STATISTICAL ANALYSIS PLAN

A Randomized, Assessor-blind Trial Comparing MENOPUR® (menotropins for injection) and Recombinant FSH (Follicle Stimulating Hormone) in a GnRH Antagonist Cycle with Single-Blastocyst Transfer in a High Responder Subject Population

000205

Investigational Product:	MENOPUR® (menotropins for injection)
Indication:	Development of multiple follicles and pregnancy in ovulatory women as part of an Assisted Reproductive Technology (ART) Cycle with ICSI
Phase:	4
Author:	Yodit Seifu, PhD
Date of issue:	March 8, 2017
Version:	4.0

Change log

Version No.	Effective Date	Reason for the Change / Revision	Supersedes

2.0	21SEP2015	SAP based on Protocol dated 22JUL2015	Version 1
3.0	16Feb2017	• For some baseline characteristics the analysis based on the safety population has been removed. The summaries based on the mITT and PP populations were considered to be sufficient. • For the primary analysis, the sensitivity analysis based on ITT population was removed. Furthermore, sub-group analysis by genetic testing was added. • Clarification on protocol deviation analysis. • Added forest plots and boxplots. • Removed the logistic regression. • Other minor clarification and corrections were made.	Version 2
4.0		In the conduct of the study, it was observed that there were a number of patients that had their primary endpoint visit too early (up to 4 weeks early). Initially, it was planned to use the visit assessment for this analysis however, because of this observation visit windows were considered to be a more accurate assessment for the endpoint. In Section 9, the visit window has now been defined. Furthermore, if a patient has pregnancy assessment beyond this visit she will also be considered as responder. To make it consistent, this type of window was also applied to the clinical pregnancy endpoint.	Version 3

Author, Reviewers, and Approver of the Statistical Analysis Plan

Author: Yodit Seifu, PhD.
Senior Project Statistician, Biometrics
Ferring International Pharmascience Center US, Inc.

Reviewers: Vladimir Yankov, M.D.
Vice President, Reproductive Health & Urology
Ferring International Pharmascience Center US, Inc.

Ralph Bobo, M.D.
Associate Vice President, US Pharmacovigilance
Ferring Pharmaceuticals Inc.

Kinesha O'Brien
Manger, Clinical Operations
Ferring International Pharmascience Center US, Inc.

Approver: Masakazu Ando, PhD.
Director, Biostatistics
Ferring International Pharmascience Center US, Inc.

TABLE OF CONTENTS

1 Introduction

This document describes the planned statistical analyses for 000205 based on the protocol dated 10 December 2015.

1.1 Definitions/ Abbreviations

1.1.1 Definition of Terms

Terms	Definitions
Randomized	Subject randomized to trial treatment
Screened	Subject who enters the screening phase

1.1.2 Abbreviations

Abbreviation	Meaning of abbreviation in document
AE	adverse event
AMH	anti-Müllerian hormone
ATC	Anatomical Therapeutic Chemical Classification System
β-hCG	Beta unit of human chorionic gonadotropin
BMI	body mass index
COS	controlled ovarian stimulation
E_2	estradiol
FSH	follicle-stimulating hormone
GnRH	gonadotropin-releasing hormone
hCG	human chorionic gonadotrohin
hMG	human menopausal gonadotrohin / menotrohin
HP-hMG	highly purified hMG
ICSI	intracytoplasmic sperm injection
IMP	investigational medicinal product
ITT	intention-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
NIMP	non-investigational medicinal product
OHSS	ovarian hyper stimulation syndrome
PP	per-protocol
PT	preferred term
SAE	serious adverse event
SOC	system organ class
TVUS	transvaginal ultrasound

2 Trial Objectives and Endpoints

2.1 Objectives

Primary Objective

- To demonstrate non-inferiority of MENOPUR® versus recombinant follicle-stimulating hormone (Gonal-f®) with respect to ongoing pregnancy rate in women undergoing controlled ovarian stimulation (COS) following a gonadotropin-releasing hormone (GnRH) antagonist protocol.

Secondary Objectives

- To evaluate the positive β-human chorionic gonadotropin (hCG) and clinical pregnancy rates after stimulation with MENOPUR® and Gonal-f® following a GnRH antagonist protocol.
- To evaluate follicular development during stimulation with MENOPUR® and Gonal-f® following a GnRH antagonist protocol.
- To evaluate the serum endocrine profile during stimulation with MENOPUR® and Gonal-f® following a GnRH antagonist protocol.
- To evaluate the number of oocytes retrieved, the fertilization rate, and embryo quality associated with MENOPUR® and Gonal-f® following a GnRH antagonist protocol.
- To evaluate the live birth rate associated with MENOPUR® and Gonal-f® following a GnRH antagonist protocol.
- To evaluate the adverse event (AE) profile of MENOPUR® and Gonal-f®

2.2 Endpoints

Primary Endpoint

- Ongoing pregnancy rate, defined as the presence of at least 1 intrauterine pregnancy with a detectable fetal heartbeat at 10–11 weeks of gestation (8-9 weeks after blastocyst transfer in the fresh cycle)

