FERRING
PHARMACEUTICALS

# Statistical Analysis Plan

| |
|---|
| **Title of trial:**<br>A randomized, double-blind, placebo-controlled, parallel groups, multicenter trial investigating the efficacy and safety of FE 999049 in controlled ovarian stimulation in women aged 35-42 years undergoing assisted reproductive technology |
| **NCT number:**<br>NCT03738618 |
| **Sponsor trial code:**<br>000002 |
| **Date:**<br>28 May 2020 |

# STATISTICAL ANALYSIS PLAN

**A randomized, double-blind, placebo-controlled, parallel groups, multicenter trial investigating the efficacy and safety of FE 999049 in controlled ovarian stimulation in women aged 35-42 years undergoing assisted reproductive technology**

**Recombinant FSH Investigation in the Treatment of Infertility with ART (RITA-2)**

**Trial 000002**

| | |
|---|---|
| **Investigational Medicinal Product** | FE 999049, human recombinant follicle-stimulating hormone (rFSH), solution for subcutaneous injection |
| **Indication:** | Development of multiple follicles and pregnancy after fresh and/or cryopreserved embryo transfer in ovulatory women undergoing assisted reproductive technology (ART) |
| **Phase:** | 3 |
| **Author:** | [redacted] |
| **Version:** | 2.0 |

# Change log

| Version No. | Effective Date | Reason for the Change / Revision | Supersedes |
|---|---|---|---|
| 1.0 | 19 Oct 2018 | Original version | None |
| 2.0 | 28 May 2020 | Additions were made to the analysis plan in order to proactively address the impact, if any, of the COVID-19 pandemic.<br><br>Included are additional summaries of subject disposition at the time of the COVID-19 pandemic trial hold, sensitivity analyses for the primary endpoint that include time periods prior to the pandemic as well as after, summaries of both secondary efficacy and safety analyses broken out by disposition during the pandemic, plans for summarizing deviations that occur due to the pandemic, and summaries for post-trial endpoints relative to subject disposition during the pandemic. | 1.0 |

# Agreement on Statistical Analysis Plan

**Author:**

[redacted]
Biostatistics
Ferring Pharmaceuticals, Inc.

**Review:**

[redacted]
Reproductive Medicine and Maternal Health
Ferring Pharmaceuticals, A/S

[redacted]
Reproductive Health and Urology, Clinical Development
Ferring Pharmaceuticals, Inc.

[redacted]
US Pharmacovigilance
Ferring Pharmaceuticals, Inc.

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

[redacted]
Clinical Operations
Ferring Pharmaceuticals, Inc.

[redacted]
Global Biometrics
Ferring Pharmaceuticals, A/S

**Approval:**

[redacted]
Biometrics
Ferring Pharmaceuticals, Inc.

# TABLE OF CONTENTS

## List of Tables



## List of Figures

# 1 Introduction

This document describes the planned statistical analyses for Trial 000002 based on the protocol dated 25 April, 2018. Additions were made to this document in concert with Protocol Amendment 01, issued 19 March, 2020, specifying a COVID-19 related temporary deferral of cryopreserved cycles.

## 1.1 Definitions/ Abbreviations

### 1.1.1 Definition of Terms

| | |
|---|---|
| Randomised | Subject randomised to trial treatment |
| Screened | Subject who signed informed consent form |

### 1.1.2 Abbreviations

| | |
|---|---|
| βhCG | β unit of human chorionic gonadotropin |
| AMH | anti-Müllerian hormone |
| ART | assisted reproductive technology |
| ATC | Anatomical Therapeutic Chemical Classification System |
| BID | "bis in die", two times daily |
| BMI | body mass index |
| COS | controlled ovarian stimulation |
| FSH | follicle-stimulating hormone |
| GnRH | gonadotropin-releasing hormone |
| hCG | human chorionic gonadotropin |
| HLT | high level term |
| ICSI | intracytoplasmic sperm injection |
| IM | intramuscular |
| IMP | investigational medicinal product |
| ITT | intention-to-treat |
| IU | international unit |
| IVF | in vitro fertilization |
| LH | luteinizing hormone |
| LLOQ | lower limit of quantification |

| | |
|---|---|
| MedDRA | Medical Dictionary for Regulatory Activities |
| mITT | modified intention-to-treat |
| NIMP | non-investigational medicinal product |
| OHSS | ovarian hyperstimulation syndrome |
| PP | per-protocol |
| PT | preferred term |
| rFSH | recombinant follicle-stimulating hormone |
| RITA | Recombinant FSH Investigation in the Treatment of Infertility with ART |
| SAE | serious adverse event |
| SC | subcutaneous |
| SD | standard deviation |
| SMQ | Standardised MedDRA Queries |
| TID | “ter in die”, three times daily |
| TSH | thyroid-stimulating hormone |
| WHO | World Health Organization |

# 2 Trial Objectives and Endpoints

## 2.1 Objectives

### Primary Objective

- To demonstrate the efficacy and safety of FE 999049 in controlled ovarian stimulation

### Secondary Objectives

- To establish the efficacy of FE 999049 in controlled ovarian stimulation with respect to pregnancy rates based on the fresh cycle and/or cryopreserved cycles
- To characterize the ovarian response, including follicular development, oocytes retrieved and endocrine profile, as well as the embryo development associated with controlled ovarian stimulation with FE 999049
- To characterize the safety profile of FE 999049 treatment, including adverse events, routine safety laboratory parameters, local tolerability and immunogenicity

### Post-trial Objectives

- To establish the efficacy of controlled ovarian stimulation with FE 999049 with respect to live birth rates based on the fresh cycle and/or cryopreserved cycles
- To characterize the neonatal health associated with controlled ovarian stimulation with FE 999049, including minor/major congenital anomalies at birth, 4 weeks and 1 year after birth in the fresh cycle and cryopreserved cycles

## 2.2 Endpoints

### Primary Endpoint

- Cumulative ongoing pregnancy rate (at least one intrauterine viable fetus 8-9 weeks after transfer) after the fresh cycle and cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation

**Secondary Endpoints**

- Ongoing pregnancy rate (at least one intrauterine viable fetus 8-9 weeks after transfer) in the fresh cycle and in the cryopreserved cycles
- Time from start of controlled ovarian stimulation to ongoing pregnancy across the fresh and cryopreserved cycles, including duration and number of cycles before achieving ongoing pregnancy
- Ongoing implantation rate (number of intrauterine viable fetuses 8-9 weeks after transfer divided by number of blastocysts transferred) in the fresh cycle, the cryopreserved cycles and cumulatively
- Clinical pregnancy rate (at least one gestational sac 5-6 weeks after transfer) in the fresh cycle, the cryopreserved cycles and cumulatively
- Vital pregnancy rate (at least one intrauterine gestational sac with fetal heart beat 5-6 weeks after transfer) in the fresh cycle, the cryopreserved cycles and cumulatively
- Implantation rate (number of gestational sacs 5-6 weeks after transfer divided by number of blastocysts transferred) in the fresh cycle, the cryopreserved cycles and cumulatively
- Positive βhCG rate (positive serum βhCG test 10-14 days after transfer) in the fresh cycle, the cryopreserved cycles and cumulatively
- Proportion of subjects in the fresh cycle with triggering of final follicular maturation (with hCG, with GnRH agonist, and in total), cycle cancellation and transfer cancellation
- Number and size of follicles on stimulation day 5 and end-of-stimulation
- Number of oocytes retrieved and proportion of subjects with <4, 4-7, 8-14, 15-19 and ≥20 oocytes retrieved
- Number and percentage of metaphase II oocytes (only applicable for those inseminated using ICSI), number of fertilized oocytes, fertilization rate as well as number and quality of blastocysts on day 5 after oocyte retrieval
- Endometrial thickness and echogenicity pattern on stimulation day 5 and end-of-stimulation
- Oocyte utilization rate (number of blastocysts transferred or cryopreserved divided by the number of oocytes retrieved) and oocyte efficiency index (cumulative number of ongoing pregnancies per oocyte retrieved)
- Number and percentage of blastocysts surviving cryopreservation and number and percentage of blastocysts with re-expansion after cryopreservation
- Number of cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation, and number and percentage of cryopreserved cycles with blastocyst transfer

- Circulating concentrations of AMH, FSH, LH, estradiol, progesterone, inhibin A and inhibin B on stimulation day 5, end-of-stimulation and oocyte retrieval, and FSH population pharmacokinetic parameters
- Total gonadotropin dose, number of stimulation days and number of 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 neutralizing 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 OHSS, overall and by grade, and proportion of subjects with moderate/severe OHSS
- Proportion of subjects hospitalized due to OHSS and proportion of subjects undergoing paracentesis due to OHSS
- Rate of multi-fetal gestation, biochemical pregnancy, spontaneous abortion, ectopic pregnancy (with and without medical/surgical intervention) and vanishing twins in the fresh cycle and in the cryopreserved cycles
- Technical malfunctions of the administration pen

**Post-trial Endpoints**

- Cumulative live birth rate after the fresh cycle and cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation
- Live birth rate in the fresh cycle and in the cryopreserved cycles
- Live birth rate of singletons born at term (≥37 weeks of gestation) in the fresh cycle, the cryopreserved cycles and cumulatively
- Time from start of controlled ovarian stimulation to live birth of a singleton born at term across the fresh and cryopreserved cycles, including duration and number of cycles before achieving a live birth of a singleton born at term
- Rate of minor/major congenital anomalies at birth, 4 weeks and 1 year after birth in the fresh cycle and cryopreserved cycles

# 3 Trial Design and Sample Size Considerations

## 3.1 Trial Design

### 3.1.1 Overall Design and Control Methods

This will be a randomized, double-blind, placebo-controlled, parallel groups, multicenter trial assessing the efficacy and safety of the rFSH preparation FE 999049 in subjects aged 35-42 years undergoing controlled ovarian stimulation for IVF / ICSI following a GnRH antagonist protocol. The primary endpoint is the cumulative ongoing pregnancy rate after the fresh cycle and cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation. Thereby, the trial is designed to capture the clinical efficacy of a single controlled ovarian stimulation cycle in a more complete manner by following outcomes from both the fresh and subsequent cryopreserved cycles. Secondary endpoints include pharmacodynamic parameters of FSH action as well as efficacy and safety parameters related to controlled ovarian stimulation from the fresh cycle and subsequent cryopreserved cycles.

*Controlled Ovarian Stimulation and Fresh Cycle*

Subjects will be screened within 90 days prior to randomization for compliance with the inclusion and exclusion criteria. On day 2-3 of the menstrual cycle, subjects will be randomized in a 10:1 ratio to FE 999049 or placebo, and controlled ovarian stimulation will be initiated. FE 999049 and placebo will be self-administered subcutaneously using a pre-filled injection pen.

Subjects assigned to treatment with FE 999049 will receive a starting dose of 15 µg daily that is fixed for the first four stimulation days. Based on ovarian response, the dose may be adjusted by 3 µg, with dose increases implemented not more frequently than once every 2 days and dose decreases implemented per investigator's judgement. The minimum daily dose is 6 µg, and the maximum daily dose is 24 µg. Subjects assigned to placebo will have the injection pen dialed to the same value (dose) as if administered FE 999049. Subjects can be treated with FE 999049 or placebo for a maximum of 20 days. Coasting, use of dopamine agonist or any other drug to prevent early OHSS with the exception of GnRH agonist for triggering of final follicular maturation, are not allowed.

