STATISTICAL ANALYSIS PLAN

Veru Inc.
48 NW 25th St., Suite 102
Miami, FL 33127

RANDOMIZED, DOUBLE-BLIND, PLACEBO CONTROLLED, DOSE FINDING PHASE 2 STUDY COMPARING ORAL DAILY DOSING OF VERU-944 AFTER A WEEK OF LOADING (DAILY DOSING) WITH PLACEBO TO AMELIORATE THE VASOMOTOR SYMPTOMS RESULTING FROM ANDROGEN DEPRIVATION THERAPY IN MEN WITH ADVANCED PROSTATE CANCER

Protocol No: V72203
Protocol Date/Version: 30OCT2019, Version 9.0, Amendment 4

Approval

Veru Inc.
48 NW 25th St., Suite 102
Miami, FL 33127

Randomized, double-blind, placebo controlled, dose finding Phase 2 study comparing oral daily dosing of VERU-944 after a week of loading (daily dosing) with placebo to ameliorate the vasomotor symptoms resulting from androgen deprivation therapy in men with advanced prostate cancer

Protocol No: V72203

K. Gary Barnette, PhD **Chief Scientific Officer** **Veru Inc.**	10/30/2019 **Approval Date**
Michael L Hancock, MS, CCRP **Consultant Biostatistician** **Veru Inc.**	**Approval Date**
• Michael L. Hancock	• 10/30/2019

REVISION HISTORY

Version 1: 7 February 2019 – Version 2, dated 30 October 2019 replaces Version 1 dated 7 February 2019. The significant changes made between Version 1 and Version 2 are:

1. The time of the primary endpoint is assessed is changed from Day 28 to Day 42. This change is being made to address the potentially long terminal half-life of cis-clomiphene in some patients (up to 30+ days).
2. The severity of hot flashes is moved from a primary endpoint to secondary endpoints.
3. Frequency and severity of the hot flashes will be conducted at Weeks 4, 6, 8, 10 and 12.
4. Wilcoxon rank sum test and t-tests are added as prospective analyses in this Phase 2 study.
5. Other clarifying edits are made.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation	**Term**
ADT	Androgen deprivation therapy
AE	Adverse event
AIC	Akaike's Information Criteria
ATC	Anatomic therapeutic chemical (classification)
BMI	Body mass index
ECG	Electrocardiogram
eCRF	Electronic Case report form
eDiary	Electronic Diary
EE	Efficacy Evaluable (population)
ITT	Intent-to-treat (population)
max	maximum
MedDRA	Medical dictionary of regulatory activities
mg	Milligram
min	minimum
MMRM	Mixed Models Repeated Measures
PK	Pharmacokinetic
PO q day	Medication is taken orally once a day
PSA	Prostate specific antigen
PT	Preferred term

SAE	Serious adverse event
SD	Standard deviation
SHBG	Sex hormone binding globulin
SOC	System organ class
TEAE	Treatment-emergent adverse event
TLF	Tables, listings, and figures
VTE	Venothromboembolism (Caprini VTE Risk Assessment)
WHODDE	World health organization drug dictionary enhanced

1. INTRODUCTION

Prostate cancer is the most common form of cancer in men, with androgen deprivation therapy (ADT) as the standard of care for many patients. The main objective of ADT is to lower the serum testosterone level in men which can result in a number of estrogen deficiency side effects the most common of which is known as hot flashes. Hot flashes are experienced by approximately 75% of men receiving ADT and are a significant component in assessing quality of life (Holzbeierlein et al., 2004). Men describe hot flashes with both physical and emotional descriptors. Men with advanced prostate cancer, being treated with ADT, have described their hot flashes as having detrimental effects on their overall quality of life to the extent that cancer-related distress caused by hot flashes is a contributing factor in the discontinuation of ADT (Engstrom, 2008; Ulloa et al., 2009; Frisk, 2010).

The precise mechanism of hot flashes in men is not yet fully understood. There are currently no approved therapies for the treatment of moderate to severe hot flashes in men with advanced prostate cancer on ADT. As a result of ADT, decreasing estrogen levels appear to be a contributing cause of the hot flashes. A nonsteroidal estrogen, VERU-944, is a hormonal therapy that can be used to alleviate vasomotor symptoms (hot flashes) in men undergoing ADT.

2. STUDY OBJECTIVES

2.1 Primary Objective

To determine an effective dose of VERU-944 (from among 10 mg, 50 mg, and possibly 100 mg PO q day), after a loading dose, for the treatment of vasomotor symptoms commonly known as hot flashes by assessing its effect on:

1. the frequency of moderate to severe hot flashes at Week 6

2.2 Secondary Objectives

To assess the effect of VERU-944 on:

1. the severity of moderate to severe hot flashes at Week 6
2. the frequency of moderate to severe hot flashes at Week 8
3. the frequency of moderate to severe hot flashes at Week 10
4. the frequency of moderate to severe hot flashes at Week 12
5. the severity of moderate to severe hot flashes at Week 8
6. the severity of moderate to severe hot flashes at Week 10
7. the severity of moderate to severe hot flashes at Week 12
8. the frequency of moderate to severe hot flashes at Week 4
9. the severity of moderate to severe hot flashes at Week 4
10. bone turnover markers

2.3 Exploratory Objectives

To assess the effect of VERU-944 on:

1. serum prostate-specific antigen (PSA) concentrations
2. serum total and free testosterone concentrations
3. serum sex hormone binding globulin (SHBG) concentrations

2.4 Safety Objective

To assess the safety and tolerability of VERU-944

3. STUDY DESIGN

This is a randomized, multicenter, double-blind, placebo-controlled dose finding study of VERU-944 to treat hot flashes (vasomotor symptoms) in men with advanced prostate cancer on ADT. The study will be conducted as a staged study with Stage 1 consisting of 3 arms (placebo, 10 mg and 50 mg of VERU-944). When all subjects in Stage 1 have completed the Week 6 of dosing, an interim analysis will be conducted to assess efficacy (unblinded) and safety (blinded) of the doses administered in Stage 1 at Week 6 frequency and severity – with the Week 6 frequency analysis being the primary analyses. Following review of the Stage 1 results, a decision will be made whether to progress with Stage 2 of the study. If initiated, Stage 2 will have 2 arms (placebo and 100 mg VERU-944). It is important to note that the placebo arm of Stage 2 will be enrolled with separate and distinct subjects from those subjects who receive placebo in Stage 1. There is no intent to combine these placebo arms for analysis purposes of this protocol.

The subjects participating in the study will have advanced prostate cancer and will be undergoing ADT with a luteinizing hormone releasing hormone therapy (agonist or antagonist) for at least 3 months prior to randomization and be experiencing regular moderate to severe hot flashes while on ADT. Subjects will continue to receive ADT over the course of the study.

Approximately 30 subjects will be enrolled into each arm, regardless of stage of study. Each subject will be randomized to receive placebo, 10 mg VERU-944, or 50 mg VERU-944 in Stage 1, and placebo, or 100 mg VERU-944 in Stage 2. Following randomization, each subject will receive a 4-day loading dose of the treatment to which they have been randomized. On Day 1, each subject will receive 5 times their maintenance dose at a single time. On Days 2-4 each subject will receive 3 times their maintenance dose at a single time each day. Then on Days 5-84, each subject will receive 1 capsule per day. Dosing should occur greater than 1 hour before or 1 hour after eating and will continue for 12 weeks (84 days).

