[redacted]

Final Analysis
Statistical Analysis Plan (SAP)

Protocol Title:	Safety and Efficacy of BIO-101 175 mg b.i.d. and 350 mg b.i.d. 26-week oral administration to patients suffering from age-related SARcopenia, including sarcopenic obesity, Aged ≥ 65 years and at risk of mobility disability. A double-blind, placebo controlled, randomized INTerventional Clinical Trial (SARA-INT).
Protocol Number:	BIO101-CL03
Protocol Version, Date	Final 1.2.11, 20 APRIL 2020
[redacted]	[redacted]
Document Version, Date:	Final 2.0, 04Jun2021

Prepared by:

[redacted]

On behalf of:

BIOPHYTIS S.A.

14 Avenue de l'Opéra, 75001 Paris France

DocUUID : b338e203-9d44-47ab-98af-c8118371eabb

SIGNATURE PAGE

Prepared at [redacted] by:

[redacted] 08:43:057+0000
REASON: I approve this document
783c07f2-8b00-491e-99d4-00e5e051d11e

[redacted]

Date
(DD Mmm YYYY)

Reviewed at [redacted] by:

[redacted] 08:41:024+0000
REASON: I approve this document
b806ff26-158d-4c7c-94b3-94cbd07de241

[redacted]

Date
(DD Mmm YYYY)

Approved at BIOPHYTIS by:

[redacted] 09:24:048+0000

REASON: I approve this document

b40da377-6548-4bce-9a84-d9a56839c3ad

[redacted] Date (DD Mmm YYYY)

[redacted]

7b20f1b6-45e9-42a1-99db-8ed31d3f6f42

[redacted]r Date (DD Mmm YYYY)

REVISION HISTORY

Version/Date	Version name	Section	Changes implemented
Final 1.0/18May2021	Initial Version 1.0	N/A	N/A
Final 2.0/04JUN2021	Final 2.0	Section 1 Section 6.4	In section 1 • Update protocol version to 1.2.11 In section 6.4 • State Non-compliance with study treatment include missing compliance; For treatment interruption of > 4 weeks includes interruption between visits or within one visit • Change to Failure to perform the 400MWT within 2 week of last dose (14 days or less after last dose) • Change to Deviation from Inclusion criteria 3, 4, 5, 6, 7 • Add Deviation from Exclusion criteria 1, 4, 6, 7d, 7e, 7f, 7g, 7h • Add Any other protocol deviation identified as major protocol deviation from above mentioned source of information.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

The following abbreviations will be used within this SAP.

Abbreviation or special term	Explanation
400MW	400-meter walking test
6MWT	6-minute (distance) walk test
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALM	Appendicular Lean body Mass
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
b.i.d	Twice a day
BMI	Body Mass Index
CFR	Code of Federal Regulations
CIRS	Cumulative Illness Rating Scale
CI	Confidence Interval
cm	Centimetre
CMH	Cochran Mantel-Haenszel
CNS	Central Nervous System
COVID-19	COronaVIrus Disease 2019
CRF	Case Report Form
CSR	Clinical Study Report
DEXA	Dual Energy X-ray Absorptiometry
DNA	Deoxyribonucleic Acid
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EU	European Union
FAS	Full Analysis Set
ICH	International Conference on Harmonisation

kg	Kilogram
LS	Least-squares
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model for Repeated Measurements
OR	Odds Ratio
PAT-D	Pepper Assessment Tool for Disability
PF-10	10 item Physical Function domain at SF-36
PK	Pharmacokinetics
PRO	Patient Reported Outcome
PT	Preferred Term
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SARA	SARcopenia and sarcopenic obesity in patients Aged ≥ 65 years
SarQoL	Sarcopenia Quality of Life (questionnaire)
SAS	Statistical Analysis System
SD	Standard Deviation
SF-36	Short Form-36 (quality of life questionnaire)
SF-MNA	Short Form Mini-Nutritional Assessment
SOA	Schedule of Activities
SOC	System Organ Class
SCPT	Stair Climb Power Test
SPPB	Short Physical Performance Battery
TEAE	Treatment Emergent Adverse Event
TFLs	Tables, Figures and Listings
TSD-OC	Test SIO Disabilità Obesità Correlata (obesity disability questionnaire)
US	United States
WHODDE	World Health Organization Drug Dictionary Enhanced

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 BIO101-CL03, Version 1.2.11, titled "Safety and Efficacy of BIO-101 175 mg b.i.d. and 350 mg b.i.d. 26-week oral administration to patients suffering from age-related SARcopenia, including sarcopenic obesity, Aged ≥65 years and at risk of mobility disability. A double-blind, placebo controlled, randomized INTerventional Clinical Trial (SARA-INT)" dated 18Oct2019 for final analysis. The list of Tables, Figures and Listings and templates for the 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 Harmonization (ICH) E3 and E9 guidelines [1, 2].

All data analyses and generation of TFLs will be performed using Statistical Analysis System (SAS®) version 9.3 or higher.

The statistical analyses plan (SAP) will be finalized & signed-off before database lock and unblinding.

2 STUDY OBJECTIVES

2.1 Primary objective

The primary objective of this study is to evaluate the effects of two daily doses of BIO101 versus placebo on mobility functions as measured by the gait speed during the 400-meter walking (400MW) test.

2.2 Secondary objectives

The key secondary objectives of this study are:

- To evaluate the effect of BIO101 on physical function from the study subjects' perspective using an adapted patient reported outcome (PRO).
- To evaluate the effect of BIO101 on muscle strength using the handgrip strength test.

The other secondary objectives of this study are:

- To assess changes in body composition and specifically on appendicular lean body mass (ALM), which is an expression of sarcopenia.
- To estimate the change of 400MW test as a dichotomous variable, for possible use in further studies.
- To evaluate the effects of two daily doses of BIO101 versus placebo on mobility functions as measured by the distance measured during the 6 minutes walking test.
- To estimate the effect on muscle strength.
- To assess the overall change of the SPPB total score as a cumulative expression of the physical frailty status.
- To assess muscle functions using a simplified function test, i.e. the change in time needed to complete the chair stand test.
- To estimate the effect using a sarcopenia-specific PRO, in preparation of future studies.

2.3 [redacted]

[redacted]

3 STUDY DESIGN

3.1 General study design

This study is a three arm (2 doses versus placebo), randomized, double-blind, placebo-controlled parallel design phase 2 clinical trial, with total study duration of 33-35 months.

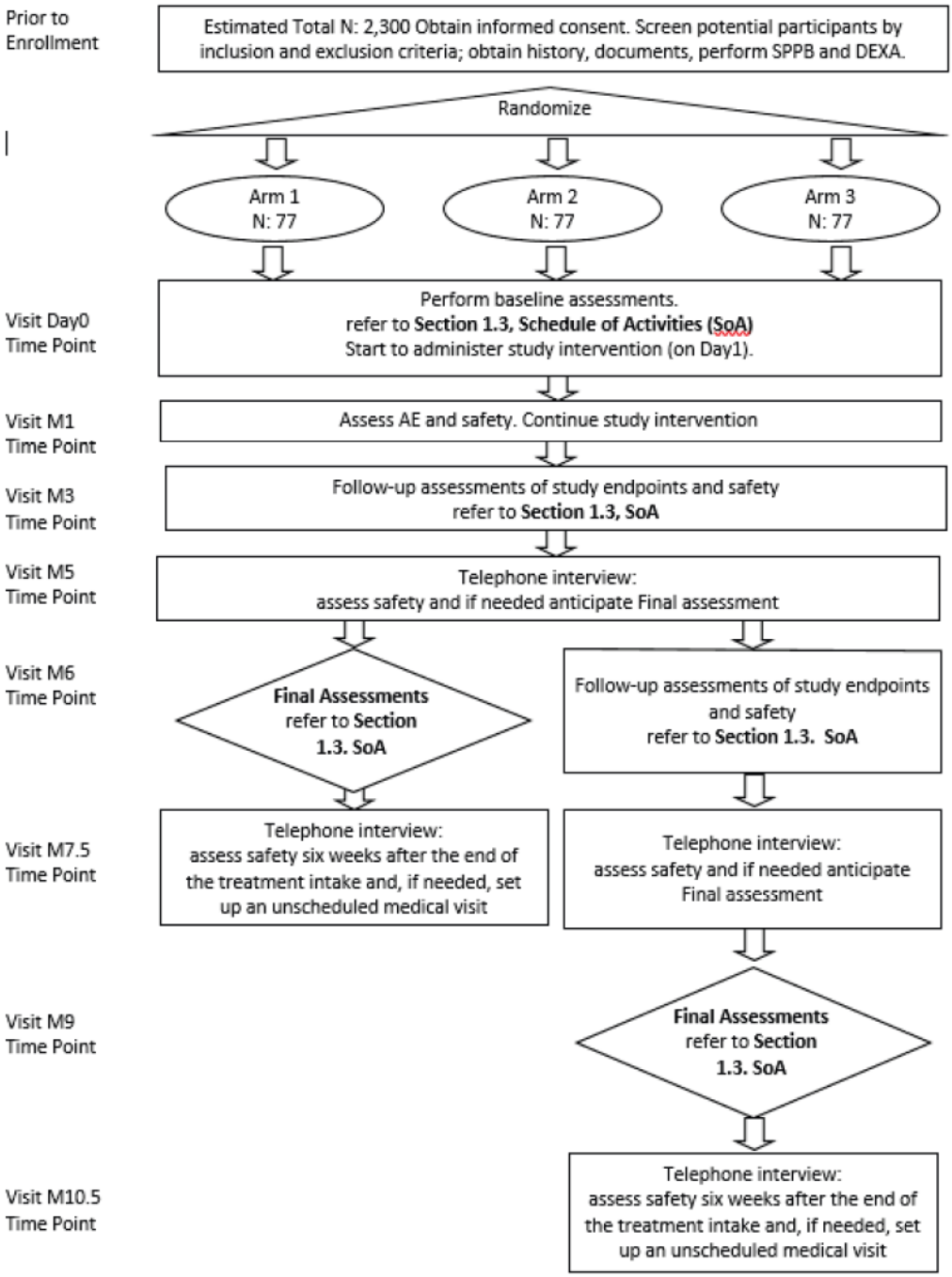
This comparative clinical trial evaluates the effects of a 6-month treatment duration, based on the hypothesis that physical function of sarcopenic, older subjects with an initial degree of mobility disability (SPPB≤8) will be meaningfully improved after 6 months of oral treatment with BIO101 (in at least one of the two tested doses) with respect to placebo.

This multi-site study will enroll 231 community dwelling older adults (men or women ≥65 years) reporting a loss of physical function over the previous 6-12 months and considered at risk of mobility disability to undergo screening tests for inclusion in the study. Included subjects will be randomized in a 1:1:1 ratio, for one of the 3 arms of treatment in a blinded manner.

Safety of study patients and investigators should be the priority during COVID-19 outbreak. Only ePROs (SF36, SARQoL, TSD-OC and Pat-D) will be assessed during this period until further guidance. An extension of the treatment period for a total period of 9 months of intervention will enable the full assessment of the safety and efficacy of the investigational product for a significant part of the population randomized in this study, while staying in compliance with good clinical practice (GCP), and minimizing risks to trial integrity during the COVID-19 pandemic.

Bias are minimized by the use of identical therapeutic units whose number is automatically assigned via an eCRF embedded algorithm after the Investigator confirms randomization of a subject. The Study Flow Chart is presented in **Figure 1**.