During stimulation, subjects will be monitored by transvaginal ultrasound on stimulation days 1 and 5 and thereafter at least every second day. When the leading follicle reaches a diameter of ≥14 mm, transvaginal ultrasound will be performed daily. To prevent a premature luteinizing hormone (LH) surge, 250 µg GnRH antagonist (ganirelix acetate, GANIRELIX, Merck Sharp & Dohme) will be initiated on stimulation day 5 for subjects with ≥3 follicles with a diameter of ≥10 mm. Subjects who fail to satisfy this GnRH antagonist criterion on stimulation day 5 will continue to be monitored at least every second day, and GnRH antagonist will be initiated when/if the criterion is met. The GnRH antagonist will be continued throughout the stimulation period. Triggering of final follicular maturation

will be done as soon as ≥2 follicles with a diameter of ≥17 mm are observed. If there are <20 follicles with a diameter of ≥12 mm, 10,000 IU hCG (NOVAREL, Ferring Pharmaceuticals) will be administered. If there are ≥20 follicles with a diameter of ≥12 mm or the serum estradiol concentration is ≥3,000 pg/mL (local laboratory), 4.0 mg GnRH agonist (leuprolide acetate, LEUPROLIDE ACETATE, Sandoz) will be administered, and the fresh blastocyst transfer will be cancelled. If after 8 days of stimulation, the investigator judges that the triggering criterion is not likely to be reached by day 20, the cycle will be cancelled. If the triggering criterion is not met after 20 days of stimulation, the cycle will be cancelled.

Oocyte retrieval will take place 36h (±2h) after triggering of final follicular maturation, and oocytes will be inseminated by IVF or ICSI 4h (±1h) after retrieval. Rescue ICSI is not allowed. Fertilization and embryo development will be assessed. For subjects who undergo triggering of final follicular maturation with hCG and have <20 oocytes retrieved, transfer will be performed on day 5 (blastocyst stage) after oocyte retrieval. Subjects will have one blastocyst transferred if at least one good-quality (i.e. grade 3BB or above) blastocyst is available, or one or two blastocysts transferred if no good-quality blastocyst is available. Remaining blastocysts will be cryopreserved by vitrification. For subjects with ≥20 oocytes retrieved following hCG administration and for subjects who undergo triggering of final follicular maturation with GnRH agonist, no transfer will take place in the fresh cycle and blastocysts will instead be cryopreserved.

A subject who fails to reach the triggering criterion due to poor ovarian response or who has ≤3 oocytes retrieved will be offered medication and financial support for an ART cycle with an approved gonadotropin preparation outside of the trial.

Vaginal progesterone inserts (progesterone, ENDOMETRIN, Ferring Pharmaceuticals) 100 mg three times daily (TID) will be provided for luteal phase support from the day after oocyte retrieval and continuing until menses, negative β human chorionic gonadotropin (βhCG test), pregnancy loss or until ongoing pregnancy has been documented.

A serum βhCG test will be performed 10-14 days after transfer, clinical and vital pregnancy will be confirmed by transvaginal ultrasound 5-6 weeks after transfer, and ongoing pregnancy will be confirmed by transvaginal or abdominal ultrasound 8-9 weeks after transfer.

Blood samples will be collected for the purpose of evaluating the endocrine profile, clinical chemistry and haematology parameters as well as anti-FSH antibodies. Endocrine parameters will be assessed at screening, stimulation day 1, stimulation day 5, end-of-stimulation and oocyte retrieval. Clinical chemistry and haematology will be assessed at screening, end-of-stimulation, and end-of-cycle. Anti-FSH antibodies will be assessed on four occasions. The first sample will be taken at the screening visit and will be used exclusively to re-establish the anti-drug antibody analytical assays. The subsequent three samples will be used for analysis of anti-FSH antibodies in individual subjects in the trial, and

taken prior to dosing on stimulation day 1 and on two occasions post-dosing: 7-10 days after the last FE 999049 or placebo dose (this may coincide with the transfer visit) and 21-28 days after the last FE 999049 or placebo dose (this may coincide with the βhCG test visit). Subjects with a treatment-induced anti-FSH antibody response will be followed until the response has become negative, returned to the pre-dosing level, or for a maximum of 1 year after the second post-dose sampling. These subjects will be called in for assessments 2 months after the last post-dosing anti-FSH antibody sampling. If required, further assessments will be made at 3, 4, 6, 9 and 12 months after the last post-dosing anti-FSH antibody sampling. The follow-up will also be terminated if the subject commences a new treatment cycle with any gonadotropin preparation.

Local tolerability of FE 999049 and placebo following subcutaneous administration will be assessed by the subject three times daily: immediately, 30 minutes and 24 hours after each injection. The presence and intensity of injection site reactions will be rated as none, mild, moderate or severe. The assessments will be made throughout the stimulation period and recorded by the subject in a diary.

*Cryopreserved Cycles*

The trial covers cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation. Either a programmed or natural cycle can be selected for any cryopreserved cycle.

Any programmed cryopreserved cycle will be initiated within 3 days of start of menses with administration of estradiol (ESTRADIOL Tablets USP, Teva Pharmaceuticals USA, Inc.) 2 mg TID or 3 mg two times daily (BID) (or 3 mg TID at the investigator's discretion, if a daily dose of 6 mg has been shown to be insufficient in a previous cycle). If after 10-12 days of estradiol treatment the endometrial thickness is ≥8 mm, the subject will initiate daily intramuscular (IM) injections of 50 mg progesterone (PROGESTERONE Injection USP, West-ward Pharmaceutical Corp or Watson Pharma, Inc.) within the next 5 days in conjunction with the estradiol treatment. The ultrasound evaluation can be repeated within 7 days if the endometrial thickness criterion is not met. In programmed cryopreserved cycles, transfer of one or two blastocysts will occur on the $6^{th}$ day from start of progesterone after warming and assessment of blastocyst survival and re-expansion. Subjects will have one blastocyst transferred if at least one good-quality (i.e. grade 3BB or above) blastocyst is available, or one or two blastocysts transferred if no good-quality blastocyst is available. Luteal phase support (estradiol and IM progesterone) will continue to be administered until menses, negative βhCG test, pregnancy loss or until ongoing pregnancy has been documented.

Any natural cryopreserved cycle will be initiated 7 days after start of menses with monitoring of urinary LH on a daily basis by the subject. The day after confirmation of LH surge by serum LH (local laboratory) and endometrial thickness of ≥8 mm, the subject will start luteal phase support with vaginal progesterone inserts (progesterone, ENDOMETRIN, Ferring Pharmaceuticals) 100 mg TID. In a

natural cryopreserved cycle, transfer of one or two blastocysts will occur on day LH surge +7 after warming and assessment of blastocyst survival and re-expansion. Subjects will have one blastocyst transferred if at least one good-quality (i.e. grade 3BB or above) blastocyst is available, or one or two blastocysts transferred if no good-quality blastocyst is available. Luteal phase support (vaginal progesterone) will continue to be administered until menses, negative βhCG test, pregnancy loss or until ongoing pregnancy has been documented.

Failure to achieve endometrial thickness ≥8 mm in the first cryopreserved cycle will result in cycle cancellation, and in the programmed cycles, administration of 100 mg IM progesterone (PROGESTERONE Injection USP, West-ward Pharmaceutical Corp or Watson Pharma, Inc.) to induce withdrawal bleeding. In subsequent cryopreserved cycles, blastocyst transfer can take place regardless of endometrial thickness at the investigator's discretion.

In both programmed and natural cryopreserved cycles, a serum βhCG test is performed 10-14 days after transfer, clinical and vital pregnancy will be confirmed by transvaginal ultrasound 5-6 weeks after transfer, and ongoing pregnancy will be confirmed by transvaginal or abdominal ultrasound 8-9 weeks after transfer.

After completion of the trial, the subject is allowed to use cryopreserved blastocysts in accordance with local guidelines and/or regulations.

*Post-trial Activities*

Post-trial activities cover pregnancy and neonatal health follow-up after the fresh cycle and cryopreserved cycles.

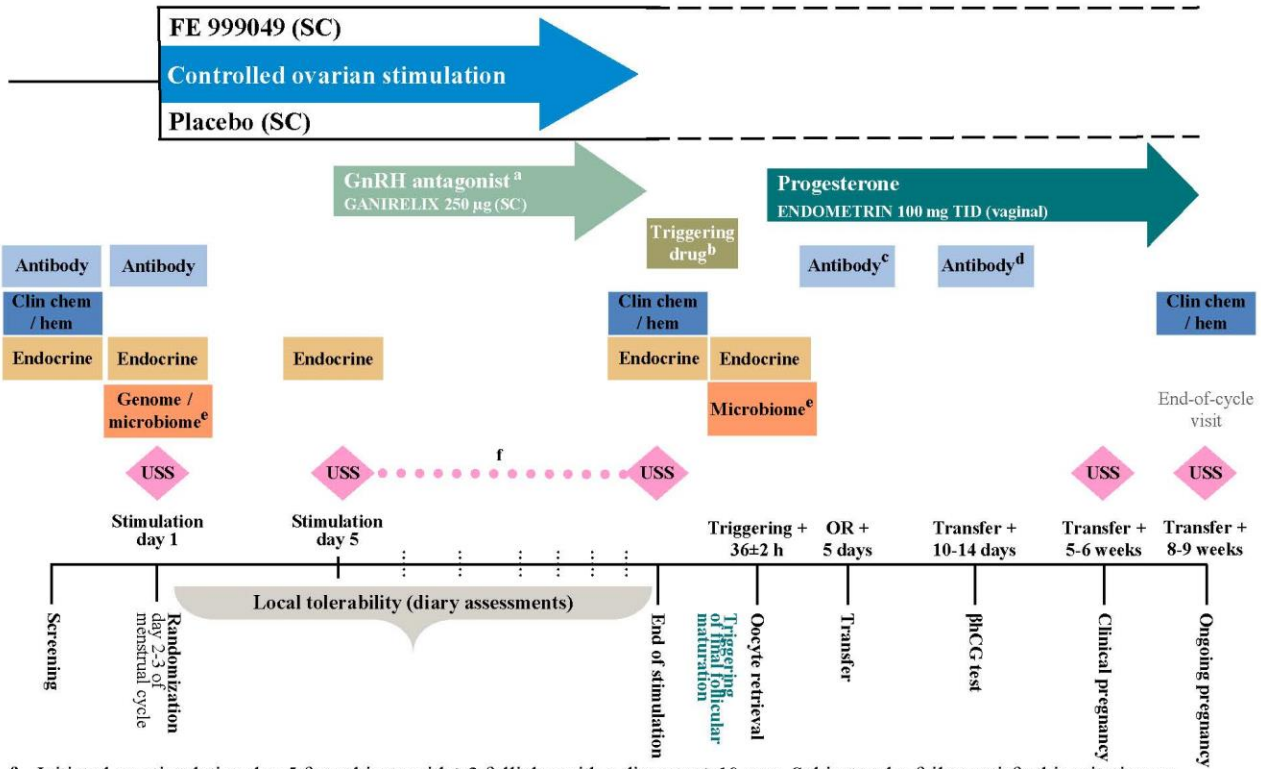
All subjects with an ongoing pregnancy obtained in the fresh cycle or in cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation will be followed until delivery to collect information on live birth rate. Furthermore, data will be collected on neonatal health, including minor/major congenital anomalies, at birth, 4 weeks and 1 year after birth.

*Optional Exploratory Analyses*

For subjects who have provided a separate informed consent, a blood sample and a saliva sample for potential future genome sequencing will be collected on stimulation day 1, and a tongue coat sample for potential future microbial profiling will be collected on stimulation day 1 and at the transfer visit(s) in the fresh and cryopreserved cycles, as applicable.

### 3.1.2 Trial Diagram

A diagram illustrating the controlled ovarian stimulation and fresh cycle is presented in Figure 1.