4. SCHEDULE OF ASSESSMENTS

The schedule of study evaluations are provided in the protocol (Appendix B) and are copied in Appendix 9.1.

4.1 Screening

Potential subjects will be screened for this study in the 28 days prior to enrollment. The following activities will be conducted at screening:

1. Signed informed consent will be obtained prior to any study-specific procedures and a copy of the signed consent form will be given to the subject.
2. Assess subject eligibility for inclusion into this study based on protocol inclusion/exclusion criteria.
3. Medical history will be obtained, including diagnosis of primary disease (date of diagnosis), clinical stage and Gleason score (individual scores and sum) at diagnosis, presence, date and extent of soft tissue and bone metastases, details and dates of primary treatment, and details and dates of prior therapies for prostate cancer, concurrent illnesses and family history.

4. The use of any medications will be recorded (including radiation). This will include medications currently being taken and those taken within the last 30 days. This will also include the type of androgen deprivation therapy (ADT) utilized and when it was initiated. Over-the-counter (OTC) medications as well as medications taken on an as-needed basis (PRN) should be recorded.
5. Vital signs (temperature/pulse/supine blood pressure)
6. Physical examination including height and weight
7. History of hot flash frequency and severity
8. Thromboembolic risk assessment including: a) Factor V Leiden gene mutation; b) Lupus Anticoagulant assessment and Cardiolipin assessment; c) Prothrombin gene mutation; and Protein C and S.

   NOTE: detection of lupus anticoagulant, cardiolipin antibody outside the normal range, prolonged PTT-LA, dRVVT and Protein C and/or Protein S deficiency are not exclusionary from the study. However, if a patient has lupus anticoagulant detected, cardiolipin outside the normal range, a prolonged PTT-LA, a prolonged dRVVT or is deficient for Protein C and/or Protein S, and remains eligible for participation in the study, prophylactic anticoagulation therapy should be considered.

   For Protein C or Protein S deficiency, consideration should be given to dietary changes to increase Vitamin K intake.
9. Electrocardiogram- 12 lead (single)
10. Caprini Venothromboembolism (VTE) Risk Assessment

    NOTE: Caprini VTE Risk is not an inclusion/exclusion criteria for the study. It is intended to assess the patient's overall risk for VTE and the change of risk over the course of the study.
11. Clinical Laboratory Tests
    - Hematology (Appendix A)
    - Serum hormones (Appendix A)
    - Serum chemistry (Appendix A)
    - Urinalysis
12. Measurement of frequency and severity of hot flashes utilizing electronic device (values over 14 contiguous days in order to establish a weekly baseline)
    - Must be at least 80% compliant during this assessment. This translates into recording frequency and severity of hot flashes at least 11 of the 14 days.
13. Bilateral Doppler ultrasound of the lower extremity (If the subject is diagnosed with a VTE in this assessment, the subject will not be eligible for the study under the exclusion criterion of history of a VTE).

Patients that have been previously screened for this study and could not complete screening for reasons other than not meeting the inclusion and exclusion criteria, may be rescreened into the study. Laboratory testing conducted under this protocol (Appendix A) within 45 days prior to randomization into this study, may be used as the baseline laboratory values for this study and repeat laboratory testing is not required

unless new coadministered medication(s) has been started, unless the investigator determines that retesting is warranted.

Patients that failed screening under previous versions of the protocol due to a positive lupus anticoagulant, but in whom lupus anticoagulant was not "detected" may be rescreened under Amendment 2, Version 7.0 of the protocol.

Patients that failed screening under previous versions of the protocol due to indeterminant cardiolipin antibody, may be rescreened under Amendment 2, Version 7.0 of the protocol.

Patients that failed screening under previous versions of the protocol due a finding related to Protein C or Protein S, including patients that were excluded from participation inappropriately due to elevated levels of Protein S and/or Protein C above the upper limit of normal, may be rescreened under Amendment 2, Version 7.0 of the protocol.

For patients that fall into the three categories outlined above, laboratory testing conducted under this protocol (Appendix A) within 45 days prior to randomization into this study, may be used as the baseline laboratory values for this study and repeat laboratory testing is not required unless new coadministered medication(s) has been started or if the investigator determines that retesting is warranted.

4.2 Enrollment

Approximately 90 subjects will be randomized into Stage 1 and 60 into Stage 2. Subjects will be randomized from 2 to 5 days prior to the Day 1 visit. Randomization should be done after the subject has completed the screening assessments and after it has been determined that the subject meets all the inclusion and exclusion criteria.

Assess eligibility status for inclusion into the trial based on protocol inclusion/exclusion criteria, including results from screening lab assessments.

4.3 Day 1 Visit

Subjects should fast for at least 8 hours (overnight) prior to this visit. The Day 1 visit assessments will serve as the baseline assessments, unless otherwise specified.

1. The medical history should be reviewed and updated to include any changes occurring since the screening visit.
2. Assessment of eligibility
3. Vital signs (temperature/pulse/supine blood pressure)
4. Physical examination (including weight)
5. Hematology
6. Serum chemistry
7. Urinalysis
8. Serum hormones
9. Bone turnover markers
10. Record the usage of any concomitant medications and ongoing treatments
11. Adverse events (ongoing on Day 1)
12. Caprini VTE Risk Assessment
13. Blood samples for pharmacokinetic (PK) assessment
14. Assess hot flashes
15. Dispense study drug
16. First dose- instruct the subject to take his capsule at the same time each day

4.4 Day 14 Visit

The following assessments will be conducted on Day 14 (±3 days):

1. Vital Signs (temperature/pulse/supine blood pressure)
2. Serum chemistry
3. Record the usage of any concomitant medications and treatments
4. Adverse events

4.5 Day 30 and Day 60 Visits

Subjects should fast for at least 8 hours (overnight) prior to these visits. The following assessments will be conducted on Day 30 (± 7 days) and Day 60 (± 7 days):

1. Vital Signs (temperature/pulse/supine blood pressure)
2. Serum chemistry
3. Hematology
4. Serum hormones
5. Blood samples for PK assessment
6. Record the usage of any concomitant medications and treatments
7. Assess hot flashes
8. Caprini Venothromboembolism Risk Assessment
    a. If there is an increase in risk from baseline (Day 1) in this assessment, appropriate preventative actions should be taken. The subject should be consulted about the risk for VTE and actions that would increase their risk further such as long periods of inactivity and surgery. Prophylactic anticoagulation therapy should be considered.
    b. If a subject has a prolonged PTT-LA or is deficient for Protein C and/or Protein S, but remains eligible for participation in the study, prophylactic anticoagulation therapy should be considered.
    c. For Protein C or Protein S deficiency, consideration should be given to dietary changes to increase Vitamin K intake.
9. Adverse events
10. Collect capsule bottles and perform capsule accountability/compliance assessment
11. Dispense study drug

4.6 Day 84 Visit (End of Study Visit)

Subjects should fast for at least 8 hours (overnight) prior to this visit. The following assessments will be conducted on Day 84 (± 7 days):

1. Vital Signs (temperature/pulse/supine blood pressure)
2. Physical examination (including weight)
3. Serum chemistry (Appendix 9.1) (central laboratory)
4. Hematology (Appendix 9.1) (central laboratory)
5. Serum hormones (Appendix 9.1) (central laboratory)
6. Blood samples for pharmacokinetic assessment
7. Bone turnover markers

8. Record the usage of any concomitant medications and treatments
9. Adverse events
10. Assess hot flashes (see Appendix 9.3)
11. Caprini Venothromboembolism Risk Assessment (Appendix D)

- It is important to note that it is expected that the patient population included in this study will have some risk of VTE due to age, presence of a malignancy, and BMI. The purpose of this assessment is a change in risk from baseline. Also, while the questionnaire is designed for the patient to complete, in this study, the Caprini VTE risk assessment should be filled out by study personnel with information from the patient and patient's chart.

