment Period (Weeks 53-End of Treatment [EOT]):** continued treatment with study medication to assess long-term safety
- **Follow-up Period (EOT + 4 weeks):** post-treatment visit (either in person or by telephone) to assess safety

The duration of the trial will depend on how long it takes for a total of 631 adjudicated MACE to occur in the 2 studies combined. A separate SAP will describe the planned MACE analyses, including the method planned for pooling data from the 2 PRO$_2$TECT studies.

The safety reporting period for a subject enrolled in this study begins upon randomization and ends at the final protocol-required study contact (visit or telephone), also known as a subject's end of study (EOS). In addition, the Investigator should report any adverse event (AE) that occurs after this period if he or she assesses it as possibly or probably related to the study medication.

According to the schedule of visits in the protocol, no laboratory assessments, physical exams, or other procedures are planned after the subject's permanent discontinuation of study medication, also known as EOT; however, AE and rescue therapy assessments, including monitoring for MACE endpoints, will continue until a subject's EOS.

The analysis time periods for efficacy divide the Correction/Conversion and Maintenance periods, Year 1 collectively, into four windows (see Table 1).

2.3 Target Hb Ranges

In several places, this SAP refers to the following target Hb ranges as specified in the 2 protocols:

- US target: 10-11 g/dL
- EU, ROW target: 10-12 g/dL

2.4 Primary Objectives of the Trials

The primary objectives of CI-0014 trial are to compare the efficacy and safety of vadadustat to darbepoetin alfa for the correction and maintenance of Hb in subjects with anemia secondary to NDD-CKD. For CI-0015, the primary objectives relate to the conversion and maintenance of Hb.

2.5 Blinding

Both trials are randomized open-label studies. Because of their open-label design, the protocols include careful steps to avoid bias.

An interactive web response (IWR) system governs treatment assignment. The investigators are not aware of which treatment will be assigned next. Because Hb values are objective and will be measured at a central laboratory for all efficacy endpoints, efficacy assessments are not subject to bias.

The studies include blinded adjudication of MACE, the use of an unblinded Independent Data Monitoring Committee (IDMC), and an identical schedule of visits, procedures, and assessments for both treatment groups as strategies to reduce the potential for bias.

To reduce subjectivity of dose adjustment, adjustments to doses for vadadustat and darbepoetin alfa are based on Hb concentration and Dose Adjustment Algorithms. The protocols contain details of the algorithms for vadadustat and darbepoetin alfa.

In order to reduce potential execution bias further, special steps will be taken to restrict access to the study data; the Blinding Procedures and Oversight Plan contains details.

2.6 Sample Size

The planned sample sizes are approximately 925 subjects per arm for both Study CI-0014 and Study CI-0015. The sample sizes have been selected to contribute adequate MACE for a pooled analysis that will take place using the results from both studies. See Sections 2.7 and 2.8 for the operating characteristics of the studies, given these samples sizes, with respect to the stand-alone efficacy analysis to be performed on change in Hb and the pooled MACE analysis.

2.7 Efficacy

For clarity, this SAP uses the word “mean” to refer to averages over the study groups and “average” to refer to the within-person average during specified evaluation periods.

The primary efficacy endpoint in each study is the change in Hb from Baseline to the average Hb over the PEP (Weeks 24 to 36). Vadadustat and darbepoetin alpha will be compared with respect to the mean change in Hb in each study arm. The primary efficacy objective is to show that vadadustat is noninferior to darbepoetin alfa where establishment of noninferiority is based on a margin of -0.75 g/dL applied to the difference in mean change: vadadustat minus darbepoetin alfa. Appendix A provides a rationale for this margin.

For the primary efficacy analysis power, the mean change from Baseline in Hb for vadadustat and darbepoetin alfa is assumed to be identical with a common standard deviation (SD) of 1.5 g/dL. For each trial, noninferiority will be established if the lower limit of the 2-sided 95% confidence interval for the difference between the mean in the vadadustat group and the mean in the darbepoetin alfa group is -0.75 g/dL or higher. Under these assumptions, a sample size of 925 subjects per treatment group in the CI-0014 trial will yield greater than 90% power to show

noninferiority. The power for the CI-0015 trial is also greater than 90% because the sample size is also planned to be about 925 subjects per group.

If the lower limit of the 2-sided 95% confidence interval for the difference between the mean in the vadadustat group and the mean in the darbepoetin alfa group is above zero, superiority will have been established and the finding will be interpreted as providing evidence of a greater change from baseline in Hb for vadadustat relative to the control arm.

2.8 Safety

The primary safety endpoint is the time from the first dose date to the first adjudicated MACE (defined as all-cause mortality, nonfatal myocardial infarction, or nonfatal stroke).

The primary safety analysis will include all first events that accrue over the pair of PRO_2TECT studies. The sample size for the MACE endpoint is based on a 2-sided 95% confidence interval for the hazard ratio (HR) (vadadustat/darbepoetin alfa) and a noninferiority margin of 1.25 (upper bound of the 95% confidence interval) for FDA decision making. Under the assumption that the MACE rate is the same in the 2 groups (i.e., the HR =1), 631 subjects with MACEs will be required in the 2 trials combined to have 80% power to establish noninferiority. If the HR is 0.95 favoring vadadustat, the power will be above 90%. Based on a noninferiority margin of 1.30 for EMA decision making, the power will be above 90% under the assumption that the HR is 1.

Additional safety presentations are described below (Section 7 and Section 10).

3 ANALYSIS POPULATIONS

3.1 The Analysis Populations

Each study has the following four analysis populations:

- Randomized population: all subjects randomized. Analyses of this population will be based on the randomized treatment.
- Full Analysis Set (FAS) population: all subjects in the randomized population who received at least 1 dose of study medication and had at least 1 post dose Hb . Analyses of this population will be based on the randomized treatment.
- Safety population: all subjects in the randomized population who received at least 1 dose of study treatment. Analysis of this population will be based on the actual treatment received. Subjects who received in error some vadadustat and some darbepoetin alfa (excluding rescue therapy) will be classified by the more frequently received drug.
- Per Protocol (PP) population: all randomized subjects who received study medication during the PEP, had at least 1 Hb assessments during the PEP, and had no critical or major protocol deviation affecting the primary endpoint analyses, i.e., prior to Week 36. Analyses of this population will be based on actual treatment received, as described for the Safety population.

Major protocol deviations or causes for site closure leading to exclusion from the PP populations will be specified prior to database lock on a blinded basis.

Efficacy analyses will utilize the randomized, FAS, and PP populations while safety analyses (including analyses of MACE) will utilize the safety population.

3.2 Characterization of the Analysis Population

Number and percent of subjects included in the analysis sets will be summarized.

4 QUALITY OF THE TRIALS

4.1 Study Drug Dosing and Compliance

Subjects track their own drug use at home. At each visit, the site records each subject's self-reported compliance.

The proportion of subjects who report at least 80% compliance at each visit will be tabulated and summarized for both safety population and FAS.

In addition to self-reported compliance, the calculated compliance rate for a given time period will be derived from exposure data as the number of days on dosing period collected on electronic case report forms (eCRFs) divided by the number of days in that time period.

4.2 Protocol Deviations

Definitions of critical, major, and minor protocol deviations will be described in master Protocol Deviation log file, including those protocol deviations leading to exclusion from the PP populations. These will be specified prior to database lock on a blinded basis.

Protocol deviations will be summarized in the randomized population as the number and percentage of subjects with a protocol deviation by treatment group and study period. A by-subject listing of protocol deviations, indicating those exclusionary from the PP population, will be provided.

4.3 Baseline Characteristics, Exposure, and Retention by Treatment Group

Baseline characteristics collected in the case report form (CRF) will be summarized in analysis populations (Section 3). Summaries of demographics, medical history, prior and concomitant medications, and treatment exposure will be displayed by treatment groups.

The total number of randomized subjects by treatment group and the number randomized but not treated will be tabulated. This summary will be performed in the randomized population.

The average weekly dosages by treatment groups will be categorized for the safety and FAS populations(see Section 5.1).

The number of subjects who discontinued treatment and the reason for treatment discontinuation will be tabulatedfor the Randomized populations.

[redacted]

5 GENERAL CONVENTIONS

Study days are defined as follows:

Study Day = [Event date – First dosing date + 1] if after first dosing date

[Event date – First dosing date] if before first dosing date.

Note that with the definition above, days of "0" will not be used. As such, the study day can be interpreted as the number of days before or after first day of dosing. Event date refers to the date associated with the result being summarized. In some cases, this is the date of an assessment or measurement; in other cases, this is the onset date of an adverse or outcome event.

For subjects whose reference date is missing, the study day will also be categorized as missing.

5.1 Visit and Analysis Time Period Classification

The time windows in this section reflect the administrative range around each visit for analysis. They are wider than those stated in the protocol to reduce missing values by capturing as many observation dates as possible. Table 1 provides the convention to classify assessment dates into protocol-defined visits and analysis time periods for assessments at every visit, such as Hb and vital signs. Table 1 will also be used for liver function assessments.

Table 2 presents visit windows for values according to a slightly less frequent assessment schedule, i.e., the iron indices. And, Table 2 will also be used for all other assessments, such as periodic complete blood count (CBC) with differential, chemistry laboratory tests, C-reactive protein (CRP), urine albumin-to-creatinine ratio (uACR) and lipids etc.

Unless otherwise specified, all assessments will be mapped to visit windows on the basis of the date of the assessment relative to the first dose date (i.e., Study Day), regardless of subject disposition or the CRF page completed, e.g., “Unscheduled”, “End of Treatment”, “End of Study”, scheduled visit number. Similarly, all values collected within a reporting period will be considered for identification of safety events of interest.

If more than 1 laboratory-based efficacy outcome value (e.g., Hb, ferritin, hepcidin, TSAT, serum iron, TIBC, and lipids) is available in an analysis time period post-baseline, the average of all observed values (at most 1 per day) will be assigned to that period. For example, the window for the primary efficacy outcome, analysis time period 3, includes any Hb value (at most 1 per day) assessed between Study Days 155 to 266, corresponding to Weeks 24, 28, 32, and 36 (Visits 10 through 13).

Efficacy evaluations will use Hb values as assessed by the central laboratory. (Local HemoCue Hb values are used only for dose adjustments.)

When more than 1 assessment is made for other measures (not part of laboratory-based efficacy) within a given visit window post-baseline, the value of the assessment closest to the end of the window is the value associated with that visit for summaries over time. Most summaries of results by visit (e.g., laboratory or vital sign results for general safety) will use the single value

for each visit determined by this convention. Notable exceptions to this convention are the primary efficacy outcome and clinically significant changes.

If no assessment is available within a time window, then the associated visit and week classifications will be missing.

Table 1. Classification of Assessments at Every Visit for Years 1 through 4

Time Period	Visit Classification	Target Week	Target Day	Actual Study Day of Visit
Baseline	Screening Visit 1[a]	-	-	-
	Screening Visit 2[a]	-	-	-
	Visit 1	Week 0	Day 1	Day 1 (first dose date)[b]
1) Weeks 2-8	Visit 2	Week 2	Day 14	Day 2-21
	Visit 3	Week 4	Day 28	Day 22-35
	Visit 4	Week 6	Day 42	Day 36-49
	Visit 5	Week 8	Day 56	Day 50-63
2) Weeks 10-20	Visit 6	Week 10	Day 70	Day 64-77
	Visit 7	Week 12	Day 84	Day 78-98
	Visit 8	Week 16	Day 112	Day 99-126
	Visit 9	Week 20	Day 140	Day 127-154
3) Primary Efficacy Period (PEP) (Weeks 24-36)	Visit 10	Week 24	Day 168	Day 155-182
	Visit 11	Week 28	Day 196	Day 183-210
	Visit 12	Week 32	Day 224	Day 211-238
	Visit 13	Week 36	Day 252	Day 239-266
4) Secondary Efficacy Period (SEP) (Weeks 40-52)	Visit 14	Week 40	Day 280	Day 267-294
	Visit 15	Week 44	Day 308	Day 295-322
	Visit 16	Week 48	Day 336	Day 323-350
	Visit 17	Week 52	Day 364	Day 351-406
-	Visit 18	Week 64	Day 448	Day 407-490
-	Visit 19	Week 76	Day 532	Day 491-574
-	Visit 20	Week 88	Day 616	Day 575-672
-	Visit 21	Week 104	Day 728	Day 673-770
-	Visit 22	Week 116	Day 812	Day 771-854
-	Visit 23	Week 128	Day 896	Day 855-938
-	Visit 24	Week 140	Day 980	Day 939-1036
-	Visit 25	Week 156	Day 1092	Day 1037-1134
-	Visit 26	Week 168	Day 1176	Day 1135-1218
-	Visit 27	Week 180	Day 1260	Day 1219-1302
-	Visit 28	Week 192	Day 1344	Day 1303-1400
-	Visit 29	Week 208	Day 1456	Day 1401-1498

a. The Screening period, which starts when the informed consent form is signed, will last a maximum of 8 weeks. Two Screening visits (Screening Visit 1 and Screening Visit 2) must be performed within 8 weeks prior to dosing (Baseline visit or Day 1). There must be a minimum of 4 days between the two Screening visits and a minimum of 4 days between Screening Visit 2 or last retest and the Baseline visit.

b. If patient has no first dose date, day 1 will be the randomization date.

Table 2. Classification of assessments at select visits for Years 1 through 4

Time Period	Visit Classification	Target Week	Target Day	Actual Study Day of Visit
Baseline	Screening Visit 1[a]	-	-	-
	Screening Visit 2[a]	-	-	-
	Visit 1	Week 0	Day 1	Day 1 (first dose date)[b]
1) Weeks 2-8	Visit 3	Week 4	Day 28	Day 2-42
	Visit 5	Week 8	Day 56	Day 43-63
2) Weeks 10-20	Visit 7	Week 12	Day 84	Day 64-112
	Visit 9	Week 20	Day 140	Day 113-154
3) Primary efficacy period (PEP) (Weeks 24-36)	Visit 11	Week 28	Day 196	Day 155-224
	Visit 13	Week 36	Day 252	Day 225-266
4) Secondary efficacy period (SEP) (Weeks 40-52)	Visit 15	Week 44	Day 308	Day 267-336
	Visit 17	Week 52	Day 364	Day 337-406
-	Visit 18	Week 64	Day 448	Day 407-490
-	Visit 19	Week 76	Day 532	Day 491-574
-	Visit 20	Week 88	Day 616	Day 575-672
-	Visit 21	Week 104	Day 728	Day 673-770
-	Visit 22	Week 116	Day 812	Day 771-854
-	Visit 23	Week 128	Day 896	Day 855-938
-	Visit 24	Week 140	Day 980	Day 939-1036
-	Visit 25	Week 156	Day 1092	Day 1037-1134
-	Visit 26	Week 168	Day 1176	Day 1135-1218
-	Visit 27	Week 180	Day 1260	Day 1219-1302
-	Visit 28	Week 192	Day 1344	Day 1303-1400
-	Visit 29	Week 208	Day 1456	Day 1401-1498

a. The Screening period, which starts when the informed consent form is signed, will last a maximum of 8 weeks. Two Screening visits (Screening Visit 1 and Screening Visit 2) must be performed within 8 weeks prior to dosing (Baseline visit or Day 1). There must be a minimum of 4 days between the two Screening visits and a minimum of 4 days between Screening Visit 2 or last retest and the Baseline visit.

b. If patient has no first dose date, day 1 will be the randomization date.

5.2 Definition of Baseline

In general, the baseline value will also be the value of assessment closest to the end of the window, i.e., Day 1 (first dose date). The baseline value must be assessed prior to initiation of study treatment. In contrast, the baseline value for laboratory-based efficacy outcomes (e.g., Hb, ferritin, hepcidin, TSAT, serum iron, TIBC, and lipids) will be an average of the last 1 or 2 values prior to or on the first dose date, as specified in Section 9.1.1.

5.3 Prior and Concomitant Medication

Prior and concomitant medications will be coded using World Health Organization Drug Dictionary (WHO DD; Mar 2017 or latest version) and summarized for each treatment group based on the safety population. Prior medication is defined as any medication taken prior to the

first dose of the study medication in the initial treatment period. Any medication taken on or after the day of first dose of the study treatment will be considered as concomitant medication for the treatment analysis.

5.4 Handling of Missing Data

Missing data will be handled using a procedure specific to each variable and particular analysis as described in the sections relevant to each endpoint. If no method for missing data is discussed, descriptive analyses will be based upon observed data without imputation. Appendix B describes the general approaches to be used for missing data.

For the analysis of safety variables, only partial dates will be imputed unless otherwise specified. The algorithms for imputation of partial dates depend upon the parameter, as follows.

Adverse event onset:

- If onset date is completely missing, date is set to date of first dose.
- If year is present and month and day are missing or year and day are present and month is missing:
  - If year = year of first dose, then set month and day to month and day of first dose.
  - If year < year of first dose, then set month and day to December 31.
  - If year > year of first dose, then set month and day to January 1.
- If month and year are present and day is missing:
  - If year = year of first dose and
    - month = month of first dose, then set day to day of first dose date.
    - month < month of first dose, then set day to last day of month.
    - month > month of first dose, then set day to first day of month.
  - If year < year of first dose, then set day to last day of month.
  - If year > year of first dose, then set day to first day of month.
- For all other cases, set date to date of first dose.

Adverse event end date:

- If year is present and month and day are missing or year and day are present and month is missing, set end month and day to December 31.
- If month and year are present and day is missing, set the day to last day of the month.
- If fatal event, date is set to minimum of imputed end date and death date.
- For all other cases, set date to missing.

Concomitant medication:

- If start date is completely missing, start date will not be imputed.
- If start year is present and month and day are missing or year and day are present and month is missing, set start month and day to January 1.
- If start year and month are present and day is missing, set start day to 1st day of month.

- If end date is completely missing, end date will not be imputed.
- If end year is present and month and day are missing or year and day are present and month is missing, set end month and day to December 31.
- If end year and month are present and day is missing, set end day to last day of the month.
- The imputed dates must be logical, ensuring that no end date is after database lock or death or before the start date.

If site queries fail to resolve partial dates for laboratory values and vital signs, including for efficacy, the date is missing and will not be imputed.

6 EFFICACY ENDPOINTS

This section lists the efficacy endpoints; Section 9 describes their definitions and the plans for the analysis of each endpoint.

6.1 Primary Efficacy Endpoint

The primary efficacy endpoint is change in average Hb between Baseline and the PEP (Weeks 24 to 36).

6.2 Key Secondary Efficacy Endpoints

Change in average Hb value between Baseline and the SEP (Weeks 40 to 52).

6.3 Other Efficacy Endpoints

This section describes the other efficacy endpoints of the studies.

6.3.1 Definition of Rescue Episode

For summaries of potential rescue therapies (i.e., red blood cell [RBC] transfusion, and ESA), exposure to the therapy will be grouped temporally into episodes, which could contain multiple administrations based on the gap in time between the end of 1 and the start of the next. For ESA medication , the longest such gap within a single episode is 30 days, while the maximum gap is 7 days for RBC transfusion.

6.3.2 Endpoints Related to Hb

The following efficacy endpoints are related to Hb:

- Change in average Hb value between Baseline and the combined PEP and SEP (Weeks 24 to 52)
- Having average Hb value in the geography-specific target range in Weeks 24 to 36 (yes/no variable)
- Having average Hb value in the geography-specific target range in Weeks 40 to 52 (yes/no variable)
- Having at least 1 Hb value in the geography-specific target range in Weeks 24 to 36 (yes/no variable)

- Having at least 1 Hb value in the geography-specific target range in Weeks 40 to 52 (yes/no variable)
- Having Hb values in the geography-specific target range for at least 1/2 of the observations in Weeks 24 to 36 (yes/no variable)
- Having Hb values in the geography-specific target range for at least 1/2 of the observations in Weeks 40 to 52 (yes/no variable)

These endpoints apply only to study CI-0014:

- Hb increase of >1.0 g/dL from Baseline to Week 52 (yes/no variable)
- Time to achieve Hb increase of >1.0 g/dL from Baseline Hb (censored at Week 52)

6.3.3 Endpoints Related to Iron

These sections address iron-related endpoints from randomization through the end of the trials. The study population will be divided into the following baseline iron groups:

- 0 – those not receiving any iron at baseline,
- I – those receiving only oral iron at baseline,
- II – those receiving only intravenous (IV) iron at baseline, and
- III – those receiving IV and oral iron at baseline.