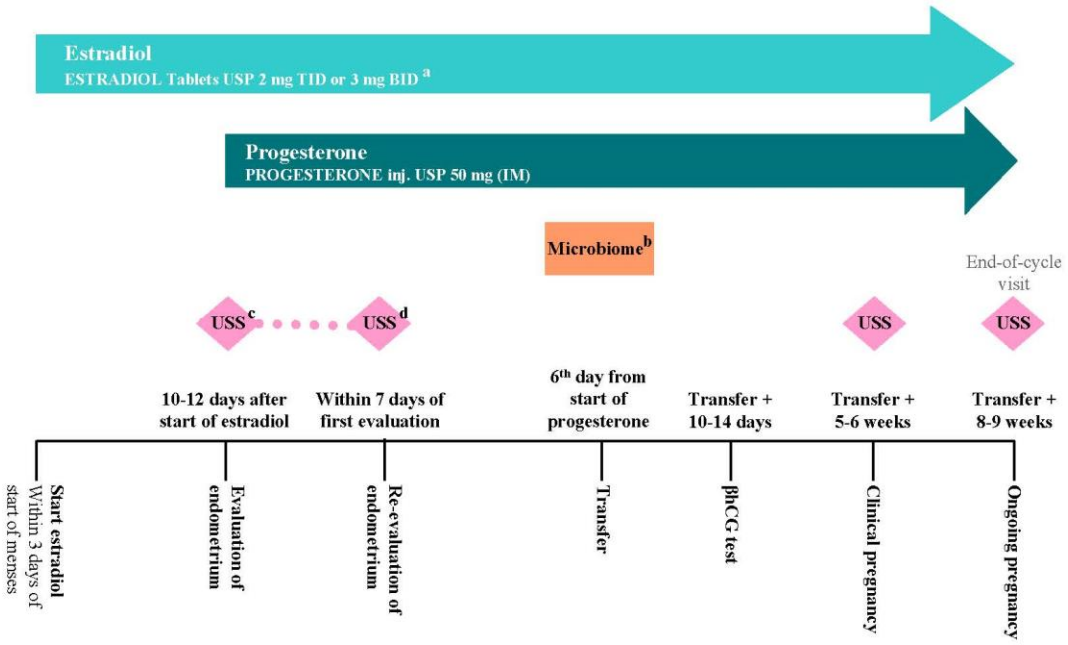



[a] Initiated on stimulation day 5 for subjects with ≥3 follicles with a diameter ≥10 mm. Subjects who fail to satisfy this criterion on stimulation day 5 will continue to be monitored at least every second day, and GnRH antagonist will be initiated when/if the criterion is met.
[b] hCG (NOVAREL 10,000 IU; SC) or GnRH agonist (LEUPROLIDE ACETATE 4 mg; SC), depending on number and size of follicles.
[c] 7-10 days after the last FE 999049 or placebo dose (may coincide with the transfer visit).
[d] 21-28 days after the last FE 999049 or placebo dose (may coincide with the βhCG test visit).
[e] Optional sampling for future potential exploratory analyses for subjects who have provided separate written informed consent.
[f] Stimulation days 1, 5 and thereafter at least every second day. When the leading follicle reaches a diameter of ≥14 mm, ultrasound will be performed daily.

SC: subcutaneous, OR: oocyte retrieval, TID: three times daily, USS: ultrasound sonography

**Figure 1 Trial Diagram – Controlled Ovarian Stimulation and Fresh Cycle**

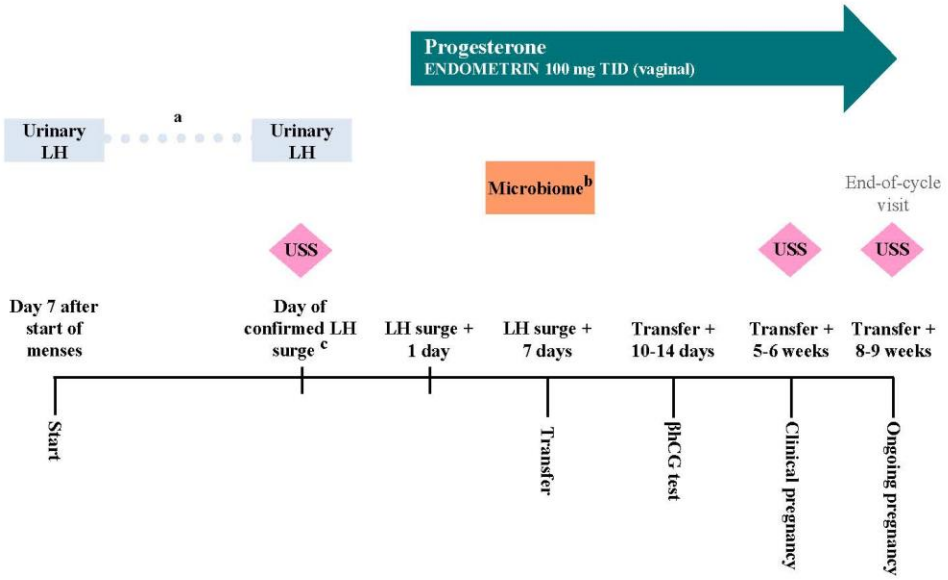
Diagrams illustrating the cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation are presented in Figure 2 for programmed cryopreserved cycles and in Figure 3 for natural cryopreserved cycles.




[a] 2 mg TID or 3 mg BID, or 3 mg TID at the investigator's discretion if a daily dose of 6 mg has been shown to be insufficient in a previous cycle.
[b] Optional sampling for future potential exploratory analyses for subjects who have provided separate written informed consent.
[c] If endometrial thickness is ≥8 mm, daily IM injections of 50 mg progesterone will be initiated within 5 days.
[d] Only for subjects with endometrial thickness <8 mm at the first evaluation. If endometrial thickness is ≥8 mm, subjects will initiate 50 mg IM progesterone within 5 days. Subjects who fail to achieve adequate endometrial thickness will discontinue estradiol and be administered 100 mg IM progesterone to induce withdrawal bleeding.

BID: two times daily, IM: intramuscular, TID: three times daily, USS: ultrasound sonography

**Figure 2 Trial Diagram – Programmed Cryopreserved Cycles Initiated Within 12 Months From the Start of Controlled Ovarian Stimulation**




[a] Monitoring of urinary LH will be done on a daily basis by the subject until confirmed LH surge.
[b] Optional sampling for future potential exploratory analyses for subjects who have provided separate written informed consent.
[c] Confirmation by serum LH.

LH: luteinizing hormone, TID: three times daily, USS: ultrasound sonography

**Figure 3 Trial Diagram – Natural Cryopreserved Cycles Initiated Within 12 Months From the Start of Controlled Ovarian Stimulation**

A diagram illustrating the post-trial activities for the fresh and cryopreserved cycles is shown in Figure 4.

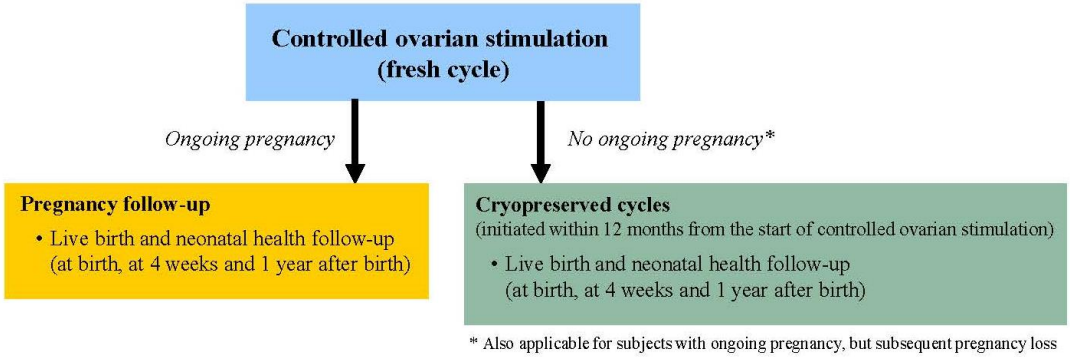



**Figure 4 Trial Diagram – Post-trial Activities**

## 3.2 Determination of Sample Size

The proposed sample size of 550 (FE 999049:placebo = 500:50) can adequately address both the efficacy and safety objectives of the trial:

- The cumulative ongoing pregnancy rate in the FE 999049-treated subjects aged 35-42 years is estimated to be approximately 30% according to results from clinical trials conducted in the U.S.[1] By comparison, the pregnancy rate in the placebo arm is expected not to exceed 3%, the monthly spontaneous pregnancy rate in infertile women.[2] Therefore, the proposed sample size of 550 (FE 999049:placebo=500:50) will provide at least 99% power for the primary efficacy comparison. Based on the estimated cumulative ongoing pregnancy rate, the cumulative live birth rate from the fresh and/or cryopreserved cycles is estimated to be approximately 27%, resulting in a power for the comparison of the cumulative live birth rate to be at least 99%.
- A key pharmacodynamic parameter in subjects aged 35-42 years is cycle cancellation due to poor follicular development, which may occur in approximately 5% of the population exposed to gonadotropins.
- The safety and tolerability of daily rFSH preparations as part of an ART treatment cycle have been well documented for the population studied in previous clinical trials.[3,4] OHSS was reported to occur in 1.7% to 3.7% of the subjects, with moderate/severe OHSS occurring in 1.4% to 2.2% of the subjects. Other adverse events were pelvic pain, pelvic discomfort and headache, with those assessed to be related to trial treatments reported in 1.5% to 7.1% of the subjects.

  Table 1 shows that the planned sample size of 500 subjects exposed to FE 999049 provides a high probability to detect a rare adverse event or safety signal occurring in 0.5% or more subjects.

**Table 1 Probabilities to Detect at Least One Rare Event**

| Incidence rate of rare event | Chance of detecting at least 1 event |
|---|---|
| 0.5% | 91.8% |
| 0.6% | 95.1% |
| 1% | 99.3% |

  Table 2 presents the margin width estimates of the 95% confidence intervals for a range of adverse event rates with the planned sample size.

**Table 2 Estimated Margin Widths of 95% Confidence Intervals**

| Adverse event rate | Margin of error width estimate |
|---|---|
| 5% | 1.9% |
| 10% | 2.6% |
| 15% | 3.1% |
| 20% | 3.5% |

# 4 Subject Disposition

All screened subjects will be accounted for. Screened subjects who discontinue from the trial prior to randomization are regarded as screening failures. Screening failures and their primary reason for screening failure will be tabulated. Screening failures will not otherwise be accounted for.

Subject disposition with respect to analysis sets will be tabulated. This table will include the number of subjects in the intention-to-treat (ITT) analysis set, the modified intention-to-treat (mITT) analysis set, the per-protocol (PP) analysis set, the safety analysis set, as well as the number of subjects who provided a separate informed consent for optional exploratory analyses. The number of subjects in the ITT analysis set will be used as denominator when calculating percentages.

A separate table will summarize the subject disposition with respect to analysis sets by trial site. This table will include the number of subjects in the ITT analysis set, the mITT analysis set, the PP analysis set, and the safety analysis set. The number of subjects in the ITT analysis set will be used as denominator when calculating percentages.

Subject disposition with respect to the timing of action taken during the COVID-19 pandemic will be tabulated using the ITT analysis set. Subject disposition will be summarized by whether a subject had exited the trial before or after the COVID-19 pandemic trial hold went into place. Of the subjects that had not yet exited the trial, the subject status (i.e.; in a cryopreserved cycle or not in a cryopreserved cycle) will be summarized. For subjects within a cryopreserved cycle, the cryopreserved cycle number will be summarized. For subjects not within a cycle, the most recent cycle completed and the number that had an ongoing pregnancy will be summarized. Additionally, for subjects that had not yet exited the trial, the time since randomization will be summarized by the following categories: <3 months, 3-<6 months, 6-<9 months, ≥9 months.

