Statistical Analysis Plan

Study title:

Dose finding phase IIb study of Bavisant to evaluate its safety and effiCacy in treAtment of exceSsive daytime sleePiness (EDS) in PARkinson's Disease (PD). CASPAR study.

Protocol Number: BB-2001-201b

Linical-Code: BEN01

Control Versions

STATUS AND VERSION NUMBER	REASON FOR CHANGE- LOCATION OF CHANGE	DATE	MODIFIED BY
Draft 0.1	First document	11MAY2018	Eduardo Sobreviela
Draft 0.2	Comments for Draft 0.1	16JUL2018	Eduardo Sobreviela
Final 1.0	Comments to Draft 0.2	25OCT2018	Eduardo Sobreviela

PREPARED BY:

Eduardo Sobreviela,
Director Biostatistics, Linical — Date

REVIEWED BY:

Dr Kevin Carroll
Director, KJC Statistics Ltd. — 31 October 2018 — Date

APPROVED:

Dr Leo James
Medical Director, BenevolentAI Bio — Date

INDEX

1 INTRODUCTION

The objective of this statistical analysis plan (SAP) is to specify the statistical analysis in more detail than stated in the protocol for the trial. The statistical analysis plan does not change the analysis described in the protocol, but it should be precise enough to serve as a guideline for statistical programming and creation of tables. The statistical analysis plan (SAP) will ensure the credibility of the study findings by specifying the statistical approaches to the analysis of the double-blind data prior to database lock.

This Statistical Analysis Plan was developed based on the International Conference on Harmonization (ICH) E3 and E9 Guidelines and with reference the valid protocol (final version v1.0 amendment 03, dated 23Aug2017). Deviations from the planned methods should also be summarized in Section 14 of this SAP. Any deviations during the analysis and reporting process from the current statistical analysis plan will be described and justified in the final report

2 RESPONSIBILITIES

The statistical analysis, interpretation and reporting will be the responsibility of the assigned project biostatistician.

All analyses and reporting will be subject to a quality check process.

3 STUDY OBJECTIVES

3.1 Primary Objectives

To assess the efficacy of Bavisant compared to placebo after a 6-week treatment period on the excessive daytime sleepiness in Parkinson's disease.

3.2 Secondary objectives

To assess the efficacy and safety assessment of Bavisant compared to placebo after 2 weeks and 6 weeks of treatment. The efficacy assessment will include excessive daytime sleepiness, motor control, and depression.

3.3 Exploratory Objectives

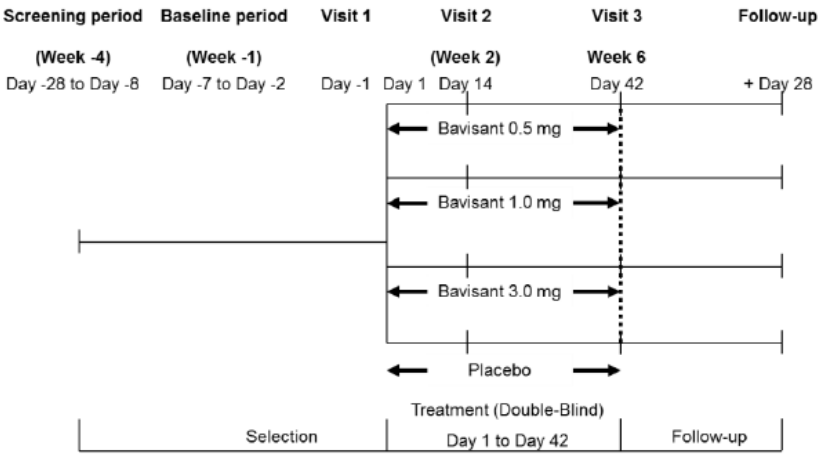
Exploratory objectives will include the assessment of free living activity in subjects on Bavisant compared to placebo after 6 weeks of treatment using a wrist-worn actigraphy.

4 DESIGN OF THE STUDY

This phase 2b study is designed as multicentre, multinational, randomized, double blind, parallel group and placebo controlled with three doses of Bavisant (0.5, 1, and 3 mg/d) in subjects with excessive daytime sleepiness with Parkinson's disease. This study will involve approximately 230

subjects from around 48 sites globally. Eligible subjects will be randomly allocated at a ratio of 1:1:1:1 to either Bavisant (at doses of 0.5, 1 and 3 mg/d) or placebo. After signing the study informed consent, subjects meeting the UK PDS Brain Bank Clinical Diagnostic Criteria for Parkinson's disease (see details in the protocol Appendix 18.2) will enter a screening period (lasting up to 3 weeks) followed by a baseline period (lasting up to 1 week). Key inclusion / exclusion criteria will be verified to determine subject eligibility at the screening, baseline periods and prior to randomisation, including the suicide risk assessment following the FDA Guidance for Industry on Suicidal Ideation and Behavior. There is a 6 weeks treatment period, starting on Day 1, followed by a safety follow-up at least 4 weeks after completion of the treatment period. Additional unscheduled visits may be required based on the investigators' judgement at the end of study visit or at the safety follow-up. For early termination, all the assessments of the treatment week 6 visit should be performed, if possible, including the safety follow-up 4 weeks after the last investigational medicinal product (IMP) administration. The figure below describes the study flow graphically.

Figure 1



Inclusion / Exclusion	X	X	X			
Randomisation			X			
Laboratory tests	X				X	
Pharmacogenomic sample	X					
Eye examination	X				X	
C-SSRS	X			X	X	X
ESS	X		X	X	X	
MoCA, FSS	X				X	
HAM-D	X			X	X	X
BPRS+	X			X	X	X
SCOPA, PDSS, UPDRS			X	X	X	
MWT, polysomnography		X			X	
Free Living Activity		X			X	
Adverse Events	X	X	X	X	X	X

4.1 STUDY WINDOWS

The process described here will be applied for visit re-allocation except any other approach was described in section 10 or 13 of this SAP for the corresponding statistical analysis. No re-allocation will be performed for nominal visits already provided in the clinical database (visit 1, visit 2 and visit 3). However, unscheduled visits or Early Termination visit may be used to complete any missing information not gathered in the scheduled visit 2 and visit 3, according to the table below.

