FERRING
PHARMACEUTICALS

Statistical Analysis Plan

Title of trial: A Randomised, Controlled, Assessor-blind, Parallel Groups, Multicentre Trial Assessing the Efficacy and Safety of FE 999049 in Controlled Ovarian Stimulation in Japanese Women Undergoing an Assisted Reproductive Technology Programme
NCT number: NCT03228680
Sponsor trial code: 000273
Date: 22 August 2019

STATISTICAL ANALYSIS PLAN

A randomised, controlled, assessor-blind, parallel groups, multicentre trial assessing the efficacy and safety of FE 999049 in controlled ovarian stimulation in Japanese women undergoing an assisted reproductive technology programme

Trial Code 000273

Investigational Product: FE 999049, human recombinant follicle-stimulating hormone, solution for subcutaneous injection

Indication: Controlled ovarian stimulation for the development of multiple follicles in women undergoing assisted reproductive technologies (ART) such as an in vitro fertilisation (IVF) or intracytoplasmic sperm injection (ICSI) cycle

Phase: 3

Author: [redacted]

Date of issue: 22 August 2019

Version: 3.0

Change log

Version No.	Effective Date	Reason for the Change / Revision	Supersedes
1.0	14 July 2017	New version	None
2.0	25 July 2019	Section 1 Introduction The text "... and is based on the final protocol version. 1.0 dated 28FEB2017." is replaced by "... is based on the final protocol version. 2.0 dated 16 March 2018 and the protocol amendment Version 1.0 (dated 13th November 2018).".	1.0
2.0	25 July 2019	Section 3.1 General Design Considerations The following clarification is added: "The randomisation strata by AMH reflect the FE 999049 dosing regimen, and therefore each subject's AMH value is rounded off to the nearest integer before allocation into one of the two strata, <15 pmol/L or ≥15 pmol/L. The rounding off rules used for AMH with respect to randomisation strata are identical to those described in Section 5.1.1 "FE 999049 Dosing Regimen" of the protocol. Consequently, the randomisation stratum AMH <15 pmol/L corresponds to the description of the AMH <15 pmol/L category in the FE 999049 dosing regimen, meaning that it includes AMH values up to and including 14.49 pmol/L. The randomisation stratum AMH ≥15 pmol/L includes AMH values from 14.50 pmol/L and higher. These rounding off rules are implemented in the eCRF and used in the automated allocation of subjects into either the AMH <15 pmol/L or ≥15 pmol/L strata."	1.0

2.0	25 July 2019	Section 4 Subject Disposition The text " ... full analysis set (FAS) overall and by AMH stratum. All subjects screened will be accounted for." is changed to "... full analysis set (FAS) overall and by AMH stratum." as sentence the "All subjects screened will be accounted for." is presented earlier in the section. The sentence "The number of subjects screened but not randomized/allocated to treatment will be presented with the reason(s) for screen failure in a data listing." is removed.	1.0
2.0	25 July 2019	Section 5 Protocol Deviations The text "The rating of protocol deviations as 'minor' and 'major' will be decided by the Ferring clinical team on the basis of a blinded …" is changed to "The rating of protocol deviations as 'minor' and 'major' will be decided by the medical officer, medical monitor and project statistician based on a blinded …".	1.0
2.0	25 July 2019	Section 7.2 Medical History The text "All medical history will be coded using MedDRA. The version of MedDRA will be documented." is changed to "All medical history will be coded using MedDRA version 20.0."	1.0
2.0	25 July 2019	Section 8.1.1 Extent of Exposure Comparison for "The total gonadotropin dose" was removed as FE 999049 and FOLLISTIM cannot be converted to be presented using a common unit. The comparison of duration of stimulation (days) between treatments for three age groups (<35, 35-37 and 38-40 years) has been removed.	1.0
2.0	25 July 2019	Section 9.1 General Considerations Under "**Analysis and Presentation of Primary and Secondary Endpoints**" The text "In addition, summary tables per age group (<35, 35-37 and 38-40 years) will be provided." is changed to "In addition, summary tables and treatment comparisons for the efficacy endpoints will be presented per age group (<35, 35-37 and 38-40 years) unless it is specified otherwise.".	1.0

2.0	25 July 2019	Section 9.2.2 Sensitivity Analyses and Additional Analysis The text “The distribution of the number of oocytes retrieved will be plotted using histograms and the percentage …” is changed to “The percentage …”.	1.0
2.0	25 July 2019	Section 9.3 Important Secondary Endpoint Under "**Clinical Pregnancy Rate**": The text "are the treatment specific standards error of the observed difference in rates within stratum *i*," is corrected to "is the standards error of the observed difference in rates within stratum *i*,". The text "Both analyses will be based on all subjects in the FAS." is changed to "The analyses will be based on all subjects as well as all subjects with transfer and they are done on the FAS.".	1.0
2.0	25 July 2019	Section 9.4 Secondary Endpoints Under "**Implantation Rate**": The text " This analysis does not adhere to the randomisation principle and is less relevant than the analysis of clinical pregnancy rate for all subjects.” is removed. Under "**Number and Size of Follicles during Stimulation**": The text "Categorical data will be compared between treatment groups using a logistic regression model with treatment and AMH stratum as fixed factors and within each AMH stratum the chi-square test." is not corresponding to any analysis and therefore it is removed. Under "**Fertilisation Rate**": The text "Categorical data will be compared between treatment groups using a logistic regression model with treatment and AMH stratum as fixed factors and within each AMH stratum the chi-square test." is not corresponding to any analysis and therefore it is removed.	1.0