Subject attendance at selected trial visits and adherence to selected trial procedures will be tabulated. A detailed listing of selected missing trial assessments will be created. Included will be an indicator of whether the corresponding visit took place, an indicator of whether the visit was expected to occur before, during, or after the COVID-19 pandemic trial hold period, and an indicator of whether the missingness was anticipated, unanticipated, or due to the COVID-19 pandemic.

Subject disposition with respect to the completion of the fresh cycle will be summarized for all randomized subjects. The summary will include the number (percentage) of subjects who completed the fresh cycle and the number (percentage) of subjects who did not complete the fresh cycle as well as the primary reason for not completing the cycle. A subject is considered to have completed the fresh cycle if all anticipated visits outlined in the protocol were attended. This includes but is not limited to the following scenarios:

- A subject with cycle cancellation due to poor ovarian response, i.e. not meeting the criterion for triggering of final follicular maturation, is anticipated to attend the end-of-stimulation visit (not anticipated to undergo triggering of final follicular maturation).
- A subject with at least one blastocyst on day 5 is anticipated to attend the transfer visit.
- A subject with transfer cancellation due to no blastocysts on day 5 is anticipated to attend the 2nd post-dosing anti-FSH antibody assessment visit (not anticipated to undergo transfer).
- A subject with blastocyst transfer is anticipated to attend the beta-hCG test visit.
- A subject with a positive beta-hCG test is anticipated to attend the clinical pregnancy visit.
- A subject with a negative beta-hCG test is anticipated to attend the end-of-cycle visit (not anticipated to attend the clinical pregnancy visit).

In addition, the following information will be summarized with respect to the controlled ovarian stimulation (COS) cycle with the investigational medicinal product (IMP, FE 999049 or placebo) and the fresh transfer:

- The number of subjects who completed the COS cycle and the number of subjects who discontinued the IMP early including the reason for discontinuation will be tabulated. Subjects who completed the COS cycle are those who underwent triggering of final follicular maturation. The percentages are based on the mITT analysis set.
- For subjects who completed the COS cycle, the status with respect to blastocyst transfer in the fresh cycle will be tabulated. This table will summarize subjects who underwent transfer and those with transfer cancellation including the reason for cancellation.

For subjects who started at least one cryopreserved cycle, as indicated by initiating estradiol for programmed cycles or by initiating LH surge monitoring for natural cycles, the subject disposition with respect to blastocyst thawing, blastocyst transfer and cycle completion will be summarized by type of cryopreserved cycle (programmed and natural) and overall in sequential order. For each cryopreserved cycle, the denominator is the total number of subjects who started the respective cycle. The transfer completion/cancellation and the cycle completion/discontinuation statuses will be summarized including the reasons for cycle cancellation and for not completing the cycle respectively.

Finally, the following summary statistics will be presented: the number and percentages of subjects who underwent stimulation with IMP, who underwent transfer in the fresh cycle, who underwent transfer in at least one cryopreserved cycle (total and separate for programmed and natural cryopreserved cycles), as well as the total number of cryopreserved cycles and the total numbers of programmed and natural cryopreserved cycles respectively. The denominator will be based on the ITT population.

Subject disposition with respect to analysis sets will be listed for all randomized subjects including information on cycle completion status by cycle, whether the cycle was completed before, coincided with, or was initiated after the COVID-19 pandemic trial hold period, and reason for cycle termination, sorted by treatment group, subject ID and cycle in overall sequential order. Subjects who did not complete the COS with IMP and who had a fresh/cryopreserved transfer cancellation will also have the early termination and transfer cancellation information listed separately.

# 5 Protocol Deviations

The rating of protocol deviations as 'minor' and 'major', as well as the criteria for major protocol deviations with the implication of exclusions from the per-protocol (PP) analysis set will be decided by the medical officer, medical monitor and statistician on the basis of a blinded review of data before declaration of clean file and lock of database.

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 summarized and listed by subject.

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.

Some of the criteria for major exclusionary protocol deviation are listed in Table 3.

## Table 3 Major Protocol Deviations

**Randomization and exposure**

IMP exposure not in accordance with randomization

**Deviation from eligibility criteria**

Inclusion criterion 6 (documented infertility)

Inclusion criterion 7 (regular menstrual cycles of 24-35 days)

Inclusion criterion 8 (uterus consistent with normal function)

Inclusion criterion 9 (ovaries without significant evidence of abnormality)

Inclusion criterion 10 (FSH levels within 1-15 IU/L within 3 months prior to randomization)

Exclusion criterion 5 (recurrent miscarriage)

Exclusion criterion 10 (known endocrine or metabolic abnormalities with the exception of pharmacologically controlled sub-clinical hypothyroidism)

Exclusion criterion 11 (known tumors which contraindicate use of gonadotropins)

Exclusion criterion 17 (congenital uterine abnormalities)

Exclusion criterion 18 (pregnancy)

Exclusion criterion 20 (use of fertility modifiers)

Exclusion criterion 21 (use of hormonal preparations, except for thyroid medication)

**IMP regimen**

IMP: wrong daily dose for more than 1 day

**NIMP regimen**

GnRH antagonist (fresh): criterion not adhered to

hCG and GnRH agonist (fresh): use of incorrect triggering drug

hCG and GnRH agonist (fresh): incorrect dose of triggering drug

Progesterone (fresh and cryo): missed daily dose for more than 1 day before beta-hCG visit or menses

**Transfer policy**

Transfer of a non-blastocyst

Double blastocyst transfer in subject fulfilling criterion for single blastocyst transfer

Single blastocyst transfer in the fresh cycle in subject fulfilling criterion for double blastocyst transfer

Blastocyst(s) available after hCG triggering but no transfer performed in the fresh cycle for reasons other than described in protocol

Blastocyst(s) available after thawing, survival and re-expansion in cryopreserved cycle but no transfer performed for reasons other than described in protocol

Transfer in the first cryopreserved cycle in subject not fulfilling endometrial thickness criterion

**Discontinuation not due to efficacy, safety, or COVID-19 pandemic reasons**

Withdrawal of consent

Discontinuation due to other protocol deviation than not attending scheduled visit

Discontinuation due to other reasons (including personal reasons)

**Unblinding**

Unblinding of investigator

# 6 Analysis Sets

## 6.1 Intention-To-Treat Analysis Set

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

## 6.2 Modified Intention-to-Treat Analysis Set

The modified intention-to-treat (mITT) analysis set comprises all randomized and exposed subjects. Subjects will be analyzed according to planned treatment.

## 6.3 Per-Protocol Analysis Set

The PP analysis set comprises all mITT 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 comprises all randomized and exposed treated subjects. Subjects will be analyzed according to actual treatment received.

# 7 Trial Population

## 7.1 Display of Trial Population Characteristics

All relevant baseline data will be tabulated 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 randomization. Baseline data will not be compared using statistical tests. Unless otherwise noted, tabulations will be produced for the mITT and PP analysis sets, as well as for the ITT analysis set if the ITT analysis set differs from the mITT 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.

In case of multiple assessments prior to the first exposure to IMP, the baseline value is defined as the last available assessed value prior to the first exposure to IMP.

All baseline data will be listed, sorted by treatment group and subject number. Listings will be produced for the mITT analysis set only.

Unless otherwise noted, missing data will not be imputed.

## 7.2 Demographics

The following demographic information will be tabulated: Age at randomization (years), ethnicity (Hispanic or Latino, Not Hispanic or Latino), and race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White).

## 7.3 Medical History

Medical history will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) using the version effective at trial start. The version of MedDRA will be documented. Medical history will be tabulated by System Organ Class sorted alphabetically and Preferred Term sorted in decreasing order of frequency. Medical history will only be tabulated for the mITT analysis set.

Medical history will be listed by treatment group and subject number for the mITT analysis set.

## 7.4 Infertility History

The following data on infertility history obtained at the screening visit will be tabulated: The primary reason for infertility, (all applicable) reasons for infertility and duration of infertility (months).

The following data on previous fertility treatment cycles obtained at the screening visit will be tabulated: Number of subjects with at least one previous fertility treatment cycle, number of previous

fertility treatment cycles per subject, number of subjects with at least one clinical pregnancy, number of clinical pregnancies per subject, number of subjects with at least one live birth, and number of live births per subject. In addition, the type of treatment (OI±IUI or ART) and medications used on the cycle level, the number of fetuses on the clinical pregnancy level and the outcome on the fetus level (live birth, still birth, ectopic pregnancy, miscarriages by trimester, and terminations) will be tabulated. Tables will be produced overall and by type of treatment.

Information on previous fertility treatment cycles will be listed by treatment group, subject number and previous fertility cycle number for the mITT analysis set. The listing will include information on start date of stimulation, type of treatment, medication used, βhCG assessment, clinical pregnancy assessment and outcome for each fetus.

### 7.5 Menstrual History

The average duration of the menstrual cycle (days) will be tabulated.

### 7.6 Reproductive History

The following data on reproductive history following natural conception obtained at the screening visit will be tabulated: Number of subjects with at least one previous clinical pregnancy, number of clinical pregnancies per subject, number of subjects with at least one live birth, and number of live births per subject. In addition the number of fetuses on a clinical pregnancy level and the outcome on a fetus level (live birth, still birth, ectopic pregnancy, miscarriages by trimester, and terminations) will be tabulated.

Information on reproductive history will be listed by treatment group, subject number and previous clinical pregnancy from natural conception number for the mITT analysis set.

### 7.7 Body Measurements

The following baseline body measurements will be summarised: Body weight (kg), height (m), and body mass index (BMI, $kg/m^2$). In addition, frequency tables for BMI (<18.5, 18.5-<25.0, 25.0-<30.0, ≥30.0) and body weight (≤75, >75) will be produced.

### 7.8 Physical Examination

Physical examination performed at screening covers the categories general appearance, central and peripheral nervous system, head and neck, respiratory system, cardiovascular system, gastrointestinal system, lymphatic system, urinary system, musculoskeletal system and skin. Each category will be evaluated as normal, abnormal not clinically significant, abnormal clinically significant or not done. Physical examination at screening will be summarised by category. Physical examination at screening will be listed as part of the safety listings (Section 10.12).

### 7.9 Gynecological Examination

Gynecological examination performed at screening covers the categories breast, external genitalia, vagina, cervix, uterus, ovaries and fallopian tubes. Each category will be evaluated as normal, abnormal not clinically significant or abnormal clinically significant. Gynecological examination at screening will be summarised by category. Gynecological examination will be listed as part of the safety listings (Section 10.13).

### 7.10 Endocrine Parameters

At the screening visit, blood samples will be collected for analysis of the following endocrine parameters: AMH (pmol/L), TSH (mIU/L) and prolactin (pmol/L). On stimulation day 1, blood samples will be collected for analysis of the following endocrine parameters: AMH (pmol/L), FSH (IU/L), LH (IU/L), estradiol (pmol/L), progesterone (nmol/L), inhibin A (ng/L) and inhibin B (ng/L). These endocrine parameters at baseline will be tabulated (AMH on stimulation day 1 is considered the baseline value).

In addition, a frequency table will be produced for AMH on stimulation day 1 (<7.5, 7.5-<15, 15-<25, ≥25 pmol/L).

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.

### 7.11 Vital Signs

Blood pressure (systolic and diastolic) (mmHg) and pulse (beats/min) measured on stimulation day 1 will be summarised. Vitals signs on stimulation day 1 will be listed as part of the safety listings (Section 10.14.4).

### 7.12 Ovarian Volume

The total and average ovarian volume ($cm^3$) observed on stimulation day 1 will be tabulated. The ovaries will be modelled as ellipsoids, implying that the average ovarian volume ($cm^3$) will be calculated as:

$$\frac{\pi}{6} \cdot \left( \frac{l_r \cdot w_r \cdot d_r}{2} + \frac{l_l \cdot w_l \cdot d_l}{2} \right) \cdot \frac{1}{1000},$$

where $l_x$, $w_x$, and $d_x$ are the length, width and depth in mm of the right ($x=r$) and left ($x=l$) ovary respectively. The total ovarian volume is then derived as twice the average volume. The total and average ovarian volume will only be reported if observations are available for both ovaries.