Secondary Endpoints

- Positive β-hCG rate (2 positive β-hCG tests; the first approximately 10-14 days after blastocyst transfer in the fresh cycle and a second confirmatory test approximately 2 days later)
- Clinical pregnancy rate (transvaginal ultrasound [TVUS] showing at least 1 intrauterine gestational sac with fetal heart beat at 5-6 weeks gestation, 3-4 weeks after blastocyst transfer in the fresh cycle)
- Early pregnancy loss (defined as 2 positive β-hCG tests but no ongoing pregnancy at 10-11 week's gestation in the fresh cycle)

- Follicular development as assessed by TVUS (total number of follicles and number of follicles with a diameter of ≤9 mm, 10-11 mm, 12-14 mm, 15-16 mm, and ≥17 mm on stimulation Day 6 and the last day of stimulation)
- Endocrine profile:
  - Serum follicle-stimulating hormone (FSH), hCG, luteinizing hormone (LH), androstenedione, total testosterone, dehydroepiandrosterone (DHEA): Day 1, Day 6, and last day of stimulation
  - Progesterone (P4), estradiol (E2): Day 1, Day 6, last day of stimulation, and Visit 4 (morning following hCG administration)
- Oocytes retrieved, metaphase II oocytes, fertilization rate, quality of embryos 3 days after oocyte retrieval, and the quality of blastocysts 5 days after oocyte retrieval
- Aneuploidy rate
- Endometrial assessment by TVUS (endometrial thickness in mm and echogenicity pattern on stimulation Day 6, the last day of stimulation, and at the time of blastocyst transfer in the fresh cycle)

Post-trial Endpoints

- Cumulative live birth rate for fresh and frozen blastocyst transfers (viable live birth >21 weeks gestation)
- Live birth rate for fresh blastocyst transfers (viable live birth >21 weeks gestation)
- Early pregnancy loss rate in frozen blastocyst transfer, defined as 2 positive β-hCG tests but no ongoing pregnancy at 10-11 weeks gestation in the frozen cycle
- Late pregnancy loss rate (defined as a confirmed ongoing pregnancy but no live birth)
- Positive β-hCG rate, clinical pregnancy rate, and ongoing pregnancy rate for frozen blastocyst transfers

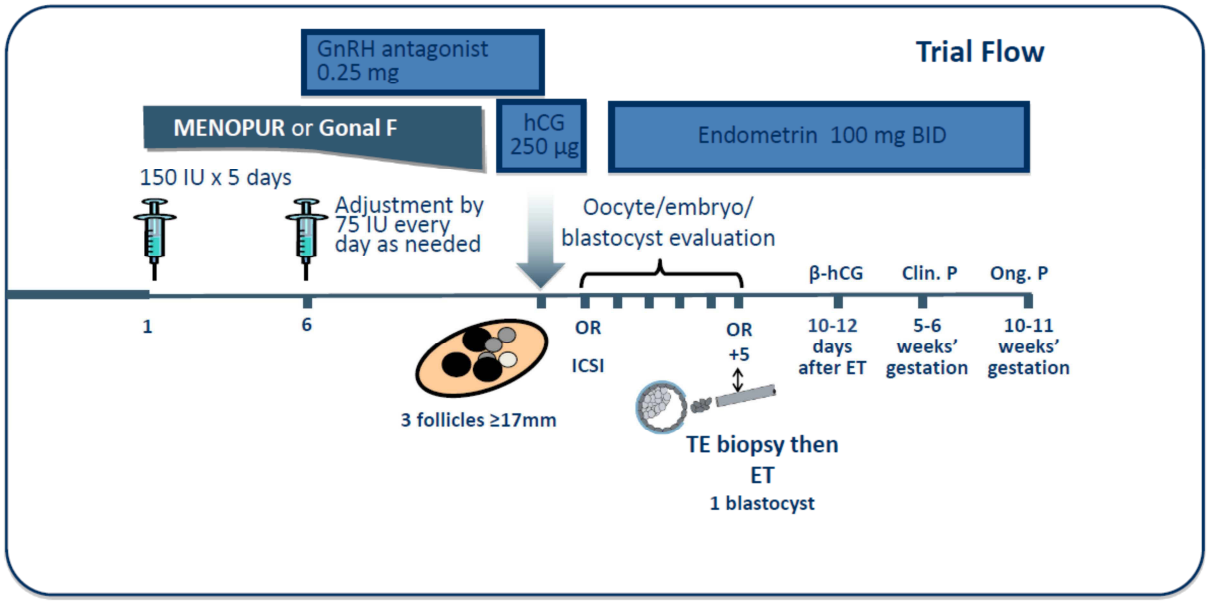
Safety Endpoints

- Adverse Events
- Frequency of ovarian hyperstimulation syndrome (OHSS) (early OHSS if the onset is ≤9 days after hCG administration and late OHSS if the onset is >9 days after hCG administration)

3 Trial design

3.1 General Design Considerations

This is a Phase 4, randomized, open-label, assessor-blind, parallel-group, multicenter study to be conducted in the United States (US). Approximately 600 females undergoing COS following a GnRH antagonist protocol will be randomized (1:1) to receive either MENOPUR® or Gonal-f®. The study is designed to demonstrate non-inferiority of MENOPUR® versus Gonal-f® for ongoing pregnancy rates in a predefined high-responder subject population. High responders will be defined as subjects who have a serum AMH ≥5 ng/mL.



