FERRING
PHARMACEUTICALS

Statistical Analysis Plan

Title of trial: A randomised, controlled, assessor-blind, parallel groups, multicentre, Pan-Asian trial comparing the efficacy and safety of FE 999049 with follitropin alfa (GONAL-F) in controlled ovarian stimulation in women undergoing an assisted reproductive technology programme
NCT number: NCT03296527
Sponsor trial code: 000145
Date: 08 December 2017

STATISTICAL ANALYSIS PLAN

A randomised, controlled, assessor-blind, parallel groups, multicentre, Pan-Asian trial comparing the efficacy and safety of FE 999049 with follitropin alfa (GONAL-F) in controlled ovarian stimulation in women undergoing an assisted reproductive technology programme

Trial Code: 000145

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:	08DEC2017
Version:	1.0

Change log

Version No.	Date	Reason for the Change / Revision	Supersedes
1.0	08DEC2017	New document	None

Signed Agreement on Statistical Analysis Plan

Author: [redacted]
Project Statistician, Global Biometrics
Ferring Pharmaceuticals A/S

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

[redacted]
Medical Director, Therapeutic Area, Reproductive Health
Ferring Pharmaceuticals A/S

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

[redacted]
Medical Writer, Medical Writing
Ferring Pharmaceuticals A/S

[redacted]
Global Regulatory Project Lead, Global Regulatory Affairs
Ferring Pharmaceuticals A/S

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

[redacted]
Director of Biostatistician, China Clinical Development
Ferring International Pharma-Science Centre (China) Company Limited

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

TABLE OF CONTENTS

1 Introduction

This document describes the planned statistical analyses for 000145 and is based on the final protocol. The protocol is amended in China to have the option of a different NIMP sourcing process. The protocol for South Korea, Taiwan and Vietnam is E-Study Protocol-19954 version 1.0 dated 18 Nov 2016. The consolidated protocol for China (including the changes introduced with the amendment) is E-Study Protocol-21768 version 1.0 dated 11 Sep 2017.

1.1 Definitions/ Abbreviations

1.1.1 Definition of Terms

Term	Definition
Assessor-blind	Assessor-blind means that all investigators, embryologists and central laboratory personnel will be blinded to treatment allocation
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
Pan-Asia	Pan-Asia refers to China and other Asian countries, potentially Taiwan, South Korea and Vietnam
Primary infertility	No previous clinical pregnancy
Randomised	Subject who was randomised to trial treatment
Screened	Subject who signed informed consent

1.1.2 Abbreviations

Abbreviation	Meaning of abbreviation in document
AE	Adverse event
AMH	Anti-Müllerian hormone
ANCOVA	Analysis of covariance
ART	Assisted reproductive technologies
βhCG	Beta unit of human chorionic gonadotropin
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
IVF	In vitro fertilisation
LH	Luteinising hormone
LLOQ	Lower limit of quantification

Abbreviation	Meaning of abbreviation in document
MedDRA	Medical Dictionary for Regulatory Activities
NIMP	Non-investigational medicinal product
OHSS	Ovarian hyperstimulation syndrome
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

2.1 Objectives

Primary Objective

- To demonstrate non-inferiority of FE 999049 compared with GONAL-F with respect to ongoing pregnancy rate in women undergoing controlled ovarian stimulation

Secondary Objectives

- To compare the clinical benefits of FE 999049 in its dosing regimen to those of GONAL-F with respect to efficacy and safety
- To compare FE 999049 with GONAL-F with respect to ovarian response including follicular development and endocrine profile, as well as with respect to embryo development
- To assess the population pharmacokinetics of FE 999049
- To compare FE 999049 with GONAL-F with respect to treatment efficiency
- To compare FE 999049 with GONAL-F with respect to safety profile, including adverse events, routine safety laboratory parameters and local tolerability
- To evaluate the immunogenicity of FE 999049 after one treatment cycle
- To perform a health economic analysis comparing FE 999049 with GONAL-F

2.2 Endpoints

Primary Endpoint

- Ongoing pregnancy rate (at least one intrauterine viable fetus 10-11 weeks after transfer)

Secondary Endpoints

- Positive βhCG rate (positive serum βhCG test 13-15 days after transfer)
- Clinical pregnancy rate (at least one gestational sac 5-6 weeks 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 embryos transferred)
- Ongoing implantation rate (number of intrauterine viable fetuses 10-11 weeks after transfer divided by number of embryos transferred)
- Proportion of subjects with extreme ovarian responses, defined as <4, ≥15 or ≥20 oocytes retrieved
- Proportion of subjects with early OHSS (including OHSS of moderate/severe grade) and/or preventive interventions for early OHSS

