Official Title: A Phase IIb, Multicenter, Randomized, Double-Blind, Parallel Group, Placebocontrolled Study to Evaluate the Efficacy, Safety and Tolerability of Basmisanil (RO5186582) as Adjunctive Treatment in Patients With Cognitive Impairment Associated With Schizophrenia Treated With Antipsychotics

NCT Number: NCT02953639

Document Date: SAP Version 1: 29-November-2019

STATISTICAL ANALYSIS PLAN FOR STUDY BP39207

STUDY TITLE: A PHASE IIB, MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PARALLEL GROUP, PLACEBOCONTROLLED STUDY TO EVALUATE THE EFFICACY, SAFETY AND TOLERABILITY OF BASMISANIL (RO5186582) AS ADJUNCTIVE TREATMENT IN PATIENTS WITH COGNITIVE IMPAIRMENT ASSOCIATED WITH SCHIZOPHRENIA TREATED WITH ANTIPSYCHOTICS

PROTOCOL NUMBERS:	BP39207
STUDY DRUG:	RO5186582
VERSION NUMBER:	1.0
SPONSOR:	F. Hoffmann-La Roche Ltd
PLAN PREPARED BY:	[redacted]
DATE FINAL:	29 November 2019

Signatures:

Name	Date	Signature
[redacted]	29/NOV/2019	[redacted]
[redacted] Statistician	29/NOV/2019	[redacted]

TABLE OF CONTENTS

1. BACKGROUND

The purpose of this document is to describe the statistical analyses to be performed at the end of the study.

One interim analysis has been performed as planned in the protocol. Based on the results of that interim analysis, randomization to the 80mg dose arm was terminated; randomization as of November 8th, 2018, continued in a 1:1 ratio to placebo and the 240mg dose group only. Patients who had already been randomized to the 80mg dose group prior to November 8th, 2018, continued on that treatment arm as per protocol.

2. STUDY DESIGN

The study design is described in the protocol and not repeated here. The most recent protocol is V5 approved on July 31, 2018.

2.1 OBJECTIVES

2.1.1 Primary Objective

The primary objective of this study is to investigate the efficacy of 24 weeks of basmisanil treatment on cognitive function as measured by the MATRICS consensus cognitive battery (MCCB) neurocognitive composite score, in stable patients with CIAS treated with antipsychotics.

2.1.2 Secondary Objectives

The secondary objectives for this study are to evaluate the effect of 24 weeks of treatment with basmisanil in stable patients with CIAS treated with antipsychotics on the following:

- Individual cognitive domains of the MCCB, namely attention, speed of processing, reasoning, working memory, visual learning, verbal learning and social cognition.
- Additional specific hippocampal and prefrontal-dependent cognitive tasks and processes (as measured by the Trail making test [TMT]-B, Wechsler memory scale - Fourth edition, verbal paired associates [WMS IV-PAL] and Wechsler memory scale - Fourth edition, logical memory test [WMS IV-LM]).
- Functional capacity and performance (as measured by the Personal and Social Performance scale [PSP] and Schizophrenia Cognition Rating Scale [SCoRS]).
- Safety and tolerability of 24 weeks of basmisanil treatment in patients with CIAS treated with antipsychotics.
- The steady-state pharmacokinetics (PK) of basmisanil and its metabolites, if appropriate, in stable patients with CIAS treated with antipsychotics using population PK modelling methods.

2.1.3 Exploratory Objectives

The exploratory objectives for this study are as follows:

- To explore the effect of 24 weeks of treatment with basmisanil on semantic priming (measured in the category fluency test of the MCCB) and primacy/recency effects (measured in the Hopkins verbal learning test of the MCCB), two experimental measures of hippocampal-dependent cognitive processes.
- To determine whether the patients' genetic profile predicts effects of treatment with basmisanil on cognitive functions.
- To evaluate the effect of 24 weeks of treatment with basmisanil on self-reported mood, sleep, subjective well-being and cognitive functioning (smartphone-based assessments).
- To evaluate the effect of 24 weeks of treatment with basmisanil on functional capacity assessed by novel computerized measures and "work readiness".
- To evaluate the effect of 24 weeks of treatment with basmisanil on symptoms of schizophrenia including positive and negative symptoms.
- To evaluate the effect of 24 weeks of treatment with basmisanil on quality of life.

2.2 OUTCOME MEASURES

Analysis details for all outcome measures are provided in Section 3.

2.2.1 Primary Efficacy Outcome Measure

The primary efficacy outcome measure of the study is the absolute change from baseline at Week 24 on the MCCB 'Neurocognitive Overall Composite T Score'.

2.2.2 Secondary Efficacy Outcome Measures

Secondary efficacy outcome measures of the study are as follows:

- Absolute change from baseline at Week 12 in the MCCB 'Neurocognitive Composite T-Score'.

  Absolute change from baseline at Weeks 12 and 24 in the Individual Domain T-Scores of the MCCB, namely the

  1. ATTENTION-VIGILANCE DOMAIN T-SCORE
  2. REASONING-PROBLEM SOLVING DOMAIN T-SCORE
  3. SOCIAL COGNITION DOMAIN T-SCORE
  4. SPEED OF PROCESSING DOMAIN T-SCORE
  5. VERBAL LEARNING DOMAIN T-SCORE
  6. VISUAL LEARNING DOMAIN T SCORE

7 WORKING MEMORY DOMAIN T-SCORE.

- Absolute change from baseline in
  - Ratio between [TMT]-B and [TMT]-A scores of the Trial Making Test (TMT). See Section 3.4.3.2 for further details.
  - Wechsler memory scale - Fourth edition, verbal paired associates [WMS IV-PAL] Available on QS domain as QSSCAT in ('VERBAL PAIRED ASSOCIATES I', 'VERBAL PAIRED ASSOCIATES II') and QSTEST in ('VPA I Total Raw Score', 'VPA II Total Raw Score'). See Section 3.4.3.3 for further details.
  - Wechsler memory scale - Fourth edition, logical memory test [WMS IV-LM]). Available on QS domain as QSSCAT in ('LOGICAL MEMORY I', 'LOGICAL MEMORY II') and QSTEST in ('LM I Total Raw Score', 'LM II Total Raw Score') See Section 3.4.3 3 for further details.
- Absolute change from baseline in the CGI Severity Score (CGI-S).
- Mean CGI Improvement Score (CGI-I)
- Absolute change from baseline at Weeks 12 and 24 in the Personal and Social Performance scale [PSP] total score.
- Absolute change from baseline at Weeks 12 and 24 in the Schizophrenia Cognition Rating Scale [SCoRS] 'Total Score' (available from Interviewer only).