3.2 Determination of Sample Size

The study has at least 80% power, with 275 subjects per treatment group, to demonstrate the non-inferiority of MENOPUR® to Gonal-f® in the ongoing clinical pregnancy rate at the 1-sided significance level of 0.025 with a -12% non-inferiority margin, by assuming an ongoing pregnancy rate of 50% for both treatment groups [1]. As the sensitivity analysis to the primary analysis is also deemed important, sufficient power is required for that analysis as well, and hence, assuming that 8% of the subjects may not be eligible for the PP analysis set, approximately 600 subjects will be randomized (1:1) into this study.

4 Subject Disposition

The number and percentage of subjects within each analysis set, randomized subjects treated with IMP, subjects who underwent a fresh cycle, and subjects prematurely discontinued from the study will be summarized. All post-baseline discontinuations will be summarized by reason for discontinuation. The subjects screened and not randomized will be presented in a separate data listing.

5 Protocol Deviations

Major protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being. A subset of these major protocol deviations will lead to exclusion of the data from the PP analysis set.

Some of the criteria for major protocol deviation are listed below:

- Actual treatment taken not in accordance with randomized treatment
- Significant non-compliance of the protocol specified visits, treatments, and/or procedures
- Violation of inclusion criteria 3 (AMH ≥5 ng/mL at screening)
- Violation of inclusion criteria 9 (at least 1 ovary)
  Violation of exclusion criteria 14 (findings at the gynecological examination that preclude gonadotropin therapy)

Blinded review of protocol deviations will be conducted prior to data base lock. The rating of protocol deviations as ‘minor’ or ‘major’ and exclusion from PP analyses will be completed prior to the declaration of ‘clean file’ and lock of database.

The number and percentage of subjects with major protocol deviations will be summarized.

6 Analysis sets

6.1 Intention-To-Treat Analysis Set

The intention-to-treat (ITT) analysis set comprises all randomized subjects and subjects will be analyzed according to their randomized treatment.

6.2 Modified Intention-To-Treat Analysis Set

The modified intention-to-treat (mITT) analysis set comprises all randomized subjects who received at least 1 dose of IMP and will be analyzed according to their randomized treatment.

6.3 Per Protocol Analysis Set

The per protocol (PP) analysis set comprises all mITT subjects except those excluded as a result of major protocol deviations and will be analyzed according to their randomized treatment.

6.4 Safety Analysis Set

The safety analysis set comprises all treated subjects and will be analyzed according to the actual treatment received.

7 Trial population

7.1 Demographics and Other Baseline Characteristics

Descriptive statistics of demographics and other baseline characteristics will be presented for the subjects in the mITT and PP analysis sets by treatment group. For patients demographic and tobacco and alcohol history this summary will also be repeated using the safety analysis set..

Categorical data will be summarised using numbers and percentages. The percentages are based on the total number of subjects within the given analysis set. Continuous data will be presented with the number of subjects, mean, standard deviation, median, interquartile range, minimum, and maximum. All demographic and baseline characteristics will also be presented in listings.

7.1.1 Demographics

Baseline demographics variables, including age, sex, race, ethnicity, height, weight, and Body Mass Index (BMI), will be summarized by treatment group. Age will be calculated as (Date of Informed Consent – Date of Birth)/365.25, rounded down to the nearest integer. BMI will be calculated as body weight (kg)/(height (m))^2, rounded to one decimal place.

7.1.2 AMH

Baseline AMH will be summarized by treatment group.

7.1.3 Endocrine Profile

The baseline endocrine profile (FSH, hCG, LH, P4, E2, androstenedione, total testosterone, dehydroepiandrosterone) will be summarized by treatment group by whether the samples were analyzed locally and centrally as appropriate.

7.1.4 Gynecological History and Infertility Diagnosis

Baseline gynecological history, infertility diagnosis, and the proportion subjects who have not had a previous clinical pregnancy will be summarized by treatment group.

7.1.5 Obstetric History

Baseline obstetric history will be summarized by treatment group.

7.1.6 Gynecological Examination

Baseline gynecological examination will be summarized by treatment group.

7.1.7 Uterine Cavity and Pregnancy Test

Baseline uterine cavity assessment and pregnancy test will be summarized by treatment group.

7.1.8 Transvaginal Ultrasound

Baseline transvaginal ultrasound will be summarized by treatment group.

7.1.9 Semen Analysis

Baseline semen analysis will be summarized by treatment group.

7.1.10 Tobacco and Alcohol

Tobacco and Alcohool Habits will be summarized by treatment group.

7.2 Medical and Surgical History

Medical and Sugical history recorded at screening visit will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 19.1 or later. Medical and Surgical history will be listed by treatment and subject and summarized by treatment group, system organ class (SOC), and preferred term (PT) for the subjects in the mITT and safety analysis sets, separately.

7.3 Prior and Concomitant Medication

Prior and concomitant medications will be coded using World Health Organisation Drug (WHODrug) Dictionary version 01SEP2013 or later. Medications will be listed by treatment and subject as well as summarized by treatment group, Anatomical Therapeutic Chemical (ATC) classification 1st level (alphabetically), and ATC classification 2nd level (in decreasing order of frequency) for the subjects in the mITT and safety analysis sets, separately. Medications will be tabulated separately for:

1) Prior medication; i.e. medication taken exclusively prior to first IMP administration, *i.e.* with stop date before date of first IMP administration
2) Concomitant medication, i.e. medication taken after IMP administration (regardless of whether the drug began before first IMP administration) i.e. medication that was not stopped before date of first IMP administration and not started after the end-of-trial visit

If the timing of the dose of a concomitant medication cannot be established in relation to the administration of IMP, it will be considered as concomitant medication.

