| | |
|---|---|
| **DOCUMENT:** | Statistical Analysis Plan |
| **PROTOCOL:** | ST10-01-304 |
| | A phase 3b, randomized, controlled, multicentre study with oral ferric maltol (Feraccru) or intravenous iron (ferric carboxy maltose; FCM), for the treatment of iron deficiency anaemia in subjects with inflammatory bowel disease |
| **SAP VERSION:** | 3.0 |
| **SAP DATE:** | 14 February 2019 |
| **PROTOCOL DATE:** | 10 July 2015, Version 1.0<br>24 August 2015, Version 2.0<br>12 February 2016, Version 3.0<br>08 August 2016, Version 4.0<br>25 August 2016, Version 4.2<br>05 January 2017, Version 5.0<br>11 July 2017, Version 6.0<br>30 April 2018, Version 7.0 |
| **SPONSOR:** | Shield TX (UK) Limited |
| **PREPARED BY:** | Clinipace Worldwide |
| **AUTHORS:** | Barbara Collins, BSc PhD<br>Senior Biostatistician<br>Clinipace Worldwide |

## APPROVAL SIGNATURES

| SIGNATURE: | DATE: |
|---|---|
| Dr Mark Sampson MBChB<br>Chief Medical Officer<br>Shield TX (UK) Ltd | 14 FEB 2019. |
| Roberta Vasari<br>Clinical Programme Manager<br>Shield TX (UK) Ltd | 14-Feb-2019 |
| Barbara Collins, BSc PhD<br>Snr Biostatistician, Clinipace | 14 - FEB - 2019 |
| Tonya Smoot, PhD<br>Director, Biostatistics US Clinipace | 19 FEB 2019 |

**TABLE OF CONTENTS**

**LIST OF ABBREVIATIONS ..... 5**

**1 INTRODUCTION ..... 6**

**2 STUDY OBJECTIVES, TREATMENTS, AND ENDPOINTS ..... 6**

2.1 STUDY OBJECTIVES ..... 6
*2.1.1 Primary Objective ..... 6*
*2.1.2 Secondary Objectives ..... 6*
2.2 TREATMENT GROUP COMPARISONS ..... 6
2.3 STUDY ENDPOINTS ..... 6
*2.3.1 Primary Efficacy Endpoint ..... 6*
*2.3.2 Secondary Efficacy Endpoints ..... 6*
*2.3.3 Pharmacoeconomic Endpoints ..... 7*
*2.3.4 Safety Endpoints ..... 8*

**3 STUDY DESIGN ..... 8**

3.1 OVERALL STUDY DESIGN ..... 8
3.2 SCHEDULE OF STUDY ASSESSMENTS ..... 8
*3.2.1 Screening Period (Visit 1) ..... 9*
*3.2.2 Treatment Period (Visits 2-7) ..... 9*
*3.2.3 Follow-up Period (Visit 8) ..... 9*

**TABLE 1. SCHEDULE OF ASSESSMENTS ..... 10**

**4 POWER AND SAMPLE SIZE CONSIDERATIONS ..... 11**

**5 ANALYSIS POPULATIONS ..... 11**

5.1 RANDOMIZED POPULATION ..... 11
5.2 SAFETY POPULATION ..... 11
5.3 INTENTION-TO-TREAT (ITT) POPULATION ..... 11
5.4 PER PROTOCOL (PP) POPULATION ..... 12

**6 CONSIDERATIONS FOR DATA ANALYSIS ..... 12**

6.1 GENERAL STATISTICAL CONSIDERATIONS ..... 12
6.2 PROGRAMMING ENVIRONMENT ..... 12
6.3 STRATA AND COVARIATES ..... 12
6.4 SUBGROUPS ..... 12
6.5 MULTIPLE COMPARISONS AND MULTIPLICITY ..... 12
6.6 SIGNIFICANCE LEVEL ..... 13
6.7 STATISTICAL NOTATION AND METHODOLOGY ..... 13

**7 DATA HANDLING METHODS ..... 13**

7.1 MISSING DATA ..... 13
*7.1.1 Date Values ..... 13*
*7.1.2 Non-Date Values ..... 14*
7.2 VISIT WINDOWS ..... 14
7.3 DATA DERIVATIONS AND DEFINITIONS ..... 15

**8 STUDY POPULATION ..... 16**

8.1 ANALYSIS POPULATIONS ..... 16
8.2 SUBJECT DISPOSITION ..... 16
8.3 INCLUSION/EXCLUSION CRITERIA ..... 16

8.4 PROTOCOL DEVIATIONS ....17
8.5 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS ....17
8.6 PRIOR AND CONCOMITANT MEDICATIONS ....18
8.7 MEDICAL HISTORY ....18
8.8 SUBJECT PROGRESS AND UNSCHEDULED VISITS ....18
8.9 STUDY DRUG DISPOSITION ....18

**9 EFFICACY ANALYSES ....19**

9.1 PRIMARY EFFICACY ANALYSES ....19
*9.1.1 Primary Analysis ....19*
*9.1.2 Sensitivity Analysis ....21*
9.1.2.1 Observed Cases (OC) Analysis ....21
9.1.2.2 Observed Cases (OC) Repeated Measures Analysis ....21
9.2 SECONDARY EFFICACY ANALYSES ....22
*9.2.1 Secondary Analysis ....22*
9.2.1.1 Key Secondary Endpoints ....22
9.2.1.2 Other Secondary Efficacy Endpoints ....24
9.2.1.3 Long-term Secondary Endpoints ....25

**10 PHARMACOECONOMIC ANALYSES ....26**

10.1 MEDICAL OUTCOMES STUDY 36-ITEM SHORT FORM (SF-36) ....26
10.2 CORRECTION AND MAINTENANCE OF IDA ....27

**11 SAFETY ANALYSES ....29**

11.1 ADVERSE EVENTS ....29
11.2 LABORATORY EVALUATIONS ....29
11.3 VITAL SIGNS ....30
11.4 PHYSICAL EXAMINATION ....30

**12 PRIMARY ANALYSIS ....30**

**13 FINAL ANALYSIS ....30**

**14 CHANGES FROM PROTOCOL ....30**

**15 REFERENCES ....31**

**16 TABLE OF CONTENTS FOR DATA DISPLAYS ....32**

16.1 TABLES ....32
16.2 FIGURES ....34
16.3 LISTINGS ....35

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AE | Adverse event |
| ANCOVA | Analysis of Covariance |
| ATC | Anatomical Therapeutic Chemical |
| bid | Two times daily |
| CD | Crohn's disease |
| CI | Confidence interval |
| CFB | Change from baseline |
| CRF | Case report form |
| CSR | Clinical Study Report |
| CTCAE | Common Toxicity Criteria for Adverse Events |
| FCM | Ferric carboxy maltose |
| GLM | Generalized Linear Model |
| GLMM | Generalized Linear Mixed Model |
| Hb | Haemoglobin |
| IDA | Iron deficiency anaemia |
| IBD | Inflammatory bowel disease |
| ITT | Intention-to-Treat |
| IV | Intravenous |
| LCL | Lower confidence limit |
| LOCF | Last Observation Carried Forward |
| LSM | Least squares mean |
| MCS | Mental Component Score |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MH | Mantel-Haenszel |
| OC | Observed Cases |
| PCS | Physical Component Score |
| PE | Pharmacoeconomic |
| PI | Prescribing Information |
| PP | Per Protocol |
| SAE | Serious adverse event |
| SAP | Statistical Analysis Plan |
| SD | Standard deviation |
| SF-36 | Medical outcomes study 36-item short form |
| TE | Treatment emergent |
| TEAE | Treatment-emergent adverse event |
| TESAE | Treatment-emergent serious adverse event |
| UC | Ulcerative colitis |
| WHO | World Health Organization |

# 1 INTRODUCTION

This document describes the statistical methods and data presentations to be used in the summary and analysis of safety, efficacy, and tolerability data from Protocol ST10-01-304. Background information is provided for the overall study design and objectives. The reader is referred to the study protocol and electronic case report forms (eCRFs) for details of study conduct and data collection.

# 2 STUDY OBJECTIVES, TREATMENTS, AND ENDPOINTS

## 2.1 Study Objectives

### 2.1.1 Primary Objective

To compare the efficacy of ferric maltol and intravenous iron (IVI, FCM) in the treatment and maintenance of iron deficiency anaemia in subjects with IBD.

### 2.1.2 Secondary Objectives

The secondary objectives of this study are:

- To evaluate the safety and tolerability of ferric maltol and intravenous iron (IVI, FCM) in subjects over a treatment duration of up to 52 weeks.
- To evaluate long-term healthcare resources utilized in the management of IDA in IBD.

## 2.2 Treatment Group Comparisons

Subjects will be randomized to one of the following two treatment arms:

- Oral ferric maltol, 30mg capsule twice daily (bid).
- Intravenous iron (IVI, FCM), administered as per the local summary of product characteristics.

Comparisons will be made between these two treatment arms.

## 2.3 Study Endpoints

### 2.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the proportion of subjects achieving either a ≥2g/dL increase in Haemoglobin (Hb) OR normalization of Hb (>12g/dL for women, >13g/dL for men) at Week 12.

### 2.3.2 Secondary Efficacy Endpoints

Secondary efficacy endpoints to be evaluated at Week 12 are:

Key secondary endpoints:

- Change in Hb concentration from baseline to Week 12
- Change in Hb concentration from baseline to Week 12 in subjects with a baseline Hb <9.5 g/dL

Other secondary endpoints:

- Proportion of subjects who experience a change from baseline in Hb concentration ≥1.0 g/dL at Week 12 (Hb rise responder)
- Proportion of subjects with baseline Hb <9.5g/dL that achieve an increase in Hb concentration of ≥1 g/dL at Week 12 (Hb rise responder)
- Proportion of subjects with Hb concentration within normal limits at Week 12 (Hb normal limit responder)
- Proportion of subjects with baseline Hb concentration <9.5 g/dL that is within normal limits at Week 12 (Hb normal limit responder)
- Change in Hb concentration from baseline to Week 4
- Change in Hb concentration from baseline to Week 4 in subjects with a baseline Hb <9.5 g/dL
- Proportion of subjects who experience a change from baseline in Hb concentration ≥2.0 g/dL at Week 12 (Hb rise responder)
- Proportion of subjects with baseline Hb <9.5g/dL that achieve an increase in Hb concentration of ≥2 g/dL at Week 12 (Hb rise responder)
- Proportion of subjects who experience a change from baseline in Hb concentration ≥1.0 g/dL at Week 4 (Hb rise responder)
- Proportion of subjects with baseline Hb <9.5g/dL that achieve an increase in Hb concentration of ≥1 g/dL at Week 4 (Hb rise responder)
- Proportion of subjects with Hb concentration within normal limits at Week 4 (Hb normal limit responder)
- Proportion of subjects with baseline Hb concentration <9.5 g/dL that is within normal limits at Week 4 (Hb normal limit responder)
- Proportion of subjects who experience a change from baseline in Hb concentration ≥2.0 g/dL at Week 4 (Hb rise responder)
- Proportion of subjects with baseline Hb <9.5g/dL that achieve an increase in Hb concentration of ≥2 g/dL at Week 4 (Hb rise responder)
- Long term efficacy endpoints i.e. proportion of subjects who are non-anaemic at 6 and 12 months (Hb normal limit responder); normalization of ferritin levels at 6 and 12 months

### 2.3.3 Pharmacoeconomic Endpoints

Pharmacoeconomic measures include the following:

- Number of hospital or clinic visits for administration of IV Iron.
- Proportion of subjects who restart IV iron during the study.
- Quality of life as assessed by the Medical Outcomes Study 36-item short form questionnaire.
- Correction and Maintenance of IDA
    - Repeat Hb or iron testing
    - Method of transportation to attend each visit.
    - The number of days off work or school/education to attend hospital or clinic visits for administration of IV Iron.
    - The time spent in the hospital or clinic receiving IV Iron.
    - The total dose of IV Iron administered over the study period.

### 2.3.4 Safety Endpoints

The safety endpoints are treatment-emergent Adverse Events (AEs), treatment-emergent Serious Adverse Events (SAEs), AEs leading to premature discontinuation from the study, and adherence to study medication.

# 3 STUDY DESIGN

## 3.1 Overall Study Design

This study will be a Phase 3b, multicentre, open-label, active-controlled, randomized, parallel group prospective study to compare the efficacy and safety of ferric maltol to IV Iron in subjects with inflammatory bowel disease. A total of 242 subjects from 35-45 European sites and around 18 United States sites will be randomized at a ratio of 1:1 to the two treatment arms of this study.

The randomization will be stratified according to screening Hb (<10g/dL or ≥ 10g/dL for Female; < 11g/dL or ≥ 11g/dL for Male) and by IBD subgroup (ulcerative colitis [UC] or Crohn's disease [CD]). Subjects who are re-screened will have their re-screen Hb substituted in lieu of screening Hb for purposes of determining the strata for randomization.

Protocol amendment 7.0 updated the study duration from 52 weeks to 12 weeks. For subjects who entered the study under the previous protocol for the 52 weeks open-label design and completed their study beyond the 12 weeks treatment period, the next scheduled visit will be their End of Study Visit. Otherwise, subjects after the eligibility screening period will enter a 12 weeks treatment period.

## 3.2 Schedule of Study Assessments

There will be three distinct periods in this trial: Screening, Treatment, and Post-

treatment safety follow-Up. Please refer to Table 1. Schedule of Assessments below for details.

### 3.2.1 Screening Period (Visit 1)

A subject will enter the Screening Period (up to 14 days) upon submission of written informed consent to participate in the study. The subject will then complete the remaining screening assessments and be evaluated for eligibility to the study. Subjects who satisfy all inclusion and exclusion criteria for the study will be scheduled for Visit 2.

### 3.2.2 Treatment Period (Visits 2-7)

The Treatment Period begins on the day for which the subject receives their first dose of a study drug (Visit 2/Day 0) and ends on the day corresponding with Visit 7 for those subjects enrolled prior to protocol amendment 7.0, and the day corresponding with Visit 4 for those subjects enrolled under protocol amendment 7.0. Eligible subjects will be randomized in a 1:1 ratio to receive treatment with either oral ferric maltol or intravenous iron.

Subjects randomized to the oral ferric maltol treatment arm will self-administer study medication twice daily; first thing in the morning at least 1 hour before food and concomitant medications, and last thing at night before bed at least 2 hours after food and concomitant medications. Capsules must be taken on an empty stomach with water only. There should be at least an 8-hour gap between doses. These subjects will maintain treatment with oral ferric maltol for the duration of the Treatment Period.

Subjects randomized to the intravenous iron treatment arm will receive IV Iron in accordance with the dosing instructions as per the local summary of product characteristics (SPC)/Prescribing Information (PI). Initial dosing will be as per local SPC/PI, with the number of IV Iron doses administered being calculated based on the subject's starting Hb and weight. For ongoing treatment decisions, ferritin will be measured at Visits 4 (Week 12) and Visits 5 and 6 (Week 24, and Week 36) for patients prior to protocol amendment 7.0. Subjects continuing after Week 12 who are iron deficient (ferritin below 100ng/mL) at any of Visits 4-6 will receive additional IV Iron doses according to the formula in the local SPC/PI.