		The text "Number of fertilised oocytes and fertilisation rate will be compared ..." is changed to "Number of fertilised oocytes, number of metaphase II oocytes, fertilisation rate, the rate of metaphase II oocytes to oocytes retrieved, and fertilisation rate relative to metaphase II oocytes will be compared ...". Under "**Circulating Levels of Endocrine Parameters**": The text "For each parameter the change from baseline will be compared between treatment groups using an analysis of covariance model (ANCOVA)." is changed to "Each parameter will be compared between treatment groups using an analysis of covariance model (ANCOVA). No subgroup analysis will be performed per age group". The text "In this model change from baseline in ln-transformed measurements will be the dependent variable and ..." is changed to "In this model ln-transformed measurements will be the dependent variable and ...". With this change it will be possible to obtain adjusted treatment means.	
2.0	25 July 2019	<u>Section 10.2 Adverse Events</u> The text "Adverse events will be coded using MedDRA. The version of MedDRA will be documented." is changed to "Adverse events will be coded using MedDRA version 20.0."	1.0
2.0	25 July 2019	<u>Section 10.2.2 Incidence of Adverse Events</u> The text "... adverse events with an incidence of ≥5% in any treatment group, and non-serious adverse events with an incidence of ≥5% in any treatment group." is changed to "... adverse events with an incidence of ≥2% in any treatment group, and non-serious adverse events with an incidence of ≥2% in any treatment group."	1.0
2.0	25 July 2019	<u>Section 10.10 Pregnancy Follow-up Information</u> This section has been deleted since it is out of the scope of this SAP.	1.0
3.0	22 Aug 2019	<u>Section 1.1 Definitions/ Abbreviations</u> The definition of "Completer" is added to "Table 1.1.1 Definition of Terms".	2.0

3.0	22 Aug 2019	Section 9.3 Important Secondary Endpoint Under "**Clinical Pregnancy Rate**": The text "The analyses will be based on all subjects as well as all subjects with transfer." is changed to "The analyses will be done based on all subjects, all subjects with oocytes retrieved, and all subjects with transfer.".	2.0
3.0	22 Aug 2019	Section 9.4 Secondary Endpoints Under "**Cycle Cancellation due to Poor or Excessive Ovarian Response or Blastocyst Transfer Cancellation due to Excessive Ovarian Response / OHSS Risk**": The text "If the expected number of observations is less than five in any of the cells in the contingency table then Fisher's exact test will be used as alternative." is changed to "In case of sparse data, treatments will be compared using the Cochran-Mantel-Haenszel test stratified for AMH stratum. Within AMH stratum the comparisons will be based on Fisher's exact test.". Under "**Extreme Ovarian Response**": The text "In addition, they will be analysed based on all subjects with triggering of final follicular maturation." is added. Under "**Number and Quality of Embryos on Day 3**": The text "Categorical data (subjects with at least one good-quality embryos) will be compared between treatment groups using a logistic regression model with treatment and AMH stratum as fixed factors and within each AMH stratum using the chi-square test." is added.	2.0
3.0	22 Aug 2019	Section 12 Deviations from Protocol Analysis The text "The description of the analysis set used for analyses, summary tables, listings, and figures are further clarified in the SAP." is added	2.0

Review and Approval of Statistical Analysis Plan

Author (e-signed in REAL):

[redacted]
Director of Biometrics
Ferring Pharmaceuticals Co., Ltd, Japan

Review (Review using PleaseReview or similar system and stored in LSAF):

[redacted]
Vice President, Therapeutic Area, Reproductive Health
Ferring Pharmaceuticals A/S

[redacted]
Medical Director, Therapeutic Area, Reproductive Medicin & Women's Health
Ferring Pharmaceuticals A/S

[redacted]
Pharmacovigilance Physician, Global Pharmacovigilance
Ferring Pharmaceuticals A/S

[redacted]
Senior Director, Medical Writing, Global Development
Ferring Pharmaceuticals A/S

[redacted]
Lead Statistical Programmer, Global Biometrics
Ferring Pharmaceuticals A/S

Approval (e-signed in REAL):

[redacted]
Senior Director, Global Biometrics
Ferring Pharmaceuticals A/S

TABLE OF CONTENTS

1 Introduction

This document describes the planned statistical analyses for 000273 and is based on the final protocol version. 2.0 dated 16 March 2018 and the protocol amendment Version 1.0 (dated 13th November 2018). No changes are done to the analyses planned in the protocol. Details and context of the analyses described in the protocol have been clarified and some additional analyses have been included to make the result description more complete.

1.1 Definitions/ Abbreviations

1.1.1 Definition of Terms

Terms	Definitions
ESTHER-1 trial	Phase 3 trial, ESTHER-1 (Evidence based Stimulation Trial with Human rFSH in Europe and Rest of World), conducted in 11 countries overseas including Europe, North America and Latin America. Ferring trial ID 000004.
Randomised	Subject randomised to trial treatment
Screened	Subject who signed informed consent
Completer	A subject is considered a completer if she attends all scheduled trial visits and undergoes the end-of-trial assessments. A subject who is not a completer will be considered a non-completer subject.

1.1.2 Abbreviations

Abbreviations	Meaning of abbreviations in document
AE	Adverse event
AMH	Anti-Müllerian hormone
ANCOVA	Analysis of covariance
ART	Assisted reproductive technologies
βhCG	Beta unit of human chorionic gonadotropin
CV	Coefficient of variation
eCRF	Electronic case report form
EMA	European Medicines Agency
FAS	Full analysis set
FSH	Follicle-stimulating hormone
GnRH	Gonadotropin-releasing hormone
hCG	Human chorionic gonadotropin
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th revision
ICSI	Intracytoplasmic sperm injection
IMP	Investigational Medicinal Product
ITT	Intention-to-treat

Abbreviations	Meaning of abbreviations in document
IVF	In vitro fertilisation
LH	Luteinising hormone
LLOQ	Lower limit of quantification
LSAF	Life Science Analytics Framework
MedDRA	Medical Dictionary for Regulatory Activities
OHSS	Ovarian hyperstimulation syndrome
OR	Oocytes retrieved
PMDA	Pharmaceuticals and Medical Devices Agency
PP	Per protocol
PT	Preferred term
SAE	Serious adverse event
SAP	Statistical analysis plan
SOC	System organ class
ULN	Upper limit of normal
ULOQ	Upper limit of quantification

2 Trial Objectives and Endpoints

Primary Objective

- To demonstrate non-inferiority of FE 999049 compared to FOLLISTIM with respect to number of oocytes retrieved in Japanese IVF/ICSI patients undergoing controlled ovarian stimulation

