Statistical Analysis Plan

Trial IDs:

P-Monofer-CKD-04 and P-Monofer-IDA-03

Phase III Trials

Author: [redacted]
Principal Statistician
[redacted]

Signature page

The signatures below confirm that the signees have read, understood and approved of the contents of the present statistical analysis plan.

Trial Statistician:

(Professional Title)

Signature:	Print Name:	Date:

CMO:

(Professional Title)

Signature:	Print Name:	Date:

Biostatistician:

(Professional Title)

Signature:	Print Name:	Date:

Table of Contents

List of Tables

1 List of Abbreviations

ADR	Adverse Drug Reaction
AE	Adverse Event
CEAC	Clinical Endpoint Adjudication Committee
CI	Confidence Interval
CKD	Chronic Kidney Disease
CRF	Case Report Form
CSR	Clinical Trial Report
DBL	Data Base Lock
ECG	Electrocardiogram
eGFR	Estimated Glomerular Filtration Rate
ESA	Erythropoiesis Stimulating Agents
FACIT	Functional Assessment of Chronic Illness Therapy
FAS	Full Analysis Set
IDA	Iron Deficiency Anaemia
ITT	Intention-To-Treat
Hb	Haemoglobin
LLOQ	Lower Limit of Quantification
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measurements
MNAR	Missing Not at Random
N	Number of Subjects
NDD	Non-Dialysis-Dependent
PP	Per Protocol
PT	Preferred Term
REML	Restricted Maximum Likelihood
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAR	Serious Adverse Reaction
SD	Standard Deviation
SOC	System Organ Class
SMQ	Standardised Medical Dictionary for Regulatory Activities Query
SUSAR	Suspected Unexpected Serious Adverse Reaction
TEAE	Treatment Emergent Adverse Event
TSAT	Transferrin Saturation

2 Introduction

The statistical analysis plan (SAP) is a generic SAP for the two trials; P-Monofer-CKD-04, based on the final protocol 4.0 dated 14 June 2017 and P-Monofer-IDA-03, based on the final protocol 3.0 dated 14 June 2017.

The SAP describes in detail the analyses to be conducted and highlights any deviations from the analysis described in the protocols (see section **Error! Reference source not found.**). Deviations from methods described in this SAP, if any, will be specified in the clinical trial reports (CSRs).

All endpoints, analysis and presentation of data will be handled the same way for the two trials, unless otherwise specified. Any differences in endpoints, analysis or presentations will be clearly described.

Before releasing data for final analysis, one or more data review and classification meetings will be held to classify subjects with respect to analysis populations. The product of the classification meetings will be a detailed description of the analysis populations, and the number and nature of unresolved data queries will also be reported.

The analysis is performed based on:

- The clinical database, which includes the electronic Case Report Forms (CRF)
- List of protocol deviations
- Analysis populations documented at the classification meetings

Reporting of the individual trials will be after the date of last subject last visit for each trial. Reporting of combined analyses of the two trials may be done after finalizing of the individual CSRs, and included as appendixes to the CSRs.

3 Trial Characteristics

3.1 Trial Objectives

3.1.1 Primary Objective

Trial P-Monofer-CKD-04
The primary safety objective of the trial is to evaluate the safety of IV iron isomaltoside compared to iron sucrose in subjects with iron deficiency anaemia (IDA) and non-dialysis-dependent chronic kidney disease (NDD-CKD). The evaluation will be based on the iron isomaltoside arm only, and conclusion will be drawn regardless of the safety observed in the iron sucrose arm.

The primary efficacy objective of the trial is to evaluate and compare the effect of iron isomaltoside to iron sucrose in its ability to increase haemoglobin (Hb) in subjects with IDA and NDD-CKD.

Trial P-Monofer-IDA-03
The primary safety objective of the trial is to evaluate the safety of IV iron isomaltoside compared to iron sucrose in subjects with IDA when oral iron preparations are ineffective or cannot be used or in whom the screening Hb measurement in Investigators' opinion were sufficiently low as to require rapid repletion of iron stores. The evaluation will be based on the iron isomaltoside arm only, and conclusion will be drawn regardless of the safety observed in the iron sucrose arm.

The primary efficacy objective of the trial is to evaluate and compare the effect of iron isomaltoside to iron sucrose in its ability to increase Hb in subjects with IDA when oral iron preparations are ineffective or cannot be used or in whom the screening Hb measurement in Investigators' opinion were sufficiently low as to require rapid repletion of iron stores.

3.1.2 Secondary Objectives

The secondary efficacy objectives of the trials are to evaluate and compare the effect of iron isomaltoside to iron sucrose on:

- Other relevant iron related biochemical parameters
- Fatigue symptoms
- Pharmacoeconomics

Note: Assessments related to pharmacoeconomics are only performed at baseline and it will not be possible to estimate a treatment effects.

3.2 Trial Design

The trials are randomised, comparative, open-label trials. In each trial 1500 subjects are planned to be randomised 2:1 to one of the following treatment groups:

- Group A: iron isomaltoside 1000 (Monofer®)
- Group B: iron sucrose (Venofer®)

In trial P-Monofer-CKD-04 subjects with IDA and NDD-CKD will be included, the randomization is stratified according to estimated Glomerular Filtration Rate (eGFR) at screening (<46, 46-59, or 60-89 mL/min/1.73m^2) and baseline cardiovascular risk (history of myocardial infarction, stroke, or congestive heart failure; yes/no).

In trial P-Monofer-IDA-03 subjects with IDA and with either up to a month run-in phase indicating intolerance or lack of response to oral iron or documented intolerance of oral iron for at least 1 month within the last 9 months or with a screening Hb measurement that was sufficiently low as to require rapid repletion of iron stores will be included. The randomization is stratified according to type of underlying disease (gastroenterology, gynecology, oncology, and 'other') and baseline cardiovascular risk (history of myocardial infarction, stroke, or congestive heart failure, yes/no).