- Proportion of subjects with cycle cancellation due to poor or excessive ovarian response or embryo transfer cancellation due to excessive ovarian response / OHSS risk
- Number and size of follicles on stimulation day 6 and end-of-stimulation
- Number of oocytes retrieved and proportion of subjects with <4, 4-7, 8-14, 15-19 and ≥20 oocytes retrieved
- Percentage of metaphase II oocytes (only applicable for those inseminated using ICSI), fertilisation rate as well as number and quality of embryos on day 3 after oocyte retrieval
- Circulating concentrations of LH, estradiol, progesterone, inhibin A and inhibin B on stimulation day 6 and end-of-stimulation
- Circulating concentrations of FSH on stimulation day 6, end-of-stimulation and oocyte retrieval as well as FSH population pharmacokinetic parameters
- Total gonadotropin dose and number of stimulation days
- Proportion of subjects with investigator-requested gonadotropin dose adjustments
- 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
- Proportion of subjects with treatment-induced anti-FSH antibodies, overall as well as with neutralising capacity
- Frequency and intensity of immune-related adverse events
- Proportion of subjects with cycle cancellations due to an adverse event, including immune-related adverse events, or due to technical malfunctions of the administration pen
- Proportion of subjects with late OHSS (including OHSS of moderate/severe grade)
- Rate of multi-fetal gestation, biochemical pregnancy, spontaneous abortion, ectopic pregnancy (with and without medical/surgical intervention) and vanishing twins
- Technical malfunctions of the administration pen

Post-trial Information

- Live birth rate and neonatal health, including minor/major congenital anomalies, at birth and at 4 weeks after birth

3 Trial Design

3.1 General Design Considerations

This is a randomised, controlled, assessor-blind, parallel groups, multicentre, Pan-Asian non-inferiority trial comparing the efficacy and safety of two rFSH preparations, FE 999049 and follitropin alfa (GONAL-F), in first cycle subjects 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, a human cell line-derived rFSH preparation, versus an approved CHO-derived rFSH preparation, i.e. GONAL-F, with ongoing pregnancy rate as the primary endpoint.

Eligible subjects will be randomised in a 1:1 ratio to controlled ovarian stimulation with FE 999049 or GONAL-F. Randomisation will be stratified by trial site and age (<35, 35-37 and 38-40 years).

The last scheduled visit in the main part of the trial is the ongoing pregnancy visit (end-of-trial). The end-of-trial form should be filled out at the subject's last visit, irrespective of whether the subject completed the trial or not.

Subjects with an ongoing 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 GONAL-F with respect to ongoing pregnancy rate in women undergoing controlled ovarian stimulation. The non-inferiority limit for the difference between treatments (FE 999049 versus GONAL-F) is -10.0% (absolute) for the primary endpoint.

Since this is a non-inferiority trial, the full analysis set (FAS) and the per-protocol (PP) analyses are equally important, i.e. non-inferiority should be established for both analysis sets in order to have a robust conclusion[1]. Therefore, the expected proportion of subjects with major protocol deviations should be taken into account when setting the total sample size for the trial. The proportion of subjects with major protocol deviations is assumed to be at most 8%.

The trial is designed to have 90% power of achieving the primary objective for the overall Pan-Asian trial population and 80% power in the Chinese population. The number of subjects needed to achieve 90% power depends on the assumptions regarding the ongoing pregnancy rate. Based on an assumed ongoing pregnancy rate of 32.2% observed in the PP population of the ESTHER 1 trial, N=918 subjects are required in the PP population to obtain a power of 90%. Anticipating at most 8% major protocol deviations this results in N=1000 subjects to be randomised. The power is reasonably robust against deviations from the ongoing pregnancy rate assumption of 32.2%: a 35% ongoing pregnancy rate would still result in a 88.9% power, while lower rates will only increase the power (e.g. 30% yields 91.1% power). Recruiting at least 740

subjects in China is expected to secure that consistency can be established between the results from the overall Pan-Asian trial population and the Chinese trial population. In addition, to obtain approval in other Asian countries at least 260 subjects are needed. Thus a total sample size of 1,000 subjects will be recruited for this trial.

A blinded sample size reassessment will take place when data on the primary endpoint are available for 70% of the planned subjects or when 800 subjects are randomised, whichever comes first. The sample size reassessment will be done without breaking the blind and without inflating the type I error of the trial[2], in line with the current regulatory guidelines for non-inferiority trials by FDA[3] and in general by EMA[4].

A restricted recalculation rule for the sample size reassessment will be used. Accordingly 1,000 subjects should be randomised to achieve sufficient power if the ongoing pregnancy rate is 32.2%. If the ongoing pregnancy rate is above the expected 32.2%, the sample size can be adjusted up to a maximum of 1,144 (572 per treatment group), corresponding to an ongoing pregnancy rate of 50%. If the rate is below the expected 32.2% the sample size will not be decreased below the planned 1,000 subjects.

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. Number of screened subjects and their primary reason for screening failure will be summarised. Screening failures will not otherwise be accounted for.

Subject disposition will be summarised by treatment group (FE 999049, GONAL-F, Total) and age stratum (All age strata, <35, 35-37, 38-40 years) for all randomised subjects. Separate tables will be produced by country.

Subject disposition with respect to analysis sets will include the number of subjects that completed the trial and discontinued the trial, including reason for trial discontinuation.

Subject disposition with respect to analysis sets and age stratum will be summarised in a separate table by trial site.

The number of subjects completing the procedures (including the reasons for not completing) will be summarised for the following trial parts: stimulation, triggering of final follicular maturation, oocyte retrieval, transfer, and pregnancy monitoring.