6.3.3.1 Hepcidin, Ferritin, Total Iron-Binding Capacity, Serum Iron, and Transferrin Saturation

For hepcidin, ferritin, total iron-binding capacity (TIBC), serum iron, and transferrin saturation (TSAT), the 2 treatment groups will be compared with respect to mean change from baseline and percentage change from baseline in the PEP and SEP. Subgroup analyses and variation with rescue and Hb may be considered.

6.3.3.2 Elemental Iron

For each baseline iron group and the whole population, a table will present by treatment group the proportion of subjects who received each of the relevant routes and the mean weekly dose of elemental iron. The calculation of mean weekly dose will include subjects with iron administration and will count only weeks in which a subject was still being followed in the denominator.

6.3.4 Other Laboratory Parameters

The following efficacy endpoints are related to laboratory chemistry:

- Change in serum glucose between Baseline and the PEP (Weeks 24 to 36).
- Change in lipid parameters between Baseline and the PEP (Weeks 24 to 36) – including total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides.

6.3.5 Dose Adjustments and Interruptions

An important aspect of evaluating the efficacy of vadadustat arm is to understand how often dose adjustments are needed to maintain target Hb levels. To that end, the pattern of dose changes in the 2 treatment arms will be described, including percent with dose adjustment or treatment interruption for the whole study and by the four analysis time periods in Year 1 (see Section 5.1).

6.3.6 Progression of CKD

Progression of chronic kidney disease (CKD) in both studies is defined as experiencing any of the following:

- Transition to chronic dialysis, or
- Receipt of a kidney transplant, or
- Estimated glomerular filtration rate (eGFR) <15 mL/min per 1.73 m^2 and confirmed by another measurement with a reduction of eGFR <15 mL/min per 1.73 m^2, which should be at least 28 days apart from the first reduction, or
- Reduction in eGFR of 40% or more from baseline (confirmed by second measurement at least 28 days later).
- The endpoint for analysis is time to progression of CKD. See Coresh, 2019; Heerspink, 2019; Levey, 2014 for a description of the endpoint.

6.3.7 Endpoints Related to RBC Transfusion

These endpoints address RBC transfusion for the entire length of study, by the four analysis time periods in Year 1 (see Section 5.1), and in the long-term treatment/follow-up period (Week 52 to EOS) unless otherwise specified.

- Receipt of any RBC transfusion
- Time to first RBC transfusion (for entire study)
- Total number of RBC transfusion episodes received
- Rate of RBC transfusions, calculated as the number of episodes divided by the duration of at-risk follow-up in person-years

RBC transfusion may be considered rescue when “Worsening Anaemia due to CKD” is the reason indicated on the CRF.

6.3.8 Endpoints Related to ESA Usage

The study will consider the following efficacy endpoints specific to ESA usage for the entire length of study, by the four analysis time periods in Year 1 (see Section 5.1), and in the long-term treatment/follow-up period (Week 52 to EOS) unless otherwise specified.

- Receipt of any ESA medication (in the darbepoetin alfa arm, use only includes ESA other than darbepoetin alfa as well as increases in darbepoetin alfa the investigator specifically designates as rescue)
- Time to first ESA medication (for entire study)
- Total number and maximum duration of ESA episodes.

ESA medication may be considered rescue when any of the following reasons are indicated on the relevant version of the CRF:

- [Original CRF] "the subject experienced a clinically significant worsening of their anaemia or symptoms of anaemia (e.g., fatigue, weakness, shortness of breath, chest pain, confusion, or dizziness) compared with Baseline"
- [Original CRF] "the subject's HGB is <9.0 g/dL"
- [CRF updated mid-2018] "Worsening of symptoms of anemia and Hb <9.0 g/dL"
- [CRF updated mid-2018] "Investigator discretion not meeting protocol-defined rescue criteria"
- [CRF updated mid-2018] "Other Specify" which include either one of the 2 reasons from the original CRF

Analyzing the impact of pre-baseline or post-baseline ESA on vadadustat efficacy requires conversion of ESA dose into common units. The epoetin alfa analogues, darbepoetin alfa and methoxy polyethylene glycol epoetin beta (Mircera), will be converted to IV epoetin equivalent units per kilogram per week (U/kg/week).

The following conversions have been derived from published literature and input from clinical experts [Paganini, 1995; Kaufman, 1998; Cremieux, 2006; Gosselin, 2006; Levin, 2008; FDA 2011; Jordan, 2012; Choi, 2013; Vega, 2014; Wright, 2015].

- Subcutaneous (SC) epoetin to IV epoetin: 1:1.25

	SC Epoetin	IV Epoetin
Darbepoetin alfa to epoetin	1:160	1:200
Methoxy polyethylene glycol-epoetin beta to epoetin	1:176	1:220

6.3.9 Rescue Therapy: Series of Definitions

Rescue will be defined from narrow to broad, based on the type, timing, intensity, and reason for treatment. These treatments may be considered rescue therapy for anemia secondary to chronic kidney disease: ESA medication, and RBC transfusion.

The possible reasons for such treatments include:

- Investigator-ordered rescue (per protocol for worsening of anemia),
- Adverse events (unrelated to anemia),
- Maintenance during prolonged interruption of study treatment, and
- Inadvertent use (at hospital or dialysis center).

If a given therapy starts the same day as permanent discontinuation of study drug, then that therapy will be considered on treatment for rescue definitions.

For the vadadustat arm, any use of ESA may be considered rescue. For the control arm, the study drug itself is an ESA and so is not necessarily rescue, yet change to another ESA may be considered rescue. Because darbepoetin alfa is titratable, increase in dose is not considered rescue unless the investigator specifically designates it as such; however, increases in dose relative to last dose will be characterized in three categories by percentage increase: <50%,

≥50% and <100%, and ≥100%. Subjects will be presented by their maximum category of increase within each of the four analysis time periods in Year 1 (see Section 5.1). All ESA use will be converted to IV epoetin alfa equivalents.

The CRFs collect reasons for ESA medication and RBC transfusion. The series of rescue definitions are defined as follows:

- Narrow: Rescue for worsening anemia with ESA medication or RBC transfusion defined in Section 6.3.7 and Section 6.3.8, not starting after permanent study treatment discontinuation.
- Broad-on-treatment: Any exposure to ESA medication (aside from darbepoetin alfa not designated as rescue in the control arm) or RBC transfusion for any reason not starting after permanent study treatment discontinuation.

In addition to the sensitivity analyses (Section 9.1.4), the report will include a summary of the incidence, timing, duration, intensity, and frequency for each definition of rescue, including a Kaplan-Meier plot of time to first rescue.

7 NON-MACE SAFETY ENDPOINTS

While the primary safety endpoints relate to MACE, defined as all-cause mortality, nonfatal myocardial infarction, or nonfatal stroke, this SAP addresses only the non-MACE endpoints. As stated above, a separate SAP will describe the analyses of MACE and the individual components of the MACE endpoint as well as expanded MACE definitions where expanded MACE will include hospitalization for heart failure and thromboembolic events.

The safety reporting period for this study begins upon randomization and ends at the final protocol-required study contact. A treatment-emergent adverse event (TEAE) is one that begins or worsens after treatment initiation. The AE CRF has been designed to capture all events after randomization. AEs that occur during the screening period are not reported on the CRF. The treatment groups will be compared with respect to the following non-MACE safety endpoints:

- Adverse events (AEs) and serious adverse events (SAEs)
- Treatment-emergent adverse events (TEAEs) and treatment-emergent SAEs
- Vital signs and clinical laboratory values (see Appendix E)
- Adverse events of special interest (AESI) (see Appendix F)
- Any value of Hb >12.0 g/dL; >13.0 g/dL, >14.0 g/dL, <9.0 g/dl or <8.0 g/dl
- An Hb increase >1.0 g/dL within any 2-week interval or >2.0 g/dL within any 4-week interval.

8 PHARMACOKINETIC/PHARMACODYNAMIC (PK/PD) ANALYSIS

Serum samples will be taken and stored for use in biomarker studies. PK sample will be collected between 15 minutes to 1 hour after vadadustat administration at baseline visit, and will also be performed along with other study laboratory samples being collected at the Weeks 4, 12, 28, and 52. For each visit, one serum sample is collected for each subject. Discriptive anlaysis of PK concentration will be provided by treatment groups. Analysis for EPO is listed in Section E.1.7

9 EFFICACY ENDPOINTS AND ANALYSES

This section briefly describes the methods to be used for each of the efficacy endpoints. Rather than providing details for each endpoint, the accompanying Appendices describe the approaches to be used for specific types of data. For all the models and methods described below, see Appendix B for handling of any missing data.

The general approach to analysis of continuous outcomes will be analysis of covariance (ANCOVA) with multiple imputation for missing data or mixed models for repeated measurements (MMRM) on observed data. For a description of the models to be used, see Appendix C.

For binary variables, the general approach will be Mantel-Haenszel estimation of risk difference stratified (Mantel, 1954) by the baseline strata with (or without) multiple imputation to deal with missing data. See Appendix D.

Some endpoints come from count data (e.g., the number of RBC transfusions) that are likely to have a spike at zero and a long tail. Such data are often complicated to analyze because of their highly skewed distributions. While the samples sizes are so large that the Central Limit Theorem ensures that the means of these distributions are essentially normal, some interpretation of the nature of the distributions may be useful clinically. To the extent reasonable, we have described the planned methods for such data, but ad hoc methods may be necessary if the skewness is extreme. See Appendix D for the general approach.

For time-to-event variables, the general approach will be to use Cox models stratified by the baseline randomization strata. See Appendix D.

9.1 Definition of Primary Efficacy Endpoint: Change from Baseline in Hb

The primary efficacy endpoint is the change in average Hb between baseline and the PEP (Weeks 24 to 36).

The primary analysis will use the randomized population. As described in Section 2.7, establishment of noninferiority will be based on a margin of -0.75 g/dL applied to the difference in mean change: vadadustat minus darbepoetin alfa.

9.1.1 Baseline Hb

Baseline Hb will be calculated as the average of the last 2 central laboratory Hb measurements of samples taken at the visits prior to or on the date of first dose date.

Subjects may be rescreened up to three times. A minimum of 1 Hb value is required for the calculation of baseline.

9.1.2 Hb in the Primary Efficacy Period

Hb for the PEP will be calculated as the average of all Hb measurements from the central laboratory within the four visit windows during Weeks 24 through 36, regardless of intercurrent events. At least 1 Hb measurement is required for the calculation; otherwise, the value will be missing and therefore imputed. The PEP corresponds to Visits 10 through 13, also called analysis time period 3 (Table 1).

9.1.3 Change in Hb From Baseline to the Primary Efficacy Period

The change from baseline will be calculated for each subject as the PEP value minus the baseline value.

9.1.4 Analysis of Change From Baseline in Hb to the Primary Efficacy Period

The primary analysis will be performed in the randomized population, as defined in Section 3.1, to address an estimand of treatment policy in accordance with the Intention to Treat philosophy. In this open-label study, the research question is to compare the assignment of a standard of care, darbepoetin, to the investigational new drug, vadadustat, regardless of intercurrent events. This analysis includes individuals who never receive treatment, discontinue treatment, are rescued with any therapy, or withdraw consent.

The primary analysis will use multiple imputation with ANCOVA as the substantive model. Missing primary endpoint data will be imputed with the group to which the subject was randomized as described in Appendix B. An ANCOVA model, as described in detail in Appendix C, will be used to compare the mean change from baseline in Hb between the 2 groups.

The primary analysis model will contain treatment group, baseline Hb level, and the 2 stratification factors (region and NYHA CHF class) as predictor variables. The stratification factor of entry Hb level will not be included in the model because of the inclusion of baseline Hb. The stratification factor assignments at randomization will be respected in the analysis.

The random seeds for all the multiple imputation runs for each trial will be generated from a single master seed. The master seeds will be 10014 for CI-0014, and 10015 for CI-0015. The generation code, resulting random seeds, and corresponding analysis assignments are listed in Appendix B.

The data will have shown noninferiority of vadadustat if the lower bound of the 95% confidence interval for the difference in estimated change from baseline in the 2 groups (vadadustat minus darbepoetin) exceeds the noninferiority margin of -0.75. This ensures a 1-sided alpha of 0.025 for the primary analysis. If the lower limit of the 2-sided 95% confidence interval for the difference between the mean in the vadadustat group and the mean in the darbepoetin alfa group is above zero, superiority will have been established and the finding will be interpreted as providing evidence of a greater change from baseline in Hb for vadadustat relative to the control arm.

9.1.5 Sensitivity Analyses to the Primary Efficacy Results

To assess the robustness of the findings from the primary analysis, the following set of sensitivity analyses will be performed using a method analogous to that of the primary analysis (including multiple imputation) unless otherwise specified:

- The primary analysis will be repeated after setting to missing all per-visit Hb values within four weeks of administration rescue therapy or after EOT visits. The definition of rescue therapy will vary across the series described in Section 6.3.9.
- Tipping point analyses (see Appendix Section B.3.2) will be performed to assess the effect of the missing data.

- The FAS population will be used instead of the randomized population without a new multiple imputation. This analysis will be performed only if the size of the FAS is less than 95% of the randomized population.
- The PP population with the actual treatment received without imputation
- A mixed model for repeated measures (MMRM) will be fit to the observed data only *without imputing missing values*. The repeated measures will be the observed averages for analysis time periods 1, 2, and 3 (See definitions in Section 5.1 and analysis details in Appendix C).

9.2 Secondary Efficacy Endpoints: Definitions and Analyses

The study has one secondary efficacy endpoint which will be analyzed formally only if the primary analysis meets the noninferiority margin.

- Change in average Hb value between Baseline and the SEP (Weeks 40 to 52)

Evaluation of the average change in Hb will employ the approach described for the primary endpoint (see Section 2.7 and Section 9.1) assessing Weeks 40 to 52 instead of Weeks 24 to 36. The power for this endpoint for a noninferiority margin of -0.75 g/dL is expected to be close to the power of the primary endpoint.

Sensitivity analyses analogous to those performed for the primary efficacy endpoint will be repeated to assess mean change in Hb from Baseline to Weeks 40 to 52. For ANCOVA models, analysis time period 4 will replace 3, and for MMRM, the models will be extended to include four post-baseline timepoints.

9.3 Other Efficacy Endpoints: Strategies for Analysis

The study has many other efficacy endpoints. Because of their exploratory nature, they will not be corrected for multiplicity.

9.3.1 Having Average Hb Value in the Geography-Specific Target Range in Weeks 24 to 36

All subjects will be defined as either being in their geography-specific target range (see Section 2.3) in Weeks 24 through 36 (“yes”) or not (“no”), based on the average Hb value during the four visit windows in Weeks 24 through 36. Subjects with no Hb value in the PEP value will be treated as missing and handled according to Appendix B.

The proportion of subjects within target range will be calculated and tested with a Mantel-Haenszel test stratified by the baseline strata. See Appendix B and Appendix D.

Noninferiority will be evaluated using a 2-sided confidence interval for the difference in proportions. Noninferiority will have been established if the lower limit of the confidence interval is above -15%. The power for this endpoint is expected to be roughly 90%.

If the lower limit of the 2-sided confidence interval for the difference in proportions is above zero, superiority will have been established and the finding will be interpreted as providing evidence of a higher proportion of subjects being within the target range for vadadustat relative to the control arm.

9.3.2 Having Average Hb Value in the Geography-specific Target Range in Weeks 40 to 52

All subjects will be defined as either being in their geography-specific target range (see Section 2.3) in Weeks 40 to 52 ("yes") or not ("no"), based on the average Hb value during the Weeks 40 to 52 visit windows. Subjects with no Hb value in the SEP value will be treated as missing and handled according to Appendix B.

The proportion of subjects within target range will be calculated and tested with a Mantel-Haenszel test stratified by the baseline strata. See Appendix D.

Noninferiority will be evaluated using a 2-sided confidence interval for the difference in proportions. Noninferiority will have been established if the lower limit of the confidence interval is above -15%. The power for this endpoint is expected to be roughly 90%.

If the lower limit of the 2-sided confidence interval for the difference in proportions is above zero, superiority will have been established and the finding will be interpreted as providing evidence of a higher proportion of subjects being within the target range for vadadustat relative to the control arm.

Table 3 lists the other efficacy endpoints. All analyses will stratify by the randomization (baseline) strata.

Table 3. Other Efficacy Endpoints

Endpoint	Approach to Analysis	Comments
Hb endpoints – Section 6.3.2		
Change in average Hb value between Baseline and the combined primary efficacy period (PEP) and secondary efficacy periods (SEP) (Weeks 24-52)	Analysis of covariance (ANCOVA) stratified by baseline strata.	Average may include imputed values for PEP and/or SEP.
Having Hb values in the geography-specific target range during PEP or SEP (2 alternate definitions)	Mantel-Haenszel test stratified by the baseline strata	This analysis includes the observed hemoglobin values.
CI-0014 only: Hb increase of >1.0 g/dL from Baseline to any time before Week 52	Mantel-Haenszel test stratified by baseline strata	This analysis includes the observed hemoglobin values.
CI-0014 only: Time to achieve Hb increase of >1.0 g/dL from Baseline Hb (censored at Week 52)	Time – to –event	This analysis will use only observed data.
Progression of CKD (Section 6.3.6)		
Progression of CKD (entire length of study)	Time – to – event	
Iron endpoints (Section 6.3.3)		
Receipt of at least 1 administration of elemental iron (IV, or oral)	Mantel-Haenszel test stratified by baseline strata	
Average weekly dose of elemental iron, IV, or oral	ANCOVA stratified by baseline strata.	If the distribution is highly skewed either because of a large spike at zero or a long tail to the right (or both), other methods may be considered.

Endpoint	Approach to Analysis	Comments
Hepcidin, Ferritin, TIBC	Four outcomes – change from baseline to PEP, change from baseline to SEP, percentage change from baseline to PEP, and percentage change from baseline to SEP. Analysis of covariance stratified by baseline strata for change from baseline to PEP or SEP.	This analysis will use only observed data
Serum Iron		
TSAT		
Other laboratory endpoints (Section 6.3.4)		
Change from baseline in serum glucose and lipid parameters at PEP and SEP	ANCOVA stratified by baseline strata.	No imputation.
RBC transfusion endpoints (Section 6.3.7)		
Receipt of any RBC transfusion	Mantel-Haenszel test stratified by the baseline strata	
Total number of RBC transfusions received (entire length of study)	See Appendix D .	If the distribution is highly skewed either because of a large spike at zero or a long tail to the right (or both), other methods may be considered.
Dose adjustment (Section 6.3.5)		
Dose adjustments (entire length of study)	Descriptive statistics	Other methods may be considered.
ESA and Rescue endpoints (Section 6.3.8 and Section 6.3.9)		
Receipt of any ESA rescue medication	Mantel-Haenszel test stratified by the baseline strata	
Receipt of any rescue therapy (series of definitions from narrow to broad)	Mantel-Haenszel test stratified by the baseline strata	

10 SAFETY ANALYSES

Unless otherwise specified, safety analyses will be performed using the safety population.

Most of the analysis of safety data will be descriptive without formal statistical testing. In some cases, 95% confidence intervals for the change from baseline as well as for the difference between the study groups may be reported.

Formal statistical methodology will be used for the MACE data; the methods are described in a separate SAP.

10.1 Mortality

The number and cause of death will be reported and summarized by treatment arm.

10.2 Hb-related Safety Endpoints

Hb-related safety endpoints will be defined using data from the central laboratory. The analyses will use all central laboratory values in the database starting from randomization. No imputation will be performed for missing data. For each endpoint in Table 4, the proportion of the study group who satisfy the definition will be calculated stratified by the baseline strata. The 95% confidence intervals for these proportions will be reported along with the difference in proportions as well as the odds ratio; a Mantel-Haenszel weighting method will be used to calculate the 95% confidence intervals for these statistics (see Appendix D).