  If there is an increase in risk from baseline (Day 1) in this assessment, appropriate preventative actions should be taken. The subject should be consulted about the risk for VTE and actions that would increase their risk further such as long periods of inactivity and surgery. Prophylactic anticoagulation therapy should be considered.

12. Collect capsule bottles and perform capsule accountability/compliance assessment

4.7 Follow-up Visit

The following assessments will be performed 30 days (± 7 days) after the End of Study visit (Day 84):

1. Vital Signs (temperature/pulse/supine blood pressure)
2. Physical examination (including height and weight)
3. Adverse events

4.8 eDiary Assessment

Subjects will be provided with an Electronic Diary (eDiary) at the Screening Visit to collect data for hot flashes during screening. If enrolled, they will continue to use the eDiary to collect hot flash information for each hot flash through Day 84/Early Termination. All subjects should be instructed to return the eDiary to the study site. Subjects must be educated to record the hot flash assessments in the eDiary in real time.

The table in Appendix 9.1 outlines the timing of assessments to be recorded throughout the study in the eDiary. Appendix 9.3 further elaborates on the hot flash assessment and provides a list of the three severity levels of hot flashes along with a description of the symptoms associated with each level.

5. ANALYSIS POPULATIONS

The subjects will be over the age of 18 and will have moderate to severe hot flashes, defined as a minimum of 4 moderate to severe hot flashes per day or 12 per week at baseline. The 4populations are as follows:

Intent-to-Treat (ITT) population:	All enrolled subjects who have taken at least 1 dose of study drug.
Efficacy Evaluable (EE) population:	All subjects in the ITT population who have analyzable baseline data and 12 weeks of hot flash data. Subjects must also be at least 80% compliant with taking study drug with no major protocol deviations, including no major inclusion/exclusion deviations.

Efficacy Evaluable (EEIA) population for IA:	All subjects in the ITT population who have analyzable baseline data and 6 weeks of hot flash data. Subjects must also be at least 80% compliant with taking study drug with no major protocol deviations, including no major inclusion/exclusion deviations.
Safety population:	All subjects who receive at least 1 dose of study drug. The safety analyses will be conducted on this population. This population is the same as the ITT population.

6. STATISTICAL METHODOLOGY

6.1 Statistical and Analytical Issues

6.1.1 Statistical Methods

For quantitative variables, descriptive statistics will include the number of subjects (n), mean, standard deviation (SD), median, minimum and maximum. Categorical variables will be summarized using the number (n) and percentage (%) of subjects for each category. Unless otherwise specified, the denominator for all percentages will be the number of subjects in the respective treatment group who have data for the respective time point in the applied analysis population. Percentage change from baseline will be calculated using the non-missing (see section 6.1.2) data of both baseline and the respective post-baseline value.

The primary efficacy analyses will only include efficacy data that were captured in the eDiary for hot flashes which were treated with study drug (VERU-944 or placebo). All baseline measurements for efficacy will be collected prior to dosing in the treatment period. By default the frequency and severity of hot flashes at baseline will be determined from the second week of the first two weeks of hot flash eDiary that were recorded; If a patient qualified for the study entry by number of hot flashes during the second week of screening, then data from that week will be used as their baseline. However, if a patient did not qualify for study entry based on the number of hot flashes during that second week then data from the screening week that was used to qualify the patient will be used as their baseline instead. Percentage change from baseline is defined as the post-baseline value of interest minus the baseline value then divided by the baseline value and finally converted to a percentage by multiplying by 100.

A 1-sided alpha level equal to 0.10 will be used to determine statistical significance for the primary and secondary efficacy objectives related to hot flash frequency and severity.

For the treatment period, baseline for the safety assessments will be defined as the Day 1 assessment, or if missing, the last assessment prior to dosing.

All data processing, summarization, and analyses will be performed using SAS® Version 9.4 or higher. Specifications for table, graph, and data listing formats can be found in the tables, listings, and figures (TLF) specifications. Adverse events will be classified using the Medical Dictionary for Regulatory Activities (MedDRA) dictionary, Version 18.1 or higher. All concomitant medications will be coded using the World Health Organization Drug Dictionary Enhanced (WHODDE), version Sep2015 or later.

6.1.2 Handling of Dropouts and Missing Data

Subjects must have entered hot flash data (frequency and severity) for 4 or more days in a given week for their data to be considered non-missing. No other imputation procedures will be used to impute missing data. If a subject enters that there are no hot flashes to report on a given day, then this will be considered a data point equal to zero. If a subject does not report any hot flash information across consecutive days,

the data management team will be alerted to query the situation. From a data perspective, if a subject fails to report hot flash information on a given day, then that day will be recorded as missing.

For determining whether adverse events (AEs) are treatment-emergent when the start date of the AE is a partial date, refer to Section 6.4.1.

For determining whether a medication is a prior medication or concomitant medication, refer to Section 6.2.5.

6.1.3 Pooling of Investigative Sites

All site data will be combined and analyzed together.

6.1.4 Determination of Sample Size

The proposed trial sample size, n=30 per arm, total N=90 in Stage 1, is based on at least a 40% reduction in the mean weekly frequency of hot flashes in a treated arm versus placebo from baseline to Week 4 (Day 28) [the planned primary endpoint at the time of the original study design]. Other assumptions to derive the sample size include a 1-sided α=0.10, a β=0.10 (i.e., power=0.9), and equal variances in the placebo and active arms. The variance was set to equal the mean of the active arm. The same assumptions hold for Stage 2, with a proposed sample size of n=30 per arm, N=60 in Stage 2.

Note, all subjects in screening at the time the 90th subject is randomized into Stage 1 will qualify to be randomized into Stage 1 up to a maximum of 120 subjects (40 per arm). In Stage 2, all subjects in screening at the time the 60th subject is randomized into Stage 2 will qualify to be randomized into the study up to a maximum of 80 subjects (40 per arm). The total enrollment, if both stages are completed, will be a minimum of N=150.

6.2 Subject Characteristics

6.2.1 Subject Disposition

The number of subjects who have been screened, screen failed, and randomized will be summarized by treatment group and overall. The number of subjects who discontinue study treatment or study follow-up will be summarized by treatment group, along with the reasons for discontinuation. The number of subjects in each analysis population will be summarized by treatment group.

Listings of the subject disposition and enrollment eligibility, analysis populations, and treatment assignment will be provided.