Secondary Objectives

- To compare FE 999049 with FOLLISTIM with respect to achieving pregnancy
- To compare FE 999049 with FOLLISTIM with respect to ovarian response and endocrine profile
- To compare FE 999049 with FOLLISTIM with respect to oocyte fertilisation and number and quality of embryos/blastocysts
- To compare FE 999049 with FOLLISTIM with respect to the frequency of OHSS and/or preventive interventions for early OHSS
- To compare FE 999049 with FOLLISTIM with respect to safety profile, including adverse events, routine safety laboratory parameters and local tolerability
- To compare FE 999049 with FOLLISTIM with respect to gonadotropin use

Primary Endpoint

- Number of oocytes retrieved

Important Secondary Endpoint

- Clinical pregnancy rate (at least one gestational sac 5-6 weeks after transfer)

Secondary Endpoints

- Positive βhCG rate (positive serum βhCG test 13-15 days after transfer)
- Vital pregnancy rate (at least one intrauterine gestational sac with fetal heart beat 5-6 weeks after transfer)
- Implantation rate (number of gestational sacs 5-6 weeks after transfer divided by number of blastocysts transferred)
- Proportion of subjects with cycle cancellation due to poor or excessive ovarian response or blastocyst transfer cancellation due to excessive ovarian response / OHSS risk
- Proportion of subjects with <4, 4-7, 8-14, 15-19 and ≥20 oocytes retrieved
- Proportion of subjects with extreme ovarian responses, defined as <4, ≥15 or ≥20 oocytes retrieved
- Proportion of subjects with preventive interventions for early OHSS
- Proportion of subjects with early OHSS (including OHSS of moderate/severe grade) and/or preventive interventions for early OHSS

- Proportion of subjects with late OHSS (including OHSS of moderate/severe grade)
- Proportion of subjects with OHSS (early and/or late) and/or preventive interventions for early OHSS
- Number and size of follicles on stimulation day 6 and end-of-stimulation
- Fertilisation rate as well as number and quality of embryos on day 3 and blastocysts on day 5 after oocyte retrieval
- Circulating concentrations of FSH, LH, estradiol, progesterone, inhibin A and inhibin B on stimulation day 6 and end-of-stimulation
- Total gonadotropin dose and number of stimulation days
- Frequency and intensity of adverse events
- Changes in circulating levels of clinical chemistry and haematology parameters and proportion of subjects with markedly abnormal changes
- Frequency and intensity of injection site reactions (redness, pain, itching, swelling and bruising) assessed by the subject during the stimulation period
- Technical malfunctions of the administration pens

Pregnancy Follow-up Information

- Ongoing pregnancy rate (at least one intrauterine viable fetus 10-11 weeks after transfer) and live birth rate
- Neonatal health at birth and at 4 weeks after birth

3 Trial Design

3.1 General Design Considerations

This is a randomised, assessor-blind, controlled, parallel groups, multicentre trial assessing the efficacy and safety of FE 999049 in its individualised dosing regimen when used in first cycle Japanese patients aged 20-40 years undergoing controlled ovarian stimulation for IVF/ICSI following a GnRH antagonist protocol.

The trial has been designed to demonstrate non-inferiority of FE 999049 versus an rFSH product approved in Japan, i.e. FOLLISTIM, with respect to number of oocytes retrieved.

Eligible subjects will be randomised in a 1:1 ratio to controlled ovarian stimulation with FE 999049 or FOLLISTIM. Randomisation will be stratified by centre and according to AMH levels at screening (<15 pmol/L and ≥15 pmol/L). The randomisation strata by AMH reflect the FE 999049 dosing regimen, and therefore each subject's AMH value is rounded off to the nearest integer before allocation into one of the two strata, <15 pmol/L or ≥15 pmol/L. The rounding off rules used for AMH with respect to randomisation strata are identical to those described in Section 5.1.1 "FE 999049 Dosing Regimen" of the protocol. Consequently, the randomisation stratum AMH <15 pmol/L corresponds to the description of the AMH <15 pmol/L category in the FE 999049 dosing regimen, meaning that it includes AMH values up to and including 14.49 pmol/L. The

randomisation stratum AMH ≥15 pmol/L includes AMH values from 14.50 pmol/L and higher. These rounding off rules are implemented in the eCRF and used in the automated allocation of subjects into either the AMH <15 pmol/L or ≥15 pmol/L strata.

The last scheduled visit in the main part of the trial is the clinical pregnancy visit (end-of-trial). Non-pregnant subjects should complete the end-of-trial assessments at time of discontinuation. Subjects with a vital pregnancy will be followed for pregnancy outcome as well as neonatal health at birth and at 4 weeks after birth.

3.2 Determination of Sample Size

The primary objective of this trial is to demonstrate non-inferiority of FE 999049 compared with FOLLISTIM with respect to number of oocytes retrieved in Japanese IVF/ICSI patients undergoing controlled ovarian stimulation. The non-inferiority margin for the difference between treatments (FE 999049 versus FOLLISTIM) is -3.0 oocytes. This margin is based on considerations of clinical relevance and overseas regulatory precedence, and has been reviewed and agreed at the follow-up meeting for the Pharmaceuticals and Medical Devices Agency (PMDA) post-phase 2 trial consultation in December 2016.

The standard deviation for the primary endpoint is estimated to 8.1 oocytes based on historical data with FOLLISTIM. The sample size is determined to obtain 90% power to achieve the primary objective for the per-protocol (PP) analysis set for a 1:1 randomisation ratio between FE 999049 and FOLLISTIM and a one-sided t-test at a 2.5% significance level. Assuming the two treatments to be equally effective and a standard deviation of 8.1 oocytes, a sample size of 155 randomised subjects per treatment group would give 90% power. The proportion of subjects with major protocol deviations is assumed to be at most 5% and hence 328 randomised subjects are needed to provide 155 subjects/group in the PP analysis set.

4 Subject Disposition

All screened subjects will be accounted for.

Screened subjects who discontinue from the trial prior to randomisation are regarded as screening failures.

Subject disposition with respect to analysis sets will be tabulated by treatment group overall and by AMH stratum (screening AMH: <15 pmol/L, ≥15 pmol/L) for all randomised subjects. This table will include the number of completed and discontinued subjects including reason for discontinuation. Screening failures and their primary reason for screening failure will also be included. Screening failures will not otherwise be accounted for.