Table 4. Hb-related Safety Endpoints from First Dose Date to Last Visit

Variable	Criterion for "yes"
Hb >12.0 g/dL	Any Hb >12.0 g/dL after Day 1
Hb >13.0 g/dL	Any Hb >13.0 g/dL after Day 1
Hb >14.0 g/dL	Any Hb >14.0 g/dL after Day 1
Hb increase >1.0 g/dL within any 2-week interval	Difference between 2 Hb values within any 2 weeks >1.0 g/dL after Day 1
Hb increase >2.0 g/dL within any 4-week interval	Difference between 2 Hb values within any 4 weeks >2.0 g/dL after Day 1
Hb<9.0 g/dL	Any Hb <9.0 g/dL after Day 1
Hb<8.0 g/dL	Any Hb <8.0 g/dL after Day 1

Hb: hemoglobin

10.3 Adverse Events and Serious Adverse Events

The AE CRF has been designed to capture all events after randomization. A TEAE is one that begins or worsens after treatment initiation. The AE reporting period for this study begins upon randomization and ends at the final protocol-required study contact. In addition, any AE that occurs after the AE reporting period is to be reported as an AE if the Investigator assesses it as possibly or probably related to the study medication.

Serious adverse events (SAEs) will be handled in a fashion analogous to the methods described for AEs.

Each AE will be summarized by Medical Dictionary for Regulatory Activities (MedDRA) system organ class (SOC) and preferred term (PT). All AE summaries will provide the number of subjects reporting at least 1 AE. Tables will present the number and percentage of subjects by treatment group reporting at least 1 of the following:

- Treatment-emergent AE (TEAE), SOC, and PT
  - TEAE leading to withdrawal of study medication, SOC, and PT
  - TEAE, severity, SOC and PT
  - Drug-related TEAE, SOC, and PT
  - Drug-related TEAE leading to withdrawal of study medication, SOC, and PT

- Treatment-emergent SAE, SOC, and PT
  - Drug-related treatment-emergent SAE, SOC, and PT
  - Fatal TEAE, SOC and PT

A summary table will show number of subjects with at least 1 of each of the following:

- TEAE
- Severe TEAE
- TEAE leading to withdrawal of study medication
- Drug-related TEAE leading to withdrawal of study medication
- Drug-related TEAE
- Treatment-emergent SAE
- Drug-related treatment-emergent SAE
- Fatal TEAE

A summary by treatment group and PT will be presented of TEAEs reported by at least 5% of subjects in either treatment group.

AE summaries will be ordered by decreasing pooled percentage for SOC, and PT within SOC.

TEAEs will be summarized by 1) worst severity, and 2) worst causality by SOC and PT. For each subject and each PT, the worst severity recorded will be used in the by-severity summaries. Similarly, the worst causality (most related to treatment) will be attributed and used in the by-causality summaries. If severity or causality is missing, data will be imputed to the worst category.

A by-subject listing of all TEAEs will be provided. This listing will be presented by treatment group and will include: center, subject identifier, age, sex, race, AE (SOC, PT, and verbatim term), study day of onset, study day of resolution, duration, severity, seriousness, relationship to the study medication, action taken, outcome and causality.

AESI, listed in Appendix F, will be summarized by treatment group. The number and percentage of subjects reporting these AESI will be provided.

10.4 Other Safety Endpoints

Vital signs and clinical laboratory values will be presented as outlined in Appendix E.

The following additional safety endpoints will be presented:

- The number of usage of hypertensive medications added after first dose by study period/visit.

The separate MACE SAP will describe the methodology for the adjudicated analyses.

11 SUBGROUPS

Descriptive summary statistics for the primary and secondary efficacy endpoints will be presented separately in the following subgroups of the Randomized Population. The methods for analysis will be those described for the respective primary and secondary endpoints. When

analyzing a subgroup that is itself a stratification variable, it will be removed from the analysis model. These studies will enroll subjects from many centers; however, most centers will have small sample sizes. Therefore, no center-specific analyses are planned. In addition, stratification variables may be removed from the models in case of small sample size for a subgroup. A table will summarize the distribution of subjects by each subgroup and treatment group.

Randomization stratification factors

- Hb stratification level at baseline
  - CI-0014
    - <9.5 g/dL
    - ≥9.5 g/dL
  - CI-0015
    - <10.0 g/dL
    - ≥10.0 g/dL
- Region (actual according to CRF)
  - US
  - EU
  - ROW
- NYHA CHF stratification level
  - 0 and 1
  - 2 and 3
- Target Hb level: These targets are completely confounded with region. The 10 to 11 g/dL target consists of subjects from the US while the 10 to 12 g/dL target applies to the EU and the ROW.
  - 10 to 11 g/dL
  - 10 to 12 g/dL

Demographics and medical history

- Age
  - <65 years
  - ≥65 years
- Sex
  - Male
  - Female
- Race
  - White
  - All others
- Diabetes mellitus
  - No diabetes mellitus
  - Diabetes mellitus

- Hypertension
  - No Hypertension
  - Hypertension
  - Medications (CI-0015 only)
- Baseline ESA dose (See Locatelli, 2004 for justification of the choice of cut-off values)
  - ≤90 U/kg/week
  - >90 and <300 U/kg/week
  - ≥300 U/kg/week

Baseline laboratory measurements

- Urine albumin-to-creatinine ratio (uACR) – reference for cutoff: Gansevoort, 2011.
  - <300 g/kg
  - ≥300 g/kg
- Estimated GFR – The threshold of 15 represents a value between CKD eGFR stages 4 and 5
  - <15 mL/min/1.73m^2
  - ≥ 15 mL/min/1.73m^2
- C-reactive protein. In Q2 laboratories, which are the labs used in these studies, the normal range is 0 to 0.6 mg/dL.
  - ≤0.6 mg/dL
  - >0.6 mg/dL
- Baseline TSAT
  - < median of Baseline TSAT (%)
  - ≥ median of Baseline TSAT (%)
- Baseline ferritin
  - < median of baseline ferritin (ng/mL)
  - ≥ median of baseline ferritin (ng/mL)

12 OTHER EXPLORATORY ANALYSES

In addition to the planned analyses described above, the study team will likely perform many other descriptive analyses of efficacy and safety. For example, the team is likely to provide descriptions of the trajectory of subjects who transition to chronic dialysis, which will include the figures for Hb and other labs after the date of initiation of dialysis.

13 DATABASE LOCK AND UNBLINDING

13.1 Blind Review of Selected Data Prior to Final Database Lock

Upon database release, protocol deviation and analysis population outputs will be produced and will be sent to Sponsor for review. An analysis population classification meeting will be arranged to discuss the outputs and to decide which subjects and/or subject data will be excluded from certain analyses. Decisions made regarding the exclusion of subjects and/or subject data

from analyses will be made prior to unblinding and will be documented and approved by Sponsor.

The following summary of analysis population will be provided:

- *A summary of the number and percentage of subjects allocated to each analysis population by blinded treatment group and overall (Analysis population: All Subjects Randomized)*

A by-subject listing of analysis population details will be provided. This listing will be presented by blinded treatment group and include: site, subject identifier, inclusion/exclusion flag for each population and reason for exclusion from each population. All subjects randomized will appear on this listing. Other data may also be selected for blind review if deemed necessary.

13.2 Final Database Lock

After completion of all Blind Data Review, validation of the project databases, and Akebia's approval of the review, the clinical database will be locked. After the database lock and the authorization for unblinding (Section 13.3), the treatment codes will be merged to the analysis datasets. Any change to the clinical database after this time will require written authorization, with explanation, by Akebia. In addition, beginning at the time of database lock, an audit trail will be maintained of all versions of the analysis datasets that may result from refinements of the algorithms for derived variables in the course of the analysis.

13.3 Authorization for Unblinding

After database lock and upon receipt of written authorization from Akebia, a blinded study team will receive the actual treatment codes directly from the group maintaining them for data analysis.

APPENDIX A. JUSTIFICATION OF THE NONINFERIORITY MARGIN

A.1. Historical Trials for Indication

Several drugs in the erythropoiesis stimulating agent (ESA) class have been studied and approved for increasing hemoglobin (Hb) levels in subjects with anemia related to chronic kidney disease (CKD). They are, in order of FDA approval:

- erythropoietin alfa (EpoA from Amgen, 1986)
- darbepoetin alfa (DA from Amgen, 1998)
- Mircera (Roche, 2004)
- Omontys (Affymax, 2007).

We have combined the results of the pivotal trials leading to the approval of these agents with the results from several other studies and performed a meta-analysis of 16 study arms to estimate the placebo-adjusted change from baseline in Hb after 24 to 36 weeks of correction treatment. The growing precedence for this endpoint plus the summary treatment effect estimated here support the noninferiority margin of -0.75 g/dL used for the primary analyses in the PRO_2TECT (CI-0014, CI-0015) and $INNO_2$VATE (CI-0016, CI-0017) trials.

A.2. Characteristics of Study Arms

The 16 study arms included (see Table A 1) vary in several key characteristics, namely: dialysis dependence (DD) or non-dialysis-dependent (NDD); the maximum allowable Hb level for inclusion of participants; the average baseline Hb level; the therapeutic drug; the timing of the primary efficacy outcome; and the target range of Hb. The timing for the PRO_2TECT and $INNO_2$VATE studies is planned for 24 to 36 weeks for the primary endpoint, and the target range of Hb is 10 to 11 g/dL in the United States (US) and 10 to 12 g/dL ex-US.

The Mircera and Omontys trials used noninferiority margins of -0.75 and -1.0 ng/dL, respectively.

Table A 1. Study Characteristics

Study Label	Drug Label[a]	Number Treated	Target Hb Level[b]	Average Baseline Hb (g/dL)	Dialaysis Dependence	Reference to Literature[c]
TREAT	DA	2012	High	10.4	No	4
Affymax	OM1	328	High	10.0	No	5
Affymax	OM2	328	High	10.0	No	5
Affymax	DA	327	High	10.0	No	5
Roche	DA	162	High	10.2	No	3
Roche	MI	162	High	10.2	No	3
Akizawa	DA	161	High	9.2	No	6
Akizawa	Epo	160	Low	9.2	No	6
Amgen202	DA	129	High	9.3	No	1,2
Amgen202	Epo	37	High	9.8	No	1,2
AmgenDial	Epo1	201	High	7.4	Yes	7
Amgen211	DA	91	High	8.8	Yes	1,2
AmgenCAN	Epo1	44	Low	6.9	Yes	8
AmgenCAN	Epo2	38	High	7.1	Yes	8
AmgenDial	Epo2	35	High	8.2	Yes	7
Amgen211	Epo	31	High	8.5	Yes	1,2

CAN: Canada; DA: darbepoetin alfa; Dial: dialysis; Epo: epoetin alfa; MI: Mircera; OM: Omontys.

a. Suffixes indicate various dosing regimens.

b. The high target Hb levels range from 11 to 13.5 g/dL, and the low levels range from 9 to 11 g/dL.

c. Reference key follows.

1. Macdougall IC. Darbepoetin alfa: a new therapeutic agent for renal anemia. Kidney Int Suppl. 2002;61(80):55-61. doi:10.1046/j.1523-1755.61.s80.11.x.
2. Locatelli F, Olivares J, Walker R, et al. Novel erythropoiesis stimulating protein for treatment of anemia in chronic renal insufficiency. Kidney Int. 2001;60(2):741-747. doi:10.1046/j.1523-1755.2001.060002741.x.
3. Macdougall, I. C., Walker, R., Provenzano, R., de Alvaro, F., Locay, H. R., Nader, P. C., … Investigators, A. S. (2008). C.E.R.A. corrects anemia in patients with chronic kidney disease not on dialysis: results of a randomized clinical trial. Clin J Am Soc Nephrol, 3(2), 337–347. https://doi.org/10.2215/CJN.00480107
4. 18 October 2010 CRDAC Meeting Briefing Document Epoetin alfa (EPOGEN®/PROCRIT®) and darbepoetin alfa (Aranesp®) https://wayback.archive-it.org/7993/20170403223814/https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/ucm192863.htm
5. CDER, FDA. Omontys (peginesatide) BLA. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/202799orig1s000statr.pdf. Published 2012.
6. Akizawa T, Gejyo F, Nishi S, et al. Positive outcomes of high hemoglobin target in patients with chronic kidney disease not on dialysis: a randomized controlled study. Ther Apher Dial. 2011;15(5):431-440. doi:10.1111/j.1744-9987.2011.00931.x.
7. Eschbach JW, Abdulhadi MH, Browne JK, et al. Recombinant human erythropoietin in anemic patients with end-stage renal disease. Results of a phase III multicenter clinical trial. Ann Intern Med. 1989;111(12):992-1000. http://www.ncbi.nlm.nih.gov/pubmed/2688507.
8. Canadian Erythropoietin Study Group, Group CES, Group CES. Association between recombinant human erythropoietin and quality of life and exercise capacity of patients receiving haemodialysis. Canadian Erythropoietin Study Group. BMJ. 1990;300(6724):573-578. doi:10.1136/bmj.300.6724.573.

Table created manually.

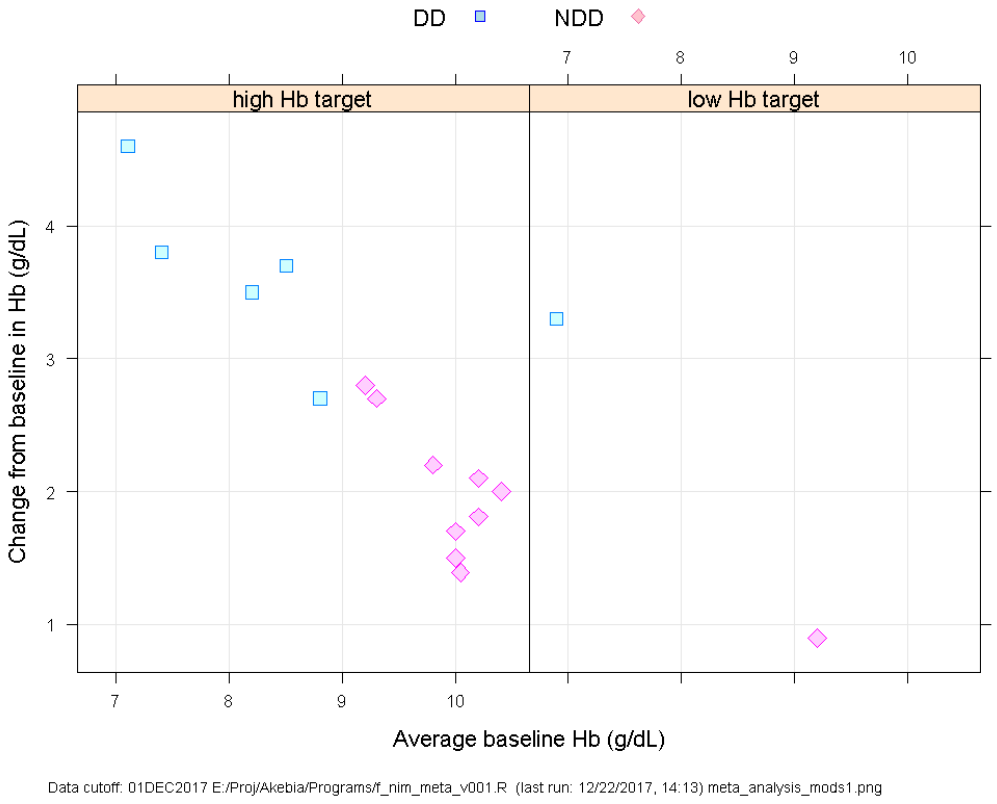
A.2.1. Placebo Control

Only 2 of these studies included a placebo arm: TREAT (for DA) and Amgen Canadian (for EpoA). In both cases, the mean estimated change from baseline in Hb was 0.3 g/dL after approximately 20 weeks. Most of the other studies were noninferiority trials that used either EpoA or DA as the reference group; they did not include a placebo group. See the methods below for handling this gap in available data.

A.2.2. Baseline and Target Hb Levels

Since the primary outcome variable is constrained by screening limits as well by therapeutic target ranges, the treatment effect for an effective drug (i.e., one increasing Hb into the target range) would naturally vary according to the average baseline Hb in the sample, guided somewhat by inclusion criteria, as well as the target Hb level. The treatment effect would be lower for the low-target studies than for the high-target ones. As Figure A 1 shows, the DD group of studies has lower average baseline Hb and higher treatment effect than the NDD studies. This suggests a moderator effect of baseline Hb in the 16 study arms included here: the higher the average baseline Hb level, the smaller the treatment effect. We interpret this result cautiously, however, because baseline Hb and DD/NDD status are almost completely confounded.

Figure A 1. Treatment Effect by Baseline and Target Hemoglobin Levels



A.3. Statistical Methods

The most important decisions in conducting a meta-analysis concern both the input data and the statistical method. The choice of input data (presented in Table A 2) depends on what studies to include and how to handle unavailable data. We faced a problem because so few previous studies had placebo arms and some of the studies had missing information about the pre and post treatment outcome variances and correlations. Our choice of statistical method, fixed effects model, was based on several packages in R (metafor, MAd) and accompanying literature.

A.3.1. Derivation of Input Statistics

We used the data collected from publications and applied the following rules to fill in the desired baseline and post-baseline means and standard deviations. Post-baseline refers to assessment after at least 20 weeks of treatment:

1. Compute the standard deviations from standard errors or half-widths of the reported confidence intervals.
2. Compute post-baseline means based on baseline and reported changes from baseline.
3. Use available data to impute missing means and SDs, making the post-baseline SD the same or slightly larger than SD for the baseline.
4. Compute correlation between baseline and post-baseline means (based on three arms with a standard error for the difference).
5. Set remaining pre-post correlations to 0.30, splitting the difference between the observed values from 0.14 to 0.58, to compute SD for the change scores. The largest study, TREAT, was “penalized” with a lower correlation of 0.10.

A.3.2. Placebo Adjustment

Three study arms had a placebo control arm. For the remaining 13 study arms, we introduced a hypothetical placebo arm with a mean of 0.3 and sample size and SD equivalent to the active treatment comparison arm. In this counterfactual scenario, we imagine each study had a placebo arm of size, baseline, and variance equal to the active treatment arm. Since there is general agreement that the placebo effect is essentially negligible, adding an imaginary placebo arm with small change from baseline will not distort the results away from the null hypothesis of no treatment effect.

Table A 2. Input Dataset for Meta-analysis

				Active treatment			Placebo control			
ID	Study Arm	DD	Hb Target	N	Mean BL (g/dL)	Mean Change (SD) (g/dL)	N	Mean BL (g/dL)	Mean Change (SD) (g/dL)	Pre-post corrl[a]
1	TREAT-DA	No	high	2012	10.4	2.00 (1.56)	2026	10.3	0.30 (1.56)	0.1
2	Affymax-OM1	No	high	328	10.0	1.50 (0.85)	328	10.0	0.30 (0.85)	0.3
3	Affymax-OM2	No	high	328	10.0	1.70 (0.85)	328	10.0	0.30 (0.85)	0.3
4	Affymax-DA	No	high	327	10.0	1.39 (0.85)	327	10.0	0.30 (0.85)	0.3
5	Roche-DA	No	high	162	10.2	1.81 (1.03)	162	10.2	0.30 (1.03)	0.3
6	Roche-MI	No	high	162	10.2	2.10 (1.03)	162	10.2	0.30 (1.03)	0.3
7	Akizawa-DA	No	high	161	9.2	2.80 (1.11)	161	9.2	0.30 (1.11)	0.3
8	Akizawa-Epo	No	low	160	9.2	0.90 (1.07)	160	9.2	0.30 (1.07)	0.3
9	Amgen202-DA	No	high	129	9.3	2.70 (0.99)	129	9.3	0.30 (0.99)	0.5
10	Amgen202-Epo	No	high	37	9.8	2.20 (1.01)	37	9.8	0.30 (1.01)	0.6
11	AmgenDial-Epo1	Yes	high	201	7.4	3.80 (2.15)	201	7.4	0.30 (2.15)	0.3
12	Amgen211-DA	Yes	high	91	8.8	2.70 (1.38)	91	8.8	0.30 (1.38)	0.1
13	AmgenCAN-Epo1	Yes	low	44	6.9	3.30 (1.18)	48	7.1	0.30 (1.40)	0.3
14	AmgenCAN-Epo2	Yes	high	38	7.1	4.60 (1.55)	48	7.1	0.30 (1.40)	0.3
15	AmgenDial-Epo2	Yes	high	35	8.2	3.50 (2.15)	35	8.2	0.30 (2.15)	0.3
16	Amgen211-Epo	Yes	high	31	8.5	3.70 (1.43)	31	8.5	0.30 (1.43)	0.3

BL: baseline ; CAN: Canada; corr: correlation; DA: darbepoetin alfa; DD: dialysis dependence; Epo: epopoetin alfa; Hb: hemoglobin; MI: Mircera; OM: Omontys SD: standard deviation

The placebo statistics for records in red are reported for a true placebo control arm.

a.The correlation between baseline and post-treatment sample means was derivable for study arms with IDs of 9, 1 0, and 12. In all other cases, it was assumed to be 0.1 or 0.3.