Table 1: Reallocation time windows definition

Scheduled Visit Post-baseline	Target Study Day	Analysis window in study days
Visit 2 (week 2)	14	10 to 21
Visit 3 (week 6)	42	22 to 45

Unscheduled visits occurring outside of the analysis windows will not be re-allocated.

5 SAMPLE SIZE AND POWER ESTIMATION

A total of 200 completed subjects will be evaluated for the study (50 completed subjects per treatment group, with expected early withdrawal rate of around 15%).

The assumptions used for the sample size calculation are a mean absolute change in the Epworth Sleepiness Scale (ESS) from baseline to the end of the 6-week treatment period of at least 3.0 points (minimum change considered as clinically relevant) and a standard deviation of 5.0 (based on a comparable study of Modafinil in a similar population recruited at one centre (3) which showed a standard deviation in the ESS between 3.5 at baseline and 4.8 at the end of the treatment period) and a reduction in ESS of 2.7 points for subjects treated with Modafinil.

The confirmation of the hypothesis of detecting a mean difference in ESS of 3.0 points vs. baseline requires a sample size of 25 subjects per group (intragroup Δ=3.0; SD=5.0; two-side α=0.05; β=0.20; power=80%; one-sample t-test).

The additional endpoint of showing a statistically significant difference of at least 3.0 points in ESS between Bavisant and placebo requires a sample size of 50 subjects per group (between treatment Δ=3.0; SD=5.0; two-side α=0.05; β =0.20; power=80%; two-sample equal-variance t-test).

The effect size of Bavisant in EDS is unknown and smaller changes in ESS may be clinically useful according to its profile, and higher variability is expected in this multicentre study; so the sample size of 200 completed subjects (50 per treatment group) is proposed in order to cover both the confirmation of an intragroup change of at least 3.0 points vs. baseline, and the between-group difference of at least 3.0 points change vs. placebo.

The Maintenance of Wakefulness Test (MWT) will be performed in a minimum of 20 subjects per group (80 overall) in order to be enough to show statistically significant differences if the expected MWT absolute change from baseline differs at least in one standard deviation (effect size = 1.0).

6 ANALYSIS SETS

The safety population (SAF) will be defined as all subjects who have received at least one dose of the study treatment.

The intent to treat population (ITT) will be defined as all randomized subjects who have taken at least one dose of study treatment.

The per protocol population (PP) will be defined as all subjects in the FAS who do not experience a major protocol deviation.

Protocol deviations will be defined as deviations from the procedure outlined in the protocol. Protocol deviations will be identified for each patient in the Data (blind) Review & Evaluability Determination meeting will be held prior to the treatment unblinding. They will be assessed as “minor” or “major” according to the possible impact expected on efficacy results. Minor and major protocol deviations will be fully documented in the Protocol Deviation Document prior to unblinding. Specific data sets will be produced and transferred to Biostatistics department in order to allow the appropriate definition for the Per Protocol Population.

Since the reasons for excluding patients from the analysis cannot be entirely foreseen at the time of writing the statistical analysis plan evolving trial conduct may require some further definition of patients to be excluded, at the end of the trial, but prior to unblinding.

Aspects to be considered (but not limited to) when determining the evaluability of data are as follows:

- Inclusion and exclusion criteria
- Acceptable timings for visit dates and measurements
- Compliance with treatment
- Incorrect randomization/treatment
- Concomitant therapies

As far as possible ITT principle will be guaranteed so it is not expected to have protocol deviations excluding patients from the ITT population.

7 PRESENTATION OF DATA

Tables and listings will be produced in accordance with the principles outlined by the ICH E3 guideline.

Categorical data will be summarized with absolute and relative frequencies (percentage), missing data will not be considered in the percentage calculation, and numerical data will be summarized with the number, mean, median, standard deviation (SD), the range (minimum and maximum), and the first, third quartiles and 95% confidence interval (if applicable). Throughout the presentation of the study results the four treatment groups will be displayed separately. For safety a new column summarizing data for the three active treatment arms together will be added.

8 SOFTWARE

All statistical analyses, listing, tabulations and figures will be producing using SAS® Version 9.4 or higher.

9 EFFICACY ENDPOINTS

Baseline efficacy value will be defined as the last measure done prior to the randomization.

9.1 Primary Endpoint for Efficacy

ESS mean absolute change from baseline to the end of the 6-week treatment period will be defined as the Total ESS score at visit 3 – Baseline ESS score. ESS questionnaire consists of 8 questions which subjects are asked to rate their usual chances of dozing off or falling asleep while engaged in eight different activities. The answers will be provided on a 4-point scale: 0=' would never doze', 1=' slight chance of dozing', 2=' moderate chance of dozing', or 3=' high chance of dozing'. Total ES is the sum of the entries for all of the 8 items-questions. If one or more item-scores are missing, it would not be imputed and this ESS total score will be considered invalid. When ESS was missing in some scheduled post-baseline visit (visit 2 or visit 3) any other ESS total score measured in whether an unscheduled visit or in the early termination visit will be considered according to the re-allocation windows rules presented in section 4.1. Remaining missing for baseline ESS will be imputed to the baseline average, and remaining post-baseline evaluations will be considered as treatment failure, a Multiple Imputation Jumping to Reference Approach will be used (see section 10.4.2 for reference). Negative absolute change will mean improvement.

9.2 Secondary Endpoints for Efficacy

9.2.1 Epworth Sleepiness Scale (ESS)

- ESS mean absolute change from baseline to the end of the 2-week treatment will be defined as Total ESS score at visit 2 – Baseline ESS score.

- ESS clinical response, defined as:

    - First approach: ESS absolute decrease from baseline of at least 3.0 points, after 2 and 6 weeks of treatment.

    - Second approach: ESS ≤ 10 after 2 and 6 weeks of treatment.

    - Third approach: Either ESS ≤ 10 after 2 and 6 weeks of treatment or ESS absolute decrease from baseline of at least 3.0 points after 2 and 6 weeks of treatment.