Table 1: Flow Chart

Visit	**1 Screening**	R1-R4[f]	**2 Baseline**	TV1[e]	**3**	TV2[e]	**4**	**5**	**6**
Time (days)	**- 14/**-49	-42 - -14	**0**	3	**7**	10	**14**	**28**	**56**
Visit window (days)	**o-**	± 2	**-**	± 1	**± 1**	± 1	**± 2**	**± 2**	**± 3**
Informed consent	X								
Demographics	X								
In/exclusion criteria	X		X						
Eligibility lab tests	X	R4							
Pregnancy test, if relevant			X						
Medical history, including history of myocardial infarction, stroke, or congestive heart failure			X		X		X	X	X
Concomitant medication			X		X		X	X	X
Physical examination			X[a]						X
Height[a]			X						
Weight			X[a]						X

Vital signs[b1]			X	(X)	X	(X)	X	X	X
Oral iron (only patients included in the run-in period)		R1-R3							
Compliance to oral iron		R2-R4							
Tolerance to oral iron		R2-R4[g]							
Randomisation			X						
ECG [b1, b2]			X	(X)	X	(X)	X	X	X
FACIT Fatigue Scale			X		X		X		X
ISDR questionnaire			X						
Health care resource use questionnaire			X						
Safety lab tests			X		X		X	X	X
Efficacy lab tests			X		X		X	X	X
Treatment with iron isomaltoside[c] or			X						
Treatment with iron sucrose[c]			X	(X)	(X)	(X)	(X)		
Adverse events			X	(X)	X	(X)	X	X	X
Final visit form[d]									X

a. Physical examination, height, and weight may be measured between the screening and baseline visit
b1. For details of the assessment, please see Section 8.1.5.
b2. An ECG will be taken for all subjects. In addition, intensive ECG monitoring for evaluation of QTc prolongations will be performed at baseline in 35 subjects in trial P-Monofer-IDA-03.
c. All trial assessments should be done before administration of the trial drug unless otherwise stated in Section 8.1.5.
d. The final visit form may be filled in at any time of the trial if the subject is withdrawn.
e. Subjects in group B may attend up to 2 additional treatment visits (TV1-TV2). The TV visits are optional treatment visit for dosing with iron sucrose, if in the opinion of the Investigator required in order to achieve the cumulative dose of iron sucrose.
f. Subjects who do not have a documented intolerance of oral iron for at least 1 month within the last 9 months or had a screening Hb measurement which in Investigators' opinion was sufficiently low as to require rapid repletion of iron stores to minimize the risk of eventual blood transfusion should have up to 1 month of run-in period of oral iron in order to document intolerance or lack of response to oral iron. Subjects enrolled in the run-in period will have 1-4 additional visits (R1-R4). Visits R1-R3 are conducted by telephone.
g. If a subject has been withdrawn during the run-in period due to intolerance, the subject may attend visit R4 before day -14.

3.3 Subject included

3.3.1 Sample Size Determination

Safety:

Data from previous trials conducted with iron isomaltoside suggests that the incidence of treatment-emergent serious and/or severe non-serious hypersensitivity adverse events (AEs) for an IDA population can be expected to be approximately 1 % (IBD-01/IBD-02/CIA-01/PP-01: n=3/341=0.9 %, 95% confidence interval (CI): 0.2 %; 2.5 %). And similar for iron sucrose (CKD-03: n=1/114=0.9 %, 95% CI: 0.0 %; 4.8 %). A common incidence of 1.5 % will be assumed for the current trial.

As these are relatively rare events, an incidence of 0 % for placebo is assumed. Iron isomaltoside should be at most 50 % worse than the difference between iron sucrose and placebo, corresponding to a maximum relative risk of 2 (1.5 % / (1.5 %-0 %)2).

This trial will primarily assess the risk of the test drug under investigation. Arguing back from a relative risk of 2, gives that the upper bound of the 95 % CI for the incidence of treatment emergent serious and/or severe non-serious hypersensitivity AEs should be at most 3 % (2×1.5 %) for the iron isomaltoside treatment group.

Efficacy:

IDA-301 was a double-blind, placebo-controlled trial designed to compare the safety and efficacy of 1000 mg IV ferumoxytol to placebo. In this trial, 608 subjects were treated with ferumoxytol and 200 received placebo. The change in Hb from baseline to week 5 was the primary efficacy endpoint for the European Union regulators, and one of the secondary endpoints for U.S. regulators. The mean change was 2.6 g/dL (standard deviation (SD): 1.5) for IV ferumoxytol and 0.1 g/dL (SD: 0.9) for placebo. The estimated treated effect of IV ferumoxytol (i.e. the difference between ferumoxytol and placebo) was estimated to be approximately 2.5 g/dL with a 95 % CI ranging from 2.3 to 2.7 g/dL. The point estimate will be used as estimate of the treatment effect. Preserving 80 % of this effect is believed to provide reasonable assurance that iron isomaltoside is efficacious. Hence, the non-inferiority margin is set to (1-0.8) × 2.5 g/dL ~ 0.5 g/dL.

Power:

The significance level is set to 5 %. With n = 1000 in the iron isomaltoside treatment group, there is 88 % power for demonstrating that the upper bound of the 95 % CI of the incidence of treatment-emergent serious and/or severe non-serious hypersensitivity AEs is less than 3 %. And with n = 500 in the iron sucrose treatment group, assuming no difference between the treatment groups, and assuming a common SD = 1.5 g/dL, there is 100 % power for demonstrating non-inferiority of the change in Hb from baseline to week 8, using a non-inferiority margin of -0.5 g/dL.