  Absolute change from baseline at Weeks 12 and 24 in the SCoRS item 'Glob Rat-Inv: Ovrall Imprsn Pt Diffclty' (available from Interviewer only).
- Safety and tolerability endpoints are described in more detail in the safety section.
- The characterization of the steady-state pharmacokinetics (PK) of basmisanil will be performed by the Modelling and Simulation team and documented elsewhere.

2.2.3 Patient Reported Outcome Measures

- Absolute change from baseline in the schizophrenia quality of life scale (SQLS).
- Change in smartphone-based Likert scales evaluating mood, sleep, subjective well-being.
- Treatment expectation questions.

2.2.4 Exploratory Outcome Measures

- Absolute change from baseline at Weeks 12 and 24 in the Total PANSS Score as well as the 8 individual subscale or factor scores
  1. Negative Symptoms Factor Score
  2. Negative Subscale Score

3. Positive Symptoms Factor Score
4. Positive Subscale Score
5. Generalized Psychopathology Subscale Score
6. Disorganized Thought/Cognition Factor Score
7. Uncontrolled Hostility/Excitement Factor Score
8. Anxiety/Depression Factor Score.

- Absolute change from baseline at Weeks 12 and 24 in the Total BNSS Score as well as the 6 individual subscale scores
  1. Anhedonia
  2. Distress
  3. Asociality
  4. Avolition
  5. Blunted Affect
  6. Alogia.
- Absolute change from baseline at Weeks 12 and 24 in the Virtual Reality Functional Capacity Assessment Tool (VRFCAT).

  Absolute change from baseline at Weeks 12 and 24 in the University of Miami Computerized Functional Assessment System (CFAS).
- Absolute change from baseline at Weeks 12 and 24 in the Work Readiness Questionnaire (WoRQ).

  In addition, there is an overall assessment '*Is this patient ready for work?*' with possible responses Y (for Yes) and N (for No).
- Absolute change from baseline at Weeks 12 and 24 in semantic priming (measured in the category fluency test of the MCCB).
- Absolute change from baseline at Weeks 12 and 24 in primacy/recency effects (measured in the Hopkins verbal learning test of the MCCB).

These latter two endpoints are considered highly exploratory and analyses of these will therefore be documented separately.

2.3 DETERMINATION OF SAMPLE SIZE

See Section 6.1 of protocol Version 5.

2.4 ANALYSIS TIMING

The primary analysis will be conducted when the double-blind 24-week treatment period ends. The database snapshot enabling the primary analysis will occur once all patients have either completed the 24-week assessment or withdrawn from the study early, and all data required for analysis have been cleaned, verified, and entered on the database.

3. STATISTICAL METHODS

3.1 GENERAL CONSIDERATIONS

Some subtests of the MCCB may be missing and will be imputed for the derivation of T-scores as described in Section 5. Based on data from April 2019, among the Baseline, Week 12, and Week 24 visits, for item specific T-Scores around 1% of the patient visits are based on imputed values and for the 1.8% of the patient visits are based for the primary endpoint, i.e., the MCCB 'Neurocognitive Composite T Score', are based on imputed values. Furthermore, as imputation follows commonly accepted procedures, the imputed values will be included in all analyses.

3.2 ANALYSIS POPULATIONS

Intent to Treat Population (ITTP): The ITT Population consists of all patients who gave informed consent, were randomized and received at least one dose of study medication, whether prematurely withdrawn from the study or not. Data will be summarized according to actual treatment arm patients were randomized to.
Note: Patients without any post-baseline assessments will not contribute a change from baseline and will hence be excluded from any analysis.

Safety Evaluable Population (SEP): The SEP is the same as the ITT population. Data will be summarized according to actual treatment arm patients were randomized to unless patients taking medication different form the one randomized to warrants allocation to another treatment for safety analyses. Such decisions are only possible after unblinding. The SEP will be used for the analysis of all safety endpoints.

Efficacy Analysis Population (EAP): The EAP consists of the ITT Population but excludes patients who have

- a recorded drug of abuse violation at any visit with cognitive testing at either BL2 week 12, or week 24 (patients [redacted] and [redacted])[1];
- documented use of benzodiazepine and other forbidden medications at any visit with cognitive testing[2] (patients [redacted] and [redacted]).

Data will be summarized according to actual treatment arm patients were randomized to The EAP will be the primary population for all analyses of the primary as well as secondary and exploratory efficacy endpoints.

---

1 Drug of abuse violations are recorded on the SU domain Select SUTRT = 'DRUGS OF ABUSE' and SUOCCUR = 'Y' and VISIT in ('Baseline','Week 12 Day 86','Week 24 Day 168) to identify patients with drug of abuse evidence at visits relevant for the interpretation of MCCB

2 Can be obtained from the LB domain, selecting

LBTEST in (Alprazolam , Alpha-Hydroxyalprazolam , Alpha-Hydroxytriazolam , Chlordiazepoxide , Clonazepam , Diazepam , Flurazepam , Midazolam , Norchlordiazepoxide , Nordiazepam , 'Oxazepam','Temazepam' , Triazolam);"

Only records where there is a numeric and positive (greater than 0) result at either the Baseline Week 12, or Week 24 visit

Per Protocol Population (PPP): The PPP is the subset of the EAP who were at least 75% compliant with study drug as measured by AiCure. This will be determined as follows:

- All records on the final AiCure file transferred will be considered if the date of this record is (i) on or after the date of first dose (from the CRF), and (ii) **on or before** the date corresponding to **one day prior** to the date of the week 24 MCCB assessment. Note that the actual date associated with the MCCB assessment and not the date corresponding to 24 weeks after randomization should be used.
- The number of days between the two dates in (i) and (ii) will be derived. For example, if a patients' 1st dose was on 26JUN2016 and the MCCB assessment on 29MAR2017, the period between the two dates in (i) and (ii) corresponds to 5 + 31 + 31 + 30 + 31 + 30 + 31 + 31 + 28 + 28 = 276 days. As doses are taken twice daily, a fully compliant patient is expected to have 2 × 276 = 552 records of drug intake on the AiCure file.
- If indeed only 427 records (regardless of whether these are am or pm records) were found within the period as given by (i) and (ii), compliance would be 427/552 = 77.4% and hence the patient would be eligible for PPP.

