Cover page for Statistical Analysis Plan

Sponsor name	Ferring Pharmaceuticals Inc.
NCT Number:	NCT04163458
Sponsor trial ID:	000303
Official title of trial:	A randomized, double-blind double-dummy trial comparing MENOPUR solution for injection in a pre-filled pen and MENOPUR powder and solvent for solution for injection (menotropins for injection) in a GnRH agonist cycle in women aged 18-42 years undergoing an assisted reproductive technology program
Document Date:	17 Aug 2021

STATISTICAL ANALYSIS PLAN

A randomized, double-blind double-dummy trial comparing MENOPUR solution for injection in a pre-filled pen and MENOPUR powder and solvent for solution for injection (menotropins for injection) in a GnRH agonist cycle in women aged 18-42 years undergoing an assisted reproductive technology program

000303

Investigational Product:	MENOPUR solution for injection in pre-filled pen, 1200 IU/1.92mL MENOPUR powder and solvent for solution for injection, 75 IU
Indication:	Development of multiple follicles and pregnancy in ovulatory women undergoing controlled ovarian stimulation as part of an assisted reproductive technology (ART) cycle
Phase:	3
Author:	[redacted] Ph.D.
Date:	17 Aug 2021
Version:	2.0

Change log

Version No.	Effective Date	Reason for the Change / Revision	Supersedes
1.0	31 Aug 2020	Original version	None
2.0	17 Aug 2021	The following describes the changes that have been made to the SAP between versions 1.0 and 2.0. The following edits were made to correct mistakes in V1.0 of the SAP. • Transfer policy deviations will not be considered during the assessment of major or minor deviations. Because transfer occurs after the primary endpoint assessment, a transfer policy deviation cannot impact the primary endpoint. • There were cases where "Stimulation Day 5" was used. These cases have been updated to reflect that measurement occurs on "Stimulation Day 6". • In V1.0 of the SAP, for the multiple imputation sensitivity analysis, the imputation model (negative binomial) and the analysis model (ANOVA) did not align. In V2.0 of the SAP, the imputation model has been changed to align to the analysis model (both are now an ANOVA model). The following edits were made to clarify the original intent of SAP V1.0. • For a subset of inferences, the within-treatment group 95% confidence intervals (adjusted for stratification factors) will be displayed. The subset of inferences is now better defined in V2.0 of the SAP. • If the Safety analysis set is identical to the mITT analysis set, tables based on the mITT analysis set will be produced. (The identical tables based on the Safety analysis set will be suppressed.) • Trial population, efficacy, and safety summaries will be produced that are broken out by age strata.	1.0

		• Presentation of NIMPs taken prior to randomization will include a summary for all subjects exposed, whether or not they were randomized. • A listing of protocol deviations due to COVID-19 will be included. • In V1.0 of the SAP, the primary endpoint model was described as a 'linear model'. In V2.0 of the SAP, the primary endpoint model has been explicitly defined as an ANOVA model. • In V1.0 of the SAP, the analysis method for the secondary pregnancy endpoints and any other dichotomous endpoint was omitted. In V2.0 of the SAP, these secondary efficacy analyses have been defined to use the Mantel-Haenszel method. • In several sections, the presentation of subgroup summaries is requested. V2.0 of the SAP clarifies that these subgroups begin with the mITT analysis set prior to subsetting. Inference is performed on complete analysis sets and not on subgroups. • The categories for the frequency table summary of the number of 2PN and the number of MII have been aligned with the same for oocytes retrieved. • The analysis of inseminated oocytes has been removed as this was never listed as an endpoint. • Additional details have been added for the modeling of change from baseline in endocrine laboratory parameters. • Ranges for markedly abnormal vital signs and laboratory parameters have been added as an appendix. • The production of within-treatment group Clopper-Pearson confidence intervals for the proportion of anti-MENOPUR antibodies has been added in V2.0 of the SAP. • Early pregnancy losses are now described as a subgroup display within the Safety section. This had previously been in the Efficacy section. • Added summaries of AEs and ADRs that occur in at least 2% of subjects. • Additional details have been added regarding the summary and analysis of post-trial endpoints. Additionally, the following minor edits have been made: • The SAP reviewers have been updated. • Minor typos have been corrected throughout.	

Agreement on Statistical Analysis Plan

Author:

[redacted]
Biostatistics
Ferring Pharmaceuticals, Inc.

Reviewers:

[redacted]
Statistical Programming
Ferring Pharmaceuticals, Inc.

[redacted]
US Pharmacovigilance
Ferring Pharmaceuticals, Inc.

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

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

[redacted]
Reproductive Medicine and Maternal Health
Ferring Pharmaceuticals, Inc.

[redacted]
Clinical Operations
Ferring Pharmaceuticals, Inc.

[redacted]
Clinical Operations
Ferring Pharmaceuticals, Inc.

Approval:

[redacted]
Biometrics
Ferring Pharmaceuticals, Inc.

TABLE OF CONTENTS

1 Introduction

This document describes the planned statistical analyses for Trial 000303 based on the protocol Version 6.0 dated 14 Oct 2019.

1.1 Definitions/ Abbreviations

1.1.1 Definition of Terms

Terms	Definitions
Baseline	The last available assessed value prior to the first exposure to IMP.

1.1.2 Abbreviations

Abbreviations	Meaning of abbreviations in document
AE	Adverse Event
AMH	Anti-Müllerian Hormone
ART	Assisted Reproductive Technology
ATC	Anatomic Therapeutic Chemical
βhCG	Beta-Human Chorionic Gonadotropin
BMI	Body Mass Index
CRF	Case Report Form
E2	Estradiol
eCRF	Electronic Case Report Form
FDA	Food and Drug Administration
FSH	Follicle Stimulating Hormone
GnRH	Gonadotropin-releasing Hormone
hCG	Human Chorionic Gonadotropin
ICSI	Intracytoplasmic Sperm Injection
IMP	Investigational Medicinal Product
ITT	Intention-To-Treat
IU	International Unit
IUI	Intrauterine Insemination
LH	Luteinizing Hormone
LLOQ	Lower Limit of Quantification
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intention-To-Treat
NIMP	Non-Investigational Medicinal Product
OI	Ovulation Induction
P4	Progesterone
PP	Per-Protocol

PT	Preferred Term
SAE	Serious Adverse Event
SD	Standard Deviation
SOC	System Organ Class
ULOQ	Upper Limit of Quantification
WHO	World Health Organization

1.2 Presentation of results

1.2.1 Presentation of descriptive results

Categorical data will be summarized, unless otherwise stated, using the frequency (n) and relative frequencies as a percentage (i.e.; 100*n/N).

Continuous data will be presented for observed values, unless otherwise stated, using the number of subjects (N), number of observations (n), mean and standard deviation (SD), median and interquartile range (IQR), and range (minimum to maximum).

1.2.2 Presentation of inferential results

Inferential results will include confidence intervals for the treatment difference. For a subset of analyses, a within-treatment group confidence interval will be presented.