This yields a total power for demonstrating both co-primary endpoints of 88 %.

The IDA-03 and CKD-04 trials will run in parallel with identical sample sizes, yielding a total of 2000 subjects to be treated with iron isomaltoside and 1000 with iron sucrose. In addition to the efficacy and safety analyses performed for this trial alone, an efficacy and safety assessment will be performed for the two trials combined. This is briefly described in Section 9.

4 Analysis Populations

The following 4 analysis sets are defined and will be used in the analyses of the data:

- Intention to treat (ITT) analysis set: The ITT analysis set will include all randomised subjects. Subjects will be included as randomised.
- Safety analysis set: The safety analysis set will include all randomised subjects who received at least one dose of the trial drug. This will be the analysis set for evaluating safety.
- Full analysis set (FAS): The FAS will consist of all randomised subjects, who received at least one dose of the trial drug, and have at least one post baseline Hb assessment. Subjects will be included as randomised.
- Per protocol (PP) analysis set: The PP analysis set will include all subjects in the FAS who do not have any major protocol deviation of clinical or statistical significance. Major protocol deviations are defined in Section 4.1. Subjects will be included as randomised.

The classification of the subjects will be performed before database lock.

Subjects included in the Safety analysis set will be evaluated according to treatment actually received.

The primary analysis population for efficacy evaluation is the ITT analysis set, and all efficacy endpoints will be analysed using the ITT and the subjects will be evaluated according to randomised treatment.

The primary efficacy endpoint will in addition be analysed using the FAS and PP. The analysis specified for the FAS and PP are to be regarded as supportive evidence.

4.1 Major protocol deviations

Deviations from the protocol will be registered as protocol deviations. Before data base lock (DBL) all protocol deviations will be evaluated and classified as minor, major, or GCP deviations. Protocol deviations classified as major will lead to exclusion from PP analysis set. The decisions will be documented.

The following will be assessed as major protocol deviation:

- Out of visit window of ≥ 14 days after the final visit (i.e. ≥ 70 days after baseline)
- Intake of prohibited medication (see section 4.1.1 for a list of prohibited medication)
- Treatment compliance outside the 80-120 % range (the calculation of compliance for iron sucrose is based upon the cumulative dose recorded at baseline)
- Other protocol deviations which are assessed as having a clinically or statistically significant effect (e.g. did not receive treatment as randomised)

All major protocol deviations will be listed by subject. Separate listings will be made for the safety analysis set of all visits outside window, intake of prohibited medication and treatment compliance outside range. Protocol deviations will be summarised by classification and category. The summary table will present number and percentages of subjects with a deviation and number of deviations.

4.1.1 Prohibited Medication

The following medication and non-drug therapy is not allowed in both trials:

- Any premedication (e.g. antihistamine or steroids) before administration of the trial drug
- Any iron supplementation other than investigational drug, except oral iron during the run-in period of trial P-Monofer-IDA-03 (nutritional supplementation including iron is allowed unless it is assumed as treatment of the subject's anaemia)
- Blood transfusion

P-Monofer-IDA-03 only:

- Erythropoiesis Stimulating Agents (ESA)

5 Planned Statistical Methods

5.1 Statistical Considerations

Baseline is defined as the last assessment with available data prior to the first administration of trial medication. No statistical tests comparing treatment groups at baseline will be performed.

Categorical data will be summarised by treatment, using number and percentages of subjects. For calculation of percentages the denominator will be the number of subjects in the analysis set, unless otherwise stated. Numerical data will be presented using the number of subjects (N), mean, SD, median, lower quartile, upper quartile, minimum and maximum. Both the absolute values and the change from baseline will be presented.

Descriptive statistics for all endpoints will be presented by treatment group and week (if applicable) using observed cases, i.e. no imputation of missing data will be performed, except for subject in the ITT with no post baseline measurements of the co-primary efficacy endpoint. For these cases the change to first post baseline visit will be set to 0. Missing data for time to event endpoints is handled using censoring.

Throughout this SAP, stratum or strata refer to the stratum the subject was randomised in. This might be different from the stratum the subject truly belongs to.

For the logistic regressions, it might be necessary to pool strata in case one or more strata are very sparse. If needed to the two smallest strata will be pooled, and if necessary the pooling will be repeated for the two smallest strata after the initial pooling.

For all analyses and presentations of efficacy endpoints scheduled time will be used, and trial visits will be labeled as follows:

Visit	Label
2 Baseline	Baseline
3	Week 1
4	Week 2
5	Week 4
6	Week 8

5.2 Subject Disposition

An overall summary table of the subject disposition will be prepared with number and percentages of subjects in the following categories (and sub-categories):

- Screened and randomised subjects
- Analysis sets (ITT, safety analysis set, FAS, and PP analysis set)
- End of study status
- Withdrawn from trial including reasons

For calculation of the percentages the denominator will be number of randomised subjects and relative to planned treatment.

Subject disposition will in addition be presented by stratum. All information will be listed. Separate listings will be made for subjects withdrawn from treatment and for screening failures.

5.3 Baseline Characteristics and Demographics

Demographics and baseline characteristics consist of age, gender, race, ethnicity, and smoking habits. Age will in addition be grouped into: 18-64 years, 65-84 years and above 84 years. Demographics will be listed and summarised using descriptive statistics, by treatment and by treatment and strata for the ITT and for the PP analysis set.

5.4 Medical History

Relevant medical history including history of myocardial infarction, stroke, or congestive heart failure, intolerance and non-response to oral iron treatment (P-Monofer-IDA-03 only) as well as underlying disorder causing IDA is collected.

Medical history will be summarised and listed. Intolerance and non-response to oral iron treatment, and underlying disorder causing IDA will be summarised separately.