Subject disposition with respect to analysis sets will be listed for all randomised subjects including information on trial completion and reason for discontinuation from the trial 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 deemed to invalidate the data and 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
- Treatment not in accordance with randomisation
- Non-compliance with IMP for two or more days
- Administration of hCG for triggering of final follicular maturation was not given despite the triggering criterion for hCG administration was met
- GnRH agonist criterion met but hCG used for triggering of final follicular maturation
- GnRH agonist criterion not met but GnRH agonist used for triggering of final follicular maturation
- Non-compliance with the number of embryos transferred

The rating of protocol deviations in 'minor' and 'major' will be decided by the Ferring clinical team on the basis of a blinded review of data before declaration of clean file and lock of database. If the blinded review identifies serious unforeseen deviations deemed to impact the primary endpoint and affect the conclusions of the trial 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 tabulated and listed by subject for all randomised subjects.

6 Analysis Sets

6.1 Intention-to-Treat Analysis Set

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

6.2 Full Analysis Set

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

6.3 Per Protocol Analysis Set

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

6.5 Subgroups

Data will in some cases be reported using the following subgroups:

- Subjects grouped according to AMH at screening (<15 pmol/L, ≥15 pmol/L)
- Subjects with at least 1 oocyte retrieved
- Subjects with transfer
- Subjects grouped according to number of embryos transferred (single, double)
- Subjects grouped according to ongoing pregnancy type (single, twin, triplet)
- Subjects grouped according to insemination method
- Subjects grouped according to primary reason for infertility
- Subjects grouped according to primary infertility
- Subjects grouped according to smoking status

7 Trial Population

All relevant baseline data will be summarised by treatment group (FE 999049, GONAL-F, Total) and age stratum (All age strata, <35, 35-37, 38-40 years) for both the FAS and the PP analysis sets. Separate tables will be produced by country. 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. 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 (smoking and alcohol habits, body measurements, ultrasound parameters, vital signs, and endocrine parameters) obtained before first exposure to IMP will be listed and presented in summary tables.

7.1.2 Infertility History, Menstrual History and Reproductive History

Infertility history (including previous fertility treatment), menstrual history and reproductive history will be listed and presented in summary tables.

7.2 Medical History

All medical history will be coded using MedDRA. The version of MedDRA will be documented. Medical history will be listed and summarised for each medical item. This summary table will be produced overall (i.e. not by age stratum) and by country for the FAS.

7.3 Prior and Concomitant Medication

Concomitant medications will be coded using the World Health Organization (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:

- 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 be produced overall (i.e. not by age stratum) and by country for the FAS.

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

Medications will be listed.

7.4 Physical Examination and Gynaecological Examination

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

8 Exposure and Treatment Compliance

8.1 Extent of Exposure

Duration of treatment (days) is defined as the number of days from first exposure to the day of last exposure (both inclusive). If a subject misses an intermediate dose she will still be considered as being under treatment.

Exposure will be summarised by treatment group (FE 999049, GONAL-F, Total) and age stratum (All age strata, <35, 35-37, 38-40 years) for the safety analysis set. Separate tables will be produced by country. Treatment comparisons will be produced for the overall Pan-Asian trial population and for the Chinese population.

Total Gonadotropin Dose and Number of Stimulation Days

Exposure to gonadotropins will be summarised as the total gonadotropin dose and the number of stimulation days. The total dose will be reported in µg for FE 999049 and both in IU and µg for GONAL-F. For subjects in the FE 999049 group the dose on stimulation day 1 will be summarised. The total gonadotropin dose and the number of stimulation days will be compared between treatments using van Elteren test stratified for age and within each age stratum using the Wilcoxon's test.

Gonadotropin Dose Adjustments

Investigator-requested decreases and increases of the gonadotropin dose will be captured during the stimulation period. The requested dose change (decrease / increase / no change) on stimulation day 6 will be tabulated. Further, the number of dose increase requests and number of dose decrease requests per subject will be tabulated. The requested dose change on stimulation day 6 will be compared between treatments using chi-square tests overall and within age stratum.

GnRH Antagonist

Exposure to GnRH antagonist will be summarised as the total dose administered (mg) and duration of treatment (days).

Triggering with hCG or GnRH agonist

Triggering criteria met (hCG, GnRH agonist or none) and drug used for triggering (hCG or GnRH agonist) will be summarised.

Progesterone

Exposure to progesterone will be summarised as the total dose administered (mg) and duration of treatment (days).

8.2 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 (ongoing pregnancy rate) based on the overall Pan-Asian trial population are essential for the non-inferiority claim. The results obtained for the Chinese trial population is considered supportive of the overall conclusion. If the results for the Chinese trial population are markedly different from those obtained based on the overall Pan-Asian trial population the reasons for this difference will be examined. The secondary endpoints positive βhCG rate, clinical pregnancy rate, vital pregnancy rate, implantation rate and ongoing implantation rate are considered supportive of the primary endpoint. However, non-inferiority does not need to be established for these secondary endpoints.

The remaining secondary endpoints are intended to provide additional characterisation of the safety and efficacy of FE 999049.

Analysis and Presentation of Primary and Secondary Endpoints

Treatment comparisons will be produced for the overall Pan-Asian trial population and for the Chinese population. Summary tables for the primary endpoint and supportive pregnancy endpoints will be prepared by treatment group and age stratum, and by treatment group and AMH group for both the FAS and PP analysis sets. Separate tables will be produced by country. For the remaining secondary endpoints, summary tables will be prepared by treatment group and age stratum for the FAS.

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 for the FAS.

