Statistical Analysis Plan (SAP)

Protocol Title:	A Phase IIb multicentre, double-blind, dose-ranging, randomised, placebo-controlled study evaluating safety and efficacy of BGS649 in male obese subjects with hypogonadotropic hypogonadism
Protocol Number:	MBGS205
Protocol Version, Date	Final V6.0, 10 March 2017
ICON ID:	3082/0005
Document Version, Date:	FINAL 2.0, 07 June 2018

Prepared by:

ICON Clinical Research Services

On behalf of:

Mereo BioPharma 2 Limited

DocUUID : 8adea619-3b33-4c9e-abad-54bd682d38a9

SIGNATURE PAGE

REVISION HISTORY

Version/Date	Version name	Section	Changes implemented
1.0 / 28FEB2018	Initial approved version	N/A	N/A
1.1/ 01MAR2018	1.1	8.9	Clarified procedure for future analysis of primary analysis assessments if they are not available at time of the primary analysis.
2.0/ 07JUN2018	2.0	7.1.3.9	Inclusion of "number of minutes worn" criteria for all actigraph derived parameters
		8.4.2	Update of categorical cut-offs for free testosterone, bioavailable testosterone, Oestradiol and Vitamin D.
		8.6.4.1	Inclusion of the high sensitivity lab values for Oestradiol analysis
		8.7.2.1	Added treatment comparison in change from baseline in renal parameters
		8.7.3	Added treatment comparison in change from baseline in systolic and diastolic blood pressure Added new table for blood pressure categorisation
		8.7.6	Site recorded DEXA results will now be summarised in addition to the overread values
		8.8.4	Additional Scatter graphs added for Oestradiol vs PRO Scores and Testosterone/oestradiol vs PRO Scores

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

The following abbreviations will be used within this SAP.

Abbreviation or special term	Explanation
AE	Adverse Event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BFI	Brief Fatigue Inventory
BMI	Body mass Index
BUN	Blood Urea Nitrogen
CDISC	Clinical Data Interchange Standards Consortium
CDS	Clinical Data Services
CRF	Case Report Form
CTx1	C-terminal telopeptide
DBP	Diastolic blood pressure
DEXA	Dual Energy X-ray Absorptiometry
DHT	Dihydrotestosterone
DMC	Data Monitoring Committee
ECG	Electrocardiogram
EOT	End of Treatment
eCRF	Electronic case report form
eGFR	Estimated glomerular filtration rate
EOT	End of Treatment
ET	Early termination
FSH	Follicle Stimulating Hormone
FT4	Free thyroxine;
HbA1c	Glycosylated haemoglobin
HDL	High Density Lipoprotein
HOMA-IR	Homeostatic assessment of insulin resistance;
hs-CRP	High sensitivity C-reactive protein
IIEF	International index of erectile function;
IRT	Interactive Response Technology
LDL	Low Density Lipoprotein
LH	Luteinising hormone
MDRD	Modification of Diet in Renal Disease
MedDRA	Medical Dictionary for Regulatory Activities
mMRC	Modified Medical Research Council
MMRM	Mixed model repeated measure
PGI-S	Patient Global Impression of Status items
PK	pharmacokinetics
P1NP	Procollagen Type 1 N-Propeptide;
PRO	Patient reported outcomes
PROMIS	Patient-Reported Outcomes Measurement Information System,
PROMIS SexFS	PROMIS Sexual Function and Satisfaction

PSA	Prostate-specific antigen
QoL	Quality of life
RBC	Red Blood Cell
RR	Respiratory Rate
RTF	Rich Text Format
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SBP	Systolic blood pressure
SHBG	sex hormone binding globulin
SI	Standard International
SOC	System Organ Class
SOP	Standard Operating Procedure
TEAE	Treatment Emergent Adverse Event
TSH	Thyroid Stimulating Hormone
WBC	White Blood Cell

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide detailed descriptions of the statistical methods, data derivations and data displays for study protocol MBGS205 Final V6.0 "A Phase IIb multicentre, double-blind, dose-ranging, randomised, placebo-controlled study evaluating safety and efficacy of BGS649 in male obese subjects with hypogonadotropic hypogonadism" dated 10 March 2017. The table of contents and templates for the Tables, Figures and Listings (TFLs) will be produced in a separate document.

Any deviations from this SAP will be described and justified in the Clinical Study Report (CSR).

The preparation of this SAP has been based on International Conference on Harmonisation (ICH) E9.

All data analyses and generation of TFLs will be performed using SAS 9.4® or higher.

Note: This plan does not address the pharmacokinetic (PK)/pharmacodynamics (PD) analyses or summarisation for this study. These analyses will be described in a separate analysis plan and results from the PK/PD modelling will be reported separately from the CSR.

Note: This plan does not include the PRO quantitative validation for the study. This will be reported separately from the CSR.

Note: This plan does not address the statistical analyses or summarisation for the extension study, MBGS206. These analyses will be described in a separate analysis plan.

2 STUDY OBJECTIVES

The study is designed to investigate the efficacy, safety and tolerability of three different doses regimens of BGS649 versus placebo in male obese subjects with hypogonadotropic hypogonadism.

2.1 Primary objective(s)

The primary objective is to demonstrate the efficacy of BGS649 to normalise total testosterone levels (300-1000 ng/dL [10.4-35 nmol/L]) in ≥ 75% of subjects after 24 weeks of treatment.

2.2 Secondary objective(s)

The secondary objectives are:

1. To follow the time-course of normalisation in total testosterone levels
2. To demonstrate the efficacy of BGS649 to normalise total testosterone levels in ≥ 90% of subjects after 24 weeks of treatment
3. To evaluate the effect of BGS649 on LH and FSH
4. To further determine the pharmacokinetics (PK) of BGS649.

2.3 Exploratory objective(s)

The exploratory objectives are:

1. To evaluate the effect of BGS649 on oestradiol, DHT and inhibins (A and B)
2. To evaluate the effect of BGS649 on testosterone/oestradiol ratio
3. To investigate the effect on patient reported outcomes (PROs) following treatment with BGS649. PRO measures will be:
    - o International Index of Erectile Function (IIEF) and PROMIS Sexual Function and Satisfaction (PROMIS SexFS): to assess improvement in erectile function.
    - o IIEF and PROMIS SexFS: to assess improvement in sexual desire and satisfaction with sex life.
    - o 36-item Short Form Health Survey (SF-36): to assess general quality of life (QoL).
    - o Brief Fatigue Inventory (BFI), PROMIS Fatigue Short Form and the SF-36 Vitality: to assess improvement in energy levels.
    - o Patient Global Impression of Status items (PGI-S): to assess the patients' overall impression of their current health status.
4. To determine any change in physical activity and sleeping pattern via wrist worn monitor

5. To determine any change in grip strength, as measured by a dynamometer
6. To determine any associations between objective findings of improvement (e.g. testosterone, time in physical activity or sleep, and grip strength) and total and domain scores on PRO measures
7. To determine body composition change: waist circumference, BMI, body composition (as measured by impedance), and gynecomastia
8. To determine any associations between change in testosterone, body fat (as measured by BMI and waist circumference) and change in grip strength and wrist monitor data
9. To determine any associations between change in testosterone, body composition (as measured by impedance) and change in grip strength and wrist monitor data
10. To determine the effect of BGS649 on cardiometabolic parameters: systolic blood pressure (SBP) and diastolic blood pressure (DBP), lipid panel (total cholesterol, low density lipoprotein [LDL], triglycerides and HDL), glycosylated haemoglobin (HbA1c), fasting glucose and insulin, HOMA-IR and high sensitivity C reactive protein (hs CRP)
11. To determine any associations between testosterone, change in body composition and cardiometabolic parameters: SBP and DBP, lipid panel (total cholesterol, LDL, triglycerides and HDL), HbA1c, fasting glucose and insulin, HOMA-IR and hs-CRP
12. To determine the change in bone mineral density (dual energy X-ray absorptiometry [DEXA] scan), bone turnover markers (C-terminal telopeptide [CTx1] and procollagen type 1 N-propeptide [P1NP], osteocalcin, bone alkaline phosphatase)
13. To explore the effects of BGS649 on bone biomarkers on subjects with and without 25 hydroxy vitamin D deficiency (vitamin D deficiency is defined as 25 hydroxy vitamin D < 20ng/ml)
14. To explore the relationship between BGS649 concentration and testosterone levels
15. To determine the effects of BGS649 on semen parameters (semen volume, sperm count, concentration, motility and morphology) and to measure BGS649 concentrations in seminal fluid in subjects that provide semen samples
16. To explore the relationship between LH/FSH change and improvement in semen analysis in subjects that provide semen samples.

2.4 Safety objective(s)

To evaluate the safety and tolerability of BGS649.

3 STUDY DESIGN

3.1 General study design

This is a phase IIb, multicentre, double-blind, randomised, placebo-controlled parallel-group 36-week study evaluating safety and efficacy of BGS649 in male obese subjects with hypogonadotropic hypogonadism (HH). The study will have three phases:

1. A Screening Phase lasting up to 28 days, including Day -28 to Day -1
2. A Randomisation Phase including the Baseline visit (Day 1) and the treatment period Day 1 to Visit 8 (Week 24)
3. A Safety Follow-Up Phase to Visit 11 (Week 36, follow-up [FU]). Visits 9 and 10 are telephone assessments

Study participation will comprise up to 2 Screening Visits, 8 visits during the treatment period up to Week 24 (End of Treatment [EOT]), FU Visits 9, 10 (telephone assessment) and 11 (at Weeks 28, 32 and 36 respectively) for safety performed to 12 weeks after Visit 8 (EOT).

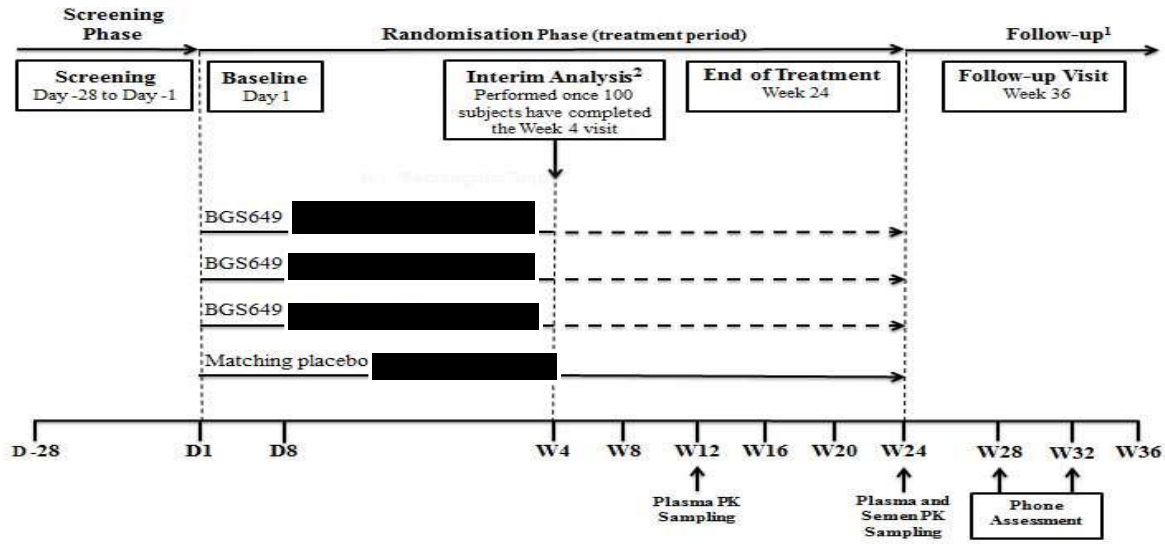
Eligible subjects will be randomised on Day 1 (Baseline) in a 1:1:1:1 ratio to one of the following treatment groups, consisting of [redacted]

- BGS649 [redacted]
- BGS649 [redacted]
- BGS649 [redacted]
- BGS649 matching placebo [redacted].