### 3.2.3 Follow-up Period (Visit 8)

At the end of the Treatment Period, subjects will enter a 14-day Follow-up Period during which AEs and concomitant medications will be monitored. At the end of the 14-day Follow-up Period, subjects will have their Follow-up visit (Visit 8). Visit 8 will be conducted by telephone interview, unless the subject has an ongoing AE that requires physical examination or investigations for assessment/management.
If an AE is reported or is ongoing at the time of the Follow-up visit, or if a significant lab abnormality is identified at the last visit prior to the Follow-up visit, the Investigator must follow the event until resolution or until there is a satisfactory

explanation of the changes observed.

## Table 1. Schedule of Assessments

| | **Screening** | **Treatment** | | | | | | **[8]Follow-up** |
|---|---|---|---|---|---|---|---|---|
| **Duration** | **Up to 14 days** | **Week** | | | | | | **14 days** |
| | | **0** | **4** | **12** | **(24)[15]** | **(36)[16]** | **(52)[17]** | **(54)** |
| **[9]Visits** | **1** | **2** | **3** | **4** | **(5)** | **(6)** | **(7)** | **(8)** |
| Informed Consent | X | | | | | | | |
| Eligibility | X | X | | | | | | |
| Demographics | X | | | | | | | |
| Medical history | X | | | | | | | |
| Physical examination | X | | | | | | | |
| [1]Vital signs | X | | | | | | | |
| [2]Height and weight | X | X | X | X | (X) | (X) | (X) | |
| [3]Basic chemistry | X | | | | | | | |
| [4]Liver function tests | X | | | | | | | |
| [5]Haematology; Iron Markers | X | X | X | X | (X) | (X) | (X) | |
| C-reactive protein | | X | | | | | | |
| [13]Retained plasma and serum samples | | X | | X | | | | |
| [6]Randomization | | X | | | | | | |
| Collect IV Iron History/IV Iron dosing detail | X | X | X | X | (X) | (X) | (X) | |
| SCCAI or CDAI[12] | X[12] | | | X | (X) | (X) | (X) | |
| [10]Collect pharmacoeconomic (PE) and SF36 data | | X | X[11] | X | (X) | (X) | (X) | |
| [7]Pregnancy test | X | | | X | | | (X) | |
| Dispense/Administer study medication [14] | | X | | (X) | (X) | (X) | | |
| Compliance check | | | X | X | (X) | (X) | (X) | |
| Adverse events | | X | X | X | (X) | (X) | (X) | X |
| Prior/Concomitant medication | X | X | X | X | (X) | (X) | (X) | X |

1 Vital signs (blood pressure & heart rate) are taken after subject has been sitting for 5 minutes.
2 Height & weight measured during Screening, weight only at other visits.
3 Basic chemistry: bicarbonate, BUN, creatinine, chloride, glucose, potassium, sodium, Vitamin B12 and folate, creatinine clearance (US-only).
4 Liver function tests: ALT, AST and total bilirubin.
5 Haematology and Iron markers all visits: haematocrit, haemoglobin, platelets, RBC count, RBC Hb concentration; WBC count and differential; total serum iron; ferritin, transferrin; TSAT.
6 Maximum 14 days between Screening and randomization.
7 Urine pregnancy test, kits provided by central laboratory.
8 Follow up is conducted by telephone, unless the subject has an ongoing AE that requires physical examination or investigations for assessment/management. Visit 8 will take place 14 days after Visit 4 (week 12/Early Termination) unless subject discontinued treatment early and end of study/Week 52 assessments occurred 14 days or more after the last dose.
9 Visit Windows: Maximum of 14 days between screening and randomization. Subjects randomized to IV Iron must complete Visit 2 ideally on the day of randomisation or within 5 days. The subject must visit to complete:
Visits 3 and 4: +/- 1 week relative to date of Visit 2/first dose administered
Visit 4 to 7: +/- 1 week relative to date of Visit 2/first dose administered
Visit 8: 14 days (+5 days) after Visit 4 (unless subject discontinued treatment early and end of study/Week 12 assessments occurred 14 days or more after the last dose).
10 SF36 to be administered at Week 0 (baseline), and Week 12, 24, 36, and 52.
11 SF36 is not collected at this visit.

12 For CD subjects, a CDAI diary card will be provided at Screening, for completion over the 7 days prior to scheduled Visit 2. The screening haematocrit value will be used for the screening CDAI calculation to assess eligibility prior to randomization and will also count as the baseline CDAI value. For CD subjects, CDAI diary cards will then be provided at Visits 3 and 4. Subjects who went beyond week 12 will receive CDAI diary at Visit 5 and 6 for completion over the 7 days prior to the scheduled dates of Visit 4, 5, 6 and 7.
13 Retained serum and plasma samples will only be collected from trial subjects at those Investigator sites who have suitable freezer facilities for storage of the samples at -70°C/-80°C, prior to subsequent periodic shipping to the central laboratory on dry-ice. Detailed instructions will be provided in the Laboratory Manual.
14. Dispense/Administer study drug: based on the randomization stratification subjects will receive FCM/Ferric Maltol on Visit 2. If subjects went beyond Visit 4 (Week 12) and their treatment period will be continuing according to previous protocol until the next scheduled visit.
15. Visit 5 (Week 24) will only occur if subjects went beyond Visit 4 according to previous protocol, otherwise this will be the End of Study Visit.
16. Visit 6 (Week 36) will only occur if subjects went beyond Visit 5 (week 24) according to previous protocol, otherwise this will be the End of Study Visit.
17. Visit 7 (Week 52) will only occur if subjects went beyond Visit 6 (Week 36) according to previous protocol, otherwise this will be the End of Study Visit.

# 4 POWER AND SAMPLE SIZE CONSIDERATIONS

This study is designed with a 'responder %' primary endpoint, where responder is a subject who either has at least 2g/dL rise in Hb from baseline or has shift from baseline of anaemic to non-anaemic status. The primary endpoint is the Hb responder rate at Week 12. The comparator group (IV Iron) responder rate is assumed to be 75% and the absolute non-inferiority margin is 20%. The non-inferiority margin is based on clinical opinion, coupled with the fact that in a previous study, the response rate in the placebo arm was less than 5% at 12 weeks.

The study is powered at 90% when the true response rate in the ferric maltol arm is 75% (i.e., the true difference in response rate to IV Iron is 0%). The primary statistical test is based on the 2-sided 95% confidence interval for the difference in response rates between the two treatment groups. This requires 108 subjects per treatment arm. In order to allow for protocol deviations, the sample size has been increased by 12% to 121 subjects per treatment arm.

# 5 ANALYSIS POPULATIONS

## 5.1 Randomized Population

The randomized population will include all subjects who are randomized.

## 5.2 Safety Population

The Safety population will consist of all subjects who have received at least one dose of a study drug. Tables summarised or analysed using the safety population will be presented by actual treatment given.

## 5.3 Intention-to-Treat (ITT) Population

The ITT population will include all subjects who are randomized. However, any efficacy measurement obtained after a patient experienced a serious adverse event (SAE) of haemorrhage (defined by MedDRA under haemorrhagic terms), received a blood transfusion, or received intravenous iron (outwith of

protocolled dose) or erythropoiesis stimulating agent will be removed from the ITT analysis.

### 5.4 Per Protocol (PP) Population

The PP population will consist of those randomized subjects who do not have major protocol deviations during the first 12 weeks of the study. Protocol deviations occurring during the study will be classified as major or minor prior to database lock. See Section 8.4 for definitions of protocol deviations. In addition, patients who do not attend week 12 and have no Hb result at week 12 will be excluded from the PP population.

The ITT and PP population will be used in the primary efficacy analysis. Tables summarised or analysed using the ITT population will be presented by randomized treatment. Tables summarised or analysed using the PP population will be presented by actual treatment given.

## 6 CONSIDERATIONS FOR DATA ANALYSIS

### 6.1 General Statistical Considerations

All analyses described in this plan are considered *a priori*, in that they have been defined prior to locking the database. All other analyses, if any, designed subsequent to locking the database, will be considered *post hoc* and, thereby, exploratory. *Post hoc* analyses will be clearly identified in the Clinical Study Report.

### 6.2 Programming Environment

All analyses will be conducted using SAS® version 9.3.

### 6.3 Strata and Covariates

Randomization will employ screening Hb (< 10g/dL or ≥ 10g/dL for Females; < 11g/dL or ≥ 11g/dL for Males) and IBD subgroup (UC or CD) as binary stratification factors. Unless otherwise indicated, IBD subgroup will be included as a binary stratification factor and baseline Hb will be included as a continuous factor in statistical modelling.

### 6.4 Subgroups

The secondary endpoint analyses involving haemoglobin (see Sections 9.2.1.1 and 9.2.1.3) will be repeated for the subgroup of subjects with baseline Haemoglobin < 9.5g/dL.

### 6.5 Multiple Comparisons and Multiplicity

There are no planned adjustments for multiple comparisons.

## 6.6 Significance Level

Unless stated otherwise, all statistical tests will be two-sided, with a significance level of 0.05. Confidence intervals (CIs) will be calculated at the 95% level, reflecting a type I error rate of 0.05.

## 6.7 Statistical Notation and Methodology

Unless stated otherwise, the term "descriptive statistics" refers to the number of subjects (n), mean, median, standard deviation, minimum, and maximum for continuous data and frequency distributions (counts and percentages) for categorical data. The denominator for percentages will use the number of subjects in the appropriate analysis population. Minimum and maximum values will be rounded to the precision of the original value, means and medians will be rounded to 1 decimal place greater than the precision of the original value, and standard deviations will be rounded to 2 decimal places greater than the precision of the original value. Percentages will be rounded to the nearest whole number (zeros are not displayed) with values of "<1%" and ">99%" shown as necessary for values falling near the boundaries (i.e., 0% and 100%). P-values will be presented with a precision of 3 decimal places, and p-values less than 0.001 will be presented as "<0.001".

Unless otherwise noted, all data collected during the study will be included in data listings and will be sorted by site number, subject number, and then by date/time for each subject number.

# 7 DATA HANDLING METHODS

## 7.1 Missing Data

### 7.1.1 Date Values

In cases of incomplete dates (e.g. AE, concomitant medication, and medical history start and/or stop dates), the missing component(s) will be assumed as the most conservative value possible. For example, if the start date has a missing day value, the first day of the month will be imputed for study day computations (i.e. treatment-emergent status, etc.) unless it is the same month and year (only day missing) or year (day and month missing) as start date of treatment. In this case, start date of treatment will be used, which assumes treatment emergence. If day is missing for an end date, the last day of the month will be imputed.

Date imputation will only be used for computational purposes such as treatment-emergent status, etc. Actual data values, as they appear in the original CRFs, will be presented within the data listings.

### 7.1.2 Non-Date Values

Every effort will be made to obtain required data at each scheduled evaluation from all subjects who have been enrolled. Missing values for the primary efficacy endpoint of Hb responder (i.e., a subject who either has at least 2g/dL rise in Hb from baseline or has shift from baseline of anaemic to non-anaemic status) at Week 12 will be imputed using a Multiple imputation approach for the primary efficacy analysis on the ITT population. If the value of Hb is missing, multiple imputation based on pre-treatment visits will be used in the baseline and multiple imputation based on post-treatment visits will be used to determine the post-baseline value to be used in the primary efficacy endpoint analysis on the ITT population.

As additional sensitivity analyses, missing values for the primary efficacy endpoint on the ITT population will also be imputed using a Last Observation Carried Forward (LOCF) approach up to Week 12.

Missing values for efficacy variables will be replaced for the secondary endpoint analyses, as follows:

- For the key endpoint, 'Hb concentration within normal limits', subjects not returning to a follow-up visit will be assumed to have Hb concentration not within normal limits.
- For the endpoint, 'Increases in Hb concentration of ≥ 1g/dL', subjects not returning to a follow-up visit will have their Hb concentration imputed using multiple imputation.
- For the endpoint, 'Increases in Hb concentration of ≥ 2g/dL', subjects not returning to a follow-up visit will have their Hb concentration imputed using multiple imputation.

Missing values for the long-term secondary efficacy endpoints, normalization of Hb at 6 and 12 months and normalization of ferritin at 6 and 12 months, will not be imputed.

Missing values for the SF-36 pharmacoeconomic analysis will be imputed by a LOCF approach up to Week 12. Missing values for all other pharmacoeconomic analyses will be kept as missing.

Safety data will be used according to availability, with no imputation for missing data.

Imputation of data values will only be carried out, where indicated, for computational purposes and fitting of models. Actual data values, as they appear in the original CRFs, will be shown in the data listings.

## 7.2 Visit Windows

Subjects should attend the study visits according to Table 1. Schedule of Assessments.

Values will be presented for all scheduled study visits according to the nominal visit obtained from the CRF. If an unscheduled visit falls in a visit window with an existing nominal visit assessment, the nominal assessment will be used for summary presentation. If no nominal visit assessment exists for a visit window with unscheduled visit(s), then the latest unscheduled visit within the visit window will be used. If multiple nominal assessments are collected within the same visit, the latest value and corresponding date will be used for summary presentation.

All values will be included in the data listings.

## 7.3 Data Derivations and Definitions

The following definitions and derivations will be used for this study:

- Baseline will be defined as the last measurement collected prior to administration of the first dose of a study drug.

- Change from baseline (CFB) will be calculated as:

  CFB = observed value – baseline value.

- As detailed in the protocol, the randomization visit is Visit 2 / Day 0 (the date of first dose with ferric maltol or FCM).
  In order to be consistent with the SDTM guidance, within SDTM and ADaM datasets, Study Day 1 will be the day corresponding to subject randomization, and Study Day -1 will be the day prior to subject randomization.

- Age in years will be as indicated within the CRF.

- A subject will be considered to have normalized Hb if their measured Hb is greater than 12.0 g/dL (females) or 13.0 g/dL (males).

- An AE will be considered treatment-emergent if it begins or worsens on the day of or after the first dose of study medication and within 14 days of the last dose of study medication. AEs will be considered treatment-emergent if all or part of the date of onset of the AE is missing and it cannot be determined if the AE meets the definition for treatment-emergent.

- Concomitant medications will be those medications that began or stopped on or after the date corresponding with the first dose of a study drug.

- Prior medications will be those medications that ended before the date corresponding with the first dose of a study drug. Concomitant and prior medications are mutually exclusive.

- Study completers will be those subjects who complete the study procedures for all visits up to and including the Follow-up visit.

- Observed Cases (OC) will include assessments collected at each scheduled or unscheduled visit and will not contain imputed values.

- Last Observation Carried Forward (LOCF) utilizes the clinical assessment value of each efficacy endpoint (primary or secondary), if available, or carries-forward the last available assessment value.