Data cut: 01Dec2017

E:\proj\akebia\programs\l_meta1.sas v.002 (last run: 01/17/2018, 13:33) l_meta1.rtf (page 1 of 1)

A.3.3. Justification of the Model

Meta-analytic techniques are often used to obtain overall estimates of treatment effects. When at least 25 studies are available for analysis, there is sufficient information to provide confidence in the combined estimates, confidence intervals, and statistical tests. When a sufficient number of studies is available, both the between and within-study variability can be estimated and used in calculating an overall treatment effect. Some statisticians favor random effects meta-analytic methods in this situation; however, most meta-analyses are not based on a large number of studies and the fixed effects approach becomes necessary. Moreover, the weights assigned to

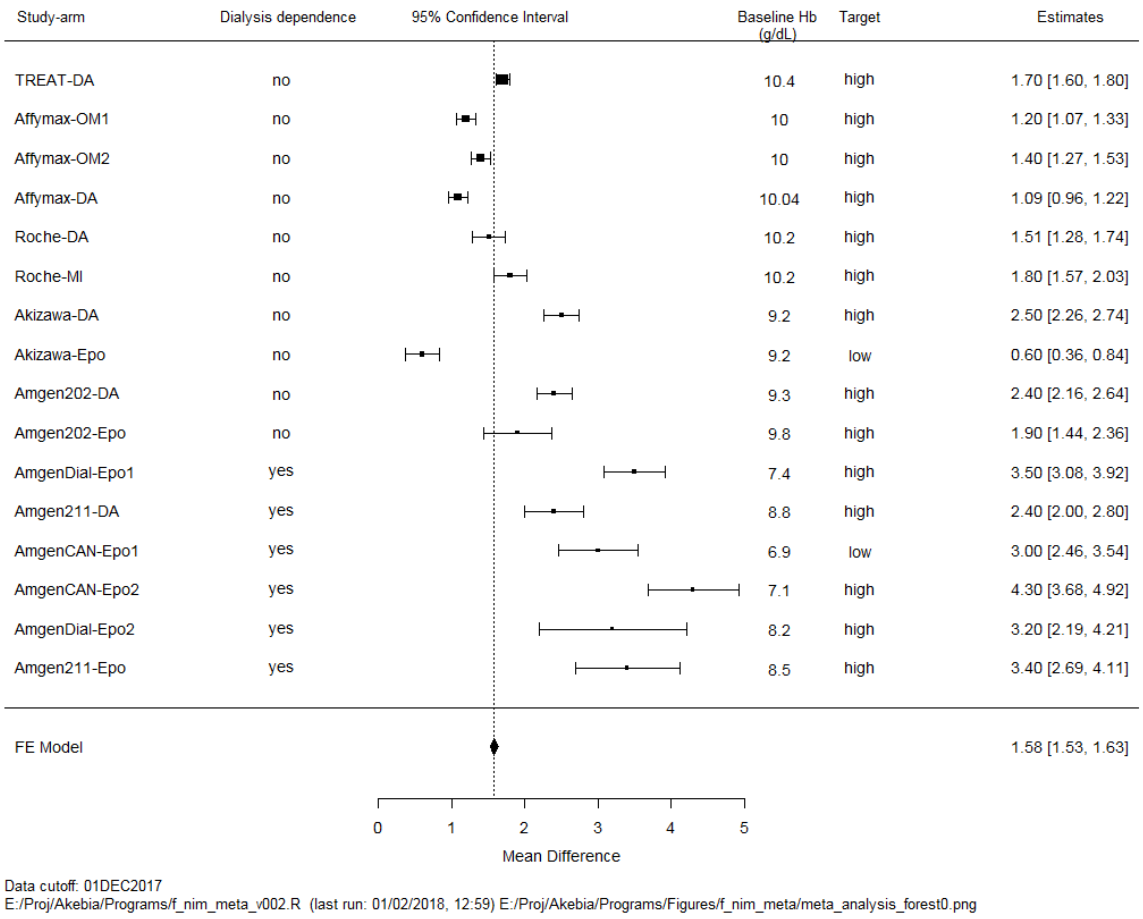
studies in random effects analyses are similar across studies allowing small studies to have the same effect on the overall estimates as large studies. Therefore, many statisticians favor fixed effects models even when a lot of studies are available.

Fixed effects meta-analysis approaches are based on the following set of assumptions: there is one true effect size; inferences are based on the studies included in the meta-analysis and cannot be formally extended beyond the collection of studies; the properties of the estimates, tests and confidence intervals are based on the total number of subjects across the studies rather than the number of studies; and the weights are approximately proportional to the sample size or follow-up time, in time-to-event studies. In this analysis, the 16 study arms range in size from 35 subjects to 2026 subjects per group favoring the use of the fixed effects approach.

A.4. Results of Meta-analysis

See Figure A 2 for the results of the specified meta-analysis model fit to the 16 study arms, augmented with placebo adjustment, with an overall mean as the sole fixed effect. The estimated placebo-adjusted change from baseline in Hb is 1.58 g/dL with a lower 95% confidence interval bound of 1.53 g/dL. We have also used this method to perform separate meta-analyses on the DD and NDD studies. The overall treatment effect estimates and 95% confidence intervals in g/dL were 1.50 (1.44, 1.55) and 3.17 (2.95, 3.38) for the NDD and DD subgroups, respectively.

Figure A 2. Forest Plot of Placebo-adjusted Treatment Effect of ESAs on Hemoglobin



Data cutoff: 01DEC2017
E:/Proj/Akebia/Programs/f_nim_meta_v002.R (last run: 01/02/2018, 12:59) E:/Proj/Akebia/Programs/Figures/f_nim_meta/meta_analysis_forest0.png

A.5. Rationale for Margin

Based on the 95% lower confidence bounds for the pooled and subgroup overview analyses, the percent of treatment effect that would be preserved varies according to the choice of noninferiority margin (see Table A 3). A noninferiority margin of -1.0 g/dL would preserve a third of the treatment effect for CI-0014 (not dialysis dependent [DD], ESA-naive) study and two-thirds of the treatment effect for the CI-0016 (dialysis dependent, -ESA-naive) study.
A noninferiority margin of -0.75 g/dL would preserve a half of the treatment effect for CI-0014 (not dialysis dependent, ESA-naive) study and three-quarters of the treatment effect for the CI-0016 (dialysis dependent, ESA-naive) study.

Table A 3. Percent of Treatment Effect Preserved by Study Type and Margin

Analysis Population (Number of Study Arms)	95% Lower Bound (g/dL)	Noninferiority Margin (g/dL)		
		-1	-0.75	-0.5
All studies pooled (16)	1.53	35%	51%	67%
Not dialysis dependent (10)	1.44	31%	48%	65%
Dialysis dependent (6)	2.95	66%	75%	83%

Table created manually.

APPENDIX B. MISSING DATA (INCLUDING MULTIPLE IMPUTATION)

This appendix describes the methods planned to handle missing data for the primary and secondary efficacy outcomes. Standard multiple imputation, where imputation of missing values is based on the group to which the subject was randomized, will be used for all analyses except the tipping point analysis. The tipping point analysis is described below to explore deviations from the assumptions that missing data would follow observed treatment arm trends.

The literature for handling missing data in noninferiority studies is limited and the considerations differ from those for a superiority study. For example, in a superiority study one can impute all missing data using the data obtained from the placebo group under a reference based imputation. Such an approach is generally considered a conservative analysis in the setting of a superiority study as imputation of missing data in the treated group is based on the assumption that no treatment was received. In this noninferiority setting, reference-based imputation is not a conservative approach as efforts to make the vadadustat group behave more like the darbepoetin alfa group are anti-conservative. We are not aware of any published statistical proposal for "conservative" imputation in a noninferiority study, but Wiens and Rosenkranz, (2013) demonstrated that standard imputation performs best for the handling of missing data in noninferiority studies.

B.1. Overall Approach

Missing data will be imputed through multiple imputation under fully conditional specification (FCS) with M = **100** imputed datasets. The general framework for the process is as follows:

- In the first step of the analysis, all missing values will be imputed for hemoglobin outcomes, all other efficacy outcomes, and any covariates to be used in the models. The primary outcome and secondary outcomes will be derived from the imputed values of Hb in the primary efficacy period (PEP) and secondary efficacy periods (SEP). The variables in the imputation model are delineated in Table B 1.
- A set of pre-assigned seeds are shown in Listing B 1. These seeds will be used for all imputations required in both efficacy and MACE safety analyses.
- All multiple imputation will be implemented using FCS.
- The type of model used for FCS will be tailored to the variable with missing outcome data. Specifically, the following types of regression will be used:
  - Continuous variables: regression
  - Binary variables: logistic regression
  - Ordinal classification variables: response logistic regression
  - Nominal classification variables: response logistic regression
- Rubin's rule will be used to obtain the final result which can be computed using PROC MIANALYZE.
- A total of 40 iterations will be used for the burn-in.
- Before inclusion in the model, variables will be assessed for collinearity using the condition index and regression coefficient variance-decomposition matrix computed using PROC REG in SAS.

This appendix provides SAS code for the imputations for all variables of interest. This code, and the associated seeds, will be used to generate the analysis datasets.

To ensure uniformity across all analyses, the change from baseline and binary outcomes derived from Hb measures will be imputed by calculating the imputed value of the continuous measure and then deriving the change score or binary outcome (Ratitch, 2016). This procedure will be adjusted if the imputation model proves inconsistent with the substantive model.

Table B 1. Outcome Variables and Covariates in the Multiple Imputation Models

Type	Variable[a]	Description	Baseline, Post, Outcome	Levels
Continuous	HBBL	Baseline Hb	Baseline	n/a
Continuous	HBS1	Hb in Weeks 2-8	Post	n/a
Continuous	HBS2	Hb in Weeks 10-20	Post	n/a
Continuous	HBS3	Hb in PEP (Weeks 24-36)	HBCHGS3, HBRNGS3	n/a
Continuous	HBS4	Hb in SEP (Weeks 40-52)	HBCHGS4, HBRNGS4	n/a
Continuous	TIBCBL	Baseline TIBC	Baseline	n/a
Continuous	HEPCNBL	Baseline hepcidin	Baseline	n/a
Continuous	FERRBL	Baseline ferritin	Baseline	n/a
Continuous	UACRBL	Baseline uACR (PRO2 only)	Baseline	n/a
Continuous	CRPBL	Baseline CRP	Baseline	n/a
Continuous	TSATBL	Baseline TSAT	Baseline	n/a
Continuous	EGFRBL	Baseline eGFR (PRO2 only)	Baseline	n/a
Continuous	TIBCS3	TIBC in PEP (Weeks 24-36)	Post	n/a
Continuous	HEPCNS3	Hepcidin in PEP (Weeks 24-36)	Post	n/a
Continuous	FERRS3	Ferritin in PEP (Weeks 24-36)	Post	n/a
Continuous	UACRS3	uACR in PEP (Weeks 24-36) (PRO2 only)	Post	n/a
Continuous	CRPS3	CRP in PEP (Weeks 24-36)	Post	n/a
Continuous	TSATS3	TSAT in PEP (Weeks 24-36)	Post	n/a
Continuous	EGFRS3	eGFR in PEP (Weeks 24-36) (PRO2 only)	Post	n/a
Continuous	ESABL	Baseline ESA dose	Baseline	n/a
Continuous	TDIALBL	Time on dialysis prior to Screening Visit 1 (INNO2 only)	Baseline	n/a
Continuous	AGE	Age	Baseline	n/a
Binary	SEX	Sex male	Baseline	2
Nominal	RACEGR2	Race: white, black, other	Baseline	3
Binary	TRT01P	Randomized treatment	Baseline	2
Binary	DIALTYP	Dialysis type (INNO2 only)	Baseline	2
Binary	DMHX	History of diabetes mellitus	Baseline	2
Binary	CVHX	History of CV disease (coronary artery disease, MI, stroke, HF)	Baseline	2
Binary	CVOTHHX	History of other CV disease (DVT, Arterial Thrombosis, PE, Vascular Access Thrombosis)	Baseline	2
Ordinal	SMOKEHX	Smoking history	Baseline	3
Binary	STRAT3P	NYHA CHF class 0/1 vs. 2/3	Baseline	2
Nominal	STRAT2A	Geographical region	Baseline	3

Table created manually.

a. Variables names are preliminary and will be adjusted to comply with CDISC.

B.2. Multiple Imputation of all Missing Values

Multiple imputation will be used to create M datasets for the primary and secondary efficacy analyses. The imputation relies on the use of FCS and model code provided in Exhibit B 1 below. The examples of code come from version 9.4 of SAS. Code will be modified slightly as needed according to software updates and compatibility with actual data.

The variables in Table B 1 will be used to populate macro variables &BINVARS for all binary variables, &NOMVARS for all nominal variables, and &ORDVARS for all ordinal variables. The variables are imputed sequentially in the order specified in the VAR statement.

Except where specified, imputation procedure will rely on the default behavior of FCS for the method and predictor covariates. The variable being imputed is sampled from a posterior distribution conditional on all the values, observed and imputed, of variables to the left of it in the VAR statement and all the observed-only values of variables to its right. If no FCS statement is included for a variable with missing data, the default method depends on the variable type. The regression method is default for continuous variables, and the discriminant function method is default for categorical variables (list in CLASS statement).

B.3. Primary Efficacy Analysis

The primary efficacy analysis is based on the difference between the changes from baseline (average pretreatment Hb) to the PEP (average of up to 4 per-visit observations in Weeks 24 to 36). The noninferiority analysis is based on a margin of -0.75 g/dL applied to the difference in mean change between vadadustat and darbepoetin alfa. To establish noninferiority, the lower limit of the confidence interval of the mean change in vadadustat minus the mean change in darbepoetin alfa must be -0.75 g/dL or higher. ANCOVA with multiple imputation for missing data will be used to calculate the 95% confidence interval of the difference.

Multiple imputation will rely on the method of fully conditional specification (FSC). Under this assumption, data for each variable with missing values will be imputed with a separate regression model allowing for all available data to be used in the imputation process. For continuous variables, the regression method will be used for imputations. For categorical variables, logistic regression models will be used for binary outcomes, ordinal logistic regression for ordinal responses, and nominal response logistic regression for nominal responses. The models for imputation under each of these conditions will be described in this appendix.

The output from the multiple imputation procedure is M complete datasets. For the primary analysis, these datasets are analyzed using ANCOVA and the code in Exhibit B 2 below. The M estimates of the difference in the change from baseline in the PEP are then "combined" using PROC MIANALYZE to obtain a final estimate and its associated significance level.

B.3.1. SAS Code for Standard Multiple Imputation using FCS

This section describes the procedures to impute missing values for the primary efficacy analysis. Missing values are imputed separately for each treatment group using a BY statement in PROC MI.

Exhibit B 1. SAS Code for Standard Imputation of Missing Values for the Primary Efficacy Analysis

```
proc mi data=&INDAT. nimpute=&M. seed=&SEED1. out=&OUTDAT.;
  by TRT;
  class &BINVARS. &NOMVARS. &ORDVARS.;

  var
    STRAT2_HF STRAT3_REG
    SEX RACE AGE HX_CV4 HX_CV_OTH HX_DM ESA_BL SMOKE_BL
    HEP_BL FERR_BL TIBC_BL CRP_BL TSAT_BL HB_BL
    HB_1 HB_2 HEP_3 FERR_3 TIBC_3 CRP_3 TSAT_3
     HB_3 HB_4;
  /* continuous - primary */
  fcs plots=trace reg ( HB_3) nbiter=40;

  /* continuous example - other */
  fcs reg ( HB_1) nbiter=40;

  /* binary or ordinal example */
  fcs logistic( SEX / link=logit likelihood=augment) nbiter=40;

/* nomial example */
  fcs logistic (RACE/link=glogit likelihood=augment) nbiter=40

       run;
```

Exhibit B 2. SAS Code to Compute the M Estimates of Treatment Effect

```
*** derive primary and other endpoints;
data &OUTDAT2.;
      set &OUTDAT.;
      DELT_HB3 = HB_3 – HBBL;
      DELT_HB4 = HB_4 – HBBL;
      if (STRAT3 eq “US” and 10 le DELT_HB3 le 11) or
          (STRAT3 ne “US” and 10 le DELT_HB3 le 12) then
            HBRNGS3=”Yes”;
      else HBRNGS3=”No”;
      if (STRAT3 eq “US” and 10 le DELT_HB4 le 11) or
          (STRAT3 ne “US” and 10 le DELT_HB4 le 12) then
```

```
                HBRNGS4=”Yes”;
        else HBRNGS4=”No”;
run;
proc genmod data=&OUTDAT2.;
        by _IMPUTATION_
        class TRT (ref="DARB") STRAT1 STRAT2 ;
        model DELT_HB3=HB_BL STRAT1 STRAT2 TRT
                / alpha=0.05 cl diagnostics residuals ;
        estimate “Primary Efficacy” TRT 1 -1;
        lsmeans TRT / cl pdiff cov;
        ods output diffs = MI_ESTIMATES lsmeans=MI_ESTIMATES2;
run;
```

The execution of the above set of SAS code will provide the datasets MI_ESTIMATES and MI_ESTIMATES2 which will contain M lines of data that include the estimates and standard errors for each imputed dataset. The final result is then obtained with application of PROC MIANALYZE provided in Exhibit B 3.

Exhibit B 3. SAS Code to Obtain the Overall Estimate and Standard Error

```
proc mianalyze data= MI_ESTIMATES2 (or MI_ESTIMATES);
        modeleffects estimate;
        stderr;
        ods output ParameterEstimates=MI_PRIMARY;
run;
```

The final results will be contained in the output dataset MI_primary. This dataset will contain the 95% confidence limits of the difference. The primary hypothesis is tested by comparing the lower limit of this confidence interval to -0.75 g/dL.

B.3.2. Tipping Point Sensitivity Analysis for Missing Data

This multiple imputation model for the primary efficacy model described above assumes that data in both treatment arms are missing at random (MAR) and would therefore follow the trend of observed data. The tipping point analysis assesses the effect of potential deviations from this assumption and explores the consequences of assuming that data in the vadadustat arm are missing not at random (MNAR) (i.e., subjects in the vadadustat arm with missing outcome are assumed to have a lower Hb values than subjects in the darbepoetin alfa arm). The specific steps are as follows:

A. The missing information on Hb in the vadadustat arm will be multiply imputed using a shift parameter S applied to lower the mean Hb values in that arm. The sample SAS code for performing multiple imputation with a shift S in the vadadustat arm is presented in Exhibit B 4. Missing data in the darbepoetin alfa arm will continue to be imputed assuming MAR, using the same code as for the primary efficacy analysis subset to the darbepoetin arm.

B. The multiply-imputed data will be analysed using standard multiple imputation combining rules to obtain an estimated treatment effect and its associated 95% confidence interval. These analyses provide a realistic estimate of variance for the treatment effect that takes into account the uncertainty caused by the missing data.

C. Steps A and B will then be repeated. The shift parameter S starts from 0, which corresponds to the primary efficacy analysis with no shift effect, and increased in the negative direction by a certain amount in each step until the analysis reaches the "tipping point", the point at which the effect of vadadustat is no longer noninferior to that of darbepoetin alfa. The more the tipping point diverges from the observed data, the more robust the conclusion based on primary efficacy analysis.

The tipping point analysis will use a single seed from Listing B 1 to generate however many seeds are needed, i.e., 1 per treatment group per shift.