- Mean relative change in the ESS from baseline to the end of the 2-week and 6-week treatment periods (percentage of absolute decrease compared to baseline ESS). For each patients the relative change will be calculates as:

    Relative change: (post-baseline evaluation - baseline evaluation) / baseline evaluation

The same missing imputation strategy as used for primary endpoint will be used for the ESS secondary endpoints (see section 9.1 and 10.4.2 for reference).

9.2.2 SCOPA-Sleep

SCOPA-Sleep mean absolute change from baseline to the end of the 6-week treatment will be defined as Total SCOPA-sleep score at visit 3 – Baseline SCOPA-sleep score.
When SCOPA-sleep is missing in some scheduled post-baseline visit (visit 2 or visit 3) any other SCOPA-sleep value measured in whether an unscheduled visit or in the early termination visit will be considered according to the re-allocation windows rules presented in the point 4.1.
The SCOPA-Sleep is a specific rating scale for assessing night-time sleep (NS) and daytime sleepiness (DS) in the past month. The NS subscale addresses NS problems in the past month and includes 5 items with 4 response options. Subjects have to indicate how much they were bothered by particular sleep problems, ranging from 0 (not at all) to 3 (a lot). The 5 items address sleep initiation, sleep fragmentation, sleep efficiency, sleep duration, and early wakening. The maximum score of this scale is 15, with higher scores reflecting more severe sleep problems. One additional question evaluates overall sleep quality on a 7-point scale (ranging from "slept very well" to "slept very badly"). The score on this item is not included in the score of the NS scale but is used separately as a global measure of sleep quality. The Daytime sleepiness (DS) subscale addresses DS problems in the past month and includes 6 items with 4 response options, the maximum score of this scale is 18 with higher scores reflecting more severe sleep problems. If more than 75% of the items for each sub-scale were missing the NS or DS will not be calculated and will be considered as missing as well. This threshold could be re-evaluated in advance to the data base lock according to the missingness rate in the data.

The same definition will apply for the 2-week evaluation.

9.2.3 Parkinson's Disease Sleep Scale (PDSS-2)

PDSS-2 mean absolute change from baseline to the end of the 6-week treatment will be defined as Total PDSS-2 score at visit 3 – Baseline PDSS-2 score. When PDSS-2 was missing in some scheduled post-baseline visit (visit 2 or visit 3) any other PDSS-2 value measured in whether an unscheduled visit or in the early termination visit will be considered according to the re-allocation windows rules presented in the section 4.1. The Parkinson's Disease Sleep Scale (PDSS) is a scale for the assessment of sleep disorders in PD in the past week. The scale has 15 items and subjects mark their response (0='Very Often', 1='Often', 2='Sometimes', 3='Occasionally', 4='Never') to each item. Per scoring calculation item 1 should be reverted (4='Very Often', 3='Often', 2='Sometimes', 1='Occasionally', 0='Never'). Total PDSS-2 SCOPA score will be calculated as the sum of the 15 items scores. If more than 75% of the items were missing the Total PDSS-2 SCOPA score will not be calculated and will be considered as missing. This threshold could be re-evaluated in advance to the data base lock according to the missingness rate in the data.

The same definition will apply for the 2-week evaluation.

9.2.4 Maintenance of Wakefulness Test (MWT)

The MWT is an evaluation used as a quantitative polysomnographic (PSG) measurement of daytime wakefulness/somnolence during soporific circumstances. During each trial, subjects will sit quietly on the bed with both back and head supported by a pillow, and subjects will be asked to look directly

ahead and try to stay awake as long as they can. If they fall asleep, they will be woken up after 90 seconds and the trial will end if the subjects do not fall asleep within 40 minutes. Start and stop times for each trial, sleep latency, total sleep time, stages of sleep achieved for each trial, and the mean sleep latency (the arithmetic mean in minutes of the four trials) will be recorded, considering sleep latency as the time from lights out until the first epoch of greater than 15 sec of cumulative sleep in a 30-sec epoch of either 3 consecutive epochs of stage 1 sleep, or one epoch of any other stage of sleep. Absolute change in the MWT from baseline to the end of the 6-week treatment period. Mean MWT sleep latency scores will be calculated for each of the 4 sleep trials. Also the total sleep latency scores average will be calculated for each patient according to the 4 round MWT sleep latency scores. If some of the latency scores are missing the total sleep latency score average will be calculated just considering the available data and dividing it by the number of non-missing tests. Sleep latency scores will be analysed and presented in minutes.

9.2.5 Polysomnography

Mean absolute change in the polysomnography parameters (as defined below) from baseline to the end of the 6-week treatment period:

Sleep Stage Parameters:
Total Recording Time (TRT) is defined as the time in minutes from “lights out” to “lights on”. Total Sleep Time (TST) is the total time in minutes asleep after sleep onset. To determine the how well the patient slept, the Sleep Efficiency (SE) is calculated by dividing the TST by the TRT and multiplying by 100. Sleep studies are recorded on 30 second “epochs”. Sleep onset is defined as the first epoch scored as any stage other than stage W. Sleep Latency (SL) is the time from “lights out” to the sleep onset. Latencies to sleep stages are determined from sleep onset to the first epoch of that sleep stage. Wake after Sleep Onset (WASO) is the time awake after sleep onset until “lights on”. To determine the percentage time spent in each of the sleep stages during the sleep study, the total minutes of the sleep stage is divided by the TST and multiplied by 100.

Sleep information:
Percentage of TST Ni (for i=1 to 3) is calculated as a percentage: (Minutes Ni / TST * 100)
Percentage of TST REM is calculated as a percentage: (Minutes REM / TST * 100)

Clinical Event Parameters:
Event (or arousal) index will be calculated as the number of events (arousal number) divided by TST and multiplied by 60 (events per hour).

Events Index= (Events/TST)*60.
Arousal Index= (Arousal number/TST)*60.