6.2.2 Protocol Deviations

Protocol deviations will be presented in a data listing, by treatment group for the Safety population. Protocol deviations will be captured through data review and site monitoring.

6.2.3 Background and Demographic Characteristics

Demographic information (e.g., age, sex, race, height, weight, body mass index [BMI], and ethnicity) will be summarized for the ITT population, by treatment group and overall with summary statistics for continuous/categorical variables.

6.2.4 Treatment Exposure and Compliance

6.2.4.1 eDiary Exposure

Exposure will be collected in the eDiary, where for Days 1-7 subjects will be asked how many pills they take each day. A listing will be provided, for the ITT population, to report treatment exposure for Days 1-7, where subjects are instructed to take 1 or more pills a day.

6.2.4.2 Overall Exposure and Compliance

Duration of exposure and compliance will be summarized for the ITT population by treatment group and overall for the active treatment groups using descriptive statistics.

Duration of exposure will be summarized as time in days from first treatment to last dose, calculated as last dose date minus first dose date plus 1 day.

Compliance will be summarized using data from the electronic case report form (eCRF). A percent compliance value (shown derived below) will be calculated and listed for each subject.

$$\frac{Number\ of\ pills\ taken}{Expected\ number\ of\ pills\ taken} * 100$$

6.2.5 Prior and Concomitant Medications and Therapies

The start and stop of each medication will be recorded by calendar date. “Prior medications” are defined as medications that have been discontinued prior to the date of first dose of study drug. “Concomitant medications” are defined as medications taken any day from the first dose of study drug until the last dose of study drug, regardless of whether the medication was initiated prior to the first dose of study drug.

The counts and percentages of subjects who took prior and concomitant medications will be summarized on the safety population, by treatment group and overall. Prior and concomitant medications will also be presented in subject data listings.

For partial start or stop dates, the following considerations and imputations will be performed.

- For a partial or missing start date where the stop date is recorded in full and is before the date of initial study treatment, it will be considered a prior medication. If the stop date is after the date of initial study treatment, it will be considered a concomitant medication.
- For a partial or missing stop date where the start date is recorded in full and the start date is after the initial study treatment date, it will be considered a concomitant medication.
- For instances where a medication has a partial stop date and a partial start date, the month and year of the stop date will be used to assess whether the medication is prior or concomitant. For instance, if month and year exist for the stop date, this data will be compared with the date of initial study treatment and if the month or year show evidence of being after this date or on the same date (same month/year), the medication will be considered concomitant.
- For medications where both the start and stop dates are missing, these medications will be considered concomitant.

The prior and concomitant medications taken during the study will be summarized using number of subjects (n) and percentage (%) in the safety population. All prior and concomitant medications will be coded using the WHODDE, version September 2015 or later, and then further coded against the

Anatomic Therapeutic Chemical (ATC) classification. Prior and concomitant medication data will be presented by ATC Level 2 and preferred term and will be presented by treatment group.

6.2.6 Medical Histories

General medical history and surgical history data will be captured during the screening period, both at screening and Day 1. A listing of medical and surgical history data will be provided on the Safety population.

6.3 Efficacy Analyses

6.3.1 Analyses of all hot flash frequency and severity endpoints both Primary and Secondary

The following conventions apply to all analyses of hot flash frequency and severity data:

Population: The EE population (efficacy evaluable) population will be the population of interest in the primary and all secondary endpoint analyses. The EEIA (efficacy evaluable interim analysis) population will be the population of interest for the analysis conducted at Week 6. The ITT population will be used in the sensitivity analyses to verify that the findings remain consistent.

Data Handling

Baseline week will be determined from the second week of the first two weeks of hot flash eDiary that were recorded; if a patient qualified for study entry by number of hot flashes during the second week of screening, then data from that week will be used as their baseline. However, if a patient did not qualify for study entry based on the number of hot flashes during that second week then data from the screening week that was used to qualify the patient will be used as their baseline instead. A week's data collection will be considered the 6 days prior to plus the day of week's data. For example, Week 6 data will include the data collected for the 7 days from Day 36 through Day 42, inclusive. For a subject's data to be considered non-missing for a given week the subject must have recorded at least 4 days of data during the respective week. For subjects who report at least 4 days of data during a week, the weekly frequency of moderate and severe hot flashes will be the sum across those days of the number of moderate to severe hot flashes. The severity level reported for each of the reported hot flashes will be calculated as follows. Baseline severity level will be the number of severe hot flashes reported times 3 + the number of moderate hot flashes reported times 2 then divided by the number of moderate and severe hot flashes reported. Post-baseline weeks will have their severity level calculated as above but with the addition of reported mild hot flashes so the post-baseline severity for a given week level will be the number of severe hot flashes reported times 3 + the number of moderate hot flashes reported times 2 + the number of mild hot flashes reported times 1 then divided by the number of mild, moderate and severe hot flashes reported. The percentage change from baseline of the frequency of moderate and severe hot flashes will be the number of moderate to severe hot flashes reported for the given week minus the number reported at baseline then divided by the number reported at baseline and converted to a percentage by multiplying by 100. The percentage change from baseline of the severity of moderate and severe hot flashes will be the calculated severity score reported for the given week minus the severity score at baseline then divided by the severity level reported at baseline and converted to a percentage by multiplying by 100.

Adjustment for Multiple Comparisons

The Hochberg procedure will be applied to control for multiplicity for each of the primary and secondary hot flash related endpoints. Each endpoint will have 4 analyses associated with them to account for the two active dose arms (10 mg and 50 mg) each compared to baseline by two methods each (MMRM test for slope and MMRM test for change from baseline at the specified week). These may be referred to as sets of analyses. Hochberg is a step-up procedure based on univariate p-values. This procedure begins

with the least significant p-value and then examines the other p-values in sequential order, making alpha level adjustments along the way, until a significant p-value is observed. Once achieved, all following lower p-values in the sequence would be considered significant. For example, if one has 4 hypotheses the following will be performed:

- Step 1: Order the p-values from largest (least significant) to smallest (most significant)
- Step 2: Using the pre-specified overall alpha level, α_0=0.10, one-sided, compare the largest p-value, P_1, to the overall alpha level. If P_1<0.10 then all (null) hypotheses are rejected; otherwise go to step 3.
- Step 3: Compare the second largest p-value, P_2, to α_1=0.10/2=0.05. If P_2<0.05 then this and all further hypotheses are rejected; otherwise repeat for the third, P_3, and fourth, P_4, largest p-values compared with α_2=0.10/3=0.033 and α_3=0.10/4=0.025 as necessary to either reject all remaining hypotheses or conclude no significant difference exists.

For this study, during Stage 1, Hochberg's procedure will rank the p-values from highest to lowest for each statistical test within each set of analyses: 10 mg versus placebo slope difference (through Week 6), 50 mg versus placebo slope difference (through Week 6), 10 mg versus placebo mean difference at week 6, 50 mg versus placebo mean difference at Week 6. Adjustment for the single comparison of 100 mg versus placebo in Stage 2 will be necessary only for the two MMRM modes, means and slopes.