### 7.13 Follicles

The number of follicles on stimulation day 1 will be summarized and tabulated using the categories 0-4, 5-9 and ≥10 follicles.

### 7.14 Endometrial Thickness

The endometrial thickness (mm) on stimulation day 1 will be summarized and tabulated using the categories ≤6 mm, 7-9 mm, ≥10 mm.

### 7.15 Prior and Concomitant Medication

Prior and concomitant medication will be will be coded using the World Health Organization (WHO) Drug Reference List. They will be summarized by anatomical therapeutic chemical (ATC) classification 1st level (alphabetically), ATC classification 2nd level (in decreasing order of frequency) and treatment group. These medications will be tabulated separately for:

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

Concomitant medication will be presented separately for the fresh cycle and the cryopreserved cycles.

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.

Prior and concomitant medication will be listed by treatment group and subject number for the mITT analysis set.

# 8 Exposure and Treatment Compliance

Tabulations will be produced for the mITT and PP analysis set.

## 8.1 Extent of Exposure

For any investigational medicinal product (IMP) or non-IMP (NIMP), duration of treatment (days) is defined as the number of days from first exposure to the day of last exposure (both inclusive).

### 8.1.1 IMP (Gonadotropins)

Exposure to gonadotropins will be summarized as the duration of stimulation, the total dose - both the actual total dose administered and the intended total dose (i.e. the sum of the dialed doses recorded by the subject), and the average daily dose defined as the total dose divided by the number of stimulation days. Frequency tables will be included to describe the duration of treatment by categories (<7, 7-8, 9-10, 11-12 and ≥13 days). The number of stimulation days will also be tabulated by treatment group.

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 5 will be tabulated. Furthermore, the total number of dose increase requests and dose decrease requests per subject will be tabulated.

In addition, summaries of the total dose administered from day 5 to last stimulation day, the maximum and minimum daily doses, the final planned/administered dose, will be presented.

These summary tables will be produced for all subjects who started stimulation, and for subjects who completed the stimulation cycle, i.e. underwent triggering of final follicular maturation. The summaries for subjects who completed the stimulation cycle will also be produced by the triggering drug (i.e., hCG or GnRH agonist).

### 8.1.2 NIMPs in the Fresh Cycle

#### GnRH Antagonist

Exposure to GnRH antagonist will be summarised as the total dose administered (mg) and duration of treatment (days). In addition, the summary of the starting day of the GnRH antagonist will be presented. This table will be produced overall and for subjects that completed stimulation, i.e. underwent triggering of final follicular maturation.

#### Drug for Triggering of Final Follicular Maturation

For subjects who underwent triggering of final follicular maturation, the following information will be summarized: triggering criterion met (yes/no) and the drug used for triggering (hCG or GnRH agonist).

**Luteal Phase Support**

For subjects who initiated luteal phase support with progesterone, the following information will be summarized: total dose administered (mg) and duration of treatment (days).

### 8.1.3 NIMPs in the Cryopreserved Cycles

In the programmed cryopreserved cycles, exposure to estradiol will be summarized by cycle as the total dose administered (mg) and duration of treatment (days), and exposure to IM progesterone will be summarized by cycle as the total dose administered (mg) and duration of treatment (days).

In the natural cryopreserved cycles, exposure to vaginal progesterone will be summarized by cycle as the total dose administered (mg) and duration of treatment (days).

## 8.2 Treatment Compliance

Non-compliance to IMP is expected to be limited and will be presented in listings. Similarly, listings will be prepared for subjects with deviations in the NIMP schedules for the fresh and cryopreserved cycles, such as

- GnRH antagonist in the fresh cycle not administered according to protocol
- hCG or GnRH agonist in the fresh cycle not administered according to protocol
- Progesterone in the fresh cycle or cryopreserved cycles not administered according to protocol
- Estradiol in the cryopreserved cycles (programmed) not administered according to protocol

# 9 Efficacy

## 9.1 General Considerations

### 9.1.1 Primary Endpoint and Secondary Efficacy Endpoints

The primary efficacy endpoint, cumulative ongoing pregnancy rate, will be compared formally between the two treatment groups. The cumulative live birth rate derived from post-trial follow-up assessments will be tested formally as a key secondary efficacy endpoint (see Section 11.1.2).

A hierarchical testing procedure is used to control the overall type 1 error rate. Upon achieving statistical significance about the primary endpoint at the 0.05 alpha level, the cumulative live birth rate will be formally tested at the 0.05 alpha level for inferential conclusions (see Section 11.1.2).

The secondary endpoints related to pregnancy and implantation parameters are considered supportive of the primary endpoint (cumulative ongoing pregnancy rate). The secondary efficacy endpoints related to ovarian response and embryo development parameters are intended to provide additional characterization of FE 999049.

## 9.2 Analysis and Presentation of Primary and Secondary Efficacy Endpoints

Summary tables and treatment comparisons for the primary endpoint will be presented for both the mITT and PP analysis sets. In case that the mITT and ITT analysis sets are not identical (i.e. at least one subject is randomized but not exposed) the summary tables and treatment comparisons will also be presented for the ITT analysis set. As applicable for the secondary endpoints, summary tables and treatment comparisons may be presented for the mITT and PP analysis sets.

All tabulations will be presented by treatment group. 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.

All 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’.

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

### 9.2.1 Multiplicity

In order to handle the formal testing procedures for the primary endpoint of cumulative ongoing pregnancy rate and the key post-trial endpoint of cumulative live birth rate, the overall type-I error will

be controlled by the use of a hierarchical inferential approach. Statistical significance of the primary endpoint at the 0.05 alpha level is required before drawing inferential conclusions about the key post-trial endpoint of the cumulative live birth rate. This fixed hierarchical approach will ensure a strong control of the overall type-I error rate at the 0.05 level.

All additional analyses of the primary endpoint and secondary efficacy endpoints are considered as supportive. Adjustments for multiplicity will therefore not be applied.

### 9.2.2 Missing Data

Missing observations for the primary endpoint of cumulative ongoing pregnancy rate will be imputed as 'negative' irrespective of the reason why data are not recorded. Similarly, for other pregnancy endpoints (βhCG, clinical pregnancy and vital pregnancy) missing data will be imputed as 'negative' with the exception of a later observation in that cycle confirming that an earlier missing observation was in fact 'positive', e.g. in case of a missing βhCG test result and a confirmed clinical pregnancy in the same cycle then the βhCG test result will be imputed as 'positive'.

However, if a subject eligible for additional transfer chooses not to come back within the one-year period from the start of stimulation due to miscellaneous personal reasons, this is not considered a censored or missing observation due to loss to follow-up for cumulative pregnancy endpoints. Rather, the observed one-year cumulative pregnancy outcomes reflect the subject's real-life decision or circumstance.

The COVID-19 pandemic induced a trial hold where new cryopreserved cycles were not allowed to be initiated. The primary endpoint analysis is based on cryopreserved cycles initiated within 12 calendar months from the date of randomization (subsequently referred to as 12 calendar months). However, for a subset of subjects, part of their 12 calendar months will have been lost due to the COVID-19 pandemic trial hold. Therefore, the pandemic could induce missing primary endpoint information. Based on the approach outlined above, these primary endpoint assessments will be imputed as 'negative'. In order to assess the impact, if any, of the COVID-19 pandemic on the primary endpoint, additional sensitivity analyses are planned. As part of one of these additional sensitivity analyses, the primary endpoint analysis will be repeated, but with the application of multiple imputation for those endpoints missing specifically due to the pandemic.

When applying the method of multiple imputation, three types of missingness are used: 1) missing the primary endpoint assessment from a recorded cryopreserved cycle because of the COVID-19 pandemic, 2) missing the primary endpoint assessment from a cryopreserved cycle that would have theoretically occurred had it not been for the COVID-19 pandemic, and 3) missing the primary endpoint assessment because of any other reason. The method of multiple imputation will only be applied to those endpoints that are missing because of the first two types of missingness. Missing endpoints because of the third type of missingness will continue to be imputed as 'negative'. If a

subject was still in the trial at the start of the COVID-19 pandemic trial hold and initiates an additional cryopreserved cycle after the end of the trial hold, but within 12 calendar months from their date of randomization, then this subject's endpoint will be considered non-missing and will not be imputed using multiple imputation.

Multiple imputation will be based on a risk-stratified hot deck imputation strategy. Subjects that are to receive imputation will be stratified by assigned treatment arm and the number of previous transfers within the study. The subjects not receiving imputation will form an imputation pool. The imputation pool will also be stratified by assigned treatment arm and the number of previous transfers within the study. If a stratum within the imputation pool contains fewer than 10 subjects, strata may be collapsed to increase the membership of a particular stratum. For each subject receiving imputation, the imputation pool will be further reduced to those subjects that have an endpoint assessment that occurred at or beyond the first date of missingness for the imputed subject's primary endpoint. Hot deck replacement commences from the reduced, stratified imputation pool 10 times, forming 10 analysis datasets from which combined estimates will be obtained using the method described by Rubin.[5]

For subjects with transfer but missing observations on the number of viable fetuses 8-9 weeks after transfer the number of viable fetuses will be imputed as zero irrespective of why data is 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 8-9 weeks after transfer.

For subjects with cycle cancellation prior to triggering of final follicular maturation, the numbers of oocytes retrieved, metaphase II oocytes, fertilized oocytes, and blastocysts on day 5 will be imputed as zero.

Missing values will not be imputed for any of the other secondary endpoints, unless otherwise noted.

### 9.3 Primary Endpoint

The primary endpoint will be determined as soon as the subject has achieved an ongoing pregnancy, or when all cryopreserved blastocysts have been exhausted, or after assessment of ongoing pregnancy status in cryopreserved cycles initiated within 12 months from the start of controlled ovarian stimulation.

The superiority hypotheses to be tested for the primary endpoint are:

$H_0$: $\pi_{\text{FE 999049}} - \pi_{\text{Placebo}} \leq 0$ against the alternative $H_A$: $\pi_{\text{FE 999049}} - \pi_{\text{Placebo}} > 0$,

where $\pi_{\text{FE 999049}}$ and $\pi_{\text{Placebo}}$ denote the cumulative ongoing pregnancy rate after the fresh cycle and cryopreserved cycles in subjects aged 35-42 years treated with FE 999049 or placebo, respectively.

### 9.3.1 Primary Analysis

$H_0$ will be tested against the alternative $H_A$ by constructing a two-sided 95% confidence interval for the difference in the cumulative ongoing pregnancy rates between the two treatment groups using the traditional Wald interval. If the lower limit of the two-sided 95% confidence interval is greater than 0, the null hypothesis $H_0$ will be rejected. In the case that the number of pregnancies observed in the placebo group is small (<5), then the one-sided Fisher's exact test will be used.

The efficacy analysis will be based on the mITT analysis set, defined as all subjects who were randomized and received at least one dose of IMP (FE 999049 or placebo).

### 9.3.2 Sensitivity Analyses

The primary analysis will be repeated for the subjects in the PP analysis set as sensitivity analysis, as well as in the ITT analysis set if the ITT analysis set differs from the mITT analysis set. The outcomes of these analyses are considered supportive.

The primary analyses described above will be repeated restricted to subjects with oocytes retrieved and to subjects undergoing blastocyst transfer, as applicable.

Missing data for the primary endpoint is expected to be limited due to the trial design, but may occur if the subject is lost to follow-up before ongoing pregnancy is assessed. If one or more such subjects have a positive pregnancy test result (βhCG test, clinical pregnancy or vital pregnancy) prior to loss to follow-up, the impact of imputing the ongoing pregnancy will be investigated in a sensitivity analysis.