Once approximately 100 of the overall planned enrolled subjects have completed the Week 4 visit, an interim analysis will be performed by an independent DMC who will determine whether the criteria for normalisation of testosterone (testosterone within or above normal reference range for healthy adult males according to FDA criteria) were met. Dosing arms which met this criterion will be deemed effective. For dosing arms deemed to be ineffective, dosing will be stopped and no further randomisation to that arm will be performed. See Section 8.8.1.2 for full details of the Interim Analysis.

A minimum of 25 subjects per treatment arm will be invited to participate in a 6 month extension study (Protocol MBGS206), starting at Visit 8 (EOT). Subjects participating in the 6 month extension will have their last study visit at Visit 8 (EOT) and will not participate in Visits 9, 10, and 11.

Figure 1: Study Flow Chart



D=day; DMC=Data Monitoring Committee; EOT= End of Treatment; IRT=interactive response technology; PK, pharmacokinetic; W=week

1 A minimum of 25 subjects per treatment arm will be invited to participate in a 6-month extension study (Protocol MBGS206), starting at Visit 8 (Week 24 EOT). This transfer will be handled by the IRT and the blind will be maintained. Subjects participating in the 6-month extension will have their last study visit at Visit 8 (Week 24 EOT)

2 If one of the BGS649 study arms at Visit 3 (Week 4) fulfils discontinuation criteria as determined per DMC, the dosing for the ineffective or unsafe arm(s) will be stopped and no further randomisation to that arm will be performed.

3.2 Randomisation and blinding

3.2.1 Randomisation

This is a double-blind randomised study. At Visit 1 (Baseline) all eligible subjects will be randomised via the IRT to one of the four treatment regimens (BGS649 0.1 mg, 0.3 mg, and 1.0 mg and matched placebo) in a 1:1:1:1 ratio. After the interim analysis (see Section 8.8.1.2), if one or more of the BGS649 study arms is discontinued by the DMC, the dosing for that arm(s) will be stopped and no further randomisation to that arm will be performed. The remaining subjects will be randomised into the remaining active arms versus placebo (ineffective or unsafe dose will not continue). The IRT will assign a medication number to the subject, which will be used to link the subject to a treatment regimen and will specify a unique medication number on the label of investigational treatment to be dispensed to the subject.

3.2.2 Blinding

Subjects, investigational staff, persons performing the assessments and data analysts (with the exception of unblinded DMC staff/members) will remain blind to the identity of the treatments [redacted] from the time of randomisation until completion of the randomisation phase. In order to maintain the blind described above, testosterone measurements will be blinded to site and sponsor staff and monitoring will be performed by an independent unblinded physician.

3.2.3 Unblinding

Upon completion of the randomization phase (final Week 24 or EOT data collected for the final subject) the blind will be broken and the primary efficacy analysis completed. The 12 week, off treatment, follow up safety data will be analyzed separately.

Emergency treatment code breaks should only be undertaken when it is essential to treat the subject safely and efficaciously. Emergency code breaks are performed using the IRT.

3.3 Study treatments and assessments

During the 24 week randomization phase, subjects in the BGS649 groups and the matching placebo group will take the required dose weekly (± 1 day from the time schedule of regular planned dose) with water. Subjects will receive three capsules per dose to deliver a total of [redacted] and two indistinguishable placebo capsules)[redacted] or placebo [redacted]

Dose changes are not permitted during the randomization phase.

The maximum study duration from screening to end of the safety follow-up period is 40 weeks.

A detailed description of procedures and assessments to be conducted during this study is summarised in the Schedule of Study Assessments in Table 1 below.

Table 1: Schedule of Study Assessments

	Screening Phase		Randomisation Phase (Treatment Period)								Follow-Up[1]		
Visit	**Screening (D -28 to D -1)**		**1 (Baseline)**	**2**	**3**	**4**	**5**	**6**	**7**	**8 (EOT)**	**9**	**10**	**11 (FU)**
Week/Day	**Screen Visit 1**	**Screen Visit 2 (Visit 1 + 3-7 days)**	**D 1**	**D 8 (±1 D)**	**W4 (±2 D)**	**W8 (±2 D)**	**W12 (±2 D)**	**W16 (±2D)**	**W20 (±2 D)**	**W24 (±2 D)**	**W28[2] (±2 D)**	**W32[2] (±2 D)**	**W36 (±2 D)**
Written informed consent	X												
Demographics	X												
Medical/surgical history	X												
Inclusion/exclusion criteria	X												
Androgen deficiency symptom checklist	X												
Randomisation			X										
Full general physical exam and prostate exam and breast exam	X						X			X			
Limited physical exam[3]			X	X	X	X		X	X				X
Vital signs[4]; height (Visit 1 only)	X		X	X	X	X	X	X	X	X			X
12-Lead ECG	X		X		X	X	X			X			
Specialised chemistry: morning cortisol, TSH, FT4, prolactin, iron, transferrin[5]	X												
25 hydroxy vitamin D			X										
Clinical laboratory tests: haematology, blood chemistry, eGFR by Cockcroft-Gault formula, dipstick urinalysis[6] and PSA[7]	X		X	X	X	X	X	X	X	X			X
Cardiometabolic parameters 1:HbA1c, fasting lipids (total cholesterol, LDL, HDL, triglycerides)[7]	X				X		X			X			X
Cardiometabolic parameters 2: Fasting glucose and insulin, hs-CRP, HOMA-IR **Bone turnover markers**[7,8]			X				X			X			X
Sex hormones 1: testosterone (total), oestradiol (total), LH	X[9, 11]	X[9,10, 11]	X[11]	X[11]	X[11]	X[11]	X[11]	X[11]	X[11]	X[11]			X
Bioavailable testosterone, SHBG			X[11]	X[11]	X[11]	X[11]	X[11]	X[11]	X[11]	X[11]			X
Sex hormones 2: FSH			X	X	X	X	X	X	X	X			X

	Screening Phase		Randomisation Phase (Treatment Period)								Follow-Up[1]		
Visit	**Screening (D -28 to D -1)**		**1 (Baseline)**	**2**	**3**	**4**	**5**	**6**	**7**	**8 (EOT)**	**9**	**10**	**11 (FU)**
Week/Day	**Screen Visit 1**	**Screen Visit 2 (Visit 1 + 3-7 days)**	**D 1**	**D 8 (±1 D)**	**W4 (±2 D)**	**W8 (±2 D)**	**W12 (±2 D)**	**W16 (±2D)**	**W20 (±2 D)**	**W24 (±2 D)**	**W28[2] (±2 D)**	**W32[2] (±2 D)**	**W36 (±2 D)**
DHT, inhibin A and B			X				X			X			X
Semen quality analysis[12]			X				X		X				
Semen PK samples[13]										X			
Plasma PK samples[14]							X			X			
Hip and waist measurement			X				X			X			
Body composition using bioimpedance[15] (by impedance)			X				X			X			
Actigraphy[16]			X				X			X			X
Grip strength			X		X	X	X			X			X
PROs: IIEF, BFI, PROMIS SexFS, PROMIS Fatigue Short Form, PGI-S		X[17]	X		X	X	X			X			X
PROs: SF-36			X				X			X			X
DEXA	X[18]									X[18]			
Study treatment							[19]						
Study drug dispensation													
Study drug accountability				X	X	X	X	X	X	X			
AE assessment[19]		X	X	X	X	X	X	X	X	X	X	X	X
Concomitant medication	X		X	X	X	X	X	X	X	X	X	X	X

1 A minimum of 25 subjects per treatment arm will be invited to participate in a 6-month extension study (Protocol MBGS206), starting at Visit 8 (EOT). Subjects participating in the 6-month extension will have their last study visit at Visit 8 (EOT)

2 Telephone Assessment Visit only

3 Limited physical examination (cardiovascular system and lower extremities oedema)

4 Vital signs assessment will include the measurement of SBP, DBP, pulse, oral temperature and body weight

5 Fasting transferrin

6 A microscopic examination including RBC and WBC will be performed only when dipstick evaluation is positive for WBC and/or blood

7 All blood samples are to be collected after 8 hours fasting, after ECG and vital sign measurements

8 Bone turnover markers include CTx1, osteocalcin, bone alkaline phosphatase and P1NP

9 At Screening serum total testosterone concentration must be < 300 ng/dL (10.4 nmol/L) based on two morning samples, taken before 11 am at least 3 days apart. Laboratory parameters from screening visit 1 do not need to be received before screening visit 2 is take place

10 Oestradiol and LH will not be measured at this visit

11 Samples to be collected in the morning before 11 am pre-dose

12 Semen quality analysis will include semen volume, sperm count, concentration, motility and morphology. Sample collection is not required to be the same day as the assessment visit. Therefore sample collection is allowed within +/-48 hours the scheduled visit for Baseline and attendance Visits during the treatment phase

13 Semen samples for seminal fluid PK is not required to be the same day as the assessment visit. Therefore semen PK collection is allowed +/- 48 hours around the attendance visits

14 Samples will be taken at pre-dose and 1h post-dose for sparse plasma PK sampling at Visit 5 (week 12). Investigators should ensure that Visit 5 where plasma PK is measured coincides with the study drug dosing day. At EOT Visit 8 (Week 24), PK will be taken only once together with other laboratory assessments taken at that visit.

15 Subjects should come well hydrated, drinking water the night before and the morning before the impedance measurements

16 Activity wrist monitors will be worn by subjects for 7 continuous days after Baseline visit and visit 5 and 7 continuous days before Visits 8 and 11, so that the monitors can be collected at Visit 8 from subjects transferring to the extension study and at Visit 11 from subjects completing FU visit.

17 Screening PRO's should be performed no more than 7 days before dosing and may be performed outside of the Screen Visit 2 to meet this timing if required

18 Screening DEXA scan will be performed during screening and up to Day 8 after randomisation. The End of treatment DEXA will be performed only in a case patient reached minimum visit 5(w12) and it can be performed within 14 days before EOT to accommodate scheduling

19 The administration of the study drug (BGS649/placebo) will be performed at the study site on Week 12 to allow measurement of PK at pre-dose and 1h post-dose

20 SAEs collected from signature of informed consent and AEs, AESIs collected from randomization.

4 STUDY ENDPOINTS

4.1 Primary efficacy endpoint(s)

The primary efficacy endpoint of this study is the normalisation of total testosterone levels in ≥ 75% of subjects at Week 24

4.2 Secondary efficacy endpoint(s)

The secondary efficacy endpoints of this study are:

1. Proportion of subjects that have normalisation of total testosterone to Week 24
2. Proportion of subjects that overshoot testosterone (total testosterone above 1000 ng/dL [35 nmol/L]) to Week 24
3. Normalisation of total testosterone in ≥ 90% subjects at Week 24
4. Change of LH and FSH to 24 weeks
5. Population PK analysis of plasma BGS649 concentrations
6. PK analysis of semen BGS649 concentrations.

4.3 Exploratory endpoint(s)

The exploratory endpoints of this study are:

1. Change in oestradiol, inhibins (A and B) levels and DHT
2. Change in testosterone/oestradiol ratio
3. Body composition changes at Week 12 and Week 24
4. Changes in markers of cardiometabolic disease: blood pressure, lipid profile, HbA1c, glucose and insulin, hs-CRP and HOMA-IR and association with change in body composition and testosterone level.
5. Change in markers of bone turnover in those with and without 25-hydroxy vitamin D deficiency
6. Change in total and domain scores on PRO measures (IIEF, PROMIS SexFS: to assess sexual function; BFI, PROMIS Fatigue Short Form, SF-36 - Vitality: to assess energy levels and fatigue; SF-36: to assess general quality of life (QoL); PGI-S: to assess the patients' impression of their current health status) over the 24 week period
7. Association of objective findings of improvement (e.g. testosterone, time in physical activity or sleep and grip strength) and domain scores on PROs
8. Change in physical activity, sleeping pattern and strength measured by wrist worn monitors and grip strength measurement at 12 and 24 weeks and association with body composition (as measured by BMI and waist circumference and impedance) and testosterone level.
9. PK/pharmacodynamic (PD) relationship between BGS649 concentrations and testosterone levels
10. For semen analysis: change in semen parameters throughout the study and association with LH/FSH level
11. Change in bioavailable testosterone

12. Time to first normal testosterone level.

4.4 Safety endpoint(s)

The safety endpoints of this study are:

1. TEAEs/SAEs (from first dose of study drug until 90 days after last treatment dose)
2. Change in PSA during 24 week treatment duration
3. Change in haematocrit during 24 week treatment duration
4. Change in Bone Mineral Density (DEXA scan T-score and density in g/cm2) from Screening and bone turnover biomarkers at 24-weeks from Baseline
5. Change in vital signs and clinical laboratory parameters, ECG
6. Change in physical examination (including general, prostate, breast and oedema of the lower extremities).