Overall, for the multiple tests of primary and secondary endpoint analyses alpha control will be achieved by a gate-keeping method such that the endpoints are ordered and the preceding endpoint must be shown to be significant in order for endpoints down the list to be eligible for significance.

The order of the secondary endpoint assessments will be:

1. Percentage change in severity of vasomotor symptoms from baseline to Weeks 6 (Day 42)
2. Percentage change in frequency of moderate to severe vasomotor symptoms from baseline to Week 8 (Day 56)
3. Percentage change in frequency of moderate to severe vasomotor symptoms from baseline to Week 10 (Day 70)
4. Percentage change in frequency of moderate to severe vasomotor symptoms from baseline to Week 12 (Day 84)
5. Percentage change in severity of vasomotor symptoms from baseline to Week 8 (Day 56)
6. Percentage change in severity of vasomotor symptoms from baseline to Week 10 (Day 70)
7. Percentage change in severity of vasomotor symptoms from baseline to Week 12 (Day 84)
8. Percentage change in frequency of moderate to severe vasomotor symptoms from baseline to Week 4 (Day 28)
9. Percentage change in severity of vasomotor symptoms from baseline to Week 4 (Day 28)
10. Percentage change in bone turnover marker concentrations at Day 84 compared with baseline

<u>Common analytical methods for all hot flash related analyses</u>

The MMRM analyses shown in the following section with SAS code written specifically to address the primary analysis of difference in percentage change of the frequency of hot flashes at Week 6 between each of the active arms and placebo are the **SAME** for all hot flash related analyses, both frequency and severity, at all weeks of interest. Simple changes to the data set to subset to the appropriate weeks of data and to the LSESTIMATE statements are all that is necessary to construct the appropriate tests. MMRM models that test for MEAN difference between active arm(s) at Week 6 and placebo and separately for a

difference in SLOPE between active arm(s) at Week 6 and placebo. There is no plan to compare the active arms to each other.

If the MMRM models does not converge or have poor fit, T-tests or non-parametric tests, if the data are deemed non-normally distributed, will compare percentage changes over time between (T-test or Wilcoxon rank sum test) each active arm and the placebo arm, respectively.

6.3.1.1 Primary Efficacy Endpoint Analysis (Hot Flash Frequency at Week 6)

STAGE 1:

The primary frequency efficacy endpoint is percentage change in frequency of moderate and severe hot flashes (vasomotor symptoms) from the baseline week to Week 6 (Day 42). The statistical tests for significant differences with regards to the percentage change between placebo and each of the active arms, 10 mg 50mg, and perhaps 100 mg comprise the main tests of interest.

The SAS code for performing the MMRM analyses to test for a significant difference in the MEAN percentage change in frequency of moderate to severe hot flashes between each of the dose arms and placebo evaluated at Week 6 (primary analysis week) is shown below. Note that this analysis will generate two of the four tests that are part of the primary analysis to which the Hochberg procedure will apply.

```
PROC MIXED DATA= XXX;
CLASS subjid tx week;
MODEL cfb = tx week_2 tx*week_2;
REPEATED week / type=AR(1) sub=subjid(tx);
LSESTIMATE tx*week '10MG VS Placebo at Week 6' [1, 10 6]  [-1, P 6];
LSESTIMATE tx*week '50MG VS Placebo at Week 6' [1, 50 6] [-1, P 6];
WHERE week le 6;
RUN;
```

The variables used in the above model will be defined as follows: tx=treatment; week (1, 2… 6); cbf=change from baseline value, week_2=week (week will be treated as a categorical variable in the class statement and week_2 will be treated as a continuous variable in the model… effectively the values of these variables will be an exact match). By creating week_2 in a preceding data step, this allows for LESTIMATES to be computed at a specific WEEK (treated as categorical in the CLASS statement) but allows week to be treated as a continuous variable, WEEK_2 in the model.

This model will be run 3 separate times using 3 separate covariance structures to determine best fit. The same model will be run, with the following specified change on the REPEATED line: (1) type=AR(1); (2) type=CS; and (3) type=UN. If a model does not converge, then disregard that model. If more than 1 model remains, then refer to the Akaike's Information Criteria (AIC) score to determine model fit. The best fit model will be chosen as having the AIC value closest to zero. The model type chosen will be referenced in the footnote for each model output.

Additional prespecified analysis of the data will be conducted regardless of whether the MMRM model(s) converge, and will be considered the definitive tests for significance in the event the MMRM model(s) do not converge:

- T-tests will be generated to compare the percentage change from baseline to each of Week 6 in the placebo arm to each of the 10 mg and 50 mg arms. If the data are determined to be non-normally distributed then the Wilcoxon rank sum test will be the definitive test of significance.

- Wilcoxon rank sum tests will be generated to compare the percentage change from baseline to Week 6 in the placebo arm to each of the 10 mg and 50 mg arms.

The SAS code for performing the MMRM analyses to test for a significant difference in the SLOPES of the percentage change in frequency of moderate to severe hot flashes between each of the dose arms and placebo by Week 6 (primary analysis week) is shown below. Note that these analyses will generate the remaining two of the four tests that are part of the primary analysis to which the Hochberg procedure will apply.

The where statement in PROC MIXED has been used to simplify the interpretation of the output so that one model will be run for the 10 mg versus placebo slope analysis and a separate one for the 50 mg versus placebo slope analysis through Week 6. Variable have the same meaning as defined in the above analysis.

```
PROC MIXED DATA=XXX;
  CLASS subjid tx;
  MODEL cfb= tx week tx*week / s;
  RANDOM intercept week/ type=AR(1) sub=subjid g;
 WHERE tx in (10, P) and week le 6; /* repeat for 'tx in (50, P)' */
RUN;
```

This model allows for both the intercept and the slope to differ randomly for each subject but the overall p- value for the interaction term is interpreted as a test of the slope between the stated active arm and placebo. Note, if the p-value is significant, the estimated coefficients must be in the appropriate direction to conclude the result is in favor of the active dose arm.

STAGE 2:

The analyses conducted in Stage 2 will be the exact same as in Stage 1, except the 10 mg and 50 mg dose of study drug will be replaced with a 100 mg dose of study drug. A separate and distinct placebo group will be used for Stage 2. The Hochberg procedure will have to address only two multiple comparisons, i.e., the MMRM slope models and the MMRM means models as there is now only one dose level comparison to placebo. The two alpha levels that may be needed to assess significance will be only α_0=0.10 and possibly $_{\alpha 1}$=0.05. Here is the appropriate code for the LSESTIMATE statement.

```
LSESTIMATE tx*week '100MG VS Placebo at Week 6' [1,100 6] [-1, P 6];
```

Also, as a reminder, the SAS code shown for the primary analysis can be used in this second stage as well for its primary analysis of hot flash frequency as well as for all secondary hot flash related endpoints.