- The non-inferiority margin is the maximum difference in the primary efficacy analysis between the two treatment arms we are prepared to tolerate if the ferric maltol treatment is not to be considered [clinically] inferior to the IV Iron treatment. The non-inferiority margin for this study is 20%.

# 8 STUDY POPULATION

Unless otherwise noted, the ITT population will be used for listings of the study population. ITT and PP populations will be used for summaries of the study population.

## 8.1 Analysis Populations

An analysis population listing will be provided for all enrolled subjects. This listing will include, for each population detailed in Section 5, whether or not the subject was included in the population and the reason for being excluded from the population. Number and percentage of subjects in each analysis population (ITT, PP, and Safety) will be summarised by treatment arm and overall (both treatment arms combined) for the Randomized population. The denominators for calculating the percentages will be based on the subjects of the Randomized population for the treatment arm summarised.

## 8.2 Subject Disposition

Subject disposition will be presented for all screened subjects. Number and percentage of subjects who enrol, complete the study, prematurely discontinue from the study, and reasons for study discontinuation will be summarised by treatment arm and overall. The denominator for calculating percentages will be based on the number of screened subjects.

## 8.3 Inclusion/Exclusion Criteria

Inclusion and exclusion criteria failures will be included in a data listing for the ITT population.

## 8.4 Protocol Deviations

Major deviations, for the purpose of the statistical analyses, are defined to be those deviations that could potentially bias either efficacy or safety conclusions of the study. Subjects associated with major protocol deviations will be identified and flagged prior to freezing (or, locking) the database at each of the two planned database freeze (or, lock) timepoints (interim analysis freeze/lock and end-of-study freeze/lock).

Major protocol deviations include, but are not limited to, the following categories:

- Subjects that received at least one dose of a study drug and who did not satisfy major inclusion/exclusion criteria.
- Subjects meeting conditions for withdrawal who were not withdrawn.
- Subjects that received a prohibited concomitant medication.
- Subjects that had at least one major required protocol procedure not being performed. Examples of major required protocol procedures include missing haemoglobin results at baseline and/or at Week 12.
- Subjects that received the opposite study drug treatment than the one to which they were randomized.
- Repeated non-compliance in self-administration of ferric maltol study drug, where non-compliance is less than 80% or greater than 120% of expected administration of study drug.
- Other relevant deviations, to be judged on an individual basis.

Protocol deviations will be summarised with frequency distributions (counts and percentages) by treatment arm, category, and deviation type (minor and major) for the ITT population. The denominators for calculating percentages will be based on the number of subjects in the analysis population for each treatment arm and overall.

Protocol deviations will be provided within a data listing by subject.

## 8.5 Demographic and Other Baseline Characteristics

The demographics listing will include year of birth, age, gender, ethnicity (Hispanic/Latino or non-Hispanic/Latino), race (American Indian/Alaskan Native, Asian, Black/African American, Native American/Pacific Islander, White, Other), and baseline Hb.

The listing of reproductive status will include fertility status (sterile, post-menopausal, potentially fertile), method of birth control (barrier method with spermicide, oral contraceptives, depot contraceptives [implants/injectables], injectables, intrauterine device, complete sexual abstinence, vasectomized partner,

none, other), date of last menstrual cycle, and whether it has been more than one year since the last menstrual cycle.

The listing of randomization will include the stratification levels for IBD subgroup (UC or CD) and screening Hb (< 10g/dL or ≥ 10g/dL for Females; < 11g/dL or ≥ 11g/dL for Males), the randomization date/time, assigned (planned) treatment arm, and the actual treatment corresponding to the study drug received.

The listing of IBD status will include date (month and year) of IBD diagnosis, date (month and year) for each of the last 3 flares of IBD preceding randomization, and IBD subgroup.

Demographic and baseline characteristics will be summarised using descriptive statistics by treatment arm and overall for the ITT and PP populations. The summary will present: each of the variables included in the demographics listing, with the exception of year of birth; fertility status, summarised separately for males and females; and, the stratification levels used in randomization. Subjects who report more than one racial category will be summarised in a separate derived race level labelled as "Multiple".

## 8.6 Prior and Concomitant Medications

Prior and concomitant medications, including: all prescription iron replacement therapies taken in the last 12 months prior to screening; all oral ferrous treatment history prior to screening; and, all IV Iron treatment history prior to screening will be coded with the World Health Organization (WHO) drug dictionary. Prior and concomitant medications will be separately summarised with descriptive statistics by drug Anatomical Therapeutic Chemical (ATC) Classification System (level one), generic name, and treatment arm for safety population. Data listings will be included that show all medications by generic name and verbatim name.

## 8.7 Medical History

A comprehensive data listing of all medical histories and ongoing conditions (including start and stop dates and status at end of study) – for the immediate 5 years preceding randomization – will be included. Medical history codes will be summarised for each treatment arm and overall for both ITT and PP populations.

## 8.8 Subject Progress and Unscheduled Visits

Progression for each scheduled visit will be displayed in a listing by subject. An unscheduled visit listing will include the date of each visit and a description of the assessments performed at the visit.

## 8.9 Study Drug Disposition

The duration of exposure (days) to study drug will be summarised by treatment arm using descriptive statistics for the Safety population. Duration of exposure will

be calculated as the difference between the last dose date of study drug and the first dose date of study drug plus 1 day. Administration dates of IV Iron will be provided in a data listing by subject, for those subjects randomized to the IV Iron treatment arm.

Compliance with ferric maltol will be assessed at each visit for which the study drug is returned by the subject, and also assessed over the entire Treatment Period. Compliance will be evaluated by comparing the expected number of capsules of ferric maltol administered – based on the total number of treatment days between visits and within the entire treatment period – with the actual number of capsules of ferric maltol taken. If the return date is missing for a particular visit, then compliance will be missing for that visit. Compliance with ferric maltol will be calculated for subjects of the Safety population who are randomized to the ferric maltol treatment arm. Compliance values that are less than 0% will be set to 0%. Summaries of compliance will be presented using descriptive statistics for subjects of the Safety population who are randomized to the ferric maltol treatment arm. A data listing of treatment compliance will be provided by subject, for those subjects of the Safety population who are randomized to the ferric maltol treatment arm.

Compliance will be calculated as follows and expressed by a percentage:

Compliance = (#caps disp - #caps ret) / (#caps exp to be administered),

where #=number of, caps=capsules, disp=dispensed, ret=returned, exp=expected.

# 9 EFFICACY ANALYSES

The ITT and PP populations will be used for summaries and analyses of all efficacy endpoints described below.

In addition to the tables described in the sections below, data for efficacy variables will be provided in data listings by subject.

## 9.1 Primary Efficacy Analyses

### 9.1.1 Primary Analysis

Haemoglobin (Hb) concentration will be collected at screening (Visit 1), pre-randomization at Visit 2 (baseline assessment), and at each visit over the Treatment Period. Hb concentration (Observed Cases [OC]) will be summarised for the ITT and PP populations using descriptive statistics by visit for each treatment arm and overall. A scatterplot of Hb concentrations (OC) at Week 12 versus baseline will be presented by treatment arm for the ITT and PP populations.

A dichotomous variable, representing response status in the degree to which Hb increases relative to baseline, will be calculated for each visit over the Treatment Period. Those subjects with at least 2g/dL increase in Hb from baseline or who

have normalization of Hb (>12.0 g/dL for women, >13.0 g/dL for men) will be classified as "Hb responders".

The null hypothesis for the primary efficacy endpoint analysis is that the proportion of Hb responders at Week 12 in the ferric maltol treatment arm is at least 20% lower than the proportion of Hb responders at Week 12 in the IV Iron treatment arm (i.e., ferric maltol is inferior to IV Iron for rate of Hb response at Week 12). The alternative hypothesis is that the proportion of Hb responders at Week 12 in the ferric maltol treatment arm is not worse than 20% lower than the proportion of Hb responders at Week 12 in the IV Iron treatment arm (i.e., ferric maltol is non-inferior to IV Iron for rate of Hb response at Week 12).

The proportion of Hb responders at Week 12 will be compared between the two treatment arms, by calculating a 2-sided confidence interval for difference in the proportions of Hb responders (risk difference) between the two treatment arms and comparing the lower confidence limit of this CI to the pre-specified non-inferiority margin. The CI will be calculated using the Delta Method approach [1] based on a logistic regression model, adjusted for treatment arm, baseline Hb (below the observed median or at least the observed median), and IBD subgroup (UC or CD).

Note that missing values for the Hb responder status at Week 12 primary efficacy analysis will be imputed using a multiple imputation approach for ITT population. BOCF will be used in the case of missing baseline data to use the screening value. Missing values of Hb concentration at Week 12 will be imputed using the SAS® procedure MI. This will generate 20 complete datasets containing imputed values of Hb concentration at Week 12 for those missing in the original dataset. Variables considered helpful in guiding this imputation are baseline (Visit 2) Hb concentration and treatment arm. The code for implementing the SAS® procedure MI is:

```
PROC MI DATA = <dataset _name> NIMPUTE = 20 SEED = <seed> OUT = MI_out;
  VAR baselineHb TreatmentArm Week12Hb;
  EM MAXITER = 10000;
RUN;
```

Each model, corresponding to a given endpoint analysis, will be fitted against each of the imputed datasets. The SAS® procedure MIANALYZE will then be used to combine the individual analysis results for the given endpoint into one result.

The null hypothesis for the primary efficacy analysis will be rejected (i.e., ferric maltol will be deemed non-inferior to IV Iron), if the LCL of the 95% CI for the difference (ferric maltol - IV Iron) in proportions of Hb responders at Week 12 (BOCF) is ≥ -20% for both the ITT analysis and the PP analysis. The p-value in testing the aforementioned hypotheses will be equal to one minus the confidence coefficient corresponding to the CI for the risk difference whose LCL equals -20% (i.e., inverting the CI for the risk difference to a test statistic and calculating the p-value from the test statistic).

The proportion of Hb responders will be summarised for the ITT and PP populations using descriptive statistics by visit over the Treatment Period for each treatment

arm and overall, utilizing all available Hb responder data (i.e., OC approach). Each summary will also present the proportion of Hb responders for the Week 12 visit (MI) using descriptive statistics, and will include the corresponding logistic regression model adjusted difference in the proportions of Hb responders between the two treatment arms, the corresponding adjusted 95% CI for the difference in proportions, and the adjusted p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. Finally, each summary will also present the estimated logistic regression model adjusted odds ratio of Hb responder at Week 12 (MI) between the two treatment arms and corresponding 95% CI for the odds ratio. A bar plot of the Hb responder rate will be presented by visit (OC) and treatment arm for each assessment.

### 9.1.2 Sensitivity Analysis

Two sensitivity analyses will be performed for the primary efficacy endpoint.

#### 9.1.2.1 Observed Cases (OC) Analysis

The first sensitivity analysis will include all available assessments of Hb responder and will not exploit imputation of missing Hb responder status (i.e., only Observed Cases [OC] of Hb responder status are considered here). The primary efficacy endpoint analysis will be repeated using the OC approach for the ITT and PP populations. The proportion of Hb responders at Week 12 (OC) will be summarised using descriptive statistics for each treatment arm and overall. The summary will include the corresponding logistic regression model adjusted difference in the proportions of Hb responders (risk difference) between the two treatment arms, the adjusted 95% CI for the difference in proportions, and the adjusted p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. The summary will also include the estimated adjusted odds ratio of Hb responder at Week 12 (OC) between the two treatment arms and corresponding 95% CI for the odds ratio.

#### 9.1.2.2 Observed Cases (OC) Repeated Measures Analysis

The second sensitivity analysis for the primary efficacy endpoint will be performed on the ITT and PP populations, in which Hb responder status will be analysed in a longitudinal analysis using a repeated measures generalized linear mixed model (GLMM) approach (i.e., a repeated measures logistic regression approach) for all available assessments of Hb (i.e., only Observed Cases [OC] of Hb responder status are considered here) over Visits 2 (Day 0; baseline) to 4 (Week 12). The approach will fit a logistic model to the Hb responder dependent variable, where an unstructured working correlation structure will be specified (e.g., by way of the appropriate random statement within the GLIMMIX SAS® procedure) to model the within-subjects correlation of repeated Hb responder status at Visits 3 and 4. If the model fit fails to converge, alternative working correlation structures will be examined. Covariates of the GLMM will include the fixed effects of treatment, baseline Hb, IBD subgroup, visit, and treatment-by-visit interaction. The estimated adjusted odds ratio for treatment effect, and its corresponding 95% CI, from the

model will be presented by visit.

To assess the sensitivity of Hb responder status over time – by way of a general linear model (GLM) approach – separate logistic regression models will be fitted to the Hb responder dependent variable at each visit during the Treatment Period up to Week 12, using all available assessments of Hb (i.e., only Observed Cases [OC] of Hb responder status are considered here). Each model will include terms for the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The estimated adjusted odds ratio for treatment effect, and its corresponding 95% CI, will be presented for each model (i.e., by visit). Sensitivity of Hb responder status over time will be assessed by comparing these by-visit treatment effect odds ratio estimates from the logistic regression models (the GLMs) to those obtained from the GLMM model.

## 9.2 Secondary Efficacy Analyses

### 9.2.1 Secondary Analysis

#### 9.2.1.1 Key Secondary Endpoints

The change from baseline (CFB) in Hb concentration at Week 12 (MI) secondary endpoint will be analysed for the ITT population using an analysis of covariance (ANCOVA) model.

For the multiple imputation, the code for implementing the SAS® procedure MIANALYZE to the CFB Hb concentration at Week 12 secondary endpoint is:

```
DATA CFB_Impute; SET MI_out;  CFBHbW12 = Week12Hb – baselineHb; RUN;

PROC SORT DATA = CFB_Impute; BY _Imputation_; RUN;

PROC REG DATA = CFB_Impute OUTEST = outreg COVOUT;
  BY _Imputation_;
  MODEL CFBHbW12 = binarybaselineHb TreatmentArm IBDGroup;
RUN;

PROC MIANALYZE DATA = outreg EDF = <complete-data degrees of freedom>;
  MODELEFFECTS Intercept binarybaselineHb TreatmentArm IBDGroup;
  ODS OUTPUT ParameterEstimates = parm1;
RUN;
```

The ANCOVA model will have terms for treatment arm, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The treatment effect of ferric maltol compared to IV Iron will be considered using differences of adjusted means. No formal statistical significance level in the treatment comparison will be defined, as the p-value from the ANCOVA model will be presented for descriptive purposes only. A summary of CFB in Hb concentration (OC) will be provided using descriptive statistics for each visit over the Treatment Period by treatment arm and overall. The summary will include the following

ANCOVA model estimates: adjusted means (i.e., Least-Square Means [LSMs]) of CFB in Hb concentration at Week 12 for each treatment arm, the difference in treatment arm LSMs, the corresponding 95% CIs for treatment arm specific LSMs and difference in treatment arm LSMs, and the p-value for the difference in treatment arm LSMs. Mean CFB in Hb concentration will be plotted against visit (OC) by treatment arm for the ITT and PP populations.