A separate table will summarise the subject disposition with respect to analysis sets by trial site overall and by AMH stratum.

The number of subjects completed and discontinued (including reason) will be summarised for the following trial parts: stimulation, triggering of final follicular maturation, oocyte retrieval, transfer, and pregnancy monitoring. This table will be produced for the full analysis set (FAS) overall and by AMH stratum.

Subject disposition with respect to analysis sets will be listed for all randomised subjects including information on trial completion and reason for discontinuation for non-completers. Subjects who discontinued from the trial will also be listed separately.

5 Protocol Deviations

Major protocol deviations, such as significant non-compliance or other serious unforeseen deviations that may affect the conclusions of the trial will lead to exclusion of data from the PP analysis set. Data will not be excluded from the PP analysis set in case of minor protocol deviations. The list of major protocol deviations include, but is not restricted to:

- Unblinding of assessor or investigator
- IMP regimen (treatment not in accordance with randomisation or non-compliance with IMP for two or more days)
- Non-compliance with the triggering criterion

The rating of protocol deviations as ‘minor’ and ‘major’ will be decided by the medical officer, medical monitor and project statistician based on a blinded review of data before declaration of clean file and lock of database. If the blinded review identifies unforeseen deviations deemed to impact the primary endpoint, these will also be rated as major deviations.

The list of major protocol deviations will be detailed and documented in the clean file document prior to database release. Major protocol deviations will be summarised and listed by subject for the intention-to-treat (ITT) analysis set.

6 Analysis Sets

6.1 Intention-To-Treat Analysis Set

The ITT analysis set is defined as all randomised subjects. Subjects will be analysed according to planned (randomised) treatment.

6.2 Full Analysis Set

The FAS is defined as all randomised and exposed (to IMP) subjects. Subjects will be analysed according to planned (randomised) treatment.

6.3 Per Protocol Analysis Set

The PP analysis set is defined as all randomised and exposed (to IMP) subjects except those excluded as a result of major protocol deviations as described in Section 5.

6.4 Safety Analysis Set

The safety analysis set is defined as all randomised and exposed (to IMP) subjects. Subjects will be analysed according to actual treatment received.

7 Trial Population

All relevant baseline data will be summarised in tables including both treatment groups and a total column. The purpose of these tabulations is to characterise the treatment groups and assess the degree of similarity achieved by the randomisation. Baseline data will not be compared using statistical tests. Unless otherwise noted, tabulations will be produced overall and by AMH stratum for both the FAS and the PP analysis set. Continuous variables will be presented with number of subjects, mean, standard deviation, median, inter-quartile range, minimum, and maximum. Categorical variables will be presented with number and percentage of subjects within each specific category.

Listings will be produced for the ITT analysis set.

7.1 Demographics and Other Baseline Characteristics

7.1.1 Demographics

Demographics and other baseline characteristics (body measurements, ultrasound parameters, vital signs, and endocrine parameters) obtained before first exposure to IMP will be listed by subject and presented in summary tables.

7.1.2 Infertility History, Menstrual History and Reproductive History

Infertility history, menstrual history and reproductive history will be listed by subject and presented in summary tables.

7.2 Medical History

All medical history will be coded using MedDRA version 20.0. Medical history will be listed by subject and summarised for each medical item. This summary table will only be produced overall (i.e. not by AMH stratum) for the FAS.

7.3 Prior and Concomitant Medication

Concomitant medications will be coded using the WHO Drug Reference List. Prior and concomitant medication will be summarised by ATC classification 1st level (alphabetically) and

ATC classification 2nd level (in decreasing order of frequency). These medications will be tabulated separately for the following parameters:

- Prior medication, i.e. medication taken exclusively prior to treatment (i.e. with stop date/time before date/time of 1st IMP administration)
- Concomitant medication, i.e. medication taken during the treatment period (i.e. medication that was not stopped before date/time of 1st IMP administration and not started after the end-of-trial visit)

These tables will only be produced overall (i.e. not by AMH stratum) for the FAS.

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.

Concomitant medications will be listed by subject.

7.4 Physical Examination and Gynaecological Examination

Physical examination and gynaecological examination performed during screening will be summarised per category. These tables will only be produced overall (i.e. not by AMH stratum) for the FAS.

8 Exposure and Treatment Compliance

8.1.1 Extent of Exposure

The total gonadotropin dose and the duration of stimulation (days) will be tabulated based on the safety analysis set. The duration of stimulation (days) will be compared between treatments on the safety analysis set using van Elteren test stratified for AMH and within each AMH stratum using the Wilcoxon's test. Descriptive summaries will be presented overall, and for the two AMH strata.

8.1.2 Treatment Compliance

Treatment non-compliance will be presented in listings as non-compliance is expected to be limited.

9 Efficacy

9.1 General Considerations

Primary and Secondary Endpoints

The results of the analyses of the primary endpoint (number of oocytes retrieved) is essential for the non-inferiority claim. The analyses of the important secondary endpoint and the other secondary endpoints are intended to provide additional characterisation of the safety and efficacy of FE 999049. However, non-inferiority does not need to be established for these endpoints.

Analysis and Presentation of Primary and Secondary Endpoints

Summary tables and treatment comparisons for the primary endpoint, the important secondary endpoint and the secondary efficacy endpoints will be presented overall and by AMH stratum for the FAS. In addition, summary tables and treatment comparisons for the efficacy endpoints will be presented per age group (<35, 35-37 and 38-40 years) unless it is specified otherwise.

All tabulations will present the treatment groups and include a total column. Continuous variables will be presented with number of subjects, mean, standard deviation, median, inter-quartile range, minimum, and maximum. Categorical variables will be presented with number and percentage of subjects within each specific category.

Visual displays will be produced as appropriate. All primary and secondary efficacy endpoints will be listed. Listings will only be produced for the FAS.

Multiplicity

No adjustment for multiplicity is required for the primary endpoint since there is only one hypothesis and one analysis set (FAS). Concerning the secondary endpoints, no formal adjustment for multiplicity will be utilised.

Missing Data

For the primary endpoint, number of oocytes retrieved, no missing data are expected. Subjects that discontinue stimulation, do not continue from stimulation to triggering of final follicular maturation, or do not undergo the oocyte retrieval procedure, will be regarded as having zero oocytes retrieved.