Multiplicity

No adjustments for multiplicity are required since there is only one primary endpoint and non-inferiority has to be established for both the FAS and the PP analysis sets. In addition, the testing for superiority based on the FAS will be made in a sequential manner, only if non-inferiority is established.[5] The evaluation of the subjects recruited in China will be regarded as supportive. Concerning the secondary endpoints no formal adjustment for multiplicity will be utilised.

Missing Data

For the primary endpoint, ongoing pregnancy rate, occurrence of missing data is unlikely but may occur in case the assessment was not done. For this endpoint, a subject's response is considered as 'negative' unless recorded as 'positive'. Missing observations for the supportive secondary endpoints positive βhCG rate, clinical pregnancy rate and vital pregnancy rate will be 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. For example, if the βhCG test result is missing but clinical pregnancy is recorded as 'positive' then the βhCG test result will be regarded as 'positive'.

For subjects with transfer but missing information on the number of viable fetuses 10-11 weeks after transfer, the number of viable fetuses will be imputed as zero irrespective of why data are not recorded. For subjects with transfer but missing information on the number of gestational sacs 5-6 weeks after transfer, the number of gestational sacs will be imputed as the number of viable fetuses at 10-11 weeks after transfer.

9.2 Primary Endpoint

9.2.1 Primary Variable Analysis

This trial has one primary endpoint: Ongoing pregnancy rate. Ongoing pregnancy is defined as at least one intrauterine viable fetus 10-11 weeks after transfer.

The primary objective of this trial is to demonstrate non-inferiority of FE 999049 compared with GONAL-F with respect to the primary endpoint in women undergoing controlled ovarian stimulation. The non-inferiority limit for the difference between treatments (FE 999049 versus GONAL-F) is -10.0% (absolute).

The non-inferiority hypothesis to be tested for the primary endpoint is

$$H_0: \pi_{FE} - \pi_{GF} \leq -10.0\% \text{ against the alternative } H_A: \pi_{FE} - \pi_{GF} > -10.0\%,$$

where π_{FE} and π_{GF} denote the ongoing pregnancy rate after treatment with FE 999049 and GONAL-F, respectively. Non-inferiority will be evaluated based on the overall Pan-Asian trial population.

The primary analyses will be adjusted for the stratification factor (age group) by using the Mantel-Haenszel method to combine results across age strata. In brief, this corresponds to deriving a weighted average across age strata where the weight depends on the number of observations in each treatment group in each age stratum.

For the primary endpoint the null hypothesis (H_0) will be tested against the alternative (H_A) by constructing a two-sided 95% confidence interval for the difference in ongoing pregnancy rates. If the lower-limit of the two-sided 95% confidence interval is greater than the non-inferiority limit (-10.0%) for both the FAS and the PP analysis set, the null hypothesis will be rejected. In that case it will be claimed that FE 999049 is non-inferior to GONAL-F with respect to ongoing pregnancy rate in women undergoing controlled ovarian stimulation.

If the lower-limit of the two-sided 95% confidence interval for the treatment difference based on the FAS not only lies above the non-inferiority limit (-10.0%) but also above zero then there is evidence of superiority in terms of statistical significance at the 5% level. In this case, the p-value from the test for superiority will be reported and it will be claimed that FE 999049 is superior to GONAL-F for the primary endpoint. The result based on the PP analysis set is not essential for the superiority claim but should lead to a comparable result for a robust interpretation.

There is no need for a multiplicity adjustment when switching from non-inferiority to superiority since it is a simple sequential test procedure. This interpretation is in line with the Food and Drug Administration (FDA) draft guidance on "Multiplicity"[5] and EMA's "Points to consider on switching between superiority and non-inferiority"[1].

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 three age strata (i.e. i = 1 to 3), $RD_i = \hat{\pi}_{FE,i} - \hat{\pi}_{GF,i}$ is the observed difference in rates (FE 999049 - GONAL-F) within stratum i,

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

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

$\hat{\pi}_{FE,i}, \hat{\pi}_{GF,i}$ are the observed rates for each treatment group and stratum, and

$n_{FE,i}, n_{GF,i}$ are the number of observations in each treatment group and stratum.

A supportive evaluation of the subjects recruited in China will be made and it is expected that the findings among the Chinese trial population will be consistent with the overall findings in the trial. If the results based on the Chinese trial population are markedly different from those obtained based on the overall Pan-Asian trial population the reasons for this difference will be examined.

9.2.2 Sensitivity Analyses

Besides stratification by age the randomisation is also stratified by trial site. However, as detailed above, the primary analysis will be based on all data pooled across sites assuming homogeneity of the risk differences across sites. This assumption will be tested using the approach described by Lipsitz et al. (1998).[6] Sites were all subjects have a positive outcome and sites where all subjects have a negative outcome will be excluded from this analysis. In addition sites with at most one subject in any treatment group will be excluded from this sensitivity analysis. If there is statistically significant heterogeneity across sites the implications for the interpretation of the primary result will be discussed.

To evaluate the robustness of the conclusions, the following factors potentially impacting ongoing pregnancy will be evaluated one by one:

- Insemination method
- Primary reason for infertility
- Primary infertility
- Smoking status

For each factor, the primary analysis will be repeated using the factor levels as strata instead of the age stratification. The outcomes of these analyses are considered supportive and should lead to similar results as the primary analysis to ensure robustness of the conclusion.