The PPP population will be used for the analyses of the MCCB 'Neurocognitive Overall Composite T Score', the SCoRS 'Total Score', and the VRFCAT 'Adjusted Total Time T Score'.

For statistical analysis there will be two subpopulations of the EAP: Subpopulation '**EAP2**' will consist of all patients randomized to either placebo or 240mg, i.e., no patients randomized to 80mg will be included. Subpopulation '**EAP3**' will include all patients randomized on or before November 8th, 2018 and will, in approximately equal proportions, include patients randomized to placebo, 80mg, or 240mg. Similarly, there will be subpopulations '**PPP2**' and '**PPP3**'.

Data will be summarized according to actual treatment arm patients were randomized to.

3.3 DEFINITION OF BASELINE

Baseline for all efficacy and safety analyses is defined as the last non-missing value recorded **prior to or on** the first study of drug administration. MCCB and TMT are assessed at visits 'BL2' and 'BL1' (up to 2 weeks prior to 'BL2'). Also in these cases the last available value will be used, i.e., 'BL2' and 'BL1' will not be averaged.

3.4 EFFIACY ANALYSES

3.4.1 Statistical Models

3.4.2 Primary Efficacy Endpoint

The primary efficacy endpoint of the study is the absolute change from baseline at Week 24 on the MCCB 'Neurocognitive Composite T Score'. The primary analysis of the primary endpoint will be performed on EAP2. Increases from baseline in the MCCB 'Neurocognitive Composite T Score' indicate improvement.

The statistical model

$$\} \cdot \quad \mu ! b \cdot \pi \quad \tau \cdot (\pi\tau)_{ik} \quad \varepsilon_{ij} \cdot \tag{1}$$

will be used for analysis. Therein μ is the general mean b_i are baseline covariates of patient i (i - 1, ..., n), π are the effects of visit j (j 12, 24), τ_k are the effects of treatment k (k = placebo, 240mg), and $(\pi\tau)_{jk}$ denotes the visit by treatment interactions. The random errors ε_{ij} assumed are assumed to be correlated across visits within each subject, i.e., $Var(\varepsilon_{i12}) = \sigma_1^2$, $Var(\varepsilon_{i24}) = \sigma_2^2$, and $Cov(\varepsilon_{i12}, \varepsilon_{i24}) = \sigma_{12}$.

Covariates included in the model are the stratification factors age at randomization (as continuous covariate), sex (two levels: male, female), and the schizophrenia cognitive subtype at screening (three levels: Type 1, Type 2, and Type 3; see Section 3.4.6 for definition and derivation). In addition, the baseline of the endpoint that is being analyzed (MCCB Neurocognitive Composite T Score in case of the primary endpoint) is added to the model as covariate as well. Statistical analyses based on EAP2 will also include a covariate for "Cohort", separating patients randomized on or before November 8th, 2018 versus those randomized after November 8th, 2018.

From the model, 90% confidence intervals comparing the 240mg dose and placebo will be derived for week 12 and week 24. The study will be considered to have met its primary endpoint if the two-sided p-value for the comparison at week 24 is below 0.1 and the result at week 24 is numerically in favor of the active arm.

As a secondary analysis, the primary endpoint will also be evaluated using EAP3, PPP2, and PPP3. The model will be same as model (1), except that the variable τ_k for the effects of treatment will have three levels, placebo, 80mg, and 240mg.

Descriptive Summaries: Based on EAP2 and EAP3, the following summaries tables will be created:

- Summary of absolute values as well as absolute change from baseline by visit (week 12, week 24) and treatment arm. Statistics included should be N, mean, median, minimum and maximum, lower and upper quartile as well as standard deviation.

- Mean graphs of absolute values by visit (week 12, week 24), standard errors of the mean should be displayed as error bars.
- Mean graphs of absolute change from baseline by visit (week 12, week 24), standard errors of the mean should be displayed as error bars. Ensure that in these graphs baseline is displayed with a mean of zero.

No listings are required.

3.4.3 Secondary Efficacy Endpoints

Unless otherwise stated, all secondary endpoint will be analyzed using EAP2. No listings are required for any of the secondary endpoints.

3.4.3.1 MCCB

Larger values in MCCB as well as its sub-scores indicate improved cognition; hence stronger increases from baseline for active treatment versus placebo would indicate a favorable treatment effect.

Results for the absolute change from baseline at Week 12 in the MCCB 'Neurocognitive Composite T-Score' will be derived as part of the primary analysis. Similar results will also be derived for the "Verbal Learning Domain T-Score" and the "Working Memory Domain T-Score" by applying model (1) to these.

Descriptive Summaries: The same types of descriptive summaries as for the primary endpoint will be created for the "Verbal Learning Domain T-Score" and the "Working Memory Domain T-Score'
Anticipated number of outputs: 2 tables, 2 mean graphs for raw values, 2 mean graphs for change from baseline.

As described in Appendix 1. imputations according to commonly accepted principles will be applied to some MCCB scales. Any such imputed values are flagged by the value 'imputed' in the variable COVAL1 on the QS domain.[3]

3.4.3.2 Trail Making Test

Smaller values in the Trail Making Test [TMT]-B over [TMT]-A ratio (see Section 2.2.2, both measured in seconds) indicate improvement; hence stronger decreases from baseline for active treatment versus placebo would indicate a favorable treatment effect.

[3] Based on data from end of April there are 92 instances where a T-Score is affected by an imputation These are the ATTENTION-VIGILANCE DOMAIN T SCORE (11 instances), CPT-IP AGE/GENDER CORRECTED T SCORE (11 instances) MATRICS OVERALL COMPOSITE T SCORE (27 instances), NEUROCOGNITIVE OVERALL COMPOSITE T SCORE (26 instances) REASONING PROBLEM SOLVING DOMAIN T SCORE (14 instances), VISUAL LEARNING DOMAIN T SCORE (3 instances) There are 7 patients for which this either affects Baseline (19/92 instances) or Week 24 (7/92 instances)

The ratio is also calculated if the two endpoints for the same visits are obtained on different days.

Absolute change from baseline at Weeks 12 and 24 for the Trail Making Test endpoint ([TMT]-B over [TMT]-A ratio[4]) will follow the same statistical analysis as described for the primary endpoint.