6 Exposure and Other Dosing Information

6.1 Exposure

Dose and compliance will be summarised using descriptive statistics and presented for the safety analysis set by treatment, and by treatment and underlying disease causing IDA.

For both treatments, planned and actual dose in mg, number of doses and compliance will be presented.

Compliance will be calculated as: $$\frac{100 * \text{actual dose}}{\text{planned dose}}$$

All information will be listed. Furthermore, subjects out of the treatment compliance range of 80-120 % will be listed.

6.2 Concomitant medication

Concomitant medication at baseline and changes in concomitant medication during the trial will be recorder. All concomitant medication will be summarised and listed using the IIT analysis set.

7 Statistical Methodology for Co-Primary and Secondary Endpoints

7.1 Analysis and Presentation of the Co-Primary Endpoints

The primary safety objective of each trial is to evaluate the safety of IV iron isomaltoside compared to iron sucrose. The primary safety analysis will be based on the iron isomaltoside arm alone, and conclusion will be drawn regardless of the safety observed in the iron sucrose arm.

The co-primary endpoints for evaluation of the primary safety and efficacy objectives are:

- Serious or severe hypersensitivity reaction starting on or after the first dose of randomised treatment (i.e. treatment emergent)
- Change in Hb from baseline to week 8

7.1.1 Co-primary Safety Endpoint

The co-primary safety endpoint will be summarised and analysed using the safety analysis set.

The hypersensitivity terms are defined as standardised Medical Dictionary for Regulatory Activities query (SMQ) terms (including four additional terms), please refer to the protocols Appendix A.

The AEs will be adjudicated in a blinded fashion by an independent Clinical Endpoint Adjudication Committee (CEAC) (described in a separate adjudication charter). Only the adjudicated serious (fatal) or severe hypersensitivity reaction as judged by the CEAC will be considered as co-primary endpoint.

The treatment emergent serious or severe hypersensitivity reactions starting on or after the first dose of treatment will be summarised by SMQ term and preferred term (PT). The summary will include number of events, number of subjects, and proportion of subjects reporting serious or severe hypersensitivity reactions.

The following hypothesis will be tested:

The number of subjects reporting treatment-emergent serious or severe non-serious hypersensitivity reactions in the iron isomaltoside treatment group is less than 3 %.

H_{01}: $\pi_{\text{iron isomaltoside}} \geq 3$ % against the

alternative:

H_{A1}: $\pi_{\text{iron isomaltoside}} < 3$ %,

where, $\pi_{\text{iron isomaltoside}}$ denote the number of subjects reporting of treatment-emergent serious or severe non-serious hypersensitivity reactions.

The null hypothesis will be tested against the alternative by constructing an unadjusted two-sided exact 95 % CI of the number of subjects. If the upper bound of the 95 % CI is <3 %, the safety objective, is considered met.

Note, In the protocols section 15.1 it is stated that the hypotheses that the incidence of treatment-emergent serious and/or severe non-serious hypersensitivity reactions in the iron isomaltoside treatment group is less than 3 % will be tested. The intension was to evaluate on the number of subjects, i.e. the hypotheses that the number of subjects, and not number of events, reporting

treatment-emergent serious or severe non-serious hypersensitivity AEs in the iron isomaltoside treatment group is less than 3 % will be tested.

In addition, as supportive information, the risk difference between iron isomaltoside and iron sucrose will be assessed by constructing a 95 % CI of the risk difference. Both an unadjusted CI (with continuity correction), and a 95 % Newcombe CI adjusted for strata using the Cochran-Mantel-Haenszel method will be produced (ref. Yeonhee K. and Seunghyun W. 2013).

7.1.2 Co-primary efficacy endpoint

Change from baseline in Hb will be summarised by treatment and week using ITT, FAS and PP, respectively. Furthermore, the change from baseline in Hb will be summarised by strata, treatment and week using ITT, FAS and PP, respectively.

All subjects in the ITT with post baseline Hb data will be included with their observed data. For subjects without post baseline Hb values, the change from baseline will be set to 0 at the first post-baseline visit. This imputation will be included in the summary tables as well as in the primary analysis. Number of subjects without post baseline Hb values will be summarised.

The co-primary efficacy endpoint will be analysed using a mixed model for repeated measurement (MMRM) with a restricted maximum likelihood (REML)-based approach.

The following hypothesis will be tested:

$$H_{02}: \mu_{\text{iron isomaltoside}} - \mu_{\text{iron sucrose}} \leq -0.5 \text{ g/dL}$$

against the alternative:

$$H_{A2}: \mu_{\text{iron isomaltoside}} - \mu_{\text{iron sucrose}} > -0.5 \text{ g/dL},$$

where, $\mu_{\text{iron isomaltoside}}$ and $\mu_{\text{iron sucrose}}$ denote the change from baseline to week 8 in Hb.

The model will include the fixed, categorical effects of treatment (iron isomaltoside and iron sucrose), week, treatment-by-week interaction, strata, as well as the continuous, covariates of baseline Hb value and baseline Hb-by-week interaction. The treatment effect will be estimated using SAS Proc Mixed:

```
proc mixed data=_data method=reml cl ;
  class trtp strata avisit subjid ;
  model chg = trtp*avisit base*avisit avisit trtp strata base  / cl ddfm=kr noint ;
  repeated avisit / type=un subject=subjid ;
  estimate 'Visit 6, Iron Isomaltoside - Iron Sucrose' trtp 1 -1  trtp*avisit 0 0 0 1 0 0 0 -1 / cl ;
run ;
```

An unstructured covariance structure will be used to model the within-subject errors. If, unexpectedly, this analysis fails to converge, the following structures will be applied, in the following order; first-order ante-dependence, heterogeneous compound symmetry, compound symmetry.