For the secondary endpoints, positive βhCG rate, clinical pregnancy rate, vital pregnancy rate, and implantation rate occurrence of missing data is unlikely but may occur in case the assessment was not done. For these endpoints, a subject's response is considered as 'negative' unless recorded as 'positive'. An exception to this is if a later observation confirms that a previous missing observation was in fact positive. If, for example, the βhCG test result is missing but clinical pregnancy is recorded as 'positive', then the βhCG test result will be imputed as 'positive'.

9.2 Primary Endpoint

9.2.1 Primary Variable Analysis

The primary endpoint 'number of oocytes retrieved' will be analysed using an analysis of variance model with treatment and AMH stratum as fixed factors. The 2-sided 95% confidence limits for the mean treatment differences (FE 999049 - FOLLISTIM) will be calculated based on the fitted model for the FAS.

The hypothesis to be tested is:

H_0: $OR_{FE\ 999049}$ - $OR_{FOLLISTIM} \leq$ -3.0 against

H_1: $OR_{FE\ 999049}$ - $OR_{FOLLISTIM}$ > -3.0

where $OR_{FE\ 999049}$ and $OR_{FOLLISTIM}$ denote the mean number of oocytes retrieved with respective treatment.

The null hypothesis (H_0) will be tested against the alternative (H_1) by constructing a two-sided 95% confidence interval for the difference in mean number of oocytes retrieved. The null hypothesis will be rejected if the lower-limit of the two-sided 95% confidence interval is greater than the non-inferiority limit (-3.0 oocytes) for the FAS. In this case it will be claimed that FE 999049 is non-inferior to FOLLISTIM with respect to number of oocytes retrieved in women undergoing controlled ovarian stimulation.

9.2.2 Sensitivity Analyses and Additional Analysis

The FAS is the primary analysis population for this trial, and the analysis of the PP analysis set is only considered supportive for the application in Japan. In the EMA guideline "Points to consider on switching between superiority and non-inferiority"[1], the FAS and the PP analysis sets are considered equally important. FE 999049 is a compound under global development. Therefore, for applications in countries applying the EMA guideline, Ferring would intend to demonstrate non-inferiority for both the FAS and PP analysis sets, and in such a case both would be considered as confirmatory analyses.

The primary analysis will therefore be repeated using the PP analysis set. This sensitivity analysis will address the assay sensitivity of the primary analysis by excluding subjects with major protocol deviations, e.g. substantial non-compliance with randomised treatment.

To investigate the possibility for different treatment differences in the two AMH strata, an analysis of variance model will be fitted where the treatment-by-stratum interaction is included. In addition, subgroup analyses will be performed for the two AMH strata separately (using a simpler model with only treatment as factor). The trial is not powered to show non-inferiority in the two strata separately and these analyses should only be regarded as descriptive.

Additional descriptions will be made based only on subjects with oocytes retrieved. The purpose of this is purely descriptive.

The percentage of subjects with <4 oocytes, 4-7 oocytes, 8-14 oocytes, 15-19 oocytes, and >20 oocytes retrieved will be tabulated both for all subjects in the FAS and for the two AMH strata.

9.3 Important Secondary Endpoint

Clinical Pregnancy Rate

Clinical pregnancy is defined as at least one gestational sac 5-6 weeks after transfer. The clinical pregnancy rate will be summarised for all subjects, for subjects with at least one oocyte retrieved and for subjects with transfer. The difference between FE 999049 and FOLLISTIM in clinical pregnancy rate will be estimated and a two-sided 95% confidence interval will be constructed using the Mantel-Haenszel method (described below) will to combine results across AMH strata. In brief, this corresponds to deriving a weighted average across AMH strata where the weight depends on the number of observations in each treatment group in each AMH stratum.

The two-sided 95% confidence intervals will be constructed based on the asymptotic normal distribution as

$$\frac{\sum w_i RD_i}{\sum w_i} \pm 1.96 \sqrt{\frac{\sum w_i^2 SE(RD_i)^2}{\left(\sum w_i\right)^2}},$$

where the sums are over the two AMH strata (i.e. $i = 1, 2$), $RD_i = \hat{\pi}_{FE,i} - \hat{\pi}_{FO,i}$ are the observed difference in rates (FE 999049 – FOLLISTIM) within stratum i,

$SE(RD_i) = \sqrt{\frac{\hat{\pi}_{FE,i}(1-\hat{\pi}_{FE,i})}{n_{FE,i}} + \frac{\hat{\pi}_{FO,i}(1-\hat{\pi}_{FO,i})}{n_{FO,i}}}$ is the standards error of the observed difference in rates within stratum i,

$w_i = \frac{n_{FE,i} n_{FO,i}}{n_{FE,i} + n_{FO,i}}$ is the weight assigned to stratum i,

$\hat{\pi}_{FE,i}, \hat{\pi}_{FO,i}$ are the observed rates for each treatment group and stratum and $n_{FE,i}, n_{FO,i}$, are the number of observations in each treatment group and stratum.

To align with the reporting of the ESTHER-1 trial, the above analysis will be repeated using age-group instead of AMH strata as stratification factor. The analyses will be done based on all subjects, all subjects with oocytes retrieved, and all subjects with transfer.

9.4 Secondary Endpoints

Secondary efficacy endpoints will be evaluated based on the FAS only.

Statistical tests will be performed using a two-sided test at a 5% significance level. Treatment differences will (where appropriate) be presented with 95% confidence intervals and p-values

corresponding to the statistical test of the hypothesis of 'equal effect' against the alternative of 'different effect'.

Positive βhCG Rate

Positive βhCG is defined as positive serum βhCG test 13-15 days after transfer. The positive βhCG rate will be summarised and analysed in a similar manner as the clinical pregnancy rate.

Vital Pregnancy Rate

Vital pregnancy is defined as at least one intrauterine gestational sac with fetal heart beat 5-6 weeks after transfer. The vital pregnancy rate will be summarised and analysed in a similar manner as the clinical pregnancy rate. For subjects with a vital pregnancy, the number of intrauterine gestational sacs with fetal heart beat and the number of fetuses with fetal heart beat will be tabulated.