Figure 1: Study Flow Chart



3.2 Randomization and blinding

At each study center, subjects who are eligible to enter the double-blind treatment period will be randomized to one of the three treatment arms (BIO101 175mg, BIO101 350mg or placebo), in a 1:1:1 ratio through [redacted] by [redacted] Randomization will be stratified by gender and study center.

The staff at the study center, the Sponsor and its representatives, and all subjects will be blinded to the identity of the investigational product.

The DSMB will be unblinded for the DSMB reviews. Full unblinding of the study will occur after database lock unless there is medical emergency. The study blinding may only be broken for an individual subject in the case of an emergency where the investigator considers it essential to know what treatment the subject was receiving. The monitor shall be notified promptly if a treatment code envelope is opened. It is recommended that the Investigator contacts the Sponsor before opening an envelope, if possible. The Investigator must document the date, time, and reason for the unblinding in the patient's medical records.

Should the code be (intentionally or unintentionally) broken by the Investigator or by someone on the clinical staff, the subject could remain in the study or not, as judged by the Investigator, according to the patients' best interest. The event will be immediately reported to the Sponsor and to [redacted] and dealt as a major protocol deviation. The clinical data for this patient will not be included in the analysis of the Per Protocol population.

3.3 Study treatments and assessments

The maximum study duration from screening to end of study is 32 months.

There will be three treatment groups in this study.

- 175mg BIO101 bid
- 350mg BIO101 bid
- Placebo

All included subjects will complete the inclusion visit and start a 6 to 9-month treatment and observation phase, with the main evaluation at the end of the study duration. Based on non-clinical considerations, the exposure is capped at 39 weeks.

A detailed description of procedures and assessments to be conducted during this study is summarized in SOA in Protocol section '1.3 SCHEDULE OF ACTIVITIES (SOA)'

4 STUDY ENDPOINTS

4.1 Primary endpoint

The primary endpoint of this study is Change from baseline to month 6 (M6)/month 9 (M9) in 400MW gait speed.

4.2 Secondary endpoints

The key secondary endpoints of this study are:

- Change from baseline to M6/M9 in Physical Function Domain (PF-10) sub-score of the Short Form Health Survey (SF-36)
- Change from baseline to M6/M9 in Handgrip strength test

The other secondary endpoints of this study are:

- Change from baseline of ALM and other parameters of body composition based on DEXA measurements
- The rate of success to complete 400 meter walking test after the 6-month/9-month treatment versus placebo
- Change from baseline of the distance during 6 Minute Walking Test (6MWT) distance
- Change from baseline in muscle strength based on the knee extension and the SCPT
- Change from baseline of the SPPB total score
- Change from baseline of the time for five chair stands as part of the SPPB assessment
- Change from baseline using the SarQol auto-evaluation questionnaire
- Change from baseline in PAT-D.
- Change from baseline in TSD-OC for obese patient
- Change from baseline in SF36 (total score)

4.3 Safety data

The safety data for this study is:

- Adverse Events (AEs)
- Clinical laboratory data (haematology, biochemistry, coagulation, urine analysis)
- Vital signs
- Physical examinations
- Electrocardiograms ECGs

[redacted]

5 SAMPLE SIZE AND POWER

Originally, the sample size was calculated for a comparison test at a 0.05 two-sided significance level, a power of 80%, to detect a difference of 0.05 meters per second (minimal clinically important difference) between active groups and placebo on the change from baseline in the 400 m gait speed at 6 months, with a standard deviation (SD) of 0.13 (estimated from the VIVE2 trial, Fielding et al, 2017[3]).

[redacted]

Based on current calculation and given a limited 20% provision for premature withdrawals or lost-to-follow up, a total number of 231 participants will be included in the study.

6 ANALYSIS POPULATIONS

The following analysis populations will be considered.

6.1 Safety population

The safety population consists of all subjects randomized in the study for whom there is any evidence that they used study medication. Subjects are included in the analysis according to the actual treatment received.

The safety population is used for the safety analyses.

6.2 Full Analysis Set (FAS)

The FAS consists of randomized subjects, excluding subjects who failed to take at least one dose of trial medication and had a very early withdrawal (first week after randomization) definitely not related to study medication and lacked any post randomization data (CPMP/ICH/363/96). Subjects are included in the analysis using the FAS according to the treatment to which they were randomized.

6.3 Per-Protocol (PP) population

The PP population consists of all FAS subjects who did not have a major protocol deviation that are related to drug administration (e.g. a long interruption period, as defined in section 7.1 of the protocol) that may have impacted the primary efficacy endpoint. The exclusion of subjects from the PP population will be determined in a blinded data review meeting (BDRM) that will be held prior to the disclosure of the randomization list. Reason(s) for exclusion from the PP population will be provided for each subject. Subjects are included in the analysis using the PP population according to the treatment to which they were randomized.

6.4 Protocol deviations and exclusions from analysis populations

The COVID-19 situation will introduce more protocol deviations than normal. All protocol deviations will be accurately recorded and identified as deviations that incurred due to the COVID-19 outbreak and as reasons why the procedures were not performed according to protocol. They will be assessed and reported in the clinical study report, following ICH E3.

All protocol deviations and exclusions of subjects from analysis populations will be identified at the BDRM prior to full study unblinding, through clinical review and input provided by Biophytis, using the following sources of information:

- Supportive subject listings, [redacted] ahead of the BDRM, based on data recorded on the eCRF.
- Protocol Deviation Logs, [redacted]

Deviations from the protocol will be classified as major or minor.

Major protocol deviations to be identified and to be mentioned in the CSR include:

- Failure to obtain informed consent prior to any study assessment
- Non-compliance with study treatment (<80% treatment compliance or missing compliance) or treatment interruption of > 4 weeks (interruption between two visits or within one visit)
- Failure to perform the 400MWT within 2 week of last dose (14 days or less after last dose)
- Deviation from Inclusion criteria 3, 4, 5, 6, 7
- Deviation from Exclusion criteria 1, 4, 6, 7d, 7e, 7f, 7g, 7h
- Wrong study treatment received (not per randomization list)
- Unintentional unblinding of the investigator and/or subject
- Any other protocol deviation identified as major protocol deviation from above mentioned source of information.

7 STATISTICAL CONSIDERATIONS AND ANALYSIS

7.1 Derived Variables

7.1.1 Derivations for demographic, other baseline characteristics, exposure and compliance

Table 1 provides the list of derived variables for this study.

Table 1: List of Derived Variables

Variables	Formula
Demographic and Baseline characteristics	
Body mass index (BMI) (kg/m^2)	weight (kg)/[height (m)]^2
Derivation of Duration	
Study day at any visit	Date of Interest - Date of first dose of study drug, if date of interest is before date of first dose of study drug If first dose of study treatment is on the same day as Randomisation (Day 0) Date of interest – date of first dose of study drug, if date of interest is on or after date of first dose of study drug. If first dose of study treatment is not on the same day as Randomisation (Day 0) Date of interest – date of first dose of study drug + 1, if date of interest is on or after date of first dose of study drug.
Extent of Exposure (days)	Date of last treatment intake– Date of first treatment intake + 1
Extent of Exposure (weeks)	Extent of exposure (days)/7
Relative day of AE Onset	AE onset date – date of first dose + 1
Duration of AE	AE end date – AE onset date + 1
Drug Compliance	
Compliance (%)	100 × [(total number of capsules dispensed) – (total number of capsules returned)] / (total number of capsules planned to be taken per day × duration of study drug exposure in days)
Baseline Derivations	
Baseline	The baseline value is defined as the last observation prior to or on the date of the first dose of study drug.
Change from baseline	Post baseline value – Baseline

7.1.2 Derivations for efficacy endpoints

7.1.2.1 PF-10 sub-score from the SF-36

SF-36[5] (version 2) Quality of Life questionnaire was developed as a general measure of perceived health status. It comprises 36 items, allowing the calculations of eight sub-scores: Physical Functioning (10 items), Role limitations due to-physical problems (4 items), Bodily Pain (2 items), General Health perceptions (5 items), Vitality (4 items), Mental Health (5 items), Role-limitations due to emotional problems (3 items) and Social Functioning (2 items). Two summary scores of health-related quality of life are also calculated; physical component summary (PCS) and the mental component summary (MCS).

SF-36 (version 2) data are recorded at baseline, M3, M6/M9 or end-of-intervention visits.

The method for calculating the SF-36 sub-scores, including PF-10 is shown in the Appendix B - Short Physical Performance Battery (SPPB)

Summary performance score: Each of the three performance measured is assigned a score ranging from 0 to 4, with 4 indicating the highest level of performance and 0 the inability to complete the test. For the test of balance, participants are assigned a score of 1 if they can hold a side-by-side standing position for 10 seconds but are unable to hold a semi-tandem position for 10 seconds; a score of 2 is assigned if they can hold a semi-tandem position for 10 seconds, but are unable to hold a full-tandem position for 3 seconds; a score of 3 is assigned if they can stand in a full-tandem position for 3 seconds but less than 10 seconds; a score of 4 is assigned if they can stand in a full tandem position for 10 seconds.

Four categories are computed for walking speed and chair stands, according to cut-off points that are based on quartiles of the time to perform each task assessed in the EPESE. The time of the faster of two walks is scored as follows: > 8.7 sec = 1; 6.21 to 8.70 sec = 2; 4.82 to 6.20 sec = 3; < 4.82 sec = 4; a score of 0 is assigned to participants unable to perform the test. The time required to perform five chair stands is scored as follows: ≥ 16.70 sec = 1; 13.70 to 16.69 sec = 2; 11.20 to 13.69 sec = 3; ≤ 11.19 = 4. A score of 0 is assigned to participants unable to perform the task. A summary score ranging from 0 (worst performers) to 12 (best performers) is calculated by adding subscores from the walking speed, chair stands and balance tests.

Appendix C

Details of how to deal with missing data are also mentioned.

Standardized sub-scale scores will be presented to one decimal place. Handling of missing data and outliers.

Missing data will not be imputed in the primary analysis of the primary efficacy endpoint. No imputation for missing data will be used for the descriptive statistics and listings. Imputation for missing SF-36 sub-scale score data is detailed in Appendix B - Short Physical Performance Battery (SPPB)

Summary performance score: Each of the three performance measured is assigned a score ranging from 0 to 4, with 4 indicating the highest level of performance and 0 the inability to complete the test. For the test of balance, participants are assigned a score of 1 if they can hold a side-by-side standing position for 10 seconds but are unable to hold a semi-tandem position for 10 seconds; a score of 2 is assigned if they can hold a semi-tandem position for 10 seconds, but are unable to hold a full-tandem position for 3 seconds; a score of 3 is assigned if they can stand in a full-tandem position for 3 seconds but less than 10 seconds; a score of 4 is assigned if they can stand in a full tandem position for 10 seconds.

Four categories are computed for walking speed and chair stands, according to cut-off points that are based on quartiles of the time to perform each task assessed in the EPESE. The time of the faster of two walks is scored as follows: > 8.7 sec = 1; 6.21 to 8.70 sec = 2; 4.82 to 6.20 sec = 3; < 4.82 sec = 4; a score of 0 is assigned to participants unable to perform the test. The time required to perform five chair stands is scored as follows: ≥ 16.70 sec = 1; 13.70 to 16.69 sec = 2; 11.20 to 13.69 sec = 3; ≤ 11.19 = 4. A score of 0 is assigned to participants unable to perform the task. A summary score ranging from 0 (worst performers) to 12 (best performers) is calculated by adding subscores from the walking speed, chair stands and balance tests.

Appendix C.

7.1.3 Missing data analysis methods for sensitivity analyses

Alternative methods to adjudicate missing data for the primary endpoint like the Bayesian Multiple Imputation (MI) methods could be considered and adopted before unblinding and preparation of the final analysis. This would be documented in a SAP amendment.