The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

The primary comparisons will be the contrasts between iron isomaltoside and iron sucrose at week 8 based on the least squares means for the treatment-by-week interaction effect. The estimated mean difference based on this model will be reported with two-sided symmetric 95 % CI.

Non-inferiority of iron isomaltoside 1000 against iron sucrose can be claimed if the lower bound of the 95% CI is above -0.5 g/dL. The hypothesis is tested by evaluating the lower CI, and p-value for the non-inferiority test will not be given.

If non-inferiority is claimed it will be investigated if iron isomaltoside 1000 is *superior* to iron sucrose. If the lower bound of the two-sided 95 % CI is above 0 g/dL superiority will be declared, and the p-value associated with a test of superiority will be presented.

As sensitivity, the analysis above will be repeated for the FAS and the PP population.

Sensitivity to missing data

Some level of data missingness is expected, and the primary analysis (MMRM) is valid under the assumption that the data is Missing at Random (MAR). Simulation trials do suggest that MMRM is robust to accommodate some level of data Missing Not at Random (MNAR). Since it is unclear how the level of missingness of this type of data will influence the outcome at this stage, choosing a pre-specified primary analysis valid under MNAR accurately will be very difficult. As such, sensitivity analyses valid under relevant cases of data MNAR will be performed.

In particular, a pattern mixture model approach mimicking an ITT scenario where withdrawn subjects are assumed to be switched to a treatment inferior to the treatment with iron sucrose after withdrawal will be applied. With 5 visits (including baseline) there are $2^5 = 32$ possible missing data patterns.

In order to reduce the number of pattern groups and obtain a monotone missing data pattern, intermittent missing values (i.e. missing values followed by a non-missing value) will initially be imputed using a Markov Chain Monte Carlo (MCMC) method. This imputation is done for each treatment group separately and 100 copies of the dataset will be generated.

For the below imputations, it might be necessary to pool strata in case one or more strata are very sparse. If needed to the two smallest strata will be pooled, and if necessary the pooling will be repeated for the two smallest strata after the initial pooling.

In the first step, for each of the 100 copies of the dataset, an analysis of variance model with strata as factor, and baseline Hb as covariate is fitted to the Hb value at week 1 for each treatment arm. The estimated parameters, and their variances, from this model are used to impute missing values at week 1 for subjects in both treatment groups, based on their strata and Hb at baseline.

In the second step, for each of the 100 copies of the dataset, an analysis of variance model with strata as factor, and baseline Hb and Hb at 1 week as covariates is fitted to Hb at 2 weeks for each treatment arm. The estimated parameters, and their variances, from this model are used to impute missing

values at 2 weeks for subjects in both treatment groups, based on their strata and Hb at baseline and 1 week.

Missing values at 4 and 8 weeks are imputed using a similar approach.

Now in step five, for each withdrawn subject in the iron isomaltoside treatment group, a value of 0.5 g/dL (the non-inferiority limit as penalty) is subtracted from the change in Hb at 8 weeks. This corresponds to the assumption that a subject that withdraws from the iron isomaltoside is switched to a treatment that is inferior to iron isomaltoside.

Step six: For each of the complete data sets, the change from baseline to visit 8 is analysed using an analysis of variance model with treatment, site, and strata as factors and the baseline Hb value as a covariate. The estimates and standard deviations for the 100 data sets are pooled to one estimate and associated standard deviation using Rubin's formula:

$$m_{MI} = \frac{1}{100}\sum_{i=1|}^{100} m_{i,} \quad SD_{MI} = \sqrt{\frac{1}{100}\sum_{i=1}^{100} SD_i{}^2 + \left(1+\frac{1}{100}\right)\left(\frac{1}{100-1}\right)\sum_{i=1}^{100}(m_i - m_{MI})^2},$$

where m_i and SD_i are the estimated means and standard deviations for the 100 copies of the dataset, and m_{MI}, SD_{MI} are the pooled estimates.

In the final step, then from m_{MI} and SD_{MI}, the 95% confidence interval for the treatment difference and the associated p-value are calculated.

Mean plots of Hb from baseline to week 8 will be presented by monotone missing pattern group and treatment.

Step five will be repeated using different delta (penalties) and the tipping-point, where the results no longer fulfil non-inferiority, will be calculated if applicable. The pattern mixture models with multiple imputations are described by C. H. Mallinckrodt, Q. Lin, and G. Molenberghs [Mallinckrodt et al., 2012]. These analyses will be performed for the ITT and for the FAS, and PP analysis sets, if applicable. Other penalties may be explored.

7.2 Analysis and Presentation of the Secondary Efficacy Endpoints

The secondary efficacy endpoints are the following:
Trial P-Monofer-CKD-04:

- Hb increase of ≥ 1 g/dL from baseline to weeks 1, 2, 4, or 8
- Time to change in Hb ≥ 1 g/dL

Trial P-Monofer-IDA-03:

- Hb increase of ≥ 2 g/dL from baseline to weeks 1, 2, 4, or 8
- Time to change in Hb ≥ 2 g/dL

Both trials:

- Hb level of > 12 g/dL at any time from week 1 to week 8

- Increase in Hb concentration ≥ 2 g/dL at any time from week 1 to week 8
- S-ferritin level of ≥ 100 ng/mL and transferrin saturation (TSAT) of 20-50 % at any time from week 1 to week 8
- Change in concentrations of Hb from baseline to weeks 1, 2, and 4
- Change in concentrations of s-ferritin, TSAT, and s-iron from baseline to weeks 1, 2, 4, and 8
- Change in fatigue symptoms from baseline to weeks 1, 2, and 8 measured by the Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue Scale
- Resources used by the subject (per visit) and health care staff (per administration) measured by the ISDR questionnaire and health care resource use questionnaire, respectively

Note, the ISDR questionnaire and the health care resource use questionnaire are only filled in at baseline.

