Statistical Analysis Plan (SAP)

Protocol Title:	A 6 month, double-blind safety extension study of MBGS205 evaluating the effects of long term treatment with BGS649 on bone mineral density
Protocol Number:	MBGS206
Protocol Version, Date	Final V2.0, 28MAR2017
ICON ID:	3082/0023
Document Version, Date:	FINAL 1.0, 20 November 2018

Prepared by:

ICON Clinical Research Services

On behalf of:

Mereo BioPharma 2 Limited

DocUUID : 8adf30e9-9345-4451-91e4-d3e0ba4d0261

SIGNATURE PAGE

[redacted]

Date
(DDMMM YYYY)

[redacted]

Date
(DDMMMYYYY)

[redacted]

Date
(DDMMMYYYY)

REVISION HISTORY

Version/Date	Version name	Section	Changes implemented
1.0 20NOV2018	Initial approved version	N/A	N/A

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 MBGS206 Final V2.0 "A 6 month, double-blind safety extension study of MBGS205 evaluating the effects of long term treatment with BGS649 on bone mineral density" dated 28MAR2017. 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.

Notes:

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

Throughout this document:

- "active study medication" refers to any BGS649 medication (i.e does not include Placebo medication taken during study MBGS205.
- MBGS205 study medication refers to any medication (BGS649 or Placebo) taken during Study MBGS205
- MBGS206 study medication refers to any medication taken during Study MBGS206

2 STUDY OBJECTIVES

The study is designed to evaluate bone mineral density in subjects after 12 months exposure with BGS649 (6 months treatment during Study MBGS205 and 6 months treatment during Study MBGS206). The study will also collect additional BGS649 safety and efficacy data on subjects that were randomised to placebo in Study MBGS205 and have 6 months active drug exposure in Study MBGS206.

2.1 Primary objective(s)

The primary objective is to evaluate the effect of 12 months treatment with BGS649 on bone mineral density (6 months treatment during Study MBGS205).

2.2 Secondary objective(s)

The secondary objectives are:

1. To follow the change of total testosterone, free and bioavailable testosterone, LH, and FSH over a total of 12 months administration of BGS649 in those on active treatment in Study MBGS205 and a total of 6 months administration of BGS649 on subjects that were switched from placebo treatment during Study MBGS205
2. Evaluation of percentage of subjects with normalised testosterone after 12 months treatment with BGS649 in those on active treatment in Study MBGS205 and total of 6 months administration of BGS649 on subjects that were on placebo treatment during Study MBGS205

2.3 Exploratory objective(s)

The exploratory objectives are:

1. To evaluate the effect of 12 months treatment with BGS649 (6 months treatment in Study MBGS205 and 6 months treatment in Study MBGS206) on:
    - Cardiometabolic parameters, body composition, physical activity, grip strength, and sleep pattern; and on sexual functioning, fatigue, and quality of life, assessed by patient reported outcomes (PROs)
    - The association between changes in PROs and changes in body composition, physical activity, grip strength, and sleep
    - The association between changes in testosterone and changes in cardiometabolic parameters, body composition, physical activity, grip strength, sleep, and PROs
2. To evaluate the 6 month efficacy data on subjects on placebo arm in Study MBGS205 treated with active study drug in Study MBGS206 (cardiometabolic parameters, body composition, physical activity, grip strength, sleep, and PROs).

2.4 Safety objective(s)

1. To evaluate the effect of BGS649 after 12 months period of treatment with active study drug (6 months treatment in Study MBGS205 and 6 months treatment in Study MBGS206) by dose group on:
   - o Bone turnover biomarkers
   - o Bone mineral density and bone biomarkers by vitamin D status (Baseline vitamin D from Study MBGS205)
   - o Oestradiol levels
   - o Testosterone outside upper limit of normal range
   - o Prostate specific antigen and haematocrit
   - o Blood Pressure
   - o Adverse events (AEs)
2. To evaluate any association between changes in bone mineral density after 12 months treatment with BGS649 with end of treatment testosterone and oestradiol levels
3. To collect safety data in subjects switching from placebo arm in Study MBGS205 and treated with 6 months active study drug in Study MBGS206

3 STUDY DESIGN

3.1 General study design

This is a 6 month, active treatment, extension study, open to subjects who have completed the full 24 week treatment period in Study MBGS205 (a Phase IIb multicentre, double-blind, dose-ranging, randomised, placebo-controlled study evaluating safety and efficacy of BGS649 in male obese subjects with HH).

Study participation will comprise of the Extension Study Baseline visit (Week 24 of Study MBGS205, where the Extension Study Baseline assessment will be performed on a same day as the EOT visit of Study MBGS205), 7 visits during the treatment period up to Week 48 (End of Treatment [EOT]), and a Follow-Up (FU) visit for safety performed at Week 60, 12 weeks after the Week 48/EOT visit.