Index will be calculated for the events and arousal detailed below:

- Apnoea
- Hypopnea
- Apnoea + Hypopnea
- Limb Movement

- Periodic Limb Movement
- Arousals index
- Spontaneous Arousals index
- Apnoea Arousals index
- Hypopnea Arousals index
- LM Arousals index
- PLM Arousals index
- Desaturation Arousals index
- Snore Arousals index
- Respiratory Arousals index
- Respiratory Effort Related Arousal (RERA) index
- User Defined Arousals index
- Total Arousals index

9.2.6 Unified Parkinson's disease Rating Scale Part III (UPDRS Part III)

UPDRS III mean absolute change from baseline to the end of the 6-week treatment will be defined as Total UPDRS III score at visit 3 – Baseline UPDRS III score. Part III contains 33 scores based on 18 items, several with right, left or other body distribution scores.

When UPDRS is missing in some scheduled post-baseline visit (visit 2 or visit 3) any other UPDRS value measured in whether an unscheduled visit or in the early termination visit will be considered according to the re-allocation windows rules presented in the section 4.1. The same analysis will be repeated for the 2-week evaluation. If more than 75% of the scores are missing then the total motor examination score will not be calculated and considered as not evaluable. The same rule will be applied for the subscales calculation. This threshold could be re-evaluated in advance to the data base lock according to the missingness rate in the data.

9.2.7 Hamilton Rating Scale for Depression (HAM-D)

Mean absolute change in the depression HAM-D score (as defined below) from baseline to the end of the 2-week and 6-week treatment period and safety follow-up (defined as 28 days ± 2 after Visit 3 or early termination visit). Although the HAM-D form lists 21 items, the scoring is based on the first 17. For the total sum, 8 items are scored on a 5-point scale, ranging from 0 (not present) to 4 (severe), and the remaining 9 are scored from 0 to 2. HAM-D total score is calculated as the sum of the first 17 items. This threshold could be re-evaluated in advance to the data base lock according to the missingness rate in the data.If more than 75% of the scores are missing then the total motor examination score will not be calculated and considered as not evaluable. The same rule will be applied for the subscales calculation.

9.2.8 Montreal Cognitive Assessment (MoCA)

Mean absolute change in the MoCA total score as defined below from screening to the end of the 6-week treatment period.
It assesses different cognitive domains: attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. Sum all sub scores listed on the right-hand side. Add one point for an individual who has 12 years or fewer of

formal education, for a possible maximum of 30 points. The total possible score is 30 points; a score of 26 or above is considered normal.

9.2.9 Fatigue Severity Scale (FSS)

Mean absolute change in the Fatigue Severity Scale, as defined below.
The FSS is a self-administered questionnaire with 9 items investigating the severity of fatigue in different situations during the past week. There are 9 items ranging from 1 to 7, where 1 indicates strong disagreement and 7 strong agreement, and the final score is calculated as the mean value of the 9 items. If more than 75% of the items are missing then the FSS total score will not be calculated as will be considered as not evaluable. This threshold could be re-evaluated in advance to the data base lock according to the missingness rate in the data

9.2.10 Berlin Questionnaire (BQ)

Percentage of patients with BQ high risk (as defined below) at week 2 and 6 will be determined.

This questionnaire consists of 3 categories (questions 2 to 6 about snoring [category 1], questions 7 to 9 about daytime somnolence [category 2], and question 10 about hypertension and BMI [category 3]) and the risk is based on the responses to individual items and overall scores in the symptom categories.

Categories and Scoring:
Category 1: items 2, 3, 4, 5 and 6;
Item 1: if '**Yes**', assign **1 point**
Item 2: if '**c**' or '**d**' is the response, assign **1 point**
Item 3: if '**a**' or '**b**' is the response, assign **1 point**
Item 4: if '**a**' is the response, assign **1 point**
Item 5: if '**a**' or '**b**' is the response, assign **2 points**
Add points. *Category 1 is positive if the total score is 2 or more points*.

Category 2: items 7, 8, 9 (item 9b should be noted separately).
Item 7: if '**a**' or '**b**' is the response, assign **1 point**
Item 8: if '**a**' or '**b**' is the response, assign **1 point**
Item 9: if '**a**' is the response, assign **1 point**
Add points. *Category 2 is positive if the total score is 2 or more points.*

Category 3 is positive if the answer to item 10 is '**Yes**' or if the BMI of the patient is greater than 30kg/m^2. *(BMI is defined as weight (kg) divided by height (m) squared, i.e.., kg/m^2).*

High Risk: if there are 2 or more categories where the score is positive.
Low Risk: if there is only 1 or no categories where the score is positive.

9.3 Exploratory Efficacy Endpoints

- Free living activity assessed by means of wrist-worn actigraphy.

9.4 Analysis sets Analysed for Efficacy

Primary Analysis, and sensitivity analysis as described above will be based in the ITT Population. In the MCP-Mod sensitivity analysis, only patients in the ITT with ESS evaluations at week 6 will be considered. Primary analysis will be repeated in the Per Protocol Population also which will be considered as a sensitivity analysis.

10 STATISTICAL METHODOLOGY FOR EFFICACY ENDPOINTS

10.1 Primary Efficacy Analysis

ESS mean absolute change from baseline to the end of the 6-week treatment period will be analysed with an ANCOVA model, with the baseline ESS as a covariate, region and treatment as a factor. Least-squared (LS) absolute mean changes will be presented by treatment group along with their associated 95% CIs and 2sided p-values; these data will then provide estimated for the within arm change from baseline. The difference between each of the 3 dose arms and the placebo arms in Least-squared (LS) absolute mean changes will also be presented together with the corresponding 95% confidence interval and 2-sided p-value. These paired comparisons will be performed without any multiplicity adjustment. Also, the 4 degrees of freedom (DF) associated with the overall F-test for differences between the treatment arms will be decomposed using Type I sums of squares into three orthogonal single DF contrasts to estimate the linear, quadratic and cubic components of any potential dose response. Baseline data will be imputed with the baseline ESS average. Multiple Imputation Jumping to Reference method will be used to impute post-baseline data (see section 10.4.2fir further reference).

10.1.1 Sensitivity analyses missing reductions imputed to 0:

Same primary analysis will be repeated imputing post-baseline missing ESS evaluations to baseline data that is reduction equal to 0.