1.2.3 Subject data listings by domain

Individual subject data will be presented in listings in accordance to ICH E3. Included will be listings of:

- Discontinued subjects
- Protocol deviations
- Subjects excluded from the efficacy analysis
- Demographic data
- Non-investigational medicinal product (NIMP) and investigational medicinal product (IMP) use (with indication of compliance and concentration)
- Individual efficacy response data
- Adverse events by subject (with indication of deaths and serious adverse events [SAEs])
- Individual laboratory measurements by subject (with indication of abnormal values)

2 Trial Objectives and Endpoints

2.1 Objectives

Primary Objective

- To demonstrate non-inferiority of MENOPUR liquid versus MENOPUR powder with respect to the number of fertilized oocytes in women undergoing controlled ovarian stimulation.

Secondary Objectives

- To evaluate the pregnancy rates after stimulation with MENOPUR liquid and MENOPUR powder.
- To evaluate the follicular development during stimulation with MENOPUR liquid and MENOPUR powder.
- To evaluate the serum endocrine profile during stimulation with MENOPUR liquid and MENOPUR powder.
- To evaluate the number of oocytes retrieved, the number and quality of embryos, and the number and quality of blastocysts, associated with MENOPUR liquid and MENOPUR powder.
- To evaluate treatment efficiency of MENOPUR liquid and MENOPUR powder.
- To evaluate the safety profile of MENOPUR liquid and MENOPUR powder, including adverse events, routine safety laboratory parameters, local tolerability and immunogenicity.

2.2 Endpoints

Primary Endpoint

- Number of fertilized (2 pronuclei [2PN]) oocytes at 19±2 hours after insemination.

Secondary Endpoints

- Positive βhCG rate (positive βhCG test 10-14 days after blastocyst transfer).
- Clinical pregnancy rate (transvaginal ultrasound showing at least 1 intrauterine gestational sac with fetal heart beat 5-6 weeks after blastocyst transfer).
- Ongoing pregnancy rate (at least one intrauterine viable fetus 8-9 weeks after blastocyst transfer).

- Early pregnancy loss (defined as a positive βhCG test but no ongoing pregnancy at 8-9 weeks after blastocyst transfer).
- Follicular development as assessed by transvaginal ultrasound on stimulation day 6 and last day of stimulation.
- Endocrine profile:
    - Serum follicle-stimulating hormone (FSH) on day 6, last day of stimulation and at oocyte retrieval, and corresponding population pharmacokinetics (PK) analysis.
    - Serum Anti-Müllerian hormone (AMH) on last day of stimulation and at end-of-trial.
    - Human chorionic gonadotropin (hCG) and luteinizing hormone (LH) on day 6 and last day of stimulation.
    - Estradiol (E2) and progesterone (P4) on day 6 and last day of stimulation.
- Number of oocytes retrieved, number of metaphase II oocytes, fertilization rate, and number and quality of blastocysts 5 days after oocyte retrieval.
- Total gonadotropin dose and number of stimulation days.
- Frequency of OHSS (early OHSS if the onset is ≤9 days after triggering of final follicular maturation and late OHSS if the onset is >9 days after triggering of final follicular maturation).
- Frequency and intensity of adverse events.
- Changes in circulating levels of clinical chemistry and hematology 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.
- Frequency of treatment-induced anti-MENOPUR antibodies, overall as well as with neutralizing capacity.
- Technical malfunctions of the pen.

Post-trial Endpoints

- Live birth rate.
- Late pregnancy loss rate (defined as an ongoing pregnancy but no live birth).

- Neonatal health including SAEs at birth, and SAEs at 4 weeks and minimum 6 months after birth.

3 Trial design

3.1 General Design Considerations

Overview of Design Elements

This is a phase 3, randomized, double-blind double-dummy, parallel-group, multicenter non-inferiority trial. Approximately 400 females aged 18-42 years undergoing controlled ovarian stimulation as a part of a gonadotropin-releasing hormone (GnRH) agonist protocol at infertility centers in the US will be randomized 1:1 to receive either MENOPUR liquid (including Placebo to MENOPUR powder) or MENOPUR powder (including Placebo to MENOPUR liquid). Randomization will be stratified by trial site and age strata (<35 years and ≥35 years).

The primary endpoint is the number of fertilized oocytes. Positive βhCG, clinical pregnancy and ongoing pregnancy are included as secondary endpoints in the trial. The secondary endpoints also cover commonly investigated pharmacodynamic parameters such as endocrine profile and ovarian response as well as standard evaluations of safety profile. Furthermore, oocyte and embryo quality will be assessed and a detailed evaluation of blastocyst quality is also included in the trial.

A blinded sample size reassessment will be done when data on the primary endpoint are available for 70% of the planned subjects or when 300 subjects are randomized, whichever comes first. The number of randomized subjects may be adjusted up to a maximum of 500 subjects, corresponding to a blinded one-sample standard deviation estimate of 5.8 and 85% power. No interim analyses intended to compare treatment groups with respect to efficacy or safety are planned.

Subjects are followed through the assessment of ongoing pregnancy. Post-trial follow-up extends to collection of delivery information (live birth and neonatal health), which will be collected for all subjects with an ongoing pregnancy. In addition, data on neonatal SAEs will be collected at 4 weeks and minimum 6 months after birth. Subjects with a treatment-induced anti-MENOPUR antibody response will be followed until the response has become negative or returned to the pre-dosing level, or for a maximum of 1 year.

Additional Design Considerations

Within 3 days of downregulation confirmation (serum E2 ≤20 pg/mL [central laboratory] and transvaginal ultrasound showing no ovarian cysts), subjects will be randomized to receive either MENOPUR liquid (including Placebo to MENOPUR powder) or MENOPUR powder (including Placebo to MENOPUR liquid) initiated at 225 international units (IU) for 5 days. From stimulation day 6 onward, based on follicular response assessed by transvaginal ultrasound, dosing can be adjusted every second day as needed by 75 IU per adjustment. However, the maximum gonadotropin dose will be 450 IU/day and the minimum dose will be 75 IU/day; gonadotropin dosing can continue for a maximum of 20 days, and coasting is not allowed.

Oocyte retrieval will take place 36±2 hours after triggering of final follicular maturation, and oocytes will be inseminated by intracytoplasmic sperm injection (ICSI) 4±2 hours after retrieval;

oocyte maturity will be recorded. Fertilization (number of pronuclei) will be assessed 19±2 hours following ICSI. Blastocyst quality will be assessed on day 5 following oocyte retrieval.

All subjects will have single blastocyst transfer if they have at least one good-quality (i.e. grade 3BB or above) blastocyst available on day 5 following oocyte retrieval. If no good-quality blastocyst is available, they may have double blastocyst transfer, if at least two blastocysts are available. Transfer of day 6 (or later) blastocyst(s) is not allowed.