7.1.4 Handling of missing or incomplete dates

Imputation rules for missing or partial AE start date are defined below:

If only Day of AE start date is missing:

If the AE start year and month are the same as that for the first dose date, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start day as the day of first dose date;
- Otherwise, impute the AE start day as 1.

If Day and Month of AE start date are missing:

If AE start year = first dose year, then:

- If the full (or partial) AE end date is NOT before the first dose date or AE end date is missing, then impute the AE start month and day as the month and day of first dose date;
- Otherwise, impute the AE start month as January and the day as 1.

If Year of AE start date is missing:

If the year of AE start is missing or AE start date is completely missing then this will be queried, with no imputation. Also compare the full (or partial) AE end date to the first dose date. If the AE end date is before the first dose date then the AE should be considered as a pre-treatment AE. Otherwise, the AE will be considered as TEAE.

Imputation rules for missing or partial medication start/stop dates are defined below:

Missing or partial medication start date:

- If only day is missing, use the first day of the month.
- If day and month are both missing, use the first day of the year.
- If day, month and year are all missing, use first dose date -1.

Missing or partial medication stop date:

- If only day is missing, use the last day of the month.
- If day and month are both missing, use the last day of the year.
- If day, month and year are all missing, assign 'continuing' status to stop date

8 STATISTICAL METHODS

8.1 General statistical conventions

All statistical procedures will be completed using SAS® version 9.3 or higher.

Unless otherwise stated, all statistical testing will be two-sided and will be performed using a significance (alpha) level of 0.05. Two-sided 95% confidence intervals (CI) will be provided for the difference between treatment groups.

Descriptive statistics (mean, standard deviation, 95% confidence interval for mean, minimum, median, maximum, and number of observations and missing cases) will be used for quantitative variables. Frequencies and percentages will be used for categorical variables. The summaries will be presented by treatment group, unless otherwise specified. Percentages will be rounded to one decimal place.

Unless stated otherwise, unscheduled data will not be included in the summaries or analysis.

8.2 Data Listing

All subject data will be presented in individual subject data listings. Unless otherwise stated, unscheduled visit results will be included in date/time chronological order, within subject listings only. All listings will be sorted by treatment group, investigational site, subject number, date/time and visit. Subject's gender and age will be stated on each listing. Unless otherwise stated, data listings will be based on all randomized subjects.

8.3 Subject disposition

All subjects who were screened as well as provided informed consent will be included in a summary of subject disposition. The number of subjects screened, the number of screen failures, the number of subjects with each reason for screen failure will be presented overall. All subjects will be included.

The number and percent of subjects randomized, re-consent to Protocol Amendment 1.2.11

(COVID-19 outbreak), took a dose of study drug, randomized and not treated, treated and not randomized will be presented. The number of subjects that discontinued prematurely by visit will be presented along with the primary reason for premature discontinuations including 'Withdraw due to the COVID-19 outbreak'. Also, the number of subjects that complete the M3, M6, M9, M6/M9 visits and completed the study will be summarized. All summaries will be provided by treatment group and overall.

The number and percent of subjects in each analysis population will be tabulated. The number of subjects randomized will be used as the denominator for the percentage calculation.

Randomization details and Subject disposition data will be listed.

The informed consent data including 'patient re-consent to Protocol Amendment 1.2.11 (COVID-19 outbreak)' will be provided in a listing.

The data of each visit including 'if the visit was modified due to COVID-19 outbreak' data will be provided in a listing.

8.4 Protocol deviations

All protocol deviations and major protocol deviations identified will be summarized by treatment group and overall.

All protocol deviations and major protocol deviations related to COVID-19 identified will be summarized by treatment group and overall.

A listing of all subjects with protocol deviations will be provided by treatment group, including flag for COVID-19 related protocol deviation. A separate listing of randomized subjects that violated inclusion/exclusion criteria will be provided.

8.5 Demographics and baseline characteristics

No formal comparison between treatment groups for demographic or baseline characteristics will be done. All demographic and baseline characteristics data will be listed.

8.5.1 Demographics

Age and other continuous demographic variables at baseline will be summarized using descriptive statistics (mean, standard deviation, median, minimum, maximum and number of missing cases for quantitative parameters and percentages and number of missing cases for the qualitative parameters) by treatment group and overall. Included will be Age category (≥65 to <75 years, ≥ 75 years), gender, race, participation in SARA OBS study, country and region (EU/US) will be summarized using the FAS.

8.5.2 Baseline characteristics

The following continuous variables, at baseline, will be summarized by descriptive statistics for the FAS.

- Height,
- Weight,
- BMI,
- Calf circumference,
- Waist circumference,
- Hip circumference,
- Mid-arm circumference,
- Tricipital plica thickness,
- Gait speed during 400m walking test (m/sec),
- PF-10 score,
- Total score (including severity index/co-morbidity index) of Cumulative Illness Rating Scale (CIRS). Severity index = results from the average of the scores of the first 13 categories (excluding the category "psychiatric / behavioral pathologies"). Comorbidity index = the

number of categories in which a score higher than or equal to 3 is obtained (excluding the category "psychiatric / behavioral pathologies")

- By domain and total score of Pepper Assessment Tool for Disability (PAT-D),
- Physical and mental component summary of SF-36
- Score of Mini Nutritional Assessment – short Form (SF-MNA). For derivation refer Appendix E.

8.5.3 Medical history

A summary (frequency and percentage) of medical history by treatment group and overall will be presented by system organ class (SOC) and preferred term (PT) using Medical Dictionary for Regulatory Activities (MedDRA) Version 21.0 or higher for the safety population. All medical history will be listed.

8.5.4 Prior and Concomitant medications

Medications used in this study will be coded by using the latest available version of the World Health Organization Drug Dictionary Enhanced (WHODDE).

Prior medications are defined as those medications with a start and end date prior to the first dose of study drug.

Concomitant medications are defined as those medications with a start date on or after the date of first dose of study drug. A medication which started prior to first dose and continued after first dose will also be considered as concomitant medications.

Prior and Concomitant medications will be summarized descriptively using frequency and percentage by Anatomical Main Group [1st level of the Anatomical Therapeutic Chemical (ATC) classification], and preferred name by treatment group using the safety population.

Subjects with multiple medications in the same ATC class or preferred name will be counted only once for that respective ATC class or preferred name.

Details for imputing missing or partial start and/or stop dates of medication are described in section 7.1.4

A listings of prior and concomitant medications will be provided.

8.6 Extent of exposure

8.6.1 Treatment duration

Study drug exposure (in days) will be summarized by treatment group for the safety population using descriptive statistics.

Exposure to study drug will be categorized in intervals (1-27 days, 28-83 days, 84-146 days, 147-

181 days, 182-223 days, 224-272 days, ≥273 days) and will be summarized by treatment group for the safety population.

8.6.2 Treatment compliance

Study drug compliance will be summarized by treatment group using descriptive statistics. The following compliance categories will be summarized "<80%" "80 – 100%" and ">100%" by frequency and percentage.

The summaries will use the safety population.

Study drug exposure and compliance data will be provided in a listing.

The data of study drug compliance at each visit will be provided in a listing.

8.7 Efficacy analyses

This section addresses separately the analyses to be conducted on the primary, key secondary and secondary efficacy endpoints.

Analyses for the primary and key secondary efficacy endpoints will be analyzed using the FAS and PP population. Analyses for all non-key secondary efficacy and exploratory endpoints will be analyzed using the FAS only.

The definition of efficacy endpoints is detailed in section 4.

Summary value and change from baseline of endpoints by visits will be summarized for Baseline, M3, M6, M9, M6/M9, including n, mean, median, standard deviation, minimum, maximum.

Analysis of Mixed Model for Repeated Measures (MMRM) will be provided for M3, M6/M9. M6/M9 data comes from M6, but M9 value will be used if M6 is missing.

Summary value and change from baseline of primary efficacy endpoint will be summarized for Last assessment on treatment. Last assessment on treatment is defined as the last non-missing value from the post-baseline treatment period.

8.7.1 Analysis methods

Treatment group comparison (175 mg BIO101 Vs Placebo and 350 mg BIO101 Vs Placebo) will be tested at a significance level of 0.05, 2-sided unless stated otherwise.

8.7.1.1 Analysis of Mixed Model for Repeated Measures (MMRM)

The analysis of the efficacy endpoints collected over time will use a MMRM analysis on change from baseline. An estimate of the contrast between each active arm and placebo (after adjusting for the Hochberg procedure, primary and key secondary endpoints only) at M6/M9 will be made. The MMRM model will include treatment, visit, center, gender (center and gender are stratification factors) as fixed effects, treatment*visit as interaction terms and baseline value as a covariate (if a

baseline value for that endpoint is collected). An unstructured covariance matrix (there is no assumption of a particular correlation structure for repeated measurements within subjects over time) will be used to estimate the within-subject covariance. The denominator degrees of freedom will be calculated according to the Kenward-Roger method.

The Least-squares [LS] means, LS mean difference between each active arm and the placebo group (reference category), standard error, two-sided 95% CIs for the LS means and difference between treatment groups and p-value for treatment differences will be presented. The p-values for the effects will also be presented. The number of patients in the analysis population and number of patients in the analysis will be provided by treatment group.

The MMRM model will give a correct estimation in the presence of missing data, under the missing at random (MAR) assumption. This analysis will provide an estimate of the efficacy of 175mg bid BIO101 and 350mg bid BIO101 in the population who stay in the study until 6/9 months.

In case of non-convergence of the above model using an unstructured covariance matrix, the following variance structure will be used in this order: Heterogeneous Toeplitz, Toeplitz and Compound Symmetry. The final covariance structure used will be documented in the CSR.

Due to the quantitative form of the primary and key secondary endpoints (400MW gait speed, PF10 and handgrip strength test) and the large number of patients, the normal assumption for the residuals of the mixed models will be assumed.

The SAS code used to implement this analysis is given in Appendix A.

8.7.1.2 Analysis of covariance (ANCOVA) model for Change from Baseline

The analysis of the efficacy endpoints will use an ANCOVA on change from baseline to M6/M9. An ANCOVA on change from baseline to last assessment on treatment for primary endpoint will be performed.

The model will include treatment, center and gender (center and gender are stratification factors) as fixed effects and the baseline measurement as a covariate.

The model will be used to derive least squares estimates of the treatment differences (each active arm versus placebo) in mean change and two-sided 95% confidence intervals. Also t-statistics corresponding to the type III sums of squares for the differences in the least squares means will be used to obtain p-values for treatment group comparisons (each active arm versus placebo group). The LS means and associated two-sided 95% CIs for the mean change within each treatment group will be calculated. The p-values for the effects will also be presented. The number of patients in the analysis population and number of patients in the analysis will be provided by treatment group.

Due to the quantitative form of the endpoints (400MW gait speed, PF10 and handgrip strength test) and the large number of patients, the normal assumption for the residuals of the mixed models will

be assumed.

8.7.1.3 Logistic Regression

Analysis of response data (responder/non-responder) at M6/M9 will use a logistic regression model. Analysis of response data at last assessment on treatment for primary endpoint will use a logistic regression model.

The model will include treatment, center and gender as fixed effects and baseline score as a covariate.

The multiplicity issue due to the two doses of active treatment will be addressed using the Hochberg procedure.

The number of subjects in the analysis population, number of subjects in the analysis, odds ratio (OR), two-sided Wald 95% CIs and p-value based on Wald chi-square test for 175 mg BIO101 Vs Placebo and 350 mg BIO101 versus placebo will be presented.