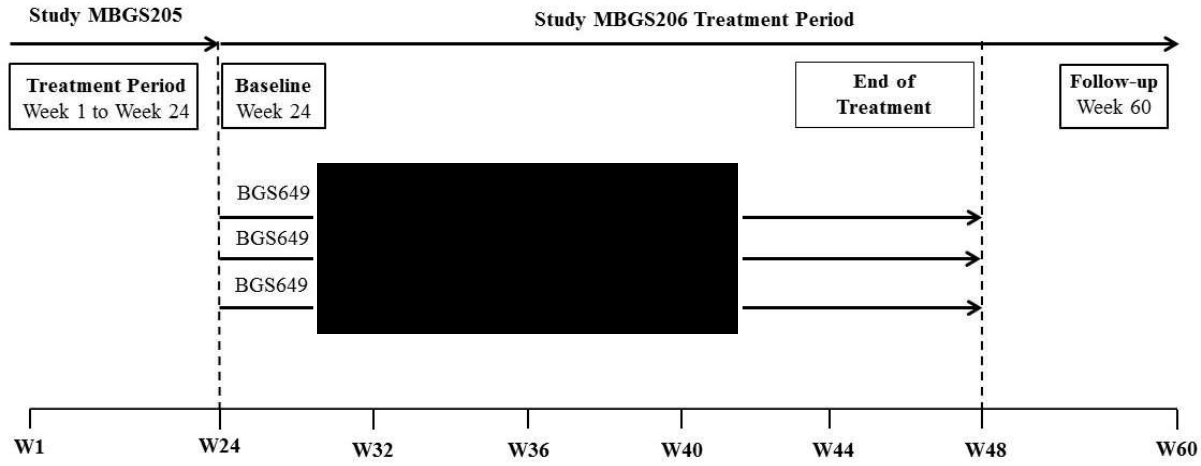
During the 24 week treatment period, subjects will take the required [redacted] time schedule of regular planned dose), starting with the first dose being taken at the baseline visit. The capsules will be taken orally by the subject with water. Subjects receiving active doses of BGS649 in Study MBGS205 will continue with the same dose in Study MBGS206; this transfer will be handled by the interactive response technology (IRT) and the blind will be maintained. Subjects receiving placebo in Study MBGS205 will be re-randomised on a 1:1:1 basis to each of the active treatment arms:

- BGS649 [redacted] capsules)
- BGS649 [redacted]
- BGS649 [redacted] capsules).

The planned interim analysis of MBGS205 determined that all active doses met efficacy and safety criteria and therefore all three doses [redacted] were progressed into MBGS206.

Enrolment will continue until 30 subjects from each of the previous active treatment arms in Study MBGS205 have entered this study.

Figure 1: Study Flow Chart



DMC=Data Monitoring Committee;

3.2 Randomisation and blinding

3.2.1 Randomisation

Subjects receiving active doses of BGS649 in Study MBGS205 will continue with the same dose in Study MBGS206 and the blind will be maintained. Subjects receiving placebo in Study MBGS205 will be randomised on a 1:1:1 basis to each of the active treatment arms. 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 of Study MBGS205. In order to maintain the blind described above, testosterone, LH, FSH and estradiol 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 of Study MBGS205 the blind will be broken and the primary efficacy analysis of Study MBGS205 completed. Given the study design, at this point the blind for MBGS206 will also be broken and the Sponsor and CRO personnel will be unblinded at the time of unblinding the MBBS205 study. However subjects, investigational staff, persons performing the assessments and data analysts (other than those described above) will remain blind to the identity of the treatments from the time of randomisation until after database lock.

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 treatment period subjects in the BGS649 groups will take the required dose [redacted] Subjects will receive three capsules per dose to deliver a total of [redacted]

Dose changes are not permitted during the treatment period.

The maximum MBGS206 study duration from baseline to end of the safety follow-up period is 36 weeks which yields to a maximum duration across both MBGS205 and MBGS206 of 60 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

			Treatment Period						
Visit		1	2	3	4	5	6	7 (EOT)	8 (FU)
Week	Assessments transferred from MBGS205	MBGS206 Baseline[1] W24	W28 (±2 D)	W32 (±2 D)	W36 (±2 D)	W40 (±2 D)	W44 (±2 D)	W48 (±2 D)	W60 (±2 D)
Confirmation of written informed consent		X							
Medical/surgical history	X								
Inclusion/exclusion criteria		X							
Treatment allocation[2]		X							
Full physical exam: general, prostate and breast exam	X				X			X	
Limited physical exam			X	X		X	X		X
Blood pressure (SBP, DBP) and heart rate	X		X	X	X	X	X	X	X
12-Lead ECG	X				X			X	
Clinical laboratory tests[3]: haematology, blood chemistry, and dipstick urinalysis[4], eGFR by Cockcroft-Gault formula, PSA	X		X	X	X	X	X	X	X
Cardiometabolic parameters[3]: HbA1c, fasting lipids (total cholesterol, LDL, HDL, triglycerides), fasting glucose and insulin, hs-CRP, HOMA-IR	X							X	
Bone turnover markers[5]	X				X			X	
Testosterone total[6]	X		X	X	X	X	X	X	X
Testosterone bioavailable, oestradiol (total), SHBG, FSH, LH[6]	X							X	
Semen PK[7]	X							X	
Plasma PK[7]	X							X	
Weight, BMI hip and waist measurement	X				X			X	
Body composition using bioimpedance[8]	X				X			X	
Actigraphy[9]	X							X	
Grip strength	X				X			X	
PROs: IIEF, BFI, SF-36 QoL PROMIS: SexFS, PGI-S[10]	X				X			X	
DEXA[11]	X							X	
Study drug dispensation		X	X	X	X	X	X		
Study drug administration[12]		[redacted]							
Study drug accountability					X				
AE assessment[13]	X		X	X	X	X	X	X	X
Concomitant medication	X		X	X	X	X	X	X	X