Blood samples will be collected throughout the trial for the purpose of evaluating the endocrine profile, clinical chemistry and hematology parameters as well as anti-MENOPUR antibodies. Endocrine parameters (FSH, LH, hCG, E2 and P4) are assessed at baseline (E2 at confirmation of downregulation and FSH, LH, hCG and P4 on stimulation day 1), day 6, and at end-of-stimulation (day of trigger), and furthermore, FSH is also assessed at oocyte retrieval. In addition, AMH is assessed on stimulation day 1, on the last day of stimulation, and at end-of-trial. Clinical chemistry and hematology are assessed at screening, last day of stimulation, and end-of-trial. Anti-MENOPUR antibodies are assessed at four occasions. The first sample is taken at the screening visit and is exclusively used to re-establish the anti-drug antibody analytical assays. The subsequent three samples are used for analysis of anti-MENOPUR antibodies in the individual subjects in the trial, and are taken prior to dosing on stimulation day 1 and at two occasions post-dosing: 7-10 days after the last MENOPUR liquid or MENOPUR powder dose (this may coincide with the transfer visit) and 21-28 days after the last MENOPUR liquid or MENOPUR powder dose (this may coincide with the βhCG test visit). Subjects with a treatment-induced anti-MENOPUR antibody response will be followed until the response has become negative or returned to the pre-dosing level.

Local tolerability of MENOPUR liquid or MENOPUR powder following subcutaneous administration will be assessed by the subject three times daily: immediately, 30 minutes and 24 hours after each injection. The assessment of injection site reactions will be made throughout the stimulation period and recorded by the subject in a diary.

Concerning post-trial follow-up, all subjects with an ongoing pregnancy will be followed. Data will be gathered on delivery information (live birth and neonatal health). In addition, data on neonatal SAEs will be collected at 4 weeks and minimum 6 months after birth.

3.2 Sample Size

The present trial is a pharmacodynamic comparison of a marketed product to a new formulation of the same. With 188 subjects per treatment group, the trial has at least 85% power to demonstrate the non-inferiority of MENOPUR liquid to MENOPUR powder in the number of fertilized oocytes at the 1-sided significance level of 0.025. This is based on the results for MENOPUR powder from the COMBINE trial (data on file) conducted in a similar setting (GnRH agonist protocol, MENOPUR starting dose of 225 IU for the first five days, insemination by ICSI etc.) and in a similar population (infertile women between 18 and 42 years). The mean (SD) of the number of fertilized oocytes were 7.88 (5.15) with similar results in the two age groups (7.92 in the <35 years age group vs. 7.80 in

the ≥35 years age group). The non-inferiority margin of -1.60 will retain 80% of the expected comparator effect.

After adjusting for 5% missing data, approximately 400 subjects will be randomized (1:1) into this trial, stratified by trial site and age (<35 years and ≥35 years).

A blinded sample size reassessment will be done when data on the primary endpoint are available for 70% of the planned subjects or when 300 subjects are randomized, whichever comes first. The sample size reassessment will be done without breaking the blind and without inflating the type I error of the trial, in line with current regulatory guidelines. The expected maximum number of subjects to be randomized is 500 (250 subjects per treatment group), corresponding to a blinded one-sample (both groups pooled) standard deviation estimate of 5.8 and 85% power.

4 Subject Disposition

All subjects screened will be accounted for. The number of subjects screened but not randomized to treatment will be presented with the reason for screen failure.

Subject disposition with respect to each analysis set will be tabulated. The number of subjects in the intention-to-treat (ITT) analysis set will be used as the denominator when calculating percentages. Randomized subjects, randomized subjects treated with investigational medicinal product (IMP; MENOPUR liquid including Placebo to MENOPUR powder or MENOPUR powder including Placebo to MENOPUR liquid), and randomized subjects prematurely discontinued from the trial will be summarized. Randomized subjects prematurely discontinued from the trial will have the reason for premature discontinuation summarized.

Subject disposition with respect to the start of the COVID-19 pandemic trial hold will be tabulated and listed using the ITT analysis set. Subject disposition will be summarized by whether a subject had exited the trial before or after the start of the COVID-19 pandemic trial hold. Of the subjects that had not yet exited the trial, the subject status (i.e.; not yet recruited, in screening, or randomized) will be summarized.

Subjects that were in screening but had not yet been randomized prior to the COVID-19 pandemic trial hold were considered screen failures. Subjects that were screen failed due to the COVID-19 pandemic trial hold could re-enter the trial, beginning again the screening process. All subjects that were screen failed due to the COVID-19 pandemic trial hold will be listed along with an indicator of whether the subject re-entered the trial or not. Those subjects that re-entered the trial will have two screening identification numbers. Both identification numbers will be included in the listing.

For subjects that were already randomized but had not yet exited the trial at the start of the COVID-19 pandemic trial hold, the last visit completed prior to the trial hold will be summarized.

A detailed listing of selected 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, and an indicator of whether the missingness was due to the COVID-19 pandemic.

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 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 with the IMP regimen, NIMP regimens, or other serious unforeseen deviations that may affect the conclusions for the primary endpoint of the trial, will lead to exclusion of data from the PP analysis set. Other protocol deviations will not lead to the exclusion from the PP analysis set.

Protocol deviations due to COVID-19 will be listed.

6 Analysis sets

6.1 Intention-to-Treat Analysis Set

The intention-to-treat (ITT) analysis set comprises of all randomized (as planned) subjects.

6.2 Modified Intention-to-Treat Analysis Set

The modified intention-to-treat (mITT) analysis set comprises all randomized (as planned) subjects who received at least 1 dose of IMP (MENOPUR liquid including Placebo to MENOPUR powder or MENOPUR powder including Placebo to MENOPUR liquid).

6.3 Per Protocol Analysis Set

The per protocol (PP) analysis set comprises all mITT subjects except those excluded as a result of major protocol deviations.

6.4 Safety Analysis Set

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

7 Trial Population

7.1 Display of Trial Population Characteristics

Descriptive statistics of demographics and other baseline characteristics will be presented by treatment group as detailed in Section 1.2.1 for the subjects in the mITT, PP, and Safety analysis sets, unless otherwise noted. In the case that the Safety analysis set is identical to the mITT analysis set, presentations will use the mITT and PP analysis sets. Summaries will be made using all subjects, subjects in the <35 years age strata, and subject in the ≥35 years age strata.

In the case of multiple assessments prior to the first exposure to IMP, the baseline value is defined as described in Section 1.1.1.

All demographic and baseline characteristics will also be presented in listings. Listings will be produced for the mITT analysis set only.

Unless otherwise noted, missing data will not be imputed.

7.2 Demographics

Descriptive statistics of baseline demographic variables (e.g.; race and ethnic origin, height, weight, BMI, and age at randomization) will be tabulated.

In addition, for age strata and trial site, the as-randomized stratification factors (according to the actual randomization) and the actual stratification factors (according to the eCRF demographic data), will also be summarized.

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 or later. 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 Menstrual History

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

7.5 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.

7.6 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 positive βhCG, number of positive βhCG 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, number of live births per subject. In addition, the type of treatment (OI/IUI/Both, assisted reproductive technology [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.

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 the start date, type of treatment, medication used, βhCG assessment, clinical pregnancy assessment, and outcome for each fetus.

7.7 Vital Signs

Blood pressure (systolic and diastolic), pulse, and temperature measured on stimulation day 1 will be summarized.

Vital signs on stimulation day 1 will be listed as part of the safety listings (Section 10.2.7.4).

7.8 Physical Examination

Each category of the physical examination is evaluated as normal, abnormal not clinically significant, abnormal clinically significant, or not done. Physical examinations at screening will be summarized by category.