A dichotomous variable, representing status of Hb concentration relative to the normal limits of Hb will be calculated for each visit over the Treatment Period. Women with Hb concentration greater than 12 g/dL and men with Hb concentration greater than 13 g/dL will be classified to be within normal limits of Hb. Those subjects with Hb concentration within the normal limits will be classified as "Hb normal limit responders". Subjects missing Hb concentration at Week 12 will be assumed to have Hb concentration not within normal limits (i.e., these subjects will be considered Hb normal limit non-responders). The logistic regression modelling and hypothesis testing for the primary analysis (Section 9.1.1) will be repeated here for the Hb normal limit responder at Week 12 secondary endpoint.

The code for implementing the SAS® procedure MIANALYZE to the Hb normal limit responder at Week 12 secondary endpoint for the ITT population is:

```
        DATA NormalHb_Impute; SET MI_out;
          NLResponder = 0;
          IF Gender =: 'F' AND Week12Hb > 12.0 THEN NLResponder = 1;
          IF Gender =: 'M' AND Week12Hb > 13.0 THEN NLResponder = 1;
        RUN;

        PROC SORT DATA = NormalHb_Impute; BY _Imputation_; RUN;

        PROC LOGISTIC DATA = NormalHb_Impute;
          MODEL NLResponder = binarybaselineHb TreatmentArm IBDGroup / covb;
          BY _Imputation_;
          ODS OUTPUT ParameterEstimates = lgsparms
                     CovB               = lgscovb;
        RUN;

        PROC MIANALYZE PARMS = lgsparms COVB(effectvar=stacking) = lgscovb;
          MODELEFFECTS Intercept binarybaselineHb TreatmentArm IBDGroup;
RUN
```

A summary will be provided for the Hb normal limit responder secondary endpoint using descriptive statistics by visit over the Treatment Period for each treatment arm and overall, utilizing all available Hb data (i.e., OC approach). The summary will also present the Hb normal limit responder rate for the Week 12 visit (MI) for ITT population using descriptive statistics, and will include the corresponding logistic regression model adjusted difference in proportions of Hb normal limit responders between the two treatment arms, the corresponding [Delta Method derived] adjusted 95% CI for the difference in proportions, and the p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. Finally, the summary will also present the estimated logistic regression model adjusted odds ratio of Hb normal limit responders at Week 12 (MI) between

the two treatment arms and corresponding 95% CI for the odds ratio.

#### 9.2.1.2 Other Secondary Efficacy Endpoints

The analysis for the key secondary endpoint, CFB in Hb concentration at Week 12 (MI), will be repeated for the secondary endpoints: CFB in Hb concentration at Week 12 (MI) in subjects with baseline Hb <9.5 g/dL; CFB in Hb concentration at Week 4 (MI); and, CFB in Hb concentration at Week 4 (MI) in subjects with a baseline Hb <9.5 g/dL.

Two dichotomous variables, each representing response for the magnitude rise in Hb from baseline, will be calculated for each visit during the Treatment Period from the Hb concentration data. Those subjects with a rise in Hb from baseline of at least 1g/dL will be classified as "Hb rise responders" for the first variable; those subjects with a rise in Hb from baseline of at least 2g/dL will be classified as "Hb rise responders" for the second variable. Separate logistic regression models will be fitted to each of the Hb rise responders at Week 12 variables (secondary endpoints) using an MI approach for the ITT populations. Each model will include terms for the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. A hypothesis test of the null hypothesis of inferiority for ferric maltol responder rate compared to IV Iron responder rate will be considered for each Hb rise responder secondary endpoint using a non-inferiority margin of 20%. No formal statistical significance level to declare non-inferiority (i.e., rejection of the null hypothesis) of ferric maltol to IV Iron is defined, as CIs and p-values will be presented for descriptive purposes only. The p-value in testing the aforementioned hypotheses will be calculated as described in the primary analysis (Section 9.1.1).

The above analysis will be repeated for the secondary endpoints: proportions of subjects with baseline Hb <9.5 g/dL that achieve an increase in Hb concentration of ≥1 or ≥2 g/dL at Week 12; proportions of subjects who experience a change from baseline in Hb concentration ≥1 or ≥2 g/dL at Week 4 (MI); and, proportions of subjects with baseline Hb <9.5 g/dL that achieve an increase in Hb concentration of ≥1 or ≥2 g/dL at Week 4 (MI). In addition, the above analysis for the Hb normal limit responder at Week 12 key secondary endpoint will be repeated for the secondary endpoints: proportion of subjects with baseline Hb <9.5 g/dL that is within normal limits at Week 12; proportion of subjects with Hb concentration within normal limits at Week 4; and, proportion of subjects with baseline Hb <9.5 g/dL that is within normal limits at Week 4 (MI).

Serum ferritin will be collected at screening (Visit 1), pre-randomization at Visit 2 (baseline assessment), and at each visit over the Treatment Period. The above ANCOVA analysis will be repeated for the CFB in serum ferritin at Week 12 (MI) secondary endpoint. A summary of serum ferritin and CFB in serum ferritin (OC) will be provided using descriptive statistics for each visit (CFB will be presented at each visit during the Treatment Period) by treatment arm. The summary will include the following ANCOVA model estimates: adjusted means (i.e., Least Squares Means [LSMs]) of CFB in serum ferritin at Week 12 for each treatment

arm, the difference in treatment arm LSMs, the corresponding 95% CIs for treatment arm specific LSMs and difference in treatment arm LSMs, and the p-value for significance of the difference in treatment arm LSMs. Normality and homogeneity of variances for the residuals of each ANCOVA model will be assessed using the procedure outlined in Section 9.2.1.1.

#### 9.2.1.3 Long-term Secondary Endpoints

A dichotomous variable, representing anaemia status (anaemic versus non-anaemic), will be calculated for each visit over the Treatment Period. Women with Hb levels greater than 12.0 g/dL and men with Hb levels greater than 13.0 g/dL will be classified as non-anaemic. Separate logistic regression models will be fitted to the long-term secondary endpoints, anaemia status at each visit over Visits 5-7 (OC), for the ITT and PP populations. Each model will include terms to capture the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The above hypothesis testing for Hb normal limit responder (Section 9.2.1.1) will be repeated here for the anaemia status secondary endpoint.

A dichotomous variable, representing status of serum ferritin level relative to the normal limits of serum ferritin, will be calculated for each visit over the Treatment Period. Women with serum ferritin between 10 and 291 ng/ml and men with serum ferritin between 22 and 322 ng/ml will be classified to be within normal limits of serum ferritin. Those subjects with serum ferritin within the normal limits will be classified as "ferritin normal limit responders". Separate logistic regression models will be fitted to the long-term secondary endpoints, ferritin normal limit responder at each visit over Visits 5-7 (OC), for the ITT and PP populations. Each model will include terms to capture the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The above hypothesis testing for rise in Hb responder (Section 9.2.1.2) will be repeated here for the serum ferritin normal limit responder secondary endpoint.

Separate summaries will be provided for each of the binary long-term secondary endpoint variables defined above. Each secondary endpoint will be summarised using descriptive statistics by visit over the Treatment Period for each treatment arm and overall, utilizing all available data (i.e., OC approach). Each summary will also present the corresponding logistic regression model adjusted difference in proportions (risk difference) of secondary endpoint Hb responders between the two treatment arms, the corresponding [Delta Method derived] adjusted 95% CI for the difference in proportions, and the p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. Finally, each summary will also present the appropriate estimated logistic regression model adjusted odds ratio of secondary endpoint at Visits 5-7 (OC) between the two treatment arms and corresponding 95% CI for the odds ratio.

Any patients who are in the PP population but have major deviations after Week 12 will be excluded from the PP analyses of the long-tern secondary endpoints.

# 10 PHARMACOECONOMIC ANALYSES

Unless otherwise noted, the ITT population will be used for all analyses, summaries, and listings of pharmacoeconomic endpoints.

## 10.1 Medical Outcomes Study 36-item Short Form (SF-36)

The SF-36 is a multipurpose, proprietary health survey with 36 questions (www.sf-36.org). It was constructed to survey health status in the Medical Outcomes Study and designed for use in clinical practice and research, health policy evaluations, and general population surveys. The SF-36 includes one multi-item scale that assesses eight health concepts (components):

- Limitations in physical activities because of health problems (SF-36 Physical Functioning Component);
- Limitations in social activities because of physical or emotional problems (SF-36 Social Functioning Component);
- Limitations in usual role activities because of physical health problems (SF-36 Role-Physical Component);
- Bodily pain (SF-36 Bodily Pain Component);
- General mental health (psychological distress and well-being [SF-36 Mental Health Component]);
- Limitations in usual role activities because of emotional problems (SF-36 Role-Emotional Component);
- Vitality (energy and fatigue [SF-36 Vitality Component]); and
- General health perceptions (SF-36 General Health Component).

These eight health component scales can be further summarised into two summary scores, the SF-36 Mental Component Score (MCS) and the SF-36 Physical Component Score (PCS).

The survey was constructed for self-administration by persons 14 years of age and older, and for administration by a trained interviewer in person or by telephone. The SF-36 survey will be administered at study visits as indicated in the schedule of assessments (Table 1), commencing pre-randomization at Visit 2. The SF-36 survey will be completed by the subjects in their native language. Missing values for each of the 36 items of the SF-36 will employ a LOCF approach in ITT analyses up to Week 12.

The process of computing SF-36 component scales involves several steps [4]. First, 10 items of the survey are reverse-coded. Next, component-specific raw scales are formed by way of simple summation over the item scores for the appropriate health component of the SF-36. The component-specific raw scales are then transformed to range between 0 and 100, followed by a norm-based (z-score) transformation.

The latter transformation is carried out such that each scale is mapped to the scale of the 1998 general United States (US) population [4]. The two summary scores, MCS and PCS, are computed by aggregating norm-based component scores using factor score coefficients from the 1990 general US population [4]. The aggregated summary scores are standardized to have a mean of 50 with a standard deviation of 10, in the general 1998 US population [4]. The full SF-36 scoring algorithm is published in [4] and is the approach we take here to score the SF-36. A summary of each of the 8 subscales and the 2 summary scores will be presented using descriptive statistics by visit (OC) for each treatment arm and overall.

The change from baseline (CFB) in each of the 8 SF-36 component subscales and the 2 summary scores will be calculated for each visit during the Treatment Period (BOCF) and analysed with separate ANCOVA models. Each ANCOVA model will include terms for treatment arm, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. A hypothesis test for the treatment effect of ferric maltol compared to IV Iron will be conducted using differences of adjusted means. Statistical significance in the treatment comparison will be declared if the p-value of the hypothesis test from the ANCOVA model is less than or equal to 0.05. Significance levels will not be adjusted for multiplicity. A summary of CFB in each SF-36 subscale and each summary score will be provided using descriptive statistics for each visit during the Treatment Period (OC and BOCF) by treatment arm. The summary will include the ANCOVA estimates: adjusted means (LSMs) of CFB in each SF-36 component and summary score for each treatment arm, the difference in treatment arm LSMs, corresponding 95% CIs for LSMs and difference in treatment arm LSMs, and the p-value for the comparison of treatment arm LSMs. Normality and homogeneity of variances for the residuals of each ANCOVA model will be assessed using the procedure outlined in Section 9.2.1.1.

A data listing of the SF-36 questionnaire will be presented by visit and subject.

## 10.2 Correction and Maintenance of IDA

Status (yes or no) of administration of IV Iron will be documented at each visit (scheduled or unscheduled) over the Treatment Period for subjects randomized to the IV Iron treatment arm, commencing with the assessment conducted post-randomization at Visit 2. The total number of visits (scheduled and unscheduled) of administration of IV Iron over the Treatment Period will be calculated for each subject randomized to the IV Iron treatment arm. A summary of administration of IV Iron will be provided using descriptive statistics by visit, including a visit labelled "unscheduled", over all visits, and over Post-Day 0 visits. The summaries over all visits and Post-Day 0 visits will capture the respective number of hospital visits for administration of IV Iron and restart of IV Iron pharmacoeconomic endpoints. Missing values of administration of IV Iron will be kept as missing in the summary. A data listing of dose administration will be provided by subject.

Details of IV Iron dosing at the study site include: the duration of time (hours) spent at the study site for IV Iron administration; the dose (mg) of IV Iron

administered at the study site; and, the number of days off work (or, school/education or childcare) that the subject is required to miss due to administration of IV Iron at the study site, will be documented at each visit (scheduled or unscheduled) over the Treatment Period for subjects randomized to the IV Iron treatment arm. A summary of these IV Iron dosing details will be provided using descriptive statistics by visit (including a visit labelled "unscheduled"). A data listing of IV Iron dosing details will be provided by subject.

Status (yes or no) of repeat Hb or Iron testing will be documented at each unscheduled visit over the Treatment Period. The total number of repeat Hb or Iron tests over all unscheduled visits of the Treatment Period will be calculated for each subject and will be analysed as a repeated time-to-event approach. A GLMM approach (i.e., a repeated measures Poisson regression) will be used to estimate the ratio (relative risk) of mean repeat Hb or Iron tests between the two treatment arms. An unstructured correlation structure will be specified (e.g., by way of the appropriate random statement within the SAS® procedure GLIMMIX) to model the within-subjects errors across unscheduled visits. If the model fails to converge, alternative correlation structures will be examined. The model will fit the natural logarithm of the mean number of repeat Hb or Iron tests to a linear combination of covariates (i.e., a log-linear model). The covariates will include: the fixed effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup; and, random effects of time (elapsed time between consecutive unscheduled visits by subject) and treatment-by-time interaction. A summary of the number of repeat Hb or Iron tests will be provided using descriptive statistics by treatment arm and overall. The summary will also include the GLMM LSM estimate of repeat Hb or Iron tests and its 95% CI by treatment arm, the GLMM estimate of the ratio (relative risk) of mean repeat Hb or Iron tests between the two treatment arms and corresponding 95% confidence interval for the ratio, and the p-value from the test of null hypothesis that the mean repeat Hb or Iron tests are the same between the two treatment arms.

The method of transportation (Public, Own transport, Hospital provided, Unknown) will be documented at each visit (scheduled or unscheduled) over the Treatment Period. For each scheduled visit, a multinomial logistic regression modelling approach will be used to estimate the visit-specific relative odds of taking a given transportation method to that of Public (i.e., Public transportation will serve as the reference level) between the two treatment arms, adjusted for the effects of baseline Hb (below the observed median or at least the observed median) and IBD subgroup. The "Unknown" method of transportation will be considered as a missing value of method of transportation for purposes of the modelling described here. A summary of Public transportation will be presented using descriptive statistics by visit (including a visit labelled "unscheduled"). The summary will also present, by scheduled visit, the estimated relative odds of taking a given transportation method to that of Public between the two treatment arms, the corresponding 95% CI for the odds ratio, and the p-value from testing the null hypothesis of no association between treatment arm and chosen transportation method.