7.2.1 Haemoglobin

HB responder

In trial P-Monofer-CKD-04 a subject is considered a Hb responder to a certain week if an increase in Hb of at least 1 g/dL from baseline to the week in question is observed. In trial P-Monofer-IDA-03 a subject is considered a Hb responder to a certain week if an increase in Hb of at least 2 g/dL from baseline to the week in question is observed. Hb response will be evaluated against the weeks 1, 2, 4 and 8. Subjects without baseline- or post baseline Hb values will be set as non-responders at the first post-baseline week.

Number and percentage of responders will be summarised by treatment and week.

Proportion of Hb responders to each week will be analysed using a repeated measures logistic regression model with treatment, visit, strata and treatment by visit interaction as fixed effects and baseline value as covariate. An unstructured covariance structure will be used to model the within-subject errors. If the model does not converge one of the following covariance structures will be applied; exchangeable, m-dependent or independent. For each visit the estimated treatment ratio of iron isomaltoside versus iron sucrose will be presented with 95 % CIs and corresponding p-value.

Time to Hb response

Time to Hb response is considered the key secondary endpoint.

For responders, time to Hb response is defined as the scheduled time from baseline until the visit where the first Hb response were measured. For non-responders, time to response will be censored at the visit with the last known Hb measurement. Subjects without baseline- or post baseline Hb values will be censored at visit 2, Baseline.

Time to Hb response will be estimated using the Kaplan-Meier method and the hypothesis of no treatment difference will be assessed by a 2-sided log-rank test. Plots including the Kaplan-Meier curves, log-rank test and median survival time will be presented overall by treatment.

Hb level of > 12 g/dL

The proportion of subjects who achieve Hb level of > 12 g/dL at any time from week 1 to week 8 will be analysed using a logistic regression model with treatment and strata as fixed effects. The estimated treatment ratio of iron isomaltoside versus iron sucrose will be presented with 95 % CIs and corresponding p-value. Subjects without postbaseline Hb values will be set as failures.

Increase in Hb concentration ≥ 2 g/dL from week 1 to week 8
The proportion of subjects who achieve an increase in Hb concentration ≥ 2 g/dL at any time from week 1 to week 8 will be analysed using a logistic regression model with treatment and strata as fixed effects. The estimated treatment ratio of iron isomaltoside versus iron sucrose will be presented with 95 % CIs and corresponding p-value.

Change in concentrations of Hb
Change in concentrations of Hb from baseline to weeks 1, 2, and 4 will be analysed using a MMRM including treatment, week, treatment-by-week interaction, and strata as factors and baseline Hb, and baseline Hb-by-week interaction as covariates. An unstructured covariance structure will be used to model the within-subject errors and the estimation method will be a REML-based approach.

The estimated treatment differences based on the least square means of the treatment-by-week interaction effects will be presented by week including 95% CIs and corresponding p-values.

Change in Hb concentrations from baseline to week 1, 2, 4, and 8 will be summarised by treatment and week using descriptive statistics. Furthermore, absolute Hb values will be presented by treatment and week.

Change in Hb concentrations over time by treatment will be presented using mean plots.

7.2.2 s-ferritin, TSAT, and s-iron

Proportion of subject who achieve a *s*-ferritin level of ≥ 100 ng/mL and a TSAT of 20-50 % at any time from week 1 to week 8 will be analysed using a logistic regression model with treatment and strata as fixed effects. The estimated treatment ratio of iron isomaltoside versus iron sucrose will be presented with 95 % CIs and corresponding p-value.

Change in concentrations of *s*-ferritin from baseline to weeks 1, 2, 4 and 8 will be analysed using a MMRM including treatment, week, treatment-by-week interaction, and strata as factors and baseline valeu, and baseline value-by-week interaction as covariates. An unstructured covariance structure will be used to model the within-subject errors and the estimation method will be a REML-based approach. Change in concentrations of TSAT and *s*-iron will be analysed using similar models.

The estimated treatment differences based on the least square means of the treatment-by-week interaction effects will be presented by week including 95% CIs and corresponding p-values.

Concentrations of *s*-ferritin will be summarised by treatment and week using descriptive statistics, including change from baseline. Change in concentrations of *s*-ferritin over time by treatment will be presented using mean plots. Change in concentrations of TSAT and *s*-iron will be presented in a similar way.

7.2.3 FACIT fatigue scale

The FACIT fatigue scale consist of 13 items ranging from 0 (not at all) to 4 (very much). All items, except items 7 and 8 are reversed scored from 4 to 0. The total score range is 0-52. A score of less than 30 indicates severe fatigue, and the higher the score, the better the QoL. If more than 50 % of the items for a subject at a given visit are missing the total score will not be calculated. If less than 50 % (6 out of 13) of the items are missing the scores are prorated using the average of the other answers in the scale.

$$Total\ score = \frac{Sum\ of\ individual\ scores\ x\ 13}{Number\ of\ items\ answered}$$

Change in fatigue symptoms will be considered a continuous measure and will be analysed and presented similar to change in Hb except that treatment differences will be given for week 1, 2 and 8.

A sensitivity analysis will be performed using observed cases only, i.e. without imputation of missing items.

7.2.4 Other Assessments

Pharmacoeconomics is assessed at baseline by the ISDR questionnaire and the health care resource use questionnaire.

The health care resource use questionnaire includes the following items:

- What is your profession (nurse, doctor, etc.)?
- How much time did you use on administration of the trial drug including preparing for the administration, administration time, and observation time?

The items of the two questionnaires will be summarised using descriptive statistics. All information will be listed.

7.3 Handling of Missing Values

The co-primary efficacy endpoint is analysed using a MMRM based on the MAR assumption and performed using observed cased, except for subject in the ITT set with no post baseline data. For these subjects change to first post baseline week is set to 0. Sensitivity analyses using pattern mixture models and MI are performed to evaluate the robustness of the MAR assumptions. Please refer to section 7.1.2 for details.