Physical examinations with abnormal findings at screening will be listed as part of the safety listings (Section 10.2.8).

7.9 Endocrine Parameters

Blood samples will be collected for analysis of the following endocrine parameters: AMH, FSH, hCG, LH, E2, and P4. The baseline values will be calculated as described in Section 1.1.1 and summarized as described in Section 1.2.1.

Endocrine parameters will be listed as part of the safety listings (Section 10.2.10.4).

7.10 Gynecological Examination

Each category of the gynecological examination is evaluated as normal, abnormal not clinically significant, or abnormal clinically significant. Gynecological examination at screening will be summarized by category.

Gynecological examination with abnormal findings will be listed as part of the safety listings (Section 10.2.9).

7.11 Follicles

The number of follicles at baseline will be summarized.

7.12 Endometrial Thickness

The endometrial thickness at baseline will be summarized.

7.13 Prior and Concomitant Medication

Prior and concomitant medication will be coded using the World Health Organization (WHO) Drug Reference List. They will be summarized by Anatomic Therapeutic Chemical (ATC) classification 1st level (alphabetically) and ATC classification 2nd level (in decreasing order of frequency). 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 trial).

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

Summaries will be produced for the mITT, PP, and Safety analysis sets. In the case that the Safety analysis set is identical to the mITT analysis set, presentations will use the mITT and PP analysis sets. For NIMP used prior to randomization (i.e.; oral contraceptives and GnRH agonist), a summary will be created that uses all screened subjects that received the corresponding NIMP.

8.1 Extent of Exposure

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

Listings associated with the IMP and NIMP summaries will be presented.

8.1.1 IMP

Exposure to MENOPUR liquid or MENOPUR powder will be summarized by treatment group as the duration of stimulation, the total dose, the maximum and minimum daily doses, and the average daily dose defined as the total dose administered 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).

Investigator-requested decreases and increases of the IMP dose will be captured during the stimulation period. The requested dose change (decrease / increase / no change) on stimulation day 6 will be tabulated. Furthermore, the total number of dose increase requests and dose decrease requests per subject will be tabulated.

8.1.2 NIMPs

Oral Contraceptives

The total dose and duration of treatment will be summarized.

GnRH Agonist

The total dose and duration of treatment will be summarized. In addition, the summary of the starting day of the GnRH agonist, relative to the last day of oral contraceptive use, will be presented.

hCG

The number of subjects that had hCG administered as well as the total dose administered for those subjects that had hCG administered will be summarized.

Luteal Phase Support

The total dose, average total daily dose, the duration of treatment, and the number of subjects that stopped due to menstrual bleeding will be summarized.

8.1.3 Treatment Compliance

Subjects that deviate from the planned treatment will be listed, including non-compliance to both IMP and NIMP.

9 Efficacy

9.1 General Considerations

The presentation of results for each analysis set will proceed as outlined in Section 1.2. Summaries will be made using all subjects, subjects in the <35 years age strata, and subject in the ≥35 years age strata.

9.1.1 Analysis Sets

All primary and secondary efficacy analyses will be conducted for the mITT analysis set. Furthermore, for the primary endpoint, sensitivity analyses will be conducted for the PP and ITT analysis sets as appropriate.

9.1.2 Analysis Methods

Because randomization is stratified by age strata and center, all statistical analyses of efficacy endpoints will incorporate fixed effects for both age strata and center. Except when specifically stated otherwise, all confidence intervals will be 2-sided 95% confidence intervals.

For the primary endpoint and other continuous secondary endpoints (excluding endocrine labs), an ANOVA model will be used and, in SAS, the confidence interval for the treatment difference may be produced either through an LSMEANS statement (with the PDIFF option) or an appropriately formulated ESTIMATE statement. Within-treatment group confidence intervals will also be produced using an LSMEANS statement.

For secondary pregnancy endpoints and other binary secondary endpoints, the Mantel-Haenszel method will be used to combine risk differences across the age strata and center stratification variables to produce the common risk difference. The common risk difference and associated confidence interval will be reported.

Recruitment within strata will be assessed. In the case where one or more strata have too small of a membership to perform inference, centers will be pooled together. Centers with the smallest membership will be pooled together into one or more pooled centers until inference is possible, taking care that no pooled center contains a membership larger than the median enrollment among the remaining centers and the pooled center(s).

9.2 Primary Endpoint

The primary endpoint is defined as the number of fertilized (2 pronuclei [2PN]) oocytes at 19±2 hours after insemination. In addition to the primary endpoint analysis, a frequency table with subjects grouped according to their number of fertilized oocytes will be prepared using these categories: ≤2, 3-5, 6-10, 11-15, and ≥16 2PN.

9.2.1 Primary Endpoint Analysis

The primary objective of the trial is to demonstrate the non-inferiority of MENOPUR liquid versus MENOPUR powder with respect to the number of fertilized oocytes in women undergoing

controlled ovarian stimulation. The null hypothesis is that MENOPUR liquid is inferior to MENOPUR powder where the non-inferiority limit for the difference between treatments (MENOPUR liquid minus MENOPUR powder) is set as -1.60. Therefore, the non-inferiority hypothesis to be tested for the primary endpoint will be:

$$H_0: \mu_{MENOPUR\ liquid} - \mu_{MENOPUR\ powder} \leq -1.60$$

against the alternative

$$H_1: \mu_{MENOPUR\ liquid} - \mu_{MENOPUR\ powder} > -1.60$$

where $\mu_{MENOPUR\ liquid}$ and $\mu_{MENOPUR\ powder}$ denote the mean numbers of fertilized oocytes in subjects treated with MENOPUR liquid and MENOPUR powder, respectively.

The null hypothesis (H_0) will be tested against the alternative by constructing a 2-sided 95% confidence interval for the difference in the least squares mean number of fertilized oocytes between the two treatment groups. If the lower-limit of the 95% confidence interval is greater than the non-inferiority limit (-1.60), the null hypothesis will be rejected and it will be claimed that MENOPUR liquid is non-inferior to MENOPUR powder with respect to the number of fertilized oocytes.

Due to the expected large sample size, the primary endpoint will be analyzed using an ANOVA model. The model will include the treatment group, the age strata, and the center as factors. The least squares mean estimate of the treatment difference in the number of fertilized oocytes and the associated 95% confidence interval will be derived as described in Section 9.1.2. For subjects who do not have any oocytes retrieved, or do not have fertilization assessment due to early withdrawal, or any other reason, the number of fertilized oocytes will be considered as zero.

If the 95% confidence interval for the treatment difference not only lies above the non-inferiority limit (-1.60) but also above zero then there is evidence of superiority in terms of statistical significance at the 2-sided 5% level. With evidence of superiority, the corresponding 2-sided p-value will be reported. There is no need for a multiplicity adjustment since it is a simple closed test procedure.

The primary endpoint analysis will be conducted for the mITT population.

9.2.2 Sensitivity Analyses

Four sensitivity analyses will be conducted. The first two sensitivity analyses will repeat the primary analysis but will use the ITT and PP analysis sets.