Descriptive Summaries: Based on EAP2, the same types of descriptive summaries as for the primary endpoint will be created for [TMT]-B/[TMT]-A.
Anticipated number of outputs: 1 table, 1 mean graph for raw values, 1 mean graph for change from baseline.

3.4.3.3 Wechsler Memory Scale

Bigger values in the Wechsler Memory Scale indicate improvement; hence increases from baseline for active treatment versus placebo would indicate a favorable treatment effect. The five scores for analysis are the 'VPA I Total Raw Score', 'VPA II Total Raw Score', 'VPA II Recognition Total Raw Score', 'LM I Total Raw Score', and 'LM II Total Raw Score' (to select records search for these 5 strings in the variable QTSTEST).

Absolute change from baseline at Weeks 12 and 24 in for each score will follow the same statistical analysis as described for the primary endpoint.

Descriptive Summaries: Based on EAP2, the same types of descriptive summaries as for the primary endpoint will be created.
Anticipated number of outputs: 5 tables, 5 mean graphs for raw values, 5 mean graphs for change from baseline.

Logarithmic transformation has been considered prior to analysis. As some results are recorded as '0', such a transformation would lead to a loss of data. Also, logarithmic transformation of the positive values appears not to improve the distributional assumptions.

3.4.3.4 CGI Severity and Improvement Scores

CGI-S: The CGI-S is assessed per protocol in all patients at baseline and weeks 2, 4, 6, 9, 12, 16, 20, and 24. Assessments on Days 2 to 6 were made in 'in-patients' at the beginning of the study only and will not be used for the analysis as results on these days are not available in all patients by design.

---

[4] Trail making test [TMT]-A Available on QS domain as QSSCAT = 'TRAIL MAKING - TIME IN SECONDS and QSCAT = 'SPR3 MCCB'. Trail making test [TMT]-B Available on QS domain as QSSCAT = 'TRAIL MAKING TEST B (TIME IN SECONDS)' and QSCAT = 'SPR2'.

Values for the CGI-S are 'NORMAL, NOT AT ALL ILL', 'BORDERLINE MENTALLY ILL', 'MILDLY ILL', 'MODERATELY ILL', 'MARKEDLY ILL', and 'SEVERELY ILL' which are encoded by the numerical values 1, 2, 3, 4, 5 and 6 respectively. Smaller values in CGI-S indicate improvements; hence stronger decreases from baseline for active treatment versus placebo would indicate a favorable treatment effect.

Based on the blinded data, the distribution of these values across all patients and visits is <1% for value 1, ~12% for value 2, ~47% for value 3, ~40% for value 4, and ~1% for value 5. Despite the categorical nature of the result it is therefore justified to use the same statistical analysis as described for the primary endpoint.

Descriptive Summaries: The same types of descriptive summaries as for the primary endpoint will be created.

Anticipated number of outputs for CGI-S: 1 table, 1 mean graph for raw values, 1 mean graph for change from baseline.

CGI-I: CGI-I is assessed per protocol in all patients at weeks 2, 4, 6, 9, 12, 16, 20, and 24. Assessments on Days 2 to 6 were made in 'in-patients' at the beginning of the study only and will not be used for the analysis as results on these days are not available in all patients by design.

There are no baseline assessments for CGI-I, hence **no change from baseline is derived for CGI-I**.

Values for the CGI-I are 'VERY MUCH IMPROVED', 'MUCH IMPROVED', 'MINIMALLY IMPROVED', 'NO CHANGE', 'MINIMALLY WORSE', 'MUCH WORSE' and 'VERY MUCH WORSE'. On the SDTM datasets, these are assigned the values 1 to 7, respectively. For ease of interpretation, new numerical values according to Table 1 will be assigned to CGI-I.

Table 1: CGI-I Responses and Assigned Numeric Values

CGI-I Result (Variable QSORRES)	Numeric Value to be Assigned for Analysis and Reporting
VERY MUCH IMPROVED	-3
MUCH IMPROVED	-2
MINIMALLY IMPROVED	-1
NO CHANGE	0
MINIMALLY WORSE	1
MUCH WORSE	2
VERY MUCH WORSE	3

This way CGI-I renders itself to the same interpretation as change from baseline in CGI-S. Positive values indicate worsening; smaller increases on drug compared to placebo would indicate a favorable treatment effect. **Error! Reference source not found.** again displays what again a summary table may look like.

Anticipated number of outputs for CGI-I: 1 table, 1 mean graph for raw values (CGI-I reflects a change from baseline already).

3.4.3.5 Personal and Social Performance Scale

The Personal and Social Performance Scale (PSP) Overall Score (QSCAT = 'PSP' and QSTEST =: 'Overall Score' on QS domain) is an integer result in the range of 0 to 100. Increases from baseline indicate improvement on the scale.

Absolute change from baseline at Weeks 12 and 24 in the PSP overall score will follow the same statistical analysis as described for the primary endpoint.

Descriptive Summaries: Based on EAP2, the same types of descriptive summaries as for the primary endpoint will be created.
Anticipated number of outputs: 1 table, 1 mean graph for raw values, 1 mean graph for change from baseline.

3.4.3.6 Schizophrenia Cognition Rating Scale

The Schizophrenia Cognition Rating Scale (SCoRS) consists of

(a) the SCoRS 'Total Score' (Interviewer only). The SCoRS 'Total Score' has non-negative values with smaller values indicating better performance.

The SCoRS 'Total Score' must be recalculated; **the version that is on the SDTM datasets must not be used**. The SCoRS 'Total Score' is recalculated as the mean of all **non-missing** (among the 20) individual interviewer assessed items at each patient visit. This mean will then be multiplied by 20. The calculation should only be performed if at least 16 or more (of the possible 20) of the individual interviewer assessed items are not missing[5]. See [redacted]

(b) the SCoRS item 'Glob Rat-Inv. Ovrall Imprsn Pt Diffclty' (interviewer only) has numerical values ranging from 1 to 10 with smaller values indicating better performance.

(c) the SCoRS item 'Difficulty changed' (interviewer). It is a change already; therefore a change from baseline does not need to be derived for that. Values are 'Much Worse', 'Moderately Worse', 'Minimally Worse', 'No Change', 'Minimally Improved',

[5] This means that for patient visits where all 20 interviewer assessed items are non-missing, the re derived score will match the Interviewer based SCoRS Total Score on the SDTM QS domain However the re-derived result will be different in case some of the items are missing Currently (September 2019) at least 16 interviewer assessed items are available for all patient visits

'Moderately Improved', and 'Much Improved'. These will be assigned the numeric -3, -2, -1, 0, 1, 2, and 3 respectively. As such, higher values indicate improvements.