AE=adverse event; AESI=adverse event of special interest; BFI=Brief Fatigue Inventory; BMI=body mass index; CTx1=C-terminal telopeptide; D=day; DEXA=dual energy X-ray absorptiometry; DBP=diastolic blood pressure; ECG=electrocardiogram; eGFR=estimated glomerular filtration rate; EOT=End of Treatment; FSH=follicle stimulating hormone; FU=Follow-Up; 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; LDL=low density lipoprotein; LH=luteinising hormone; P1NP=procollagen type 1 N-propeptide; PGI-S=Patient Global Impression of Status; PK=pharmacokinetics; PRO=patient reported outcome; PROMIS=Patient-Reported Outcomes Measurement Information System; PSA=prostate-specific antigen; QoL=quality of life; RBC=red blood cell; SAE=serious adverse event; SBP=systolic blood pressure; SexFS= Sexual Function and Satisfaction; SF-36=36-item Short Form Health Survey; SHBG=sex hormone binding globulin; W=Week; WBC=white blood cell.

1 Baseline assessment will be performed on a same day as the EOT visit of Study MBGS205. As applicable, EOT assessments of Study MBGS205 will be used for Baseline assessments of Study MBGS206 and do not need to be repeated. The results of liver enzymes, PSA, haematocrit, and ECG assessments performed at the Baseline visit will not be available for eligibility assessment, therefore results of these parameters from last valid visit of Study MBGS205 will be used. This includes any of these parameters from last scheduled or unscheduled visit measure but no longer than 1 month before Baseline visit/EOT of MBGS205 .

2 Only subjects receiving placebo in Study MBGS205 to be randomised. Subjects already on active in Study MBGS205 will maintain their treatment allocation

3 All blood samples are to be collected after 8 hours fasting, after ECG and vital sign measurements have been performed

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

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

6 Has to be collected before 11 am. If the sample was inadvertently taken after this time, a retest is allowed after a discussion with the Medical Monitor.

7 All subject that are not vasectomised will be asked to provide a semen sample. Subjects that provided semen sample for PK will also have plasma for PK drawn at the assessment visit Semen sample collection is not required to be the same day as the assessment visit. Therefore, semen PK collection is allowed ± 48 hours around the plasma PK attendance visits.

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

9 Activity wrist monitors will be collected from subjects at EOT visit of the Study MBGS205. If the activity wrist monitor was not performed before EOT of Study MBGS205 and collected at EOT Study MBGS205, it can be applied at the Baseline visit of current study and worn for 7 c consecutive days and collected at Visit 2. An activity wrist monitor will also be worn for 7 continuous days after Visit 7 (EOT) and collected at Visit 8 (FU)

10 Some questionnaires will be new for some subjects and completed first time at the Baseline visit

11 For subjects that discontinued early, EOT DEXA scan will be performed only if they reached minimum Week 36 (Visit 4) active treatment period. In order to accommodate scheduling, EOT DEXA can be performed within 5 days before the EOT visit

12 First dose of study drug will be administered at the site at Baseline visit

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

4 STUDY ENDPOINTS

4.1 Primary Endpoint

Percentage change in lumbar bone mineral density measured by DEXA (g/cm2) from Baseline in Study MBGS205 to Week 48 in Study MBGS206 by dose group in subjects randomised to active treatment.

4.2 Secondary Endpoints

4.2.1 Secondary Safety Endpoints

1. In subjects that were randomized to active treatment in Study MBGS205 by dose group:
    - Percentage change in hip (total and femoral neck) bone mineral density measured by DEXA (g/cm2) from Baseline in Study MBGS205 to Week 48 in Study MBGS206
    - Percentage change in bone turnover markers (C-terminal telopeptide [CTx1], osteocalcin, bone alkaline phosphatase, and procollagen type 1 N-propeptide [P1NP]) from Baseline in Study MBGS205 to Week 48 in Study MBGS206
    - Proportion of subjects with T-scores $\leq$ -2.5 at Week 48
    - Percentage change in bone density and bone biomarkers adjusted for vitamin D deficiency from Baseline in Study MBGS205 to Week 48 in Study MBGS206
    - Change of oestradiol (absolute and percentage) from Baseline in Study MBGS205 to Week 48
    - Proportion of subjects that overshoot testosterone (total testosterone above 1000 ng/dl [35 nmol/L], from first dose of study drug in Study MBGS205 until study completion)
    - Change in PSA from Baseline in Study MBGS205 to Week 48 in Study MBGS206
    - Change in haematocrit from Baseline in Study MBGS205 to Week 48 in Study MBGS206
    - Change in blood pressure (systolic and diastolic) Baseline in Study MBGS205 to Week 48 in Study MBGS206
    - Treatment emerged AE, SAE, and AESI (from first dose of study drug in Study MBGS205 until study completion)
    - Relationship between change in bone mineral density from Baseline in Study MBGS205 to Week 48 in Study MBGS206 and absolute levels of oestradiol and total testosterone at Week 48 in MBGS206