5 SAMPLE SIZE AND POWER

The sample size for this study is calculated using exact binomial methods for comparison of a single proportion to a performance goal.

The statistical hypothesis is the proportion of subjects that have normalisation of testosterone at Week 24 within a dose group is greater than the pre-specified performance goal of 75%:

H0: Normalisation ≤ 75%

H1: Normalisation > 75%

The calculation assumes a one-sided test conducted at the 2.5% significance level, for which 67 subjects will need to be randomised in each dose group of BGS649.

If the true normalisation rate in a dose group of BGS649 is 90% then the power of the study is 85%.

6 ANALYSIS POPULATIONS

All primary and secondary efficacy endpoints will be analysed using the intention to treat (ITT) and per-protocol (PP) populations.

Safety and tolerability will be analysed using the safety population.

Pharmacokinetic data will be analysed using the PK population.

6.1 Intention-To-Treat population (ITT)

The ITT population includes all subjects who:

1. Are randomised, and
2. Receive at least one dose of study medication, and
3. Provide a Baseline and at least one post-Baseline testosterone value.

Subjects will be analysed according to the treatment to which they were randomised.

6.2 Safety population (Safety)

The safety population includes all subjects who received at least one administration of the study medication.

Subjects will be analysed according to the treatment actually received. Subjects who did not receive the treatment planned by the randomisation will be analysed according to the treatment received.

6.3 Per-Protocol population (PP)

The PP population is a subset of the ITT population and includes all randomised subjects who have been treated according to the protocol and fulfil the following criteria (to be further described in the classification meeting plan):

1. Specific inclusion/exclusion criteria satisfied
2. Absence of relevant protocol violations with respect to factors likely to affect the efficacy of treatment where the nature of protocol violation will be defined before breaking the blind
3. Adequate study medication compliance
4. Adequate measurement of the primary variable.

Subjects will be analysed according to the treatment to which they were randomised.

6.4 PK population

For the nonlinear mixed effects modelling, all subjects who are randomised and received at least one administration of the study medication and have at least 1 quantifiable plasma concentration (either plasma or semen) will be included. Subjects will be analysed according to their randomised treatment.

6.5 Other Populations Defined for Tables and Listings

For the purposes of tables and listings a further two populations are utilized:

- All screened subjects (for use of subject disposition only)
- Randomised population (all randomised subjects)

6.6 Protocol Deviations/Violations and Exclusions from Analysis Sets

All violations and exclusions of subjects from analysis populations will be identified and documented at the Classification Meeting prior to study unblinding. The review of each subject's data will be conducted using (but not limited to) the following sources of information:

- Supportive subject listings, provided by the ICON ahead of the Classification Meeting, based on data recorded on the eCRF.
- Protocol Deviation Logs, retrieved from Clinical Trial Management System (CTMS).

7 STATISTICAL CONSIDERATIONS AND ANALYSIS

7.1 Derived Variables

7.1.1 General Variables

Study Day

Study day will be relative to the date of first study medication (Study Day 1) and calculated as:

- (assessment date – date of first dose of study medication) + 1, for assessments on or after the date of first dose of study medication date
- (assessment date – date of first dose of study medication), for assessments prior to date of first dose of study medication

Follow-up time

Follow-up time (days) will be calculated as:

(date of last contact - date of first dose of study medication) + 1

7.1.2 Definitions relative to demographic and other baseline characteristics

Age

Age at informed consent will be calculated as:
Age (years) = (date of informed consent - date of birth + 1) / 365.25

Weight, height and BMI

Weight, recorded in pounds on the eCRF, will be converted in kilograms (1 pound = 0.45359 kg). Height, recorded in inches on the eCRF, will be converted in centimeters (1 inch = 2.54 cm) (International System of Units).

Body mass index (BMI) will be calculated in kg/m^2 as: weight (kg)/(height (m))^2.

Temperature

Temperature, recorded in Fahrenheit degrees on the eCRF, will be converted in Celsius degrees: Celsius degrees = (Fahrenheit degrees - 32) x (5/9)

HH disease duration

Duration of HH disease will be calculated in years as:

(date of informed consent – date of initial diagnosis of HH + 1) / 365.25

7.1.3 Definitions relative to efficacy criteria

7.1.3.1 Total Testosterone

Testosterone Normalisation

A subject is deemed to have achieved testosterone normalisation if their total testosterone value is between 300-1000 ng/dL (10.4-35 nmol/L) inclusive.

To allow for a potential 10-25% decrease in total testosterone that can occur due to diurnal variation in the afternoons (given current testing is restricted to mornings), a stricter secondary definition of normalisation will be used. A subject is deemed to have achieved the secondary definition of testosterone normalisation if their total testosterone value is between 350-1170 ng/dL (12.15-40.6 nmol/L) inclusive. The secondary definition boundaries were derived by averaging the testosterone value that a 25% decrease in would result in the primary definition boundary with the primary definition boundary itself.

Testosterone Overshoot

A subject is deemed to have testosterone overshoot if their total testosterone value is greater than 1000 ng/dL (35 nmol/L).

Time to first Testosterone Normalisation

Time to first Testosterone Normalisation is defined as the time (in days) from the date of first dose of study medication until the date of the first occurrence of testosterone normalisation during the treatment period.

Subjects who do not achieve testosterone normalisation will be censored at their final testosterone assessment during the treatment period (Up to Visit 8).

End of Treatment Testosterone

End of treatment testosterone is defined as the Week 24 total testosterone assessment for subjects that completed the treatment period. For subjects that discontinued the treatment period early, the last on-treatment testosterone assessment will be used as their end of treatment testosterone value.

7.1.3.2 Body Composition

Gynecomastia

The size of glandular breast tissue, recorded in inches on the eCRF, will be converted in centimeters (1 inch = 2.54 cm).

7.1.3.3 International Index of Erectile Function (IIEF)

The IIEF is a 15-item self-administered questionnaire with each question being scored between 0-5 (see Appendix A for scoring conversion). Five sexual function domains will be derived by summing the scores of the associated questions (noted in parentheses below):

1. Erectile function (1,2,3,4,5,15)
2. Orgasmic Function (9,10)
3. Sexual Desire (11,12)
4. Intercourse Satisfaction (6,7,8)
5. Overall Satisfaction (13,14)

If at least one of the questions within a domain is missing then the domain will not be calculated. A higher score denotes a higher level of satisfaction/function

In addition the erectile function domain scores will be classified into the below severity categories:

- No Dysfunction (25-30)
- Mild (22-24)
- Moderate (11-21)
- Severe (1-10)

The minimal meaningful difference in the change from baseline in the erectile function domain score is defined as a change in 2 points for subjects who are mild at baseline, 5 points for subjects who are moderate at baseline and 7 points for subjects who are severe at baseline.

7.1.3.4 Patient-Reported Outcomes Measurement Information System(®) Sexual Function and Satisfaction (PROMIS(®) SexFS)

The PROMIS SexFS PRO measure will be utilized by collecting a 10 item self-administered questionnaire compiled from relevant items in the PROMIS Sexual Function and Satisfaction domains (Flynn et al., 2013). Each question is scored between 0 and 5 (See Appendix B for scoring conversion) and three sexual function domains will be derived by summing the scores of the associated questions (noted in parentheses below) and converting to a standardised T-score (standardised to a scale with mean 50 and Standard Deviation 10 – See Appendix B for conversion tables):

1. Interest in Sexual Activity (1,2)
2. Erectile Function (3)
3. Satisfaction with Sex Life (4,5,6,7,8)

If at least one of the questions within a domain is missing then the domain will not be calculated. In addition if a subject answers any question within a domain with a score of 0 (indicating the subject has not had the item in question) then the domain will also not be calculated.

Questions 9 and 10 (Interfering Factors) will be summarized by raw score as no calibration is available for interfering factors.

7.1.3.5 PROMIS(®) Fatigue Short Form

The PROMIS Fatigue Short Form measure is an 8-item self-administered questionnaire assessing the extent of fatigue and its impact on work and functioning in the last 7 days. Each question is scored between 1 and 5 (See Appendix C for scoring conversion) and a total raw score will be calculated by summing the values of the response to each question. A standardised T-Score will then be calculated by standardizing the raw total score to a scale with mean 50 and SD 10 (See Appendix C for conversion tables).

If between 1 and 4 items (inclusive) have a missing response then the raw total score will be

estimated from the non-missing responses using the below formula:

$$\text{Total Raw Score} = \frac{\text{Sum of scores of items answered} \times 8}{\text{Number of items answered}}$$

If the result is a fraction, round up to the nearest whole number. If greater than 4 items have a missing response then the total raw score will not be calculated

7.1.3.6 Brief Fatigue Inventory (BFI)

The BFI is an 9-item self-administered questionnaire, providing an assessment of the severity of fatigue and its impact on the subject's ability to function at present and in the previous 24 hours. A global fatigue score is calculated by deriving the mean score of each of the non-missing questions.

If more than 4 questions have a missing response, then the global fatigue score will not be calculated

7.1.3.7 36-Item Short Form Health Survey (SF-36)

The SF-36 questionnaire is a multidimensional instrument that evaluates quality of life. It measures 8 general health concepts or domains: Vitality, Physical Functioning, Bodily Pain, General Health Perceptions, Physical Role Functioning, Emotional Role Functioning, Social Role Functioning and Mental Health. These domains will also be summarised as physical and mental component scores. Please see Appendix D for detailed scoring of each domain and component scores.

7.1.3.8 Patient Global Impression of Status items (PGI-S)

Three patient global impression items will be included to assess the patients' overall impression of their current health status and thus to evaluate the performance of the other PROs in this patient population. Each response will be converted to a numeric score for summarising using the below conversions:

Response	Score
Not at all	1
A little bit	2
Somewhat	3
Quite a bit	4
Very much	5

7.1.3.9 Actigraphy Derived Activity and Sleep Parameters

The Autograph Link is a portable device that measures gross motor movements featuring a validated 3-axis accelerometer and data filtering technology that captures and records continuous, high resolution physical activity and sleep/wake information.

Subjects are instructed to wear the monitor for 7 days consecutively starting with the day of the scheduled Visit 1, and 5 and 7 days consecutively before Visits 8 and 11. Only data collected during the specified visit windows (See Appendix E) on days were the device had been worn for >=1152 minutes (i.e 80% daily compliance) and deemed (either algorithmically or hardware detected by the device) to have been collected while the device was being worn will be used. All evaluable data within the specified visit window will be used to derive each parameter at each visit. However if there are less than 3 evaluable days' worth of data collected within the window the parameter will be set to missing.

Activity Parameters

The average number of minutes spent per day within each activity category below will be calculated for each visit:

- Sedentary
- Light
- Lifestyle
- Moderate
- Vigorous
- Very Vigorous
- MVPA (Moderate to Very Vigorous Physical Activity)

In addition the average daily total number of steps and total activity counts (across all 3 axis) will be calculated as well as the maximum daily total activity count across each visit.

Sleep Parameters

The average length of awakenings (minutes), number of awakenings, number of minutes asleep (minutes), number of minutes awake (minutes) and proportion of time spent asleep during the sleep period will be calculated per day for each scheduled visit. Sleep periods starting before 5am will be assigned to the previous day.

7.1.3.10 Grip Strength

Grip strength assessment using a hand held dynamometer will be used as a surrogate measure of muscle strength. During this assessment, the subject will squeeze the device three times with each hand. An average (mean) of three measurements for each hand will be calculated at each visit.