The SAS code used to implement this test is given in Appendix A.

8.7.1.4 Cochran Mantel-Haenszel (CMH) Test

Analysis of success rate or risk ratio will use a CMH test with center and gender as stratification (adjustment) variables.

The adjusted relative risk (adjusted for center and gender) based on the ratio of success rates for each treatment group versus placebo, 95% CI for the adjusted relative risk and CMH p-value will be presented.

The SAS code used to implement this test is given in Appendix A.

8.7.1.5 Pearson Correlation co-efficient

In order to the strength and direction of the linear relationship between the two variables, a Pearson correlation co-efficient along with the p-value will be presented.

The SAS code used to implement this test is given in Appendix A.

8.7.1.6 Multiplicity

To strictly control overall type I error rate for the analysis of the primary and the key secondary endpoints to a two-sided 5% level, a Hochberg procedure will be used.

The aim of the study is to show that at least one dose is statistically superior to placebo while the family-wise type I error rate is maintained at the two-sided α =0.5 for the following primary and key secondary endpoints.

- Change from baseline to month 6 (M6)/ month 9 (M9) in 400MW gait speed
- Change from baseline to M6/M9 in Physical Function Domain (PF-10) sub-score of the Short Form Health Survey (SF-36)
- Change from baseline to M6/M9 in Handgrip strength test

The null hypotheses H_{01} and H_{02} given below will be tested against the alternative hypotheses H_{A1} and H_{A2} associated with the 175 mg BIO101 and 350 mg BIO101 groups respectively:

$$H_{01}: \mu_1 - \mu_P \leq 0, \qquad H_{02}: \mu_2 - \mu_P \leq 0,$$

$$H_{A1}: \mu_1 - \mu_P > 0, \qquad H_{A2}: \mu_2 - \mu_P > 0,$$

Where μ_1 and μ_2 denote the mean change from baseline at M6/M9 in the 175 mg BIO101 and 350 mg BIO101 groups. With μ_p denoting the mean change from baseline at M6/M9 in the placebo group.

A statistical rejection of H_{01} (or H_{02}) will indicate that the 175 mg BIO101 (or 350 mg BIO101) is superior to placebo or we say that the 175 mg BIO101 (or 350 mg BIO101) is significantly better than placebo. When at least one dose is significantly better than placebo, we say that the study is positive.

Two-sided p-values for comparing the 175 mg BIO101 to placebo and comparing the 350 mg BIO101 to placebo are denoted by p_{01} and p_{02}, respectively. The first step is to compare max (p_{01}, p_{02}) to 0.05. If max (p_{01}, p_{02}) $\leq$ 0.05, both doses are claimed to be statistically better than placebo. If max (p_{01}, p_{02})> 0.05, the second step is needed to compare min (p_{01}, p_{02}) to 0.5/2 =0.025. If min (p_{01}, p_{02}) $\leq$ 0.025 at the second step, the dose with is concluded to be statistically better than placebo. Otherwise no dose can be claimed to be better than placebo.

Multiplicity concerns for the non-key secondary endpoints are kept under control by the pre-specification of key-secondary endpoints and a low number of secondary endpoints and analyses. Analysis of the non-key secondary endpoints will be considered supportive and no adjustments for multiplicity will be made.

8.7.1.7 Treatment by interaction analysis

A center by treatment interaction will not be included in any of the analysis models.

8.7.1.8 Pooling of centers for analysis

Centers will be pooled by country depending on the distribution of randomized patients. The algorithm for the pooling of centers, if required, will be detailed in a separate document prior to unblinding.

8.7.2 Analysis of primary efficacy endpoint

The 400MW test used to calculate the 400MW gait speed will be performed at Screening, M3, M6/M9 and last assessment on treatment visits. The Screening value will be considered to be the baseline value. If the screening value is missing the baseline value will be considered as missing.

The main analysis of the primary endpoint (change from baseline to month 6 (M6)/month 9 (M9) in 400MW gait speed) will be performed using the FAS.

Change from baseline to M6/M9 in 400WM gait speed will be summarized by treatment group.

The main hypothesis to be tested in this study is the superiority of at least one dose of active versus placebo on the change from baseline in 400m gait speed after 6 months of treatment.

The null hypotheses H_{01} and H_{02} given below will be tested against the alternative hypotheses H_{A1} and H_{A2} respectively:

$$H_{01}: \mu_1 - \mu_P < 0.1, \qquad H_{02}: \mu_2 - \mu_P < 0.1,$$

$$H_{A1}: \mu_1 - \mu_P \geqslant 0.1, \qquad H_{A2}: \mu_2 - \mu_P \geqslant 0.1,$$

Where μ_1 and μ_2 denote the mean change from baseline in meters per second in the 400MW gait speed at M6/M9 in the 175 mg BIO101 and 350 mg BIO101 groups. With μ_p denoting the mean change from baseline at M6/M9 in the placebo group. A minimum clinically significant benefit is set at 0.1 meter per second in the mean difference between groups (i.e., superiority margin).

A Mixed Effect Model Repeat Measurement (MMRM) model with fixed factors treatment and center and the baseline score and the stratification factor (gender) will produce an estimate of the contrast at M6/M9 in the change from baseline of the 400MW test gait speed between each active arm and the placebo group (section 8.7.1.1) (after adjustment by the Hochberg procedure (Section 8.7.1.6)).

The same MMRM model will be also fitted for the Per-Protocol population to give an estimand of the efficacy of 175mg BIO101 and 350mg BIO101 in the population with a good compliance to the protocol. The FAS is the main population/set of interest.

A by-subject listing of 400MW data will be provided.

Alternative methods to adjudicate missing data for the primary endpoint like the Bayesian Multiple Imputation (MI) methods will also be considered. Chen at al (2017) demonstrated the interest of applying Bayesian MI method to 400MW non-completers in the LIFE study[9].More detail please refer to section 8.7.6.

Change from baseline to last assessment on treatment will be analyzed using an ANCOVA method as defined in Section 8.7.1.12, and will be performed using the FAS and Per-Protocol population separately.

Additionally, a responder analysis will be performed, using a logistic regression (method defined in section 8.7.1.3) to model 400MW gait speed test response at M6/M9. A responder is defined as an improvement (increase) of 0.1 m/s or more in 400MW gait speed test compared to baseline. A non-responder will be defined as a subject that is not a responder. Subjects with a missing value at the visit and/or missing baseline value will be considered as non-responders. The number and rate (%) of responders and non-responders will be provided by treatment group for the M3 and M6/M9 visits.

8.7.3 Analysis of secondary efficacy endpoints

8.7.3.1 Analysis of key secondary efficacy endpoints

The key secondary endpoints (change from baseline to M6/M9 in Physical Function Domain (PF-10) and change from baseline to M6/M9 in Handgrip strength test) will be summarized by treatment group and analyzed using the FAS. With sensitivity analysis performed using the PP population. The FAS is the main population/analysis set of interest.

Details of how to calculate PF-10 from the SF-36 questionnaire is in Appendix B - Short Physical Performance Battery (SPPB)

Summary performance score: Each of the three performance measured is assigned a score ranging from 0 to 4, with 4 indicating the highest level of performance and 0 the inability to complete the test. For the test of balance, participants are assigned a score of 1 if they can hold a side-by-side standing position for 10 seconds but are unable to hold a semi-tandem position for 10 seconds; a score of 2 is assigned if they can hold a semi-tandem position for 10 seconds, but are unable to hold a full-tandem position for 3 seconds; a score of 3 is assigned if they can stand in a full-tandem position for 3 seconds but less than 10 seconds; a score of 4 is assigned if they can stand in a full tandem position for 10 seconds.

Four categories are computed for walking speed and chair stands, according to cut-off points that are based on quartiles of the time to perform each task assessed in the EPESE. The time of the faster of two walks is scored as follows: > 8.7 sec = 1; 6.21 to 8.70 sec = 2; 4.82 to 6.20 sec = 3; < 4.82 sec = 4; a score of 0 is assigned to participants unable to perform the test. The time required to perform five chair stands is scored as follows: ≥ 16.70 sec = 1; 13.70 to 16.69 sec = 2; 11.20 to 13.69 sec = 3; ≤ 11.19 = 4. A score of 0 is assigned to participants unable to perform the task. A summary score ranging from 0 (worst performers) to 12 (best performers) is calculated by adding subscores from the walking speed, chair stands and balance tests.

Appendix C.

The handgrip strength test will be performed at Day 0 and at the M6/M9 visit or the end-of-intervention visit and will be measured 3 times for both hands. Grip strength is considered as the dominant hand so the highest of all 3 attempts for the dominant hand will be used for the analysis. The highest value for each hand will be summarized by treatment group. If there has been current flare-up of pain or has undergone fusion, arthroplasty, tendon repair, synovectomy, or other related surgery for the dominant hand or wrist in the past 3 months, the other hand will be tested. For the handgrip strength test, a minimum clinically significant benefit is set at 2 kg difference in the change at M6/M9 from baseline versus placebo in the mean difference between treatment groups.

Fatigue strength will be summarized by treatment group.

The analysis of change from baseline to M6/M9 in Physical Function Domain (PF-10) will use the same method as for analysis of the primary efficacy endpoint as detailed in section 8.7.1.1.

Missing data will not be replaced in these analyses but the MMRM model on observed data is known to perform well under a MAR hypothesis.

An ANCOVA method detailed in section 8.7.1.12 will be used to analyse change from baseline to M6/M9 in Handgrip strength test.

Alternative methods to adjudicate missing data for the primary endpoint like the Bayesian Multiple Imputation (MI) methods will also be considered for Physical Function Domain (PF-10). Chen at al (2017) demonstrated the interest of applying Bayesian MI method to 400MW non-completers in the LIFE study[9]. More detail please refer to section 8.7.6.

A responder analysis will be performed, using a logistic regression (method defined in section 8.7.1.33) to model PF-10 response at M6/M9. A responder is defined as an improvement of PF-10 greater or equal to 2 points compared to the baseline value, at an individual level. A non-responder for a visit (M3 or M6/M9) will be defined as a subject that is not a responder at that visit. Subjects with a missing value at a visit (M3 or M6/M9) will be considered as non-responders. The number and rate (%) of responders and non-responders will be provided by treatment group for the M3 and M6/M9 visits.

Handgrip strength test will be summarized and analyzed in the same way as PF-10 response. Where response is defined as $\geq$ 2kg improvement in handgrip strength test at M6/M9 compared to the baseline value, at an individual level. A non-responder is a subject that is not a responder. Subjects with a missing value at M6/M9 will be considered as non-responders. The number and rate (%) of responders and non-responders will be provided by treatment group.

A by-subject listing of PF-10 and Handgrip strength test will be provided.

8.7.3.2 Analysis of other secondary efficacy endpoints

The other secondary endpoints will be analyzed using the FAS only.

Observed and change from baseline values for the following secondary efficacy endpoints will be summarized using descriptive statistics by treatment group. These endpoints will be analyzed using an ANCOVA method as defined in Section 8.7.1.12. This analysis will be performed without any correction for multiplicity.