Formal statistical analysis will only be performed for the absolute change from baseline in the SCoRS 'Total Score' and will follow the same statistical analysis as described for the primary endpoint. The analysis will be done using EAP2, EAP3, PPP2, and PPP3.

Descriptive Summaries: Based on EAP2, the same types of descriptive summaries as for the primary endpoint will be created for (a), (b), and (c). Of note, change from baseline is not required for SCoRS item (c).

For the change from baseline, this will lead to two tables, as a change from baseline should be derived for all endpoints **except** SCoRS item (c), 'Difficulty changed'.

Anticipated number of outputs: 3 tables, 3 mean graph for raw values, 2 mean graph for change from baseline.

3.4.4 Patient Reported Endpoints

3.4.4.1 Schizophrenia Quality of Life Scale

The SQLS consists of 33 items (not listed here, see values of QSTEST for QSCAT in: ('SQLS-R4') on QS domain). For all items results are coded as 'Never', 'Rarely', 'Sometimes', 'Often', 'Always', represented by the numerical values 1, 2, 3, 4, and 5, respectively. Of note, for some items larger values indicate improvement, while for other items larger values indicate worsening, see below for more detailed instructions.

Prior to any analyses, the range from 1 to 5 will be transformed to a range of 0 to 4, (similarly to PANSS).

Items 7 ('Able Carry Out Day To Day Activities'), 12 ('Felt I Could Cope'), 14 ('Slept Well') and 26 ('Felt Happy') ask whether patients engage in positive aspects of life. Therefore these four items must be recoded as follows: 4 → 0, 3 → 1, 1 → 3 and 0 → 4 before any scale is calculated. There in "x → y" means that an original result of x is to be recoded as y.

After the transformations as described above, three scores will be derived:

- **SQLS Total Score (SQLS-T):** Derived as the mean of all 33 items, hence the range is from 0 to 4. The **mean should not be derived** at a visit for a patient if the individual item responses for that patient at that visit are missing more than 8 (i.e., 9 or more) of the individual items. SQLS-T should be derived as

  SQLS-T = (Mean of all **non-missing** items among the 33)*25.

  SQLS-T will have a result between 0 and 100. When deriving the mean ensure that missing values for a patient at that visit are truly excluded and NOT erroneously treated as '0'.

- **SQLS Cognition & Vitality Score (SQLS-CV):** Consists of the 13 items
  - Item 1 (Lacked Energy To Do Things)
  - Item 2 (Couldn't Be Bothered To Do Things)
  - Item 7 (Able Carry Out Day To Day Activities)
  - Item 9 (Hard To Concentrate)
  - Item 12 (Felt I Could Cope)
  - Item 14 (Slept Well)
  - Item 20 (Trouble Remembering Things)
  - Item 23 (Trouble Thinking Clearly)
  - Item 26 (Felt Happy)
  - Item 28 (Felt Drowsy)
  - Item 31 (Felt Tired)
  - Item 32 (Felt Physically Weak)
  - Item 33 (Felt Wasn't Leading Normal Life).

  SQLS-CV is derived as

  SQLS-CV = (**Mean** of all **non-missing** items among the 13)*25.

  The **mean should not be derived** at a visit for a patient if the individual item responses for that patient at that visit are missing for more than 3 (i.e., 4 or more) of the individual items As the range of results for each item is from 0 to 4, SQLS-CV will have a result between 0 and 100. When deriving the mean ensure that missing values for a patient at that visit are truly excluded and NOT erroneously treated as '0'.

  SQLS Psychosocial Score (SQLS-P): Consists of the 20 items that not >=included in the SQLS-CV (not listed here). SQLS-P is derived as

  SQLS-P = (Mean of all **non-missing** items among the 20)*25.

  The **mean should not be derived** at a visit for a patient if the individual item responses for that patient at that visit are missing for more than 5 (i.e., 6 or more) of the individual items. As the range of results for each item is from 0 to 4, SQLS-P will have a result between 0 and 100. When deriving the mean ensure that missing values for a patient at that visit are truly excluded and NOT erroneously treated as '0'.

Descriptive Summaries: Based on EAP2 and EAP3, the same types of descriptive summaries as for the primary endpoint will be created for the three endpoints described. *Anticipated number of outputs:* 2×3 tables, 2×3 mean graphs for raw values, 2×3 mean graphs for change from baseline.

Formal statistical analysis will be performed for the absolute change from baseline of the three endpoints and will follow the same statistical analysis as described for the primary endpoint. The analysis will be done using EAP2 and EAP3.

3.4.4.2 Likert Scales

The records can be found on the QS domain via QSCAT = 'SPR:LIKERT SCALE ASSESSMENT'. The following outcome measures will be considered

- **Mood:** Recorded as response to "*How Are Feeling Today?*" (QSSCAT = 'MOOD') Results are on a scale from 0 to 5, with higher values indicating better mood. See [redacted]
- **Sleep:** Recorded as response to "*How Did You Sleep Last Night?*" (QSSCAT = 'SLEEP'). Results are on a scale from 0 to 5, with higher values indicating better mood. See [redacted]
- **Subjective Cognitive Functioning:** Recorded as response to "*How is Your Concentration/Memory Today?*" (QSSCAT = 'SUBJECTIVE COGNITIVE FUNCTIONING'). Results are on a scale from 0 to 5, with higher values indicating better mood. See [redacted]

For the aforementioned three smartphone-based Likert scales evaluating the data will be aggregated across 3 week intervals, i.e., ≤ Day 1, (Day 1, Day 21] (Day 21, Day 42], (Day 42, Day 63], . . ., (Day 147, Day 175]. If more than one value is in an interval, the last value in the interval will be used. Based on EAP2, a statistical model similar to the one for the primary analysis will be used except that the variable for "visit" will have eight levels. An unspecified covariance structure will be used to model repeated observations in the same patient. Random coefficient models based on the original data (i.e., without assigning to the above windows) will be considered if the data can be considered linear within each patient.

No descriptive summaries will be provided for any of these three endpoints.