7.1.3.11 Semen Parameters

Semen concentration will be assigned into the below categories:

- Normal (>15 million sperm/mL)

- Oligospermia: Mild (10-15 million sperm/mL)
- Oligospermia: Moderate (5-10 million sperm/mL)
- Oligosperma: Severe (<5 million sperm/mL)
- Azoospermia (complete absence of sperm)

In addition total motile count will be derived by the below formula:

Total Motile Count = Volume of Semen x Semen Concentration x percentage motile

7.1.3.12 Metabolic syndrome (APP III) Definition

A subject will be deemed as meeting the definition of metabolic syndrome if they meet 3 out of 5 the below criteria at the same visit:

- Waist Circumference >102cm (>40inches)
- Triglycerides > 150mg/dl (>1.69 mmol/L)
- HDL Cholesterol <40mg/dl (<1.036 mmol/L)
- Blood pressure >130mmHg (systolic) or >85mmHg (diastolic)
- Glucose >110 mg/dl (>6.11 mmol/L)

7.1.4 Definitions relative to safety parameters

7.1.4.1 Adverse Event (AE)

Treatment Emergent Adverse Event (TEAE)

A Treatment Emergent AE (TEAE) is defined as an AE occurring or worsening on or after the first dose of study medication. For subjects that enter the extension study, only AE's occurring prior to entering the extension study will be considered a TEAE and summarised.

TEAEs will also be split between AEs that start during the treatment period and those that start on or after the first day of the follow-up period (for subjects not entering the extension study).

Duration of AEs

The duration of an AE will be calculated as the resolution date minus the start date plus 1.

7.1.4.2 Treatment Compliance

Duration of Exposure

The duration of Exposure to study medication (number of days) will be defined as:

(date of last dose of study medication – date of first dose of study medication) + 1

Treatment Compliance

Overall treatment compliance will be defined as:

$$\frac{100\% \times [\text{Total Number of capsules dispensed} - \text{Total Number of capsules returned}]}{[(\text{Duration of Exposure (weeks)} \times \text{Number of capsules expected to be taken (3)}]}$$

Where Duration of Exposure (weeks) = [duration of exposure (days) / 7] rounded up to the nearest integer to derive the number of weeks started medication.

If either the duration of exposure is missing or the number of (Number of capsules dispensed – Number of capsules returned) cannot be calculated then the compliance calculation will not be computed.

7.1.4.3 Prior, Concomitant and Follow-up Medications/Procedures

Medications and Procedures will be assigned as being prior to study treatment, concomitant with study treatment or taken during the follow-up phase based on the start and stop dates of the medication and dosing dates.

If the medication/procedure stop date is before the date of the first dose of study medication, the medication/procedure will be assigned as being prior to study treatment. Otherwise, the medication/procedure will be assigned as being concomitant with study treatment unless the start date of the medication/procedure is after the Visit 8 date, when it will then be classified as occurring in the follow-up phase. For subjects that enter the extension study, only medications/procedures with a start date prior to entering the extension study will be categorised and summarised.

7.1.4.4 Osteopenia and Osteoporosis

DXA Scans assessments for each region (Spine and Hip) will be assigned into the below severity cateogories:

- Normal: (T score >-1.0)
- Osteopenia: (T Score between -1.0 and -2.5)
- Osteoporosis: (T score < -2.5 or Z-Score (under 50 years) <-2.0)

7.1.4.5 Potential Hy's Law Criteria

Subjects will be deemed as meeting potential Hy's law Criteria if their ALT or AST > 3x ULN and bilirubin > 2 x ULN at the same visit.

7.1.4.6 Estimate Glomerular Filtration Rate (eGFR)

A derived calculation of eGFR will be calculated using the 4 variable Modification of Diet in Renal Disease (MDRD) equation (Lavey 1999 and 2000):

$$\text{eGFR (mL/min/1.73m}^2\text{)} = 175\ (\text{Serum Creatinine in µmol/l} \times 0.011312)^{-1.154} \times (\text{age})^{-0.203} \times (0.742 \text{ if female}) \times (1.212 \text{ if African/American Black})$$

7.2 Handling of missing data and outliers

7.2.1 Missing data analysis methods

Unless specified, data in summary tables will be presented using Observed Case (OC) data and therefore no missing data will be imputed.

The primary endpoint will be performed using both a last observation carried forward (LOCF) imputation and non-responder imputation (NRI). For the LOCF approach, subjects with missing testosterone value at a visit will be imputed with the last non-missing, post-baseline testosterone value. This value will then be used to determine if the subject achieved normalisation at that visit. For the NRI approach, subjects with missing testosterone value at a visit will be deemed as not achieving testosterone normalisation at that visit.

For each continuous secondary endpoint, the change from Baseline will be assessed with a mixed model repeated measure (MMRM) analysis.

For data listings, unless specified, all data will be presented as they have been recorded (e.g. missing and partial dates will not be replaced).

7.2.2 Handling of missing or incomplete data

7.2.2.1 Partial dates of first HH diagnosis

In order to calculate the HH disease duration (i.e. time since first HH diagnosis), the following rules will be applied for partial dates for first diagnosis of HH:

- if the day of the month is missing it is imputed to be the 15th
- if both the day and month are missing, they are imputed to be June 30th
- missing years will be left as missing.

The above will be flagged and described as a footnote in the appropriate listings.

7.2.2.2 Missing items within Questionnaire data

The method of handling missing data for each questionnaire will be based upon the author's recommendation. Details can be found in Section 7.1.3.

7.2.2.3 Missing or incomplete concomitant medication dates

Should the start date for a medication/procedure be missing or incomplete to the extent that it could be before or after the time of start of study medication, then it will be assumed that the medication/procedure began after the start of study medication (i.e. reported as concomitant medication/procedure). Similarly, if it is not clear whether the medication/procedure start date was on or before, or after the Visit 8 date, then it will be assumed the medication/procedure began on or before the Visit 8 date (i.e. reported as a concomitant medication/procedure) (worst case approach).

7.2.2.4 Definition of treatment-emergent AEs and handling of missing or incomplete dates

In the event of an incomplete onset date, the event will be considered to be treatment-emergent unless the partial onset date information or complete or partial end date confirms onset or end prior to the first dose of study medication date (or after the date entering the extension study for those that continue into the extension study).

8 STATISTICAL METHODS

8.1 General statistical conventions

All statistical procedures will be completed using SAS version 9.4 or higher.

8.1.1 Populations for analysis

Demographic and baseline characteristics will be summarised by Safety, ITT and PP populations, unless otherwise stated.

Analyses of the primary efficacy endpoint will be performed on the ITT population, and also on the PP population as a sensitivity analysis.

The secondary and exploratory efficacy endpoints will be analysed using the ITT population.

The safety and tolerability variables will be analysed using the Safety population.

PK data will be analysed using the PK population.

8.1.2 Treatment groups


t regimens in the below order:

8.1.3 Descriptive statistics

Continuous variables will be summarised using descriptive statistics including number of non-missing observations (n), arithmetic mean (Mean), median (Median), standard deviation (SD), minimum value (Min) , maximum value (Max) and number of missing observations (if any). One additional decimal point for mean and median and 2 additional decimal points for SD will be used.

For categorical variables, summaries will include the number of non-missing observations (n) or the number of patients in the population (N) as applicable, the counts of subjects and percentages. Percentages will be rounded to one decimal place. The number of missing values will be presented as a separate category with no percentage, but only if any missing value is recorded in the data for that summary.

Summary statistics will only be presented at each visit for which the parameter is scheduled to be collected.

8.1.4 Statistical significance

Unless otherwise stated, all statistical testing will be two-sided and conducted at the significance (alpha) level of 0.05. Two-sided 95% confidence intervals (CIs) will be provided when relevant. All four treatment regimens will be assessed by pairwise comparisons, and no adjustment for

multiplicity will be made.

8.1.5 Definition of Baseline

For summary purposes, the baseline value will be defined as the last non-missing value collected prior to the first dose of study medication, unless other specified.

Change from Baseline is defined as (value at assessment date – baseline value).

For Actigraphy Derived Activity and Sleep Parameters, as data is not collected prior to administrating the study drug, data collected during the 7 days from Visit 1 will be considered baseline data for all associated analysis.

For Semen quality analysis/parameters, samples at baseline can be taken up to 48 hours after the Baseline Visit. As such baseline will be defined as the last non-missing value collected up to 48 hours post the date of first dose of study medication.

For DXA parameters, assessments at baseline can be taken up to day 8 after the Baseline Visit. As such baseline will be defined as the last non-missing value collected up to 8 days post the date of first dose of study medication.

For selected renal parameters change from Screening will be calculated. Change from Screening will be defined as (value at assessment date – screening value), where the screening value is defined as the last non-missing value collected at a Screening Visit (Day -28 to Day -1).

8.1.6 Data re-allocation

The following general rules to handle repeated assessments will be considered:

- 12-lead ECG:
  - in case of multiple measurements associated to the same timepoint (e.g. triplicate ECGs), the average value will be considered for all post baseline HR, RR interval, PR interval, QRS duration, QT interval, QTcB interval and QTcF interval.
- Semen quality analysis/parameters:
  - in case of multiple measurements associated to the same timepoint as permitted by the protocol the average value will be considered for all post baseline semen parameters.

8.1.7 Data listings

All relevant subject data, including those derived, will be presented in individual subject data listings. All listings will be sorted by treatment regimen, investigational site, subject number, date/time and visit. The subject's age will be stated on each listing. Unless otherwise stated, data listings will be based on all subjects randomised.

Unscheduled visit results will be included in date/time chronological order within subject listings, but will not be tabulated.

8.2 Subject disposition

All subjects who provided informed consent will be included in a summary of subject

accountability. The number of subjects screened, the number of screen failures, the frequency and percentage of subjects randomised, in the Safety population, in the ITT population, in the PP population and in the PK population will be summarised by treatment regimen and overall.

Subject disposition information will be summarised by treatment regimen and overall. The number of subjects completing and withdrawing from the Randomisation Phase (up to and including Visit 8), number of subjects entering the extension study and the number of subjects completing and withdrawing from the study will be tabulated. Reasons for treatment discontinuation and discontinuation from the study will also be presented.

The follow-up time will be summarised descriptively.

Finally, the reasons for screen failure as reported on the eCRF will be tabulated.

8.3 Protocol deviations

The number of subjects excluded from ITT, Safety, PP and PK populations and reasons for exclusion will be summarised by treatment regimen and overall.

All protocol deviations identified will be summarised descriptively by treatment regimen and overall.

Summaries will be conducted on all subjects that were randomised

8.4 Demographics and baseline characteristics

No formal comparison between treatment regimens on demographics and baseline characteristics will be conducted.

8.4.1 Demographics

Demographic variables will be listed and summarised by treatment regimen and overall. This will include age, sex, race, weight, height and BMI. Separate summaries will be produced using the Safety, ITT and PP populations.

8.4.2 Baseline and disease characteristics

Duration of HH disease (years) and the number of subjects receiving previous treatment and each item on the Androgen Deficiency Symptom Checklist will be summarised descriptively by treatment regimen and overall.

Total testosterone at initial diagnosis, first and second screening visits and the average testosterone over the first and second screening visits will be tabulated using the Safety, PP and ITT population.

In addition baseline total testosterone (absolute values and number of subjects above and below 200 ng/dl), free testosterone (absolute values and number of subjects above and below 47 pg/mL), bioavailable testosterone (absolute values and number of subjects above and below 130 ng/dl), oestradiol, DHT, inhibin A, inhibin B, HBA1c, sex hormone binding globulin (SHBG), fasting lipid profile (total cholesterol, low density lipoprotein [LDL], triglycerides and HDL), Vitamin D, IIEF domain scores, IIEF erectile function domain severity and PROMIS sexFS domains will be summarised descriptively by treatment regimen and overall.

All Summaries will be produced using the Safety Population and PP Population unless stated.