It is also possible that a subject without pregnancy assessments following a transfer has her birth/miscarriage/elective termination outcome available at a later time. The information may be evaluated to help determine the ongoing pregnancy status in the sensitivity analysis on a case-by-case basis.

In order to assess the impact, if any, of the COVID-19 pandemic on the primary endpoint, additional sensitivity analyses are planned which will assess both the timing of the COVID-19 pandemic trial hold and the imputation method applied to missing data that arises due to the pandemic. The primary endpoint analysis is based on cryopreserved cycles initiated within 12 calendar months from the date of randomization (subsequently referred to as 12 calendar months). However, for a subset of subjects, part of their 12 calendar months will have been lost due to the COVID-19 pandemic trial hold. These subjects will be allowed to initiate cryopreserved cycles within 12 months of active study days (subsequently referred to as 12 study months). Because the last randomization occurred on 17 September 2019 and the trial hold began subsequent to the issuing of Protocol Amendment 01 on 19 March 2020, the first COVID-19 pandemic sensitivity analysis will repeat the same methodology as the primary endpoint analysis (i.e.; same analysis set, same inference method, and same imputation method), but will instead use the cumulative ongoing pregnancy rate from cycles initiated within the

first six calendar months following randomization. In this manner, all subjects can be evaluated using an equivalent time period. The second COVID-19 pandemic sensitivity analysis will again repeat the same methodology as the primary endpoint analysis, but with ongoing pregnancies observed within 12 study months. The third COVID-19 pandemic sensitivity analysis will, like the primary endpoint analysis, use data from all cycles initiated within 12 calendar months, but will apply a different imputation method for endpoint data that are missing due to the COVID-19 pandemic. The method of multiple imputation, as described in Section 9.2.2, will be applied to those endpoints that are missing specifically due to the pandemic.

**Subgroup Displays**

The primary endpoint will also be displayed by the following subgroups:

- Primary reason for infertility (unexplained infertility, tubal infertility, male factor, endometriosis stage I/II, other)
- Method of insemination (IVF, ICSI, IVF/ICSI)
- Age (35-37, 38-40, 41-42 years; ≤ median, > median)
- Body weight (≤75, >75 kg; ≤ median, > median)
- AMH on stimulation day 1 (<7.5, 7.5-<15, 15-<25, ≥25 pmol/L; ≤ median, > median)
- Center

For each subgroup factor, descriptive statistics for the cumulative ongoing pregnancy rate will be presented by subgroup level and treatment group. These subgroup displays will be prepared for the mITT and PP analysis sets.

## 9.4 Secondary Efficacy Endpoints

Secondary efficacy endpoints will be analyzed based on the mITT analysis set. As specified below, for selected endpoints, analyses will be repeated on the PP analysis set.

For secondary efficacy endpoints that are collected in the cryopreserved cycles, additional summaries will be tabulated by whether or not the cycle occurred before, concurrently during, or after the conclusion of the COVID-19 pandemic trial hold. It is anticipated that the number of cryopreserved cycles that are concurrent with or initiated after the COVID-19 pandemic trial hold will be small therefore only descriptive statistics, and not inferential statistics, will be applied to the three groupings.

### 9.4.1 Ongoing Pregnancy Rate

The ongoing pregnancy rate will be calculated and tabulated for the fresh cycle and for each cryopreserved cycle sequentially (overall and within programmed and natural cryopreserved cycles separately), and for cryopreserved cycles cumulatively based on both the total number of subjects in the treatment group and the total number of subjects who started the respective cycle.

Note that, as a result of late pregnancy loss (after confirmation of ongoing pregnancy), a subject is allowed to come back for another transfer within one year from the start of stimulation, as long as there is at least one blastocyst remaining. The eCRF will be kept open for the subject to capture potentially further cryopreserved cycle information up to one year from the start of stimulation unless the subject withdraws consent or when all her blastocysts are exhausted. Therefore, a subject may achieve more than one ongoing pregnancy within two different cycles in the trial, but for cumulative ongoing pregnancy rates, a subject can only be counted once in the numerator.

Analyses similar to that used for the primary endpoint will be applied for the fresh cycle, and for cryopreserved cycles as applicable i.e. where there is at least one subject who started the respective cycle in both treatment groups and will be applied to both the mITT and PP analysis sets.

For subjects with an ongoing pregnancy, the number of intrauterine viable fetuses will be tabulated overall and by ongoing pregnancy in sequential order.

### 9.4.2 Time to Ongoing Pregnancy

For subjects who achieve ongoing pregnancy following a fresh or cryopreserved transfer, time to the first ongoing pregnancy across the fresh and cryopreserved cycles will be calculated as the number of days from the start of controlled ovarian stimulation to the first visit confirming ongoing pregnancy, as well as the number of cycles until ongoing pregnancy is achieved. The number of days will be presented in two ways: total number of days and number of days within cycles (i.e. excluding the periods between the end-of-cycle visit and the next cycle initiation). Descriptive statistics will be presented by treatment group, and will be applied to both the mITT and PP analysis sets.

### 9.4.3 Ongoing Implantation Rate

The ongoing implantation rate will be calculated for the fresh cycle, for each cryopreserved cycle sequentially (overall and within programmed and natural cycles separately), cumulatively for the cryopreserved cycles, and cumulatively across the fresh and cryopreserved cycles for subjects with at least one transferred blastocyst in the applicable cycles.

For subjects where the number of viable fetuses 8-9 weeks after transfer is greater than the number of blastocysts transferred, the number of viable fetuses will be set to the number of blastocyst transferred in the analysis. Descriptive statistics will be presented by treatment group and the two treatment groups will be compared using the nonparametric Wilcoxon rank-sum test as applicable. The analyses will be applied to both the mITT and PP analysis sets.

### 9.4.4 Clinical Pregnancy Rate

The clinical pregnancy rate will be calculated for the fresh cycle, for each cryopreserved cycle sequentially (overall and within programmed and natural cycles separately), cumulatively for the cryopreserved cycles, and cumulatively across the fresh and cryopreserved cycles based on both the total number of subjects in the treatment group and the total number of subjects who started the respective cycle. The analyses will be conducted in a similar manner as ongoing pregnancy rate and will be applied to both the mITT and PP analysis sets. For subjects with clinical pregnancy, the type of clinical pregnancy (intrauterine or ectopic) will be tabulated.

### 9.4.5 Vital Pregnancy Rate

The vital pregnancy rate will be calculated and analyzed in a similar manner as ongoing pregnancy rate and will be applied to both the mITT and PP analysis sets.

### 9.4.6 Implantation Rate

The implantation rate will be calculated and analyzed in a similar manner as ongoing implantation rate and will be applied to both the mITT and PP analysis sets.

### 9.4.7 Positive βhCG Rate

The positive βhCG rate will be calculated and analyzed in a similar manner as ongoing pregnancy rate and will be applied to both the mITT and PP analysis sets.

### 9.4.8 Triggering of Final Follicular Maturation, Cycle Cancellation and Transfer Cancellation

The proportion of subjects with triggering of final follicular maturation will be tabulated by triggering drug (hCG, GnRH agonist) and in total. The proportion of subjects with cycle cancellation or transfer cancellation in the cycle will also be tabulated, including the reasons as described in Section 4 for the summary of subject disposition with respect to cycle and transfer cancellations. In addition, the proportion of subjects with transfer cancellation in cryopreserved cycles will be tabulated.

### 9.4.9 Number and Size of Follicles during Stimulation

The follicle cohort on stimulation day 5 and end-of-stimulation will be summarized 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 Wilcoxon's test. Categorical data will be compared between treatment groups using the chi-square test or Fisher's exact test in case of sparse data. The analyses will be applied to both the mITT and PP analysis sets.

### 9.4.10 Number and Distribution of Oocytes Retrieved

The number of oocytes retrieved will be tabulated including summary statistics. Furthermore, a frequency table with subjects grouped according to number of oocytes retrieved will be prepared using these categories: <4, 4-7, 8-14, 15-19 and ≥20 oocytes. Subjects with cycle cancellation due to poor ovarian response will be included in the <4 oocytes group. Continuous data will be compared between treatment groups using Wilcoxon's test. Categorical data will be compared between treatment groups using the chi-square test or Fisher's exact test in case of sparse data. The analyses will be applied to both the mITT and PP analysis sets.

### 9.4.11 Number of Metaphase II Oocytes

Oocytes undergoing ICSI will have their maturity stage assessed prior to insemination. Maturity stage will be categorized as germinal vesicle, metaphase I, metaphase II, degenerated or other. The assessment of maturity stage will be tabulated on the oocyte level, i.e. the denominator will be the number of oocytes with assessment of maturity stage.

Analyses of MII oocytes will be restricted to subjects where all inseminated oocytes are inseminated using ICSI.

The number of MII oocytes per subject will be tabulated including both summary statistics and a frequency table using the categories ≤6, 6-12, ≥13 MII oocytes. Furthermore, the percentage of MII oocytes to oocytes retrieved for subjects where all oocytes are inseminated using ICSI will be tabulated. Continuous data will be compared between treatment groups using Wilcoxon's test.

Categorical data will be compared between treatment groups using the chi-square test or Fisher's exact test in case of sparse data. The analyses will be applied to both the mITT and PP analysis sets.

### 9.4.12 Number of Fertilized Oocytes and Fertilization Rate

The number of pronuclei will be counted at 19h (±2h) after insemination and recorded as 0, 1, 2 or >2 (or damaged). An oocyte is defined as fertilized if it is scored as 2PN at 19h (±2h). The assessment of pronuclei will be tabulated on the oocyte level, i.e. the denominator will be the number of oocytes retrieved.

The number of fertilized oocytes per subject will be tabulated including both summary statistics and a frequency table using the categories <3, 3-5, 6-10, 11-15 and ≥16 fertilised oocytes. Furthermore, for subjects with oocytes retrieved, the rate of fertilized oocytes to oocytes retrieved (and also the rate of fertilized oocytes to metaphase II oocytes for those inseminated using ICSI) will be tabulated overall and by method of insemination. Continuous data will be compared between treatment groups using Wilcoxon's test as applicable. Categorical data will be compared between treatment groups using the chi-square test or Fisher's exact test in case of sparse data. The analyses will be applied to both the mITT and PP analysis sets.

### 9.4.13 Number and Quality of Blastocysts on Day 5

The evaluation on day 5 after oocyte retrieval will consist of assessment of embryo stage and classification of blastocysts according to blastocyst expansion and hatching status (1-6), blastocyst inner cell mass grading and trophectoderm grading (for those embryos with blastocyst expansion and hatching status 3-6). Furthermore, the destiny of the embryo will be recorded.

Embryo stage is classified as blastocyst, morula, degenerated or cleavage stage. For embryos still at the cleavage stage the number of blastomeres will be recorded.

Destiny at day 5 is either transferred, cryopreserved, out of trial or continue to day 6 after oocyte retrieval.

In the event of continued culture, blastocyst grading will be recorded on day 6 after oocyte retrieval.

Based on the blastocyst expansion and hatching status, blastocyst inner cell mass grading and trophectoderm grading, the blastocyst will be classified as a good-quality blastocyst if the grade is 3BB or above as illustrated in Figure 5.

| | | | |
|---|---|---|---|
| 3AA | 4AA | 5AA | 6AA |
| 3AB | 4AB | 5AB | 6AB |
| 3AC | 4AC | 5AC | 6AC |
| 3BA | 4BA | 5BA | 6BA |
| 3BB | 4BB | 5BB | 6BB |

| | | | |
|---|---|---|---|
| | 4BC | 5BC | 6BC |
| | 4CA | 5CA | 6CA |
| | 4CB | 5CB | 6CB |
| | 4CC | 5CC | 6CC |

**Figure 5 Good-quality Blastocysts**

The embryos on day 5 and day 6 after oocyte retrieval will be summarised on both the embryo level and on the subject level.