The third sensitivity analysis will repeat the primary analysis using the mITT analysis set, but will apply a different method of imputation for missing data. The primary analysis treats all missing primary endpoints as having observed zero 2PN. There are instances, however, where the number

of 2PN for a subject with a missing primary endpoint could theoretically have been greater than zero. Such instances may include: withdrawal of consent, being lost to follow-up, having the study terminated by the sponsor, not attending key visits, due to the COVID-19 pandemic trial hold, or other reasons that would cause a subject with a potential non-zero 2PN count to have a missing value. The temporal nature of the above stated reasons for missing a non-zero 2PN count are such that a subject would need to have taken one or more doses of IMP and have a missing primary endpoint. A sensitivity analysis for the primary endpoint will be performed by using the multiple imputation approach which could impute a non-zero value for missing primary endpoints where it can be reasonably expected that the true value would have been non-zero. The following steps will be taken to implement the multiple imputation method:

- Imputation phase – An ANOVA model with factors for treatment group, age strata, and center will be fitted to data from subjects in the mITT analysis set, less those subjects with missing primary endpoints. For each missing primary endpoint, a total of 1000 draws will be made from a normal distribution that is characterized by the estimated parameters from the fitted model and the factor values associated with the missing record. A total of 1000 complete datasets will be constructed.

- Analysis phase – Each of the 1000 complete datasets will be analyzed using the ANOVA model described in Section 9.1.2.

- Pooling phase – The estimated mean differences and standard errors from the analysis phase will be combined using Rubin's formula (Rubin 1987):

$$m_{MI} = \frac{1}{1000}\sum_{i=1}^{1000} m_i$$

$$se_{MI} = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000} se_i^2 + \left(1+\frac{1}{1000}\right)\left(\frac{1}{1000-1}\right)\sum_{i=1}^{1000}(m_i - m_{MI})^2}$$

  where m_i and se_i are the estimated mean treatment difference and standard error for the treatment difference from the i^{th} complete dataset and m_{MI} and se_{MI} are the pooled estimates. The mean difference in the number of 2PN between the MENOPUR liquid and the MENOPUR powder groups will be estimated as m_{MI} and the associated 95% confidence interval will be calculated as $m_{MI} \pm 1.96 * se_{MI}$.

The fourth sensitivity analysis will again repeat the primary endpoint analysis using the mITT analysis set but will instead use a 95.08% two-sided confidence interval to assess non-inferiority. This sensitivity analysis assesses the impact of any Type I error inflation incurred due to performing a blinded sample size recalculation during a non-inferiority study. See Appendix 1 for more details.

Furthermore, the number of fertilized oocytes will be summarized for several subgroups:

- By reason for infertility
- By center
- By number of oocytes retrieved (<4, 4-7, 8-14, 15-19, and ≥20)
- By AMH level on stimulation day 1 (<7.5, 7.5-<15, 15-<25, ≥25 pmol/L; ≤ median, > median)

9.3 Secondary Efficacy Endpoints

Secondary efficacy endpoints will be analyzed based on the mITT analysis set and summarized based on both the mITT and PP analysis sets. For several endpoints, summaries consisting of subgroups of the mITT analysis set will be presented. Descriptive summaries will be tabulated as described in Section 1.2.1.

9.3.1 Positive βhCG Rate

The positive βhCG rate will be presented by treatment group and analyzed using the method described for secondary pregnancy endpoints in Section 9.1.2. Subjects who do not have a confirmed positive βhCG test will be counted as not having positive βhCG unless a positive clinical or ongoing pregnancy outcome is observed later. For example, if the outcome βhCG is missing, but the clinical pregnancy outcome is 'positive' then βhCG will be imputed as 'positive'.

9.3.2 Clinical Pregnancy Rate

The clinical pregnancy rate will be presented by treatment group and analyzed using the method described for secondary pregnancy endpoints in Section 9.1.2. Subjects who do not have at least 1 intrauterine gestational sac, confirmed by TVUS, with a fetal heart beat at 5-6 weeks after blastocyst transfer will be counted as not having clinical pregnancy unless a positive ongoing pregnancy outcome is observed later. For example, if the clinical pregnancy assessment is missing, but the ongoing pregnancy outcome is 'positive' then clinical pregnancy will be imputed as 'positive'.

9.3.3 Ongoing Pregnancy Rate

The ongoing pregnancy rate will be presented by treatment group and analyzed using the method described for secondary pregnancy endpoints in Section 9.1.2.

9.3.4 Number and Size of Follicles During Stimulation

The follicle cohort on stimulation day 6 and the last day of stimulation will be summarized by treatment on the follicle level (number of follicles ≤9 mm, 10-11 mm, 12-14 mm, 15-16 mm, and ≥17 mm) and the subject level (largest follicle size and average number of follicles ≥17 mm, ≥15

mm, and ≥12 mm). The subject level parameters will be analyzed as described in Section 9.1.2 for the primary endpoint.

9.3.5 Levels of Endocrine Parameters

The levels of endocrine parameters will be presented by treatment group for both the scheduled visit and the change from baseline for post-baseline visits. Baseline values for endocrine parameters are calculated as defined in Section 1.1.1. 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.

The following list describes which post-baseline visits (and, consequently, change from baseline calculations) are summarized for each of the endocrine parameters.

- AMH: End-of-stimulation and end-of-trial.
- FSH: Stimulation day 6, end-of-stimulation, and oocyte retrieval.
- hCG, LH, Estradiol, and Progesterone: Stimulation day 6 and end-of-stimulation.

For each parameter the change from baseline will be compared between treatment groups using a log-normal model. In this model, the follow-up measurement will be the dependent variable and the linear predictors will include treatment, age strata, and center as factors as well as the baseline measurement (log-transformed) as a covariate. Inferential results will be presented as described in Section 1.2.2.

9.3.6 Number of Oocytes Retrieved

The number of oocytes retrieved will be presented by descriptive statistics by treatment group and analyzed as described in Section 9.1.2 for the primary endpoint. Furthermore, a frequency table with subjects grouped according to their number of oocytes retrieved will be prepared using these categories: ≤3, 4-7, 8-14, 15-19, and ≥20 oocytes. Subjects without oocyte retrieval will be imputed to have 0 oocytes retrieved. For example, if the cycle is cancelled due to poor ovarian response, the number of oocytes retrieved is imputed as 0. Additional subgroup summaries of the number and distribution of oocytes retrieved for subjects with oocytes retrieved will also be presented.

9.3.7 Number of Metaphase II Oocytes

Oocytes will have their maturity stage assessed prior to insemination. Maturity stage will be categorized as germinal vesicle, metaphase I, metaphase II, degenerated, or other. For subjects that have oocytes retrieved, but are missing maturity stage information, the maturity stage of their oocytes will be regarded as unknown (i.e.; not metaphase II).

The number and distribution (≤3, 4-6, 7-12, 13-17, and ≥18) of MII oocytes will be presented by descriptive statistics by treatment group. Furthermore, the number of inseminated oocytes, the sperm source, and the sperm type will also be presented by descriptive statistics by treatment group.

9.3.8 Fertilization Rate

For subjects with >0 oocytes retrieved, the rate of fertilized oocytes to oocytes retrieved will be presented by descriptive statistics by treatment group. Furthermore, for subjects with >0 metaphase II oocytes, the rate of fertilized oocytes to metaphase II oocytes will be presented by descriptive statistics by treatment group. Fertilization rates will be summarized based on the subset of the mITT analysis set.