8.4.3 Medical history

A summary of all medical history will be presented by Medical History Code (as per CRF), system organ class (SOC) and preferred term (PT), by treatment regimen and overall using Medical Dictionary for Regulatory Affairs® (MedDRA) Version 19.1 or higher. Separate summaries will be produced using the Safety and ITT populations.

In addition the number of subjects with an MRI of pituitary performed, the MRI finding (Normal, Abnormal NCS or Abnormal CS) and the number of subjects who have had a vasectomy will be summarised.

All Summaries will be produced using the Safety Population.

8.4.4 Prior, Concomitant and Follow-up medications

Prior, concomitant and follow-up medications will be coded using the World Health Organization (WHO) Drug Dictionary (March 2016 or later).

Prior, concomitant and follow-up medications will be summarised by Anatomical Therapeutic Chemical (ATC) classification level 4 and Preferred Name by treatment regimen and overall for the Safety population.

8.4.5 Concomitant Procedures

Concomitant procedures will be summarised by Procedure Name, treatment regimen and overall for the Safety population.

8.5 Extent of exposure

8.5.1 Treatment duration

Duration of treatment (in days) will be categorised in intervals (1-28, 29-56, 57-84, 85-112, 113-140, >140) and summarised descriptively by treatment regimen on the Safety population using descriptive statistics.

8.5.2 Treatment compliance

The overall treatment compliance (in %) will be presented by treatment regimen using the Safety population. The number of subjects with overall compliance <80% or >120% will be presented

8.6 Efficacy analyses

This section addresses the analyses to be conducted on the primary, secondary and exploratory efficacy variables.

All definitions relative to efficacy and exploratory endpoints are detailed in Section 7.1.3.

8.6.1 Analysis methods

For binary outcomes the proportion of subjects with a response will be summarised and compared between treatment regimens using Fisher's Exact Test.

For all continuous outcomes, where appropriate, a mixed model repeated measure model (MMRM) analysis will be performed. Analyses will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of the associated baseline value and baseline-by-visit interaction. An unstructured covariance matrix will be used to model the within-subject errors. If the model with the unstructured covariance matrix fails to converge, other covariance structures, including Toeplitz, compound symmetry, and spatial power, will be considered. The denominator degrees of freedom will be calculated according to the Kenward-Roger method. Significance tests for treatment differences (individual BGS649 doses vs. Placebo and between BGS649 doses) will be based on least-squares means.

8.6.1.1 Multiplicity

All four treatment regimens will be assessed by pairwise comparisons. No adjustment will be made for multiple comparisons.

8.6.1.2 Treatment by centre interaction analysis (multi-centre study)

No analysis will be made to assess the treatment-by-centre interaction.

8.6.2 Analysis of primary efficacy endpoint(s)

The Primary Endpoint will be considered to have been met for a dose if ≥ 75% of subjects in the ITT population have normalisation of total testosterone levels (testosterone value is between 300-1000 ng/dL inclusive) at Week 24.

The number and proportion of subjects with normal testosterone levels will be summarised by treatment regimen at each visit using the total testosterone assessed by ICON central lab. The Week 12 total testosterone values assessed by the GLP lab will also be presented. In addition, a P-value for the comparison between each dose will be provided using Fisher's exact test.

Moreover, a bar chart will be presented per treatment regimen for the normalisation of testosterone at each visit showing the percentage of subjects normalised and confidence intervals.

The primary analysis of efficacy will be performed with the ITT population using LOCF imputation. Additionally for exploring the robustness of the intent-to-treat results, a supportive analysis using the PP population using LOCF, ITT population using non-responder analysis (subjects with missing testosterone values will be deemed as not achieving testosterone normalisation) and ITT population using observed cases will be carried out.

8.6.3 Analysis of secondary efficacy endpoint(s)

All analysis of secondary efficacy endpoints will be performed on the ITT population and the PP Population (as required).

8.6.3.1 Testosterone Normalisation (Secondary Definition)

The secondary definition of testosterone normalisation (350-1170 ng/dL inclusive) that allows a decrease in total testosterone due to diurnal variation will be analysed in the same manner as the primary efficacy endpoint.

8.6.3.2 Change from Baseline in Total Testosterone

A descriptive summary table will be presented for Total Testosterone values and change from baseline in Total Testosterone by treatment regimen and visit using both ICON central lab and GLP lab assessed testosterone.

Change and percentage change from baseline in Total Testosterone up to Week 24 using the ICON central lab assessed values will be analysed using an MMRM model as described in Section 8.6.1. Least squares means and standard errors will be presented for each visit, p-value and 2-sided 95% confidence interval for treatment differences (BGS649 doses vs. Placebo) will be presented for each visit. Significance tests will be based on least-squares means using a two-sided $\alpha = .05$.

A figure presenting the least square mean change from baseline in total testosterone at each visit by treatment regimen will also be provided.

Change from baseline in Total Testosterone up to Week 12 using the GLP lab assessed values will be analysed using an ANCOVA model with treatment as a fixed effect and the associated baseline value as a covariate. Least squares means and standard errors will be presented for each visit, p-value and 2-sided 95% confidence interval for treatment differences will be presented for each visit. Significance tests will be based on least-squares means using a two-sided $\alpha = .05$.

8.6.3.3 Proportion of subjects that Overshoot Testosterone

The number and proportion of subjects that overshoot testosterone (total testosterone above 1000 ng/dL [35 nmol/L]) at each visit and at least once during the study will be summarised similar to that of the primary endpoint.

In addition the number and proportion of subjects whose maximum testosterone value falls into each of the below categories will be presented by visit and overall during the study:

- <=1500ng/dl
- >1500 – <1800 ng/dL
- =>1800 - <= 2500 ng/dL
- >2500 ng/dL

8.6.3.4 Change in LH and FSH

Descriptive summary tables will be presented for LH and FSH values and change from baseline in LH and FSH by treatment regimen and visit.

The change from baseline in LH and FSH up to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

8.6.4 Analysis of exploratory endpoint(s)

The analysis of exploratory endpoints will be performed on the ITT population.

8.6.4.1 Change in oestradiol, inhibins (A and B) levels, DHT and testosterone/oestradiol ratio

Descriptive summary tables will be presented for oestradiol, inhibin A, inhibin B, DHT, testosterone/oestradiol ratio and DHT/testosterone ratio values,change from baseline values and

percentage change from baseline values by treatment regimen and visit. High sensitivity Oestradiol values will summarised together with the Oestradiol values assessed by the Central lab.

The change and percentage change from baseline in each parameter up to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1). Additionally a figure of absolute oestradiol will be provided by treatment regimen.

In addition the number and proportion of subjects with oestradiol values < 11 pg/mL will be summarised by treatment regimen at each visit.

8.6.4.2 Change in Body composition

Descriptive summary tables will be presented for Body Weight, BMI, Waist Measurement, Hip Measurement and Waist to Hip Ratio values and change from baseline values by treatment regimen and visit.

The change from baseline in each parameter up to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

The number and proportion of subjects with gynecomastia will be summarised by treatment regimen at each visit. A P-value for the comparison of the presence of gynecomastia between each dose of BGS649 and placebo will be provided using Fisher's exact test. A descriptive summary table for the size of glandular breast tissue will be presented for those subjects with gynecomastia.

Descriptive summary tables will be presented for all body composition bio impedance parameters [Total Fat (kg and percentage), Visceral adipose tissue, Total fat free mass (kg and percentage) and Skeletal muscle mass whole body, right arm, right leg, left arm, left leg and torso] values and change from baseline values by treatment regimen and visit.

The change from baseline in each parameter up to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

A subgroup analysis of body composition parameters by baseline total testosterone (<200ng/dL vs >=200ng/dL) as assessed by ICON central labs will be conducted. Descriptive summaries and analysis up to Week 24 as described above will be presented.

8.6.4.3 Changes in Cardiometabolic Disease Markers

Descriptive summary tables will be presented for blood pressure (systolic and diastolic blood pressure), fasting lipid profile (total cholesterol, low density lipoprotein [LDL], triglycerides and HDL), HbA1c, fasting glucose, fasting insulin, high sensitivity C reactive protein (hs CRP) and HOMA-IR values and change from baseline values by treatment regimen and visit.

The change from baseline in each parameter up to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

Non-fasting values of all lipid profile parameters, glucose and insulin will be excluded from both the descriptive summary and analysis.

In addition the number and proportion of subjects who met the definition of metabolic syndrome will be summarised by treatment regimen at each visit (baseline and post-baseline).

8.6.4.4 Change in bone turnover markers

Descriptive summary tables will be presented for all bone turnover marker (C-terminal telopeptide [CTx1] and procollagen type 1 N-propeptide [P1NP], osteocalcin and bone alkaline phosphatase) values, change from baseline and percentage change from baseline by treatment regimen and visit.

The change from baseline and percentage change from baseline in each parameter up to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1). Additionally line graphs of absolute bone turnover marker parameters will be provided by treatment regimen and boxplots of observed values by treatment and visit will be generated for all bone turnover markers.

In addition subgroup descriptive summary and analysis tables of bone turnover markers by subjects with and without 25 hydroxy vitamin D deficiency at baseline (vitamin D deficiency is defined as 25 hydroxy vitamin D < 30ng/ml) will be presented.

8.6.4.5 Change in PRO measures

For some PRO Measures (PROMIS SexFS, PROMIS Fatigue Short Form and PGI-S), not all protocol versions mandated their collection. Subjects originally enrolled using protocol versions 2.0 were not required to collect these measures. To handle such cases, denominators used for analysis will only include the subjects mandated to collect the particular PRO.

IIEF

Descriptive summary tables will be presented for each of the 5 sexual function domain scores (Erectile function, Orgasmic function, Sexual desire, Intercourse satisfaction and Overall satisfaction) and change from baseline scores by treatment regimen and visit.

The change from baseline in each sexual function domain score to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1), however 90% confidence intervals (instead of 95%) will be presented for all inferences.

The number and proportion of subjects achieving a minimal meaningful difference in the change in IIEF erectile function domain score will be summarised by treatment regimen at each visit and overall.

In addition the change from baseline in the IIEF domains will also be summarised and analysed as described above by erectile function baseline severity (No dysfunction, mild, moderate and severe), baseline total testosterone (<200ng/dL vs >=200ng/dL) as assessed by ICON central labs and end of treatment total testosterone (<300ng/dL, 300-500 ng/dL and >500ng/dL) as assessed by ICON central labs.

PROMIS SexFS

Descriptive summary tables will be presented for each of the 3 sexual function domains scores (Interest in Sexual Activity, Erectile Function, Satisfaction with Sex Life), 2 interfering factors raw

scores and change from baseline scores (domain and interfering factor raw scores) by treatment regimen and visit.

The change from baseline in each sexual function domain score and 2 interfering factor raw scores to Week 24 will be analysed via MMRM in the same manner as the change from baseline in IIEF domain scores including subgroup analysis by baseline total testosterone (<200ng/dL vs >=200ng/dL) and end of treatment total testosterone (<300ng/dL, 300-500 ng/dL and >500ng/dL) as assessed by ICON central labs.

PROMIS Fatigue Short Form

A descriptive summary table will be presented for the PROMIS Fatigue Short Form standardised T-Score and change from baseline score by treatment regimen and visit.

The change from baseline to Week 24 will be analysed via MMRM in the same manner as the change from baseline in IIEF domain scores including subgroup analysis by baseline total testosterone (<200ng/dL vs >=200ng/dL) and end of treatment total testosterone (<300ng/dL, 300-500 ng/dL and >500ng/dL) as assessed by ICON central labs.

BFI

A descriptive summary table will be presented for the global fatigue score and change from baseline score by treatment regimen and visit.

The change from baseline to Week 24 will be analysed via MMRM in the same manner as the change from baseline in IIEF domain scores including subgroup analysis by baseline total testosterone (<200ng/dL vs >=200ng/dL) and end of treatment total testosterone (<300ng/dL, 300-500 ng/dL and >500ng/dL) as assessed by ICON central labs.

SF-36

Descriptive summary tables will be presented for each of the 8 general health domain scores and 2 component scores and change from baseline scores by treatment regimen and visit.