- **Change from baseline in ALM and other parameters of body composition by DEXA:** The DEXA parameters will be ALM, ALM BMI, percentage lean mass for left/right arm, trunk and left/right leg and percentage fat mass for trunk, right/left leg and right/left arms.
- **Change from baseline in muscle strength (knee extension) and stair climb power test (SCPT):**
  Knee extension will be performed at Day 0 and at the M6/M9 visit or the end-of-intervention visit and is measured using the maximum peak torque measurement (Nm) for 60°/s, 90°/s and 180°/s for the tested leg (either right or left). It is usually measured as the force applied to the ankle with the subject seated in an adjustable straight-back chair, the lower leg unsupported and the knee flexed to 90°. Knee extension will be measured 3 times for the chosen leg. The highest of all 3 attempts will be used for the summary and analysis. The assay with the highest torque output (highest value) will be taken for analyses.
  The SCPT will be performed at Day 0 and at the M6/M9 visit and is the measure of ability to ascend and descend stairs which tests the body strength and balance. SCPT measures the time (in seconds) taken to ascend and descend a flight of stairs (10 steps with a 20cm step height). Step heights should be suitable (between 16 and 20 cm).
- **Change from baseline in Pepper Assessment Tool for Disability (PAT-D) for obese patient.**
  The PAT-D is a disability questionnaire that was developed to assess difficulty with functioning in both discrete tasks and social/role functioning. There are three domains: mobility, activities of daily living, and instrumental activities of daily living. This questionnaire asks respondents how much difficulty they have had with a range of activities in the past month due to their health. For each item, respondents answer on a 5-point Likert-type scale whether they experience: (1) no difficulty, (2) a little difficulty, (3) some difficulty, (4) a lot of difficulty, or (5) unable to do. For each item there is also a 'not applicable' option if the respondent usually did not do that activity for reasons other than disability. The PAT-D produces a score for each of the three domains (basic ADLs, mobility and IADLs) as well as a total score, with higher scores indicating greater disability. See Appendix D for method of calculation. The PAT-D will be performed at Day 0 and at the M6/M9 visit.

- **Change from baseline to month 6 (M6)/month 9 (M9) in 6MWT distance**
  The 6MWT is a test for functional exercise capacity and involves measuring the distance a participant can cover within the allotted time of 6 minutes. The 6MWT will be performed at Day 0 and at the M6/M9 visit.

The following secondary efficacy endpoints will be summarized using descriptive statistics by treatment group and analyzed using a MMRM analysis as defined in Section 8.7.1.11. This analysis will be performed without any correction for multiplicity.

- **Change from baseline in the Short Physical Performance Battery (SPPB) total score:**
  The SPPB is evaluated at Screening, M3 and M6/M9 or end of study visit and is a brief performance battery based on timed short distance walk, repeated chair stands and balance test. Change from baseline of SPPB will provide an estimate of the overall clinical impact of treatment on the burden of physical frailty. The Screening value will be considered to be the baseline value. If the screening value is missing the baseline value will be considered as missing. The scores for the individual tests and the total SPPB score are recorded in the CRF. Total gait speed test score, total chair stand test score, total balance test score and total SPPB score will be summarized by visit for each treatment group.
- **Change from baseline of the time for five chair stands as part of the SPPB assessment.**
  Time will be in seconds. The Screening value will be considered to be the baseline value. If the screening value is missing the baseline value will be considered as missing. The total time to five chair stands will be summarized by visit for each treatment group.
- **Change from baseline in SarQol auto evaluation questionnaire:**
  SarQol sarcopenia-specific is evaluated at Screening, M3 and M6/M9 or end of study visit and is self-administered, quality of life questionnaire, designed for community-dwelling elderly subjects aged 65 years and older.
  The questionnaire contains 22 questions, which covers 7 domains: physical and mental health; locomotion; body composition; functionality; activities of daily living; leisure activities; and fears. Most questions are answered on a 4 point or 5-point Likert scale. The total and domain score for the SarQol will be summarized by visit for each treatment group.
- **Change from baseline in TSD-OC for obese patient**
  The TSD-OC is considered a reliable and valid instrument for measuring self-reported disability in obese subjects. The questionnaire is a set of multiple questions in 7 dimensions to measure pain; stiffness; function and autonomy in daily activities, housework, outdoor activities, occupational activities, and social life. Each question is scored by the subject from 0 (best) to 10 (worst). Each dimension score is calculated and then all dimension scores are added together for a total TSD-OC score. The TSD-OC will be performed at Day 0, M3 and

M6/M9 or End of intervention Visit. The total score will be summarized by visit for each treatment group.

The 400MW test is a dichotomous test, failure being defined as the inability to walk 400 meters in 15 minutes without needing to sit or require help. The number and percentage of success and failure to complete the 400 MW test at M6/M9 will be presented by treatment group.

The rate of success to complete the 400 meter walking (400 MW) test after 6/9 months of treatment will be analyzed using the Cochran Mantel-Haenszel (CMH) test adjusted for center and gender per Section 8.7.1.4.

A by-subject listing will be provided for secondary efficacy endpoint data.

[redacted]

8.7.5 Analysis of other efficacy datas

All other efficacy data described here will be summarized using the FAS.

Descriptive statistics of the observed scores and change from baseline by treatment group for each scheduled visit will be provided.

- Time to walk 400m, 200m, 100m
- SF-36: 8 subscales and two summary scores. The 8 subscales are: physical functioning, role limitations due to physical problems, bodily pain, general health perceptions, vitality, social functioning, role-limitations due to emotional problems, and mental health. The two summary scores are the physical component summary and the mental component summary. Each scale is directly transformed into a 0-100 scale on the assumption that each question carries equal weight. The lower the score, the more disability; the higher the score the, less disability (i.e., a score of zero is equivalent to maximum disability and a score of 100 is equivalent to no disability).

By-subject listings of other efficacy data will also be provided.

8.7.6 Impact of the COVID-19 Outbreak

Due to the COVID-19 outbreak, an updated strategy for the final analysis will be implemented, to account for the missing data on safety and efficacy which will have impact on the final analysis. This will include sensitivity analyses comprising before/after COVID_19 outbreak.

Subjects will be classified into two COVID-19 outbreak classification, including:

1. End of Study before 16MAR2020
2. End of Study after 16MAR2020

End of study date will be used in determining the COVID-19 outbreak classification, the last contact date will be used if no end of study date is available.

The COVID-19 outbreak classification will be calculated for primary endpoint and key secondary endpoints by using FAS population and PP population. The COVID-19 outbreak classification will be calculated for adverse events by safety population including all AEs, all TEAEs, treatment related TEAEs, treatment related Serious TEAEs, TEAEs leading to treatment discontinuation, serious

AEs/TEAEs, Adverse events of special interest (AESI), Serious Adverse events of special interest (AESI), Treatment related Adverse events of special interest (AESI), and TEAEs leading to death by system organ class and preferred term.

8.7.6.1 Sensitivity Analyses of Missing data by Using Bayesian Multiple Imputation (MI) Methods

Bayesian MI method will be used to impute 400-m gait speed for the non-completers. The technical details of the Bayesian MI method have been described in publication[10]. Briefly, assumed that the 400-m gait speed data collected at 3, 6/ 9 months followed a multivariate normal distribution. The 400-m gait speed for the non-completers was imputed subject to the constraint that the value is less than 0.44 m/s. All imputations were performed separately for the two intervention groups. The SAS macro used to implement the procedure is detailed in Supplementary Appendix S1.

Two different Bayesian imputation models were used. The unadjusted Bayesian imputation model relied solely on the assumption that the 400-m gait speed of the non-completers was left-censored. . Non-completers include subjects who did not complete the test and discontinued from study, including discontinued due to COVID-19. The adjusted Bayesian imputation model used the baseline 400-m gait speed as a covariate in the imputation process.

Subsequently, we conducted analyses for the longitudinal 400-m gait speed data using both the simplistic imputation and the Bayesian MI methods. For the Bayesian MI method, analyses were performed on each of the 25 multiply imputed datasets obtained from the adjusted Bayesian imputation models. The results were then combined using Rubin's method carried out in SAS PROC MIANALYZE for valid statistical inference[11]. In all analyses, linear mixed-effects models were fit with unstructured variance-covariance matrices. The models contained main effects for intervention assignment and visit and the intervention by visit interaction. Baseline 400-m gait speed was used as a covariate. Least square means were estimated for each intervention group at each follow-up visit. An example of the SAS program is included in Supplementary Appendix S2.

8.8 Safety analyses

All analysis of safety data will use the Safety population.

All safety data will be summarized/analyzed by treatment group.

No statistical hypothesis testing will be performed.

Due to the COVID-19 situation and safety concerns over viral transmission, every investigational visit within the protocol is now be performed remotely (e.g. by phone) for all participants still active in the study by the investigational site staff. The investigational site will continue to monitor the health status of the participants by collecting information on general health status, Concomitant Medication and Adverse Events.

M7.5 and M10.5 phone calls include a pre-diabetic questionnaire as recommended by the DSMB committee.

8.8.1 Adverse events

All AEs or treatment emergent adverse events (TEAEs) will be classified by Primary SOC and PT according to MedDRA Version 21.0 or higher.

AEs will be classified as pre-treatment AEs, TEAEs (treatment-emergent adverse events), or post-treatment AEs, defined as follows:

Pre-treatment AE: any AE that started before the first dose date of study drug intake (AE onset date < date of first study drug intake).

Treatment-emergent: An AE will be considered a TEAE if it starts on or after the first dose date of study drug up to the last dose date + 6 weeks (date of first randomized study medication intake ≤ AE onset date ≤ last dose date + 6 weeks).

Post-treatment AE: any AE that was newly developed after the last dose date + 6 weeks (AE onset date > last dose date + 6 weeks).

Details for imputing missing or partial start dates of adverse events are described in Section 7.1.4.

Summaries of the following will be presented.

An overall summary table with number and percentage of subjects and number of events for the following will be presented by treatment group and overall:

- All AEs (excluding pre and post treatment events)
- All TEAEs
- Treatment related TEAEs (AE will be defined as related if causality is either definitely, probably or potentially)
- Treatment related Serious TEAEs (AE will be defined as related if causality is either definitely, probably or potentially)
- TEAEs by maximum severity (if a subject has several AEs with the same SOC/PT. The AE will be included only once in the summary, this being the AE with maximum severity
- TEAEs leading to treatment discontinuation (action taken = 'Discontinued')
- Serious AEs/TEAEs
- TEAEs leading to death.
- Adverse events of special interest (AESI) (falls and injurious falls and orthostatic hypotension)
- Serious Adverse events of special interest (AESI) (falls and injurious falls and orthostatic hypotension

- Treatment related Adverse events of special interest (AESI) (falls and injurious falls and orthostatic hypotension)

Pre-treatment AEs and post-study AEs will be presented in the listings only.

The number and percentage of subjects with at least one AE and the number of AEs will be presented by SOC and PT for the categories above. Two AEs with the same PT and classified in the same category (pre-treatment AE, TEAE or post-study AE) will be considered as two different events when calculating the "number of events" in the summaries. The same rule will apply for SOCs summaries by SOC and PT will be presented in alphabetical order.

All information about AEs collected on the eCRF will be listed alongside the treatment, preferred term, and SOC. For Serious AEs, deaths, AEs related to the study treatment, TEAEs leading to treatment discontinuation and AESI, a separate listing will also be provided. In addition, where applicable, a separate flag to define the treatment period (pre-treatment/TEAE/post-treatment) will be included.

8.8.2 Clinical laboratory evaluations

SI units will be used for all summaries and presentations (including listings) of laboratory data.

Descriptive statistics of observed values and change from baseline will be presented for clinical laboratory (hematology, biochemistry and coagulation) and continuous Urinalysis evaluations (by treatment group and visit.

Categorical urinalysis (protein and occult blood) parameters will be tabulated by visit and treatment group in frequency tables.