9.3.9 Number and Quality of Blastocysts on Day 5

The embryos on Day 5 after oocyte retrieval will be summarized 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, destiny, and grading. 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 will be derived for all subjects in the analysis set. For any subject that did not have a blastocyst, did not have oocyte retrieval, or otherwise had a missing subject-level endpoint, a value of 'NO' or 0 will be imputed, as appropriate.

- At least one blastocyst available
- At least one good-quality blastocyst available
- Number of blastocysts
- Number of good quality blastocysts
- Number of blastocysts transferred

Among the above variables, the number of blastocysts and the number of good quality blastocysts will be analyzed as described in Section 9.1.2 for the primary endpoint. The dichotomous variables of having at least one blastocyst available and of having at least one good quality blastocyst available will be analyzed as described in Section 9.1.2 for secondary pregnancy endpoints. No analysis will be performed for the remaining variables in this section.

At the subject level the following variables will be derived for mITT analysis set subjects that had oocytes retrieved:

- Rate of blastocysts to oocytes retrieved

- Rate of good-quality blastocysts to oocytes retrieved

At the subject level the following variables will be derived for mITT analysis set subjects that had at least one fertilized oocyte:

- Rate of blastocysts to fertilized oocytes
- Rate of good quality blastocysts to fertilized oocytes

10 Safety

10.1 General Considerations

Safety summaries, unless otherwise stated, will be performed using the Safety analysis set and will be summarized as described in Section 1.2.1. Summaries will be made using all subjects, subjects in the <35 years age strata, and subject in the ≥35 years age strata.

All individual subject data will be listed per subject and treatment as observed including any derived values. Adverse event listings will also include an indicator of whether the adverse event occurred before, during, or after the COVID-19 pandemic trial hold.

10.1.1 Imputation rules

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 Safety endpoints

10.2.1 Adverse Events

Adverse events will be coded using MedDRA. Adverse events will be grouped according to the start of both NIMP and IMP as follows:

- Pre-NIMP adverse event, i.e. any adverse event occurring after signed informed consent and before start of NIMP, or a pre-existing medical condition that worsens in intensity after signed informed consent but before start of NIMP.

- Pre-treatment adverse event, i.e. any adverse event occurring after the start of NIMP and before start of IMP, or a pre-existing medical condition that worsens in intensity after start of NIMP but before start of IMP.

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

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

Written narratives will be issued for all SAEs and adverse events leading to discontinuation. Adverse events judged by the investigator as being reasonably possibly related to IMP will be termed adverse drug reactions.

10.2.1.1 Overview of Treatment Emergent Adverse Events

A summary table for treatment-emergent adverse events will be presented, including for each treatment, the number of subjects reporting an adverse event, the percentage of subjects (%) with an adverse event, and the number of events (E) reported, for the following categories:

- Treatment-emergent adverse events.
- Treatment-emergent severe adverse events.
- Treatment-emergent non-serious adverse events.
- Treatment-emergent serious adverse events.
- Treatment-emergent adverse drug reactions.
- Treatment-emergent adverse events leading to discontinuation.
- Treatment-emergent deaths.

10.2.1.2 Incidence of Adverse Events

Treatment-emergent adverse events in each treatment group will be tabulated by system organ class (SOC) and preferred term (PT). 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 (E) reported. AEs will be presented by SOC sorted alphabetically and PT sorted in decreasing frequency of occurrence.

Summary tables will be prepared for:

- All treatment-emergent adverse events.
- Treatment-emergent adverse events occurring in ≥2% and ≥5% of subjects in any treatment group.
- Treatment-emergent adverse drug reactions occurring in ≥2% and ≥5% of subjects in any treatment group.
- Treatment-emergent adverse events by causality.
- Treatment-emergent adverse events by intensity.
- Treatment-emergent adverse drug reactions by intensity.
- Treatment-emergent adverse events leading to discontinuation.

- Treatment-emergent serious adverse events.
- Treatment-emergent adverse events leading to death.

Supporting data listings will be provided for:

- All pre-NIMP adverse events sorted by SOC and PT.
- All pre-treatment adverse events sorted by SOC and PT.
- All treatment-emergent adverse events sorted by SOC and PT.
- All treatment-emergent adverse events sorted by center and subject number.
- All unique verbatim and coded terms.
- All SAEs.
- All non-Serious adverse events.
- All adverse events leading to discontinuation.
- All adverse events leading to death.

10.2.2 OHSS

OHSS for each treatment group will be tabulated by classification (mild, moderate, severe) and grade (1, 2, 3, 4, 5). These tabulations will be made for OHSS overall as well as separately for early OHSS and late OHSS.

10.2.3 Early Pregnancy Loss

Early pregnancy loss is defined as a positive βhCG test but no ongoing pregnancy at 8-9 weeks after blastocyst transfer. The early pregnancy loss rate will be presented by treatment group. The denominator for the early pregnancy loss rate consists of those subjects with a positive βhCG.

10.2.4 Injection Site Reactions

For each injection site reaction (redness, pain, itching, swelling, and bruising), the number of events and the number of subjects experiencing those events will be tabulated by time (immediately, 30 minutes, and 24 hours), reaction type, and intensity (none, mild, moderate, and severe). Because of the double-dummy design, injection site reactions for IMP (MENOPUR liquid and MENOPUR powder) will be summarized together and injection site reactions for Placebo (Placebo to MENOPUR liquid and Placebo to MENOPUR powder) will be summarized together.

10.2.5 Treatment-induced Anti-MENOPUR Antibodies

The proportion of subjects with treatment-induced anti-MENOPUR antibodies as well as the proportion of subjects with treatment-induced anti-MENOPUR antibodies with neutralizing capacity will be tabulated. Within each treatment group, a 95% Clopper-Pearson confidence interval will be constructed for both the proportion of subjects with treatment-induced anti-MENOPUR antibodies and for the proportion of subjects with treatment-induced anti-MENOPUR antibodies with neutralizing capacity.

10.2.6 Technical Malfunction of the Pen

The frequency of reported malfunctions of the administration pen will be presented in a summary table.

10.2.7 Vital Signs

The baseline for the vital sign analyses will be calculated as described in Section 1.1.1.

10.2.7.1 Summary Statistics

Mean change from baseline at each time point will be presented for each vital sign variable using the process outlined in Section 1.2.1.

10.2.7.2 Changes Relative to Normal Range

Each vital sign variable value is classified as either ‘Low’, ‘Normal’, or ‘High’. Shift tables will be prepared to show the number and percentage of subjects that experience a shift from baseline to the end-of-stimulation visit and from baseline to the end-of-trial visit.

10.2.7.3 Markedly Abnormal Changes

Summary tables 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 vital signs are specified in Appendix 2.

10.2.7.4 Data Listings

Data listings will be prepared by treatment group, center, subject number, and time point for all vital sign values (including screening and baseline) and abnormal values will be flagged.

10.2.8 Physical Examination

Physical examination at the end-of-trial visit will be compared to baseline (as defined in Section 1.1.1) and summarized in shift tables. All subjects with any abnormal finding will be listed per subject.