7.4 Physical Examination

Abnormal physical examination results will be listed by treatment, subject, visit, and body system for the Safety analysis set.

7.5 Other Procedures During the Study

Information regarding other procedures logged during the study will be listed by treatment, subject, and start date for the Safety analysis set.

8 Exposure and Treatment Compliance

8.1 Extent of Exposure

All tabulations will be prepared for the Safety analysis set.

8.1.1 IMP (Gonadotropins)

Exposure to gonadotropins will be summarized as the total gonadotropin dose administered, dose administered from day 6 to last stimulation day, duration of treatment (days). Frequency tables will be included to describe the duration of treatment, the maximum dose, the dose on stimulation day 6, the final dose, and the number of dose adjustments.

This table will be produced overall and for subjects who completed the stimulation cycle, i.e. underwent triggering of final follicular maturation.

8.1.2 GnRH Antagonist

Once both the lead follicle measures ≥14 mm and serum E2 levels are ≥300 pg/mL, the GnRH antagonist (ganirelix acetate) will be initiated at a daily dose of 0.25 mg and continued throughout the gonadotropin treatment period.

The total dose, starting day of GnRH antagonist, and duration of treatment (days) will be tabulated. This table will be produced overall and for subjects who completed the stimulation cycle, i.e. underwent triggering of final follicular maturation.

8.1.3 hCG

A single injection of 250 µg hCG (Ovidrel®) will be administered to induce final follicular maturation as soon as 3 follicles of ≥17 mm are observed on TVUS. The following will be tabulated: number of subjects who met the hCG criteria, number of subjects that had hCG administered, number of subjects who met the hCG criteria and had hCG administered, the number of days from when the hCG criteria was met to when the hCG administered, and the number of subjects who had the hCG trigger replaced with a GnRH agonist trigger.

8.1.4 Progesterone

Vaginal progesterone inserts (200 mg/day – ENDOMETRIN®; Ferring) will be initiated for luteal phase support and will continue until the day of the β-hCG test (10-14 days after blastocyst transfer) if negative for pregnancy or if pregnancy is confirmed, will be continued for a maximum of 10 weeks.

The total dose and duration of treatment (days) will be tabulated for all subjects, for subjects with two positive β-hCG, and for subjects that do not have two positive β-hCG results.

8.2 Treatment Compliance

As treatment compliance is assumed to be high, only subjects not receiving the scheduled dose will be listed by center and subject number.

Separate listings will be prepared for

- violations of the IMP schedule
- violations of the antagonist schedule
- hCG not administered according to the protocol
- Agonist not administered according to the protocol
- Progesterone not administered according to the protocol

In addition, a summary table displaying the number and frequency of subjects with violations of the IMP schedule, violations of the antagonist schedule, hCG not administered according to the protocol, agonist not administered according to the protocol and progesterone not administered according to the protocol will be presented.

9 Efficacy

9.1 General Considerations

Efficacy analysis will be conducted for the mITT population unless otherwise specified. All statistical tests will be performed using a two-sided test at a 5% significance level. No adjustments will be made for multiple tests. Baseline for all efficacy analyses will be the values obtained at the last assessment prior to the first dose of IMP.

9.1.1 Tabulations

All summaries will be presented by treatment group and overall. Categorical data will be summarized using numbers and percentages. The percentages are based on the total number of subjects with a corresponding assessment. Continuous data will be presented using the number of subjects with an observation, mean and standard deviation, median, interquartile range, minimum, and maximum.

9.1.2 Listings

All primary and secondary endpoints will be listed sorted by center, treatment, and subject id.

9.1.3 Handling of Drop-Outs and Missing Values

For the primary and secondary endpoints positive β-hCG and clinical pregnancy missing observations will be imputed to negative unless a positive result is observed at a later pregnancy assessment. For example, if the outcome of β-hCG is missing but clinical pregnancy is 'positive' then β-hCG will be imputed as 'positive'.

Missing values will not be imputed for the any of the other secondary endpoints.

9.2 Primary Endpoint

The primary objective of the trial is to demonstrate that MENOPUR® is non-inferior to recombinant FSH with respect to ongoing pregnancy rate in a single fresh treatment cycle following a GnRH antagonist protocol. Ongoing pregnancy is defined as presence of at least 1 intrauterine pregnancy with a detectable fetal heartbeat at 10-11 weeks gestation (8-9 weeks after blastocyst transfer in the fresh cycle). The non-inferiority limit for the difference between treatments (MENOPUR® minus Gonal-f®) is -12% (absolute). The non-inferiority hypothesis to be tested for the primary endpoint is:

$$H_0: \pi_{MENOPUR} - \pi_{Gonal\text{-}f^{®}} \leq -12\%$$

against the alternative

$$H_1: \pi_{MENOPUR} - \pi_{Gonal\text{-}f^{®}} > -12\%,$$

where $\pi_{MENOPUR}$ and $\pi_{Gonal\text{-}f^{®}}$ denote the ongoing pregnancy rate of subjects treated with MENOPUR® and Gonal-f®, respectively, in a single fresh embryo transfer following a GnRH antagonist protocol.