The clinical safety laboratory tests that will be summarized are listed below.

a. **Haematology**: haemoglobin, HbA1c, haematocrit, red blood cell (RBC) count, WBC count, platelet count, mean corpuscular haemoglobin, mean corpuscular haemoglobin concentration, mean corpuscular volume, absolute basophils, absolute eosinophils, absolute lymphocytes, absolute monocytes, absolute neutrophils.

b. **Biochemistry**: sodium, potassium, chloride, bicarbonate, urea (BUN), uric acid, creatinine (enzymatic), albumin, glucose, cholesterol (total cholesterol, LDL [calc and mes], HDL fractions), triglycerides, aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase, bone-specific alkaline phosphatase, lipase, amylase, gamma glutamyl aminotransferase (GGT), bilirubin (total, indirect, direct), lactate dehydrogenase, total protein, hs-CRP and eGFR calculation;

c. **Coagulation**: activated partial thromboplastin time, prothrombin time, and international normalised ratio.

d. **Urine analysis:** urinalysis by dipstick glucose, protein, and occult blood.

Shift tables from Baseline to the end of treatment, with regard to normal range (low, normal, high), will be presented by treatment group for each of the laboratory parameters

All laboratory and urinalysis data will be listed.

8.8.3 Vital signs

Absolute value and changes from baseline for vital sign measurements (heart rate, systolic and diastolic blood pressure) will be summarized by treatment group and position (at each scheduled visit using descriptive statistics. For systolic and diastolic blood pressure the values collected when sitting/lying will be summarized separately to those collected whilst standing.

Vital signs data will be listed.

8.8.4 Physical examinations

Physical examination findings (Normal, Abnormal) will be summarized using frequencies and percentages by Body System, treatment group and scheduled visit.

Physical examination data will be listed.

8.8.5 Anthropometry

Anthropometry measurements (height, weight, BMI, calf/waist/Hip/mid-arm circumference and tricipital plica thickness) will be summarized by treatment group at each scheduled visit using descriptive statistics.

Anthropometry data will be listed.

8.8.6 Electrocardiograms (ECG)

ECG data collected at screening visit will be listed.

8.8.7 Short Form Mini-Nutritional Assessment

A summary of SF-MNA score as mentioned in the Appendix at baseline using descriptive statistics and the nutritional status using frequency and percentage will be presented.

SF-MNA data will be listed.

8.9 Other analysis

8.9.1 Subgroup analysis

The subgroup analyses will use the FAS.

Subgroup analyses will be performed for the primary endpoint (change from baseline in 400MW gait speed at M6/M9) and the key secondary endpoints (change from baseline in PF-10 sub-score of the SF-36 and change from baseline to M6/M9 in Handgrip strength test).

These pre-identified, at-high risk of worsening, subpopulations have been identified:

- Low gait speed (4 meter walk with a gait speed <= 0.8 m/s) at Screening
- Sarcopenic obesity (having a percentage of body fat mass of >25% in men and >35% in women) as measured by the DEXA at Screening
- Study subjects with a chair stand sub-score of ≤2 of the SPPB at Screening
- Study subjects who experience a deterioration. Deterioration is defined as a decrease of > 2% in the ALM or BMI value at M6/M9 compared to the Screening value.

The subgroup analyses based on these factors will be performed in order to better characterize treatment benefits in subjects with increased risk of mobility disability. The analysis will be performed in both the subpopulations identified above and the group of FAS patients that are not in the subgroup. Summaries of each subgroup by treatment will be presented and MMRM analysis per section 8.7.1.1 will be performed for the primary endpoint and PF-10. An ANCOVA method detailed in section 8.7.1.12 will be used for the analysis of change from baseline to M6/M9 in Handgrip strength test. Parameters of gait speed (400MW with a gait speed <= 0.8 m/s/400MW with a gait speed > 0.8 m/s) and sarcopenic obesity (percentage of body fat mass of >25% in men and >35% in women/ percentage of body fat mass of ≤25% in men and ≤35% in women) will be included in the models. Important demographics or baseline characteristics such as age, country, etc. may be defined prior to the database lock and unblinding and will be included in the model.

Least square mean difference (175mg BIO101 Vs Placebo and 350mg BIO101 Vs Placebo) and 95% CIs for each of the subgroups will be presented in forest-plots.

No adjustment for multiplicity will be performed.

A by-subject listing will be provided.

8.10 Interim analysis

Not Applicable.

8.11 Data Safety Monitoring Board

The DSMB will be responsible for reviewing accumulated safety and efficacy data at intervals throughout the study.

The DSMB is periodically reviewing all safety data and raise alerts in case of negative or dangerous findings. Members of the DSMB will be selected based on their expertise by the SARA managing Board.

During the COVID-19 outbreak, the DSMB is meeting at least every other week to review its recommendation based on update of guidance from the national health agencies and local status.

In particular, the DSMB will evaluate the results of a formal pre-planned interim futility analysis and sample size reassessment. The DSMB will recommend to the Sponsor whether to continue, modify or stop the clinical trial based on futility but also on safety issues or other considerations not related to efficacy or safety. In case the study continues and the original sample size is retained, the DSMB will not communicate to anyone whether this is due to unfavorable or favorable results.

Documents for DSMB review will be provided by a group external to [redacted]. [redacted] will not be involved in providing support for the DSMB.

9 CHANGES TO PLANNED ANALYSIS FROM STUDY PROTOCOL

- The text from the protocol ''and for whom any follow-up information is available'' is not included in the definition of the safety population, thus more closely aligned with the definition in CPMP/ICH/363/96.
- The protocol mentioned '' The study would stop for futility at the interim analysis if a gait speed difference below the threshold defined in the SAP is observed.'' The interim analysis (IA) will no longer include the possibility of stopping for futility as all patients will have been recruited before the IA occurs. The details of the IA will be contained in the IA SAP.
- 'Month 6' is updated to 'Month 6/Month 9' in SAP due to COVID-19 except for section '5 SAMPLE SIZE AND POWER'
- 'Additional sensitivity analyses may be added in the SAP during the blind-review meeting if the number of missing data is important (more than 10%).' is removed from SAP, and section '8.7.6.1 Sensitivity Analyses of Missing data by Using Bayesian Multiple Imputation (MI) Methods' is added due to COVID-19.
- Additional analysis for primary endpoint is planned. Last assessment on treatment will be derived and analyzed.

10 REFERENCES

1. ICH Topic E3: Structure and Content of Clinical Study Reports (CPMP/ICH/137/95-adopted December 1995).

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

3. Fielding RA, Travison TG, Kirn DR et al. Effect of Structured Physical Activity and Nutritional Supplementation on physical Function in Mobility-Limited Older ADULTS: results from the VIVE2 Randomized trial. J Nutr Healthy Aging. 2017. Online First Articles. 1-7

4. Chen H, Ambrosius WT, Murphy TE, Fielding R, Pahor M, Santanasto A, Tudor-Locke C, Jack Rejeski W, Miller ME; LIFE Study. Imputation of Gait Speed for Noncompleters in the 400-Meter Walk: Application to the Lifestyle Interventions for Elders Study. J Am Geriatr Soc. 2017 Sep 8. doi:10.1111/jgs.15078. [Epub ahead of print] PubMed PMID: 28884789

5. Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). I: Conceptual framework and item selection. Med Care. 1992;30(6):473-483.

6. Ware JE, Kosinski M and Dewey JE. How to score Version 2 of the SF-36 Health Survey. Lincoln, RI: QualityMetric Incorporated. 2001.

7. Huhmann MB, Perez V, Alexander DD, Thomas DR 2013. A self- completed Nutrition screening tool for community dwelling older adults with High Reliability: A Comparison study. The Journal of Nutrition, Health & Aging Volume 17, Number 4, 2013

8. Jianjun (David) Li, Simon Kirby. Testing strategy in phase 3 trials with multiple doses. Biometrical Journal. 2019;61:115–125.

9. Chen, H. et al. Imputation of Gait Speed for Noncompleters in the 400-Meter Walk: Application to the Lifestyle Interventions for Elders Study. J. Am. Geriatr. Soc. 65, 2566– 2571 (2017).

10. Chen H, Quandt SA, Grzywacz JG, Arcury TA. A Bayesian multiple imputation method for handling longitudinal pesticide data with values below the limit of detection. Environmetrics. 2013;24:132–142. [PMC free article] [PubMed] [Google Scholar]

11. Little RJA, Rubin DB. Statistical Analysis with Missing Data. 2. John Wiley and Sons; Hoboken, NJ: 2002. [Google Scholar]

[redacted]

ANCOVA

```
PROC MIXED DATA = dataset;
CLASS Trt country;
MODEL dependent_variable = Trt center <covariates> / SOLUTION;
ESTIMATE '175 versus Placebo grp at M6/M9' Trt 1 0 -1/CL ALPHA=0.05;
ESTIMATE '350 versus Placebo grp at M6/M9' Trt 0 1 -1/CL ALPHA=0.05;
RUN;
```

Logistic Regression

```
PROC GENMOD DATA=dataset ORDER=data;
   CLASS Trt gender center;
   MODEL Responder(event='1')  =base Trt gender center / DIST=bin LINK=logit LRCI type3
   wald;
   ESTIMATE 'Arm1 versus Placebo grp at M6/M9' Trt 1 0 -1 / EXP;
   ESTIMATE 'Arm2 versus Placebo grp at M6/M9' Trt 0 1 -1 / EXP;
RUN;
```

Normality Check

```
PROC UNIVARIATE PLOT NORMAL DATA = dataset;
   BY Trt;
```

```
    VAR chg;
RUN;
```

Pearson Correlation Co-efficient and Scatter Plot

```
PROC CORR DATA = dataset;
    VAR PF  10 biomarker  1;
    ODS OUTPUT PearsonCorr = PE;
RUN;

Data _NULL_;
 SET PE;
 IF   n   = 1 THEN CALL SYMPUT ('pear', PUT (PF  10, BEST6.));
 ELSE              CALL SYMPUT ('pval', PUT (pbiomarker_1, PVALUE8.4));
RUN;

PROC SGPLOT DATA=dataset;
  SCATTER x= biomarker  1 y=PF_10;
  REG x= biomarker  1 y=PF  10;
  INSET ("Pearson" = "&pear" "P-value" = "&pval") /
      BORDERITLE="Correlation co-efficient" POSITION=topleft;
RUN;
```

Note: Repeat the same for the each treatment group and the correlation between each endpoint specified and specified biomarkers.

Appendix B - Short Physical Performance Battery (SPPB)

Summary performance score: Each of the three performance measured is assigned a score ranging from 0 to 4, with 4 indicating the highest level of performance and 0 the inability to complete the test. For the test of balance, participants are assigned a score of 1 if they can hold a side-by-side standing position for 10 seconds but are unable to hold a semi-tandem position for 10 seconds; a score of 2 is assigned if they can hold a semi-tandem position for 10 seconds, but are unable to hold a full-tandem position for 3 seconds; a score of 3 is assigned if they can stand in a full-tandem position for 3 seconds but less than 10 seconds; a score of 4 is assigned if they can stand in a full tandem position for 10 seconds.

Four categories are computed for walking speed and chair stands, according to cut-off points that are based on quartiles of the time to perform each task assessed in the EPESE. The time of the faster of two walks is scored as follows: > 8.7 sec = 1; 6.21 to 8.70 sec = 2; 4.82 to 6.20 sec = 3; < 4.82 sec = 4; a score of 0 is assigned to participants unable to perform the test. The time required to perform five chair stands is scored as follows: ≥ 16.70 sec = 1; 13.70 to 16.69 sec = 2; 11.20 to 13.69 sec = 3; ≤ 11.19 = 4. A score of 0 is assigned to participants unable to perform the task. A summary score ranging from 0 (worst performers) to 12 (best performers) is calculated by adding subscores from the walking speed, chair stands and balance tests.