The change from baseline in each domain and component score to Week 24 will be analysed via MMRM in the same manner as the change from baseline in IIEF domain scores including subgroup analysis by baseline total testosterone (<200ng/dL vs >=200ng/dL) and end of treatment total testosterone (<300ng/dL, 300-500 ng/dL and >500ng/dL) as assessed by ICON central labs.

PGI-S

Descriptive summary tables will be presented for each of the 3 status items and associated change from baseline scores by treatment regimen and visit.

The change from baseline in each status item to Week 24 will be analysed via MMRM in the same manner as the change from baseline in IIEF domain scores including subgroup analysis by baseline total testosterone (<200ng/dL vs >=200ng/dL) and end of treatment total testosterone (<300ng/dL, 300-500 ng/dL and >500ng/dL) as assessed by ICON central labs.

8.6.4.6 Change in physical activity, sleeping pattern and strength

Physical Activity

Descriptive summary tables will be presented for the average number of minutes spent per day over the 7 days within each activity category, total daily steps, total daily activity counts and maximum daily activity counts by treatment regimen and visit. The change from baseline in the average number of minutes spent within each activity category, total daily steps and total activity counts to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

Sleeping Pattern

Descriptive summary tables will be presented for each sleeping pattern parameter (The average length of awakenings [minutes], number of awakenings, number of minutes asleep [minutes], number of minutes awake [minutes] and proportion of time spent asleep during the sleep period) and associated change from baseline by treatment regimen and visit. The change from baseline in each parameter to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

Grip Strength

A descriptive summary table will be presented for the average grip strength and associated change from baseline by treatment regimen and visit. The change from baseline in average grip strength to Week 24 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

A subgroup analysis of the change from baseline in grip strength by baseline total testosterone (<200ng/dL vs >=200ng/dL) as assessed by ICON central labs will be conducted. Descriptive summaries and analysis up to Week 24 as described above will be presented.

8.6.4.7 Change in Semen Parameters

Descriptive summary tables will be presented for each semen parameter (semen volume, sperm count, concentration, motility percentage, total motile count and morphology percentage) and change from baseline scores by treatment regimen and visit.

The change from baseline in each semen parameter to Week 20 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

A subgroup analysis of the change from baseline in semen parameters by baseline assessment timing (prior to first dose of study medication vs after first dose of study medication) will be conducted. Descriptive summaries and analysis up to Week 20 as described above will be presented.

A shift table of change from baseline in sperm concentration categories to post baseline categories will be generated by treatment regimen and visit.

The number and proportion of subjects moving from azoospermia at baseline to oligospermia post baseline and oligospermia at baseline to normal post baseline will be summarised by treatment regimen at each visit and overall.

8.6.4.8 Change in Bioavailable testosterone and Free testosterone

Descriptive summary tables will be presented for bioavailable and free testosterone values and change from baseline values by treatment regimen and visit.

The change from baseline in bioavailable and free testosterone up to Week 24 will be analysed via MMRM in the same manner as the change from baseline in IIEF domain scores.

8.6.4.9 Time to first normal testosterone level

Time to first normal total testosterone level will be analysed using the Kaplan-Meier method. As per Section 7.1.3.1, subjects that did not achieve testosterone normalisation (300-1000 ng/dL) or who discontinued the treatment period before achieving it will be considered as "censored" subjects. The number of subjects are normalised at baseline will be presented but excluded from the analysis.

For the following time points: Week 4, Week 8, Week 12, Week 16, Week 20 and Week 24 the number of subjects not having achieved a normal total testosterone value from baseline to the timepoint, the cumulative number of subjects who have achieved a normal total testosterone value at the end of the period and the probability of achieving a normal total testosterone level at the end of the period with the associate two-sided 95% CIs will be presented by treatment regimen. The point estimates and the relative two-sided 95% CIs will be presented by treatment regimen for the 75^{th}, 50^{th} and 25^{th} percentiles. A Kaplan-Meier plot will also be presented. Comparisons between treatment regimens will be performed by means of the log-rank test.

8.7 Safety analyses

All definitions relative to safety endpoints are detailed in Section 7.1.4.

All safety analyses will be based on the Safety population and will be performed for all safety variables specified below.

No statistical tests will be performed.

8.7.1 Adverse events

All AEs will be classified by SOC and PT according to the MedDRA Version 19.1 or higher.

Details for imputing missing or partial start dates of adverse events are described in Section 7.2.2.4.

All AE summaries will be generated by treatment period, follow-up period and overall (i.e including those that started in both the treatment period and follow-up period).

Notes:

- Two AEs with the same PT will be considered as two different events when calculating the "number of events" in the tables.

- Where a subject has the same AE, based on preferred terminology, reported multiple times in the same category the subject will only be counted once at the preferred terminology level in AE frequency tables.
- Where a subject has multiple AEs within the same SOC in the same category, the subject will only be counted once at the SOC level in AE frequency tables.
- AEs where the intensity is missing will be assumed to be "Severe"
- AEs where the causality is missing will be assumed to have "Reasonable possibility of relatedness

An overall summary of AEs will be provided. The total number of events and number and proportion of subjects experiencing any AEs, TEAEs, TEAEs of special interest, AEs related to study medication (i.e where the investigator has recorded "Reasonable possibility of relatedness"), SAEs, SAEs related to study medication, severe AEs, AEs leading to permanent discontinuation of study treatment and AEs leading to death will be tabulated for each treatment regimen and overall.

Additionally, TEAEs, TEAEs of special interest, related TEAEs, serious TEAEs, related SAEs, AEs leading to permanent discontinuation of study treatment and AEs leading to death will be summarised by SOC and PT for each treatment regimen and overall (number and percentage of subjects experiencing at least one AE per PT as well as the number of observed events per PT). All TEAEs will also be summarised separately by maximum intensity for each SOC and PT and by causality for each SOC and PT.

A table presenting the number and percentage of subjects with at least one AE and the number of AEs for the most common treatment-emergent AEs (reported in ≥ 1% of patients in any treatment regimen) will be provided. PTs will be used for tabulation, sorted by decreasing overall frequency.

All AEs will be presented in full in a comprehensive listing including subject number, treatment regimen, intensity, seriousness, actions taken, outcome, causality, onset/stop and duration. Details of all TEAEs of special interest, SAEs, AEs leading to permanent discontinuation of study treatment and AEs leading to death will be listed separately.

8.7.2 Laboratory evaluations

8.7.2.1 Clinical laboratory

Haematology, blood chemistry, and urinalysis assessments will be conducted at Screening, Baseline, each visit during the treatment phase and Week 36.

The following laboratory parameters will be assessed:

- Blood chemistry
  - Sodium
  - Potassium
  - Chloride
  - Bicarbonate/CO_2
  - Blood urea nitrogen

  - Creatinine
  - Fasting Glucose
  - Albumin
  - Alkaline phosphatase
  - AST
  - ALT
  - GGT
  - PT/INR
  - Total bilirubin
  - Total protein
  - Calcium
  - Lipid panel (total cholesterol, LDL, HDL, triglycerides)
  - PSA
  - eGFR (calculated based on Cockcroft-Gault formula)
  - eGFR (derived from 4-v MDRD)
- Haematology
  - Red Blood Cell Count [RBC]
  - White Blood Cell Count [WBC]
  - Neutrophils
  - Lymphocytes
  - Monocytes
  - Eosinophils
  - Basophils
  - Haemoglobin
  - Haematocrit
  - Platelets
- Urinalysis
  - Specific gravity
  - pH
  - Protein
  - Bilirubin
  - Glucose
  - Blood
  - Ketones
  - Leukocytes

For the purposes of summarisation in both the tables and listings, all clinical laboratory data will be reported in Standard International (SI) units.

If a lab value is reported using a nonnumeric qualifier e.g., less than (<) a certain value, or greater than (>) a certain value, the given numeric value will be used in the summary statistics, ignoring the nonnumeric qualifier.

Descriptive statistics will be presented for quantitative clinical laboratory parameters for each treatment regimen and time-point. Similarly, changes from baseline (and changes from screening for selected renal parameters; creatinine, urea, and glomerular filtration rate) will be summarised.

The change from baseline in each renal parameter to Week 24 will also be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

Values outside the normal range will be categorised as H (above the normal range) or L (below the normal range) based on the laboratory's reference range and these will be flagged in the individual data listings along with the Investigator's assessment.

In addition boxplots of observed values by treatment and visit will be generated for the below lab parameters:

- Hematocrit
- Hemoglobin
- Creatinine
- PSA

Qualitative urinalysis parameters (e.g protein, glucose and blood) will only be listed.

8.7.2.2 Specialised Chemistry

The following laboratory parameters will be measured at Screening and will be only be listed:

Morning cortisol
thyroid stimulating hormone (TSH)
free thyroxine (FT4)
Prolactin
25-hydroxy vitamin D
Fasting transferrin saturation

8.7.2.3 Liver Function

The number and proportion of subjects with laboratory values for liver function and enzymes of clinical concern as follows will be presented descriptively by treatment regimen:

Parameter	Level for clinical concern
ALT	> 1.5 times the upper limit of the normal range (ULN) > 2xULN > 3xULN > 5xULN > 10xULN
AST	> 1.5xULN > 2xULN > 3xULN > 5xULN > 10xULN
Bilirubin Total	> 2xULN > 3xULN
ALT/AST and Bilirubin Total	ALT/AST> 3xULN and Bilirubin Total> 2xULN (assessments to occur at the same visit)

8.7.3 Vital signs

Vital sign assessments will be performed at each visit in the Randomisation Phase and Week 36.

Descriptive statistics (observed values and changes from baseline) will be presented for each treatment regimen and time-point for vital sign measurements (body temperature, pulse rate, systolic blood pressure, diastolic blood pressure, weight and BMI).

The change from baseline in systolic and diastolic blood pressure to Week 24 will also be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.1).

Boxplots of observed and change from baseline values by treatment and visit will also be generated for both systolic and diastolic blood pressure.

In addition the number and percentage of subjects meeting each of the criteria below will be presented:

- Normal: SBP <120mm Hg and DBP < 80 mm Hg
- Elevated BP: 120≤SBP ≤ 129 mm Hg and DBP < 80 mmHg
- Hypertension Stage 1: 130≤SBP ≤ 139 mmHg or 80≤DBP ≤ 89 mmHg
- Hypertension Stage 2: SBP ≥ 140 mm Hg or DBP ≥ 90 mmHg

- SBP Change from baseline >=20 mmHg
- DBP Change from baseline >=10 mmHg

8.7.4 Physical examinations

A full physical examination will be performed at Screening, Week 12 and Week 24 whereas a limited physical examination (cardiovascular system and lower extremities oedema only) will be performed at all other visits during the randomisation phase and Week 36.

The number and percentage of patients reporting an abnormal physical examination finding (Abnormal CS and Abnormal NCS) will be presented per body system for each treatment regimen and time-point.

In addition the grade of pitting oedema will be summarised descriptively for each treatment regimen and time-point.

8.7.5 Electrocardiograms

12-Lead ECG assessments will be performed at Screening, Baseline, Week 4, Week 8, Week 12, Week 24 and Week 36.

Descriptive statistics (observed values and changes from baseline) will be presented for the 12-lead ECG measurements for each treatment regimen and time-point for all ECG parameters (Heart Rate, PR interval, QT interval, RR interval, QRS duration and QTcF interval). In addition, the overall ECG interpretation will be summarised by presenting the number and percentage of subjects with "Normal", "Abnormal, NCS" and "Abnormal, CS" for each treatment regimen and time-point.

8.7.6 DXA Scan

DXA Scan assessments will be performed at screening (up to day 8 after randomisation) and Week 24 and will be over-read by IMI. Both the site assessed and over-read data will be used for all statistical analysis and utilise the visit windows specified in Appendix F. If multiple assessments are recorded within the window, the closet to the target date will be used for analysis.

Descriptive statistics (observed values, changes from baseline and percentage change from baseline) for bone density in both the lumbar spine and hip (total and femoral neck) will be presented . In addition descriptive statistics (observed values and changes from baseline) for the associated T-scores in both the lumbar spine and hip will be presented for each treatment regimen and time-point.