10.1.2 Sensitivity analyses MMRM Model:

A mixed-effect model with repeated measures (MMRM) approach will be used, under the missing at random framework carried out using an adequate contrast at Week 2 or Week 6, accordingly. The model will include fixed categorical effects of treatment group, region, visit and treatment-by-visit interaction as well as the continuous fixed covariates of mean baseline ESS. This MMRM model will be run with an unstructured correlation matrix to model the within subject errors. In this case missing values will not be imputed, since (MMRM) is handling missing data through the Maximum Likelihood estimation method.

Unstructured covariance matrix is assumed to explain within subject covariance in MMRM method in protecting the type I error rates. In the event if the assumption of unstructured covariance (UN) matrix results in non-convergence of the model, only then the following covariance structures will be considered in the order mentioned below.

The order of covariance structure to be considered is:

1. Compound Symmetry (CS)

10.1.3 Sensitivity analysis using MCP-Mod

This sensitivity analyses will be done for patients in the ITT population having non-missing ESS evaluation in the week 6.

MCP-Mod will be produced following the steps below, (see Pinheiro J 2014 (1) for reference):

I. Candidate dose-response shapes
II. Derive optimum contrast coefficients to maximize power to finding dose-response association
III. Identify whether dose-response signal could be established
IV. Model Selection

I. Candidate dose-response shapes

The three models detailed below will be considered in the dose-response evaluation.

a. $Emax_1$ (D_{50}=0.2): according to Thomas et al. Emax model can adequately describe the observed data in many situations. The Emax model dose–response curve is defined as:

$$R = E_0 + (D \times E_{max}) / (D + ED_{50})$$

where:

R = response
D = dose
E_0 = basal effect corresponding to the response when dose of drug is 0.
E_{max} = maximum effect attributable to the drug
ED_{50} = dose which produces half of the Emax

b. Exponential (δ=1):

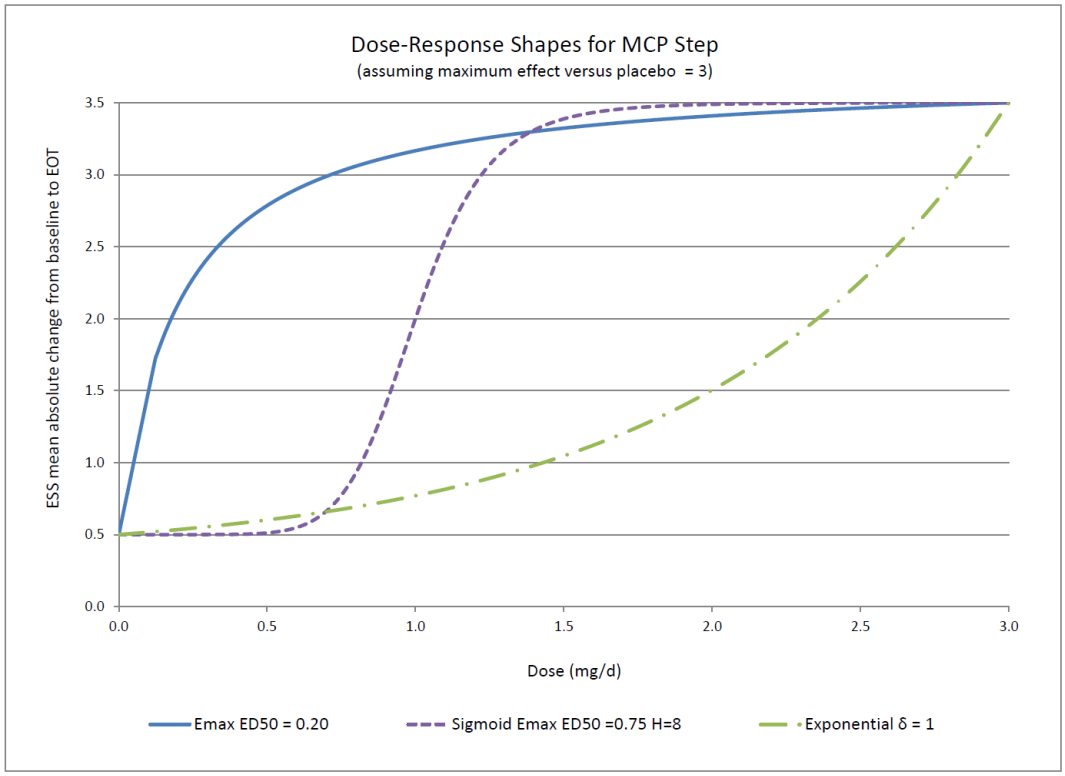
$$R=E_0+ E_1*(EXP(D/\delta)-1)$$

E_0= basal effect corresponding to the response when dose of drug is 0
$E_{1=}$ slope parameter for exponential model
D=dose
δ= parameter, controlling the convexity of the model

c. Sigmoid Emax (ED50=1, H=8):

E_0= basal effect corresponding to the response when dose of drug is 0
E_{max}= maximum effect attributable to the drug
ED_{50}= dose which produces half of the Emax
H=Hill parameter, determining the steepness of the model at the ED50

The following plot illustrates visually the shape of these three models used in the MCP part to test the dose-response signal:



II. Optimal Contrast identification

In order to obtain the vector of optimal coefficients $\hat{c}$ for each dose-response signal contrast, the following equation will be maximized as proposed by Pinheiro, Bornkamp, Bretz (2006): $c'\mu/\sqrt{\sum c_i^2/n_i}$. The following steps will be performed for each curve (e.g. $Emax_1$):

1. A dataset with one single record and four variables μ_1, μ_2, μ_3 and μ_4 will be created where μ_1, μ_2, μ_3 and μ_4 are means for placebo, 0.5, 1, and 3 mg/d respectively compatible with the shape of the given curve (e.g. $Emax_1$).
2. The goal is to obtain the values c_1, c_2, c_3 and c_4 that maximize $(c'\mu)^2/(\sum c_i^2/n_i)$ where n_1, n_2, n_3 and n_4 are the sample sizes for placebo, 0.5, 1, and 3 mg/d respectively. To do so, the NLMIXED procedure will be applied on the above-mentioned one-record dataset to obtain the "maximum likelihood" parameters c_1, c_2, c_3 and c_4 using the following likelihood function:

```
ll=(mu1*c1 + mu2*c2 + mu3*c3 + mu4*c4)**2/
(c1**2/n_1 + c2**2/n_2 + c3**2/n_3 + c4**2/n_4);
```

   There are infinite solutions to the above problem, so we will hold one coefficient to an arbitrary value (e.g. c_1=-10).
3. Finally, the obtained coefficients will be normalized as follows (shown for c_1):

```
nc1=c1/(sqrt(c1**2 + c2**2 +c3**2 + c4**2));
```

For validation purposes, the obtained coefficients will be compared with the ones obtained using the dose-finding R package developed by te authors of MCP-Mod.