The null hypothesis (H_0) will be tested against the alternative by constructing a 2-sided 95% confidence interval for the difference in ongoing pregnancy rates. If the lower-limit of the 95% confidence interval is greater than the non-inferiority limit (-12%), the null hypothesis will be rejected and it would be claimed that MENOPUR® is non-inferior to Gonal-f® with respect to ongoing pregnancy rate in a single fresh embryo transfer following a GnRH antagonist protocol.

If the 95% confidence interval for the treatment difference not only lies above the non-inferiority limit (-12%) but also above zero then there is evidence of superiority in terms of statistical significance at the 2-sided 5% level. With evidence of superiority, the corresponding 2-sided p-value will be reported. There is no need for a multiplicity adjustment since it is a simple closed test procedure.

Due to the expected large sample size, the confidence interval will be established based on the asymptotic normal distribution as follows:

$$\hat{\pi}_{MENOPUR} - \hat{\pi}_{Gonal-f} \pm Z_{(1-\alpha/2)} \sqrt{\frac{\hat{\pi}_{MENOPUR}(1-\hat{\pi}_{MENOPUR})}{N_{MENOPUR}} + \frac{\hat{\pi}_{Gonal-f}(1-\hat{\pi}_{Gonal-f})}{N_{Gonal-f}}}$$

where

$\hat{\pi}_{MENOPUR}$: is the observed ongoing pregnancy rate of subjects randomized to MENOPUR® in a single fresh treatment cycle following a GnRH antagonist protocol.

$\hat{\pi}_{Gonal-f}$: is the observed ongoing pregnancy rate of subjects randomized to Gonal-f® in a single fresh treatment cycle following a GnRH antagonist protocol.

$N_{MENOPUR}$: is the number of subjects randomized to MENOPUR®.

$N_{Gonal-f}$: is the number of subjects randomized to Gonal-f®.

$Z_{(1-\alpha/2)}$: is the 1-α/2 percentile in the standard normal distribution.

α : is the significance level, i.e., 0.05.

Subjects who do not have at least 1 intrauterine pregnancy with a detectable fetal heartbeat at 10–11 weeks gestation (8-9 weeks after blastocyst transfer in the fresh cycle) due to missing data, early withdrawal, or any other reason will be considered treatment failures (i.e., not having ongoing pregnancy) unless they have an assessment of detectable fetal heartbeat at a later date. Hence, for the analysis of this endpoint all positive (detectable fetal heartbeat) ultrasound assessments that occur in the interval from the end of 7th week from fresh tranfer till the end of study will be considered as responder for the primary endpoint (i.e., ultrasound assessment date – fresh transfer date + 1 $\geq$ 7 *7 = 49). Forty nine days was used because the 8-9 weeks visit window included 50

(beginning of the 8th week) to 63 days after fresh transfer. The primary endpoint analysis will be based on the mITT analysis set.

9.2.1 Sensitivity Analyses

Sensitivity analyses for the primary endpoint analysis will be conducted for the PP analysis set in the same manner as the mITT analysis set.

In addition, the ongoing pregnancy rate will be presented by center for the mITT analysis set, for subjects in the mITT analysis set with oocytes retrieved, and for subjects in the mITT analysis set with an embryo transfer.

Furthermore, the ongoing pregnancy rate will be presented for the following break downs for all mITT subjects, for mITT subjects with oocytes retrieved and for mITT subjects with a blastocyst transfer:

- By reason for infertility
- By center
- By number of oocytes retrieved (<=3, 4-7, 8-10, 11-14, >=15)
- By quality of blastocyst transferred.
- By AMH level at screening
- For patients that received fresh blastocyst, by results of the genetic testing

For these sub-group analyses, treatment differences and associated 95% CIs will be generated. Furthermore, these summaries will be displayed in a forest plot.

9.3 Secondary Endpoints

All secondary endpoints will be based on the mITT analysis set unless otherwise specified. Positive β-hCG rate and clinical pregnancy rate, in the fresh cycle, will be presented by treatment group with the corresponding 95% confidence intervals of the treatment difference. This will be calculated in the same manner as the primary endpoint.

9.3.1 Positive β-hCG rate in the fresh cycle

Positive β-hCG in the fresh cycle is defined as 2 positive β-hCG tests; the first performed approximately 10-14 days after the fresh blastocyst transfer with a second positive test at least 2 days later. The positive β-hCG rate in the fresh cycle will be presented by treatment group with the corresponding 95% confidence intervals of the treatment difference. This analysis will also be repeated for the PP set.

Subjects who do not have 2 positive β-hCG tests due to missing data, early withdrawal, or any other reason will be counted as not having positive β-hCG unless a positive result is observed at a later pregnancy assessment. For example, if the outcome of β-hCG is missing but clinical pregnancy is ‘positive’ then β-hCG will be imputed as ‘positive’.

9.3.2 Clinical pregnancy rate in the fresh cycle

Clinical pregnancy in the fresh cycle is defined as at least 1 intrauterine gestational sac, confirmed by TVUS, with a fetal heart beat at 5-6 weeks gestation (3-4 weeks after the fresh blastocyst transfer). The clinical pregnancy rate in the fresh cycle will be presented by treatment group with the corresponding 95% confidence intervals of the treatment difference. This analysis will also be repeated for the PP set.