2. In subjects that were randomized to placebo in Study MBGS205 by dose group:
    - Percentage change in lumbar bone mineral density measured by DEXA (g/cm2) from Baseline in Study MBGS206 to Week 48
    - Percentage change in hip (total and femoral neck) bone mineral density measured by DEXA (g/cm2) from Baseline in Study MBGS206 to Week 48
    - Percentage change in bone turnover markers (CTx1, osteocalcin, bone alkaline phosphatase, and P1NP) from Baseline in Study MBGS206 to Week 48
    - Change of oestradiol (absolute and percentage) from Baseline in Study MBGS206 to Week 48
    - Proportion of subjects with T-scores ≤ -2.5 at Week 48 of Study MBGS206
    - Proportion of subjects that overshoot testosterone (total testosterone above 1000 ng/dl [35 nmol/L] from first dose of study drug in Study MBGS206 until study completion)
    - Change in PSA from Baseline in Study MBGS206 to Week 48
    - Change in haematocrit from Baseline in Study MBGS206 to Week 48
    - Change in blood pressure (systolic and diastolic) Baseline in Study MBGS206 to Week 48
    - Treatment emerged AE, SAE, and AESI (from first dose of study drug in Study MBGS206 until study completion).

4.2.2 Secondary Efficacy Endpoints

1. In subjects that were randomized to active treatment in Study MBGS205 by dose group:
    - Percentage of subjects that have normalisation of total testosterone at Week 48 in Study MBGS206
    - Change in total, free and bioavailable testosterone from Baseline in Study MBGS205 to Week 48 in Study MBGS206
    - Change of LH and FSH from Baseline in Study MBGS205 to Week 48 in Study MBGS206
2. In subjects that were randomized to placebo in Study MBGS205 by dose group:
    - Percentage of subjects that have normalisation of total testosterone at Week 48 of Study MBGS206
    - Change in total, free and bioavailable testosterone from Extension Study Baseline to Week 48 in Study MBGS206
    - Change of LH and FSH from Extension Study Baseline to Week 48 in Study

4.3 Exploratory endpoint(s)

1. In subjects that were randomized to active treatment in Study MBGS205 by dose group:
    - Change in exploratory parameters from Baseline in Study MBGS205 to Week 48 in Study MBGS206 by dose group:
        - Body composition (weight, BMI, waist and hip circumference, and parameters measured by bio-impedance)
        - Markers of cardiometabolic disease (blood pressure, lipid profile, HbA1c, glucose and insulin, hs-CRP, and homeostatic assessment of insulin resistance [HOMA-IR])
        - Physical activity, sleeping pattern, and strength (measured by wrist worn monitors and grip strength measurement)
        - Total and domain scores on PRO measures (Sexual function (IIEF, PROMIS SexSF), Energy/fatigue (BFI, PROMIS Fatigue), SF-36
    - The relationship between changes in PROs and changes in body composition, physical activity, grip strength, and sleep
    - The relationship between changes in testosterone and changes in cardiometabolic parameters, body composition, physical activity, grip strength, sleep, and PROs.
2. In subjects that were randomized to placebo in Study MBGS205 by dose group:
    - Change in exploratory parameters from Baseline in Study MBGS206 to Week 48 in Study MBGS206 by dose group:
        - Body composition (weight, BMI, waist and hip circumference, and parameters measured by bio-impedance)
        - Markers of cardiometabolic disease (blood pressure, lipid profile, HbA1c, glucose and insulin, hs-CRP, and homeostatic assessment of insulin resistance [HOMA-IR])
        - Physical activity, sleeping pattern, and strength (measured by wrist worn monitors and grip strength measurement)
        - Total and domain scores on PRO measures (Sexual function (IIEF, PROMIS SexSF), Energy/fatigue (BFI, PROMIS Fatigue), SF-36

5 SAMPLE SIZE AND POWER

The statistical hypothesis is the percentage decrease from Baseline in MBGS205 to EOT in MBGS206 in lumbar bone mineral density measured in g/cm2 by DEXA within a dose group is greater than [redacted]:

H0: [redacted]

H1:

The null hypothesis will be rejected for a dose group and the endpoint considered met if the lower bound of the 2-sided 95% confidence interval (CI) for the percentage decrease from Baseline in lumber spine is greater than [redacted].

Assuming a 2-sided test at [redacted] significant level and a standard deviation (SD) of [redacted] g/cm2, to achieve 90% power N=[redacted] subjects in each dose group would be needed to detect a [redacted] decrease from Baseline in bone mineral density at 48 weeks.

Enrolment will continue until [redacted] subjects from each of the previous active treatment arms in Study MBGS205 have entered this study. Subjects previously treated with placebo will be enrolled until target subject population is met in all active arms.

6 ANALYSIS POPULATIONS

All safety endpoints will be analysed using the safety population. All efficacy endpoints will be analysed using the intention to treat (ITT) population.