10.2.9 Gynecological Examination

Gynecological examination at the end-of-trial visit will be compared to baseline (as defined in Section 1.1.1) and summarized in shift tables. All subjects with any abnormal finding will be listed per subject.

10.2.10 Safety Laboratory Variables

The baseline for the safety laboratory variable analyses will be calculated as described in Section 1.1.1.

10.2.10.1 Summary Statistics

The circulating levels of clinical chemistry and hematology parameters, including change from baseline, will be tabulated for each time point for each laboratory variable. Mean change from baseline to the end-of-stimulation visit and to the end-of-trial visit will be presented for each laboratory variable using the process outlined in Section 1.2.1.

10.2.10.2 Changes Relative to Normal Range

Each safety laboratory variable value is classified as either 'Low', 'Normal', or 'High'. Shift tables will be prepared to show the number and percentage of subjects that experience a shift from baseline to the end-of-stimulation visit and from baseline to the end-of-trial visit.

10.2.10.3 Markedly Abnormal Changes

Summary tables 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 hematology variables are specified in Appendix 2.

10.2.10.4 Data Listings

Data listings will be prepared by treatment group, center, subject number, and time point for all safety laboratory values (including screening) and abnormal values will be flagged along with their determination of clinically significant or not clinically significant. Laboratory variables will be grouped under "Hematology" and "Clinical Chemistry".

11 Post-trial Assessments

The post-trial assessments will be reported as an addendum to the 000303 clinical study report.

11.1 Live Birth Rate

The live birth rate will be presented by treatment group and analyzed using the method described for secondary pregnancy endpoints in Section 9.1.2. Subjects with no information on live birth will be defaulted to a negative response.

11.2 Late Pregnancy Loss

Late pregnancy loss is defined as an ongoing pregnancy but no live birth. The rate of late pregnancy loss will be presented by treatment group.

11.3 Neonatal Health

The gestational age, gender, weight, length, and Apgar scores (1 minute, 5 minutes, and 10 minutes) will be presented by treatment group. Furthermore, admissions to care units (NICU, NCU, or PCU), admissions to care units for longer than 2 hours, malformations or congenital anomalies, neonatal death, and important medical events will be presented by treatment group at birth, at 4 weeks after live birth, and after 6 months following live birth.

Neonatal SAEs will be summarized from birth through a minimum of 6 months after birth. Written narratives will be issued for all SAEs.

Parameters relating to maternal health will also be presented by treatment group. These will include clinically relevant pregnancy, labor, or delivery complications, COVID-19 infection, exposure to risk factors of congenital anomalies or malformations, and maternal death in connection with pregnancy or labor.

12 Interim analyses

A sample size reassessment will be performed without breaking the blind and without meaningful inflation of the Type I error of the trial (see Appendix 1), in line with the current regulatory guidelines for non-inferiority trials by the Food and Drug Administration (FDA). The one-sample standard deviation of the primary efficacy endpoint will be evaluated in a blinded manner in relation to the assumptions underlying the sample size calculation.

The sample size reassessment is to take place when 70% of the planned sample size has provided a primary endpoint (280 primary endpoints) or when 300 subjects have been randomized, whichever occurs first.

A restricted recalculation rule for the sample size reassessment will be used. The trial will originally target 400 randomized subjects. At the time of the interim analysis, the estimated one-sample standard deviation will be used to recalculate the required sample size. If the recalculated sample size is less than 400, then the targeted sample size will remain at 400 (i.e.; the sample size will never decrease below 400). If the recalculated sample size is greater than 400, however, then the targeted sample size will be increased to either the recalculated sample size or 500, whichever is smaller.

13 Deviations from the protocol

There is one deviation from the planned analysis in the protocol. Center will be incorporated into statistical analyses as a fixed effect instead of a random effect.

Because randomization is stratified by age strata and center, it was planned in the protocol to perform analyses by adjusting for both age strata and center. Age strata was to be incorporated as a fixed effect and center was to be incorporated as a random effect.

In this statistical analysis plan, however, we incorporate center as a fixed effect. The reason for doing so is twofold. First, centers are rarely chosen at random. Second, the interpretation of an interaction between center and any other variable of interest would be complicated by the use of a random effect. Interactions between fixed effects are much more readily interpretable. While there are no pre-planned analyses that incorporate an interaction that includes center, it is possible that a post-hoc analysis could.

14 Tables, Listings and Figures

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

15 References

1. Rubin, D.B. (1987) Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons.

Appendix 1

Assessment of Type I Error Inflation

In November 2019, the FDA released a guidance for industry document titled, "Adaptive Designs for Clinical Trials of Drugs and Biologics." Of relevance for the present trial, Section IV: Adaptive Designs Based on Non-Comparative Data states, "These adaptations generally do not inflate the Type I error probability. However, there is the potential for limited Type I error probability inflation in trials incorporating hypothesis tests of non-inferiority or equivalence (Friede and Kieser 2003). Sponsors should evaluate the extent of inflation in these scenarios." The present trial is a non-inferiority study that includes a form of non-comparative adaptive design, namely a blinded sample size recalculation.

Using the methods described in Appendix 1 of Friede and Kieser (2003), a simulation study was performed. The simulation study assessed the potential Type I error of the present trial given the design considerations for the sample size recalculation (See Section 3.2). The 'true' pooled standard deviation was set to range from 3.0 to 15.0 using increments of 0.1. For each increment, 10,000,000 simulations were performed. Based upon the design considerations of the present trial, the minimum sample size was set to 400, the maximum sample size was set to 500, and the sample size recruited at the time of the sample size recalculation was set to 280. Likewise, when recalculating the sample size during each simulation rep, an alternative treatment difference of 0 was assumed, the non-inferiority margin was set to -1.6, a two-sided alpha of 0.05 was used, and the power was set to 85%.

The results of the simulation can be seen in Figure 0-1 below. For true pooled standard deviations that are less than 5 and greater than 6, the Type I error is well controlled. For true pooled standard deviations between 5 and 6, there is a small amount of Type I error inflation. The maximum Type I error observed was 0.0253705, which occurred at a true pooled standard deviation of 5.6.