Subjects who do not have at least 1 intrauterine gestational sac, confirmed by TVUS, with a fetal heart beat at 5-6 weeks gestation (3-4 weeks after the fresh blastocyst transfer) due to missing data, early withdrawal, or any other reason will be counted as not having clinical pregnancy unless a positive result is observed at a later pregnancy assessment. For example, if TVUS test at 5-6 weeks is missing but clinical pregnancy is 'positive' at 10-11 weeks of gestation, it will be imputed as 'positive'. Hence, for the analysis of this endpoint all positive (detectable fetal heartbeat) ultrasound assessments that occur in the interval from the end of 2nd week from fresh tranfer till the end of study will be consdered as responder for this secondary endpoint (i.e., ultrasound assessment date – fresh transfer date + 1 $\geq$ 2 *7 = 14). Fourteen days was used because the 3-4 weeks visit window included 15 (beginning of the 3th week) to 28 days after fresh transfer.

9.3.3 Early pregnancy loss in the fresh cycle

Early pregnancy loss in the fresh cycle is defined as 2 positive β-hCG tests, but no ongoing pregnancy at 10-11 week's gestation, in the fresh cycle. The early pregnancy loss rate in the fresh cycle will be calculated for mITT subjects with 2 positive β-hCG tests and presented by descriptive statistics for each treatment group. Furthermore, the difference in event rates between the two treatments will be presented together with the associated 95% CI.

9.3.4 Follicles

The follicles within each treatment group on stimulation day 6 and last day of stimulation will be summarized by frequency at the follicle level (i.e. the number of follicles within the treatment group that are ≤9mm, 10-11 mm, 12-14 mm, 15-16 mm, and ≥17 mm). The follicles for each subject within each treatment group on stimulation day 6 and last day of stimulation will also be summarized by the largest follicle size, average follicle size, average size of the 3 largest follicles, and the rate of follicles ≥17 mm, ≥15 mm, and ≥12 mm. Tables will be produced for all mITT subjects and for all mITT subjects with oocytes retrieved.

9.3.5 Endocrine Profile

The endocrine profile will be summarized using descriptive statistics by scheduled visit, as well as for the change from baseline for post-baseline visits;

- Serum follicle-stimulating hormone (FSH), hCG, luteinizing hormone (LH), androstenedione, total testosterone, dehydroepiandrosterone will be summarized on Day 1, Day 6, and last day of stimulation.

- Progesterone (P4), estradiol (E2) will be summarized on Day 1, Day 6, last day of stimulation, and Visit 4 (morning following hCG trigger)

Any subject who receives an antagonist earlier than Day 6 will be excluded from Day 6 endocrine analysis. Associated box-plots will also be generated.

9.3.6 Oocytes

The number of oocytes retrieved, the number of metaphase II oocytes, the number of oocytes undergoing ICSI, and the number of normally fertilized (2PN) oocytes will be summarized by frequency distribution and by descriptive statistics for each treatment group.

9.3.7 Fertilization Rate

The fertilization rate will be expressed as a percentage for each subject and calculated as 100 times the ratio of the number of fertilized 2PN oocytes to the number of oocytes retrieved. In addition for subjects undergoing ICSI the fertilization rate will be expressed as a percentage for each subject and calculated as 100 times the ratio of the number of fertilized 2PN oocytes to the number of metaphase II oocytes. Descriptive statistics will be provided by treatment group.

9.3.8 Quality of Embryos 3 Days after Oocyte Retrieval

The quality of embryos 3 days after oocyte retrieval will be assessed by cleavage stage, blastomere uniformity, cell size, the degree of fragmentation, and visual signs of multinucleation. Frequency distributions and descriptive statistics will be provided for each treatment group at the embryo level.

9.3.9 Quality of Blastocysts 5 Days after Oocyte Retrieval

The quality of blastocysts 5 days after oocyte retrieval will be assessed by blastocyst expansion and hatching status, blastocyst inner cell mass grading, and trophectoderm grading. Blastocyst information collected on day 6 and 7 will be summarized similarly. Furthermore, the overall quality of the blastocysts on day 5 will also be summarized. Frequency distributions will be provided for each treatment group at the blastocyst level.

9.3.10 Best Quality Blastocysts 5 Days after Oocyte Retrieval

The best quality blastocyst 5 days after oocyte retrieval at the subject level will be summarized for excellent-quality blastocysts and good-quality blastocysts separately by treatment using descriptive statistics.

9.3.11 Aneuploidy Rate

The aneuploidy rate will be expressed as a percentage for each subject and calculated as 100 times the ratio of the number of aneuploid blastocysts to the total number of blastocysts..Descriptive statistics will be provided by treatment group.

9.3.12 Endometrial Thickness and Echogenicity Pattern

Endometrial thickness will be summarized by descriptive statistics and the echogenicity pattern will be summarized by frequency distribution , at stimulation Day 6, the last day of stimulation, and at the time of blastocyst transfer.