**Embryo Level**

At the embryo level all embryos evaluated will be included in the tables when reporting embryo stage and destiny of the embryo. Tables will be prepared for overall and by method of insemination (IVF or ICSI). Frequency tables will be produced for the embryo stage, destiny at day 5, blastocyst expansion and hatching status, blastocyst inner cell mass grading and trophectoderm grading.

**Subject Level**

At the subject level the following variables on day 5, day 6 and cumulatively will be derived for all subjects:

- Number of blastocysts after oocyte retrieval
- Number of good-quality blastocysts
- Rate of blastocysts to oocytes retrieved, defined as the ratio of blastocysts to number of oocytes retrieved
- Rate of good-quality blastocysts to oocytes retrieved, defined as the ratio of good-quality blastocysts to number of oocytes retrieved

The cumulative day 5 and day 6 assessment takes the last available quality assessment value on the two days for the blastocyst. These derived variables will be summarised and compared between treatment groups using the Wilcoxon's test. Tables will be produced for all subjects, for subjects with at least 1 oocyte retrieved, and by insemination method (IVF or ICSI).

Further, at the subject level the following will be derived for all subjects:

- At least one blastocyst available on day 5, day 6 and cumulatively
- At least one good-quality blastocyst available on day 5, day 6 and cumulatively
- Number of blastocysts transferred, total and by fresh and cryopreserved transfers,
- Number of blastocysts cryopreserved

These derived variables will be reported in frequency tables. The availability of blastocysts will be compared between treatment groups using the chi-square test or Fisher's exact test in case of sparse

data. The total numbers if blastocysts transferred and cryopreserved will be compared the Wilcoxon test for ordinal. Tables will be produced for all subjects, for subjects with at least 1 oocyte retrieved, and by insemination method (IVF or ICSI).

For subjects where all inseminated oocytes are inseminated using ICSI the following will be derived for each subject:

- Number of blastocysts on day 5, day 6 and cumulatively after oocyte retrieval relative to the number of MII oocytes
- Number of good-quality blastocysts on day 5, day 6 and cumulatively relative to the number of MII oocytes

These derived variables will be summarised and compared between treatment groups. The overall comparisons will be based on the Wilcoxon's test.

The analyses will be applied to both the mITT and PP analysis sets.

### 9.4.14 Endometrial Thickness and Echogenicity Pattern

Endometrial thickness on stimulation day 5 and at end-of-stimulation will be tabulated including summary statistics. Endometrial echogenicity pattern on stimulation day 5 and at end-of-stimulation will be tabulated. Continuous data will be compared between treatment groups using Wilcoxon's test. Categorical data will be compared between treatment groups using the chi-square test or Fisher's exact test in case of sparse data. The analyses will be applied to both the mITT and PP analysis sets.

### 9.4.15 Oocyte Utilization Rate and Oocyte Efficiency Index

The oocyte utilization rate (number of blastocysts transferred or cryopreserved divided by the number of oocytes retrieved) will be analyzed in a similar manner as ongoing implantation rate and will be applied to both the mITT and PP analysis sets.

The oocyte efficiency index (cumulative number of ongoing pregnancies per oocyte retrieved) will be analyzed in a similar manner as ongoing implantation rate and will be applied to both the mITT and PP analysis sets.

### 9.4.16 Blastocyst Survival and Re-expansion after Cryopreservation

The number and percentage of blastocysts surviving cryopreservation will be tabulated for cryopreserved cycles cumulatively, overall and by programmed/natural cycles separately. Further,

among the blastocysts that survived cryopreservation, the number and percentage of blastocysts with re-expansion will be tabulated.

### 9.4.17 Number of Cryopreserved Cycles

The number of cryopreserved cycles initiated (i.e. programmed cycles with subjects starting estradiol treatment and natural cycles with subjects initiating LH surge monitoring), as well as the number and percentage of cryopreserved cycles with blastocyst transfer will be tabulated. Further, the total number of blastocysts transferred in fresh and cryopreserved cycles will be tabulated.

### 9.4.18 Circulating Levels of Endocrine Parameters

Blood samples drawn on stimulation day 1, stimulation day 5, end-of-stimulation, and oocyte retrieval will be analyzed for AMH, 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.

Each endocrine parameter and the change from baseline for post-baseline measurements will be tabulated for stimulation day 1 (baseline), stimulation day 5, end-of-stimulation and oocyte retrieval. For each parameter the change from baseline will be compared between treatment groups using a log-normal model. In this model change from baseline in log-transformed measurements will be the dependent variable and the linear predictor will include treatment as factor and baseline measurement (log-transformed) as covariate. The estimated treatment difference with 95% confidence interval will be presented on the original scale of measurement (i.e. exp-transformation applied to the log-transformed measurement) and accompanied by the p-value for test of no treatment difference.

Further, a population pharmacokinetic model will be used for evaluating the effects of subject characteristics on FSH concentrations and for assessing the variability in exposure. The modelling analysis will be described in a separate modelling analysis plan and the results reported separately from the clinical trial report.

### 9.4.19 Total Gonadotropin Dose, Number of Stimulation Days, and Number of Dose Adjustments

The analysis strategies are described in Section 8.1.1.

# 10 Safety

## 10.1 General Considerations

Analyses for the safety analysis set will be conducted according to the actual treatment received.

Data will be presented by summary tables and listings. Categorical data will be summarized by treatment using the number and percentage of subjects in each category. For calculation of percentages, the denominator will be the total number of subjects in the respective treatment group in the safety analysis set. Continuous data will be summarized by treatment using number, mean, standard deviation, median, minimum, and maximum.

All individual subject data will be listed per subject and treatment as observed including any derived values.

### Missing Data

Missing values will be treated as missing, except for causality, intensity, seriousness, and outcome of adverse events. A worst-case approach will be used: if causality is missing, the adverse event will be regarded as related to the IMP; if the intensity of an adverse event is missing, the adverse event will be regarded as severe; if seriousness is missing, the adverse event will be regarded as serious; if outcome is missing, and no date of outcome is present, the outcome is regarded as 'ongoing'.

## 10.2 Adverse Events

Adverse events will be coded using MedDRA using the version effective at trial start. Adverse events will be 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 in the fresh cycle, i.e. any adverse event occurring after start of IMP and before the end-of-cycle 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-cycle visit.
- Adverse events in cryopreserved cycles, i.e. any adverse event occurring after initiation of a cryopreserved cycle and before the end-of-cycle visit, or a pre-existing medical condition that worsens in intensity after initiation of a cryopreserved cycle and before the end-of-cycle visit.

Treatment-emergent adverse events and adverse events in cryopreserved cycles will be presented in summary tables and listings. Separate summary tables will be produced for treatment-emergent adverse events in the fresh cycle and adverse events in cryopreserved cycles cumulatively (overall and by programmed and natural cycles separately). Summaries will also be provided for the trial across fresh and cryopreserved cycles.

Pre-treatment adverse events will be presented in a listing only.

Treatment-emergent adverse events will be tabulated by system organ class alphabetically and 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 treatment-emergent adverse events, treatment-emergent adverse events by causality (reasonable possibility / no reasonable possibility), treatment-emergent adverse events leading to death, adverse events by intensity (mild / moderate / severe), treatment-emergent adverse reactions by intensity (mild / moderate / severe), serious treatment-emergent adverse events, serious treatment-emergent adverse reactions, treatment-emergent adverse events leading to discontinuation, treatment-emergent adverse events with an incidence of ≥5% in any treatment group, treatment-emergent adverse reactions with an incidence of ≥5% in any treatment group, and non-serious treatment-emergent adverse events with an incidence of ≥5% in any treatment group.

Similar displays will be prepared for the adverse events in cryopreserved cycles. Listings of adverse events in the cryopreserved cycles will indicate whether or not the adverse event occurred before, during, or after the COVID-19 pandemic trial hold.

### 10.2.1 Overview of Treatment-Emergent Adverse Events and Adverse Events in Cryopreserved Cycles

Adverse events overview summary tables for treatment-emergent adverse events in the fresh cycle and adverse events in cryopreserved cycles will be prepared including the number of subjects reporting an adverse events, 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, defined as adverse events judged by the investigator to be related to IMP with a reasonable possibility
- Adverse events leading to discontinuation
- Serious adverse events
- Deaths

In addition to overview summary tables for the fresh and cryopreserved cycles, an overview summary table for the trial will also be produced.

## 10.3 Clinical Chemistry and Hematology Parameters

Blood samples for clinical chemistry and haematology parameters will be drawn at screening, end-of-stimulation and end-of-cycle in the fresh cycle.

### 10.3.1 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.

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

### 10.3.2 Changes Relative to Normal Range

Shift tables will be prepared to compare baseline values to the end-of-stimulation and end-of-cycle values, using a categorization 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.

### 10.3.3 Markedly Abnormal Changes

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 clinical chemistry and haematology variables are specified in Appendix 1.

### 10.3.4 Data Listings

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

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

## 10.4 Injection Site Reactions

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 type and intensity (none, mild, moderate and severe).

## 10.5 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 neutralizing capacity will be tabulated. Furthermore, all assessments of anti-FSH antibodies will be listed for subjects with a positive result in assay 2.

## 10.6 Immune-related Adverse Events

All treatment-emergent adverse events will be analyzed to identify those that potentially are immune-related. To identify all possible cases, a broad-scope search on Standardized MedDRA Queries (SMQs), including 'Hypersensitivity', 'Anaphylactic reactions', 'Angioedema' and 'Severe cutaneous adverse reactions' will be considered. Moreover, to identify the potential cases manifested by non-specific symptoms and not covered by these SMQs, other MedDRA PTs like 'Musculoskeletal pain', 'Asthenia', 'Pyrexia', 'Chills', 'Body temperature increased', 'Influenza like illness', 'Injection related reaction', 'Presyncope' and 'Syncope' will also be taken into account. Hypersensitivity reactions manifested by local symptoms will be identified using the MedDRA high level term (HLT) 'injection site reactions'. The SMQs may include very specific as well as less specific terms; hence a narrow-scope search on these SMQs will be carried out to identify those cases that are highly likely to represent an immune-related etiology.

Potential immune-related adverse events will be tabulated using the SMQs and PTs, for the fresh cycle, cumulatively for the cryopreserved cycles (overall and by programmed and natural cycles separately), and for the trial across the fresh and cryopreserved cycles.

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

A summary table will be prepared showing the proportion of subjects with cycle cancellations including reason for cancellation for the fresh cycle. If the cycle is cancelled due to an adverse event, it will be further categorized as immune-related or not.

Any cancellations in the cryopreserved cycles due to the COVID-19 pandemic will also be summarized.

## 10.8 Ovarian Hyperstimulation Syndrome (OHSS),

OHSS will for each treatment group be tabulated by classification (mild, moderate, severe) and grade (1, 2, 3, 4, 5). The tabulation will also include the total incidence of OHSS and the incidence of moderate/severe OHSS. OHSS will be presented overall and by timing (early, late, and early and late combined).

For the fresh cycle, 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. Late OHSS is defined as OHSS with onset >9 days after triggering of final follicular maturation.

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

Separate tabulations will be made for the fresh cycle, cumulatively for the cryopreserved cycles (overall and by programmed and natural cycles separately), and for the trial across the fresh and cryopreserved cycles.

### 10.9 Hospitalization and Paracentesis due to Ovarian Hyperstimulation Syndrome

The number and percentage of subjects hospitalized due to OHSS as well as the number and percentage of subjects who underwent paracentesis due to OHSS will be tabulated. Separate tabulations will be made for the fresh cycle, cumulatively for the cryopreserved cycles (overall and by programmed and natural cycles separately), and for the trial across the fresh and cryopreserved cycles.