Part of the Likert scales are also the questions "Expect Study Medication to Help", "Is Mediation Helping You", and "Think Taking Placebo or Drug". These will be summarized as follows (all can be found on QS domain in variable QSTEST for QSSCAT = 'POSITIVE TREATMENT EXPECTANCY'):

(a) "Expect Study Medication to Help" is only available at baseline; results are on a scale from 0 to 5, with larger values indicating more positive expectations.

(b) "Is Mediation Helping You?" is only available at various intervals after baseline. Results are also on a scale from 0 to 5, with larger values indicating a more favorable assessment. The results will be assigned to time windows as follows:

Start Day	End day	Visit Label
2	10	Week 1 (Day 7)
11	21	Week 2 (Day 14)

22	35	Week 4 (Day 28)
36	52	Week 6 (Day 42)
53	73	Week 9 (Day 63)
74	98	Week 12 (Day 84)
99	126	Week 16 (Day 112)
127	182	Week 20 (Day 140)

In case of multiple assessments in a time window the one closest to day listed in the column "Visit Label" will be used.

(c) "Think Taking Placebo or Drug" is only available at the end of the study; results in one the two values "DRUG" or "PLACEBO".

The responses to (a) will be summarized descriptively to summarize the expectations at baseline. The number and percentage of patients providing answers 0 to 5 will be reported by treatment arm (to assess how comparable treatment arms were at study start) as well as overall.

For the analysis of (b) we derive a change from baseline as the difference in the response at any post-baseline visit in the table above as well as the response to the question "Expect Study Medication to Help" provided at baseline. A separation between treatment arms would be expected for an active drug. The same display will also be provided for patients who responded '4' to '5' at baseline (close to 70% of patients), i.e., patients who had "high" expectations in terms of the drug helping them.

The question "*Think Taking Placebo or Drug*" (collected at week 24) will be tabulated as shown in Table 2: Summary of "Think Taking Placebo or Drug"Table 2 below based on the **ITT population**. All patients in the ITT population should be accounted for in Table 2. In this table, the last post-baseline assessment per patient should be included. That means, study day need to be re-derived based on date of first dose.

Table 2: Summary of "Think Taking Placebo or Drug" (ITT)

Randomized Treatment Arm	Think Taking Placebo or Drug? (Week 24)			
	Drug	Placebo	Missing	Total
Placebo	n10	n11	n12	n1•
80mg	n20	n21	n22	n2•
240mg	n30	n31	n32	n3•
Total	n•0	n•1	n•2	n••

Programming notes (not to be added as footnote to the final table) In each cell, percentages should be added as well Percentages for nx• should be derived using n•• as denominator percentages for n•x should be derived using n•• as denominator Percentages for nij should use ni• as denominator

Part of the Likert Scales is also the assessment "*How did you take your study medication?*" with the possible answers "DIRECTLY INTO THE MOUTH" and "MIXED

WITH SOFT FOOD". The results will be tabulated as shown in Table 3 below based on the **ITT population**. Only patients who have 20 or more responses will be included in this assessment and only records where a response was provided will be considered. For Table 7, only records with QSSTRESC = 'DIRECTLYINTOTHEMOUTH' and date **on or after** September 11, 2018, should be used as that is the date when a patient took the drug directly into the mouth for the first time.

Table 3: Summary of "How did you take your study medication?"

Randomized Treatment Arm	Number and Percentage of Patients Using Available Routes of Intake for Study Medication							
	0%	>0% to ≤20%	>20% to ≤40%	>40% to ≤60%	>60% to ≤80%	>80% to <100%	100%	Total
DIRECTLY INTO THE MOUTH	n11 (%)	n12 (%)	n13 (%)	n14 (%)	n15 (%)	n16 (%)	n17 (%)	N
MIXED WITH SOFT FOOD	n21 (%)	n22 (%)	n23 (%)	n24 (%)	n25 (%)	n26 (%)	n27 (%)	N

Example: n13 is the number of patients who took the drug directly into the mouth on more than 40% but no more than 60% of all recorded drug intakes. Likewise n11 is the number of patients who never took drug directly into the mouth while n17 are those who always took drug directly into the mouth. It is expected that n11 = n27 and n17 = n21

Table 3 summarizes the number of patients who took the stated percentage of administrations by each method. As only records where either of the two types of responses was available will be used.

3.4.5 Exploratory Efficacy Endpoints

3.4.5.1 Positive and Negative Symptoms Score

The Positive and Negative Symptoms Score (PANSS) is a 30-item scale designed to capture the degree of severity for many symptoms in schizophrenia. The symptoms are rated on a 7-point scale capturing absent to extreme psychopathology and the tool demonstrated sensitivity to effects seen with medication. Results for PANSS are 'absent', 'minimal', 'mild', 'moderate', 'moderate/severe', 'severe', and 'extreme' which are encoded as 1 to 7. To ensure comparability with previous studies, for all analyses of PANSS data, the scores will be transformed into 0 to 6 points with 'absent' expressed as 0. Higher values for the PANSS indicate worsening. A favorable drug effects would manifest itself by a smaller change from baseline for active treatment relative to placebo, i.e., a negative placebo corrected response.

The assessment for each of the 30 items (see Appendix 2) will be provided in the eCRF to allow calculation of a total score describing overall symptomatology as well as scores for positive, negative and general psychopathology subscales. The factor scores listed in Table 4 below will be derived and analyzed. **If at least one item score is missing, then the total/factor/subscale scores that include the missing item will be set to**

missing. Although typically the scores are derived as sums, they will be derived as averages for the statistical analysis. This will not impact the statistical analysis but will make displaying the results more efficient as all scores will be on the same scale.

Table 4: Factor Scores for the PANSS

PANSS (Sub)Score	Derivation (See Appendix 2)
Total PANSS Score	Mean of all 30 items
Negative Symptoms Factor Score	(N1 + N2 + N3 + N4 + N6 + G7 + G16)/7
Negative Subscale Score	(N1 + N2 + N3 + N4 + N6 + N7)/6
Positive Symptoms Factor Score	(P1 + P3 + P5 + P6 + N7 + G1 + G9 + G12)/8
Positive Subscale Score	(P1 + P2 + P3 +P4 + P5 + P6 + P7)/7
Generalized Psychopathology Subscale Score	(Sum of G1 to G16)/16
Disorganized Thought/Cognition Factor Score	(P2 + N5 + G5 + G10 + G11 + G13 + G15)/7
Uncontrolled Hostility/Excitement Factor Score	(P4 + P7 + G8 + G14)/4
Anxiety/Depression Factor Score	(G2 + G3 + G4 + G6)/4
Avolition Score	(N1 + N2 + N4 + G16)/4
Expressive Deficit Score	(N1 + N3 + N6 + G7)/4

Note The acronyms Nx, Gx and Px used in the table above refer as given as the last three characters of the variable QSTESTCD on the qs domain where QSCAT = 'PANSS'

Formal statistical analysis will be performed for the nine scores in Table 4 and will follow the same statistical analysis as described for the primary endpoint. The analysis will be done using EAP2 and EAP3.