9.2.3 Supplementary Descriptions

The ongoing pregnancy rate will be summarised for subjects with at least one oocyte retrieved and for subjects with transfer. The purpose of this is purely descriptive.

As described in the protocol Section 6.5.1, transfer of one or two embryo(s) is to be performed on day 3 after oocyte retrieval. The decision to transfer either one or two embryo(s) is based on the subject's age and the quality and availability of embryo(s) on the day of transfer. Since this decision is based on observations performed after randomisation it can be affected by the treatment allocation. It is therefore not planned to perform a formal statistical analysis of the ongoing pregnancy rate adjusting for number of embryos transferred. The rationale for not including these adjustments in the primary analysis is that such an adjustment may hide or exaggerate the differences between treatments.[7] Instead the ongoing pregnancy rate will be presented by number of embryos transferred. Further, a frequency table will be produced to compare treatments with respect to number of embryos transferred.

9.3 Secondary Endpoints

For each secondary endpoint the analysis based on the FAS will be used to assess the statistical significance of differences between treatments. The analysis based on the PP analysis set will be considered supportive and should lead to a similar but not necessarily statistically significant result in order to have a robust interpretation.

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 is considered supportive to the primary endpoint and will therefore be analysed in a similar manner as the primary endpoint.

Clinical Pregnancy Rate

Clinical pregnancy is defined as at least one gestational sac 5-6 weeks after transfer. The clinical pregnancy rate is considered supportive to the primary endpoint and will therefore be analysed in a similar manner as the primary endpoint. For subjects with clinical pregnancy, the type of clinical pregnancy (intrauterine or ectopic) will be tabulated.

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 is considered supportive to the primary endpoint and will therefore be analysed in a similar manner as the primary endpoint. 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 embryos transferred. For implantation rate, the experimental unit will be the transferred embryos, i.e., the analysis will compare the proportions of embryos transferred that results in gestational sacs 5-6 weeks after transfer. The implantation rate will be analysed in a similar manner as the primary endpoint. Note that this endpoint is less relevant than the pregnancy endpoints since it is only defined in the subgroup of subjects that had an embryo transfer, and hence does not adhere to the randomisation principle.

Ongoing Implantation Rate

Ongoing implantation rate is defined as the number of intrauterine viable fetuses 10-11 weeks after transfer divided by number of embryos transferred. For ongoing implantation rate, the experimental unit will be the transferred embryos, i.e., the analysis will compare the proportions of embryos transferred that results in intrauterine viable fetuses 10-11 weeks after transfer. The ongoing implantation rate will be analysed in a similar manner as the primary endpoint.

Ovarian Response and Extreme Ovarian Response

The ovarian response will be defined based on the number of oocytes retrieved, and if the cycle was cancelled, the reason for cycle cancellation. Subjects will be grouped according to ovarian response as:

- Low response (<4 oocytes retrieved or cancellation of cycle due to poor ovarian response)
- Moderate response (4-7 oocytes retrieved)
- Targeted response (8-14 oocytes retrieved)
- Hyperresponse (15-19 oocytes retrieved)
- Severe hyperresponse (≥20 oocytes retrieved or cancellation of cycle due to excessive ovarian response)

Extreme ovarian response will be defined as:

- <4 oocytes retrieved or cancellation of cycle due to poor ovarian response
- ≥15 oocytes retrieved or cancellation of cycle due to excessive ovarian response
- ≥20 oocytes retrieved or cancellation of cycle due to excessive ovarian response
- <4 or ≥15 oocytes retrieved or cancellation of cycle due to poor or excessive ovarian response
- <4 or ≥20 oocytes retrieved or cancellation of cycle due to poor or excessive ovarian response

For each definition of ovarian response and extreme ovarian response the proportion of subjects will be tabulated. Treatments will be compared using a logistic regression model with treatment and age stratum as 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. 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. The possibility of a treatment-by-age interaction will be investigated.

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

The number of subjects with cycle cancellation or embryo transfer cancellation including the reasons for cancellation will be summarised.

The following endpoints will be defined:

- Cycle cancellation or embryo transfer cancellation
- Cycle cancellation due to poor or excessive response, or transfer cancellation due to excessive ovarian response/OHSS risk
- Cycle cancellation due to poor ovarian response
- Cycle cancellation due to excessive ovarian response
- Cycle cancellation due to poor or excessive ovarian response
- Transfer cancellation due to excessive ovarian response/OHSS risk
- Cycle cancellations due to excessive response or transfer cancellation due to excessive ovarian response/OHSS risk

These endpoints will be tabulated and analysed in a similar manner as extreme ovarian response.

Number and Size of Follicles during Stimulation

Based on the transvaginal ultrasound on stimulation day 6 and at end-of-stimulation the following endpoints are defined:

- Total number of follicles
- Size of the largest follicle (mm)
- Average follicle size (mm)
- Average size of three largest follicles (mm)
- Number of follicles ≥8 mm, ≥10 mm, ≥12 mm, ≥15 mm and ≥17 mm

Number and size of follicles will be compared between treatments using the van Elteren test stratified for age and within each age stratum using the Wilcoxon's test.

The follicle cohort on stimulation day 6 and end-of-stimulation will be summarised on the follicle level (number of follicles 2-7 mm, 8-9 mm, 10-11 mm, 12-14 mm, 15-16 mm and ≥17 mm).