For endpoints describing time to response missing values will be handled using censoring.

If less than 50 % of the items in the FACIT questionnaire are missing the scores are prorated using the average of the other answers in the scale.

7.4 Multiplicity adjustments

In these two-armed trials there is no multiplicity issue with more comparators. The trials contain two co-primary endpoints and both needs to be fulfilled in order to claim overall success of the individual trials. Superiority for the will only be tested if non-inferiority has been claimed.

For the limited number of secondary endpoints, no adjustment for multiplicity will be done. For secondary efficacy endpoints analysed using MMRM estimated treatment differences, CIs and corresponding p-values are presented for each week, but no adjustment for multiplicity is applied.

7.5 Sub-group and Centre Effects

No sub-group analyses are planned. The trials are multi centre trials, but no adjustment for centre effects is planned.

8 Statistical Methodology for Safety Endpoints

Safety will be considered for subjects included in the Safety Analysis Set and reported by actual received treatment, unless otherwise stated. No imputation of missing data is planned for safety endpoints.

8.1 Safety Endpoints

The secondary safety endpoints are the following:

- Composite cardiovascular AEs starting on or after the first dose of randomised treatment. The adjudicated composite AEs includes the following:
  - Death due to any cause
  - Non-fatal myocardial infarction
  - Non-fatal stroke
  - Unstable angina requiring hospitalization
  - Congestive heart failure requiring hospitalization or medical intervention
  - Arrhythmias
  - Hypertension
  - Hypotension
- Time to first composite cardiovascular safety AE
- Serum (*s*-) phosphate < 2 mg/dL at any time from baseline to weeks 1, 2, 4, or 8

Additional safety assessments are physical examinations and measurements of vital signs, height, weight, electrocardiogram (ECG), and safety laboratory parameters will be measured as part of standard safety assessments. Furthermore, in trial P-Monofer-IDA-03, intensive ECG monitoring for evaluation of QTc prolongations will be performed in 35 subjects treated with iron isomaltoside.

8.1.1 Analysis and Presentation of Safety Endpoints

8.1.2 Adverse Events

AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) current version. AEs will be regarded as treatment emergent AEs (TEAEs) if they occur after administration of randomised treatment. Non-treatment emergent AEs (Non-TEAE) are defined as AEs collected before dosing including AEs reported in randomised subjects never exposed to trial drug. Related or possible related AEs are defined as adverse drug reactions (ADRs).

An overall summary table by treatment will be made. The overall summary table will include number of events, number of subjects, and proportion of subjects reporting Non-TEAEs, TEAEs, Treatment emergent serious adverse events (SAEs), Fatal SAEs, ADRs, serious adverse reactions (SAR), TEAEs by severity, and TEAEs by outcome. The number of TEAEs, SAEs, ADRs, and SARs will be compared between treatment groups using a Fishers Exact test.

All TEAEs will be summarised by system organ class (SOC) and PT. The summaries will include number of events, number of subjects, and proportion of subjects reporting these events and will be tabulated, by treatment for:

- TEAEs

- Non-serious TEAEs, frequency ≥ 5% (or lower frequency as applicable)
- Treatment emergent SAEs
- TEAEs by severity (mild, moderate, severe)
- TEAEs by relationship (related, possible, unlikely, not related, unknown)
- Composite cardiovascular TEAEs
- Serious or severe hypersensitivity reactions

The following listings of AEs will be made based on the safety analysis set:

- Non-TEAEs
- TEAEs (non-serious and serious)
- TESAEs
- ADRs (including SUSARs)
- AEs leading to dose reduction or withdrawal from treatment
- Fatal SAEs
- Serious or severe hypersensitivity reactions
- Composite cardiovascular AEs

In addition, a listing of SAEs for subjects not in the safety analysis set will be made.

All data from adjudication review will be listed.

8.1.3 Composite cardiovascular AEs

The overall incidence of adjudicated composite cardiovascular AEs will be tabulated and compared between the treatment groups by a Fisher's exact test.

Time to first composite cardiovascular AE is defined as the actual time in days from first dose of treatment till the date of the adjudicated AE. For subjects not reporting a composite cardiovascular AE, the time will be censored at the date of the last attended visit.

The time to first composite cardiovascular AEs will be estimated using the Kaplan-Meier method and the hypothesis of no treatment difference will be assessed by a log-rank test. Plots including the Kaplan-Meier curves, and log-rank test will be presented.

8.1.4 Serum Phosphate

Subjects with serum (*s*-) phosphate < 2 mg/dL at any time from baseline to weeks 1, 2, 4, or 8 will be listed. The listing will include phosphate values for all weeks.

8.1.5 Other Safety Assessments

Vital Signs

Vital signs (heart rate, systolic and diastolic blood pressure) will be measured at baseline, weeks 1, 2, 4 and 8, and at the additional treatment visits, if applicable.

At the treatment visits, vital signs will be measured at the following time points:

- For subjects receiving an infusion of iron isomaltoside: approximately 0-10 minutes before infusion, during infusion, 5-15 minutes, and 20-40 minutes after the infusion has ended

- For subjects receiving a push injections of iron sucrose: approximately 0-10 minutes before injection, 5-15 minutes, and 20-40 minutes after the injection has ended

Vital signs will be summarised by week and time point, including change from pre- infusion/injection at treatment visits.

Height and Weight

Height is measured at baseline. Weight is measured at baseline and at the final visit.

ECG

Standard 12 lead ECG is assessed at baseline, weeks 1, 2, 4 and 8, and at the additional treatment visits, if applicable. At the treatment visits, two ECGs will be recorded; one before administration of the trial drug and one approximately 30 minutes after start of the dosing.