Further safety and tolerability data 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 MBGS206 study medication, and
3. Provide a baseline (see section 8.1.5 for full definition of baseline) primary endpoint safety value (lumbar spine) and at least 1 available evaluation of primary endpoint safety value (lumbar spine) post- Baseline in MBGS206.

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

6.1.1 Modified Inention-To-Treat Population (MITT)

The mITT population includes all subjects who are included in the ITT but were not enrolled from site 1156

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 MBGS206 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.

6.2.1 Modified Safety Population (MSafety)

The modified safety population includes all subjects who are included in the Safety Population but were not enrolled from site 1156.

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 Completer Population (CP)

The completer population includes all subjects in the safety population who completed the MBGS206 study and subsequently provided a Week 48 primary endpoint point assessment.

Subjects will be analysed according to the treatment to which they received.

6.4 PK population

For the summary of semen concentration of BGS649, all subjects who received at least 1 administration of MBGS206 study medication and have at least 1 quantifiable concentration will be included in the PK population.

6.5 Other Populations Defined for Tables and Listings

For the purposes of tables and listings a further population will be utilized:

- 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 active study medication (Study Day 1, either from MBGS205 or MBGS206) and calculated as:

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

Follow-up time

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

(date of last contact - date of first dose of MBGS206 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 of MBGS205 - 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 of MBGS205 – 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).

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 - **International Index of Erectile Function (IIEF)**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 - **Patient-Reported Outcomes Measurement Information System(®) Sexual Function and Satisfaction (PROMIS(®) SexFS)**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 - **PROMIS(®) Fatigue Short Form**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 - **PROMIS(®) Fatigue Short Form**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 - **36-Item Short Form Health Survey (SF-36)**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 (if not collected as part of the EOT assessment for MBGS205), and Visit 7 (EOT). Only data collected during the specified visit windows (See Appendix E – Actigraphy Visit Windows) 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 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 active study medication (from either MBGS205 or MBGS206).

TEAEs will also be split between AEs that start during the subject's active treatment period (for subjects randomised to active in MBGS205 this includes AE's starting in study MBGS205) and those that start on or after the first day of the MBGS206 follow-up period. A further category of TEAEs that started in MBGS206 but prior to the MBGS206 follow up period will also be utilised for summary tables (TEAEs started in MBGS205 would be excluded from this category).

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 active study medication over both MBGS205 and MBGS206 (number of days) will be defined as:

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

The duration of Exposure to MBGS206 study medication (number of days) will be defined as:

(date of last dose of MBGS206 study medication – date of first dose of MBGS206 study medication) + 1

Treatment Compliance

Overall (MBGS205 and MBGS206) active treatment compliance will be defined as:

[redacted]

MBGS206 treatment compliance will be defined as:

[redacted]

7.1.4.3 Prior, Concomitant and Follow-up Medications/Procedures

Medications and Procedures will be assigned as being prior to active study medication, concomitant with active 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 active study medication (for subjects randomised to Placebo in MBGS205, this includes medication/procedures starting in study MBGS205), the medication/procedure will be assigned as being prior to active study medication. Otherwise, the medication/procedure will be assigned as being concomitant with active study treatment (for subjects randomised to active in MBGS205, this includes medications/procedures starting in study MBGS205) unless the start date of the medication/procedure is after the MBGS206 end of treatment date, when it will then be classified as occurring in the follow-up

phase.

In addition medications will be flagged as being concomitant during the MBGS206 treatment period (i.e medication started on or after the first dose of MBGS206 study medication but before the follow up period or was ongoing at MBGS206 baseline).

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):

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

This derivation will be used as a supporitve analysis to the eGFR as derived by the lab.

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.

For specified endpoints a last observation carried forward (LOCF) imputation will be utilised. In this approach subjects with a missing observed value at a visit will be imputed with the last non-missing, post-baseline value.

For each continuous exploratory 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 MBGS206 study medication, then it will be assumed that the medication/procedure began after the start of MBGS206 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 7 date, then it will be assumed the medication/procedure began on or before the Visit 7 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 active study medication date.

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, Completers and ITT populations, unless otherwise stated.

Analyses of the primary endpoint and secondary safety BMD endpoints will be performed on ITT and completer population. Analysis of all other safety endpoints will use the safety population.

The secondary efficacy endpoints and exploratory 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

Statistics will be displayed for the following treatment regimens in the below order:

- BGS649 [redacted]
- BGS649 [redacted]
- BGS649[redacted]
- Active Total (Includes all continuous active groups where applicable)
- Placebo -> BGS649 [redacted]
- Placebo -> BGS649 [redacted]
- Placebo -> BGS649 [redacted]
- Plabebo-> Active Total (Includes all Placebo-> Active Groups where applicable)
- Overall Total (Includes all groups where applicable)

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. 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 is defined as the last non-missing value collected on or before the first active study treatment administration (from either Study MBGS205 or MBGS206). Therefore subjects randomised to active treatment in Study MBGS205, baseline will be the last value prior to first dose of study medication in Study MBGS205. For subjects randomised to placebo in Study MBGS205, baseline will be the last value prior to first dose of study medication in Study MBGS206 (expected to be from the MBGS206 baseline visit)..