A data listing of correction and maintenance of IDA will be provided by subject.

# 11 SAFETY ANALYSES

Unless otherwise noted, the Safety Population will be used for all analyses, summaries, and listings of safety endpoints.

## 11.1 Adverse Events

An AE is any unfavourable or unintended sign, symptom, or disease temporally associated with the use of an investigational product or protocol-imposed intervention, regardless of attribution. All observed or volunteered AEs regardless of suspected causal relationship to the investigational product(s) will be reported.

All reported terms (investigator descriptions) for AEs will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) and summarised with frequencies and percentages by treatment group, system organ classification, and preferred term.

The number, intensity, relation to study medication and action taken will be described by incidence tables. Subjects with multiple intensities for the same AE will be categorized by the maximum intensity experienced. Subjects with instances of the same AE recorded as both related and unrelated to study medication will be categorized as related. Subjects with multiple actions taken within the same AE will be categorized according to the worst case. Actions taken are discontinued, interrupted, dose reduced, and none (listed from worst case to best case).

The most relevant AE dataset will be the treatment emergent adverse effects (TEAEs), defined as those that occur on or after the day of first dose of study medication and up to 14 days after the last dose of study medication. Multiple instances of the TEAE in each system organ class and multiple occurrences of the same preferred term will be counted only once per subject. When an adverse event occurs more than once for a subject, the maximum severity, causality, and action will be used. The number and percentage of subjects in the following categories will be summarised by group and overall: subjects with AE, TEAE, SAE, Deaths, TESAE, TE Deaths, TEAE related to study treatment, TESAE related to study treatment, and TE deaths related to study treatment.

The following tables will also be presented by treatment arm:

- All TEAEs (including TEAEs by intensity, relation to study medication, and action taken)
- Serious TEAEs
- AEs leading to premature discontinuation of study medication.

Separate data listings will be presented for all AEs, all SAEs, and all deaths.

## 11.2 Laboratory Evaluations

Clinical laboratory evaluations include the haematology, iron marker, basic

chemistry, liver function, and urine pregnancy test results (for all females of childbearing potential; data will be provided in a data listing).

Laboratory values, change from baseline (haematology and iron markers only) will be summarised using descriptive statistics by category, laboratory test, and visit for each treatment arm and overall.

Laboratory data, including normal ranges and abnormal laboratory flags, will be provided in data listings.

### 11.3 Vital Signs

Vital signs (blood pressure and heart rate), along with height and weight, will be collected at screening and will be summarised using descriptive statistics. Vital signs, height, and weight will be provided in a data listing by subject.

### 11.4 Physical Examination

Findings from the screening physical examination will be listed by subject.

## 12 PRIMARY ANALYSIS

Not applicable

## 13 FINAL ANALYSIS

The final, complete analysis will be conducted after the last subject completes or discontinues the study and the resulting clinical database has been cleaned, quality checked, and locked. The assignment of subjects to the analysis populations, as described in Section 5, will be the same as in the primary analysis.
The final analysis will then be performed and will include all safety data and efficacy results up-to-and including Follow-up. Results of this analysis will be reported in the CSR.

## 14 CHANGES FROM PROTOCOL

The Safety population is defined in the protocol as all subjects who have had at least one dose of study drug and one subsequent contact with the Investigator. This has been updated so that the safety population will only consist of subject who received at least one dose of a study drug for consistency with ICH definitions.

A two stage analysis of the safety and efficacy data was detailed in the protocol, with the first stage being conducted when all subjects have completed 12 weeks of treatment, and the second stage being the final analysis. Due to the shortening of the study period, it is anticipated that when all patients have completed 12 weeks

of treatment there will be no patients ongoing in the study, and so only one analysis will be performed.

Other pharmacoeconomic endpoints have been specified here as they are not explicitly mentioned in the protocol. These are:

- Quality of life as assessed by the Medical Outcomes Study 36-item short form questionnaire.
- Correction and Maintenance of IDA
    - Repeat Hb or iron testing
    - Method of transportation to attend each visit.
    - The number of days off work or school/education to attend hospital or clinic visits for administration of IV Iron.
    - The time spent in the hospital or clinic receiving IV Iron.
    - The dose of IV Iron administered.

## 15 REFERENCES

1. Agresti A (2002). Categorical data analysis, 2nd edition. New York: Wiley.
2. D'Agnostino RB, Belanger A, D'Agnostino RB Jr. (1990). A suggestion for using powerful informative tests of normality. *The American Statistician*; 35(3).
3. Glaser RE (1982). Levene's robust test of homogeneity of variances. Encyclopaedia of Statistical Sciences. New York: John Wiley and Sons; 608-610.
4. Ware JE, Kosinski M, Dewey JE (2000). How to score version 2 of the SF-36 Health Survey, Quality Metric Incorporated.

# 16 TABLE OF CONTENTS FOR DATA DISPLAYS

## 16.1 Tables

Study Population


| | |
|---|---|
| Table 14.1.1.1 | Summary of Analysis Populations – All Randomized Subjects |
| Table 14.1.1.2 | Summary of Subject Disposition – All Screened Subjects |
| Table 14.1.1.3.1 | Summary of Protocol Deviations in the First 12 Weeks – ITT Population |
| Table 14.1.1.3.2 | Summary of Protocol Deviations After the First 12 Weeks – ITT Population |
| Table 14.1.1.3 | Summary of Protocol Deviations – ITT Population |
| Table 14.1.2.1 | Summary of Demographic Characteristics – ITT Population |
| Table 14.1.2.2 | Summary of Demographic Characteristics – PP Population |
| Table 14.1.3.1 | Summary of Medical History – ITT Population |
| Table 14.1.3.2 | Summary of Medical History – PP Population |


Efficacy/PE


| | |
|---|---|
| Table 14.2.1.1 | Summary of Haemoglobin Concentration (g/dL) – ITT Population |
| Table 14.2.1.2 | Summary of Haemoglobin Concentration (g/dL) – PP Population |
| Table 14.2.1.3 | Summary of Haemoglobin Concentration (g/dL) in Subjects with Baseline Hb < 9.5 g/dL – ITT Population |
| Table 14.2.1.4 | Summary of Haemoglobin Concentration (g/dL) in Subjects with Baseline Hb < 9.5 g/dL – PP Population |
| Table 14.2.2.1 | Summary of Haemoglobin Responder Rate – ITT Population |
| Table 14.2.2.2 | Summary of Haemoglobin Responder Rate – PP Population |
| Table 14.2.2.3 | Statistical Analysis of Haemoglobin Responder Rate Using a Repeated Measures Approach – ITT Population |
| Table 14.2.2.4 | Statistical Analysis of Haemoglobin Responder Rate Using a Repeated Measures Approach – PP Population |
| Table 14.2.3.1 | Summary of Rise in Haemoglobin Responder Rate – ITT Population |
| Table 14.2.3.2 | Summary of Rise in Haemoglobin Responder Rate – PP Population |
| Table 14.2.3.3 | Summary of Haemoglobin Rise Responder in Subjects with Baseline Hb < 9.5 g/dL – ITT Population |
| Table 14.2.3.4 | Summary of Haemoglobin Rise Responder in Subjects with Baseline Hb < 9.5 g/dL – PP Population |
| Table 14.2.4.1 | Summary of Haemoglobin Normal Limit Responder Rate – ITT Population |
| Table 14.2.4.2 | Summary of Haemoglobin Normal Limit Responder Rate – PP Population |
| Table 14.2.4.3 | Summary of Haemoglobin Normal Limit Responder Rate in Subjects with Baseline Hb < 9.5 g/dL – ITT Population |

| | |
|---|---|
| Table 14.2.4.4 | Summary of Haemoglobin Normal Limit Responder Rate in Subjects with Baseline Hb < 9.5 g/dL – PP Population |
| Table 14.2.5.1 | Summary of Anaemic Status – ITT Population |
| Table 14.2.5.2 | Summary of Anaemic Status – PP Population |
| Table 14.2.6.1 | Summary of Serum Ferritin Concentration (ng/ml) – ITT Population |
| Table 14.2.6.2 | Summary of Serum Ferritin Concentration (ng/ml) – PP Population |
| Table 14.2.7.1 | Summary of Ferritin Normal Limit Responder Rate – ITT Population |
| Table 14.2.7.2 | Summary of Ferritin Normal Limit Responder Rate – PP Population |
| Table 14.2.8.1 | Summary of Medical Outcomes Study SF-36 Scores – ITT Population |
| Table 14.2.8.2 | Summary of Change from Baseline in Medical Outcomes Study SF-36 Scores – ITT Population |
| Table 14.2.9.1 | Summary of Administration of IV Iron – ITT Population |
| Table 14.2.9.2 | Summary of IV Iron Dosing Details – ITT Population |
| Table 14.2.10 | Summary of Number of Repeat Haemoglobin or Iron Tests – ITT Population |
| Table 14.2.11 | Summary of Method of Transportation by Visit – ITT Population |

<u>Safety</u>

| | |
|---|---|
| Table 14.3.1.1 | Overall Summary of Adverse Events |
| Table 14.3.1.2.1 | Summary of Treatment-Emergent Adverse Events |
| Table 14.3.1.2.2 | Summary of Treatment-Emergent Adverse Events by Maximum Intensity |
| Table 14.3.1.2.3 | Summary of Treatment-Emergent Adverse Events by Relation to Study Medication |
| Table 14.3.1.2.4 | Summary of Treatment-Emergent Adverse Events by Action Taken |
| Table 14.3.1.3.1 | Summary of Study Drug Exposure |
| Table 14.3.1.3.2 | Summary of Ferric Maltol Compliance by Visit |
| Table 14.3.2.1 | Summary of Treatment-Emergent Serious Adverse Events |
| Table 14.3.2.2 | Summary of Adverse Events Leading to Premature Discontinuation of Study Medication |
| Table 14.3.3.1 | Summary of Prior Medications |
| Table 14.3.3.2 | Summary of Concomitant Medications |
| Table 14.3.4.2.1.1 | Summary of Laboratory Evaluations (Haematology) |
| Table 14.3.4.2.1.2 | Summary of Laboratory Evaluations (Basic Chemistry) |
| Table 14.3.4.2.1.3 | Summary of Laboratory Evaluations (Clinical Chemistry) |
| Table 14.3.4.2.1.4 | Summary of Laboratory Evaluations (Liver Function) |
| Table 14.3.4.3 | Summary of Vital Signs at Screening |

## 16.2 Figures

| | |
|---|---|
| Figure 14.2.1.1 | Haemoglobin Concentration (g/dL) at Week 12 Versus Baseline – ITT Population |
| Figure 14.2.1.2 | Mean Change from Baseline Haemoglobin Concentration (g/dL) by Visit – ITT Population |
| Figure 14.2.1.3 | Haemoglobin Responders by Visit – ITT Population |
| Figure 14.2.1.4 | Haemoglobin Concentration (g/dL) at Week 12 Versus Baseline – PP Population |
| Figure 14.2.1.5 | Mean Change from Baseline Haemoglobin Concentration (g/dL) by Visit – PP Population |
| Figure 14.2.1.6 | Haemoglobin Responders by Visit – PP Population |
| Figure 14.2.1.7 | Haemoglobin Concentration (g/dL) at Week 24 Versus Baseline – ITT Population |
| Figure 14.2.1.8 | Haemoglobin Concentration (g/dL) at Week 24 Versus Baseline – PP Population |
| Figure 14.2.1.9 | Haemoglobin Concentration (g/dL) at Week 36 Versus Baseline – ITT Population |
| Figure 14.2.1.10 | Haemoglobin Concentration (g/dL) at Week 36 Versus Baseline – PP Population |
| Figure 14.2.1.11 | Haemoglobin Concentration (g/dL) at Week 52 Versus Baseline – ITT Population |
| Figure 14.2.1.12 | Haemoglobin Concentration (g/dL) at Week 52 Versus Baseline – PP Population |
| Figure 14.2.1.13 | Haemoglobin Responders (1g/dL) by Visit – ITT Population |
| Figure 14.2.1.14 | Haemoglobin Responders (1g/dL) by Visit – PP Population |
| Figure 14.2.2.1 | Ferritin Concentration (ng/mL) at Week 24 Versus Baseline – ITT Population |
| Figure 14.2.2.2 | Ferritin Concentration (ng/mL) at Week 24 Versus Baseline – PP Population |
| Figure 14.2.2.3 | Ferritin Concentration (ng/mL) at Week 24 Versus Baseline – ITT Population |
| Figure 14.2.2.4 | Ferritin Concentration (ng/mL) at Week 24 Versus Baseline – PP Population |
| Figure 14.2.2.5 | Ferritin Concentration (ng/mL) at Week 36 Versus Baseline – ITT Population |
| Figure 14.2.2.6 | Ferritin Concentration (ng/mL) at Week 36 Versus Baseline – PP Population |
| Figure 14.2.2.7 | Ferritin Concentration (ng/mL) at Week 52 Versus Baseline – ITT Population |
| Figure 14.2.2.8 | Ferritin Concentration (ng/mL) at Week 52 Versus Baseline – PP Population |

## 16.3 Listings

Study Population

| | |
|---|---|
| Listing 16.2.1.1 | Listing of Subject Disposition |
| Listing 16.2.1.2.1 | Listing of Scheduled Visit Dates |
| Listing 16.2.1.2.2 | Listing of Unscheduled Visits |
| Listing 16.2.2.1 | Listing of Informed Consent |
| Listing 16.2.2.2 | Listing of Inclusion/Exclusion Criteria Deviations |
| Listing 16.2.2.3 | Listing of Protocol Deviations |
| Listing 16.2.3.1 | Listing of Subject Randomization |
| Listing 16.2.3.2 | Listing of Analysis Populations |
| Listing 16.2.4.1.1 | Listing of Demographics |
| Listing 16.2.4.1.2 | Listing of Reproductive Status |
| Listing 16.2.4.1.3 | Listing of IBD Status |
| Listing 16.2.4.2 | Listing of Medical History |
| Listing 16.2.5.1 | Listing of Study Drug Exposure and Compliance |
| Listing 16.2.5.2 | Listing of Overall Compliance |

Efficacy/PE

| | |
|---|---|
| Listing 16.2.6.1 | Listing of Haemoglobin Results |
| Listing 16.2.6.2 | Listing of Ferritin Results |
| Listing 16.2.6.3 | Listing of Medical Outcomes Study SF-36 Scores |
| Listing 16.2.6.4 | Listing of Correction and Maintenance of IDA |