In addition, for P-Monofer IDA-03, intensive ECG monitoring for evaluation of QTc prolongations will be performed at baseline in 35 consecutive subjects in relation to treatment with iron isomaltoside at selected sites.

All ECG will be evaluated as normal or abnormal, and if abnormal clinical significant or not.

The 3 categories for the standard 12 lead ECG will be summarised by week, time point and treatment. All data will be listed.

Physical Examination

Physical examination is assessed at baseline and at week 8. Each body system of the physical examination will be evaluated as normal or abnormal, and if abnormal clinical significant or not.

The body system is not collected in the CRF and physical examination will be presented as an overall worst case of the 3 categories by treatment using shift tables from baseline till week 8. All data will be listed.

8.1.6 Laboratory Safety Data

Eligibility laboratory haematology and biochemistry parameters are measured at screening. Safety laboratory parameters (haematology and biochemistry) are measured at baseline and weeks 1, 2, 4, or 8.

Laboratory safety data will be summarised using descriptive statistics, including means, medians, SDs, quartiles, and minimum and maximum by week and treatment, including changes from baseline. In addition, laboratory safety data will be plotted using boxplots by week and treatment. For the summary tables and boxplots, values below lower limit of quantification (LLOQ) will be set to half times the limit.

Values outside normal ranges will be flagged. All safety laboratory data will be listed by subject, including eligibility laboratory assessments. Clinical significant laboratory will be listed including corresponding AEs.

8.2 Interim Analysis

No interim analysis is planned.

9 Combined Analyses

Data for the two trials, P-Monofer IDA-03 and P-Monofer CKD-04 will be pooled and combined analyses will be performed for the co-primary endpoints, and all secondary efficacy endpoints that appears in both trials.

9.1 Pooled analyses of the co-primary safety endpoint

The primary safety analysis described in section 7.1.1 will be repeated on the pooled data.

For safety, in addition to estimating the absolute risk (number of subjects and 95 % CI) of iron isomaltoside, a non-inferiority analysis will be performed of iron isomaltoside versus iron sucrose. For iron sucrose, a similar rate of 1.5 % will be assumed. As iron isomaltoside should be no worse than 3 %, the non-inferiority margin will be set to 1.5 % points. The non-inferiority assessment will be performed by estimation of the risk difference and the associated 95 % CI, adjusting for trial using the Cochran-Mantel-Haenszel method. With a total of 2000 subjects treated with iron isomaltoside, and 1000 subjects treated with iron sucrose, assuming common incidences of 1.5 %, and using a significance level of 5 %, the power to demonstrate non-inferiority is 87 %.

As sensitivity, the cumulative incidence will be estimated from a Kaplan-Meier curve of actual time from first treatment to serious and/or severe hypersensitivity AEs. Subjects without serious and/or serious hypersensitivity AEs will be censored at the last know time point without event. The log-log transformation will be used to estimate the standard errors for each treatment group, and the 95 % CI will be constructed by pooling these standard errors.

9.2 Pooled analyses of the co-primary efficacy endpoint

For efficacy, the same non-inferiority assessment as in the individual trials will be performed, including the specified sensitivity analyses. A similar MMRM model as for the individual trials will be applied, with trial added as fixed factor. The power for demonstrating efficacy is 100 %.

Hence, with a total of 2000 subjects treated with iron isomaltoside, and 1000 subjects treated with iron sucrose, the combined power of demonstrating the two co-primary endpoints is 87 %.

Similar to the individual trials, if the upper bound of two-sided 95 % CI of the risk difference (iron isomaltoside - iron sucrose) < 0 for the incidence of serious and/or severe hypersensitivity, or the lower bound of two-sided 95 % CI of the mean difference (iron isomaltoside - iron sucrose) in change in Hb from baseline to week 8 is 0 g/dL superiority will be declared, and the p-values for the test of superiority will be presented.

9.3 Pooled Analyses of the Secondary Efficacy Endpoints

Analyses combining data from the two trials will be made for the following secondary endpoints:

- Hb responder from baseline to weeks 1, 2, 4, or 8 (see section 7.2.1 for the definition of a responder)
- Hb level of > 12 g/dL at any time from week 1 to week 8
- Increase in Hb concentration ≥ 2 g/dL at any time from week 1 to week 8

- S-ferritin level of ≥ 100 ng/mL and transferrin saturation (TSAT) of 20-50 % at any time from week 1 to week 8
- Change in concentrations of Hb from baseline to weeks 1, 2, and 4
- Change in concentrations of s-ferritin, TSAT, and s-iron from baseline to weeks 1, 2, 4, and 8
- Change in fatigue symptoms from baseline to weeks 1, 2, and 8 measured by the FACIT Fatigue Scale

The analyses for the individual endpoints will be performed as described in section 7.2.1 to 7.2.3 above. For each analysis trial will be added as fixed factor in the models.

9.4 Pooled Analyses of the Composite Cardiovascular AEs

The overall incidence of adjudicated composite cardiovascular AEs will be tabulated and compared between the treatment groups by a Fisher's exact test.

The time to first composite cardiovascular AEs will be estimated using the Kaplan-Meier method and the hypothesis of no treatment difference will be assessed by a log-rank test. Plots including the Kaplan-Meier curve and log-rank test will be presented.

10 Software

All statistical calculations described in this SAP will be done by [redacted] Aps using SAS, release 9.4 or later (SAS Institute, Cary, NC, USA).

11 Reference List

Yeonhee K. and Seunghyun W. (2013). "Adjusted proportion difference and confidence interval in stratified randomized trials", *PharmaSUG 2013 – Paper SP04*.

Mallinckrodt CH, Lin Q, Molenberghs M. (2012). A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. Pharmaceutical Statistics; 12:1-6