III. Dose Response Test

For each of the candidate models, dose-response association will be evaluated using the optimal coefficients identified in the previous step. These contrasts will be tested using the PROC GLIMMIX procedure at an one-sided alpha level of 0.025 and multiplicity adjusted p-values will be obtained through parametric resampling ESS mean absolute change from baseline to the end of the 6-week treatment period will be included in the model as dependent variable, dose as fixed factor and baseline ESS as covariate. SAS code like the code below will be used:

```
PROC GLIMMIX;
CLASS DOSES;
MODEL ESS = DOSES BASELINE /NOINT;
LSMEANS DOSES;
ESTIMATE "EMAX1"        DOSES &C11. &C12. &C13. &C14.,
         "EXPONENTIAL"  DOSES &C21. &C22. &C23. &C24.,
         "SIGMOID EMAX" DOSES &C31. &C32. &C33. &C34.
          / UPPERTAILED ADJUST=SIMULATE(NSAMP=100000 SEED=1234) ;
RUN;
```

IV. Model Selection

The maximum contrast test method will be used to select the best model to be used in the Mod stage. The selected model will be fitted with PROC NLMIXED procedure using SAS code similar to the script below (shown for the Emax model only):

```
PROC NLMIXED;
BOUNDS S2>0, ED50 >0;
PARMS  S2 BETA E0 EMAX ED50;
IF DOSE >0 THEN MEAN = E0+(EMAX*DOSE)/(ED50+DOSE) + BETA*BASELINE;
ELSE IF DOSE = 0 THEN MEAN= E0 + BETA*BASELINE;
MODEL ESS ~ NORMAL(MEAN, S2);
RUN;
```

A graph will be produced presenting the modelled dose-response curve; point estimate and 95% confidence limits will be displayed.

10.1.4 Sensitivity analyses Rank ANCOVA

Similar to the primary analysis a Rank ANCOVA will be modelled in order to provide evidence of the treatment effect in case of violation of normality assumption.
BLOM Ranks will be calculated with the whole sample, regardless of the treatment group. Ties will be ranked with the mean for tied values.

10.1.5 Sensitivity three-parameter sigmoidal Hill dose-response curve analyses

A three-parameter sigmoidal Hill dose-response curve will be fitted to the data using SAS PROC NLIN. The form of the fitted model will be:

$$y = \frac{akD^{e^n}}{(1-k)b^{e^n} + kD^{e^n}} + \epsilon$$

where:

y is the outcome

D is dose

n is the log shape parameter

ε is the random error

a is the maximum effect

k is a fixed constant where $0 < k \leq 1$

b is the dose that provides an effect equal to k x a

for k=0.5 then b=ED50

for k=0.9 then b=ED90

The parameters a, b and n will be estimated along with their associated SEs and 95% CIs. The dose-response will be displayed in terms of the estimated value for outcome, together with 90% confidence and prediction limits.

A variation on the dose response model will also be evaluated that corrects for placebo:

$$y = P + \frac{akD^{e^n}}{(1-k)b^{e^n} + kD^{e^n}} + \epsilon$$

where:

P is the outcome in placebo treated subjects.

10.1.6 Primary Analysis with new missing data allocation windows

The primary analysis will be repeated but considering now the windows below to reallocate missing data:

- Visit 2: 7 to 21 days
- Visit 3: 35 to 49 days

10.2 Secondary Efficacy Analysis

All secondary endpoints evaluated through a mean absolute change will follow a similar statistical analysis as for the primary variable (ANCOVA, considering baseline corresponding endpoint evaluation as covariate and treatment and region as fixed factors):

- Mean changes of ESS (up to 2-week treatment period)
- Mean relative change in the Epworth Sleepiness Scale (ESS) from baseline to the end of the 2-week and 6-week treatment periods (percentage of absolute decrease compared to baseline ESS).
- SCOPA-Sleep (up to 2-week and 6-week treatment period)
- PDSS-2 (2-week and 6- week)
- MWT (6-week)
- Polysomnography (6-week)
    - Sleep Efficiency
    - Percentage of TST REM
- UPDRS Part III (2-week and 6-week)
- Motor function assessed by the wrist worn device (6-week)
- HAM-D (2-week, 6-week and follow-up)
- MoCA (6-week), and FSS (6-week).
- Fatigue Severity scale (FSS)
- Polysomnography (6-week): Baseline index will not be considered in these Logistic models.
    - Events Index (Increase in Events Index)
    - Arousal Index (Increase in Arousal Index)

Categorical variables as detailed below will be analysed through fitted logistic regression model adjusted on baseline corresponding endpoint evaluation as a covariate and treatment and region as fixed factors. This model will estimate the adjusted Odds Ratio between treatment arms, its corresponding 95% and P-value.

- ESS clinical response, defined as ESS absolute decrease from baseline of at least 3.0 points, after 2 and 6 weeks of treatment.

- ESS clinical response, defined as ESS $\leq$ 10 after 2 and 6 weeks of treatment.

- ESS clinical response, defined as either ESS $\leq$ 10 after 2 and 6 weeks of treatment or ESS absolute decrease from baseline of at least 3.0 points after 2 and 6 weeks of treatment.

10.3 Exploratory Efficacy Analysis

We aim to characterise the changes in free-living activity by means of wrist-worn actigraphy. GENEActiv accelerometers are worn on the wrist and measure tri-axial acceleration with a sample

frequency of 100Hz. The device also measures temperature and light levels. The data is collected pre- and post- dosing (V3), each over a period of seven days. We aim to measure differential physical activity behaviour, and also the length and frequency of daytime sleep episodes.