Descriptive Summaries: Based on EAP2 and EAP3, the same types of descriptive summaries as for the primary endpoint will be created for each of the nine scores shown in Table 4.

Anticipated number of outputs: 2×9 tables, 2×9 mean graphs for raw values, 2×9 mean graphs for change from baseline.

3.4.5.2 Brief Negative Symptom Scale

The Brief Negative Symptom Scale (BNSS) is a 13-item instrument designed for clinical trials and other studies that measures the 6 domains blunted affect, alogia, asociality, anhedonia, avolition, and distress. The domains consist of the following items:

- **Anhedonia.** Three items: 'Intensity of Pleasure during Activities', 'Frequency of Pleasure During Activities', 'Intensity of Expected Pleasure from Future Activities'.
- **Distress.** One item: 'Distress'.
- **Asociality.** Two items: 'Asociality: Behavior', 'Asociality: Internal Experience'.
- **Avolition.** Two items: 'Avolition: Behavior', 'Avolition: Internal Experience'.

- **Blunted Affect.** Three items: 'Facial expression', 'Vocal expression', 'Expressive gestures'.
- **Alogia.** Two items: 'Quantity of Speech', 'Spontaneous Elaboration'.

The individual 13 items are available in the variable QSTEST on the QS domain, however called slightly different there. A **Total BNSS score** is calculated as the average of the 13 individual items. Each of the 13 items in the BNSS is rated on a 7-point (0–6) scale, with anchor points generally ranging from the symptom's being absent (0) to severe (6). Thus, the Total BNSS score has possible total scores ranging from 0 to 6. Subscale scores are calculated as the average of the individual items within each subscale. If at least one item score is missing, then the total/factor/subscale scores that include the item will be set to missing. The distress subscale has only one item, which quantifies the absence of distress, but this subscale is otherwise treated in the same manner as the other subscales [2]. Higher values for the BNSS indicate worsening. A favorable drug effects would manifest itself by a smaller change from baseline for active treatment relative to placebo i.e., a negative placebo corrected response.

Formal statistical analysis will be performed for the six domain scores as well as the Total BNSS score and will follow the same statistical analysis as described for the primary endpoint. The analysis will be done using EAP2.

Descriptive Summaries: Based on EAP2, the same types of descriptive summaries as for the primary endpoint will be created for each of the six domain scores and the Total BNSS score.
Anticipated number of outputs: 7 tables, 7 mean graphs for raw values, 7 mean graphs for change from baseline.

3.4.5.3 Virtual Reality Functional Capacity Assessment Tool

The three items for analysis of the Virtual Reality Functional Capacity Assessment Tool (VRFCAT) are the 'Adjusted Total Time T Score' (most relevant score for VRFCAT), the 'Total Errors T Score', and the 'Total Forced Progressions T Score'. All are on ZA domain under ZACAT = 'VRFCAT'. Results of all scores are recorded on a continuous scale and may include negative numbers. For each variable, a bigger T-score indicates a higher performance and an increase in T-score indicates improvement.

Formal statistical analysis will be performed for all three scores and will follow the same statistical analysis as described for the primary endpoint. The analysis will be done using EAP2, EAP3, PPP2, PPP3.

Descriptive Summaries: Based on EAP2, the same types of descriptive summaries as for all three scores will be created, summary tables and graphs as described for the primary endpoint.

Anticipated number of outputs: 3 tables, 3 mean graphs for raw values, 3 mean graphs for change from baseline.

3.4.5.4 University of Miami Computerized Functional Assessment System

For the University of Miami Computerized Functional Assessment System (CFAS) there are 7 endpoints in total. For 'FORMS SUMMARY' there is one item, 'Total Time to Completion'. For each of 'ATM SUMMARY' and 'PRESCRIPTION SUMMARY' there are the items 'Total Time to Completion', 'Ratio – Total Correct/Total Responses', and 'Rate – Total Correct/Total Time'. Results for all endpoints are non-negative numbers on a continuous scale. A decrease in time to complete, a decrease in rate and an increase in ratio all indicate improvement.

Descriptive Summaries: Based on EAP2, descriptive summaries will only be created for the summary of the total completion times for all forms, i.e., ZATEST = 'Total Time to Completion' and ZASCAT = 'FORMS SUMMARY'. No formal statistical analysis will be performed.
Anticipated number of outputs: 1 table, 1 mean graph for raw values, 1 mean graph for change from baseline.

3.4.5.5 Work Readiness Questionnaire (WoRQ)

The main endpoint of the WoRQ is the overall dichotomous assessment "*Based on your clinical judgment, is this patient ready for work?*" with possible responses Y (for Yes) and N (for No).

Descriptive Summaries: Based on EAP2 for all patients with an assessment at baseline and week 12/24 the following "shift table" will be presented.

Table 5: Shift Table for Work Readiness Questionnaire (EAP2)

	Changes in Patient Status "Is this patient ready for work?"				Total
	YES → YES	YES → NO	NO → YES	NO → NO	
Change from Baseline to Week 12					
Placebo	n11 (%)	n12 (%)	n13 (%)	n14 (%)	N1
240mg	n21 (%)	n22 (%)	n23 (%)	n24 (%)	N2
Change from Baseline to Week 24					
Placebo	n31 (%)	n32 (%)	n33 (%)	n34 (%)	N3
240mg	n41 (%)	n42 (%)	n43 (%)	n44 (%)	N4

Note: The above is only an illustration of what we need. The totals per row should add up to the total number of patients with an assessment at baseline as well as week 12/24.

No formal statistical analyses will be performed.

Anticipated number of outputs: 1 table, see Table 5 above.