Number of Oocytes Retrieved

The number of oocytes retrieved will be compared between treatments using the van Elteren test stratified for age and within each age stratum using the Wilcoxon's test.

Additional summarises will be made for subjects with at least one oocyte retrieved. The purpose of this is purely descriptive.

Metaphase II Oocytes

Oocytes undergoing ICSI will have their maturity stage assessed prior to insemination. The percentage of MII oocytes to oocytes retrieved for subjects where all oocytes are inseminated using ICSI will be tabulated. In the subgroup of subjects where all oocytes are inseminated using ICSI the number of MII oocytes will be compared between treatments using the van Elteren test stratified for age and within each age stratum using the Wilcoxon's test. Subjects will be grouped according to number of MII oocytes as: ≤6, 6-12, ≥13 and the corresponding frequencies tabulated.

Fertilisation Rate

An oocyte is defined as fertilised if it is scored as 2 pronuclei on day 1 after oocyte retrieval. 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. The number of fertilised oocytes and the fertilisation rate will be compared between treatment groups using the van Elteren test stratified for age and within each age stratum using the Wilcoxon's test. Subjects will be grouped according to number of fertilised oocytes as: <4, 4-7, 8-14, 15-19, ≥20 and the corresponding frequencies tabulated. 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

Based on the embryo assessment day 3 the following endpoints are defined:

- Number of embryos
- Number of good-quality embryos
- Subjects with at least one embryo
- Subjects with at least one good-quality embryo
- Rate of embryos to oocytes retrieved
- Rate of good-quality embryos to oocytes retrieved
- Rate of embryos to MII oocytes
- Rate of good-quality embryos to MII oocytes
- Number of cryopreserved embryos
- Subjects with at least one cryopreserved embryo

Numbers and rates will be compared between treatments using the van Elteren test stratified for age and within each age stratum using the Wilcoxon's test. For each definition of 'subjects with at least one' the proportion of subjects will be tabulated and analysed in a similar manner as extreme ovarian response.

Embryo quality will be summarised including a breakdown by selected quality parameters (degree of fragmentation, embryo stage, blastomere uniformity, visual sign of multi-nucleation) and insemination method. Separate summaries will be made for the transferred embryos and for the cryopreserved embryos.

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. Furthermore, blood samples drawn at the oocyte retrieval visit are also analysed for FSH. 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.

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 (and also the oocyte retrieval visit for FSH). For each parameter the change from baseline will be compared between treatments using an analysis of covariance model (ANCOVA). The ANCOVA will be fitted to the relative change from baseline in ln-transformed measurements with treatment and age stratum as fixed factors and the ln-transformed baseline measurement as covariate. The estimated treatment difference with 95% confidence interval will be presented on the scale of measurement (i.e. back-transformed using base e) and accompanied by the p-value for test of no treatment difference.

Summary tables will be prepared for LH and/or progesterone displaying the proportion of subjects who have a markedly abnormal value. These tables will also include a break-down by classification of the baseline value. The markedly abnormal criteria is specified in Appendix 1.

9.4 Other Endpoints

FSH Population Pharmacokinetics

A population pharmacokinetic model describing FSH concentrations following repeated dosing of FE 999049 will be prepared under the responsibility of the Ferring Translational Medicine Department.

These results will be reported separately.

10 Safety

10.1 General Considerations

Safety endpoints will be summarised by treatment group for the safety analysis set. Separate tables will be produced by country.

Missing Data

For adverse events a worst-case approach will be used to impute missing values for causality, intensity and seriousness of the event.

10.2 Adverse Events

Adverse events will be coded using MedDRA. The version of MedDRA will be documented.

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.

If the timing of an AE cannot be established in relation to the administration of IMP, it will be considered as a treatment emergent AE.

In addition to the grouping of adverse events described in the protocol, adverse events will be categorised as 'other significant adverse events'[8] and 'unexpected non-serious adverse drug reactions'. These events will be identified in a blinded review of the data before declaration of clean file and lock of database.

Treatment-emergent adverse events will be presented in summary tables and listings. Pre-treatment adverse events will be presented in listings 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
- Adverse events leading to death

An adverse drug 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 drug reactions by intensity (mild / moderate / severe)
- Serious adverse events
- Adverse events leading to discontinuation
- Other significant adverse events[8]
- Unexpected non-serious adverse drug reactions
- Adverse events with an incidence of ≥5% in any treatment group
- Non-serious adverse events with an incidence of ≥5% in any treatment group

10.3 OHSS

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

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

Early OHSS is defined as OHSS with onset ≤9 days after triggering of final follicular maturation. Note that this includes OHSS with onset before triggering and OHSS with onset during stimulation where triggering is not performed.

The following endpoints are defined:

- 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
- Subjects with early OHSS of moderate or severe grade and/or preventive interventions for early OHSS

For each definition the proportion of subjects will be tabulated. Treatments will be compared using a logistic regression model with treatment and age stratum as factors. Treatment comparisons will be produced for the overall Pan-Asian trial population and for the Chinese population.

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 GONAL-F. 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: $\text{LOGIT}(\pi) = \beta_0 + \beta_1 \cdot \ln(\text{AMH}) + \beta_{2,i} + \beta_{3,i} \ln(\text{AMH})$,
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 GONAL-F) and associated 95% Wald confidence interval 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 should 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)
Late OHSS is defined as OHSS with onset >9 days after triggering of final follicular maturation.