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

For DXA parameters, assessments at MBGS205 baseline can be taken up to 8 days 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 subjects randomised to active treatment in Study MBGS205).

For selected renal parameters change from MBGS205 Screening and Baseline 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.

8.1.7 Data listings

All relevant subject data collected during the MBGS206 study, including those derived will be presented in individual subject data listings. Selected data from study MBGS205 will also be presented to aid in interpretation (e.g MBGS205 baseline) 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 entered the extension study will be included in a summary of subject accountability. The number of subjects entering the extension study, the frequency and percentage of subjects randomised, in the Safety population, in the ITT 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 treatment period (up to and including Visit 7) 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.

8.3 Protocol deviations

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

All protocol deviations identified during MBGS206 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 captured at the start of study MBGS205 will be listed and summarised for subjects entering the extension study by treatment regimen and overall. This will include age, sex, race, weight, height and BMI. Separate summaries will be produced using the Safety and ITT 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 captured at the start of study MBGS205 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 from study MBGS205 will be tabulated.

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 ITT Population.

8.4.3 Medical history

A summary of all medical history captured from Study MBGS205 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.

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

All medications (taken across both MBGS205 and MBGS206) will be coded using the World Health Organization (WHO) Drug Dictionary (March 2016 or later) and will be summarised by Anatomical Therapeutic Chemical (ATC) classification level 4 and Preferred Name by treatment regimen and overall for the Safety population.

Summaries of Prior medications to active study medication, concomitant to active study medication (i.e across both MBGS205 and MBGS206), concomitant to MBGS206 study medication and follow-up medications will be presented.

8.4.5 Concomitant Procedures

Concomitant procedures to active study medication (i.e across both MBGS205 and MBGS206) and concomitant procedures to MBGS206 study medication 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) for active study medication and just MBGS206 study medication will be categorised in intervals and summarised descriptively by treatment regimen on the Safety population using descriptive statistics.

8.5.2 Treatment compliance

The overall treatment compliance (in %) of active study medication and just MBGS206 study medication 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 secondary and exploratory efficacy endpoints.

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

Analyses will be conducted separately for subjects that were randomised to active treatment and subjects that were randomised to placebo treatment in study MBGS205.

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 continuous parameters, the change from baseline will be summarised by treatment arm and visit. A single sample 2-sided t-test conducted at the 5% significance level will be used to analyse the change from baseline to Week 48 within each treatment arm. In addition, the change from baseline will be assessed with a mixed model repeated measure (MMRM) analysis. 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 will be based on least-squares means.

Associations between changes in PROs and changes in body composition, physical activity, sleep, and grip strength will be assessed with correlation coefficients and with separate linear regression models in which changes in PRO measure total or subscale scores will be taken as the predictor variable, and baseline score as a covariate. The associations between changes in testosterone and changes in PROs, body composition, physical activity, grip strength, and sleep will be assessed in the same way. Correlation coefficients, Beta coefficients, standard errors, and 95% CIs will be presented.

8.6.1.1 Multiplicity

Treatment regimens will be assessed by pairwise comparisons. No adjustment will be made for multiple comparisons and all analyses are considered exploratory.

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 endpoint

The Primary Endpoint is defined as the change in lumbar bone density measured by DEXA from baseline in Study MBGS205 to Week 48 in Study MBGS206 by dose group in subjects randomised to active treatment in Study MBGS205.

DXA Scan assessments will only be performed at the EOT visit (for subjects that reached at least Week 36) and will be over-read by IMI. The DEXA over-read data will be used for the primary endpoint statistical analysis.

Descriptive statistics (observed values, changes from baseline and percentage change from baseline) in lumbar bone mineral density will be presented for each treatment regimen and time-point. A single sample 1-sided t-test conducted at the 2.5% significance level will be used to analyse percentage change from Baseline (MBGS205) to Week 48 (MBGS206) LOCF in each treatment arm to test the null hypothesis that the percentage decrease from Baseline is ≤ 3%. P-values and associated CI will be presented.

Only subjects included in the safety population that have a post-Baseline DEXA assessment in MBGS206 and were randomised to active treatment in Study MBGS205 will be included in this analysis. The anlysis will be repeated for subjects included in the completer population.

Observed and percentage change from baseline values at each visit will be presented in a line graph by treatment group.

8.6.3 Analysis of secondary efficacy endpoint(s)

All analysis of secondary efficacy endpoints will be performed on the ITT population.

8.6.3.1 Testosterone Normalisation (Primary Definition)

The number and proportion of subjects with normal testosterone levels (testosterone value is between 300-1000 ng/dL inclusive) will be summarised by treatment regimen at each visit using the total testosterone assessed by ICON central lab. Observed values and LOCF values will be presented. In addition, a P-value for the comparison between each dose will be provided using Fisher's exact test.

8.6.3.2 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 primary definition of testosterone normalisation.

8.6.3.3 Change from Baseline in Total Testosterone

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

A single sample 2-sided t-test conducted at the 5% significance level will be used to analyse the percentage change from Baseline at Week 48 within each treatment arm. The p-value and associated confidence internal will be presented.