9.4 Post-Trial Endpoints

All post-trial endpoints will be summarized using tables and figures, as appropriate. Additional details will be specified in the separate Statistical Analysis Plan for the post-trial endpoints. The post-trial endpoints include:

- Cumulative live birth rate for fresh and frozen blastocyst transfer (defined as the proportion of subjects with at least 1 viable live birth greater >21 weeks gestation).
- Live birth rate for fresh blastocyst transfer (defined as the proportion of subjects with at least 1 viable live birth greater >21 weeks gestation).
- Early pregnancy loss rate in frozen blastocyst transfer is defined as 2 positive β-hCG tests but no ongoing pregnancy at 10-11 weeks gestation in the frozen cycle.
- Late pregnancy loss rate (defined as a confirmed ongoing pregnancy but no viable live birth greater >21 weeks gestation).
- Positive β-hCG rate, clinical pregnancy rate, and ongoing pregnancy rate for frozen blastocyst transfers.

Descriptive statistics will be provided for each of the respective post-trial endpoints.

10 Safety

10.1 General Considerations

Safety parameters will be evaluated for the safety analysis set and subjects will be included in tabulations according to actual treatment received.

10.2 Adverse Events

Adverse events will be recorded from signed informed consent until the end-of-trial visit.

Adverse events (AEs) are classified according to the Medical Dictionary for Regulatory Activities (MedDRA) version 19.1 or later.

Missing values will be treated as missing, except for causality, intensity, seriousness, and outcome of AE. A 'worst case' approach will be used:

- if causality is missing, the AE will be regarded as related to the IMP
- if the intensity of an AE is missing, the AE will be regarded as severe
- if seriousness is missing, the AE will be regarded as serious
- if outcome is missing and no date of outcome is present, the outcome is regarded as 'Not yet recovered'

Missing event start and event stop dates will be imputed in the most conservative way, i.e. the event should start as early as possible based on the (partially) missing start date while remaining treatment emergent (if relevant), and last as long as possible, i.e. until last visit or one day prior to the next event with the same preferred term, if any. In the latter case both consecutive AEs (with the same PT) will remain to be regarded as two separate events (despite being adjacent).

A treatment emergent adverse event is any adverse event occurring in the period from first administration of IMP till end-of-trial visit.

Pre-treatment adverse events will be reported in a listing only. This listing will be sorted by center and subject ID and date of onset.

10.2.1 Overview of Treatment-Emergent Adverse Events

An AE overview summary table will be prepared including the number of subjects reporting an AE, the percentage of subjects (%) with an AE, and the number of events (E) reported, for the following categories:

- Treatment-emergent adverse events
- Deaths
- Serious adverse events
- Adverse events leading to withdrawal
- Severe and life threatening adverse events
- Adverse drug reactions

10.2.2 Incidence of Treatment-Emergent Adverse Events

Treatment-emergent adverse events will be summarized in summary tables. The tables will display the total number of subjects reporting an AE and the percentage of subjects (%) with an AE. AEs will be presented alphabetically by system organ class (SOC) and in decreasing frequency of occurrence by preferred term (PT).

Summary tables will be prepared for:

- All adverse events
- Adverse events with an incidence [≥ 5%] of subjects in any treatment group
- Adverse events by causality (related/unrelated)
- Adverse events leading to death
- Adverse events by intensity (mild/moderate/severe)
- Serious adverse events
- Adverse events leading to withdrawal

A supporting data listings will be provided for:

- All adverse events sorted by center, subject number, and date of onset
- Adverse drug reactions
- Adverse events leading to death
- Severe adverse events
- Serious adverse events
- Adverse events leading to withdrawal

Supporting data listing by SOC sorted alphabetically and PT sorted in decreasing frequency of occurrence will be provided for:

- All adverse events

The all adverse events table will summarize all the treatment-emergent AEs by SOC and Preferred Term displaying the total number of subjects (N) reporting an AE, the percentage of subjects (%) with an AE, and the number of events (E) reported will also be provided.

10.3 Chemistry and Hematology Laboratory Variables

Mean change and mean percentage (%) change from baseline at the end of trial will be presented for each laboratory variable. In addition, descriptive statistics, i.e., the number of subjects with data, mean, standard deviation, median, minimum, and maximum values, will be presented for observed values at screening and end of treatment for each laboratory variable.

Data listings will also be prepared by center for all subjects with any markedly abnormal laboratory safety values at any time-point (including screening/baseline) (appendix 1).

10.4 Vital Signs

10.4.1 Vital Signs

Baseline for all vital signs analyses will be the values obtained at the screening visit. The only planned post-baseline assessment of vital signs is at the End of trial visit.

10.4.2 Summary Statistics

Mean change and mean percentage (%) change from baseline at end of trial will be presented for each vital signs variable. In addition, descriptive statistics, i.e., the number of subjects with data, mean, standard deviation, median, minimum, and maximum values, will be presented for observed values and change from baseline at screening and end of treatment for each vital signs variable.

10.4.3 Data Listings

Data listings will be prepared by center for all subjects with any markedly abnormal vital signs value at any time-point (including screening/baseline) (appendix 1).

10.5 Other Safety Variables

10.5.1 OHSS

A summary table will be prepared showing the incidence and severity of early OHSS, late OHSS, and overall OHSS. This table will display the number and percentage of subjects (%) and the number of events. Early OHSS is defined as OHSS with onset ≤9 days after triggering of final follicular maturation. Note this includes OHSS with onset before triggering and OHSS with onset during stimulation where triggering is not performed.