The following endpoints are defined:

- Subjects with late OHSS
- Subjects with late OHSS of moderate or severe grade

These endpoints will be tabulated and analysed in a similar manner as the endpoints defined for early OHSS.

OHSS (Early and/or Late) and/or Preventive Interventions for Early OHSS
The following endpoints are defined:

- 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
- Subjects with early and/or late OHSS of moderate or severe grade and/or preventive interventions for early OHSS

These endpoints will be tabulated and analysed in a similar manner as the endpoints defined for early OHSS.

10.4 Safety Laboratory Variables

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

The baseline is based on the blood sample drawn at stimulation day 1. Treatment-emergent laboratory data will be obtained at end-of-stimulation and end-of-trial.

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

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.

Summary tables by visit will be prepared displaying the proportion of subjects who have a markedly abnormal value. These tables 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.

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

10.5 Injection Site Reactions

Injection site reactions will be tabulated by symptom (redness, pain, itching, swelling, bruising), time after injection (immediately after, 30 minutes after, 24 hours after) and intensity (none, mild, moderate, severe, moderate/severe).

10.6 Treatment-induced Anti-FSH Antibodies

The proportion of subjects with treatment-induced anti-FSH antibodies as well as the proportion of subjects with treatment-induced anti-FSH antibodies with neutralising capacity will be tabulated. Furthermore, for subjects with treatment-induced anti-FSH antibodies, all antibody assessments will be listed.

10.7 Immune-related Adverse Events

Potential immune-related adverse events will be tabulated by SOC and PT using the SMQs for: anaphylactic reaction, angioedema, hypersensitivity reactions and severe cutaneous reactions and the Ferring defined PTs including: asthenia, pyrexia, body temperature increased, chills, influenza-like illness, musculoskeletal pain, syncope and pre-syncope.

10.8 Cycle Cancellations due to an Adverse Event, including Immune-related Adverse Events, or due to Technical Malfunctions of the Administration Pen

Adverse events leading to cycle cancellation will be listed.

10.9 Multi-fetal Gestation, Biochemical Pregnancy, Spontaneous Abortion, Ectopic Pregnancy and Vanishing Twins

The incidence of biochemical pregnancy, spontaneous abortion, ectopic pregnancy and vanishing twins will be reported as part of the adverse events summaries described in Section 10.2.2 .

Early pregnancy loss is defined as pregnancy loss occurring before ongoing pregnancy (βhCG to ongoing pregnancy). The incidence of early pregnancy loss will be tabulated. The denominator will be the subjects at risk for early pregnancy loss, i.e. number of subjects with a positive βhCG. Early pregnancy losses will also be grouped as: 'βhCG to vital pregnancy' and 'vital pregnancy to ongoing pregnancy'. Summaries will be made overall, by vital pregnancy type and by ongoing pregnancy type. The pregnancy types are defined as 'singleton', 'twin', 'triplets' and 'higher-order multiples' based on number of intrauterine viable fetus(es) at the respective assessment.

10.10 Pen Malfunction

Number of subjects reporting technical malfunctions of the administration pen and the type of malfunction will be tabulated.

10.11 Additional Safety Evaluations

10.11.1 Physical Examinations

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

10.11.2 Gynaecological Examination

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

10.11.3 Changes in Body Weight

Body weight will be measured at screening, on stimulation day 1 (baseline) and at end-of-trial. The actual measurements and the change from baseline will be tabulated.

10.11.4 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 reference values. Low, normal and high reference values is specified in the Appendix 1.

For systolic and diastolic blood pressure, a summary table will be prepared displaying the proportion of subjects with a markedly abnormal value at end-of-trial. A break-down of the classification at baseline will be included. The markedly abnormal criteria is specified in Appendix 1. Vital signs will be listed and values categorised as low, high or markedly abnormal will be flagged.

10.12 Pregnancy Follow-up

Pregnancy follow-up information will be collected for all subjects with ongoing pregnancy. All intrauterine viable fetuses at the ongoing pregnancy visits will be accounted for. Neonatal health data will be collected at birth and at 4 weeks after birth for all children born.

All data captured during the pregnancy follow-up will be listed.

Live Birth

Live birth is defined as the birth of at least one live baby. For all live births the type of pregnancy (singleton, twin, triplets and higher-order multiples) will be summarised. Subjects with no information ona live birth will be regarded as not having live birth.

The live birth rate is considered supportive to the primary endpoint and will therefore be analysed in a similar manner as the primary endpoint. Treatment comparisons will be produced for the overall Pan-Asian trial population and for the Chinese population. Summary tables will be prepared by treatment group and age stratum for both the FAS and PP analysis sets. Separate tables will be produced by country. In addition, live birth will be summarised based on the subgroups defined in Section 6.5.

Live Rate at 4 Weeks after Birth

The live rate is defined as all subjects with a live neonate at 4 weeks after birth. The live rate at 4 weeks after birth will be analysed in a similar manner as the primary endpoint. Treatment comparisons will be produced for the overall Pan-Asian trial population and for the Chinese population. Summary tables will be prepared by treatment group and age stratum for both the FAS and PP analysis sets. Separate tables will be produced by country.