The data is to be processed using the calibrated GENEActiv tracker data pre-processing package in R. We extract 12-hour daytime records and convert the accelerometer to gravity-subtracted Signal Vector Magnitudes (SVMgs), which is our measure of instantaneous physical intensity. The time-series is downsampled (100x) to one second intervals by averaging the instantaneous SVMgs across the 100 time points. The averaged SVMgs are grouped into four intensity levels groups: Sedentary (< 4.5), Light (4.5-16.5), Moderate (16.5-42), and Vigourous (> 42). These threshold boundaries may change as thresholds used in existing studies are calibrated for different populations. The summary statistics we aim to extract are bout lengths and frequencies, where a bout is a continuous episode of physical activity at a specific intensity levels band. Using paired t-tests, we aim to demonstrate a reduction in sedentary activity and an increase in light to vigourous activity in the dosed population. Using automatic changepoint detection methods (such binary segmentation methods and its extensions, with the fused lasso total variation regulariser, together with Bayesian approaches such as Dirichlet process hidden Markov models.) for detection of sleep boundary times, we also, independently of activity level measurements, aim to quantify the frequency and lengths of daytime sleep episodes.

10.4 Efficacy analysis Considerations

10.4.1 Pooling of Regions

Regions where the low number of subjects is making the statistical model questionable (<20 patients) will be merged with the closer region according to the regional location. If some of the investigative sites are able to enrol a sufficient number of subjects (≥20) they will not be merged with the others in the region and will be considered as an independent fixed effect region category in the model. Any required pooling will also be applied to the per-protocol population to allow a valid comparison of the study results.

10.4.2 Handling of missing data

- For primary analysis, as describes in section 9.1, post-baseline ESS will be imputed through a Multiple Imputation Jumping to Reference Model.

11 SAFETY ENDPOINTS

The safety analysis will be based on the reported adverse events and other safety information, such as clinical laboratory data and physical examination as described in the protocol.

The safety of Bavisant compared to placebo will be assessed by the evaluation of the following:

- Incidence of Adverse Events (AEs), Serious Adverse Events (SAEs), Serious and related Adverse Events and Adverse Events of Special Interest (AESIs) headache, nausea and insomnia
- Incidence of suicidal ideation (C-SSRS) findings from screening/baseline to the end of the 2-week and the 6-week treatment period and safety follow-up
- Incidence of positive psychotic symptoms (BPRS+) findings from screening/baseline to the end of the 2-week and the 6-week treatment period and safety follow-up
- Incidence of physical examination, vital signs and laboratory tests changes from normal to abnormal (haematology and biochemistry)
- Incidence of cardiovascular safety findings (blood pressure, heart rate, ECG including QT/QTc)

11.1 Analysis Sets Analysed for Safety

Safety endpoints will be evaluated in the SAF population as defined in section 6.0, and will be presented by treatment group.

12 STATISTICAL METHODOLOGY FOR SAFETY ENDPOINTS

The baseline value is defined as the last available value before the first dose of double-blind IMP taken.

The analysis of the safety variables will be essentially descriptive and no systematic testing is planned.

The observation period of safety data will be divided into 2 sub-periods:

- The pre-treatment period is defined as the time between the date of signed informed consent and the first dose of double-blind IMP.
- The on-Treatment period (TEAE period): is defined as the Date/ time from first dose of double-blind IMP.

12.1 Incidence of Adverse Events (AEs), Serious Adverse Events (SAEs), and Adverse Events of Special Interest (AESIs) such as headache, nausea and insomnia

All adverse events will be analysed and determined to be either pre-treatment or treatment-emergent. If an adverse event date/time of onset (occurrence, worsening, or becoming serious) is incomplete, an imputation algorithm will be used to classify the adverse event as pre-treatment or treatment-

emergent. The algorithm for imputing date/time of onset will be conservative and will classify an adverse event as treatment emergent unless there is definitive information to determine it is pre-treatment.

All adverse events (including serious adverse events and adverse events of special interest) will be coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the version of Medical Dictionary for Regulatory Activities (MedDRA) version 20.1. Lower-level term (LLT) will not be presented in the analysis. Number of patients in the SAF will be used as denominator for the percentages calculation.

12.2 Incidence of suicidal ideation (C-SSRS) findings from screening/baseline to the end of the 2-week and the 6-week treatment period and safety follow-up

Patients with a positive response on the C-SSRS (an answer of 'yes' to any of the 6 questions) will be described through absolute frequency and percentage at visit 2 (day 14) , visit 3 (day 42) and globally (at any time in the follow-up period) in the study. This percentage will be calculated using the number of patients with C-SSRS evaluation done at the corresponding visit as denominator.

12.3 Incidence of positive psychotic symptoms (BPRS+) findings from screening/baseline to the end of the 2-week and the 6-week treatment period and safety follow-up

The BPRS consists of 4 symptom constructs (*(#4, #11, #12, and #15)* which ranges from 1 (not present) to 7 (extremely severe). 0 is entered if the item is not assessed. The total score is calculated as the sum of the scores of the 4 items. Incidence of patients with an increase and incidence of patients with a decrease from baseline in the BPRS+ total score will be described at the corresponding visit. Also the change from baseline will be descriptively analysed as a numerical data (see section 7 for further reference) for all patients, for patients with BPRS+ total score reduction and for patients with BPRS+ total score increase.

12.4 Incidence of physical examination, vital signs

Physical examinations will consist of the following body systems: (1) general, appearance; (2) extremities; (3) skin; (4) head and neck; (5) eyes, ears, nose and throat; (6) lungs and chest; (7) heart/ cardiovascular; (8) neurological; (9) abdomen / gastrointestinal; (10) liver; (11) musculoskeletal; and (12) other. Each system should be assessed as normal or abnormal and in the last case if that is clinically relevant or not. The incidence of patients with at least one change from normal at baseline to abnormal at Visit 3 will be calculated for each of the body systems.

Potentially Clinically Significant Abnormality (PCSA) are detailed below for Vital signs:

- WEIGHT change from baseline >5%
- Number of breaths per minute <12 or >25. Percentages of patients with a normal respiratory rate (RR) at baseline and an abnormally high or low RR at Visit 3 will be summarized by treatment.