Implantation Rate

Implantation rate is defined as the number of gestational sacs 5-6 weeks after transfer divided by number of blastocysts transferred. The implantation rate will be summarised and analysed in a similar manner as the vital pregnancy rate. The transferred blastocysts are the experimental unit. Note that since single blastocyst transfer is mandatory in this trial, these descriptions and analyses are likely to be identical to a description and analysis of the clinical pregnancy rate in the subgroup of subjects that had a blastocyst transfer.

Cycle Cancellation due to Poor or Excessive Ovarian Response or Blastocyst Transfer Cancellation due to Excessive Ovarian Response / OHSS Risk

For this endpoint, treatment groups will be compared using a logistic regression model with treatment and AMH stratum as fixed factors. The difference between treatments will be reported as an odds ratio including 95% confidence interval and p-value for test of no treatment difference. In case of sparse data, treatments will be compared using the Cochran-Mantel-Haenszel test stratified for AMH stratum. Within AMH stratum the comparisons will be based on Fisher's exact test.

If both the factors treatment and AMH stratum are significant in the logistic regression analysis the possibility of a treatment-by-AMH interaction will be investigated.

The analysis will be performed for the composite endpoint (all cancellations of either the cycle due to poor or excessive response, or of the transfer due to excessive ovarian response/OHSS risk) and each of the separate components defined as: the proportion of subjects with cycle cancellation due to poor ovarian response, the proportion of subjects with cycle cancellation due to excessive ovarian response, the proportion of subjects with cycle cancellation due to poor or excessive ovarian response, the proportion of subjects with blastocyst transfer cancellation due to excessive ovarian response/OHSS risk, and the combination of cycle cancellations due to excessive response and transfer cancellation due to excessive ovarian response/OHSS risk.

Ovarian Response

The number of oocytes retrieved will be tabulated with subjects grouped according to number of oocytes retrieved (<4 (low response), 4-7 (moderate response), 8-14 (targeted response), 15-19 (hyperresponse) and ≥20 (severe hyperresponse)). Subjects with cycle cancellation due to poor ovarian response will be included in the <4 oocytes group. Subjects with cycle cancellation due to excessive ovarian response will here be included in the ≥20 oocytes group. For each definition the proportion of subjects will be tabulated. These endpoints will be analysed in a similar manner as the proportion of subjects with cycle cancellation.

Extreme Ovarian Response

The analyses of extreme ovarian response will be performed using the following definitions of extreme ovarian response: <4 oocytes retrieved, ≥15 oocytes retrieved, ≥20 oocytes retrieved, <4 or ≥15 oocytes retrieved and <4 or ≥20 oocytes retrieved. Subjects with cycle cancellation due to poor ovarian response will be included as <4 oocytes retrieved. Subjects with cycle cancellation due to excessive ovarian response will be included as ≥15 and ≥20 oocytes retrieved.

For each definition the proportion of subjects with extreme ovarian response will be tabulated. These endpoints will be analysed in a similar manner as the proportion of subjects with cycle cancellation. In addition, they will be analysed based on all subjects with triggering of final follicular maturation.

Number and Size of Follicles during Stimulation

The follicle cohort on stimulation day 6 and end-of-stimulation will be summarised by treatment on the follicle level (number of follicles 8-9 mm, 10-11 mm, 12-14 mm, 15-16 mm and ≥17 mm) and on the subject level (total number of follicles, size of largest follicle, average follicle size, average size of three largest follicles, and number of follicles ≥8 mm, ≥10 mm, ≥12 mm, ≥15 mm and ≥17 mm). Continuous data will be compared between treatment groups using van Elteren test stratified for AMH and within each AMH stratum using the Wilcoxon's test.

Fertilisation Rate

An oocyte is defined as fertilised if it is scored as 2 pronuclei at 19 hours. For subjects with oocytes retrieved, the rate of fertilised oocytes to oocytes retrieved (and also the rate of fertilised oocytes to metaphase II oocytes for those inseminated using ICSI) will be tabulated. Further, the number of fertilised oocytes per subject will be tabulated including both summary statistics and a frequency table. Number of fertilised oocytes, number of metaphase II oocytes, fertilisation rate, the rate of metaphase II oocytes to oocytes retrieved, and fertilisation rate relative to metaphase II oocytes will be compared between treatment groups using van Elteren test stratified for AMH and within each AMH stratum using the Wilcoxon's test. Note that the analysis of fertilisation rates excludes subjects without oocytes retrieved and do therefore not adhere to the randomisation principle.

Number and Quality of Embryos on Day 3

The number of embryos on day 3 including a breakdown by selected quality parameters will be tabulated including both summary statistics and frequency tables. Further, for subjects with oocytes retrieved, the rate of embryos to oocytes retrieved will be summarised overall and by selected quality parameters. Numbers and rates will be compared between treatment groups using van Elteren test stratified for AMH and within each AMH stratum using the Wilcoxon's test. Categorical data (subjects with at least one good-quality embryos) will be compared between treatment groups using a logistic regression model with treatment and AMH stratum as fixed factors and within each AMH stratum using the chi-square test.

Number and Quality of Blastocysts on Day 5

The number of blastocysts on day 5 including a breakdown in quality scores will be tabulated including both summary statistics and frequency tables. The number of subjects with at least one good-quality blastocyst, i.e. of grade 3BB or higher, will be reported. Further, for subjects with oocytes retrieved, the rate of blastocysts to oocytes retrieved will be summarised overall and by quality score. Numbers and rates will be compared between treatment groups using van Elteren test stratified for AMH and within each AMH stratum using the Wilcoxon's test. Categorical data (subjects with at least one good-quality blastocyst) will be compared between treatment groups using a logistic regression model with treatment and AMH stratum as fixed factors and within each AMH stratum using the chi-square test.

Circulating Levels of Endocrine Parameters

Blood samples drawn at stimulation days 1 and 6 and end-of-stimulation are analysed for FSH, LH, estradiol, progesterone, inhibin A and inhibin B. Values below the lower limit of quantification (LLOQ) will be included as LLOQ/2. Values above the upper limit of quantification (ULOQ) will be included as ULOQ. For LH, progesterone and LH surge (both LH and progesterone) a summary table will be prepared displaying the proportion of subjects who have at least one markedly abnormal value. The table will also be produced based on data from day 6 only and for end-of-stimulation only. The markedly abnormal criteria is specified in Appendix 1.