6.3.1.1.1 Additional frequency analyses

The t-test or the or the Wilcoxon rank-sum test, if the data are determined to be non-normally distributed, will be used for between group treatment comparisons comparing the percentage change from baseline in the placebo arm to the percentage change from baseline in each treatment group (10 mg and 50 mg) at Day 42. The paired t-test or the Wilcoxon signed-rank test, if the data are determined to be non-normally

distributed, will be used for comparing baseline and Day 42 within each treatment group. The comparisons are as follows:

- Stage 1: Change from Baseline week to Week 6 between placebo and each of the 10 mg and 50 mg arms
- Stage 2: Change from Baseline week to Week 6 between placebo and the 100 mg
- Stage 1: Baseline week versus Week 6 for each treatment arm (placebo, 10 mg, and 50 mg)
- Stage 2: Baseline week versus Week 6 for each treatment arm (placebo and 100 mg)

6.3.1.1.2 Additional frequency analyses – sensitivity analyses among the ITT population

Sensitivity analyses will also be conducted on the ITT population as described above.

6.3.1.1.1 Additional frequency analyses – sensitivity analyses including all severity levels

Sensitivity analyses on the percentage change in frequency of hot flashes from baseline to Week 6 will be conducted where a subject's frequency of hot flashes for a given week will include any hot flash designated as mild, moderate, or severe. This will be conducted among subjects in the EEIA population and among those in the ITT population.

6.3.2 Secondary Efficacy Variables

As noted previously, the same MMRM models - one for means and one for slopes - specified for the primary analysis will be used for each of the following hot flash related secondary endpoints including both frequency of hot flashes and severity of hot flashes.

6.3.2.1 Percentage Change in severity of moderate and severe hot flashes from baseline to week 6

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Please see the data handling description (Section 6.3.1) for the specifics of severity calculation Repeat for Stage 2 as necessary.

6.3.2.2 Percentage Change in frequency of moderate and severe hot flashes from baseline to week 8

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Repeat for Stage 2 as necessary.

6.3.2.3 Percentage Change in frequency of moderate and severe hot flashes from baseline to week 10

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Repeat for Stage 2 as necessary.

6.3.2.4 Percentage Change in frequency of moderate and severe hot flashes from baseline to week 12

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Repeat for Stage 2 as necessary.

6.3.2.5 Percentage Change in severity of moderate and severe hot flashes from baseline to week 8

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Please see the data handling description (Section 6.3.1) for the specifics of severity calculation Repeat for Stage 2 as necessary.

6.3.2.6 Percentage Change in severity of moderate and severe hot flashes from baseline to week 10

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Please see the data handling description (Section 6.3.1) for the specifics of severity calculation Repeat for Stage 2 as necessary.

6.3.2.7 Percentage Change in severity of moderate and severe hot flashes from baseline to week 12

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Please see the data handling description (Section 6.3.1) for the specifics of severity calculation Repeat for Stage 2 as necessary.

6.3.2.8 Percentage Change in frequency of moderate and severe hot flashes from baseline to week 4

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Repeat for Stage 2 as necessary.

6.3.2.9 Percentage Change in severity of moderate and severe hot flashes from baseline to week 4

Analyzed with the same methods and considerations as for the analysis of frequency at week 6 (primary endpoint). Please see the data handling description (Section 6.3.1) for the specifics of severity calculation Repeat for Stage 2 as necessary.

6.3.2.10 Bone Marker Turnovers

STAGE 1:

The secondary efficacy endpoint for bone marker turnovers is percentage change in bone turnover marker concentrations at Day 84 compared with baseline (Day 1, pre-dose). The bone markers to be assessed will include: serum C-telopeptide and bone specific alkaline phosphatase. For each marker, the mean, SD, median, minimum, and maximum will be summarized at each time point by treatment. The percentage change from baseline to Day 84 will also be summarized.

The t-test or the Wilcoxon rank-sum test, if the data are determined to be non-normally distributed, will be used for between group treatment comparisons comparing the percentage change from baseline in the placebo arm to the percentage change from baseline in each treatment group (10 mg and 50 mg) at Day 84. The paired T-test or the Wilcoxon signed-rank test, if the data are determined to be non-normally distributed, will be used for comparing baseline and Day 84 within each treatment group. Analyses will be conducted on the ITT population.

STAGE 2:

The same methods and analyses described directly above for Stage 1, will be repeated in Stage 2 for the placebo and 100 mg dose.

6.3.3 Exploratory Analysis

6.3.3.1 Change in Serum PSA Concentrations

STAGE 1:

A percentage change in serum PSA concentrations comparing baseline (Day 1, pre-dose) and each scheduled assessment (Days 30, 60, and 84) will be reported for each treatment group. The mean, SD, median, minimum, and maximum will be summarized at each time point by treatment. The percentage change from baseline to each scheduled assessment will also be summarized.

The t-test or the or the Wilcoxon rank-sum test, if the data are determined to be non-normally distributed, will be used for between group treatment comparisons comparing the percentage change from baseline in the placebo arm to the percentage change from baseline in each treatment group (10 mg and 50 mg) at Days 30, 60, and 84. The paired t-test or the Wilcoxon signed-rank test, if the data are determined to be non-normally distributed, will be used for comparing baseline and Day 84 within each treatment group. Analyses will be conducted on the ITT population.

.

STAGE 2:

The same methods and analyses described directly above for Stage 1, will be repeated in Stage 2 for the placebo and 100 mg dose.

6.3.3.2 Change in Serum Total Testosterone and Serum Free Testosterone Concentrations

STAGE 1:

A percentage change in serum total testosterone concentrations comparing baseline (Day 1, pre-dose) and each scheduled assessment (Days 30, 60, and 84) will be reported for each treatment group. The mean, SD, median, minimum, and maximum will be summarized at each time point by treatment. The percentage change from baseline to each scheduled assessment will also be summarized.

The t-test or the or the Wilcoxon rank-sum test, if the data are determined to be non-normally distributed, will be used for between group treatment comparisons comparing the percentage change from baseline in the placebo arm to the percentage change from baseline in each treatment group (10 mg and 50 mg) at Days 30, 60, and 84. The paired t-test or the Wilcoxon signed-rank test, if the data are determined to be non-normally distributed, will be used for comparing baseline and Day 84 within each treatment group. Analyses will be conducted on the ITT population.

Similar analyses will be implemented for serum free testosterone concentrations.

STAGE 2:

The same methods and analyses described directly above for Stage 1, will be repeated in Stage 2 for the placebo and 100 mg dose.

6.3.3.3 Change in Serum SHBG Concentrations

STAGE 1:

A percentage change in serum SHBG concentrations comparing baseline (Day 1, pre-dose) and each scheduled assessment (Days 30, 60, and 84) will be reported for each treatment group. The mean, SD, median, minimum, and maximum will be summarized at each time point by treatment. The percentage change from baseline to each scheduled assessment will also be summarized.

The t-test or the or the Wilcoxon rank-sum test, if the data are determined to be non-normally distributed, will be used for between group treatment comparisons comparing the percentage change from baseline in the placebo arm to the percentage change from baseline in each treatment group (10 mg and 50 mg) at Days 30, 60, and 84. The paired t-test or the Wilcoxon signed-rank test, if the data are determined to be non-normally distributed, will be used for comparing baseline and Day 84 within each treatment group. Analyses will be conducted on the ITT population.

STAGE 2:

The same methods and analyses described directly above for Stage 1, will be repeated in Stage 2 for the placebo and 100 mg dose.

6.4 Safety Analysis

6.4.1 Adverse Events

A treatment-emergent adverse event (TEAE) is defined as any event not present before the date of treatment with the first dose of study drug, or any event already present that worsens in severity on or after the date of the first dose of study drug.