Safety

| | |
|---|---|
| Listing 16.2.7.1 | Listing of Adverse Events |
| Listing 16.2.7.2 | Listing of Serious Adverse Events |
| Listing 16.2.7.3 | Listing of Deaths |
| Listing 16.2.8.1 | Listing of Laboratory Evaluations (Haematology) |
| Listing 16.2.8.2 | Listing of Laboratory Evaluations (Iron Markers) |
| Listing 16.2.8.3 | Listing of Laboratory Evaluations (Basic Chemistry) |
| Listing 16.2.8.4 | Listing of Laboratory Evaluations (Liver Function) |
| Listing 16.2.8.5 | Listing of Pregnancy Test Results |
| Listing 16.2.9.1 | Listing of Vital Signs at Screening |
| Listing 16.2.9.2 | Listing of Physical Examination at Screening |
| Listing 16.2.9.3 | Listing of Activity Index |
| Listing 16.2.9.4 | Listing of Initial Pregnancy Report |
| Listing 16.2.9.5 | Listing of Follow-up Pregnancy Report |
| Listing 16.2.9.6 | Listing of Prior Medications |
| Listing 16.2.9.7 | Listing of Concomitant Medications |
| Listing 16.2.9.8 | Listing of Medical Procedures Completed During Study |

| | |
|---|---|
| **DOCUMENT:** | Post-hoc Statistical Analysis Plan |
| **PROTOCOL:** | ST10-01-304 |
| | A phase 3b, randomized, controlled, multicentre study with oral ferric maltol (Feraccru) or intravenous iron (ferric carboxy maltose; FCM), for the treatment of iron deficiency anaemia in subjects with inflammatory bowel disease |
| **SAP VERSION:** | Final 1.0 of Post-hoc Statistical Analysis Plan |
| **SAP DATE:** | 23 July 2020 |
| **PROTOCOL DATE:** | 10 July 2015, Version 1.0<br>24 August 2015, Version 2.0<br>12 February 2016, Version 3.0<br>08 August 2016, Version 4.0<br>25 August 2016, Version 4.2<br>05 January 2017, Version 5.0<br>11 July 2017, Version 6.0<br>30 April 2018, Version 7.0 |
| **SPONSOR:** | Shield TX (UK) Limited |
| **PREPARED BY:** | Clinipace Worldwide |
| **AUTHORS:** | Barbara Collins, BSc PhD<br>Senior Biostatistician<br>Clinipace Worldwide |

## APPROVAL SIGNATURES

SIGNATURE: DATE:

Julian Howell — 24-JLY-2020

Dr Julian Howell MBBS, FFPM
Clinical Consultant to:
Shield TX (UK) Ltd

Vasari Roberta — 24-July-2020

Roberta Vasari
Clinical Programme Manager
Shield TX (UK) Ltd

Signed by: Barbara Collins
Senior Statistician
Reason: I am the author of this document
Date & Time: Jul 23 2020 10:44 PM

DocuSign

Barbara Collins, BSc PhD
Snr Biostatistician, Clinipace

**TABLE OF CONTENTS**

**LIST OF ABBREVIATIONS** ..... **4**

**1 INTRODUCTION** ..... **5**

**2 STUDY OBJECTIVES, TREATMENTS, AND ENDPOINTS** ..... **5**

**3 STUDY DESIGN** ..... **5**

**4 POWER AND SAMPLE SIZE CONSIDERATIONS** ..... **5**

**5 ANALYSIS POPULATIONS** ..... **5**

**6 CONSIDERATIONS FOR DATA ANALYSIS** ..... **6**

**7 DATA HANDLING METHODS** ..... **6**

**8 STUDY POPULATION** ..... **7**

**9 EFFICACY ANALYSES** ..... **8**

*9.1 PRIMARY EFFICACY ANALYSES* ..... *8*
*9.1.1 Primary Analysis* ..... *8*
*9.1.2 Sensitivity Analysis* ..... *10*
*9.1.2.1 Observed Cases (OC) Analysis* ..... *10*
*9.1.2.2 Observed Cases (OC) Repeated Measures Analysis* ..... *10*
*9.1.2.3 Week 12 compliance analysis* ..... *11*
*9.1.2.4 Week 12 non-windowed analysis* ..... *11*
*9.1.2.5 Week 12 immunosuppressant analysis* ..... *12*
*9.2 SECONDARY EFFICACY ANALYSES* ..... *12*
*9.2.1 Secondary Analysis* ..... *12*
*9.2.1.1 Key Secondary Endpoints* ..... *12*
*9.2.1.2 Other Secondary Efficacy Endpoints* ..... *13*
*9.2.1.3 Long-term Secondary Endpoints* ..... *14*

**10 PHARMACOECONOMIC ANALYSES** ..... **15**

**11 FINAL ANALYSIS** ..... **17**

**12 SUMMARY OF POST HOC CHANGES** ..... **17**

**13 REFERENCES** ..... **18**

**14 TABLE OF CONTENTS FOR DATA DISPLAYS** ..... **19**

*14.1 TABLES* ..... *19*
*14.2 FIGURES* ..... *20*
*14.3 LISTINGS* ..... *21*

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ANCOVA | Analysis of Covariance |
| CI | Confidence interval |
| CFB | Change from baseline |
| CRF | Case report form |
| CSR | Clinical Study Report |
| FCM | Ferric carboxy maltose |
| GLM | Generalized Linear Model |
| GLMM | Generalized Linear Mixed Model |
| Hb | Haemoglobin |
| IDA | Iron deficiency anaemia |
| IBD | Inflammatory bowel disease |
| ITT | Intention-to-Treat |
| IV | Intravenous |
| LCL | Lower confidence limit |
| LOCF | Last Observation Carried Forward |
| LSM | Least squares mean |
| MH | Mantel-Haenszel |
| OC | Observed Cases |
| PE | Pharmacoeconomic |
| PP | Per Protocol |
| SAP | Statistical Analysis Plan |
| SD | Standard deviation |
| SF-36 | Medical outcomes study 36-item short form |
| UC | Ulcerative colitis |
| WHO | World Health Organization |

# 1 INTRODUCTION

This document describes the statistical methods and data presentations to be used in the revised and post-hoc summary and analysis of efficacy data from Protocol ST10-01-304. The reader is referred to the study protocol and electronic case report forms (eCRFs) for details of study conduct and data collection, and the final statistical analysis plan version 3.0 dated 14 February 2019 for background information and details of the original analysis where applicable.

# 2 STUDY OBJECTIVES, TREATMENTS, AND ENDPOINTS

For details of the study objectives, treatment group comparisons and non-primary endpoints refer to Statistical Analysis Plan v3.0, dated 14 February 2019.

The primary efficacy endpoint is the proportion of subjects achieving either a ≥2g/dL increase in Haemoglobin (Hb) OR normalization of Hb (≥12g/dL for women, ≥13g/dL for men based on the WHO definition (WHO/NMH/NHD/MNM/11.1) at Week 12 [*Post-hoc change from previous SAP, noted in the original CSR, from >12.0 g/dL (women) or 13.0 g/dL (men)*].

Note: the WHO definition of Hb normal limit (≥12g/dL for women, ≥13g/dL for men) was applied in the primary efficacy endpoint but not the secondary efficacy endpoints in the original analysis for CSR v1.0.

In this post-hoc analysis, all secondary endpoints that use normal limits will also be based on the WHO definition.

# 3 STUDY DESIGN

For details of the overall study design and the schedule of study assessments, refer to Statistical Analysis Plan v3.0, dated 14 February 2019.

# 4 POWER AND SAMPLE SIZE CONSIDERATIONS

Refer to Statistical Analysis Plan v3.0, dated 14 February 2019.

# 5 ANALYSIS POPULATIONS

For details of the randomized, safety, and ITT populations, refer to Statistical Analysis Plan v3.0, dated 14 February 2019, Section 5.1 to 5.3.

For details of the PP population, refer to Statistical Analysis Plan v3.0, dated 14 February 2019, Section 5.4 and Section 8 below. To clarify, data of the subjects with major protocol deviations occurring after Week 12 were excluded from the PP analysis of the long-term endpoints from that point of deviation onwards in the original analysis for CSR v1.0.

Patients data are re-evaluated with a focus on visit window, drug compliance, missing data, and the allowance of immunosuppressants. A re-evaluated PP population is defined for the post hoc analyses.

Two sensitivity PP population will be determined for the analysis of the primary efficacy data:

The modified PP population for compliance (mPP compliance) contains all patients from the re-evaluated PP population, along with all patients who are only excluded from the re-evaluated PP population due to compliance issues.

The modified PP population for windowing (mPP windowing) contains all patients from the re-evaluated PP population, along with all patients who are only excluded from the re-evaluated PP population due to Week 12 windowing issues for haemoglobin.

Data of the subjects with major protocol deviations occurring after Week 12 are excluded from the re-evaluated PP analysis of the long-term endpoints from that point of deviation onwards.

# 6 CONSIDERATIONS FOR DATA ANALYSIS

For details of the programming environment, strata and covariates, subgroups, multiple comparisons and multiplicity, significance level, and statistical notation and methodology, refer to Statistical Analysis Plan v3.0, dated 14 February 2019.

Any analysis based on the PP population in the original SAP will be rerun based on the re-evaluated PP population.

*Post hoc* analyses will be clearly identified as such both in this document and in the Clinical Study Report. Any outputs that are generated post hoc due to re-evaluations of the statistical methods or re-evaluated PP populations will be labelled "Post-Hoc" followed by the month/year run and the table number, Eg "Post Hoc July 2020 Table 14.1.1". Outputs that are updated due to a correction in methods to that described in the Statistical Analysis Plan v3.0, dated 14 February 2019 will be labelled "Revised" followed by the month/year run and the table number, Eg "Revised July 2020 Table 14.1.1".

# 7 DATA HANDLING METHODS

For details of the data handling methods, refer to Statistical Analysis Plan v3.0, dated 14 February 2019.

Missing data for adherence to ferric maltol (i.e. no returned bottles and no pill count) will not be imputed and will not be assumed to be zero adherence or all tablets taken.

The code for implementing the SAS® procedure MI is:

```
PROC MI DATA = <dataset _name> NIMPUTE = 20 SEED = <seed> OUT = MI_out;
  VAR baselineHbtrans gender ibd Week4Hbtrans Week12Hbtrans;
  EM MAXITER = 10000;
RUN;
```

For the multiple imputation, the code for implementing the SAS® procedure MIANALYZE to the CFB Hb concentration at Week 12 secondary endpoint is:

```
DATA CFB_Impute; SET MI_out;  CFBHbW12 = Week12Hb – baselineHb; RUN;

PROC SORT DATA = CFB_Impute; BY _Imputation_; RUN;

PROC REG DATA = CFB_Impute OUTEST = outreg COVOUT;
  BY _Imputation_;
  MODEL CFBHbW12 = binarybaselineHb TreatmentArm IBDGroup;
RUN;

PROC MIANALYZE DATA = outreg EDF = <complete-data degrees of freedom>;
  MODELEFFECTS Intercept binarybaselineHb TreatmentArm IBDGroup;
  ODS OUTPUT ParameterEstimates = parm1;
RUN;
```

The code for implementing the SAS® procedure MIANALYZE to the Hb normal limit responder at Week 12 secondary endpoint for the ITT population is:

```
DATA NormalHb_Impute; SET MI_out;
  NLResponder = 0;
  IF Gender =: 'F' AND Week12Hb => 12.0 THEN NLResponder = 1;
  IF Gender =: 'M' AND Week12Hb => 13.0 THEN NLResponder = 1;
RUN;

PROC SORT DATA = NormalHb_Impute; BY _Imputation_; RUN;

PROC LOGISTIC DATA = NormalHb_Impute;
  MODEL NLResponder = binarybaselineHb TreatmentArm IBDGroup / covb;
  BY _Imputation_;
  ODS OUTPUT ParameterEstimates = lgsparms
             CovB               = lgscovb;
RUN;

PROC MIANALYZE PARMS = lgsparms COVB(effectvar=stacking) = lgscovb;
  MODELEFFECTS Intercept binarybaselineHb TreatmentArm IBDGroup;
RUN
```

## 8 STUDY POPULATION

For details of the listings and summaries for the study population, refer to Statistical Analysis Plan v3.0, dated 14 February 2019.

In addition to the existing categories described in Statistical Analysis Plan v3.0, dated 14 February 2019, additional major protocol deviations include the following categories:

- Repeated non-compliance in self-administration of ferric maltol study drug, where non-compliance is less than 80% or greater than 120% of expected administration of study drug in the first 12 weeks. Where data were missing the compliance may be assessed based on individual visit pill returns, where data exists the compliance would be based on the average compliance in the first 12 weeks of study.
- Subjects that had a Week 12 haemoglobin value measured more than 2 weeks (+/- 14 days) outside the planned Week 12 assessment, ie an additional 1 week window beyond the 1 week window in the schedule of assessments.
- IV Iron subjects who did not follow the specific country dosing regimens. An IV iron dose up to Week 12 that was $\geq 2$ or $\leq .5$ times the calculated dose based on weight and Hb and per the local Prescribing Information, will not be included in the re-evaluated PP population.
- Immunosuppressant use will not be a criterion for exclusion from the re-evaluated PP population, based on protocol version 6, dated 11 July 2017.

Protocol deviations deemed to be major, but not recorded on the database, will be included by means of a spreadsheet which will be agreed by Shield and Clinipace based on the minutes of the subject evaluability meeting.

No imputation was performed for missing data in compliance calculations.

# 9 EFFICACY ANALYSES

## *9.1 Primary Efficacy Analyses*

### 9.1.1 Primary Analysis

Haemoglobin (Hb) concentration will be summarised at screening (Visit 1), pre-randomization at Visit 2 (baseline assessment), and at each visit over the Treatment Period. Hb concentration (Observed Cases [OC]) will be summarised for the ITT (post hoc table 14.2.1.1) and re-evaluated PP (post hoc table 14.2.1.2) populations using descriptive statistics by visit for each treatment arm and overall. A scatterplot of Hb concentrations (OC) at each visit versus baseline will be presented by treatment arm for the re-evaluated PP population (post hoc figures 14.2.1.4, 14.2.1.5.1, 14.2.1.5.2, 14.2.1.6, 14.2.1.8, 14.2.1.10, 14.2.1.12.1, 14.2.1.12.2).

A dichotomous variable, representing response status in the degree to which Hb increases relative to baseline, will be calculated for each visit over the Treatment Period. Those subjects with at least 2g/dL increase in Hb from baseline or who

have normalization of Hb from baseline (≥12.0 g/dL for women, ≥13.0 g/dL for men) will be classified as "Hb responders". [*Post-hoc change from > 12.0 g/dL for women or > 13.0 g/dL for men*]. A shift from baseline of anaemic to non-anaemic status (ie Hb concentration shift from below to within the normal limits will be classified as "Hb normal limit responders". [*Post-hoc change to align Hb normal limit responder definition with the definition given in section 2*]. Patients who are normal at baseline cannot be classified as a responder. Ferric maltol patients who take IV Iron cannot be classified as responders after the intake of IV iron.

The null hypothesis for the primary efficacy endpoint analysis is that the proportion of Hb responders at Week 12 in the ferric maltol treatment arm is at least 20% lower than the proportion of Hb responders at Week 12 in the IV Iron treatment arm (i.e., ferric maltol is inferior to IV Iron for rate of Hb response at Week 12). The alternative hypothesis is that the proportion of Hb responders at Week 12 in the ferric maltol treatment arm is not worse than 20% lower than the proportion of Hb responders at Week 12 in the IV Iron treatment arm (i.e., ferric maltol is non-inferior to IV Iron for rate of Hb response at Week 12).