A shift table of baseline severity categories (Normal, Osteopenia and Osteoporosis) to post baseline severities will be generated by treatment regimen and visit.

Change from baseline in bone density (in both the lumbar spine and hip) to Week 24 will be analysed using an ANCOVA model with treatment as a fixed effect and the associated baseline value as a covariate. Least squares means and standard errors will be presented for each visit, p-value and 2-sided 95% confidence interval for treatment differences across all doses will be presented for each visit. Significance tests will be based on least-squares means using a two-sided $\alpha = .05$.

In addition subgroup descriptive summary tables of change in bone density by subjects with and without 25 hydroxy vitamin D deficiency at baseline (vitamin D deficiency is defined as 25

hydroxy vitamin D < 30ng/ml) will be presented.

8.8 Other analysis

8.8.1 DMC

An independent, unblinded external DMC will periodically review accumulating safety data. This will include data evaluation of accumulating unblinded safety data of BGS649, testosterone monitoring assessment and performing the interim analysis. Full details of composition, operational aspects, and data to be reviewed and recommendation of the DMC is provided in a separate DMC charter and DMC SAP.

8.8.1.1 Testosterone Monitoring

To ensure adequate safety monitoring, serum total testosterone levels will be evaluated at each visit by an independent unblinded physician and if subject meets discontinuation criteria (total testosterone is ≥ 1500 ng/dL (52 nmol/L) at any 2 consecutive time points during the study), the subject will be discontinued from treatment. In addition, if one of the active BGS649 study arms presents with ≥ 2 subjects meeting this discontinuation criterion at any time point within the first 24 randomized subjects to that arm or > 15% subjects meet this criterion at any time point ≥ 25 randomized subjects, then the arm will be evaluated by an independent unblinded, external DMC for potential dosing discontinuation based on FDA guidance (FDA Advisory Committee Industry Briefing Document Testosterone Replacement Therapy, Sept 2014). Subjects may be discontinued from the study arm and no further randomisation to that arm will be performed.

Any treatment arm discontinued due to this process will still be presented in summary tables however no formal analysis vs Placebo will be conducted.

8.8.1.2 Interim Analysis

An interim analysis was conducted once approximately 100 of overall planned subjects completed their Week 4 visit. The analysis was performed by the independent DMC who reviewed un-blinded data to identify:

1. Any safety signals across the dose groups which would preclude further continuation of dosing
2. Whether an individual arm has met the criteria of ≥ 75% of subjects reaching the normalisation of testosterone (testosterone within or above normal reference range for healthy adult males according to the FDA)
3. Whether in the opinion of the DMC following unblinded review that when the arm is fully recruited it is expected to achieve a ≥ 75% response at Week 24.

Dosing arms which met either criterion 2 or 3 were deemed to be effective. For arms that were determined to be ineffective or unsafe, dosing was to be stopped and no further randomisation to that arm was performed. Subjects assigned to any ineffective or unsafe arm(s) were to be discontinued from the treatment, attend an EOT visit and followed-up for an additional 12 weeks

(FU, Visit 11). Any treatment arm discontinued due to this process would still be presented in summary tables however no formal analysis vs Placebo will be conducted.

Following the results of the Interim Analysis no dose arm was dropped from the study.

8.8.2 PK analyses

Pharmacokinetic samples will be collected as follows:

1. Plasma PK samples will be taken pre-dose and 1h post-dose at Visit 5. At Visit 8 (EOT; Week 24) it will be taken only once.
2. Semen samples for seminal fluid PK will be taken at Visit 8 (EOT)

This plan does not address the PK analyses of BGS649 (Nonlinear mixed effects modelling will be utilised to develop a population PK model for BGS649 plasma concentrations and to develop a PK/PD model linking BGS649 plasma concentrations to changes in testosterone levels). These analyses will be performed by ICON Pharmacokinetics, Pharmacodynamics, Modelling and Simulation Department and will be described in a separate analysis plan. Results from the PK/PD modelling will be reported separately from the CSR.

BGS649 PK plasma concentrations will be summarised for the PK population by descriptive statistics by treatment regimen and time-point, and including the geometric mean and coefficient of variation.

A graphical representation of the PK plasma concentrations distribution using mean plots (mean and SD) will be provided over time per treatment regimen for the PK population. All dose levels will be plotted on the same graph.

8.8.3 Subgroup analysis

To determine the impact of other baseline measures on the primary efficacy analysis the following covariates may be added to the analysis independently and subgroup analysis performed to test for significance:

- Age
- BMI
- Other baseline parameters reviewed at the classification meeting

All parameters to be tested will be documented at the classification meeting prior to unblinding.

8.8.4 Association Testing

To determine any correlation between endpoints at Week 24, various association testing will be conducted using the PP Population. For the below sets of endpoints scatter graphs will be generated annotated with the coefficient of determination (R^2 statistic):

- Total testosterone vs PRO scores (total and domain)
- Total oestradiol vs PRO score (total and domain)

- Testosterone/oestradiol vs PRO scores (total and domain)
- Activity parameters vs PRO scores (total and domain)
- Sleep parameters vs PRO scores (total and domain)
- Change in grip strength vs PRO scores (total and domain)
- Change in total testosterone vs Change in Grip Strength
- Change in BMI vs Change in Grip Strength
- Change in waist Circumference vs Change in Grip Strength
- Change in bio impedance parameters vs Change in Grip Strength
- Total testosterone vs cardiometabolic parameters: SBP and DBP, lipid panel (total cholesterol, LDL, triglycerides and HDL), HbA1c, fasting glucose and insulin, HOMA-IR and hs-CRP
- Change in BMI vs cardiometabolic parameters: SBP and DBP, lipid panel (total cholesterol, LDL, triglycerides and HDL), HbA1c, fasting glucose and insulin, HOMA-IR and hs-CRP
- Change in waist circumference vs cardiometabolic parameters: SBP and DBP, lipid panel (total cholesterol, LDL, triglycerides and HDL), HbA1c, fasting glucose and insulin, HOMA-IR and hs-CRP
- Change in semen parameters vs Change in LH and FSH levels
- Change in semen parameters vs Change in oestradiol
- Change in systolic blood pressure vs Total Testosterone, Oestradiol and Testosterone/Oestradiol ratio (at Week 12 and Week 24)
- Change in diastolic blood pressure vs Total Testosterone, Oestradiol and Testosterone/Oestradiol ratio (at Week 12 and Week 24)
- Change in Creatinine vs Blood Pressure (Absolute systolic and diastolic blood pressure)
- Change in Creatinine vs Change in Blood Pressure (Systolic and Diastolic blood pressure)

8.8.4.1 PROMIS SexFS Domains Testing

End of treatment testosterone levels (<300ng/dL vs 300-500 ng/dL vs >500 ng/dL) will be tested against each PROMIS sexFS domain score (excluding the domain “interfering factors”) by a MMRM model. Analyses will include the fixed, categorical effects of end of treatment testosterone level, visit, and end of treatment testosterone level-by-visit interaction, as well as the continuous, fixed covariates of the associated baseline value and baseline-by-visit interaction. An unstructured covariance matrix will be used to model the within-subject errors. The end of treatment testosterone factor effect p-value aswell as the Week 24 least square mean differences and 95% CI of the pairwise comparisons of each end of treatment testosterone level group will be presented.

This process will be repeated to test the association of baseline testosterone level (<200 ng/dL vs >200 ng/dL) against the change from baseline of each PROMIS sexFS Domain (excluding the domain “interfering factors”).

8.8.4.2 IIEF Erectile Function Domain Testing

End of treatment testosterone levels (<300ng/dL vs 300-500 ng/dL vs >500 ng/dL) will be tested

against IIEF erectile function domain response of achieving a minimal meaningful difference at Week 24 by a logistic regression model. The end of treatment testosterone factor effect p-value aswell as the odds ratio and 95% CI of the pairwise comparison of each end of treatment testosterone level group will be presented.

This process will be repeated to test the association of baseline testosterone level (<200 ng/dL vs >200 ng/dL) against IIEF erectile function domain response of achieving a minimal meaningful difference.

8.9 CSR

The primary efficacy analysis and study unblinding will be performed when all data up to Visit 8 (Week 24) has been collected. Given the study has been powered at Week 24 and there will not be multiple comparisons, there is no requirement to adjust the Type I error rate.

The Clinical Study report (CSR) will be prepared based upon the results of this analysis.

The data collected during the off treatment 12 week follow-up will be separately analysed and published in an addendum to the CSR.

If for any reason a primary analysis assessment (Total Testosterone) is not available at the time of the primary analysis, such values will not be used for any future efficacy endpoint analysis at any stage (i.e they will continue to be assumed missing). However such values will be included for future safety endpoints analysis (e.g Total Testosterone Overshoot).

9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

Vitamin D deficiency is defined in the protocol as <20ng/ml. For all analyses Vitamin D deficiency has been defined as <30ng/ml.

10 REFERENCES

1. ICH Topic E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96 – adopted March 1998).

2. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med. 1999; 130:461-70.

3. Levey AS, Coresh J, Greene T, et al. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. Ann Intern Med. 2006; 145:247-54.

4. Landon W. Trost and John P. Mulhall. Challenges in Testosterone Measurement, Data Interpretation, and Methodological Appraisal of Interventional Trials. J Sex Med. 2016 July ; 13(7): 1029–1046.

11 APPENDICES

Appendix A - International Index of Erectile Function (IIEF)

Question Number	Question	Response	Score
1	Over the past 4 weeks how often were you able to get an erection during sexual activity?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
		No sexual activity	0
2	Over the past 4 weeks when you had erections with sexual stimulation, how often were your erections hard enough for penetration?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
		No sexual stimulation	0
3	Over the past 4 weeks when you attempted sexual intercourse how often were you able to penetrate (enter) your partner?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
		Did not attempt intercourse	0
4	Over the past 4 weeks during sexual intercourse how often were you able to maintain your erection after you had penetrated (entered) your partner?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
		Did not attempt intercourse	0
5	Over the past 4 weeks during sexual intercourse how difficult was it to maintain your erection to completion of intercourse?	Not difficult	5
		Slightly difficult	4
		Difficult	3
		Very difficult	2
		Extremely difficult	1
		Did not attempt intercourse	0
6	Over the past 4 weeks how many times have you attempted sexual intercourse?	11+ attempts	5
		7-10 attempts	4
		5-6 attempts	3
		3-4 attempts	2
		1-2 attempts	1
		No attempts	0
7	Over the past 4 weeks when you attempted sexual intercourse how often was it satisfactory for you?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
		Did not attempt intercourse	0

8	Over the past 4 weeks how much have you enjoyed sexual intercourse?	Very highly enjoyable	5
		Highly enjoyable	4
		Fairly enjoyable	3
		Not very enjoyable	2
		Not enjoyable	1
		No intercourse	0
9	Over the past 4 weeks when you had sexual stimulation or intercourse how often did you ejaculate?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
		No sexual stimulation or intercourse	0
10	Over the past 4 weeks when you had sexual stimulation or intercourse how often did you have the feeling of orgasm with or without ejaculation?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
		No sexual stimulation or intercourse	0
11	Over the past 4 weeks how often have you felt sexual desire?	Almost always or always	5
		Most times (much more than half the time)	4
		Sometimes (about half the time)	3
		A few times (much less than half the time)	2
		Almost never or never	1
12	Over the past 4 weeks how would you rate your level of sexual desire?	Very high	5
		High	4
		Moderate	3
		Low	2
		Very low or none at all	1
13	Over the past 4 weeks how satisfied have you been with your overall sex life?	Very satisfied	5
		Moderately satisfied	4
		About equally satisfied and dissatisfied	3
		Moderately dissatisfied	2
		Very dissatisfied	1
14	Over the past 4 weeks how satisfied have you been with your sexual relationship with your partner?	Very satisfied	5
		Moderately satisfied	4
		About equally satisfied and dissatisfied	3
		Moderately dissatisfied	2
		Very dissatisfied	1
15	Over the past 4 weeks how would you rate your confidence that you could get and keep an erection?	Very high	5
		High	4
		Moderate	3
		Low	2
		Very	1