The calendar date is to be collected for the start and stop of each AE. In the event that only a partial end date (month/year) is available, and the month/year occurs before Day 1 in the treatment period of the study, the AE will not be considered treatment-emergent. However, if the onset date is a partial date (month/year) and the month/year occurs on or after Day 1 in the treatment period of the study, the following cases will be considered:

- The AE will be considered as treatment-emergent if:
    - The month/year of the onset date is later than the month/year of Day 1 in the treatment period.
    - The month/year of the onset date is equal to the month/year of Day 1 in the treatment period of study, and the AE is ongoing at the end of study or the end date and time are present in full or are partially provided such that they place the stop of the AE after the start of exposure to study treatment on Day 1 in the treatment period of the study.

  Otherwise, the AE will be considered to be not treatment-emergent and will not be summarized; however, it will be included in the listing.

AEs will be summarized based on subjects having at least 1 occurrence of the AE at the level of the MedDRA preferred term, with the subject as the unit of analysis. If a subject has more than 1 AE of the same preferred term and there are different grades of severity, only the highest grade will be represented in the summary of severity. If the severity assessment is missing, the severity will be reported as 'unknown', and the known severity will be reported over the unknown severity. Each AE will be assessed for relatedness to the study treatment. If a subject has more than 1 AE of the same preferred term and there are different levels of relatedness to study treatment, only the highest level of relatedness for each will be represented in the summary of relatedness. Missing relationships will be included as related.

The following summaries of TEAEs by system organ class (SOC) and preferred term (PT) for the safety population will be provided. Summaries will be displayed by number of subjects with the event, percentage of subjects with the event and number of events reported by treatment group and overall for the following groups:

- TEAEs
- Serious adverse events (SAEs)

Similarly, summaries will also be displayed by number and percentage of subjects with the event reported by treatment group and overall for the following groups:

- TEAEs by maximum severity and incidence
- TEAEs by strongest relationship to study treatment number and incidence

Any adverse events, serious adverse events, actions taken due to adverse events, discontinuation from the study due to adverse events, adverse event outcomes, and adverse events resulting in study discontinuation will be listed.

Summary tables and listings will be presented by treatment group for the safety population.

6.4.2 Physical Examination

Physical exam data will be listed for the safety population. The listing will include exam date, body system, and result.

6.4.3 Vital Signs

The following vital sign test results and changes from baseline will be descriptively summarized by study visit for the safety population. Summaries will be displayed by treatment group and overall.

- Temperature (ºF)
- Systolic and diastolic blood pressure (mmHg)
- Heart Rate (beats/minute)

A listing of the vital signs and weight for each study visit will be presented.

6.4.4 Electrocardiogram

An electrocardiogram (ECG) is performed at screening. A listing of ECG date, parameters and interpretation will be provided for each subject in the ITT population.

6.4.5 Laboratory Parameters

Laboratory data (Section 9.2) will be displayed by treatment group and overall for the safety population. Continuous data for laboratory tests of serum chemistry, hematology, urinalysis, and serum hormone including change from baseline will be descriptively summarized at each visit, and urinalysis categorical data will be summarized by visit. The number of subjects who have values outside the normal range will also be displayed descriptively at each time point.

A listing of serum chemistry, hematology, urinalysis, and serum hormone test results will be presented separately by visit. The numeric values will be flagged/indicated with LN = Low Normal, LP = Low Panic, N = Normal, HN = High Normal, HP = High Panic in the listing.

6.5 PK Analysis

Samples of venous blood for the determination of trough plasma concentrations of VERU-944 will be collected on Days 1 (prior to dose), 30, 60, and 84. VERU-944 trough concentrations will be determined in each subject and summarized by mean, median, standard deviation, and coefficient of variation, maximum, and minimum for each treatment arm.

6.6 Caprini VTE Risk Assessment

The VTE risk assessment data will be summarized for the safety population at each visit, by treatment group and overall.

6.7 Interim Analysis

During Stage 1, after the last subject has completed the Day 42 visit, an unblinded interim analysis (IA) of the Week 6 (Day 42) hot flash primary endpoint (frequency) and first secondary endpoint (severity) of VERU-944 will be conducted. A blinded analysis of the safety will also be conducted. Both the Sponsor and CRO study team will remain blinded to patient level data until completion of the study stage. An

unblinded statistical support group, independent of the study team (within the CRO), along with a representative of Veru Inc., will review the IA results and determine steps forward. It is of note that due to the material nature of the IA results of this study in the business of Veru Inc., a publicly traded company, these results may be required to be released in a press release prior to the completion of Stage 1 of this study.

The analyses to be included in the IA will be the analyses of hot flash frequency at Week 6 (Section 6.3.1.1) and the hot flash severity at Week 6. These analyses will include the summary statistics and MMRM model for both the ITT and EEIA populations. The sensitivity analyses associated with these sections will not be conducted during the IA.

In addition to the efficacy analyses, a blinded safety report will be reviewed to assess drug safety. All IA analyses will be conducted on the EEIA population.

If an effective dose of VERU-944 is achieved in Stage 1, as defined as a statistically significant result), the Sponsor reserves the right to not initiate Stage 2 of the protocol. However, if an effective dose is not achieved in the Stage 1 IA, then the study will commence with Stage 2.

Regardless of IA outcome, unless safety issues are observed, the Stage 1 section of the study will continue to completion at 12 Weeks with Follow-up.

Assuming Stage 2 is implemented, an additional replicate IA will be conducted after the last subject in Stage 2 has completed the Day 42 visit. This IA will follow the same procedures as done for Stage 1.

6.8 Data Monitoring Committee

There will be no formal Data Monitoring Committee for this study. Safety data, including all SAEs, will be reviewed on an ongoing basis by a Safety Review Team comprised of the Medical Monitor, and representatives of the Sponsor. This Safety Review Team will meet at least monthly. Additional ad hoc meetings will be scheduled if required to evaluate the safety and/or thromboembolic events further

6.9 Changes to Methods Planned in the Protocol

At the time the SAP was finalized, there were no changes to methods planned in the protocol. Any other changes from the SAP, including any ad-hoc analysis, will be detailed in the clinical study report.

7. TABLES, LISTINGS AND FIGURES

The standard operating procedures of Worldwide Clinical Trials will be followed in the creation and quality control of all tables, listings and figures.

Format of Output

1. Unless otherwise specified, all computer-generated output should be produced in landscape mode. Required margins: 1.25 inches on top and bottom and 1.00 inch on the left and right; required font: Courier New; and required font size: 9. All output should have the following header at the top of the page:

   VERU Inc. Page n of N
   Protocol Number: V72203

All output should have date (date output was generated) and page number. Tables/listings/figures should be internally paginated in relation to total length (i.e., page number should appear sequentially as page n of N, where N is the total number of pages in the table).

2. Each output should be identified by a numeral, and the output designation (e.g., Table 1) should be listed on the same line, before the title. A decimal system (x.y and x.y.z) should be used to identify tables and listings with related contents. The title is centered in initial capital characters.

Table No.
Table Title
Study Population

The study population should be identified immediately following the title.