The proportion of Hb responders at Week 12 will be compared between the two treatment arms, by calculating a 2-sided confidence interval for difference in the proportions of Hb responders (risk difference) between the two treatment arms and comparing the lower confidence limit of this CI to the pre-specified non-inferiority margin. The CI will be calculated using the Delta Method approach [1] based on a logistic regression model, adjusted for treatment arm, baseline Hb (below the observed median or at least the observed median), and IBD subgroup (UC or CD). Note: For the CI calculation in the primary efficacy endpoint, the risk difference model was applied in the original analysis for CSR v1.0. The analysis method will be corrected to the Delta Method approach in the current post hoc analysis.

Post-hoc change: Note that missing values for the Hb responder status at Week 12 primary efficacy analysis will be imputed using a multiple imputation (MI) approach for ITT population. Missing values of Hb concentration at Week 12 will be imputed using the SAS® procedure MI. This will generate 20 complete datasets containing imputed values of Hb concentration at Week 12 for those missing in the original dataset. Variables considered helpful in guiding this imputation are deviation of baseline (Visit 2) Hb concentration from response threshold, deviation of Week 4 Hb concentration from response threshold, gender and IBD subgroup. A separate model will be computed for each treatment arm.

The Week 12 imputed value will then be back transformed from the Week 12 deviation from the response threshold.

Each model, corresponding to a given endpoint analysis, will be fitted against each of the imputed datasets. The SAS® procedure MIANALYZE will then be used to combine the individual analysis results for the given endpoint into one result.

The null hypothesis for the primary efficacy analysis will be rejected (i.e., ferric maltol will be deemed non-inferior to IV Iron), if the LCL of the 95% CI for the

difference (ferric maltol – IV Iron) in proportions of Hb responders at Week 12 is ≥ -20% for both the ITT analysis and the re-evaluated PP analysis. The p-value in testing the aforementioned hypotheses will be equal to one minus the confidence coefficient corresponding to the CI for the risk difference whose LCL equals -20% (i.e., inverting the CI for the risk difference to a test statistic and calculating the p-value from the test statistic).

The proportion of Hb responders will be summarised for the ITT and re-evaluated PP populations using descriptive statistics by visit over the Treatment Period for each treatment arm and overall, utilizing all available Hb responder data (i.e., OC approach). Each summary will also present the proportion of Hb responders for the Week 12 visit (MI) using descriptive statistics, and will include the corresponding logistic regression model adjusted difference in the proportions of Hb responders between the two treatment arms, the corresponding adjusted 95% CI for the difference in proportions, and the adjusted p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. Finally, each summary will also present the estimated logistic regression model adjusted odds ratio of Hb responder at Week 12 (MI) between the two treatment arms and corresponding 95% CI for the odds ratio(post hoc tables 14.2.2.1 and 14.2.2.2). A bar plot of the Hb responder rate will be presented by visit (OC) and treatment arm for each assessment for the re-evaluated PP population (post hoc figure 14.2.1.6).

### 9.1.2 Sensitivity Analysis

Two sensitivity analyses will be performed for the primary efficacy endpoint; in addition, three post-hoc sensitivity analyses will be performed using the re-evaluated PP population.

#### 9.1.2.1 Observed Cases (OC) Analysis

The first sensitivity analysis will include all available assessments of Hb responder and will not exploit imputation of missing Hb responder status (i.e., only Observed Cases [OC] of Hb responder status are considered here). The primary efficacy endpoint analysis will be repeated using the OC approach for the ITT population. The proportion of Hb responders at Week 12 (OC) will be summarised using descriptive statistics for each treatment arm and overall. The summary will include the corresponding logistic regression model adjusted difference in the proportions of Hb responders (risk difference) between the two treatment arms, the adjusted 95% CI for the difference in proportions, and the adjusted p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. The summary will also include the estimated adjusted odds ratio of Hb responder at Week 12 (OC) between the two treatment arms and corresponding 95% CI for the odds ratio. (post hoc table 14.2.2.1)

#### 9.1.2.2 Observed Cases (OC) Repeated Measures Analysis

The second sensitivity analysis for the primary efficacy endpoint will be performed on the ITT and re-evaluated PP populations, in which Hb responder status will be

analysed in a longitudinal analysis using a repeated measures generalized linear mixed model (GLMM) approach (i.e., a repeated measures logistic regression approach) for all available assessments of Hb (i.e., only Observed Cases [OC] of Hb responder status are considered here) over Visits 2 (Day 0; baseline) to 4 (Week 12).

Post-hoc change: The approach will fit a logistic model to the Hb responder dependent variable, where an unstructured working correlation structure will be specified for each arm (e.g., by way of the appropriate random statement within the GLIMMIX SAS® procedure) to model the within-subjects correlation of repeated Hb responder status at Visits 3 and 4. If the model fit fails to converge, an auto-regressive(1) structure will be specified. If this also fails to converge, alternative working correlation structures will be examined. Covariates of the GLMM will include the fixed effects of treatment, baseline Hb (below the observed median or at least the observed median)", IBD subgroup, visit as a categorical variable, and treatment-by-visit interaction. The estimated adjusted odds ratio for treatment effect, and its corresponding 95% CI, from the model will be presented by visit.

To assess the sensitivity of Hb responder status over time – by way of a general linear model (GLM) approach – separate logistic regression models will be fitted to the Hb responder dependent variable at each visit during the Treatment Period up to Week 12, using all available assessments of Hb (i.e., only Observed Cases [OC] of Hb responder status are considered here). Each model will include terms for the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The estimated adjusted odds ratio for treatment effect, and its corresponding 95% CI, will be presented for each model (i.e., by visit). Sensitivity of Hb responder status over time will be assessed by comparing these by-visit treatment effect odds ratio estimates from the logistic regression models (the GLMs) to those obtained from the GLMM model. (post hoc tables 14.2.2.3 and 14.2.2.4)

#### 9.1.2.3 Week 12 compliance analysis

Post hoc change: The first post-hoc sensitivity analysis will be performed on the mPP compliance population (post hoc table 14.2.2.5). The analysis performed will be the same as for the primary efficacy. This analysis is relevant to HEOR and more of the real-world interpretation of the data. In addition, a waterfall plot of overall compliance in the first 12 weeks based on this subset of patients will be included (post hoc figure 14.3.2.3).

#### 9.1.2.4 Week 12 non-windowed analysis

Post hoc change: The second post-hoc sensitivity analysis will be performed on the mPP windowing population (post hoc table 14.2.2.6). Patients will be considered as a responder if they have a haemoglobin response at any point up to and including Week 12 visit. The analysis performed will be the same as for the primary efficacy.

#### 9.1.2.5 Week 12 immunosuppressant analysis

Post hoc change: The third post-hoc sensitivity analysis will be performed on the re-evaluated PP population. Patients will be included into subgroups based on the protocol version they first consented under (prior to version 6, version 6 onwards) (post hoc table 14.2.2.7). The analysis performed for each subgroup will be the same as for the primary efficacy.

## *9.2 Secondary Efficacy Analyses*

### 9.2.1 Secondary Analysis

#### 9.2.1.1 Key Secondary Endpoints

The change from baseline (CFB) in Hb concentration at Week 12 (MI) secondary endpoint will be analysed for the ITT population using an analysis of covariance (ANCOVA) model.

The ANCOVA model will have terms for treatment arm, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The treatment effect of ferric maltol compared to IV Iron will be considered using differences of adjusted means. No formal statistical significance level in the treatment comparison will be defined, as the p-value from the ANCOVA model will be presented for descriptive purposes only. A summary of CFB in Hb concentration (OC) will be provided using descriptive statistics for each visit over the Treatment Period by treatment arm and overall. The summary will include the following ANCOVA model estimates: adjusted means (i.e., Least-Square Means [LSMs]) of CFB in Hb concentration at Week 12 for each treatment arm, the difference in treatment arm LSMs, the corresponding 95% CIs for treatment arm specific LSMs and difference in treatment arm LSMs, and the p-value for the difference in treatment arm LSMs (post hoc table 14.2.2.1). Mean CFB in Hb concentration will be plotted against visit (OC) by treatment arm for the re-evaluated PP populations (post hoc figures 14.2.1.5.1 and 14.2.1.5.2).

A dichotomous variable, representing status of Hb concentration relative to the normal limits of Hb will be calculated for each visit over the Treatment Period. Female subjects with Hb concentration greater than or equal to 12 g/dL and male subjects with Hb concentration greater than or equal to 13 g/dL will be classified to be within normal limits of Hb [*Post-hoc change from >12.0 g/dL (females) or >13.0 g/dL (males)*]. Those subjects with a shift from baseline of anaemic to non-anaemic status (ie Hb concentration shift from below to within the normal limits will be classified as "Hb normal limit responders". [*Post-hoc change to align Hb normal limit responder definition with the definition given in section 2*]. Subjects missing Hb concentration at Week 12 will be assumed to have Hb concentration not within normal limits (i.e., these subjects will be considered Hb normal limit non-responders). The logistic regression modelling and hypothesis testing for the primary analysis (Section 9.1.1) will be repeated here for the Hb normal limit responder at Week 12 secondary endpoint.

A summary will be provided for the Hb normal limit responder secondary endpoint using descriptive statistics by visit over the Treatment Period for each treatment arm and overall, utilizing all available Hb data (i.e., OC approach). The summary will also present the Hb normal limit responder rate for the Week 12 visit (MI) for ITT population using descriptive statistics, and will include the corresponding logistic regression model adjusted difference in proportions of Hb normal limit responders between the two treatment arms, the corresponding [Delta Method derived] adjusted 95% CI for the difference in proportions, and the p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. Finally, the summary will also present the estimated logistic regression model adjusted odds ratio of Hb normal limit responders at Week 12 (MI) between the two treatment arms and corresponding 95% CI for the odds ratio (post hoc tables 14.2.4.1 and 14.2.4.2).
Note: For the CI calculation in the secondary efficacy endpoints, the risk difference model was applied in the original analysis for CSR v1.0. The analysis method will be corrected to the Delta Method approach in the current post hoc analysis.

#### 9.2.1.2 Other Secondary Efficacy Endpoints

The analysis for the key secondary endpoint, CFB in Hb concentration at Week 12 (MI), will be repeated for the secondary endpoints: CFB in Hb concentration at Week 12 (MI) in subjects with baseline Hb <9.5 g/dL; CFB in Hb concentration at Week 4 (MI); and, CFB in Hb concentration at Week 4 (MI) in subjects with a baseline Hb <9.5 g/dL (post hoc tables 14.2.1.3 and 14.2.1.4).

Two dichotomous variables, each representing response for the magnitude rise in Hb from baseline, will be calculated for each visit during the Treatment Period from the Hb concentration data. Those subjects with a rise in Hb from baseline of at least 1g/dL will be classified as "Hb rise responders" for the first variable; those subjects with a rise in Hb from baseline of at least 2g/dL will be classified as "Hb rise responders" for the second variable. Separate logistic regression models will be fitted to each of the Hb rise responders at Week 12 variables (secondary endpoints) using an MI approach for the ITT populations. Each model will include terms for the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. A hypothesis test of the null hypothesis of inferiority for ferric maltol responder rate compared to IV Iron responder rate will be considered for each Hb rise responder secondary endpoint using a non-inferiority margin of 20%. No formal statistical significance level to declare non-inferiority (i.e., rejection of the null hypothesis) of ferric maltol to IV Iron is defined, as Cis and p-values will be presented for descriptive purposes only. The p-value in testing the aforementioned hypotheses will be calculated as described in the primary analysis (Section 9.1.1).

The above analysis will be repeated for the secondary endpoints: proportions of subjects with baseline Hb <9.5 g/dL that achieve an increase in Hb concentration of ≥1 or ≥2 g/dL at Week 12; proportions of subjects who experience a change from baseline in Hb concentration ≥1 or ≥2 g/dL at Week 4 (MI); and, proportions of

subjects with baseline Hb <9.5 g/dL that achieve an increase in Hb concentration of ≥1 or ≥2 g/dL at Week 4 (MI) (post hoc tables 14.2.3.1 to 14.2.3.4). In addition, the above analysis for the Hb normal limit responder at Week 12 key secondary endpoint will be repeated for the secondary endpoints: proportion of subjects with baseline Hb <9.5 g/dL that is within normal limits at Week 12; proportion of subjects with Hb concentration within normal limits at Week 4; and, proportion of subjects with baseline Hb <9.5 g/dL that is within normal limits at Week 4 (MI) (post hoc tables 14.2.4.1 to 14.2.3.4).

A bar plot of the Hb responder rate (1g/dL) will be presented by visit (OC) and treatment arm for each assessment for the re-evaluated PP population (post hoc figure 14.2.1.14).

Serum ferritin will be collected at screening (Visit 1), pre-randomization at Visit 2 (baseline assessment), and at each visit over the Treatment Period. The above ANCOVA analysis will be repeated for the CFB in serum ferritin at Week 12 (MI) secondary endpoint. A summary of serum ferritin and CFB in serum ferritin (OC) will be provided using descriptive statistics for each visit (CFB will be presented at each visit during the Treatment Period) by treatment arm. The summary will include the following ANCOVA model estimates: adjusted means (i.e., Least Squares Means [LSMs]) of CFB in serum ferritin at Week 12 for each treatment arm, the difference in treatment arm LSMs, the corresponding 95% Cis for treatment arm specific LSMs and difference in treatment arm LSMs, and the p-value for significance of the difference in treatment arm LSMs. Normality and homogeneity of variances for the residuals of each ANCOVA model will be assessed using the procedure outlined in Section 9.2.1.1. (post hoc table 14.2.6.2).

#### 9.2.1.3 Long-term Secondary Endpoints

A dichotomous variable, representing anaemia status (anaemic versus non-anaemic), will be calculated for each visit over the Treatment Period. Female subjects with Hb levels greater than or equal to 12.0 g/dL and male subjects with Hb levels greater than or equal to 13.0 g/dL will be classified as non-anaemic [*Post-hoc change from >12.0 g/dL (females) or >13.0 g/dL (males)*]. Separate logistic regression models will be fitted to the long-term secondary endpoints, anaemia status at each visit over Visits 5-7 (OC), for the ITT and re-evaluated PP populations. Each model will include terms to capture the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The above hypothesis testing for Hb normal limit responder (Section 9.2.1.1) will be repeated here for the anaemia status secondary endpoint (revised table 14.2.5.1 and post hoc table 14.2.5.2).