In addition the percentage change from baseline in Total Testosterone up to Week 48 for subjects randomised to active in Study MBGS206 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 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.

The above analyses will be repeated for subjects that were randomised to placebo in Study MBGS205 by analysing the change from MBGS206 baseline (Week 24) to Week 48.

8.6.3.4 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 Testosterone normalisation.

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:

- >1000 - <=1500ng/dl
- >1500 – <1800 ng/dL
- =>1800 - <= 2500 ng/dL
- >2500 ng/dL

8.6.3.5 Change from Baseline in Free and Bioavailable Testosterone

Descriptive summary tables will be presented for Free and Bioavailable Testosterone values and change from baseline in Free and Bioavailable Testosterone by treatment regimen and visit.

The change from baseline in Free and Bioavailable Testosterone will be analysed in the same manner as the change from baseline in total testosterone for subjects that were randomised to active treatment in study MBGS205 (see Section 8.6.3.3). For subjects that were randomised to placebo in study MBGS205, the change from baseline will be analysed using an ANCOVA model with treatment as a fixed effect and 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 $\alpha = .05$.

8.6.3.6 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 48 will be analysed in the same manner as the change from baseline in bioavailable testosterone (see Section 8.6.3.5). Analyses will be conducted separately for subjects that were randomised to active treatment (MMRM analysis) and subjects that were randomised to placebo (ANCOVA analysis) in study MBGS205.

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

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 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 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.

The change from baseline in each parameter up to Week 48 will be analysed via t-test and MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3). Analyses will be conducted separately for subjects that were randomised to active treatment and subjects that were randomised to placebo treatment in study MBGS205.

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 48 will be analysed via t-test and MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3). Analyses will be conducted separately for subjects that were randomised to active treatment and subjects that were randomised to placebo treatment in study MBGS205.

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 48 as described above will be presented.

8.6.4.2 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 48 will be analysed via t-test and MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3). Analyses will be conducted separately for subjects that were randomised to active treatment and subjects that were randomised to placebo treatment in study MBGS205.

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).

8.6.4.3 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 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 in moderate to very vigorous activity, total daily steps and total activity counts to Week 48 will be analysed in the same manner as the change from baseline in bioavailable testosterone (see Section 8.6.3.5). Analyses will be conducted separately for subjects that were randomised to active treatment (MMRM analysis) and subjects that were randomised to placebo (ANCOVA analysis) in study MBGS205.

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 48 will be analysed in the same manner as the change from baseline in bioavailable testosterone (see Section 8.6.3.5). Analyses will be conducted separately for subjects that were randomised to active treatment (MMRM analysis) and subjects that were randomised to placebo (ANCOVA analysis) in study MBGS205.

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 48 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3). Analyses will be conducted separately for subjects that were randomised to active treatment and subjects that were randomised to placebo treatment in study MBGS205.

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 48 as decribed above will be presented.

8.6.4.4 Change in PRO measures

For some PRO Measures (PROMIS SexFS, PROMIS Fatigue Short Form and PGI-S), not all MBGS205 protocol versions mandated their collection. Thus for subjects originally enrolled into

study MBGS205 using protocol version 2.0 MBGS206, baseline will be the first time these measures were collected. To handle such cases, these subjects will be excluded from the change from MBGS205 baseline to Week 48 analysis (if randomised to active treatment in Study MBGS205), however will still be included in the descriptive summary tables.

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 48 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3), however 90% confidence intervals (instead of 95%) will be presented for all inferences. Analyses will be conducted separately for subjects that were randomised to active treatment and subjects that were randomised to placebo treatment in study MBGS205. A figure presenting the least square mean change from baseline in each sexual function domain at each visit by treatment regimen will also be provided.

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 to Week 48 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 48 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 48 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 48 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 48 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.

8.7 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 endpoints specified below.

8.7.1 DXA Scans

DXA Scan assessments will only be performed at the EOT visit (for subjects that reached at least Week 36) and will be over-read by IMI. Both the site assessed and the 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) in hip (total and femoral neck) bone density 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 48 for subject randomised to active treatment in Study MBGS205 will analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3).

Change from baseline in bone density (in both the lumbar spine and hip) to Week 48 for subjects randomised to placebo in Study MBGS205 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.

8.7.2 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 48 will be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3). Analyses will be conducted separately for subjects that were randomised to active treatment and subjects that were randomised to placebo treatment in study MBGS205.

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.

8.7.3 Oestradiol and testosterone/oestradiol ratio

Descriptive summary tables will be presented for oestradiol and testosterone/oestradiol ratio values, change from baseline values and percentage change from baseline values by treatment regimen and visit. High sensitivity Oestradiol values collected during MBGS205 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 48 will be analysed in the same manner as the change from baseline in bioavailable testosterone (see Section 8.6.3.5). Analyses will be conducted separately for subjects that were randomised to active treatment (MMRM analysis) and subjects that were randomised to placebo (ANCOVA analysis) in study MBGS205. 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.

8.7.4 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 active treatment period (across both MBGS205 and MBGS206), MBGS206 treatment period, MBGS206 follow-up period and overall (i.e including those that started in both the active 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.