Exhibit B 4. SAS Code for Tipping Point Analysis

```
proc mi data=&INDAT. (where=(TRT="VADA")) nimpute=&M.
        seed=&SEED2. out=&OUTDAT.;
   class &BINVARS. &NOMVARS. &ORDVARS.;
   var
    STRAT2_HF STRAT3_REG
    SEX RACE AGE HX_CV4 HX_CV_OTH HX_DM ESA_BL SMOKE_BL
    HEP_BL FERR_BL TIBC_BL CRP_BL TSAT_BL HB_BL
    HB_1 HB_2 HEP_3 FERR_3 TIBC_3 CRP_3 TSAT_3 HB_3
    HB_4;
   fcs reg ( HB_3) nbiter=40;
   mnar adjust ( HB_3 / shift=&S. );
run;
```

B.4. Secondary Efficacy Analyses

The secondary endpoint, change in Hb at Time 4, is continuous and will be analyzed in a manner similar to the method to be used for the primary efficacy analysis.

B.5. Other Efficacy Analyses

For the below 2 efficacy endpoints of having average Hb value in the geography-specific target range in weeks 24-36 and having average Hb value in the geography-specific target range in weeks 40-52, their endpoints are binary; they will be analyzed according to the Mantel-Haenszel estimate of stratified risk differences. These binary outcome measures will be computed from the continuous imputed data which will serve as the input data set for the computation of the estimate of risk difference. The code for this analysis is provided in Exhibit B 5. Rubin's rule requires that the input statistic is normally distributed, which we assume for this estimate.

Exhibit B 5. SAS Code to Obtain the Mantel-Haenszel Estimate of Risk Difference

```
*** compute the MH estimate;
proc freq data=&OUTDAT2.;
   by _IMPUTATION_;
   tables STRAT1*STRAT2*STRAT3*TRT*&OUTCOME/riskdiff(common) cmh;
   ods output CommonPdiff=CommonPdiff;
run;
*** combine the results;
proc mianalyze data= CommonPdiff;
        modeleffects Value;
        stderr;
        ods output parameterestimates=MI_SEC;
run;
```

Multiple imputation is not currently planned for the other efficacy analyses, except in the case that outcomes for these analyses are part of the multiple imputation for the primary efficacy analysis. If analysis of a covariate included in the overall imputation model (Table B 1) is desired, follow the procedure for the primary or secondary analyses, depending on the variable type.

B.6. Random Seed Specification

This section presents the generation code, resulting random seeds, and corresponding analysis assignments for all analyses with multiple imputation. The random seeds for all the multiple imputation runs for each trial will be generated from a single master seed. The master seeds will be 10014 for CI-0014, and 10015 for CI-0015.

Exhibit B 6. SAS Code for Random Seed Generation

```
data SEED;
 call streaminit(&MASTER.);
 do I = 1 to &NUMSEED. ;
   SEED = put(ceil( ((2**31) - 1) * rand("UNIFORM") ), best.);
   output;
 end;
run;
```

The seeds and their analysis assignments are shown in Listing B 1. These will be used for all imputations required in both efficacy and MACE safety analyses for this set of trials.

Listing B 1. Random Seed Assignment for all Planned Analyses

			PRO$_2$TECT		INNO$_2$VATE	
Analysis Category	**Analysis ID**	**Analysis Detail**	**CI-0014[a]**	**CI-0015[b]**	**CI-0016[c]**	**CI-0017[d]**
Primary Efficacy, Secondary, & All Subgroups	PE	Impute missing Hb & covariates	635320903	1987393372	32552474	1864814524
Primary Efficacy - Sensitivity	PES1	Rescue definition 1 (narrow)	429585122	1373196336	1027066416	928246872
	PES2	Rescue definition 2 (broad-on-treatment)	549618248	1156395436	477818637	2071955541
	PES3	Tipping point	340408944	1747908400	1450082522	216479710

Analyses that use a shared seed between CI-0014 and CI-0015 or CI-0016 and CI-0017 are highlighted in yellow.
a. Seeds for CI-0014 are generated based on a master seed of 10014.
b. Seeds for CI-0015 are generated based on a master seed of 10015.
c. Seeds for CI-0016 are generated based on a master seed of 10016.
d. Seeds for CI-0017 are generated based on a master seed of 10017.

APPENDIX C. ANALYSIS OF CONTINUOUS ENDPOINTS

This appendix provides details related to the analysis of the continuous endpoints in the studies. It includes a description of the statistical approaches and provides examples of the SAS code for the analysis of specific continuous endpoints. The primary efficacy analysis will be an analysis of covariance (ANCOVA) with multiple imputation for missing data based on standard imputation. The details of this approach are provided in Appendix B.

Additional analyses will use the ANCOVA method including some sensitivity analyses for the primary efficacy endpoint, secondary efficacy endpoint analyses, and other continuous exploratory endpoint analyses. Other sensitivity analyses for the primary efficacy endpoint will use a mixed model repeated measures (MMRM) without imputing missing data.

C.1. Analyses using ANCOVA

C.1.1. Primary Analysis

The primary analysis is based on the change in average Hb between the baseline and the primary efficacy period (PEP), Weeks 24 to 36. The noninferiority analysis will be based on a margin of -0.75 g/dL applied to the mean change in vadadustat minus darbepoetin alfa. To establish noninferiority, the lower limit of the confidence interval of the mean change in vadadustat minus darbepoetin alfa must be -0.75 g/dL or higher. If the lower limit of the 2-sided 95% confidence interval for the difference between the mean in the vadadustat group and the mean in the darbepoetin alfa group is above zero, superiority will have been established and the finding will be interpreted as providing evidence of a greater change from baseline in Hb for vadadustat relative to the control arm.

The primary efficacy analysis is based on the estimated confidence interval estimated from an ANCOVA model. The computation of this confidence interval is described in this section. Modifications for additional continuous endpoints are outlined in a separate section.

The ANCOVA model will be fit using PROC GENMOD in SAS with the following variables:

- HB_BL – this is the average of the last 2 Hb measure on or prior to the first dose date
- DELT_HB3 – this is computed as the average Hb measure computed from the PEP (Week 24 to 36) visits minus HB_BL
- TRT – denotes treatment group; vadadustat or darbepoetin alfa
- STRAT1 – stratification factor of geographic region (US, EU, versus Rest of World)
- STRAT2 – NYHA CHF class 0 or I versus II or III

The basic SAS code to fit the model is in Exhibit C 1, but the full process includes multiple imputed datasets as seen in Exhibit B 2.

Exhibit C 1. SAS Code for Primary Analysis

```
proc genmod data=&OUTDAT2.;
        class TRT (ref="DARB") STRAT1 STRAT2 ;
        model DELT_HB3=HB_BL STRAT1 STRAT2 TRT
                / alpha=0.05 cl diagnostics residuals ;
        estimate “Primary Efficacy” TRT 1 -1;
        lsmeans TRT / cl pdiff cov;
        ods output diffs = MI_ESTIMATES lsmeans=MI_ESTIMATES2;
run;
```

The confidence limits for the coefficient associated with TRT in the above model are used to compute the confidence intervals based on M imputed datasets.

C.1.2. Select Sensitivity Analyses for the Primary Efficacy Endpoint

To assess the robustness of the primary efficacy analysis, the sensitivity analyses outlined below will be conducted using the same method in the primary efficacy analysis, but starting from different missing data patterns, imputation models, or analysis populations. As with the primary analysis, the following sensitivity analyses will use multiple imputation and ANCOVA.

- The primary analysis will be repeated after setting to missing all per-visit hemoglobin values within four weeks of administration rescue therapy. The definition of rescue therapy will vary across the series described in Section 6.3.9.
- Tipping point analyses (see Appendix Section B.3.2) will be performed to assess the effect of the missing data.
- The Full Analysis Set (FAS) population will be used instead of the randomized population. This analysis will be performed only if the size of the FAS is less than 95% of the randomized population.

C.1.3. Analysis of Continuous Secondary Endpoint

The secondary efficacy endpoint is the change in average Hb value between Baseline and the secondary efficacy period (SEP), Weeks 40 to 52. Analyses of the other efficacy endpoints are described in Appendix D.

The analysis of this secondary continuous endpoint will be similar to the analysis approach for the primary efficacy shown in Exhibit B 2 and Exhibit C 1, except the outcome of interest in the model statement DELT_HB3 will be replaced with DELT_HB4, the change in Hb from baseline to the SEP. This is computed as the average Hb measure computed from the Week 40 to 52 visits minus HB_BL.

The assessment of noninferiority will be tested using the same approach with the lower limit of the confidence interval compared to -0.75 g/dL.

Note that of all the continuous efficacy endpoints, only the primary efficacy endpoint and the secondary efficacy endpoint of the change in Hb between baseline and the SEP will be formally tested. All other continuous endpoint analyses will be descriptive.

C.1.4. Analysis of Additional Continuous Endpoints

The analysis of additional continuous endpoints will be similar to the analysis approach for the primary efficacy analysis (see either Exhibit C 1 or Exhibit B 2) depending on whether using multiple imputation), and adjustments to these analyses approaches are outlined for each example. The team may consider alternative methods or links to analyze the certain endpoints as noted below. Several additional continuous endpoints are to be evaluated:

- Change in average Hb value between Baseline and the combined PEP and SEP (Weeks 24 to 52).
- Average weekly dose of elemental iron administered. This model will include the average weekly dose of elemental iron by route groups (oral, IV) for each analysis time period in Year 1 and Weeks 52 to EOS. Note that some variants will not be multiply imputed, in which case the analysis will be a single ANCOVA of the observed values.
- Change in hepcidin, ferritin, TSAT, serum iron, and TIBC over time. The code for this analysis is identical to that for the primary analysis with the new outcome measure of interest. Four outcomes will be tested for each lab value – the change from baseline to PEP, the change from baseline to SEP, the percentage change from baseline to PEP, and the percentage change from baseline to SEP.

C.2. Analyses Using MMRM

Additional sensitivity analyses will be performed under an alternate framework to assess the robustness of the primary efficacy analysis with respect to missing data.

A mixed model repeated measures (MMRM) will be fit to the observed data only *without imputing missing values*. The repeated measures will be the observed averages for analysis time periods 1, 2, and 3 (See definitions, Section 5.1). Note that analysis time period 4 will be included in the sensitivity analysis for the continuous secondary efficacy endpoint (Section 9.2).

The following variables will be used in the model:

- DELT_HB – this is computed as the Hb measure minus HB_BL
- ASPER – refers to the subperiod number post-baseline (1, 2, 3)
- SUBJID – denotes subject identifier

The SAS code to fit the model is as follows:

Exhibit C 2. SAS Code for Sensitivity Analysis for the Primary Efficacy Endpoint

```
proc mixed data = &LONGDAT. method=reml alpha=0.05 covtest;
  where ASPER le 3;
  class SUBJID TRT (ref="DARB") ASPER (ref=last)
STRAT1 STRAT2 ;
  model DELT_HB3 = HB_BL STRAT1 STRAT2 TRT ASPER
          HB_BL*ASPER TRT * ASPER / s ddfm=kr
  repeated ASPER / subject = SUBJID type = UN r;
  estimate "MMRM Sensitivity" TRT 1 -1 TRT*ASPER 0 0 1 0 0 -1/ cl;
  ods output estimates=MMRM_EST;
run;
```

The test of efficacy is obtained from the ESTIMATE statement. The model will have the following characteristics:

- The response variable will be the vector of observed change (average of per-window changes) in the primary efficacy endpoint from baseline to the average value in each analysis time period. The endpoint will consist of the three repeated measures defined as Analysis Time Period X Hb – baseline Hb.
- Repeated post-baseline measurements from each subject will be identified by the subject identifier denoted as SUBJID in the PROC MIXED code.
- Within-subject correlations will be modeled using an unstructured covariance structure.
- In the unlikely situation that this model does not converge, the model will use the heterogeneous Toeplitz structure, which assumes the correlation between 2 repeated measurements depends solely on their lag in visit number and that the variance of the endpoint may differ over time.
- If the model still does not converge with the heterogeneous Toeplitz structure, the model will use a homogeneous Toeplitz structure.
- Finally, if the model using the homogeneous Toeplitz structure does not converge, the model will use a compound symmetry structure which assumes equal correlation for a subject's measurements, regardless of how far apart in time they were taken.
- The Kenward-Roger degrees of freedom approximation will be used in calculating the ESTIMATE contrasts for the sensitivity analysis.

APPENDIX D. ANALYSIS OF NON-NORMAL ENDPOINTS

D.1. Analysis of Binary Endpoints

This section describes the methods for analyzing other binary endpoints.

For each binary endpoint, a 95% confidence interval will be calculated for each proportion by treatment group using the Clopper-Pearson method [Clopper, 1934].

For analyses conducted in the randomized population, the weighted difference in proportions between treatment groups, adjusting for the stratification factors, and the corresponding 2-sided 95% confidence interval will be calculated by weighting according to the Mantel-Haenszel method The associated odds ratio and corresponding 2-sided 95% confidence interval will also be calculated. Subgroup analysis will be conducted similarly.

The below efficacy endpoints have binary outcome measures and will use the code structure presented in Exhibit D 1. For the first 4 endpoint, noninferiority will have been established if the lower limit of the confidence interval for the risk difference is above -15%.

- At least 1 Hb value in the geography-specific target range in Weeks 24 to 36
- At least 1 Hb value in the geography-specific target range in Weeks 40 to 52
- Hb values in the geography-specific target range for at least 1/2 of the observations in Weeks 24 to 36
- Hb values in the geography-specific target range for at least 1/2 of the observations in Weeks 40 to 52
- Hb increase of >1.0 g/dL in Year 1 (CI-0014)
- Receipt of at least 1 administration of elemental iron, repeated using each route, intravenous (IV), or oral
- Receipt of any red blood cell (RBC) transfusion
- Receipt of any erythropoietin-stimulating agent (ESA) rescue medication
- Receipt of any rescue therapy (series of definitions from narrow to broad)
- Hb values >12, 13, or 14 g/dL
- Hb values <8 or 9 g/dL
- Hb increase of >1.0 g/dL within any 2-week period (CI-0015)
- Hb increase of >2.0 g/dL within any four-week period (CI-0015)

Exhibit D 1. SAS Code for Analyzing Binary Endpoints

```
proc freq data=&OUTDAT2.;
   tables STRAT1*STRAT2*STRAT3*TRT*&OUTCOME/riskdiff(common) cmh;
   ods output CommonPdiff=CommonPdiff;
run;
```

D.2. Analysis of Count Data

This section describes the methods for analyzing endpoints measured as counts. These endpoints include:

- Number of ESA episodes
- Number of RBC transfusion episodes

The initial analysis approach will be descriptive, reporting the proportion of subjects with counts greater than 0 and the mean number for this population. Statistical modeling based on a zero-inflated Poisson regression model, as outlined in Exhibit D 2, will be used to account for stratification factors.

Exhibit D 2. SAS Code for Analyzing Count Endpoints

```
proc genmod data=&INDAT.;
class TRT (ref="DARB") STRAT1 STRAT2;
model COUNT = TRT HB_BL STRAT1 STRAT2 / dist=zip
                                        offset=LN_FUT;
zeromodel;
output out=ZIP predicted=PRED pzero=PZERO;
estimate "Treatment Effect” TRT 1 -1 @ZERO TRT 1 -1;
ods output modelfit=FIT estimates=ZIPEST;
run;
```

D.3. Analysis of Time to Event Data

This section describes the methods for analyzing time-to-event endpoints, for example, time to:

- achieve 1g/dL increase in Hb in Year 1 (CI-0014)
- first rescue therapy (by series of definitions)
- first RBC transfusion
- first ESA use (aside from control arm medication not specified as rescue)
- progression of chronic kidney disease (CKD).
- study discontinuation

Time-to-event distributions will be estimated using the Kaplan-Meier product-limit method and log rank test.

The hazard ratio and corresponding 95% confidence interval for the comparison of vadadustat to darbepoetin alfa will be estimated using a Cox proportional hazards model, with treatment group and other predictors as covariates. Analyses in the randomized population will be stratified by the stratification factors; analyses in subgroups will be conducted similarly. The predictor variables (&PREDVAR. in Exhibit D 3) for hemoglobin-related events are baseline ESA dose (three

levels specified in Section 11) and continuous baseline Hb. For CKD progression, the predictor variable is continuous baseline eGFR.

Exhibit D 3. SAS Code for Time to Event Data Analysis

```
*** compute median times with KM product limit method;
proc lifetest data=&TTEDAT. alphaqt=0.05 conftype=LOGLOG confband=ALL
timelim=OBSERVED outsurv=RATES timelist= 0 24 36 40 52 reduceout;
      by TRT;
      time AVAL * CNSRN(1);
run;

*** compute hazard ratios with Cox proportional hazards model;
proc phreg data=TTE;
class TRT;
model AVAL * Status(0) = TRT HB_BL STRAT1 STRAT2 &PREDVAR. / ties=Efron;
hazardratio 'Cox proportional hazard' TRT;
ods output HazardRatios=hazard;
run;
```

For laboratory-based time-to-event endpoints, subjects who do not experience an event will be censored on the date of the last evaluable assessment of the laboratory parameter. For time to progression of CKD, subjects who do not experience CKD progression will be censored on date of last visit with no kidney transplant or dialysis.

APPENDIX E. CLINICAL LABORATORY AND VITAL SIGN SAFETY ENDPOINTS

This appendix lists endpoints and cut-off points for vital signs, physical findings, and other observations related to safety. Summary tables by visit, and by-subject listings in select cases, will be provided for all outcomes, observed and change from baseline, in this section. All listings will be sorted by treatment group, subject, parameter, and time period as appropriate.

If the central laboratory uses assays that have lower limits of detection (LLD), all laboratory results below the LLD will be imputed with the LLD. Results reported as greater than a value (i.e., "> value") will be imputed as 1.5 × that value.

E.1. Clinical Laboratory Evaluation

E.1.1 Liver Function Abnormality

A summary of liver function abnormalities by analysis period will be provided by treatment group.

Any subject with at least 1 of the following liver function abnormalities will be summarized:

- Alanine aminotransferase (ALT) >2 × and ≤3 × upper limit of normal (ULN); ALT >3 × and ≤5 × ULN; ALT >5 × and ≤10× ULN; ALT >10 × ULN
- Aspartate aminotransferase (AST) >2 × and ≤3 × ULN; AST > 3× and ≤ 5 × ULN; AST >5 × and ≤10 × ULN; AST >10 × ULN
- Bilirubin >2× and ≤3× ULN; Bilirubin >3× ULN

In addition, a table will summarize the occurrence of events that satisfy the following versions of Hy's Law:

- (ALT or AST >3× ULN and <= 5xULN) and total bilirubin >2× ULN;
- (ALT or AST >5× ULN and <= 10xULN) and total bilirubin >2× ULN;
- ALT or AST >10× ULN and total bilirubin >2× ULN

E.1.2 Serum and Urine Pregnancy Tests

Serum pregnancy test for females of childbearing potential will be performed at screening visit 2 and will be used for the subject's inclusion or exclusion. A urine pregnancy test will be performed during the screening visit and the results must be negative before subject's initiation of study drug. Additional pregnancy tests may be conducted during the study to establish the absence of pregnancy based on the investigator's clinical judgment or as required by local regulations.

A by-subject listing of serum and urine pregnancy tests will be provided.

E.1.3 Urine Albumin-to-creatinine Ratio (uACR)

A random urine spot sample is collected at the investigative site during the Baseline, Weeks 28, 52, 104, 156, 208 and end of treatment (EOT) to assess the urine albumin-to-creatinine ratio (uACR). A by-subject listing of change and percentage change from baseline in uACR will be provided, as well as geometric means (logarithmic scale) over time. The number and percent of subjects with a percent change in uACR exceeding 30% (in either direction) will be presented.

E.1.4 Complete Blood Count

The following components of the complete blood count (CBC) will be analyzed:

- mean corpuscular volume
- mean corpuscular Hb
- mean corpuscular Hb Concentration
- red cell distribution width
- white blood cell count with differential (neutrophils, lymphocytes, monocytes, eosinophils, and basophils)
- platelets
- hematocrit
- RBC counts

Summary statistics will be provided for the following:

- Each component of the CBC by treatment group.
- Change from baseline in each component of the CBC by treatment group.
- Changes from baseline category to the worst reported Common Terminology Criteria for Adverse Events (CTCAE) grade in each analysis time period in Year 1 and Weeks 52 to EOS.