Each endocrine parameter and the change from baseline for post-baseline measurements will be tabulated for stimulation day 1 (baseline), stimulation day 6 and end-of-stimulation. Each parameter will be compared between treatment groups using an analysis of covariance model (ANCOVA). No subgroup analysis will be performed per age group. In this model ln-transformed measurements will be the dependent variable and the linear predictor will include treatment and AMH stratum as fixed factors and the baseline measurement (ln-transformed) as covariate. The estimated treatment ratio (FE 999049/FOLLISTIM) with 95% confidence interval will be presented accompanied by the p-value for test of no treatment difference.

10 Safety

10.1 General Considerations

Safety parameters will be evaluated for the safety analysis set.

10.2 Adverse Events

Adverse events will be coded using MedDRA version 20.0.

Adverse events are grouped according to start of IMP as follows:

- Pre-treatment adverse event, i.e. any adverse event occurring after signed informed consent and before start of IMP, or a pre-existing medical condition that worsens in intensity after signed informed consent but before start of IMP.
- Treatment-emergent adverse event, i.e. any adverse event occurring after start of IMP and before the end-of-trial visit, or a pre-treatment adverse event or pre-existing medical condition that worsens in intensity after start of IMP and before the end-of-trial visit.

Treatment-emergent adverse events will be presented in summary tables and listings. Pre-treatment adverse events will be presented in a listing only.

10.2.1 Overview of Treatment-Emergent Adverse Events

A treatment-emergent adverse event overview table will be prepared including the number of subjects reporting an adverse event, the percentage of subjects with an adverse event, and the number of events reported, for the following categories: all adverse events, severe adverse events, adverse drug reactions, adverse events leading to discontinuation, serious adverse events and deaths. An adverse reaction is an adverse event judged by the investigator to be related to IMP with a reasonable possibility.

10.2.2 Incidence of Adverse Events

Treatment-emergent adverse events will be tabulated by system organ class (SOC) alphabetically and preferred term (PT) in decreasing order of frequency. The following will be presented: number of subjects reporting an adverse event, the percentage of subjects with an adverse event, and the number of events reported.

Summary tables will be produced for the following: all adverse events, adverse events by causality (reasonable possibility / no reasonable possibility), adverse events leading to death, adverse events by intensity (mild / moderate / severe), adverse reactions by intensity (mild / moderate / severe), serious adverse events, adverse events leading to discontinuation, adverse events with an incidence of ≥2% in any treatment group, and non-serious adverse events with an incidence of ≥2% in any treatment group.

10.3 OHSS

Early OHSS (Including OHSS of Moderate/Severe Grade) and/or Preventive Interventions for Early OHSS

OHSS will for each treatment group be tabulated by classification (mild, moderate, severe, moderate or severe) and grade (1, 2, 3, 4, 5). 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.

This endpoint will be tabulated and analysed in a similar manner as the proportion of subjects with cycle cancellation, i.e. using a logistic regression model with treatment and AMH stratum as fixed factors. Analyses will be performed for subjects with early OHSS, subjects with early OHSS of moderate or severe grade, subjects with preventive interventions for early OHSS, subjects with early OHSS and/or preventive interventions for early OHSS, and subjects with early OHSS of moderate or severe grade and/or preventive interventions for early OHSS. The difference between treatments will be reported as an odds ratio including a 95% confidence interval and a p-value for the test of no difference between treatments.

The risk of preventive interventions for early OHSS and/or early OHSS is related to the ovarian response potential. AMH is a well-established predictor of ovarian response to gonadotropin treatment and was confirmed to be the best endocrine marker of ovarian response to FE 999049 treatment in the phase 2 trial. Since FE 999049 is dosed based on AMH it is likely that the relationship between AMH and the risk of experiencing preventive interventions for early OHSS and/or early OHSS differs between FE 999049 and FOLLISTIM. To further explore this for each of the endpoints above, two different logistic regression models will be fitted assuming an increasing risk with increasing AMH with two nested models as follows:

Model 1: $\text{LOGIT}(\pi) = \beta_0 + \beta_1 \cdot \ln(\text{AMH})$,

where π denotes the risk and the β's denotes the regression coefficients.

Model 2: $\text{LOGIT}(\pi) = \beta_0 + \beta_1 \cdot \ln(\text{AMH}) + \beta_{2,i} + \beta_{3,i} \cdot \ln(\text{AMH})$, where i=1, 2 indicates the two treatments. This is model 1 with addition of treatment and treatment-by-ln(AMH) interaction terms.

The two models will be compared using the likelihood ratio test. Adjusted odds ratio estimates (comparing FE 999049 to FOLLISTIM) and associated 95% Wald CI will be provided. A statistically significant improvement with model 2 compared to model 1 should be interpreted as a treatment difference in that the relationship between AMH and the risk of experiencing preventive interventions for early OHSS and/or early OHSS.

The estimated risks based on the models will be plotted as a function of AMH, i.e. overall and for each treatment group. The plots will include observed incidences for AMH-based subgroups of

subjects. The adequacy of the model fits will be evaluated using the Hosmer-Lemeshow goodness-of-fit test and by examining how well the model fit describes the data in the plot.

It can be noted that the likelihood ratio test using two nested models above corresponds to a test of the contrast $\beta_{2,1} = \beta_{2,2} = \beta_{3,1} = \beta_{3,2} = 0$ in model 2 above.

Late OHSS (Including OHSS of Moderate/Severe Grade)

OHSS will for each treatment group be tabulated by classification (mild, moderate, severe, moderate and severe) and grade (1, 2, 3, 4, 5). Late OHSS is defined as OHSS with onset >9 days after triggering of final follicular maturation.

This endpoint will be tabulated and analysed in a similar manner as the proportion of subjects with early OHSS. The analysis will be performed for subjects with late OHSS, and subjects with late OHSS of moderate or severe grade.