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

Summary tables will be prepared for these endpoints. Separate tabulations will be made for the fresh cycle, cumulatively for the cryopreserved cycles (overall and by programmed and natural cycles separately), and for the trial across the fresh and cryopreserved cycles.

### 10.11 Pen Malfunction

The frequency of reported malfunctions of the administration pen will be presented in a summary table. For confirmed pen malfunctions, the categories include: not able to detach the pen cap, not able to click / screw-on the needle due to damaged thread on cartridge holder, not able to dial the dose, not able to inject the dose, and other.

### 10.12 Physical Examination

Physical examination at end-of-cycle in the fresh cycle compared to screening will be summarized in shift tables and all subjects with any abnormal finding will be listed per subject. For each cryopreserved cycle, physical examination at end-of-cycle in the cryopreserved cycle compared to cycle initiation in the respective cycle will be summarized in shift tables and all subjects with any abnormal finding will be listed per subject.

### 10.13 Gynecological Examination

Gynecological examination at end-of-cycle in the fresh cycle compared to screening will be summarized in shift tables and all subjects with any abnormal finding will be listed per subject. For each cryopreserved cycle, gynecological examination at end-of-cycle in the cryopreserved cycle compared to cycle initiation in the respective cycle will be summarized in shift tables and all subjects with any abnormal finding will be listed per subject.

### 10.14 Vital Signs

Vital signs (systolic and diastolic blood pressure and pulse) will be measured at screening, on stimulation day 1 prior to the first exposure to IMP and at end-of-cycle in the fresh cycle as well as at cycle initiation and at end-of-cycle in any cryopreserved cycle.

In the fresh cycle, the baseline for the vital sign analyses will be the values obtained at the last assessment prior to the first exposure to IMP.

### 10.14.1 Summary Statistics

Mean change and mean percentage (%) change from baseline at each time point will be presented for each vital sign variable by cycle in a sequential order (overall and within programmed and natural cycles separately). In addition, descriptive statistics, i.e., the number of subjects with data, mean (standard deviation), median, minimum, and maximum values, will be presented for observed values and change from baseline at each time point by cycle sequentially (overall and within programmed and natural cycles separately) for each vital sign variable.

### 10.14.2 Changes Relative to Normal Range

Shift tables will be prepared to compare baseline values to the end-of-cycle values, using a categorization of low, normal and high values at each visit by cycle sequentially (overall and within programmed and natural cycles separately). Low, normal and high are defined in Appendix 1.

### 10.14.3 Markedly Abnormal Changes

Summary tables will be prepared displaying the proportion of subjects who have at least one markedly abnormal value in the fresh cycle, cumulatively for the cryopreserved cycles (overall and by programmed and natural cycles separately), and for the trial across the fresh and cryopreserved cycles.. The table will also include a break-down by classification of the baseline value. Markedly abnormal criteria for vital signs are specified in Appendix 1.

### 10.14.4 Data Listings

Data listings will be prepared by treatment group, subject number for all subjects with any abnormal vital sign value at any time point.

# 11 Post-trial Assessments

## 11.1 Post-trial Efficacy Assessments

### 11.1.1 Cumulative Live Birth Rate

The live birth rate will be calculated cumulatively across the fresh and cryopreserved cycles. Subjects with no information on live birth will be defaulted to a negative response.

The cumulative live birth rate is regarded as a key secondary endpoint and will be formally tested for treatment efficacy between the FE999049 treatment group and the placebo group using a hierarchical procedure. Upon achieving statistical significance about the primary endpoint at the 0.05 alpha level, the cumulative live birth rate will be tested at the 0.05 alpha level for inferential conclusions, using a similar method to the primary efficacy analysis.

Similarly to the primary efficacy endpoint, sensitivity analysis will be conducted for the subjects in the PP analysis set as sensitivity analysis, as well as in the ITT analysis set if the ITT analysis set differs from the mITT analysis set. Furthermore, the additional sensitivity analyses due to the COVID-19 pandemic will be repeated. Descriptive statistics will be presented by treatment group.

### 11.1.2 Live Birth Rate

The live birth rate will be calculated for the fresh cycle, for each cryopreserved cycle sequentially (overall and within programmed and natural cycles separately), cumulatively for the cryopreserved cycles, and cumulatively across the fresh and cryopreserved cycles based on both the total number of subjects in the treatment group and the total number of subjects who started the respective cycle. Subjects with no information on live birth will be defaulted to a negative response. Descriptive statistics will be presented by treatment group.

### 11.1.3 Live Birth Rate of Singletons Born at Term

The live birth rate of singletons born at term will be calculated for the fresh cycle, for each cryopreserved cycle sequentially (overall and within programmed and natural cycles separately), cumulatively for the cryopreserved cycles, and cumulatively across the fresh and cryopreserved cycles based on both the total number of subjects in the treatment group and the total number of subjects who started the respective cycle. Subjects with no information on live birth will be defaulted to a negative response. Descriptive statistics will be presented by treatment group.

### 11.1.4 Time to Live Birth of a Singleton Born at Term

For subjects who achieve live birth of a singleton born at term, time to live birth of a singleton born at term across the fresh and cryopreserved cycles will be calculated as the number of days from the start of controlled ovarian stimulation to delivery, as well as the number of cycles until delivery. The number of days will be presented in two ways: total number of days and number of days within cycles

(i.e. excluding the periods between the end-of-cycle visit and the next cycle initiation). sDescriptive statistics will be presented by treatment group.

The time to live birth of a singleton born at term will also be summarized by whether each subject had any cycles leading to birth that occurred during or after the COVID-19 pandemic trial hold. Descriptive statistics will be presented by treatment group.

## 11.2 Post-trial Safety Evaluations

### 11.2.1 Minor/Major Congenital Anomalies

The rate of minor/major congenital anomalies at birth, 4 weeks and 1 year after birth in the fresh cycle and cryopreserved cycles will be calculated. Descriptive statistics will be presented by treatment group.

## 11.3 Other Post-trial Evaluations

Gender, birth weight and length, and Apgar score will be tabulated. Descriptive statistics will be presented by treatment group.

# 12 Optional Exploratory Analyses

Genome sequencing and microbial profiling data obtained from subjects who provided a separate informed consent may be used in exploratory analyses aimed at identifying novel relationships between ovarian response/associated outcomes with subjects' genomic variants and microbiome. The analysis plans will be described in a separate document with the trial results reported separately.

## 13 Interim Analysis

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

## 14 Deviations from Protocol

There are no deviations from the planned analyses described in the clinical trial protocol.

The following additions and deviations from the planned displays described in the clinical trial protocol have been implemented in this statistical analysis plan:

- Trial population characteristics will also be tabulated for the PP analysis set (see Section 7.1).
- Subgroup displays of the primary endpoint will also include breakdown by median for the factors age, body weight, and AMH on stimulation day 1 (see Section 9.3.2).
- Time to ongoing pregnancy will also be calculated using the number of days spent in fresh and cryopreserved cycles (see Section 9.4.2).
- Summary tables will also be produced for serious treatment-emergent adverse reactions and treatment-emergent adverse reactions with an incidence of ≥5% in any treatment group (see Section 10.2).
- Potential immune-related adverse events will be tabulated using the SMQs and PTs, instead of using the SMQs, HLTs and PTs (see Section 10.6).
- Time to live birth of a singleton born at term will also be calculated using the number of days spent in fresh and cryopreserved cycles (see Section 11.1.4).

# 15 Tables, Listings and Figures

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

## 16 References

[1] Boostanfar R, Gates D, Guan Y, Gordon K, McCrary Sisk C, Stegmann BJ. Efficacy and safety of frozen-thawed embryo transfer in women aged 35 to 42 years from the PURSUE randomized clinical trial. Fertil Steril 2016; 106: 300-305.

[2] Jones HW, Toner JP. The infertile couple. N Eng J Med 1993; 329: 1710–1715.

[3] Nyboe Andersen A, Nelson SM, Fauser BCJM, Garcia-Velasco JA, Klein BM, Arce J-C. Individualized versus conventional ovarian stimulation for in vitro fertilization: a multicenter, randomized, controlled, assessor-blinded, phase 3 noninferiority trial. Fertil Steril 2017; 107: 387-396.

[4] Boostanfar R, Shapiro B, Levy M, Rosenwaks Z, Witjes H, Stegmann BJ, Elbers J, Gordon K, Mannaerts B; Pursue investigators. Large, comparative, randomized double-blind trial confirming noninferiority of pregnancy rates for corifollitropin alfa compared with recombinant follicle-stimulating hormone in a gonadotropin-releasing hormone antagonist controlled ovarian stimulation protocol in older patients undergoing in vitro fertilization. Fertil Steril 2015; 104: 94-103.

[5] Rubin DB. Multiple imputation for nonresesponse in surveys. New York, John Wiley & Sons, Inc., 1987.

# Appendix 1 Markedly Abnormal Laboratory Safety Values and Vital Signs

## Table 4 Markedly Abnormal Criteria for Laboratory Tests – Clinical Chemistry

| Variable | Unit | Markedly abnormal criteria | |
|---|---|---|---|
| | | **Low** | **High** |
| Alanine transaminase (ALT) | IU/L | Not applicable | > 3xULN |
| Albumin | g/L | < 20 | Not applicable |
| Alkaline phosphatase | IU/L | Not applicable | > 3xULN |
| Aspartate aminotransferase (AST) | IU/L | Not applicable | > 3xULN |
| Bicarbonate | mmol/L | < 15.1 | > 34.9 |
| Bilirubin direct | µmol/L | Not applicable | > 2xULN |
| Bilirubin total | µmol/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 total | mmol/L | Not applicable | > 10.34 |
| Creatinine | µmol/L | Not applicable | > 3xULN |
| Gamma-glutamyl transpeptidase (GGT) | IU/L | Not applicable | > 3xULN |
| Glucose | mmol/L | < 2.5 | > 16.7 |
| Lactate dehydrogenase (LDH) | IU/L | Not applicable | > 3xULN |
| Phosphorus | 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 | µmol/L | Not applicable | > 595 |

## Table 5 Markedly Abnormal Criteria for Laboratory Tests – Hematology

| Variable | Unit | Markedly abnormal criteria | |
|---|---|---|---|
| | | **Low** | **High** |
| Red blood cells | $10^{12}$/L | < 3.0 | Not applicable |
| White blood cells | $10^{9}$/L | < 2.5 | > 15.0 |
| Hemoglobin | g/L | < 100 | > 190 |
| Hematocrit | Ratio | < 0.28 | > 0.50 |
| Platelets | $10^{9}$/L | < 110 | > 600 |
| Red blood cell morphology | | | |
| MCV | fL | Not applicable | Not applicable |
| MCH | Pg | Not applicable | Not applicable |
| MCHC | g/L | Not applicable | Not applicable |
| White blood cell morphology | | | |
| Eosinophils | % | Not applicable | ≥ 10 |
| Neutrophils | % | ≤ 15 | ≥ 90 |
| Lymphocytes | % | ≤ 10 | ≥ 80 |
| Monocytes | % | Not applicable | ≥ 20 |
| Basophils | % | Not applicable | ≥ 5 |
| Bands | % | Not applicable | ≥ 20 |

**Table 6 Markedly Abnormal Criteria for Vital Signs***

| Variable | Criterion Value | Change from Baseline |
|---|---|---|
| Systolic blood pressure | ≥ 180 mmHg<br>≤ 90 mmHg | Increase of ≥ 20 mmHg<br>Decrease of ≥ 20 mmHg |
| Diastolic blood pressure | ≥ 105 mmHg<br>≤ 50 mmHg | Increase of ≥ 15 mmHg<br>Decrease of ≥ 15 mmHg |
| Pulse | ≥ 120 bpm<br>≤ 50 bpm | Increase of ≥ 15 bpm<br>Decrease of ≥ 15 bpm |

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