A dichotomous variable, representing status of serum ferritin level relative to the normal limits of serum ferritin, will be calculated for each visit over the Treatment Period. Female subjects with serum ferritin between 10 and 291 ng/ml and male subjects with serum ferritin between 22 and 322 ng/ml will be classified to be within normal limits of serum ferritin. Those subjects with serum ferritin within the

normal limits will be classified as "ferritin normal limit responders". Separate logistic regression models will be fitted to the long-term secondary endpoints, ferritin normal limit responder at each visit over Visits 5-7 (OC), for the ITT and re-evaluated PP populations. Each model will include terms to capture the effects of treatment, baseline Hb (below the observed median or at least the observed median), and IBD subgroup. The above hypothesis testing for rise in Hb responder (Section 9.2.1.2) will be repeated here for the serum ferritin normal limit responder secondary endpoint. (post hoc table 14.2.7.2)

A scatterplot of serum ferritin concentrations (OC) at each visit versus baseline will be presented by treatment arm for the re-evaluated PP population (post hoc figures 14.2.2.2, 14.2.2.4, 14.2.2.6, 14.2.2.8.1, 14.2.2.8.2, 14.2.2.10.1, 14.2.2.10.2).

Separate summaries will be provided for each of the binary long-term secondary endpoint variables defined above. Each secondary endpoint will be summarised using descriptive statistics by visit over the Treatment Period for each treatment arm and overall, utilizing all available data (i.e., OC approach). Each summary will also present the corresponding logistic regression model adjusted difference in proportions (risk difference) of secondary endpoint Hb responders between the two treatment arms, the corresponding [Delta Method derived] adjusted 95% CI for the difference in proportions, and the p-value from testing the null hypothesis of inferiority in risk difference with non-inferiority margin 20%. Finally, each summary will also present the appropriate estimated logistic regression model adjusted odds ratio of secondary endpoint at Visits 5-7 (OC) between the two treatment arms and corresponding 95% CI for the odds ratio.
Note: For the CI calculation in the long term secondary efficacy endpoint, the risk difference model was applied in the original analysis for CSR v1.0. The analysis method will be corrected to the Delta Method approach in the current post hoc analysis.

Any patients who are in the re-evaluated PP population but have major deviations after Week 12 will be excluded from the re-evaluated PP analyses of the long-term secondary endpoints from the date of the major deviation onwards.

## 10 PHARMACOECONOMIC ANALYSES

For details of the pharmacoeconomic analyses, refer to Statistical Analysis Plan v3.0, dated 14 February 2019.

The IV iron quantity that should have been prescribed was also calculated based on Weight and Hb and the local FCM Prescribing Information

Post hoc analysis: The total amount of IV Iron taken by patients who completed 52 weeks of treatment will be summarised for the IV Iron group using descriptive statistics on the safety population. (post hoc table 14.2.12)

Post hoc analysis: A Kaplan-Meier plot of the time to first additional IV Iron will be provided for both ITT and re-evaluated PP population. The time to first additional IV

Iron will be the time from first treatment to first additional or rescue IV iron taken on or after Visit 4. Patients originally consented under amendment 7 will be excluded from this plot. Patients who do not receive additional or rescue IV Iron will be censored at the end of study visit. (post hoc figure 14.3.2.2)

Post hoc listing: A listing for all patients detailing the amount of IV Iron up to Week 12 in subjects in the IV Iron group will be provided, detailing whether patients received the country specific amounts up to Week 12 and also if they received more than the maximum weekly amount at any point up to Week 12. A summary of this data will be provided, indicating the number of patients who were dosed as planned, those who were not but were considered minor deviations, and those who were not but were considered major deviations (post hoc table 14.2.13).

Country specific guidance is detailed below:

| Country | Hb (g/dl) | Patient weight | maximum by weight | Maximum weekly dose | Maximum recommended dose for a course |
|---|---|---|---|---|---|
| **Germany** | <10 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| | ≥10 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1000 mg |
| | <10 | over 70 kg | 20 mg/kg | 1000 mg | 2000 mg |
| | ≥10 | over 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| **Belgium** | <10 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| | ≥10 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1000 mg |
| | <10 | ≥ 70 kg | 20 mg/kg | 1000 mg | 2000 mg |
| | ≥10 | ≥ 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| **France** | <10 | < 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| | ≥10 | < 70 kg | 20 mg/kg | 1000 mg | 1000 mg |
| | <10 | ≥ 70 kg | 20 mg/kg | 1000 mg | 2000 mg |
| | ≥10 | ≥ 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| **Spain** | <10 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| | ≥10 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1000 mg |
| | <10 | ≥ 70 kg | 20 mg/kg | 1000 mg | 2000 mg |
| | ≥10 | ≥ 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| **Hungary** | <10 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| | 10-< 14 | between 35 and 70 kg | 20 mg/kg | 1000 mg | 1000 mg |
| | ≥14 | between 35 and 70 kg | 20 mg/kg | 500 mg | 500 mg |
| | <10 | ≥ 70 kg | 20 mg/kg | 1000 mg | 2000 mg |

| | | | | | |
|---|---|---|---|---|---|
| | 10-< 14 | ≥ 70 kg | 20 mg/kg | 1000 mg | 1500 mg |
| | ≥14 | ≥ 70 kg | 20 mg/kg | 500 mg | 500 mg |
| **USA** | | ≥ 50 kg | | 750 mg | 1500mg in 2 doses separated 7 days apart. 750 mg x2 |
| | | <50 kg | 15 mg/kg | 15 mg/kg | 15 mg/kg 2 doses separated by 7 days |

## 11 FINAL ANALYSIS

The corrections and updates to the analyses will be performed and will include data as described above. Results of this analysis will be reported in version 2 of the CSR.

## 12 SUMMARY OF POST HOC CHANGES

The definition of Hb normalisation, for Hb responder, normal limit responder and anaemic response has been updated to align with the WHO definition.

The visit window around the haemoglobin evaluation at week 12 has been extended from 1 week, as in the schedule of visits, to 2 weeks, as per protocol Version 1.

IV Iron subjects who did not follow the specific country dosing regimens were considered to be major protocol deviations. (Based on at least 2x deviation from recommended dosing)

Dose changes or initiation of immunosuppressants was permitted in Protocol Version 6 and later; this rule was applied to all patients included in earlier Protocol versions.

Protocol deviations deemed to be major, but not recorded on the database, will be included by means of a spreadsheet which will be agreed by Shield and Clinipace based on the minutes of the subject evaluability meeting

The following summaries and analyses were included post-hoc:

The total amount of IV Iron taken by patients who completed 52 weeks of treatment will be summarised for the IV Iron group using descriptive statistics.

A Kaplan-Meier plot of the time to first additional IV Iron will be provided for both ITT and re-evaluated PP population. The time to first additional IV Iron will be the time from first treatment to first additional or rescue IV iron taken on or after Visit 4. Patients originally consented under amendment 7 will be excluded from this plot. Patients who do not receive additional or rescue IV Iron will be censored at the end of study visit.

A listing detailing the amount of IV Iron up to Week 12 in subjects in the IV Iron group will be provided, detailing whether patients received the country specific amounts up to Week 12 and also if they received more than the maximum weekly amount at any point up to Week 12. A summary of this data will be provided, indicating the number of patients who were dosed as planned, those who were not but were considered minor deviations, and those who were not but were considered major deviations.

A sensitivity analysis will be performed on the mPP compliance population. The analysis performed will be the same as for the primary efficacy. This analysis is relevant to HEOR and more of the real-world interpretation of the data. In addition, a waterfall plot of overall compliance in the first 12 weeks based on this subset of patients will be included.

A sensitivity analysis will be performed on the mPP windowing population containing all patients from the re-evaluated PP population, along with all patients who are only excluded from the re-evaluated PP population due to haemoglobin windowing issues will be included. Patients will be considered as a responder if they have a haemoglobin response at any point up to the Week 12 visit. The analysis performed will be the same as for the primary efficacy.

A sensitivity analysis will be performed on the re-evaluated PP population. Patients will be included into subgroups based on the protocol version they first consented under (prior to version 6, version 6 onwards). The analysis performed for each subgroup will be the same as for the primary efficacy.

## 13 REFERENCES

1. Agresti A (2002). Categorical data analysis, 2nd edition. New York: Wiley.

# 14 TABLE OF CONTENTS FOR DATA DISPLAYS

## *14.1 Tables*

Study Population

| | |
|---|---|
| Post hoc Table 14.1.1.1 | Summary of Analysis Populations – All Randomized Subjects |
| Post hoc Table 14.1.1.3.1 | Summary of Protocol Deviations in the First 12 Weeks – ITT Population |
| Post hoc Table 14.1.1.3.2 | Summary of Protocol Deviations After the First 12 Weeks – ITT Population |
| Post hoc Table 14.1.2.2 | Summary of Demographic Characteristics – Re-evaluated PP Population |
| Post hoc Table 14.1.3.2 | Summary of Medical History – Re-evaluated PP Population |

Efficacy/PE

| | |
|---|---|
| Post hoc Table 14.2.1.1 | Summary of Haemoglobin Concentration (g/dL) – ITT Population |
| Post hoc Table 14.2.1.2 | Summary of Haemoglobin Concentration (g/dL) – Re-evaluated PP Population |
| Post hoc Table 14.2.1.3 | Summary of Haemoglobin Concentration (g/dL) in Subjects with Baseline Hb < 9.5 g/dL – ITT Population |
| Post hoc Table 14.2.1.4 | Summary of Haemoglobin Concentration (g/dL) in Subjects with Baseline Hb < 9.5 g/dL – Re-evaluated PP Population |
| Post-hoc Table 14.2.2.1 | Summary of Haemoglobin Responder Rate – ITT Population |
| Post hoc Table 14.2.2.2 | Summary of Haemoglobin Responder Rate – Re-evaluated PP Population |
| Post hoc Table 14.2.2.3 | Statistical Analysis of Haemoglobin Responder Rate Using a Repeated Measures Approach – ITT Population |
| Post hoc Table 14.2.2.4 | Statistical Analysis of Haemoglobin Responder Rate Using a Repeated Measures Approach – Re-evaluated PP Population |
| Post hoc Table 14.2.2.5 | Summary of Haemoglobin Responder Rate – mPP Compliance Population |
| Post hoc Table 14.2.2.6 | Summary of Haemoglobin Responder Rate – mPP Windowing Population |
| Post hoc table 14.2.2.7 | Summary of Haemoglobin Responder Rate – Protocol Version Subgroup - Re-evaluated PP Population |
| Post-hoc Table 14.2.3.1 | Summary of Rise in Haemoglobin Responder Rate – ITT Population |
| Post-hoc Table 14.2.3.2 | Summary of Rise in Haemoglobin Responder Rate – Re-evaluated PP Population |

| | |
|---|---|
| Post-hoc Table 14.2.3.3 | Summary of Haemoglobin Rise Responder in Subjects with Baseline Hb < 9.5 g/dL – ITT Population |
| Post-hoc Table 14.2.3.4 | Summary of Haemoglobin Rise Responder in Subjects with Baseline Hb < 9.5 g/dL – Re-evaluated PP Population |
| Post-hoc Table 14.2.4.1 | Summary of Haemoglobin Normal Limit Responder Rate – ITT Population |
| Post-hoc Table 14.2.4.2 | Summary of Haemoglobin Normal Limit Responder Rate – Re-evaluated PP Population |
| Post-hoc Table 14.2.4.3 | Summary of Haemoglobin Normal Limit Responder Rate in Subjects with Baseline Hb < 9.5 g/dL – ITT Population |
| Post-hoc Table 14.2.4.4 | Summary of Haemoglobin Normal Limit Responder Rate in Subjects with Baseline Hb < 9.5 g/dL – Re-evaluated PP Population |
| Revised Table 14.2.5.1 | Summary of Anaemic Status – ITT Population |
| Post-hoc Table 14.2.5.2 | Summary of Anaemic Status – Re-evaluated PP Population |
| Post-hoc Table 14.2.6.2 | Summary of Serum Ferritin Concentration (ng/ml) – Re-evaluated PP Population |
| Post-hoc Table 14.2.7.2 | Summary of Ferritin Normal Limit Responder Rate – Re-evaluated PP Population |
| Post hoc Table 14.2.12 | Total Amount of IV Iron in Patients who Completed 52 Weeks of Study – Safety Population |
| Post hoc Table 14.2.13 | Planned vs Actual Amount of IV Iron up to Week 12 – Safety Population |

## *14.2 Figures*

| | |
|---|---|
| Post hoc Figure 14.2.1.4 | Haemoglobin Concentration (g/dL) at Week 12 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.1.5.1 | Mean Change from Baseline Haemoglobin Concentration (g/dL) by Visit – Re-evaluated PP Population |
| Post hoc Figure 14.2.1.5.2 | Mean Change from Baseline Haemoglobin Concentration (g/dL) by Visit with Early Termination – Re-evaluated PP Population |
| Post hoc Figure 14.2.1.6 | Haemoglobin Responders by Visit – Re-evaluated PP Population |
| Post hoc Figure 14.2.1.8 | Haemoglobin Concentration (g/dL) at Week 24 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.1.10 | Haemoglobin Concentration (g/dL) at Week 36 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.1.12.1 | Haemoglobin Concentration (g/dL) at Week 52 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.1.12.2 | Haemoglobin Concentration (g/dL) at Week 52/Early Termination Versus Baseline – Re-evaluated PP Population |

| | |
|---|---|
| Post hoc Figure 14.2.1.14 | Haemoglobin Responders (1g/dL) by Visit – Re-evaluated PP Population |
| Post hoc Figure 14.2.2.2 | Ferritin Concentration (ng/mL) at Week 12 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.2.4 | Ferritin Concentration (ng/mL) at Week 24 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.2.6 | Ferritin Concentration (ng/mL) at Week 36 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.2.8.1 | Ferritin Concentration (ng/mL) at Week 52 Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.2.8.2 | Ferritin Concentration (ng/mL) at Week 52/Early Termination Versus Baseline – Re-evaluated PP Population |
| Post hoc Figure 14.2.2.10.1 | Mean Change from Baseline Ferritin Concentration (ng/mL) by Visit – Re-evaluated PP Population |
| Post hoc Figure 14.2.2.10.2 | Mean Change from Baseline Ferritin Concentration (ng/mL) by Visit with Early Termination – Re-evaluated PP Population |
| Post hoc Figure 14.3.2.2 | Time to First Additional IV Iron – Re-evaluated PP Population |
| Post hoc Figure 14.3.2.3 | Waterfall Plot of Compliance in the First 12 Weeks – Safety Population |

### *14.3 Listings*

<u>Study Population</u>

| | |
|---|---|
| Post hoc Listing 16.2.2.3 | Listing of Protocol Deviations |
| Post hoc Listing 16.2.3.2 | Listing of Analysis Populations |
| Post Hoc Listing 16.2.5.4 | Listing of Planned vs IV Administration In First 12 Weeks |

<u>Efficacy/PE</u>

| | |
|---|---|
| Post hoc Listing 16.2.6.1 | Listing of Haemoglobin Results |
| Post hoc Listing 16.2.6.2 | Listing of Ferritin Results |
| Post hoc Listing 16.2.6.4 | Listing of Correction and Maintenance of IDA |