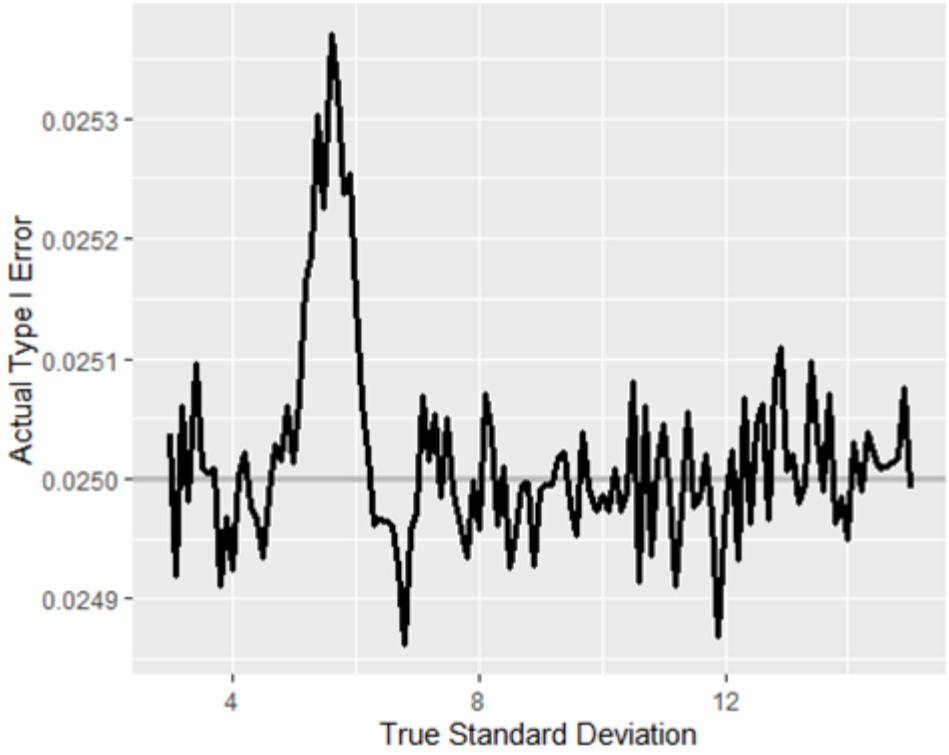

Figure 0-1: Actual Type I error of the 000303 trial given the true pooled standard deviation.

One conservative method for controlling the maximum Type I error inflation was illustrated by Friede and Stammer (2010). An iterative procedure is performed, adjusting alpha until the maximum observed Type I error is within a certain distance, epsilon, of the predefined desirable level of Type I error. This method was applied to the present trial by setting epsilon equal to 0.0001 (i.e.; a maximum Type I error no greater than 0.0251).

The results of applying the procedure can be seen in Figure 0-2 below. When adjusting the one-sided Type I error from 0.025 to 0.0246 (or, equivalently, the two-sided Type I error from 0.05 to 0.0492), the maximum observed Type I error is within 0.0001 of the pre-specified Type I error of 0.025. The maximum observed Type I error was 0.0250291 and again occurred at a standard deviation of 5.6.

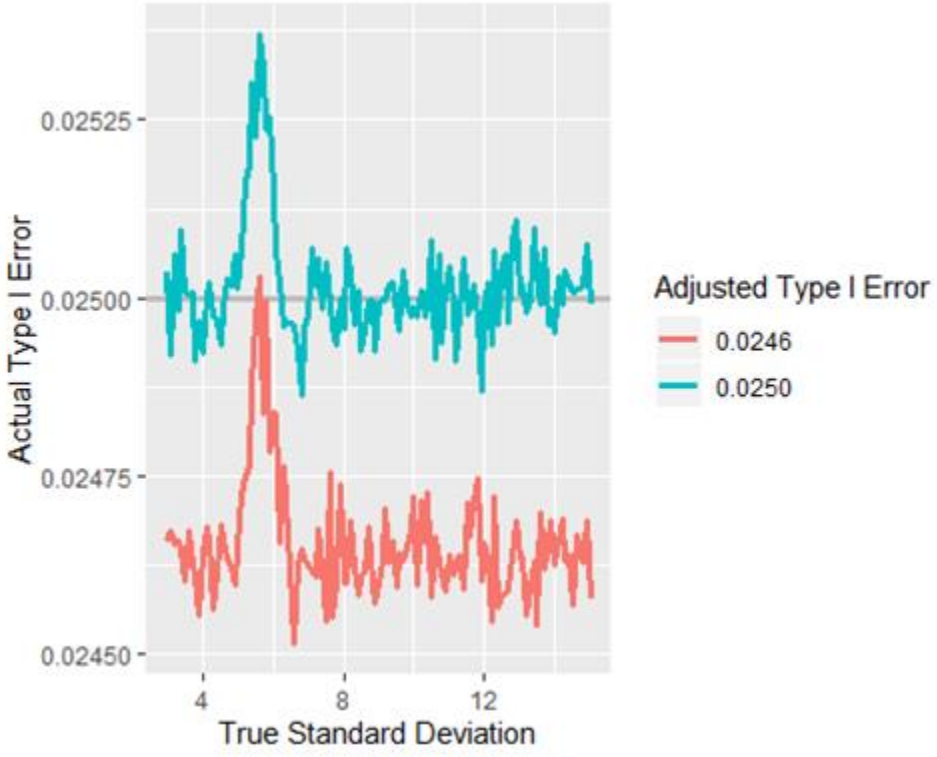

Figure 0-2: Adjusted Type I error after applying an alpha correction.

The Type I error inflation observed in Figure 0-1 is considered to be small and inconsequential. Therefore, no adjustments to the primary analysis are taken. A sensitivity analysis, however, is added. As described in Section 9.2.2, the primary analysis will be repeated using the same analysis set and imputation approach, but a different alpha level when constructing the two-sided confidence interval. Instead of a two-sided 95% confidence interval, a two-sided 95.08% confidence interval will be constructed. This confidence interval reflects the adjustment necessary to contain the maximum Type I error to be within 0.0001 of the pre-specified desired level of Type I error.

Below are the references made in the appendix.

1. Friede, T. & Kieser, M. (2003) Blinded sample size reassessment in non-inferiority and equivalence trials. *Statistics in Medicine*, 22:995-1007. DOI: 10.102/sim.1456.

2. Friede, T. & Stammer, H. (2010) Blinded sample size recalculation in noninferiority trials: A case study in dermatology. *Drug Information Journal*, 44:598-608.

Appendix 2

Table 1 Markedly Abnormal Criteria for Laboratory Tests

Variable Code	Variable	Units	Low	High
RBC	Erythocytes	10^12/L	< 3.0	N.A
WBC	Leukocytes	10^9/L	< 2.5	> 15.0
HGB	Haemoglobin	g/L	< 100	> 190
HCT	Haematocrit	RATIO	< 0.28	> 0.5
EOSLE	Eosinophils/Leukocytes	%	N.A	≥ 10
NEUTLE	Neutrophils/Leukocytes	%	≤ 15	≥ 90
LYMLE	Lymphocytes/Leukocytes	%	≤ 10	≥ 80
MONOLE	Monocytes/Leukocytes	%	N.A	≥ 20
BASOLE	Basophils	%	N.A	≥ 5
PLAT	Platelets	10^9/L	< 110	> 600
AST	AST	IU/L	N.A	> 3xULN
BUN	Blood urea nitrogen	mmol/L	N.A	> 12.5
CA	Calcium	mmol/L	< 1.75	> 2.74
CL	Chloride	mmol/L	N.A	N.A
ALT	ALT	IU/L	N.A	> 3xULN
CREAT	Creatinine	umol/L	N.A	> 3xULN
GGT	GGT	IU/L	N.A	> 3xULN
GLUC	Glucose	mmol/L	< 2.5	> 16.7
K	Potassium	mmol/L	< 3.0	> 5.8
SODIUM	Sodium	mmol/L	< 125	> 155
ALP	Alkaline phosphatase	IU/L	N.A	> 3xULN
LH	Luteinizing hormone	IU/L	N.A	>10

Table 2 Markedly Abnormal Criteria for Vital Signs*

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