OHSS (Early and/or Late) and/or Preventive Interventions for Early OHSS

This endpoint will be tabulated and analysed in a similar manner as the proportion of subjects with early OHSS and/or preventive interventions for early OHSS. Analyses will be performed for subjects with early and/or late OHSS, subjects with early and/or late OHSS of moderate or severe grade, subjects with early and/or late OHSS and/or preventive interventions for early OHSS, and subjects with early and/or late OHSS of moderate or severe grade and/or preventive interventions for early OHSS.

10.4 Injection Site Reactions to IMP

For each injection site reaction (redness, pain, itching, swelling and bruising), the number of events and number of subjects experiencing those events will be tabulated by time (immediately, 30 minutes, 24 hours), reaction and intensity (none, mild, moderate and severe).

10.5 Pen Malfunction

The frequency of technical malfunctions of the administration pens will be tabulated.

10.6 Safety Laboratory Variables

Safety laboratory variables will be grouped under “Haematology” and “Clinical Chemistry”.

The baseline is based on the blood sample drawn at screening. Treatment-emergent laboratory data will be obtained at end-of-stimulation and end-of-trial.

10.6.1 Summary Statistics

The circulating levels of clinical chemistry and haematology parameters including change from baseline will be tabulated for each time-point for each laboratory variable.

10.6.2 Laboratory Variable Changes Relative to Normal Range

Shift tables will be prepared to compare baseline values to the end-of-stimulation and end-of-trial values, using a categorisation of low, normal and high values at each visit. Low, normal and high will be defined according to the reference ranges provided by the central laboratory.

10.6.3 Markedly Abnormal Changes

For each laboratory variable, a summary table will be prepared displaying the proportion of subjects who have at least one markedly abnormal value. The table will also include a break-down by classification of the baseline value. Markedly abnormal criteria for the safety laboratory variables is specified in Appendix 1.

10.6.4 Data Listings

All laboratory values will be listed by subject number and time point. Values outside the reference range and markedly abnormal values will be flagged.

10.7 Vital Signs

10.7.1 Vital Signs

Vital signs and their change from stimulation day 1 (baseline) to end-of-trial will be summarised. Shift tables will be prepared to compare the baseline values with the end-of-trial values using the categorisation of low, normal and high values. Low, normal and high values will are specified in Appendix 1. All vital signs values will be listed per subject. Values outside the reference range will be flagged.

10.8 Physical Examination

Physical examination at end-of-trial compared to baseline (screening) will be summarised in shift tables and all subjects with any abnormal finding will be listed per subject. The list will include both baseline and end-of-trial assessment for comparison.

10.9 Gynaecological Examination

Gynaecological examination at end-of-trial compared to baseline (screening) will be summarised in shift tables and all subjects with any abnormal finding will be listed by subject. The list will include both baseline and end-of-trial assessment for comparison.

11 Interim Analyses

No interim analysis is planned.

12 Deviations from Protocol Analysis

There are no deviations from the analyses planned in the protocol. Details and context of the analyses described in the protocol have been clarified in this SAP and some additional analyses

have been included to make the result description more complete. The description of the analysis set used for analyses, summary tables, listings, and figures are further clarified in the SAP.

13 References

1 European Medicines Agency (EMA). Points to consider on switching between superiority and non-inferiority. CPMP/EWP/482/99.

14 Tables, Listings and Figures

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

Appendix 1 Markedly Abnormal Laboratory Values and Vital Signs

Table 1: Markedly Abnormal Criteria for Haematology

Variable	Units	Markedly abnormal criteria	
		Low	**High**
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

RBC: Red blood cells, WBC: White blood cells

Table 2: Markedly Abnormal Criteria for Clinical Chemistry (including Coagulation)

Variable	Units	Markedly abnormal criteria	
		Low	**High**
Albumin	g/L	< 20	Not applicable
ALT	IU/L	Not applicable	> 3xULN
Alkaline phosphatase	IU/L	Not applicable	> 3xULN
AST	IU/L	Not applicable	> 3xULN
Bicarbonate	mmol/L	< 15.1	> 34.9
Bilirubin direct	umol/L	Not applicable	> 2xULN
Bilirubin total	umol/L	Not applicable	> 2xULN
Blood urea nitrogen	mmol/L	Not applicable	> 12.5
Calcium	mmol/L	< 1.75	> 2.74
Chloride	mmol/L	Not applicable	Not applicable
Cholesterol (30-40Y)	mmol/L	Not applicable	> 10.34
Creatinine	umol/L	Not applicable	> 3xULN
GGT	IU/L	Not applicable	> 3xULN
Glucose	mmol/L	< 2.5	> 16.7
LDH	IU/L	Not applicable	> 3xULN
Phosphate	mmol/L	< 0.6	Not applicable
Potassium	mmol/L	< 3.0	> 5.8
Sodium	mmol/L	< 125	> 155
Total protein	g/L	< 20	> 90
Uric acid	umol/L	Not applicable	> 595

MCV: Mean corpuscular volume, MCH: Mean corpuscular haemoglobin, MCHC: Mean corpuscular haemoglobin concentration, ALT: Alanine Aminotransferase, AST: Aspartate Aminotransferase, GGT: Gamma Glutamyl Transferase, LDH: Lactate Dehydrogenase.

Table 3: Markedly Abnormal Criteria for Endocrine Parameters

Variable	Units	Markedly abnormal criteria (High)
LH	IU/L	> 10
Progesterone	nmol/L	> 3.18
LH surge	Not applicable	LH > 10 IU/L and Progesterone > 3.18 nmol/L on the same day

LH: Luteinizing Hormone

Table 4: Reference Ranges and Markedly Abnormal Criteria for Vital Signs

Variable	**Reference range**		**Markedly abnormal criteria**	
	Low	**High**	**Low**	**High**
Systolic blood pressure (mmHg)	<90	>140	< 80 and decrease of ≥ 20 from baseline	> 155 and increase of ≥ 20 from baseline
Diastolic blood pressure (mmHg)	<60	>90	< 50 and decrease of ≥ 15 from baseline	> 100 and increase of ≥ 15 from baseline
Pulse rate (bmp)	<50	>100	< 45 and decrease of ≥ 15 from baseline	> 130 and increase of ≥ 15 from baseline