Incidences will be calculated using the number of patients in the safety population as denominator.

12.5 Incidence of laboratory tests findings (haematology and biochemistry)

Clinical laboratory values after conversion will be analysed into standard international units and international units will be used in all listings and tables. Shift tables will be produced presenting the intra-patient change between below normal, normal and above normal for each of the laboratory parameters, according to the appropriate laboratory ranges.

Incidences will be calculated using the number of patients in the safety population as denominator.

12.6 Incidence of cardiovascular safety findings (blood pressure, heart rate, ECG QT/QTc)

Incidence of Potentially Clinically Significant Abnormality (PCSA) as detailed below will be described for blood pressure and ECG:

- Blood Pressure
  - SBP: ≤95 mmHg or decrease from baseline ≥20 mmHg or ≥160 mmHg or increase from baseline ≥20 mmHg
  - DBP: ≤45 mmHg or decrease from baseline ≥10 mmHg or ≥110 mmHg or increase from baseline ≥10 mmHg
- Heart rate
  - <50 bpm or decrease from baseline ≥20 bpm or ≥100 or increase ≥20 bpm from baseline.
- ECG
  - QTc (> 450 for males and > 470 for females) for both Bazzet and Friedericia.
  - QT > 500

Incidences will be calculated using the number of patients in the safety population as denominator.

12.7 Incidence of eye exam findings

Incidence of patients with any new finding on ophthalmological examination in the follow up period will be calculated. Detailed listings of all baseline and on-treatment ophthalmological findings will be provided

Incidences will be calculated using the number of patients in the safety population as denominator.

13 OTHER ANALYSES

13.1 Demographic and Baseline Characteristics

Demographic characteristics, age, gender and race will be summarized by treatment group and overall using descriptive statistics. Age will be summarized using the number of available data, mean, standard, deviation (SD), median, minimum and maximum and 95%CI whenever needed for each treatment group. Gender and race will be summarized using the number and percentage of patients in each treatment group.

The baseline value is defined as the last available value prior to the randomization.

All baseline safety and efficacy parameters are presented along with the summary statistics in the efficacy and safety sections (Section 9. and Section 11.).

13.2 Prior or Concomitant Medications

All medications taken within the past 4 weeks before the date of informed consent and until the end of the study are to be reported in the case report form pages.

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD_ Sep2017) and will be presented in the analysis by the Anatomic and Therapeutic Class.

Prior medications are those the patient used within 4 weeks prior to the date of informed consent and from screening to first investigational medicinal product (IMP) intake. Prior medications can be discontinued before first administration or can be ongoing during treatment phase. Those ongoing will be considered as Concomitant medication also.

Concomitant medications are any treatments received by the patient after first intake of IMP. A given medication can be classified both as a prior medication and as a concomitant medication.

13.3 Prohibited Medication

The following primary PD medications will be permitted at any time during the study if taken as stable medication for at least 4 weeks before screening (date of IC signature):

- Dopaminergics
- L-dopa
- Monoamine oxidase inhibitors (MAOI)
- Short-acting antipsychotics if taken in the evening
- Catechol-O-methyl transferase (COMT) inhibitors
- Amantadine if not used as an alerting agent

At a minimum, the following medications will be prohibited at any time during the study:

- Alerting agents, including r-modafinil, modafinil or methylphenidate
- Benzodiazepines
- Histamine active agents
- Hypnotics

- Cholinergics
- Skeletal muscle relaxants
- Clozapine
- Atomoxetine
- Amitriptyline
- Any other daytime medications which affect sleep

13.4 Medical History

All Medical History will be coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the version of Medical Dictionary for Regulatory Activities (MedDRA) version 20.1. Number of patients with at least one medical history and percentage in each HLT will be presented by SOC. Denominator used for percentage calculation will be the number of patients in the SAF population.

13.5 Extent of investigational medicinal product exposure and compliance

Average number of doses taken per day will be calculated. Percentage of patients will be obtained for each of the categories defined below:

- <0.9 average doses/day taken
- ≥ 0.9 and <1.1 average doses/day taken
- ≥1.1 average doses/day taken

Duration of IMP exposure is defined as last dose date – first dose date, regardless of unplanned intermittent discontinuations. Duration of IMP exposure will be summarized descriptively as a quantitative variable (number, mean, SD, median, minimum, and maximum).

In addition, duration of treatment exposure will also be summarized categorically by numbers and percentages for each of the following categories and cumulatively according to these categories: Up to 1 weeks, >1 to 2 weeks; >2 to 3 weeks; >3 to 4weeks; >4 to 6 weeks; >6 weeks

14 CHANGES FROM THE PROTOCOL

There are no changes to the statistical content specified in the protocol.

15 INTERIM ANALYSES

No interim analysis is planned to be done in this study.

16 REFERENCES

(1) Model-based dose finding under model uncertainty using general parametric models. Pinheiro J, Bornkamp B, Glimm E, Bretz F, Stat Med. 2014 May 10;33(10):1646-61. doi:10.1002/sim.6052. Epub 2013 Dec 3.

(2) Combining multiple comparisons and modeling techniques in dose-response studies. Bretz F, Pinheiro JC, Branson M. Biometrics. 2005 Sep;61(3):738-48.

(3) Ondo WG, Fayle R, Atassi F, Jankovic J. Modafinil for daytime somnolence in Parkinson's disease: double blind, placebo controlled parallel trial. Journal of neurology, neurosurgery, and psychiatry. 2005;76(12):1636-9. Epub 2005/11/18.

(4) Qualification Opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. EMA/CHMP/SAWP/757052/2013. 23 January 2014.

(5) Pinheiro JC, Bornkamp B, Bretz F. Design and analysis of dose finding studies combining multiple comparisons and modeling procedures. Journal of Biopharmaceutical Statistics 2006; 16:639–656.

(6) Request for Qualification of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. Janssen Pharmaceuticals and Novartis Pharmaceuticals. 22 April, 2015.

17 GENERAL FORMAT OF TABLES, FIGURES AND SUBJECT DATA

Table, listings and figures shell documents will be sent to the sponsor for separate review and approval.