Late Pregnancy Loss

A late pregnancy loss is defined as an ongoing pregnancy not followed by live birth. The incidence of late pregnancy losses will be tabulated by treatment group and age stratum for the safety analysis set. Separate tables will be produced by country. The denominator will be the subjects at risk of a late pregnancy loss, i.e. number of subjects with an ongoing pregnancy.

The reason for no live birth (still birth, miscarriage, elective termination) will be tabulated at the fetus/neonate level. In case of elective termination the reason (social circumstances, fetal anomaly, other) will be included. Further, the table will summarise if an autopsy was taken (yes/no). In case the reason for elective termination is 'other' the listing will include the specification as reported by the investigator in the eCRF.

Neonatal Health at Birth and at 4 Weeks after Birth

Neonatal health at birth and at 4 weeks after birth, including minor/major congenital anomalies, will be summaries for all live-born neonates. Summaries includes delivery details, gestational age, gender, birth weight and birth length. Admission to neonatal intensive care unit and/or neonatal care unit will be summaries at birth and at 4 weeks after birth, including the reasons for admissions. The Apgar score at 1, 5 and 10 minutes after birth will be summarised. Neonates will be grouped according to the Apgar score for each assessment time as: 0-3, 4-6, 7-10 and the corresponding frequencies tabulated. These data will be summarised by the treatment group and age stratum, and by treatment group and pregnancy type (singletons, twins), for the safety analysis set. Separate tables will be produced by country.

Incidence of Serious Adverse Events

All serious adverse events will be coded using MedDRA. In addition, congenital anomalies on the fetus/neonates will be coded using ICD-10 and these events will be classified as minor or major in accordance with the EMA guideline.

All serious adverse events will be listed.

Summary tables will be produced for the following:

- Maternal serious adverse events by SOC and PT
- Neonatal serious adverse events by SOC and PT
- Congenital anomalies on the fetus/neonates by ICD-10
- Congenital anomalies on the fetus/neonates classification as major by ICD-10
- Congenital anomalies on the fetus/neonates classification as minor by ICD-10

11 Interim Analyses

No interim analysis intended to compare treatments with respect to efficacy or safety is planned.

A sample size re-assessment is planned and will be done without breaking the blind and without inflating the type 1 error of the trial, in line with the current regulatory guidelines (see Section 3.2).

12 Changes Compared to the Analyses Described in the Trial Protocol

There are no changes to the primary analysis planned in the protocol for the primary endpoint. The sensitivity analyses planned on subgroups of subjects (subjects with oocytes retrieved and subjects with transfer) have been excluded since these analyses do not adhere to the randomisation principle. Descriptive statistics will be used to present the results in the subgroups.

For the secondary endpoints, 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. The proposed analyses for ordinal data have been excluded since there are no ordinal endpoints. For example a grouping based on increasing number of oocytes retrieved is ordinal, however, the corresponding ovarian response (low response, moderate response, target response, hyperresponse, severe hyperresponse) is not ordinal. The treatment comparisons will be carried out as planned based on the ungrouped data.

13 References

[1] The European Agency for the Evaluation of Medicinal Products (EMEA), Committee for Proprietary Medicinal Products (CPMP). Points to consider on switching between superiority and non-inferiority. CPMP/EWP/482/99.

[2] Friede T, Mitchell C, Müller-Velten G. Blinded Sample Size Reestimation in Non-Inferiority Trials with Binary Endpoints. Biometrical Journal 2007; 6, 903-916.

[3] U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Draft guidance March 2010, Non-inferiority Clinical Trials.

[4] The European Agency for the Evaluation of Medicinal Products (EMEA), Committee for Proprietary Medicinal Products (CPMP), ICH E9, Statistical Principles for Clinical Trials, CPMP/ICH/363/96.

[5] U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Draft guidance January 2017, Multiple Endpoints in Clinical Trials.

[6] Lipsitz SR, Dear KGB, Laird NM, Molenberghs G. Tests for Homogeneity of the Risk Difference When Data Are Sparse. Biometrics 1998; 54: 148-160.

[7] The European Agency for the Evaluation of Medicinal Products (EMEA), Committee for Proprietary Medicinal Products (CPMP). Points to consider on adjustments for baseline covariates. CPMP/EWP/2863/99.

[8] The European Agency for the Evaluation of Medicinal Products (EMEA), Committee for Proprietary Medicinal Products (CPMP), ICH E3, Structure and Contents of Clinical Study Reports. CPMP/ICH/137/95.

14 Tables, Listings and Figures

Shells for tables, figures and listings will be presented in a separate document.

Appendix 1 Markedly Abnormal Laboratory Safety Values and Vital Signs

The markedly abnormal laboratory safety values and vital signs presented in this Appendix are FE 999049 project standards. The same criteria were applied to the data captured in the ESTHER trial.

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

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

ALT: Alanine Aminotransferase, AST: Aspartate Aminotransferase, GGT: Gamma Glutamyl Transferase, LDH: Lactate Dehydrogenase.

Table 3: Markedly Abnormal Criteria for Endocrine Parameters

Variable	Unit	Markedly abnormal criteria (high)
LH	IU/L	> 10
Progesterone	nmol/L	> 3.18
LH and progesterone	Not applicable	LH > 10 IU/L and progesterone > 3.18 nmol/L on the same day

LH: Luteinising 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 ≥150 from baseline