10.5.2 Signs of Excessive Response

A table will be prepared summarizing signs of excessive response. The table will present the number of subjects in following categories:

- OHSS early onset defined as at most 9 days after hCG administration
- Cancellation of cycle due to excessive response
- Preventive action (leuprolide acetate)
- Received a medical or surgical intervention for OHSS

11 Interim analyses

No interim analysis is planned.

12 Deviations from protocol analysis

For some of the baseline characteristics, the analysis based on the safety population has been removed. The summaries based on mITT and PP populations were considered to be sufficient.

For the primary analysis, the sensitivity analysis based on ITT populations was removed. Furthermore, sub-group analysis by genetic testing was added.

For the primary analysis, it was added that patients that have positive pregnancy test at any time after the 8-9 week window post-fresh transfer will also be considered as positive for this primary endpoint.

13 References

1. Yeh JS, Steward RG, Dude AM, et al. Pregnancy outcomes decline in recipients over age 44: an analysis of 27,959 fresh donor oocyte in vitro fertilization cycles from the Society for Assisted Reproductive Technology. Fertil Steril. 2014;101(5):1331-6.

14 Tables, Listings and Figures Shells

The document with tables, figures and listings (TLF) shells presented in a separate document.

Appendix 1 Markedly Abnormal Laboratory Safety Values and Vital Sign

Table 1: Markedly abnormal Criteria for Laboratory Tests

Variable	Units	Markedly abnormal Criteria	
		Low	**High**
Haematology			
Haemoglobin	g/L	≤ 115	Not applicable
Haematocrit	Ratio	≤ 0.37	≥ 0.56
Total WBC	10^9/L	≤ 2.8	≥ 16.0
Eosinophils	%	Not applicable	≥ 10
Neutrophils	%	≤ 15	≥ 90
Lymphocytes	%	≤ 10	≥ 80
Monocytes	%	Not applicable	≥ 20
Basophils	%	Not applicable	≥ 5
Bands	%	Not applicable	≥ 20
Platelets	10^9/L	≤ 75	≥ 700
Total RBC	10^{12}/L	≤ 3.5	Not applicable
Clinical Chemistry			
AST	IU/L	Not applicable	> 3xULN
ALT	IU/L	Not applicable	> 3xULN
Alkaline phosphatase	IU/L	Not applicable	> 3xULN and 25% increase from baseline
GGT	IU/L	Not applicable	> 3xULN
LDH	IU/L	Not applicable	> 3xULN
Total bilirubin	µmol/L	Not applicable	≥ 1.5xULN
Urea nitrogen	mmol/L	Not applicable	≥ 10.7
Creatinine	µmol/L	Not applicable	≥ 177
Total protein	g/L	≤ 45	≥ 90
Albumin	g/L	≤ 25	≥ 65
Sodium	mmol/L	≤ 130	≥ 155
Potassium	mmol/L	≤ 3.0	≥ 5.8
Chloride	mmol/L	≤ 90	≥ 115
Phosphorus	mmol/L	≤ 0.5	≥ 1.9
Calcium	mmol/L	≤1.8	≥ 3.9
Uric acid	mmol/L	Not applicable	≥ 0.62
Glucose	mmol/L	≤ 2.8	≥ 10
Total cholesterol	mmol/L	Not applicable	≥ 8.0
Urinalysis: Quantitative			

Variable	Units	Markedly abnormal Criteria	
		Low	**High**
pH	None	≤ 4	Not applicable
Specific gravity	None	≤ 1.005	Not applicable
Urinalysis: Dipstick Chemistries			
Glucose	0 - 4+	Not applicable	Increase of 2 or more units from baseline
Casts		Not applicable	Increase of 2 or more units from baseline
Protein	0 - 4+	Not applicable	Increase of 2 or more units from baseline
Ketones	0 - 4+	Not applicable	Increase of 2 or more units from baseline
Blood (Hgb)	0 - 4+	Not applicable	Increase of 2 or more units from baseline
Urinalysis: Microscopic Variables			
RBC	no./hpf	Not applicable	≥ 10
WBC	no./hpf	Not applicable	≥ 20
Casts	no./hpf	Not applicable	Neg at baseline to positive on-treatment
Bacteria	no./hpf	Not applicable	Neg at baseline to positive on-treatment
Cells	no./hpf	Not applicable	Neg at baseline to positive on-treatment
Crystals	no./hpf	Not applicable	Neg at baseline to positive on-treatment

Table 2: Markedly abnormal Criteria for Vital Signs*

Variable	Criterion Value	Change from Baseline
Systolic blood pressure	≥ 180 mmHg ≤ 90 mmHg	Increase of ≥ 20 mmHg Decrease of ≥ 20 mmHg
Diastolic blood pressure	≥ 105 mmHg ≤ 50 mmHg	Increase of ≥ 15 mmHg Decrease of ≥ 15 mmHg
Pulse rate	≥ 120 bpm ≤ 50 bpm	Increase of ≥ 15 bpm Decrease of ≥ 15 bpm
Body weight	None	Increase of ≥ 7% Decrease of ≥ 7%

* To be identified as markedly abnormal, a treatment value must meet the criterion value and also the specified change from baseline.