Appendix C - SF-36v2 Coding and Scoring

All questions will be scored as per the raw data valued collected on the CRF with the following exceptions. Seven questions will have their coding inversed so that 5=1, 4=2, 3=3, 2=4 and 1=5. These questions are: 6, 9a, 9d, 9e, 9h, 11b and 11d.

The SF-36v2 scoring system relies on an assumption of linearity among the responses. However, for three of the 36 questions, it was found that the intervals were not evenly spaced among some of the qualitative responses so the values were recoded to preserve the linearity assumption. The questions affected are question 1 (General Health) and questions 7 and 8 (Bodily Pain).

Question 1 (General Health)

Q1 (verbatim responses)	Q1 (raw value)	Q1 (recoded)
Excellent	1	5.0
Very good	2	4.4
Good	3	3.4
Fair	4	2.0
Poor	5	1.0

Question 7 (Bodily Pain)

Q7 (verbatim responses)	Q7 (raw value)	Q7 (recoded)
None	1	6.0
Very mild	2	5.4
Mild	3	4.2
Moderate	4	3.1
Severe	5	2.2
Very severe	6	1.0

Question 8 will have its score inversed too, but also depends on the response given for question 7, in the following manner:

Q7 (raw value)	Q8 (verbatim responses)	Q8 (raw value)	Q8 (recoded)
1	Not at all	1	6
2, 3, 4, 5 or 6	Not at all	1	5
any	A little bit	2	4
any	Moderately	3	3
any	Quite a bit	4	2
any	Extremely	5	1

If question 7 is not answered then question 8 will have its score recoded to preserve linearity, in the following manner:

Q7 (raw value)	Q8 (verbatim responses)	Q8 (raw value)	Q8 (recoded)
missing	Not at all	1	6.0
missing	A little bit	2	4.75
missing	Moderately	3	3.5
missing	Quite a bit	4	2.25
missing	Extremely	5	1.0

If more that 50% of items for a sub-score then, for the purposes of calculating the subscale score, the sub-scale score is considered as missing. If less that 50% are missing then the missing items are replaced by the individual subjects mean score for the remaining items in the sub-scale. Imputed item scores are only used for calculating the sub-scale score.

This transformation converts the lowest and highest possible sub-scale scores to 0 and 100, respectively (0 represents worst possible health state and 100 represents best possible health state):

$$\text{Standardised Score} = \frac{(\text{sub-scale score} - \text{Lowest Possible Sub-scale Score})}{\text{Possible Sub-scale Score Range}} \times 100$$

The lowest sub-scale value and the range of sub-scales values for each subscale of SF-36 are as follows:

Description of item	Lowest possible score	Range of Sub-scale score
PF-10	10	20
Role limitations due to-physical problems	4	16
Bodily Pain	2	10
General Health perceptions	5	20
Vitality	4	16
Mental Health	5	20
Role-limitations due to emotional problems	3	12
Social Functioning	2	8

The following is an example of the SAS® code used to divide the 36 questions (recoded as necessary) into the 8 sub-classifications and to transform the values to a 0 – 100 scale.

```
data SF36;

   tot_PF=sum(of Q3A Q3B Q3C Q3D Q3E Q3F Q3G Q3H Q3I Q3J);
   tot_RP=sum(of Q4A Q4B Q4C Q4D);
   tot_BP=sum(of Q7 Q8);
   tot_GH=sum(of Q1 Q11A Q11B Q11C Q11D);
   tot_VT=sum(of Q9A Q9E Q9G Q9I);
   tot_SF=sum(of Q6 Q10);
   tot_RE=sum(of Q5A Q5B Q5C);
   tot_MH=sum(of Q9B Q9C Q9D Q9F Q9H);

   sf36_PF=((tot_PF - 10)/(20))*100;
   sf36_RP=((tot_RP - 4)/(16))*100;
   sf36_BP=((tot_BP - 2)/(10))*100;
   sf36_GH=((tot_GH - 5)/(20))*100;
   sf36_VT=((tot_VT - 4)/(16))*100;
   sf36_SF=((tot_SF - 2)/ (8))*100;
```

```
    sf36_RE=((tot_RE - 3)/(12))*100;
    sf36_MH=((tot_MH - 5)/(20))*100;

    label sf36_PF = "Physical functioning score (10 items)"
        sf36_RP = "Role-physical score (4 items)"
        sf36_BP = "Bodily pain (2 items)"
        sf36_GH = "General health score (5 items)"
        sf36_VT = "Vitality score (4 items)"
        sf36_SF = "Social functioning score (2 items)"
        sf36_RE = "Role-emotional score (3 items)"
        sf36_MH = "Mental health score (5 items)"
        ;
run;
```

Two summary measures of physical and mental health are constructed from the eight sub-scales, If any item in the calculation is missing, then the summary measure score is missing:

- Physical Health = (sf36_PF + sf36_RP + sf36_BP + sf36_GH) / 4
- Mental Health = (sf36_VT + sf36_SF + sf36_RE + sf36_MH) / 4

Appendix D - PAT-D Calculation

- Total score = sum all items and divide by 19
- Basic ADLs = sum items 7, 10, 11, 13, 14, 15, & 16 and divide by 7
- Mobility = sum items 2, 3, 5, 6, 18, 19 and divide by 6
- IADLs = sum items 1, 4, 8, 9, 12, 17 and divide by 6

*Items that receive a score of 6 (usually did not do for other reasons) are not included in scoring.

No fewer than 4 items should be used in calculating any subscale score. If fewer than 4 items are available then the score is considered as missing.

Appendix E - Mini Nutritional Assessment – Short Form

The Mini Nutritional Assessment – short Form (SF-MNA®) is a 6-question short form of the Mini Nutritional Assessment (MNA)[7].

The possible scores for each of the six questions A to F in the CRF are as follows:

Question A: Food intake declined over the past 3 months due to loss of appetite, digestive problem, chewing or swallowing difficulties
0 = Severe decrease in food intake
1 = Moderate decrease in food intake
2 = No decrease in food intake

Question B: Weight loss during the last 3 months
0 = Weight loss greater than 3 kg (6.6 lbs)
1 = Does not know
2 = Weight loss between 1 and 3 kg (2.2 and 6.6 lbs)
3 = No weight loss

Question C: Mobility
0 = Bed or chair bound
1 = Able to get out of bed/chair, but does not go out
2 = Goes out

Question D: Has suffered psychological stress or acute disease in the past 6 months?
0 = Yes
2 = No

Question E: Neuropsychological problems
0 = Severe dementia or depression
1 = Mild dementia
2 = No psychological problems

Question F: Body Mass Index (BMI) (weight in kg) / (height in m2)
0 = BMI less than 19
1 = BMI 19 to less than 21
2 = BMI 21 to less than 23
3 = BMI 23 or greater

Total score = sum of scores from all questions. If more than 50% of question responses are missing then the total score is considered as missing.

The total score ranges from 0 to 14, where score 0-7 represents malnourished, 8-11 represents at risk of malnutrition and 12-14 represents normal nutritional status.

The total score is collected in the CRF.

Nutritional Status
1 = Normal nutritional status
2 = At risk of malnutrition
3 = Malnourished

Appendix G - SarQoL Calculation

The SarQoL® questionnaire is composed of 55 items translated into 22 questions rated on a 4-point Likert scale. The questionnaire is scored on 100 points. Higher score reflects a higher quality of life. Items are organized into seven domains: domain 1 'Physical and Mental Health' with 8 items; domain 2 'Locomotion' with 9 items; domain 3 'Body Composition' with 3 items; domain 4 'Functionality' with 14 items; domain 5 'Activities of daily living' with 15 items, domain 6 'Leisure activities' with 2 items, and, at last, domain 7 'Fears' with 4 items. It produces 7 domain scores and 1 overall quality of life score, all between 0 and 100 points.

Appendix F - TSD-OC Calculation

The TSD-OC is composed of 36 items divided into seven sections (pain: 5 items; stiffness: 2 items; activities of daily living (ADL) and indoor mobility: 7 items; housework: 7 items; outdoor activities: 5 items; occupational activities: 4 items and social life: 6 items), which reflect the domains in which individuals experience the most common problems. Individuals are requested to provide a subjective assessment of their disability for each item on a 0–10 visual analogue scale (VAS), where 10 indicates the highest level of disability and 0 indicates no difficulties in performing the task.

The final 'disability scores' is defines as the sum of each item's raw score divided by the maximum possible score, expressed as a percentage according to the following linear transformation: (raw score/max score)*100. For example, a raw score of 90, being 25% of the total possible score (which corresponds to 360), results in a final disability score of 25.

In this way, disability scores can be calculated for each subsections pain, stiffness, ADL and indoor mobility.