A by-subject listing of all components of the CBC will be provided by treatment group.

In addition for white blood cells (WBC), lymphocytes, neutrophils, and platelets, a summary of the number and percentage of subjects experiencing abnormal values will be presented by treatment group and analysis time period. This summary will indicate the worst reported Common Terminology Criteria for Adverse Events (CTCAE) grade for the analysis time period for all subjects experiencing CTCAE grade 3 values or higher.

E.1.5 Chemistry and Estimated Glomerular Filtration Rate (eGFR)

The following laboratory parameters (with preferred units in parentheses) will be analyzed:

- Sodium (mmol/L)
- Potassium (mmol/L)
- Bicarbonate (mEq/L)
- Chloride (mmol/L)
- Calcium (mmol/L)
- Magnesium (mEq/L)
- Phosphorus (mmol/L)
- Phosphate (mmol/L)
- Glucose (mg/dL)
- Creatinine (umol/L)
- Blood urea nitrogen (mg/dL)
- Creatine phosphokinase (U/L)
- Uric acid (umol/L)
- Albumin (g/dL)
- Total protein (g/L)
- Total bilirubin (umol/L)
- Alkaline phosphatase (U/L)
- Alanine aminotransferase (ALT) (U/L)
- Aspartate aminotransferase (AST) (U/L)
- Lactate dehydrogenase (U/L)

- Total cholesterol (mg/dL)
- Low-density lipoprotein (LDL) cholesterol (mg/dL)
- High-density lipoprotein (HDL) cholesterol (mg/dL)
- Triglycerides (mg/dL)
- Derived parameter: non-HDL-cholesterol (= total cholesterol minus HDL cholesterol) (mg/dL)
- Estimated glomerular filtration rate (eGFR) (calculated from serum creatinine) (mL/min/1.73 m^2).

Summary statistics will be provided for the following:

- All chemistry parameters and eGFR by treatment group.
- Change (absolute and percent) from baseline in chemistry and eGFR by treatment group.
- Changes from baseline category to the worst reported Common Terminology Criteria for Adverse Events (CTCAE) grade in each analysis time period in Year 1 and Weeks 52 to EOS.

The eGFR will be calculated from serum creatinine by the formula of the 2009 Chronic Kidney Disease Epidemiological Collaboration (CKD-EPI). See Appendix B of the protocols.

A by-subject listing of chemistry and eGFR will be provided by treatment group. The categories for the shift tables will be based on CTCAE grades if available.

In addition, a summary of the number and percentage of subjects experiencing clinically significant values for each of the following parameter definitions will be presented by analysis time period:

- Decrease in eGFR ≥40% from baseline.
- Potassium >6.0 mmol/L.

E.1.6 C-reactive protein and Vascular Endothelial Growth Factor

Summary statistics will be provided for the following:

- C-reactive protein (CRP) and vascular endothelial growth factor (VEGF) by treatment group.
- Change (absolute and percent) from baseline in CRP and VEGF by treatment group.

A by-subject listing of CRP and VEGF will be provided by treatment group.

E.1.7 Erythropoietin

Blood samples for erythropoietin (Epo) analysis are obtained at Baseline and at Weeks 4, 12, 28, and 52.

A by-subject listing and a summary table of Epo (absolute level and change from baseline) by treatment group and time of collection will be provided.

E.2. Vital Signs, ECGs, and Physical Findings

E.2.1 Vital Signs and Weight

Body weight and the following vital signs will be summarized:

- Systolic BP
- Diastolic BP
- Heart rate
- Respiratory rate
- Body temperature.

Summary statistics will be provided for the following:

- Vital Signs and Weight by treatment group.
- Change from baseline in Vital Signs and Weight by treatment group.
- The number and percentage of subjects experiencing the following findings by treatment group and by analysis time periods:
  - Systolic blood pressure (SBP) ≥ 160 mmHg or diastolic blood pressure (DBP) ≥ 110 mmHg
  - SBP ≤ 90 mmHg or DBP ≤ 50 mmHg
  - change in SBP or DBP in either direction ≥ 20 mmHg
  - Systolic blood pressure (SBP) ≥ 180 mmHg

E.2.2 12-lead ECG

The following variables will be summarized at the Baseline Visit:

- Heart rate
- PR interval
- QT interval
- QRS interval
- QTc (corrected)
- Overall result

A summary of each electrocardiography (ECG) parameter will be provided by treatment group. In addition, a by-subject listing of 12-lead ECG results will be provided.

APPENDIX F. ADVERSE EVENTS OF SPECIAL INTEREST

F.1. Potential Risks

The following potential risk factors will be presented as adverse events of special interest (AESI):

- Hypersensitivity: Hypersensitivity Standardised MedDRA Query (SMQ) Narrow
- Hyperkalemia: Hyperkaliemia MedDRA PT, Blood potassium abnormal MedDRA PT, Blood potassium increased MedDRA PT
- Hypertension: Hypertension MedDRA SMQ Narrow

F.2. Events of Special Interest

The following will be presented as AESI:

- Hepatotoxicity: Drug related hepatic disorders Comprehensive SMQ broad
- Pulmonary Hypertension: Pulmonary Hypertension SMQ narrow
- Cardiac Valve disorders: Cardiac Valve disorders high-level group term (HLGT)
- Adrenal disorder: high-level term (HLT) Adrenal gland disorders not elsewhere classfied (NEC), HLT Adrenal cortex tests

F.3. Events Under Monitoring

The following will be presented as AESI:

- Malignancies: Malignancies MedDRA SMQ Narrow
- Congestive heart failure: Cardiac failure MedDRA SMQ Narrow
- Retinal effects due to Vascular Endothelial Growth Factor (VEGF) expression: Retinal disorders SMQ Narrow

14 REFERENCES

Choi P, Farouk M, Manamley N, Addison J. Dose conversion ratio in hemodialysis patients switched from darbepoetin alfa to peg-epoetin beta: AFFIRM study. *Adv Ther* 2013; 30:1007-17.

Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; 26:404-23.

Coresh J, Heerspink HJL, Sang Y, et al. Change in albuminuria and subsequent risk of end-stage kidney disease: an individual participant-level consortium meta-analysis of observational studies. *Lancet Diabetes Endocrinol* 2019; 7:115-27.

Cremieux PY, Van Audenrode M, Lefebvre P. The relative dosing of epoetin alfa and darbepoetin alfa in chronic kidney disease. *Curr Med Res Opin* 2006; 22:2329-36.

Food and Drug Administration. Aranesp (Darbepoetin alfa) label. 2011. Available at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2011/103951Orig1s5173_103951Orig1s5258lbl.pdf.

Gansevoort RT, Matsushita K, van der Velde M, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. *Kidney Int* 2011; 80:93-104.

Gosselin A, McKenzie R, Lefebvre P, et al. Dose-conversion ratio for epoetin alfa and darbepoetin alfa in chemotherapy patients with anemia and cancer. *P&T* 2006; 31:594-6, 601-3.

Heerspink HJL, Greene T, Tighiouart H, et al. Change in albuminuria as a surrogate endpoint for progression of kidney disease: a meta-analysis of treatment effects in randomised clinical trials. *Lancet Diabetes Endocinol* 2019; 7:128-39.

Jordan J, Breckles J, Leung V, et al. Conversion from epoetin alfa to darbepoetin alfa: effects on patients' hemoglobin and costs to Canadian dialysis centres. *Can J Hosp Pharm* 2012; 65:443-9.

Kaufman JS, Reda DJ, Fye CL, et al. Subcutaneous compared with intravenous epoetin in patients receiving hemodialysis. Department of Veterans Affairs cooperative study group on erythropoietin in hemodialysis patients. *N Engl J Med* 1998; 339:578-83.

Levey AS, Inker LA, Matsushita K, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis* 2014; 64(6):821-35.

Levin NW, Fishbane S, Cañedo FV, et al. Intravenous methoxy polyethylene glycol-epoetin beta for haemoglobin control in patients with chronic kidney disease who are on dialysis: a randomised non-inferiority trial (MAXIMA). *Lancet* 2007;370:1415-21. Erratum in: *Lancet* 2008; 371:386.

Mantel N, Haenszel W. Statistical aspects of analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719–48.

Paganini EP, Eschbach JW, Lazarus JM, et al. Intravenous versus subcutaneous dosing of epoetin alfa in hemodialysis patients. *Am J Kidney Dis* 1995;6:331-40.

Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharm Stat* 2013;12(6):337-347.

Vega A, Abad S, Verdalles U, et al. Dose equivalence between continuous erythropoietin receptor activator (CERA), darbepoetin, and epoetin in patients with advanced chronic kidney disease. *Hippokratia* 2014;18:315-18.

Wiens BL, Rosenkranz GK. Missing data in noninferiority studies. *Stat Biopharm Res* 2013;5:383-93.

Wright DG, Wright EC, Narva AS, et al. Association of erythropoietin dose and route of administration with clinical outcomes for patients on hemodialysis in the United States. *Clin J Am Soc Nephrol* 2015;10:1822-30.

15 ATTACHMENTS

Approval Sheet

Product:	Statistical Analysis Plan
Protocol Number:	AKB-6548-CI-0014 and -0015
SAP Version:	1.0

The individuals signing below have reviewed and approve this statistical analysis plan.

[redacted]

Jan 29, 2020
Date

Senior Director, Biostatistics

[redacted]

JAN 29, 2020
Date

Senior Medical Director

Statistical Analysis Plan

MAJOR ADVERSE CARDIOVASCULAR EVENTS (MACE) IN SUBJECTS WITH NON-DIALYSIS-DEPENDENT CHRONIC KIDNEY DISEASE (NDD-CKD) (PROTOCOLS AKB-6548-CI-0014 AND AKB-6548-CI-0015: PRO$_2$TECT)

AND

MAJOR ADVERSE CARDIOVASCULAR EVENTS (MACE) IN SUBJECTS WITH DIALYSIS-DEPENDENT CHRONIC KIDNEY DISEASE (DD-CKD) (PROTOCOLS AKB-6548-CI-0016 AND AKB-6548-CI-0017: INNO$_2$VATE)

Sponsored by: Akebia Therapeutics, Inc.

Version: 1.0

Date Approved: FINAL

Confidentiality Statement

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviation	**Definition**
CEC	Clinical Endpoint Committee
CHF	congestive heart failure
CKD	chronic kidney disease
CRP	C-reactive protein
CV	cardiovascular
DD	dialysis-dependent
DVT	deep vein thrombosis
EAS	Endpoint Adjudication System
EOT	end of treatment
ESA	erythropoietin-stimulating agent
EU	European Union
GSCD	global study completion date
HF	heart failure
HR	hazard ratio
IDMC	Independent Data Monitoring Committee
MACE	major adverse cardiovascular events
MAR	missing at random
MI	myocardial infarction
MNAR	missing not at random
NDD	non-dialysis dependent
NI	noninferiority
PE	pulmonary embolism
RMST	restricted mean survival time
ROW	rest of world
SAP	statistical analysis plan
SMQ	Standard MedDRA Queries
uACR	urine albumin creatinine ratio
US	United States

1 INTRODUCTION

This statistical analysis plan (SAP) describes the methods to be used to analyze the major adverse cardiovascular events (MACE) in Akebia's PRO$_2$TECT Protocols (AKB-6548-CI-0014 and AKB-6548-CI-0015) and INNO$_2$VATE Protocols (AKB-6548-CI-0016 and AKB-6548-CI-0017). CI-0014 and CI-0015 are studying vadadustat in subjects with non-dialysis dependent (NDD) chronic kidney disease (CKD) while CI-0016 and CI-0017 are studying vadadustat in dialysis-dependent (DD) CKD. These four trials, referred to hereafter as CI-0014, CI-0015, CI-0016, and CI-0017, constitute Akebia's Phase 3 program for vadadustat.

Note that while this SAP covers all four trials, the analyses will be performed by study pair, PRO$_2$TECT (CI-0014 and CI-0015) or INNO$_2$VATE (CI-0016 and CI-0017), in particular at such time as the specified target is met in each respective pair, independent of the progress in the other pair of studies. Thus, the analyses in this SAP will not necessarily be conducted on the same calendar date.

In addition to the plans for analysis of MACE, this SAP also describes plans for analysis of other cardiovascular and thrombotic events.

(For simplicity, sometimes MACE refers to the singular and sometimes the plural. The context is always clear.)

Protocol CI-0014 is a "correction study". CI-0014 enrolls subjects not on an erythropoietin-stimulating agent (ESA).

Protocols CI-0015 and CI-0017, the "conversion studies", which are designated here as CI-0015/17, enroll subjects on any dose of ESA.

Protocol CI-0016 is a "correction/conversion study", enrolling incident dialysis subjects who may be on an ESA or not.

For each safety parameter described in this SAP, the outcomes of most interest are estimates that combine data from each pair CI-0014/15 and CI-0016/17 calculated by considering as a stratum each component study of the pair. In addition, estimates will be presented for each study separately.

The protocols for the four studies, as well as the general SAPs for the PRO$_2$TECT and INNO$_2$VATE Protocols, provide information about the plans for the studies and their analyses exclusive of the MACE analyses.

This SAP contains language and programming code that specifies the intent of each analysis. The sponsor will finalize and sign this document prior to locking the database and unblinding. Many of the analyses described in this SAP are quite complex. A blinded team of clinicians and statisticians will review the data carefully to develop conventions and analytic methods not anticipated in writing this SAP.

2 STUDY DESCRIPTION

CI-0014 and CI-0015 are studying vadadustat in subjects with NDD-CKD while CI-0016 and CI-0017 are studying vadadustat in DD-CKD. Darbepoetin alfa is the control drug in all four studies.

The Phase 3 program is designed so that meta-analyses of the PRO_2TECT and $INNO_2$VATE pairs of studies will have a sufficient number of independently adjudicated endpoints to allow a meaningful comparison of vadadustat and darbepoetin alfa with respect to MACE in each pair of trials. The PRO_2TECT pair of trials will continue until at least 631 subjects experience an adjudicated MACE in the pair of trials. Similarly, the $INNO_2$VATE pair of trials will continue until at least 631 subjects experience an adjudicated MACE in that pair. The individual trials in each pair will continue after the planned number of MACE have occurred in order to ensure that each subject is followed the minimum amount of time specified in the individual protocols.

2.1 Study Periods

2.1.1 The Correction (CI-0014) and Correction/Conversion Study (CI-0016)

Following randomization, CI-0014 and CI-0016 will have five periods:

- Correction
  - CI-0014: Correction Period (Weeks 0-23): initial period on study medication for correction of Hb
  - CI-0016: Correction and Conversion Period (Weeks 0-23): initial period on study medication for correction of Hb or conversion to study treatment
- Maintenance Period (Weeks 24-52): this period will be divided into weeks 24-36 (primary efficacy evaluation period) and Weeks 40-52 (secondary efficacy evaluation period)
- Long-Term Treatment Period (Weeks 53-End of Treatment [EOT]): continued study medication to assess long-term safety
- Follow-Up (EOT + 4 weeks): post-treatment visit (either in person or by telephone) for safety
- Continued MACE Follow-up: each subject, except those who are lost to follow-up or who withdraw consent, will be followed for MACE until the end of the study.

2.1.2 The Conversion Studies (CI-0015 and CI-0017)

Following randomization, each conversion study will have four periods:

- Conversion and Maintenance Period (Weeks 0-52): conversion to study medication for maintaining Hb (Weeks 0-23), primary efficacy evaluation (Weeks 24-36), and secondary efficacy evaluation (Weeks 40-52)
- Long-term Treatment Period (Week 53-EOT): continued study medication to assess long- term safety
- Follow-up Period (EOT + 4 weeks): post-treatment visit (either in person or by telephone) for safety
- Continued MACE Follow-up: each subject, except those who are lost to follow-up or who withdraw consent, will be followed for MACE until the end of the study.

Each pair of trials (CI-0014/15 and CI-0016/17) will continue until at least 631 events occur in the respective pair.

2.2 Method of Assigning Subjects to Treatment Groups

Subjects will be randomized 1:1 using permuted block randomization. Randomization will be stratified in each of the four studies by the following:

- Geographic region
  - United States (US)
  - European Union (EU)
  - Rest of World (ROW)
- New York Heart Association congestive heart failure (CHF) class
  - Class 0 or I
  - Class II or III
- Baseline Hb (g/dL)
  - <9.5 versus ≥9.5 for CI-0014 and 0016
  - <10 versus ≥10 for CI-0015 and 0017

2.3 Primary Safety Outcome: Major Adverse Cardiovascular Event (MACE)

The primary safety outcome in each study is MACE, defined as all-cause mortality, nonfatal myocardial infarction (MI), or nonfatal stroke, occurring at any time on or following first dose date and prior to each subject's end of study (EOS) date.

2.4 Blinding

All four studies are randomized open-label trials. An interactive web response system governs treatment assignment. Investigators are not aware of which treatment will be assigned next.

The studies involve blinded adjudication of MACE, the use of an unblinded independent data monitoring committee (IDMC), and an identical schedule of visits, procedures, and assessments for both treatment groups in order to reduce the potential for bias.

The protocols discuss the method of guiding dose adjustment for vadadustat and darbepoetin alfa.

As further mechanisms to reduce potential execution bias in the trials, special steps have been taken to restrict access to the study data (see the Blinding Procedures and Oversight Plan).

2.5 Sample Size

The primary safety endpoint is the time from first dose date to the first adjudicated MACE. This section describes the justification for the sample sizes of the four studies.

2.5.1 The PRO$_2$TECT studies (CI-0014 and CI-0015)

The primary safety analysis will be based upon all first events that accrue over the two PRO$_2$TECT trials. If the noninferiority (NI) margin is 1.25 and the event rate is the same in the two treatment groups (i.e., the hazard ratio (HR) is 1.0), then 631 events are needed to establish NI with 80% power; 631 events yields more than 90% power to exclude an NI margin of 1.30. If the HR is 0.95 favoring vadadustat, the power is above 90% when the NI margin is 1.25.

Justification for the NI margin of 1.25 for FDA decision making and 1.30 for EMA decision making is provided in Appendix C.

An annual MACE rate of 10% is anticipated. This rate is based on a comprehensive review of available epidemiology and prospective clinical studies in NDD-CKD. Therefore, a projected total of 3700 subjects will be enrolled in studies CI-0014 and CI-0015. With 1850 subjects planned per treatment group in studies CI-0014 and CI-0015 and a planned accrual period of approximately 20 months, and up to 36 months of follow-up (with expected mean follow-up about 2 years), the number of MACE in each study will be a function of the actual pattern and size of enrollment as well as the duration of follow-up.

2.5.2 The INNO$_2$VATE studies (CI-0016 and CI-0017)

The primary safety analysis will be based upon all first events that accrue over two INNO$_2$VATE studies (CI-0016 and CI-0017). As described in the previous section, to have 80% power to establish NI if the NI margin is 1.25 and the event rate is the same in the two treatment groups (i.e., the HR=1.0), 631 events are needed; this number of events yields more than 90%

power to exclude an NI margin of 1.30. If the HR is 0.95 favoring vadadustat, the power is above 90% when the NI margin is 1.25.

An annual MACE rate of 12% is anticipated. This rate is based on a comprehensive review of available epidemiology and prospective clinical studies in DD-CKD. Therefore, an estimated total of 3700 subjects will be enrolled in studies CI-0016 and CI-0017. With 1850 subjects planned per treatment group enrolled in study CI-0016 and CI-0017, a planned accrual period of approximately 20 months, and up to 36 months of follow-up (with expected mean follow-up about 2 years), the number of MACE in each study will be a function of the actual pattern and size of enrollment as well as the duration of follow-up.

3 SAFETY OUTCOMES

This section defines the outcomes to be used for the analyses of MACE and related events.

3.1 Adjudicated Outcomes

Study investigators will solicit potential cardiovascular events at each study visit. Each of these potential events will be recorded on the eCRF. Potential cardiovascular events will also be identified through periodic review of Standard MedDRA Queries (SMQs) of adverse events and serious adverse events irrespective of the source of identification. An Endpoint Management Plan will provide details describing the SMQs terms and processes related to identifying these potential events. Events not reported by study investigators identified through SMQ search or CEC will be manually entered into EAS (Endpoint Adjudication System) database.