Appendix B - Patient-Reported Outcomes Measurement Information System(®) Sexual Function and Satisfaction (PROMIS(®) SexFS)

Question Number	Question	Response	Score
1	How interested have you been in sexual activity?	Very	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
2	How often have you felt like you wanted to have sexual activity?	Always	5
		Often	4
		Sometimes	3
		Rarely	2
		Never	1
3	Please rate your ability to have an erection.	Very good	5
		Good	4
		Fair	3
		Poor	2
		Very Poor	1
4	How satisfied have you been with your sex life?	Very	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
5	How much pleasure has your sex life given you?	A lot	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		None	1
6	How often have you thought that your sex life is wonderful?	Always	5
		Often	4
		Sometimes	3
		Rarely	2
		Never	1
7	How satisfied have you been with your sexual relationship(s)?	Very	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
		Have not had a sexual relationship in last 30 days	0
8	When you have had sexual activity, how much have you enjoyed it?	Very much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
9	How much has fatigue or lack of energy affected your satisfaction with your sex	Very Much	5
		Quite a bit	4
		Somewhat	3

	life?	A little bit	2
		Not at all	1
		Have not had fatigue or lack of energy in the last 30 days	0
10	How much has weight gain affected your satisfaction with your sex life?	Very Much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
		Have not had weight gain in the last 30 days	0

T-Score Conversion Tables :

1. Interest in Sexual Activity (1,2)

Q1 Raw	Q2 Raw	T Score
1	1	33.4
1	2	40.1
1	3	44.5
1	4	50.3
1	5	55.3
2	1	39.8
2	2	43.6
2	3	47.6
2	4	52.7
2	5	57.4
3	1	43.1
3	2	46.7
3	3	51.1
3	4	55.5
3	5	60.1
4	1	45.4
4	2	49.3
4	3	54.5
4	4	59
4	5	63.8
5	1	47.1
5	2	51.2
5	3	57.5
5	4	63.2
5	5	70

2. Erectile Function (3)

Q3 Raw	T score
1	31.4
2	36.6
3	40.4
4	45.6
5	56.3

3. Satisfaction with Sex Life (4,5,6,7,8)

Satisfaction with Sex Life Tscore.xlsx

Appendix C - PROMIS(®) Fatigue Short Form

Question Number	Question	Response	Score
1	I feel fatigued	Very much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
2	I have trouble starting things because I am tired	Very much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
3	How run-down did you feel on average?	Very much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
4	How fatigued were you on average?	Very much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
5	How much were you bothered by your fatigue on average?	Very much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
6	To what degree did your fatigue interfere with your physical functioning?	Very much	5
		Quite a bit	4
		Somewhat	3
		A little bit	2
		Not at all	1
7	How often did you have to push yourself to get things done because of your fatigue?	Always	5
		Often	4
		Sometimes	3
		Rarely	2
		Never	1
8	How often did you have trouble finishing things because of your fatigue?	Always	5
		Often	4
		Sometimes	3
		Rarely	2

		Never	1

T-Score Conversion Table :

Fatigue 8a *Short Form Conversion Table*		
Raw Score	**T-score**	**SE***
8	33.1	4.8
9	38.5	2.7
10	41.0	2.2
11	42.8	2.0
12	44.3	1.9
13	45.6	1.8
14	46.9	1.8
15	48.1	1.8
16	49.2	1.8
17	50.4	1.8
18	51.5	1.7
19	52.5	1.7
20	53.6	1.7
21	54.6	1.7
22	55.6	1.7
23	56.6	1.7
24	57.5	1.7
25	58.5	1.7
26	59.4	1.7
27	60.4	1.7
28	61.3	1.7
29	62.3	1.7
30	63.3	1.7
31	64.3	1.7
32	65.3	1.7
33	66.4	1.7
34	67.5	1.7
35	68.6	1.7
36	69.8	1.8
37	71.0	1.8
38	72.4	2.0
39	74.2	2.4
40	77.8	3.7

*SE = Standard Error

Reference : https://www.assessmentcenter.net/

Appendix D - 36-Item Short Form Health Survey (SF-36)

VARIABLE	DERIVATION
SF-36 PF scale score	raw score = sum (items 3A, 3B, 3C, 3D, 3E, 3F, 3G, 3H, 3I, 3J) PF = (raw score -10) * 5 PF_Z = (PF – 82.62455) / 24.43176 **PF scale score = (PF_Z*10) + 50** When calculating the raw score, if 5 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 5 of the items are non-missing then PF scale score is missing. The response scale for each activity ranges from 1 to 3 where 1=limited a lot, 2=limited a little, and 3=not limited at all. A higher PF scale score indicates better physical functioning.
SF-36 RP scale score	raw score = sum (items 4A, 4B, 4C, and 4D) RP = [(raw score -4)/16] * 100 RP_Z = (RP – 82.65109) / 26.19282 **RP scale score = (RP_Z * 10) + 50** When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then RP scale score is missing. The response scale for each item ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher RP scale score indicates better role-physical functioning.

SF-36 BP scale score	raw score = sum (reversed item 7 and reversed item 8)
	BP = (raw score -2) * 10
	BP_Z = (BP – 73.86999) / 24.00884
	BP scale score = (BP_Z * 10) + 50
	
	Reverse direction of Item 7 as follows: if =1, set to 6; if =2, set to 5.4; if =3, set to 4.2; if=4, set to 3.1; if=5, set to 2.2 if=6, set to 1.
	
	Reverse direction of item 8 as follows:
	if =1 and original value of item 7=1, set to 6;
	if =1 and original value of item 7>=2, set to 5; if=2, set to 4;
	if=3, set to 3; if=4, set to 2; if=5, set to 1.
	
	If item 7 is answered and item 8 is missing, set 8 = reversed 7 as defined above. If 8 is answered and 7 is missing, set 7 as reverse item 8 as follows:
	if =1, set to 6;
	if =2, set to 4.75; if=3, set to 3.5; if=4, set to 2.25;
	if=5, set to 1.
	
	If 1 or more questions were answered, calculate BP scale score as defined above. If neither question was answered then BP scale score is missing.
	
	The scale for Question 7, amount of bodily pain, ranges from 1 to 6 where 1=None, 2=Very mild, 3=mild, 4=Moderate, 5=Severe, and 6=Very severe. The scale for Question 8, the degree to which pain interfered with normal work, ranges from 1 to 5 where 1=Not at all, 2=A little bit, 3=Moderately, 4=Quite a bit, and 5=Extremely.
	
	A higher BP scale score indicates lack of bodily pain.

SF-36 GH scale score	raw score = sum (reversed item 1, item 11A, reversed 11B, 11C and reversed 11D) GH = (raw score -5) * 5 GH_Z = (GH – 70.78372) / 21.28902 **GH scale score = (GH_Z * 10) + 50** Reverse direction of Item 1 as follows: if =1, set to 5; if =2, set to 4.4; if =3, set to 3.4; if=4, set to 2; if=5, set to 1. Reverse direction of item 11B and 11D by subtracting score from 6. When calculating the raw score, if 3 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 3 of the items are non-missing then GH scale score is missing. Responses for Question 1, an assessment of self-perceived health status, range from 1 to 5 where 1=Excellent, 2=Very good, 3=Good, 4=Fair, and 5=Poor. Responses for the items in Question 11 range from 1 to 5 where 1=Definitely true, 2=Mostly true, 3=Don't know, 4=Mostly false, and 5=Definitely false and reflect the subject's perception of their relative health and expectations of their future health status. A higher GH scale score indicates better general health perceptions.

SF-36 VT scale score	raw score = sum (reversed item 9a, reversed 9e, 9g and 9i) VT = [(raw score -4)/16] * 100 VT_Z = (VT – 58.41968) / 20.87823 **VT scale score = (VT_Z * 10) + 50** Reverse direction of Items 9a and 9e by subtracting score from 6. When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then VT scale score is missing. The scale for these items ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher VT scale score indicates more vitality.
SF-36 SF scale score	raw score = sum (reversed 6 and 10) SF = [(raw score -2) / 8] * 100 SF_Z = (SF – 85.11568) / 23.24464 **SF scale score = (SF_Z * 10) + 50** Reverse direction of score for item 6 by subtracting score from 6. When calculating the raw score, if 1 of the items is missing then substitute the missing score with the score on the non- missing item. If both items are missing then SF scale score is missing. Responses to Question 6, an assessment of the extent to which health/emotional problems interfered with social activities, range from 1 to 5 where 1=Not at all, 2=Slightly, 3=Moderately, 4=Quite a bit, and 5=Extremely. Responses to Question 10 reflect the amount of time that health/emotional problems interfered with social activities and range from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher SF scale score indicates better social functioning.

SF-36 RE scale score	raw score = sum (items 5A, 5B, and 5C) RE = [(raw score -3) / 12] * 100 RE_Z = (RE – 87.50009) / 22.01216 **RE scale score = (RE_Z * 10) + 50** When calculating the raw score, if 2 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 2 of the items are non-missing then RE scale score is missing. Responses to the items in Question 5 range from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher RE scale score indicates better role-emotional functioning.
SF-36 MH scale score	raw score = sum (items 9B, 9C, reversed 9D, 9F and reversed 9H) MH = (raw score - 5) * 5 MH_Z = (MH – 75.76034) / 18.04746 **MH scale score = (MH_Z * 10) + 50** Reverse direction of scores for 9D and 9H, by subtracting score from 6. If 3 or more of the items are non-missing then replace any missing values as follows: Calculate the mean of the answered questions, using the number of answered questions as the denominator. Substitute the mean for any missing scores. Otherwise, if less than 3 of the items are non-missing then MH scale score is missing. The scale for these items ranges from 1 to 5 where 1=All of the time, 2=Most of the time, 3=Some of the time, 4=A little of the time, and 5=None of the time. A higher MH scale score indicates better mental health.
SF-36 PCS score	PCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT. PF1= (PF-82.62455)/24.43176; RP1=(RP-82.65109)/26.19282; BP1=(BP-73.86999)/24.00884; GH1 = (GH-70.78372)/21.28902; VT1= (VT-58.41968)/20.87823; SF1=(SF-85.11568)/23.24464; RE1= (RE-87.50009)/22.01216; MH1=(MH-75.76034)/18.04746; Raw Score = ((GH1*.24954)+(PF1* .42402)+(RP1*.35119)+ (RE1*-.19206)+(SF1*-.00753)+(MH1*-.22069)+(BP1*.31754)+ (VT1*.02877)) **PCS Summary Scale Score = (raw score *10) + 50** Raw Score is missing if one of the component scale scores is missing.

SF-36 MCS score	MCS score includes the 8 scales for GH, PF, RP, RE, SF, MH, BP, and VT. PF1= (PF-82.62455)/24.43176; RP1=(RP-82.65109)/26.19282; BP1=(BP-73.86999)/24.00884; GH1 = (GH-70.78372)/21.28902; VT1= (VT-58.41968)/20.87823; SF1=(SF-85.11568)/23.24464; RE1= (RE-87.50009)/22.01216; MH1=(MH-75.76034)/18.04746; Raw Score =((GH1*-.01571)+(PF1*-.22999)+(RP1*-.12329)+ (RE1*.43407)+(SF1*.26876)+(MH1*.48581)+(BP1*-.09731)+ (VT1*.23534)) **MCS Summary Concept Score = (raw score *10) + 50** Raw Score is missing if one of the component scale scores is missing.

Appendix E – Actigraphy Visit Windows

Visit	Window Start	Window End
Baseline	Day 1	Day 7
Week 12	Day 56	Day 112
Week 24	Day 140	Day 196
Week 36	Day 224	End of Study

Where Day 1 is first dose of study medication

Appendix F – DEXA Visit Windows

Visit	Window Start	Window End	Target Day
Baseline	Day -28	Day 8	Day 1
Week 24	Day 140	Day 196	Day 168

Where Day 1 is first dose of study medication