3. Column headings should be in initial upper-case characters.
4. For numeric variables, include "unit" in column or row headings when appropriate.
5. Footnotes should be single spaced, but separated by at least a double space from the bottom line of the table. The notes are aligned vertically by the left vertical border of the table.
6. If the categories are not ordered (e.g., race), then only those categories for which there is at least 1 subject represented in 1 or more groups should be included.
7. An Unknown or Missing category should be added to any parameter for which information is not available for 1 or more subjects.
8. Listings should be sorted by subject number and study visit.
9. In a listing, display the subject number only once for the subject with multiple records. If a subject's records run across multiple pages, display the subject number once for every page.

Data format

1. Unless otherwise specified, the estimated mean and median for a set of values should be printed out to 1 more significant digit than the individual units of measurement and the standard deviation for a set of values should be printed out to 2 more significant digits than the individual units of measurement. The minimum and maximum should report the same significant digits as the original values. For example, for age (with raw data in whole years):

n	XX
Mean (STD)	XX.X (XX.XX)
Median	XX.X
Min, Max	XX, XX

2. Unless otherwise specified, data in columns of a table should be formatted as follows:

- Alphanumeric values are left-justified.
- Whole numbers (e.g., counts) are right-justified.
- Numbers containing fractional portions are decimal aligned.

3. Unless otherwise specified, percentage values should be printed with 1 digit to the right of the decimal point (e.g., 12.8%, 5.4%). Less-than-signs "<0.1%" should be printed when values are >0.0 and <0.1% (not 0.0%).
4. Unless otherwise specified, missing data should be represented on subject listings as either a hyphen ("-") with a corresponding footnote (" - = unknown or not evaluated"), or as "N/A," with the footnote "N/A = not applicable," whichever is appropriate.
5. Dates should be printed in SAS® DATE9.format ("DDMMMYYYY": 01JUL2000). Missing portions of dates should be represented on subject listings as dashes (--JUL2000). Dates that are missing because they are not applicable for the subject are output as "N/A", unless otherwise specified.

Unless otherwise specified, time should be printed in SAS® TIME5.format ("HH:MM": 17:30). Missing portions of time should be represented on subject listings as dashes (--:30). Times that are missing because they are not applicable for the subject are output as "N/A", unless otherwise specified.

7.1 Stage 1 versus Stage 2 reporting

TLFs reported for Stage 1 will be as shown in the TLF specification document, reporting only for Stage 1 of the study. Assuming Stage 2 is implemented, all data for Stage 1 and 2 will be combined for safety and demographic tables and listings. However, the placebo groups for Stage 1 and Stage 2 will not be combined and will remain treated as 2 separate treatment groups. See TLF specifications for more detail on reporting.

8. REFERENCES

1. Engstrom, C.A. (2008). Hot flashes in prostate cancer: state of the science. American journal of men's health *2*, 122-132.

2. Frisk J. (2010). Managing hot flushes in men after prostate cancer--a systematic review. Maturitas *65*, 15-22.

3. Holzbeierlein JM, Castle E, Thrasher JB. 2004. Complications of Androgen Deprivation Therapy: Prevention and Treatment. *Oncology (Williston Park).* Mar; 18(3):303-309, discussion 310, 315, 319-321.

4. Ulloa E.W., Salup, R., Patterson, S.G., and Jacobsen, P.B. (2009). Relationship between hot flashes and distress in men receiving androgen deprivation therapy for prostate cancer. Psycho-oncology *18*, 598-605.

9. APPENDICES

9.1 Schedule of Study Evaluations

Day	Screen[a]	Pre-Randomization during screening period	1	14	30	60	84 End of study	Follow-up
Informed Consent	X							
Medical History	X		X					
Assessment of Eligibility *(For Screening visit capture history of hot flashes frequency and severity*	X	X	X					
Physical Exam	X		X				X	X
Vital signs	X		X	X	X	X	X	X
12-lead ECG (single)	X							
Clinical Laboratory Tests								
Hematology	X		X		X	X	X	
Urinalysis	X		X					
Serum Chemistry	X		X	X	X	X	X	
Serum Hormones	X		X		X	X	X	
Blood sample for pharmacokinetic assessment[c]			X		X	X	X	
Bone turnover markers			X				X	
Thromboembolic risk assessment	X							
Bilateral Doppler ultrasound of the lower extremities		X						
Assess hot flashes (electronic diaries)		X	X		X	X	X	
First dose			X					
Dispense study drug			X		X	X		
Collect capsule bottles (even if empty) and perform accountability/compliance assessment [b]					X	X	X	
Assessment of conmeds	X		X	X	X	X	X	
Assessment of AEs			X	X	X	X	X	X
Caprini VTE risk assessment [d]	X		X		X	X	X	

a. Screening evaluations to be conducted within 28 days prior to Day 1.
b. Collect capsule bottles (even if empty) and perform accountability/compliance assessment throughout the participation of the subject in the study.
c. The pharmacokinetic blood samples should be done prior to receiving the dose of study drug for that day.
d. It is important to note that it is expected that the patient population included in this study will have some risk of VTE due to age, presence of a malignancy, and BMI. The purpose of this assessment is a change in risk from baseline. Also, while the questionnaire is designed for the patient to complete, in this study, the Caprini VTE risk assessment should be filled out by study personnel with information from the patient and patient's chart.
e. The same schedule of study evaluations will be used in Stage 1 and Stage 2 of this protocol.

9.2 Clinical Laboratory Tests (central laboratory)

Hematology:	***Urinalysis:***
Hemoglobin	pH
Hematocrit	Specific Gravity
Red Blood Cell Count	Protein
White Blood Cell Count	Glucose
White Blood Cell Differential	Leucocytes
Platelet Count	Nitrates
Reticulocyte Count	Ketones
	Blood
"Serum Hormones"	Microscopic Examination (only if urinalysis results are abnormal)
LH	
FSH	
Total testosterone (LC-MS/MS)	***Serum Chemistry:***
Free testosterone (LC-MS/MS)	Sodium
Sex Hormone Binding Globulin	Potassium
Serum PSA	Chloride
	Bicarbonate
	BUN
Bone turnover markers	Creatinine
C-telopeptide (CTX)	Calcium
Bone specific alkaline phosphatase	Phosphorus
	Total Protein
	Albumin
Thromboembolic Risk Assessment	Total Bilirubin
Factor V Leiden gene mutation	SGOT (ALT)
Antiphospholipid antibodies for Lupus	SGPT (AST)
Anticoagulant and Anticardiolipin	Alkaline phosphatase
Prothrombin gene mutation	LDH
Protein C and S	GGT
	Glucose

9.3 Hot Flash Assessments

Assessments of hot flashes will be made at baseline (14 contiguous days during screening) and daily from Day 1 (first dose) to the end of the study. Patients will record frequency and severity of hot flashes using an electronic diary (eDiary) to ensure attributable, time stamped data collection in an unsupervised environment. When experiencing a hot flash the subject will indicate on the app that they have had a hot flash and then will assign a severity to the hot flash based upon the following scale:

- MILD: sensation of heat without sweating
- MODERATE: sensation of heat with sweating, able to continue activity
- SEVERE: sensation of heat with sweating, causing cessation of activity

The hot flashes will then be compiled to examine the frequency and severity of hot flashes over the course of the study.