The Clinical Endpoint Committee (CEC) adjudicates all potential cardiovascular events. The formal cardiovascular safety analyses will only include CEC-adjudicated events (i.e., events confirmed by the CEC).

The CEC adjudicates the following events CEC Manual of Operations (MOP) outlines the program endpoint definitions and requirements for committee adjudication.

- All-cause mortality (The CEC will attribute cause of death to responsible underlying disease cause).
  - Non-cardiovascular death
  - Unknown death
  - Cardiovascular death
    - Sudden death
    - Non-sudden cardiovascular death

  - Fatal or nonfatal MI
  - Fatal or nonfatal stroke
- Thromboembolic event
  - Arterial thrombosis
  - Deep vein thrombosis (DVT)
  - Pulmonary embolism (PE)
  - Vascular access thrombosis
- Hospitalization for heart failure (HF)

3.2 Time to Adjudicated MACE: Primary Safety Outcome

An adjudicated MACE, the primary safety outcome, is defined as any death, CEC-confirmed nonfatal MI, or CEC-confirmed nonfatal stroke occurring between first dose date and each subject's EOS date. The following subsections describe variations on, and components of, the primary safety outcome.

3.2.1 Potential MACE: Descriptive Safety Outcome

All potential MACE are all cases adjudicated as MACE by the CEC (i.e., primary safety outcome events confirmed by the CEC) plus those potential MACE cases reported by the investigator, submitted for adjudication, but then adjudicated as non-MACE (i.e., events refuted by the CEC). All deaths are included in this descriptive safety outcome.

3.2.2 Time to MACE Within 4 Weeks of End of Treatment

Only a MACE that occurs up to and within 28 days inclusive after permanent discontinuation of study treatment (End of Treatment, EOT) will be included as a MACE Within 4 weeks of End of Treatment.

3.2.3 Individual CEC-Adjudicated Components of MACE: Descriptive Safety Outcome

The components of CEC-Adjudicated MACE are as follows:

- All-cause mortality
- CEC-confirmed nonfatal MI
- CEC-confirmed nonfatal stroke

The total number of these outcomes equals the total number of CEC-adjudicated MACE. They will be reported simply to help interpret the primary MACE outcome. No formal analysis of these outcomes is planned.

3.2.4 Time to Components of Adjudicated MACE

The following CEC-confirmed MACE will be reported and analyzed as time-to-first event. As with the primary safety outcome, the CEC adjudication of cause does not exclude any death from all-cause mortality.:

- All-cause mortality: death from cardiovascular (CV), non-CV, and unknown causes.
  - Cardiovascular death
    - Fatal MI
    - Fatal stroke
    - Sudden death
  - Non-cardiovascular death
- Non-fatal MI
- Non-fatal stroke

3.3 Time to Expanded MACE

The studies have several outcomes that expand the definition of MACE. Each of these outcomes includes MACE (all-cause mortality, nonfatal MI and nonfatal stroke) plus one or more other type of event. The events are as follows:

- MACE plus thromboembolic event
- MACE plus thromboembolic event excluding vascular access thrombosis
- MACE plus hospitalization for heart failure (HF)
- MACE plus hospitalization for HF or thromboembolic event
- MACE plus hospitalization for HF or thromboembolic event excluding vascular access thrombosis

3.4 Time to cardiovascular death, nonfatal MI, or nonfatal stroke

This alternate MACE definition is the same as the primary safety outcome, except includes only deaths positively adjudicated by the CEC as cardiovascular (i.e., only CEC-confirmed CV deaths). This excludes non-cardiovascular deaths as well as unknown deaths.

4 STATISTICAL ANALYSIS

For each subject in the four studies, duration of follow-up will be calculated as the number of days between date of first dose date and last contact date during the study plus one. For each outcome, the number of events that occur during that period may be included in the analysis. The event rate will be calculated as the number events divided by the duration of at-risk follow-up in

person-years. Survival analyses will include only the first event for each subject in that period. For each type of event, the event rate will be summarized by treatment group for each study and for each pair of studies (CI-0014/15 and CI-0016/17).

For each type of event that includes all-cause death (including the primary endpoint), the method of analysis for each pair of studies will be a Cox regression with the randomized treatment assignment as a predictor. The models will be stratified by study. Subjects without an event will be censored at their last study contact (visit or telephone contact) at which all components of the outcome were assessed. Deaths that occur between withdrawal of consent or loss to follow up and the global study completion date (GSCD) and that are identified, prior to database lock, through public records will be included in the analyses as an event. The models will use other covariates as well, as described below.

For all other events, analyses will use competing risk methods. These methods produce valid estimates of the marginal probability of the event in case other events preclude observation of the event of interest. For each pair of studies, the models include the randomized treatment assignment as a predictor and, should a competing event not occur, subjects will be censored on their last study contact (visit or telephone contact) at which all components of the outcome were assessed. The models will use other covariates as well, as described below. Models for the overall analysis of PRO$_2$TECT and of INNO$_2$VATE will also stratify by study. Appendix A provides SAS code for the competing risk analysis.

The studies will enroll subjects from many centers; however, most centers will have small sample sizes. Therefore, as described in the general SAPs for PRO$_2$TECT and INNO$_2$VATE, no center-specific analyses are planned.

4.1 Analysis Population

The analyses will be based on a Safety Population, which includes randomized subjects who receive at least one dose of study treatment and which classifies subjects by the more frequently received drug.

4.2 Global Study Completion Date

The sponsor will notify the sites approximately three months prior to the global study completion date (GSCD) for each study so that they can schedule visits with all subjects who are alive and have not withdrawn consent. At that visit the investigators should try to assess whether the subjects have had any MACE, Hospitalized for HF or Thromboembolic event. At a minimum the investigators will attempt to contact each subject to determine whether that subject is still alive. Additionally, investigators will request that subjects who continued study treatment until this visit come in for another follow-up visit four weeks later.

Operationally this plan is designed to capture all MACE, hospitalized for HF and Thromboembolic events that occurred in subjects still being followed and with onset within three

months of GSCD in the clinical database. Should subjects who remained on study treatment until that visit subsequently experience MACE, Thromboembolic events or hospitalized for HF between that visit and four weeks later, these events may also be captured.

Analytically, any outcome event with onset on or before the last contact date will be included in the primary safety outcome analysis, and all subjects without events will be censored at the last contact date.

The earliest GSCD of a pair of studies will be a date that ensures each pair of studies will have the planned number of safety outcome events (i.e., 631) and each study in each pair has the protocol-required amount of safety follow-up.

4.3 Reference dates

This SAP uses the following terms to refer to dates used to calculate time.

Date of last visit for MACE and other adjudicated events is the date of the last visit when all events could be assessed. This visit may be conducted in person at the site or by telephone.

Date of last contact is the last date when the site had any contact at all with the subject.

4.4 Primary MACE Analysis

The endpoint for the primary analysis is defined as the first occurrence of all-cause mortality, adjudicated nonfatal MI, or adjudicated nonfatal stroke after taking the first dose of investigational drug. Treatment comparisons will be based on the estimated hazard ratio and its 95% CI from the Cox model described below. As specified earlier, there will be a primary MACE analysis for each pair of trials, one for the CI-0014 and CI-0015 NDD-CKD trials and one for the CI-0016 and CI-0017 DD-CKD trials.

Counts and proportions of patients who experience a primary endpoint event will be calculated for each treatment arm, as well as the proportion difference (95% CI) in proportions between the groups. Mantel-Haenszel confidence interval will be reported for the proportion difference. In addition to event rate, person-years of follow-up for the primary endpoint and the incidence rate, calculated by dividing the number of patients who developed the event during the study period by the event-specific person-years of at-risk follow-up, will be provided.

For each subject who has experienced a MACE, the time to the first adjudicated MACE is defined as the date of the first MACE – date of first dose +1.

Subjects who have not experienced a MACE will be censored at the date of the last study contact (visit or telephone contact) at which all components of MACE were assessed.

For each of the four studies and for each pair of studies, Kaplan-Meier curves for the time to first adjudicated MACE will be generated.

For each pair of studies, analysis of time to first MACE will be based on a stratified Cox regression model with study as a stratification factor. The model will also include covariates of baseline Hb, randomization strata of region (US; EU; ROW) and NYHA (0 or I; II or III), sex (male; female), age (> 65; ≤65), race (white or non-white), preexisting cardiovascular disease as defined in Section 5.1 (yes/no), and diabetes mellitus (yes/no). The following kernel of SAS code (details in Appendix A) will be used for PRO$_2$TECT (for INNO$_2$VATE, the study strata will be replaced with CI-0016/17):

```
proc phreg;
    strata /*study pair*/ CI-0014/15;
    model time to first MACE * censor = trtgrp
    /*covariates*/ age sex race pre-existing
    cardiovascular disease DM region NYHA
    Hb_base;
run;
```

The hazard ratio (vadadustat/darbepoetin alfa) and its 95% confidence interval will be presented for each study and each pair of studies. For a given pair of studies (PRO$_2$TECT and INNO$_2$VATE), Vadadustat will be considered to have non-inferior risk of MACE relative to darbepoetin alfa if the upper limit of this 95% CI of the pair is less than 1.25 (for FDA assessment) or 1.30 (for EMA assessment).

Vadadustat will be considered be superior to darbepoetin alfa with respect to the MACE outcome in a particular disease population if the upper limit of this 95% CI of the pair of studies is less than 1.

The stratified Cox regression model (primary safety analysis) allows the underlying hazard to vary by stratum (i.e. by study) but it makes the assumption that the effects of covariates are proportional over time within each stratum. The proportionality assumption will be assessed graphically and visually. If the assumption is violated, alternative method may be explored.

4.5 Secondary MACE endpoints: Definition and Analysis

The study has the below key secondary MACE endpoints which will be analyzed formally only if the primary analysis meets the noninferiority margin. A hierarchical testing scheme will be used to correct for the multiplicity of these endpoints. These endpoints are prioritized according to the following ordering:

- MACE plus hospitalization for HF or thromboembolic event excluding vascular access thrombosis
- Cardiovascular death, non-fatal MI or non-fatal stroke
- Cardiovascular death
- All cause death

For each of these secondary MACE endpoints, the hazard ratio (vadadustat/darbepoetin alfa) and its 95% confidence interval will be presented for each study and each pair of studies by the similar approach described for the primary MACE analysis. Kaplan-Meier or cumulative incidence curves will be generated too.

4.6 Analyses of Secondary and Other Cardiovascular Safety Outcomes

The following table summarize the cardiovascular safety outcomes related to MACE, the associated definitions, and the model that will be applied to each outcome.

Adjudicated Outcome	Definition of Outcome First bullet: time to event for subjects with a MACE Second bullet: time to censoring for subject who do not experience a MACE	Model
Time to MACE within 4 weeks of End of Treatment	• Date of the first MACE occurring within 4 weeks of EOT – date of first dose +1 • Date of last study visit (including telephone contact) within 4 weeks of EOT where all the components of MACE were assessed – date first dose +1	Cox regression model with same stratification factors and covariates as primary MACE analysis
Individual CEC-adjudicated components of MACE	The frequency of each MACE component (all-cause mortality, nonfatal myocardial infarction, and nonfatal stroke) will be shown. For each subject, the total number of these outcomes (i.e., total number of MACE) will be calculated, and descriptive statistics for total number of MACE will be shown.	Not applicable
Time to components of adjudicated MACE	• Date of the first MACE components – date of first dose +1 • Date of last study visit (included telephone contact) when where all the components of MACE were assessed – date of first dose +1	Cox regression model or competing risk analysis with same stratification factors and covariates as primary MACE analysis
Time to expanded MACE	• Date of the first MACE (expanded definition) – date of first dose +1 • Date of last study visit (included telephone contact) when where all the components of MACE (expanded definition) were assessed – date of first dose +1	Cox regression model with same stratification factors and covariates as primary MACE analysis
Time to Cardiovascular death, non-fatal MI or non-fatal stroke	• Date of first MACE – date of first dose +1 • Subject's date of last contact – date of first dose +1	Competing risk analysis with same stratification factors and covariates as primary MACE analysis. Non-cardiovascular and unknown deaths are competing events.
All-cause death	• Date of death – date of first dose +1 • Subject's date of last contact – date of first dose +1	Cox regression model with same stratification factors and covariates as primary MACE analysis

Adjudicated Outcome	Definition of Outcome First bullet: time to event for subjects with a MACE Second bullet: time to censoring for subject who do not experience a MACE	Model
Cardiovascular death	• Date of cardiovascular death – date of first dose +1 • Subject's date of last contact – date of first dose +1	Competing risk analysis with same stratification factors and covariates as primary MACE analysis. Non-cardiovascular and unknown deaths are competing events.
Non-cardiovascular death	• Date of non-cardiovascular death – date of first dose +1 • Subject's date of last contact – date of first dose +1	Competing risk analysis with same stratification factors and covariates as primary MACE analysis. Cardiovascular and unknown deaths are competing events.
Time to any thromboembolic event	• Date of first occurrence of adjudicated arterial thrombosis, DVT, PE, or vascular access thrombosis – date of first dose +1 • Date of last study visit (including telephone contact) where such events were assessed – date of first dose +1	Competing risk analysis with same stratification factors and covariates as primary MACE analysis. All –cause death is a competing event.
Time to arterial thrombosis, DVT, or PE	• Date of first occurrence of adjudicated arterial thrombosis, DVT, or PE thrombosis – date of first dose +1 • Date of last study visit (including telephone contact) where such events were assessed – date of first dose +1	
Time to venous thromboembolic event (DVT or PE)	• Date of first occurrence of adjudicated DVT or PE – date of first dose +1 • Date of last study visit (including telephone contact) where such events were assessed – date of first dose + 1	
Hospitalization for heart failure	• Date of first occurrence of adjudicated hospitalization for heart failure – date of first dose +1 • Date of last study visit (including telephone contact) where such events were assessed – date of first dose + 1	

The following outcomes will be summarized as proportions and are purely descriptive as they do preserve the randomized groups. That is, only subjects with CEC-confirmed MI or strokes are included in these calculations, respectively.

30-day MI fatality	The proportion is calculated as the number of deaths occurring within 30 days of an CEC-confirmed MI divided by the number of subjects who had an CEC-confirmed MI.
30-day stroke fatality	The proportion is calculated as the number of deaths occurring within 30 days of an CEC-confirmed MI divided by the number of subjects who had an CEC-confirmed MI.

5 SUBGROUPS

The effects of vadadustat compared to darbepoetin alfa on the incidence of the primary safety outcome will be examined across the following subgroups. All subgroup analyses are regarded as exploratory.

5.1 Prespecified Subgroups

The main MACE analysis will be presented separately within the following subgroups. In each case, if the subgroup is a stratification factor or a covariate (indicated in italics), the model will be the same as the overall model with the deletion of the respective stratification factor or covariate. Subgroup analyses will be performed only if the total number of outcomes in a stratum, combined over the two treatment groups, is at least 100.

Randomization stratification factors

- Hb stratification level at baseline
  - PRO_2TECT
    - Low level group (<9.5 g/dL for CI-0014; <10.0 g/dL for CI-0015)
    - High level group (≥9.5 g/dL for CI-0014; ≥10.0 g/dL for CI-0015)
  - $INNO_2VATE$
    - Low level group (<9.5 g/dL for CI-0016; <10.0 g/dL for CI-0017)
    - High level group (≥9.5 g/dL for CI-0016; ≥10.0 g/dL for CI-0017)
- Region
  - US
  - EU
  - ROW
- NYHA CHF stratification level
  - 0 and 1

  - 2 and 3
- Target Hb level: These targets are completely confounded with region. The 10-11 g/dL target consists of subjects from the US while the 10-12 g/dL target applies to the EU and the ROW.
  - 10-11 g/dL
  - 10-12 g/dL

Demographics and medical history

- Age
  - <65 years
  - ≥65 years
- Sex
  - Male
  - Female
- Ethnicity
  - Hispanic
  - Non-Hispanic
- Race
  - White
  - All others
- Diabetes mellitus
  - No diabetes mellitus
  - Diabetes mellitus
- History of cardiovascular disease (defined as medical history of coronary artery disease, myocardial infarction, stroke, or heart failure as captured by four separate questions on the Case Report Form)
  - Yes
  - No
- For CI-0016 and CI-0017 only, type of dialysis

- Hemodialysis
- Peritoneal dialysis

Medications (CI-0015 and CI-0017 only)

- Baseline ESA dose (See Locatelli, 2004 for justification of the choice of cut-off values)
  - <90 U/kg/week
  - ≥90 U kg/week
  - >300 U/kg/week

Baseline laboratory measurements

- For CI-0014 and CI-0015 only, urine albumin creatinine ratio (uACR) (See Gansevoort, 2011 for a justification of the choice of cut-off values)
  - <300 mg/g
  - ≥300 mg/g
- For CI-0014 and CI-0015 only, estimated GFR (The threshold of 15 represents a value between CKD eGFR Categories 4 and 5)
  - <15 ml/min/1.73m^2 (CKD Stage 5)
  - ≥ 15 ml/min/1.73m^2 (CKD stage 4 or lower)
- C-reactive protein. In Q2 laboratories, which are the laboratories used in these studies, the normal range is 0-0.6mg/dL.
  - ≤0.6 mg/dL
  - >0.6 mg/dL
- Baseline transferrin saturation (TSAT)
  - CI-0014 and CI-0015
    - < median of CI-0014 and CI-0015 combined
    - ≥ median of CI-0014 and CI-0015 combined
  - CI-0016 and CI-0017
    - < median of CI-0016 and CI-0017 combined
    - ≥ median of CI-0016 and CI-0016 combined
- Baseline ferritin

  - CI-0014 and CI-0015
    - < median of CI-0014 and CI-0015 combined
    - ≥ median of CI-0014 and CI-0015 combined
  - CI-0016 and CI-0017
    - < median of CI-0016 and CI-0017 combined
    - ≥ median of CI-0016 and CI-0017 combined
- Incident Dialysis Patients
  - CI-0016 and CI-0017
    - All patients in CI-0016 plus patients who have ≤16 weeks of dialysis history from the date of informed consent in CI-0017
- Pooled analysis of patients with CKD G5 and incident dialysis patients (those who initiated dialysis within 16 weeks) after all studies indicted below are completed
  - <15 ml/min/1.73m^2 (CKD Stage 5) at baseline from CI-0014 and CI-0015
  - All patients in CI-0016 plus patients who have ≤16 weeks of dialysis history from the date of informed consent in CI-0017

In these summary statistics, only the primary analysis method will be presented. Hazard ratios and their 95% confidence intervals for subgroups will be displayed using forest plots.

5.2 Subgroup Analyses of Other Outcomes

Exploratory post-hoc analyses of other outcomes may be performed to gain insight into the effect of the drugs in other subgroups.

6 SUPPORTIVE ANALYSIS - RESTRICTED MEAN SURVIVAL TIME

Recent papers on time to event, especially regarding trials assessing cardiovascular safety of non-cardiovascular drugs [e.g., Uno, 2015, Zhao, 2016] have recommended use of the restricted mean survival time (RMST), rather than hazard ratio, to assess the effect of drugs. As a supportive analysis, we will compare the RMST in the two treatment arms for the PRO$_2$TECT and INNO$_2$VATE pairs of studies for the primary MACE event.

To specify RMST, one needs to select a time T. An event that occurs prior to time T "counts" in the analysis; events that occur after T do not enter the analysis. T will be selected as the 75th percentile of follow-up time.

7 DATA SAFETY MONITORING

An independent statistical analysis center performs analyses in support of the IDMC which meets approximately every 4 to 6 months throughout the course of the study.

The IDMC and the independent statistician will be the only people who will be unblinded to these MACE data.

The IDMC will also consider early termination if the quality of trial conduct is such that the trial will not be able to provide a timely and reliable answer to the questions it was designed to address.

8 HANDLING OF MISSING DATES

For all adjudicated events, the CEC-determined date of the event will be used in all analyses. In the unlikely event that this date is missing then the investigator-reported date will be used.

For an incomplete date of an endpoint event, imputation will be performed as outlined below:

- If the event date is completely missing, date is set to date of first dose.
- If year is present and month and day are missing or year and day are present and month is missing:
  - If year = year of first dose, then set month and day to month and day of first dose.
  - If year < year of first dose, then set month and day to December 31.
  - If year > year of first dose, then set month and day to January 1.
- If month and year are present and day is missing:
  - If year = year of first dose and
    - month = month of first dose, then set day to day of first dose date.
    - month < month of first dose, then set day to last day of month.
    - month > month of first dose, then set day to first day of month.
  - If year < year of first dose, then set day to last day of month.
  - If year > year of first dose, then set day to first day of month.
- For all other cases, set date to date of first dose.