Appendix S1 - A SAS macro for Implementing the Bayesian Multiple Imputation Method

```
/*************************************************************************
*************************************************************************
Program: impute car only data.sas
Purpose: Bayesian MI method for left-censored data
Author: [redacted]
Macro Note:
multivariate normal data is in wide format including id, varlist
values that are < LOD and need imputed are replaced with LOD
varlist: variables need imputation, whether MAR or CAR (coarsening at
random)
covlist: variables that have no missing data are considered as covariates
indexlist: binary variables that identify whether the observations of the
corresponding variable in the varlist needs imputed
this is useful when some values in the varlist are <LOD but do not need
to be imputed; otherwise it is just an indicator for data missing or not
noimpute: the value that indexlist variables takes if the observations of
the variables in the varlist do not need to be imputed
lboundlist: impute values > lbound, . represent no bound
LODlist: the limit of detection for the normal data, . represents no LOD
iteration-# of iterations for Gibbs sampler to converge
ndata-# of multiply imputed data sets for each chain
nchain-# of independent Markov chain
interval-# thinning parameter of the MCMC chain
output ndata imputed data sets from each of the nchain chains:
can use ndata data sets from nchain different chains or ndata data sets
from a single chain (data set called midata_org)
varlist, indexlist, lboundlist,lodlist need to have the same dimension
this program does not track parameter estimates due to memory constraints
*************************************************************************
*************************************************************************
*/

%macro coarsen(data,id,varlist, covarlist,indexlist, noimpute,
lboundlist,lodlist,iteration,ndata,nchain,interval, initialseed);
proc iml;

* read in data;
use &data;
  read all var {&varlist} into zorg;
  read all var {&id} into pid;
  %if &covarlist ne  %then %do;
  read all var {&covarlist} into xcov;
  %end;
  %if &indexlist ne  %then %do;
  read all var {&indexlist} into indexorg;
  %end;
close &data;
```

```
* sample size;
n=nrow(zorg);
call symput('totaln',left(char(n)));
* dimension of multivariate;
p=ncol(zorg);
call symput('dim',left(char(p)));

* dimension of covariates with no missing data;
%if &covarlist ne  %then %do;
pcov=ncol(xcov);
%end;
%else %do;
pcov=0;
%end;

* identify LOD for each variable;
%let li=1;
%do %while (%scan(&lodlist, &li, ' ') ne );
%let lod&li= %scan(&lodlist, &li, ' ');
lodorg=lodorg || &&lod&li;
%let li = %eval(&li+1);
%end;

* identify lbound for each variable;
%let lb=1;
%do %while (%scan(&lboundlist, &lb, ' ') ne );
%let lb&lb= %scan(&lboundlist, &lb, ' ');
lborg=lborg||&&lb&lb;
%let lb = %eval(&lb+1);
%end;


* identify the number of observed values for each z;
nobs=j(p,1,0);
* calculate mean and sd of observed data;
mean0=j(p,1,0);
sd0=j(p,1,0);
do k=1 to p;
obs=loc(zorg[,k] >lodorg[k]);
zobs=zorg[,k][obs];
mean0[k]=sum(zobs)/ncol(obs);
sd0[k]=sqrt(1/(ncol(obs)-1)*sum((zobs-mean0[k])##2));
nobs[k]=ncol(obs);
end;
nobs=shape(1:p,p,1)||nobs;
* sort by the number of observed values, descending order;
call sort(nobs,2,2);
varorder=nobs[,1];
* rearrange the original data matrix by the number of observed values;
z=j(n,p,0);
lod=j(1,p,0);
lbound=j(1,p,0);
```

```
%if &indexlist ne  %then %do;
indorder=j(n,p,0);
%end;
do l=1 to p;
z[,l]=zorg[,varorder[l]];
* rearrange lod accordingly;
lod[l]=lodorg[varorder[l]];
* rearrange lbound accordingly;
lbound[l]=lborg[varorder[l]];
* rearrange the indicator for index accordingly;
%if &indexlist ne  %then %do;
indorder[,l]=indexorg[,varorder[l]];
%end;
end;

* obtain order information to rearrange to the original variable order
later;
rorder=shape(1:p,p,1)||varorder;
call sort (rorder,2);
rvarorder=rorder[,1];

* starting number for labeling the imputed data sets;
mi=1;

seed=&initialseed;
* start multiple chains;
%do chain=1 %to &nchain;

* set the nmi to the number of MI for each chain;
%let nmi=&ndata;

* initialize, impute z_car and z_mar based on observed z;
* null matrix;
z0=   j(n,p,0);

do i=1 to n;
   do j=1 to p;
   if zorg[i,j]>lodorg[j] then z0[i,j]=zorg[i,j];
   else do;
   z0[i,j]=mean0[j]+ sd0[j] * normal(seed)   ;
   seed=seed+1;
   end;
   end;
end;

* rearrange the initialized data matrix by the number of observed values;
* start Gibbs sampler;
znew=j(n,p,0);
do l=1 to p;
znew[,l]=z0[,varorder[l]];
end;
```

```
* save memory space;
store z0;
free z0;

* gibbs sampler for each chain;
%do t=1 %to &iteration;
* for the (t+1) iteration;
* cycle through all p variables;
%do parm=1 %to &dim;
* regression for zj;
* design matrix for zj;
* multivariate in varlist;
%if &dim ne 1 %then %do;
index=setdif(1:p, &parm);
%if &covarlist ne  %then %do;
x=j(n,1,1)||znew[,index]||xcov;
%end;
%else %do;
x=j(n,1,1)||znew[,index];
%end;
%end; /* this goes with the dim ne 1 loop */

* univariate in varlist;
%if &dim = 1 %then %do;
%if &covarlist ne  %then %do;
x=j(n,1,1)||xcov;
%end;
%else %do;
x=j(n,1,1);
%end;
%end; /* this goes with the dim = 1 loop */

betahat=solve(x`*x, x`*znew[,&parm]);
residual=znew[,&parm]-x*betahat;
* standard deviation of residual variances;
sighat=sqrt((residual`*residual)/(n-p));
* generate a chi-square random variate;
call streaminit(seed);
chisq=rand('CHISQUARE', n-p);
seed=seed+1;
* a random draw of sig_jj from inverse chi-square;
sigtilda=sqrt((n-p))*sighat/sqrt(chisq);
* compute Cholesky decomposition;
chol=root(inv(x`*x));
* generate p+pcov random standard normal variates;
bzdev=j(p+pcov,1,0);
call randseed(seed);
call randgen(bzdev,'NORMAL',0,1);
seed=seed+1;
* a random draw of beta_j from multivariate normal;
betatilda=betahat+sigtilda*t(chol)*bzdev;
```

```
* compute mu_j;
mu=x*betatilda;

* save memory space;
store  x residual;
free  x residual;

* identify the observed data, including <LOD but do not need to be
imputed;
* note that z is the rearranged original matrix;
%if &indexlist ne  %then %do;
iobs=loc(z[,&parm] >lod[&parm] | indorder[,&parm]=&noimpute);
%end;
%else %do;
iobs=loc(z[,&parm] >lod[&parm]);
%end;
zobs=z[,&parm] [iobs];

* identify the MAR in z j;
imar=loc(z[,&parm] = .);
if ncol(imar) > 0 then do;
mumar=mu[imar];
* impute MAR, greater than lbound if there is one;
if lbound[&parm] ^= . then do;
* generate random uniform variates;
uniform_mar=j(ncol(imar),1,0);
call randseed(seed);
call randgen(uniform_mar,'UNIFORM');
seed=seed+1;
lmar=(lbound[&parm]-mumar)/sigtilda;
vmar=CDF('NORMAL', lmar, 0,1)+ uniform_mar # (1-CDF('NORMAL', lmar,
0,1));
zmar=mumar+sigtilda*probit(vmar);
end; /* this goes with the lbound ne . loop */

* impute MAR, if there is no lbound ;
if lbound[&parm] = . then do;
* generate random standard normal variates;
normal=j(ncol(imar),1,0);
call randseed(seed);
call randgen(normal,'NORMAL',0,1);
seed=seed+1;
zmar=mumar+sigtilda*normal;
end; /* this goes with the lbound = . loop */

end; /* this goes with the imar loop */

* no need to impute CAR;
if ncol(iobs)+ncol(imar)=n then do;
* combine the imputed data and update the design matrix;
z&parm.new=(t(iobs)||zobs) //  (t(imar)||zmar);
end;
```

```
if ncol(iobs)+ncol(imar) < n then do;

* identify CAR in z_j;
%if &indexlist ne  %then %do;
icar=loc(z[,&parm] <= lod[&parm] & z[,&parm] ^= . & indorder[,&parm] ^=
&noimpute);
%end;
%else %do;
icar=loc(z[,&parm] <= lod[&parm] & z[,&parm] ^= .);
%end;
if ncol(icar) >0 then do;
* generate random uniform variates;
uniform=j(ncol(icar),1,0);
call randseed(seed);
call randgen(uniform,'UNIFORM');
seed=seed+1;
mucar=mu[icar];
tcar=(lod[&parm]-mucar)/sigtilda;
*  impute CAR within the interval lbound and LOD;
if lbound[&parm] ^= . then do;
lcar=(lbound[&parm]-mucar)/sigtilda;
vcar=CDF('NORMAL', lcar, 0,1)+ uniform # (CDF('NORMAL', tcar, 0,1)-
CDF('NORMAL', lcar, 0,1));
end; /* this goes with the lbound ne . loop */

*  impute CAR < LOD if there is no lbound ;
if lbound[&parm] = . then do;
vcar=uniform # CDF('NORMAL', tcar, 0,1);
end;

* Note that occasionally vcar is close to 1 and cause invalid argument to
probit function;
* print lcar tcar vcar mucar;
test=loc(vcar > 0.9999999);
if ncol(test) > 0 then do;
vcar[test]=0.9999999;
end;

zcar=mucar+sigtilda*probit(vcar);
end; /* this goes with the icar loop */

* combine the imputed data and update the design matrix;
if ncol(imar)>0 & ncol(icar) >0 then do;
z&parm.new=(t(iobs)||zobs) // (t(imar)||zmar) //  (t(icar)||zcar);
end;
else if ncol(imar)=0 & ncol(icar) >0 then do;
z&parm.new=(t(iobs)||zobs) //  (t(icar)||zcar);
end;
else if ncol(imar)>0 & ncol(icar) =0 then do;
z&parm.new=(t(iobs)||zobs) //  (t(imar)||zmar);
end;
```

```
end; /* this goes with the < n loop */

* sort the imputed data by id;
call sort(z&parm.new,1);
* remove id;
z&parm.mi=z&parm.new[,2];

* update the data matrix;
* repalce z j from iteration t with z_j from iteration (t+1);
%if &dim ne 1 %then %do;
zcom= znew[,index]||z&parm.mi;
* organize the order of the columns from z1 to zp;
order=index|| &parm ;
%end; /* this goes with the dim ne 1 loop */

%if &dim = 1 %then %do;
zcom= z&parm.mi;
order= &parm ;
%end; /* this goes with the dim = 1 loop */

zcom=t(order//zcom);
call sort (zcom,1);
znew=t(zcom[,2:n+1]);

* save memory space;
store mu z&parm.new z&parm.mi zcom;
free mu z&parm.new z&parm.mi zcom;

%end; /* this goes with the parm loop */

* output the m data sets after the convergence for specified intervals
apart;
%if (&t=%eval(&iteration-&interval * (&nmi-1)) ) %then %do;
impute=j(n,1,&chain)||j(n,1,mi)||pid||znew;
* rearrange to the original variable order;
impute_org=j(nrow(impute),p+3,0);
do m=1 to p;
impute_org[,m+3]=impute[,rvarorder[m]+3];
end;
impute_org=impute[,1:3]||impute_org[,4:p+3];

midata org=midata_org//impute_org;
mi=mi+1;

%let nmi=%eval(&nmi-1);

* save memory space;
store impute impute_org;
free impute impute_org;

%end; /* this goes with the output loop */
```

```
%end; /* this goes with the iteration loop */

%end; /* this goes with the nchain loop */

* get variable names for output dataset;
vlist="&varlist";
 %do vi=1 %to &dim;
 vname&vi='mi_'+scan(vlist, &vi);
 call symputx("varname&vi", vname&vi);
%end;

* note that all variables in the output dataset contained imputation of
missing and left-censored data;
create midata_org from midata_org
[colname={'chain' '_imputation_' "&id"
 %do vi=1 %to &dim;
"&&varname&vi"
%end;
 }];

append from midata_org;

quit;

%mend;


* %coarsen(data=walk,
           id=pidn,
           varlist=speed_m3 speed_m69 speed_lod_m3 speed_lod_m69,
           indexlist=complete_m3 complete_m69 complete_m3 complete_m69

           noimpute=1,
           covarlist=speed_m0,
           lboundlist=0 0 0 0,
           lodlist=. . 0.44 0.44,
           iteration=2000,
           ndata=5,
           nchain=5,
           interval=100,
           initialseed=155504946);
```

Appendix S2 - Sample SAS Code for PROC MIANALYZE

```
* analyze each imputed dataset;
proc mixed data=midata method=reml;
class pid Arm month;
model speed_mi = speed_base Arm month Arm*month /solution;
repeated /subject=pid type=un;
lsmeans Arm month Arm*month;
estimate 'PA-HE month 3'  Arm -1 1 Arm*month -1 0 1 0;
estimate 'PA-HE month 69' Arm -1 1 Arm*month 0 -1 0 1;
estimate 'PA-HE overall'  Arm -1 1;
ods output SolutionF=TEMP1 LSMeans=TEMP2  Estimates=TEMP3;
by  imputation_;
run;

* obtain overall significance of the effects;
proc mianalyze parms(classvar=full)=TEMP1 ;
class Arm month;
modeleffects Intercept speed_base Arm month Arm*month;
ods output ParameterEstimates=TEMPRESULT1;
run;

* obtain least square means;
proc mianalyze parms(classvar=full)=TEMP2;
class Arm month;
modeleffects Arm month Arm*month;
ods output ParameterEstimates=TEMPRESULT2;
run;

* obtain estimates for the intervention effect;
proc sort data=TEMP3 ;
by label _imputation_;
run;
proc mianalyze data=TEMP3;
MODELEFFECTS Estimate;
StdErr StdErr;
by label;
ods output ParameterEstimates=TEMPRESULT3;
run;
```