8.7.5 Laboratory evaluations

8.7.5.1 Clinical laboratory

Haematology, blood chemistry, and urinalysis assessments will be conducted at each visit during the treatment phase and Week 60.

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 48 will also be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3).

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.

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

8.7.5.2 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)

8.7.6 Vital signs

Vital sign assessments will be performed at each visit.

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 48 will also be analysed via MMRM in the same manner as the change from baseline in total testosterone (see Section 8.6.3.3).

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

8.7.7 Physical examinations

A full physical examination will be performed at Week 36 and Week 48 whereas a limited physical examination (cardiovascular system and lower extremities oedema only) will be performed at all other.

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.

8.7.8 Electrocardiograms

12-Lead ECG assessments will be performed at Week 36 and Week 48.

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.

8.8 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 dat

a 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. This would also include situation when first testosterone value that meets this threshold of ≥ 1500 ng/dL [52 nmol/L] was measured during MBGS205 and consecutive testosterone during study MBGS206.

8.8.1.2 Interim Analysis

No interim analysis is planned for this study.

8.8.2 PK analyses

Plasma and Semen PK will be collected at Week 48/EOT only.

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

8.8.3 Association Testing

To determine any correlation between endpoints after 12 months active treatment, various association testing will be conducted using the ITT Population in subjects that were randomised to active treatment in Study MBGS205 at Week 48. For the below sets of endpoints separate linear regression models in which changes in PRO measure total or subscale scores will be taken as the predictor variable, associated PRO baseline score as a covariate and the endpoint to be tested against as the regressor variable. Pearson's Correlation coefficients, coefficients of determination (R^2 statistic) Beta coefficients, standard errors, and 95% CIs will be presented. In addition scatter graphs will be generated annotated with the coefficient of determination (R^2 statistic):

- Change in PRO measures vs Change in Body Composition Parameters
- Change in PRO measures vs Change in Bio Impedance Parameters
- Change in PRO measures vs Change in Physical Activity Parameters
- Change in PRO measures vs Change in Grip Strength
- Change in PRO measures vs Change in Sleep Parameters
- Change in PRO measures vs Change in Oestradiol
- Change in PRO measures vs Change in Testosterone/oestradiol Ratio

Similar association testing will be conducted on the below endpoints using total testosterone as the predictor variable:

- Change in Total Testosterone vs Change in Cardiometabolic parameters
- Change in Total Testosterone vs Change in Body Composition parameters
- Change in Total Testosterone vs Change in Bio Impedance Parameters
- Change in Total Testosterone vs Change in Physical Activity Parameters
- Change in Total Testosterone vs Change in Grip Strength
- Change in Total Testosterone vs Change in Sleep Parameters
- Change in Total Testosterone vs Change in PRO measures

Association testing will also be conducted on the below

- Change in systolic blood pressure vs Total Testosterone, Oestradiol and Testosterone/Oestradiol ratio at Week 48
- Change in diastolic blood pressure vs Total Testosterone, Oestradiol and Testosterone/Oestradiol ratio at Week 48
- Change in Creatinine vs Blood Pressure (Absolute systolic and diastolic blood pressure) at Week 48
- Change in Creatinine vs Change in Blood Pressure (Systolic and Diastolic blood pressure) at Week 48

8.8.4 Sensitivity Analysis of Site 1156

To assess the impact of site 1156 on key endpoints the below analyses will be conducted utilising the modified populations (see section 6.0) which removes subjects enrolled from site 1156:

- Analysis of Proportion of Subject Achieving Normalisation of Total Testosterone (300-1000 ng/dl). Analysis will be conducted on the Central lab testosterone using LOCF imputation.
- Summary of Change in LH and FSH
- Summary of Vitals signs Parameters and Change from Baseline
- Summary of Blood Pressure Categorisation
- Summary of Body Composition

8.9 CSR

The primary efficacy analysis will be performed when all data has been collected.

9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

The below are changes to the study protocol analysis:

The Percentage change in lumbar bone mineral density measured by DEXA (g/cm2) from Baseline in Study MBGS206 to Week 48 for subjects that were randomised to placebo in Study MBGS205 has been added as a secondary safety endpoint. This was inadvertently absent from the 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.

Following data integrity concerns at site 1156, sensitivity analyses were added to assess the impact of the site on key parameters.

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	Very Much	5
		Quite a bit	4

	satisfaction with your sex life?	Somewhat	3
		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 gainin 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

For subjects randomised to active medication in MBGS205

Visit	Window Start	Window End
Baseline	Day 1	Day 7
Week 12	Day 56	Day 112
Week 24	Day 140	Day 196
Week 48	Day 308	End of Study

For subjects randomised to Placebo in MBGS205

Visit	Window Start	Window End
Baseline	Day - 28	Day 7
Week 48	Day 140	End of Study

Appendix F – DEXA Visit Windows

For subjects randomised to active medication in MBGS205

Visit	Window Start	Window End	Target Day
Baseline	Day -28	Day 8	Day 1
Week 24	Day 140	Day 196	Day 168
Week 48	Day 308	End of Study	Day 336

For subjects randomised to Placebo in MBGS205

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