If the above imputations produce an illogical date, a logical date will be selected by a statistical team blinded to everything but the dates.

For an incomplete death date, imputation will be performed using the same rule as described above.

9 SUBJECT DISPOSITION, TREATMENT ADHERENCE, AND TREATMENT EXPOSURE

The main SAPs for PRO$_2$TECT and INNO$_2$VATE summarize methods for describing subject disposition, adherence to treatment, and exposure to treatment.

10 REFERENCES

Anderson PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; 10:1100–1120.

Besarab A, Bolton WK, Browne JK, et al. The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. *N Engl J Med* 1998; 339:584-90.

Centers for Disease Control. Anemia or Iron Deficiency. *National Center for Health Statistics* [web]. March 17, 2017.Available at: http://www.cdc.gov/nchs/fastats/anemia.htm. Accessed August 13, 2015.

De Zeeuw D, Akizawa T, Audhya P, et al. Bardoxolone methyl in type 2 diabetes and stage 4 chronic kidney disease. *N Engl J Med* 2013; 369:2492-503.

EVOLVE trial investigators, Chertow GM, Block GA, et al. Effect of cinacalcet on cardiovascular disease in patients undergoing dialysis. *N Engl J Med* 2012; 367:2482-2494.

Fellstrom BC, Jardine AG, Schmieder RE, et al. Rosuvastatin and cardiovascular events in patients undergoing hemodialysis. *N Engl J Med* 2009; 360:1395-1407.

Fishbane S, Shiller B, Locatelli F, et al. Peginesatide in patients with anemia undergoing hemodialysis. *N Engl J Med* 2013; 368:307-319.

Fleiss JL, Bigger Jr. M, McDermott M, et al. Nonfatal myocardial infarction is, by itself, an inappropriate end point in clinical trials in cardiology. *Circulation* 1990; 81:684-5.

Food and Drug Administration. FDA drug safety communication: Modified dosing recommendations to improve the safe use of erythropoiesis-stimulating agents (ESAs) in chronic kidney disease. *Drug Safety and Availability* [web]. June 24, 2011. Available at: http://www.fda.gov/Drugs/DrugSafety/ucm259639.htm. Accessed September 1, 2011.

Food and Drug Administration. Providing clinical evidence of effectiveness for human drug and biological products. *FDA Guidance for Industry* [web]. May 1998. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatory%20Information%20/Guidances/UCM072008.pdf.

Ilya Lipkovich, Bohdana Ratitch, Michael O'Kelly. Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. Pharmaceutical Statistics 2016; 15: 216-229.

Little, RJ, Wang J, Sun X, et al. The treatment of missing data in a large cardiovascular clinical outcomes study. *Clinical Trials* 2016; 13:344-351.

Locatelli F, Covic A, Chazot C, et al. Hypertension and cardiovascular risk assessment in dialysis patients. *Nephro Dial Transplant* 2004; 19:1058-1068.

Macdougall IC, Provenzano R, Sharma A, et al. Peginesatide for anemia in patients with chronic kidney disease not receiving dialysis. *N Engl J Med* 2013; 368:320-332.

O'Kelly M, Ratitch B. Clinical trials with missing data: a guide for practitioners. West Sussex, United Kingdom: John Wiley & Sons, 2014.

Pfeffer MA, Burdmann EA, Chen CY, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *N Engl J Med* 2009; 361:2019-32.

Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley and Sons, 1987.

Singh AK, Szczech L, Tang KL, et al. Correction of anemia with epoetin alfa in chronic kidney disease. *N Engl J Med* 2006; 355:2085-98.

Swedberg K, Young JB, Anand ID, et al. Treatment of anemia with darbepoetin alfa in systolic heart failure. *N Engl J Med* 2013; 368:1210-1219.

Thamer M, Zhang Y, Kshirsager O, et al. Erythropoiesis-stimulating agent use among nondialysis--dependent CKD patients before and after the trial to reduce cardiovascular events with aranesp therapy (TREAT) using a large US health plan database. *Am J Kidney Dis* 2014; 64:706-713.

United States Renal Data Systems. 2014 Annual Data Report. *Annual Data Report archive*. 2014. Available at: https://www.usrds.org/2014/view/.

Uno H, Wittes J, Fu H, et al. Alternatives to hazard ratios for comparing efficacy or safety of therapies in noninferiority studies. *Ann Intern Med* 2015; 163:127-134.

Yuan Y. Sensitivity analysis in multiple imputation for missing data. Paper presented at: SAS Global Forum; 2014; Washington, DC.

Zhao L, Claggett B, Tian L, et al. On the restricted mean survival time curve in survival analysis. *Biometrics* 2016; 72:215-221.

APPENDIX A. SAS CODE

This below SAS codes are for competing risk analysis and restricted mean survival time analysis.

Competing Risk

```
proc phreg data=&INDAT.;
  class study TRT(ref="DARB") STRAT2_HF STRAT3_REG AGE SEX
        RACE HX_CV_ALL HX_DM;
  strata study;
  model T*Status(0)= HB_base STRAT2_HF STRAT3_REG TRT AGE SEX RACE
HX_CV_ALL HX_DM / eventcode=1 ties=Efron rl;
  Ods ouput ParameterEstimates = HR_EST;
Run;
```

Restricted Mean Survival Time (RMST)

```
proc lifetest data=&INDAT. plots=(rmst) rmst(tau=&fut);
   time SurvTime*Censor(1);
   strata study /group=TRT;
run;

where &fut is the 75th percentile of follow-up time.
```

APPENDIX B. JUSTIFICATION OF THE NONINFERIORITY MARGIN

This appendix provides the rationale for the noninferiority margin of 1.25 in MACE analysis for FDA decision making and for the noninferiority margin of 1.30 for EMA decision making.

C.1. Background

Many scientific questions are unanswered to date and there remains an unmet medical need for a new therapy for anemia in CKD. Among the unanswered questions are:

- Can Hgb be safely raised with a novel agent other than an ESA without increasing CV risk?
- Are hypoxia-inducible factor prolyl-hydroxylase inhibitors (HIF-PHIs) as safe as or safer than ESAs?

The primary safety outcome in each study is MACE, defined as all-cause mortality, nonfatal myocardial infarction (MI) or nonfatal stroke, occurring at any time following randomization and prior to each subject's end-of-study date. The endpoint for the primary analysis will be the first occurrence after randomization of all-cause mortality, adjudicated nonfatal MI, or adjudicated nonfatal stroke. Treatment comparisons will be based on the estimated hazard ratio (vadadustat/darbepoetin alfa) and its 95% CI from a Cox proportional hazards model for the pair of NDD studies and for the pair of DD studies.

Vadadustat will be considered to have non-inferior risk of MACE relative to darbepoetin alfa if the upper limit of this 95% CI of each pair of studies is less than or equal to the prespecified noninferiority margin.

C.2. FDA Decision Making

C.2.1. Feasibility of and MACE Event Rates in Trials of Subjects with Anemia Secondary to CKD

While the number of adults diagnosed with CKD in the US is 3.9 million (CDC 2015), the number of individuals with anemia and NDD-CKD is significantly lower. According to recent claims data, 10% of NDD-CKD stage 3 patients and 20% of NDD-CKD stage 4 patients have Hgb<10 g/dL (Thamer 2014). Moreover, the number of patients being treated for anemia continues to decline. This is consistent with recent claims data demonstrating that 11% of NDD-CKD stage 3 patients and 27% of NDD-CKD stage 4 patients are being treated with an ESA (Thamer 2014). This low utilization of ESA reflects changes in practice patterns as a result of TREAT (Pfeffer 2009) as well as the change in the ESA labels in the US. As a result, the impact on the feasibility of successfully executing a clinical trial of anemia in CKD must be considered with these changing practices. In fact, recent trials have taken longer to execute than did trials in the past. For example, the CHOIR (Singh 2006) and PEARL (Macdougall 2013) trials had broad inclusion/exclusion (I/E) criteria (e.g., Hgb<11.0 g/dL) and were able to recruit patients at a rate of 0.30-0.40 patients per site per month (p/s/m). In contrast, more recent trials requiring Hgb inclusion of <10.0 g/dL have had recruitment rates of 0.10-0.15 p/s/m (ClinicalTrials.gov NCT00922587 and NCT01887600). Put into practical terms, a trial needing 2000 CKD patients with Hgb<10.0 g/dL and recruiting at a rate of 0.10 p/s/m would require nearly 850 sites to recruit over 24 months. This is impractical given the limited number of nephrology

investigators; utilizing so many sites risks making a trial unfeasible. As a consequence, future trials require an extended timeline for recruitment, leading to a delay in the availability of new therapies for patients with CKD. In addition, these delays limit the potential clinical applicability as patient populations and practice patterns evolve.

Of the nearly 500,000 patients with dialysis-dependent CKD (DD-CKD) in the US, more than 80% require therapy for anemia and would potentially qualify for inclusion into a trial. However, addition of specific I/E criteria, competition with other trials, and the limited number of nephrology research centers available to perform research again negatively impact execution of DD-CKD trials. For example, for an incident dialysis trial where only 0 or 1 new patient starts dialysis per month in a given research center, anticipated recruitment rate is <0.20 p/s/m. Thus, a trial requiring enrollment of 400 patients would need at least 80 sites enrolling for over 24 months, not including time for start-up.

In addition to the aforementioned considerations, the event rate must also be factored into the feasibility of a trial. In development of our trials, while we have assumed a 10% annualized MACE event rate in NDD-CKD and 12% in DD-CKD, we understand that this event rate may decrease over time with improved medical care. Our event rates are based on review of contemporary population level data (USRDS), recent CKD trials (BEACON [de Zeeuw 2013]), and anemia in CKD trials. For NDD-CKD, given the narrower inclusion criteria of Hgb<10 g/dL, we assumed most patients would have later stage CKD with MACE event rates similar to that in TREAT at 10% overall. For MACE in DD-CKD, we reviewed the event rates from contemporary population-level data (USRDS), recent ESRD trials, EVOLVE (Chertow 2012), AURORA (Fellstrom 2009), and anemia in ESRD trials EMERALD (Fishbane 2013) and NHT (Besarab 1998). Only EMERALD and EVOLVE had applicable information on MACE event rate. While EMERALD had a MACE event rate of ~14.5%, EVOLVE had an 11.5% rate and thus 12% was conservatively selected. While Akebia considered enrichment to reduce the overall sample size knowing the difficulty we may have enrolling these trials, the feasibility trade-offs ruled out this option. For example, RED-HF (Swedberg 2013), an anemia trial in patients with heart failure, recruited at <0.10 p/s/m and had to extend enrollment to nearly 6 years. Rather than enrich the Akebia trials, we have assumed moderate event rates with the knowledge that the trials may need to be extended if the event rates decrease over time. For example, if an annual event rate of 8% is observed rather than our estimated 10%, the NDD-CKD studies will need to be extended by approximately 6 months given our baseline assumptions.

Taken together, we have considered the prevalence of anemia in CKD, narrower I/E criteria, and lower event rates over time when determining a reasonable trial design and sample size for our development program.

C.2.2. Safety Plan for Development of a New Anemia Therapy in CKD

The question around how to develop a new therapy for anemia in CKD is challenging, as we need to ensure the product is both efficacious and safe in the setting of the clinical trial logistical limitations cited above. For a new therapy to be approved, it must not increase risk, and ideally would decrease CV risk, compared to the currently approved ESAs. However, the long timelines and increased cost of developing such products need to be weighed against the practicality of executing the development programs. Furthermore, the design of these anemia CV outcome trials must assess event rates in the setting of current ESA use, recognizing the difference in ESA

practice patterns since publication of CHOIR and TREAT and subsequent modification to the ESA labels (Thamer 2014). While we are familiar with the increased risk of stroke seen with targeting higher Hgb levels with ESAs in TREAT, the population- level risk associated with targeting lower Hgb levels and the use of lower doses of ESAs as currently prescribed remains unknown. Therefore, while a CV safety superiority study with vadadustat is desirable, the uncertainty about the CV event rate of ESAs prescribed under their current labeling requirements (e.g., dosing to avoid transfusions) make planning a well-powered superiority study challenging. Data from the TREAT trial provide some information on acceptable risk for a new therapy. The overall HR for MACE within TREAT when targeting a higher Hgb level compared to placebo was 1.10 (95% CI 0.97-1.24). However, it was the individual component of stroke that raised concerns with targeting higher Hgb with ESAs (HR=1.923, CI 1.379-2.681).

We anticipate that a new treatment for anemia in CKD should not meaningfully increase the risk of MACE relative to ESAs. The currently available trial data suggest that the HR for a new treatment compared to an ESA could be near or slightly below 1.0 with the upper limit of the two-sided 95% CI being no more than 1.25 (this upper limit in TREAT was 1.24). See Table C1 for examples of sample size calculations based on a variety of non-inferiority (NI) margins with similar assumptions. If the true HR is 0.95, the study will have roughly 90% power. The table includes a line for an NI margin of 1.3, the margin approved by the EMA and the margin typically used in trials of Type II diabetes.

Table C 1. Summary of statistical assessment of MACE safety endpoint

NI Margin	Hazard Ratio	Power	MACE Events (Per Indication)	NDD-CKD Sample Size (Event Rate 10%)	DD-CKD Sample Size (Event Rate 12%)	Total Sample Size
1.3	1.00	85%	522	2564	2162	4726
1.25	0.95	85%	477	2398	2020	4418
		90%	559	2810	2367	5177
1.25	1.00	80%	631	3100	2612	5712
		85%	721	3546	2988	6534
1.20	0.95	80%	576	2950	2500	5450
		85%	659	3314	2792	6106
1.20	1.00	85%	1080	5310	4476	9786

C.2.3. Conclusion

We have developed an evidence-based clinical development program designed to exclude a clinically meaningful increased risk associated with vadadustat compared to ESA. We believe the proposed event driven studies which utilize an NI margin of 1.25 for the upper limit of the two-sided 95% CI will allow for a rigorous analysis of the safety of vadadustat compared to the current standard of care (ESAs), while being logistically feasible.

An earlier version of this rationale for a non-inferiority margin of 1.25 for the MACE analysis was provided to the FDA in Akebia's Type C Meeting Information Package dated 31 Aug 2015. FDA agreed with the use of 1.25 as the non-inferiority margin in Meeting Minutes dated 1 Oct 2015 and 12 Jan 2016.

C.3. EMA Decision Making

In materials submitted to EMA/CHMP for Scientific Advice on Vadadustat (AKB-6548) in November 2015, Akebia proposed to assess MACE, time to first event, using Cox regression via a non-inferiority analysis; e.g., a margin of 1.3 for the 95% upper confidence interval for the hazard ratio (vadadustat/epoetin alfa) for the MACE safety endpoint in the pooled analysis of each pair of studies (CI-0016/17 for DD-CKD and CI-0014/15 for NDD-CKD).

It was calculated that 631 events will be required overall in a pair of studies to have >90% power to establish non-inferiority with a margin of 1.3 when evaluated with a 2- sided 95% confidence interval assuming no difference between the treatments. The planned sample size (totalling an estimated enrolment of 2600 in the Phase 3 DD-CKD program and 3100 in the Phase 3 NDD-CKD program, one-half of which will receive vadadustat) and planned duration (about 20 months recruitment and up to 36 months of treatment - average follow-up duration 1.8 years) of the Phase 3 studies will far exceed what is required to adequately assess efficacy and safety of a novel agent for the treatment of anaemia of CKD.

C.3.1. Rationale

The margin chosen for assessing MACE was proposed based upon a review of historical trials assessing cardiovascular outcomes associated with the use of ESAs in CKD (patients receiving haemodialysis and not receiving haemodialysis). In particular, randomized controlled trials were identified where an increased risk has been associated with ESA use. As the choice of a margin for a safety endpoint is inherently based upon a subjective assessment of risk benefit, these trials were selected due to their size, length, and degree of follow-up to provide insight into what levels of increased risk have been found to be perceived as unacceptable. With the exception of the peginesatide trials (see Table D2), these trials all compared targeting higher versus lower target haemoglobin levels. In the peginesatide trial, a similar target was used for the treatment arms being compared.

Table C 2. Overview of Cardiovascular Safety Outcomes for Erythropoiesis-Stimulating Agents in Chronic Kidney Disease

Trial	ESA	Design	CV Safety Assessment	Result
Normal Hematocrit Study	Epoetin alfa	1265 subjects with CKD receiving hemodialysis with a history of either chronic heart failure or ischemic heart disease; randomization to maintain a target HCT of 42 ± 3% ("normal" HCT) or to maintain a target HCT of 30 ±3% (low HCT)	Time to death or first nonfatal myocardial infarction	Hazard ratio for the composite endpoint of 1.3 (95% CI 0.9, 1.9) favoring low HCT treatment
Correction of Hemoglobin and Outcomes in Renal Insufficiency CHOIR	Epoetin alfa	1432 subjects with CKD not receiving dialysis; randomization to maintain a target HGB of either 13.5 g/dL (high HGB group) or 11.3 g/dL (low HGB group)	Time to the composite of death, non-fatal myocardial infarction, hospitalization for congestive heart failure, and stroke	Hazard ratio for the composite endpoint of 1.34 (95% CI 1.03, 1.74) favoring lower HGB target
CREATE	Erythro-poietin beta	603 subjects with CKD not receiving dialysis; randomization to randomized to a high hemoglobin target (13 to 15 g/dL) or a low hemoglobin target (10.5 to 11.5 g/dL); with rescue ESA therapy	Composite of eight cardiovascular events: sudden death, myocardial infarction, acute heart failure, stroke, transient ischemic attack, angina pectoris resulting in hospitalization for 24 hours or more or prolongation of hospitalization, complication of peripheral vascular disease (amputation or necrosis), or cardiac arrhythmia resulting in hospitalization for 24 hours or more.	Hazard ratio of 1.28 (95% CI, 0.88 to 1.89) favoring lower HGB target
TREAT	Darbe-poetin alfa	4038 subjects with CKD not receiving dialysis; randomized to a high hemoglobin target (13.5 g/dL) or a placebo; with rescue ESA therapy if hemoglobin level falls below 9 g/dl	CV composite endpoint and renal composite endpoint	Hazard ratio of 1.06 (95% CI-0.95, 1.19) for composite favoring placebo. Hazard ratio for stroke of 1.92 (95% CI 1.38, 2.68) favoring placebo.
AFX01-11/13	Pegines-atide	983 subjects with CKD not receiving dialysis; randomization to peginesatide or darbepoetin	Sponsor-defined CSE events: death, stroke, MI, congestive heart failure, unstable angina, and arrhythmia. MACE endpoint: death, stroke and MI.	Hazard ratios CSE 1.32 (90% CI 1.02, 1.72) MACE 1.28 (95% CI 0.84, 1.94)

Source: FDA Briefing Document for Peginesatide Injection Oncology Drugs Advisory Committee Meeting (December 7, 2011)

C.3.2. Conclusion

Thus, Akebia proposed the MACE hazard ratio (HR) for a new treatment compared to ESA to be equal to or less than 1.0 with the upper limit of the two-sided 95% confidence interval less than or equal to the noninferiority margin of 1.3 for the DD-CKD program (Studies AKB-6548-CI-0016 and AKB-6548-CI-0017), as well as for the NDD-CKD program (Studies AKB-6548-CI-0014 and AKB-6548-CI-0015).

The CHMP endorsed (response dated 28JAN2016) the plan to rule out an excess risk for MACE with a non-inferiority margin of 1.3. However, it was noted in the response that the final assessment of the MAA will not be merely based on just the efficacy from the primary outcome and the exclusion of harm based on this safety analysis, but on the overall benefit-risk assessment. CHMP agreed with Akebia's rationale that the point estimate of the HR for this product should be below 1 considering the ESA cardiovascular safety profile.

Approval Sheet

The individuals signing below have reviewed and approve this statistical analysis plan.

[redacted]

[redacted], Biostatistics

Jan 29, 2020

Date

[redacted]

JAN 29, 2020

Date