3.4.6 Subgroup Analyses

For the MCCB 'Neurocognitive Composite T-Score' as well as the SCORS 'Total Score' mean values and absolute changes from baseline will be provided for the following two subgroups:

a) age at baseline ('≤ 35 years' vs > '35 years'),

b) schizophrenia cognitive subtype (Type 1, 2, or 3; see Figure 1 of protocol),

The statistical analysis for the subgroups will follow the model as described in Section 3.4.2. A categorical factor describing the subgrouping variable will be added to the model together with its interaction with treatment. From this model, treatment effect estimates versus placebo and 90% confidence intervals will be derived. Only stratification variables different from the subgrouping variable will be retained in the model.

Based on Figure 1 of the protocol, the schizophrenia cognitive subtype will be derived in the following steps. First, the difference WRAT4 (Word Reading Standard Score) minus FSIQ is calculated. If this difference is

- ≥ 10, the patient is of Subtype 1 ('deteriorated');
- < 10 and WRAT4 ≤ 90, the patient is of Subtype 2 ('compromised');
- < 10 and WRAT4 > 90, the patient is of Subtype 3 ('preserved').

Patients with Subtype 1 ('deteriorated') are expected to benefit most from treatment with basmisanil. Likewise, younger patients are expected to benefit more from treatment with basmisanil compared to older patients.

Subgroup analyses will be performed for the primary endpoint, i.e., the absolute change from baseline in the MCCB 'Neurocognitive Composite T Score' as well as the SCoRS 'Total Score', and the VRFCAT 'Adjusted Total Time T Score'.

The following two subgroups will only be evaluated for the absolute change from baseline in the MCCB 'Neurocognitive Composite T Score':

- **Median Split on Baseline Cognitive Performance**. Baseline is defined in Section 3.3 and the median will be derived based on the EAP population. This leads to a value of 32.0 (N = 208), hence the subgroups being 'BL ≤ 32.0' and 'BL > 32.0'. As patients were not selected based on a pre-existing cognitive deficit, this subgroup analysis is to test for any ceiling effects at baseline with the underlying hypothesis that low performers at baseline (i.e., the '≤ BL 32' subgroup) have more room for improvement and may be more responsive to treatment than high performers.

- **Median Split on Baseline Learning Capacity.** Baseline learning capacity (BLC) is defined as the change from Screening to Baseline 2 (i.e., the difference 'Baseline 2

Value' minus 'Screening Value') in the MCCB 'Neurocognitive Composite T Score'. This leads to a value of 4.0 (N = 200), hence the subgroups being 'BLC ≤ 4' and 'BLC > 4'. The analysis compares treatment effects in non-learners ('BLC ≤ 4) and learners ('BLC > 4). This subpopulation analysis tests the hypothesis that a preserved learning capacity in the absence of drug treatment may predict capacity to respond to a pro-cognitive treatment, i.e. that learners may be more responsive to basmisanil.

The following subgroup will only be evaluated for the absolute change from baseline in the Total BNSS score (see Section 3.4.5.2):

- **Median Split on Baseline Total BNSS score.** Baseline is defined in Section 3.3 and evaluated based on the EAP population. This leads to a value of 2.0 (N = 213), hence the subgroups being 'BL ≤ 2 and 'BL > 2'. Note that this derived from the average of the 13 items, see Section 3.4.5.2. As more diseased patients have higher values for BNSS, larger treatment effects are expected in the 'BL > 2' subgroup as there is more room for improvement.

3.5 GENETIC ANALYSES

These analyses will be documented separately and also reported separately from the CSR.

3.6 SAFETY ANALYSES

All safety related analyses will be performed on the Safety Evaluable Population (SEP)

The following **listings** of safety data will be provided.

- Listing of Adverse Events (l_ae_SE),
- Listing of Serious Adverse Events (l_ae_ser_SE),
- Summary of All Adverse Events (t_ae_SE),
- Summary of All Serious Adverse Events (t_ae_SER_SE),
- Summary of All Severe or Worse Adverse Events (t_ae_SEV_SE).
- Listing of Previous/Concomitant Medications (l_cm_SE),
- Listing of Columbia Suicide Severity Rating Scale (C-SSRS) (l_cssrs_SE),
- Listing of Deaths (l_dd_AP),
- Listing of Patients who Discontinued from Study (l_ds_trt_SE),
- Listing of QTcF Interval (l_eg_qtcf_SE; including absolute change from baseline),
- Listing of Extrapyramidal Symptom Rating Scale (ESRS) (l_esrs_SE),
- Listing of Laboratory Abnormalities (l_lb_abn_SE),
- Listing of Marked Laboratory Abnormalities (l_lb_mabn_SE),
- Listing of Vital Signs (l_vs_SE).

The following **summary tables** of safety data will be provided.

- Summary of Adverse Events by Greatest Intensity (t_ae_int_SE),
- Summary of Serious Adverse Events by Greatest Intensity (t_ae_ser_int_SE),
- Summary of Previous/Concomitant Medications (t_cm_SE),
- Summary of Demographic and Baseline Characteristics (t_dm_SE),
- Summary of ECG Parameters Change from Baseline by Visit (t_eg_cb_SE),
- Summary of QTcF Interval (Categorical Absolute Values) (t_eg_qtcf_SE),
- Summary of Laboratory Tests and Change from Baseline by Visit (t_lb_cb_SE),
- Summary of Marked Laboratory Abnormalities (t_lb_mabn_SE),
- Summary of Out of Range Laboratory Assessments (t_lb_oor_SE),
- Previous and Concurrent Medical History (t_mh_SE),
- Summary of Abnormal Vital Sign Results (t_vs_abn_SE),
- Vital Signs Change from Baseline by Visit (t_vs_cb_SE).

Graphical displays of safety data will be provided. All graphs should show the three visits 'Baseline', 'Week 12' and 'Week 24'. Means by visit should be shown with error bars indicating the standard errors of the means. Graphs for the change from baseline should start at 0 for 'Baseline'. The graphical displays (raw values and absolute change form baseline) will be provided for the following parameters:

- **Vital Signs:** Systolic blood pressure, diastolic blood pressure, heart rate.
- **ECGs:** QTcF, QRS interval, Heart Rate.

4. INTERIM ANALYSIS

In accordance with the protocol, one interim analysis has been performed when approximately 30 patients per arm have completed the first 12 weeks of treatment. The results of this interim analysis were reviewed by an Internal Monitoring Committee (IMC). Following the recommendation of the IMC, randomization to the 80mg dose arm was terminated; randomization as of November 8th, 2018, continued in a 1:1 ratio to placebo and the 240mg dose group only. Patients who had already been randomized to the 80mg dose group prior to November 8th, 2018, continued on that treatment arm as per protocol.

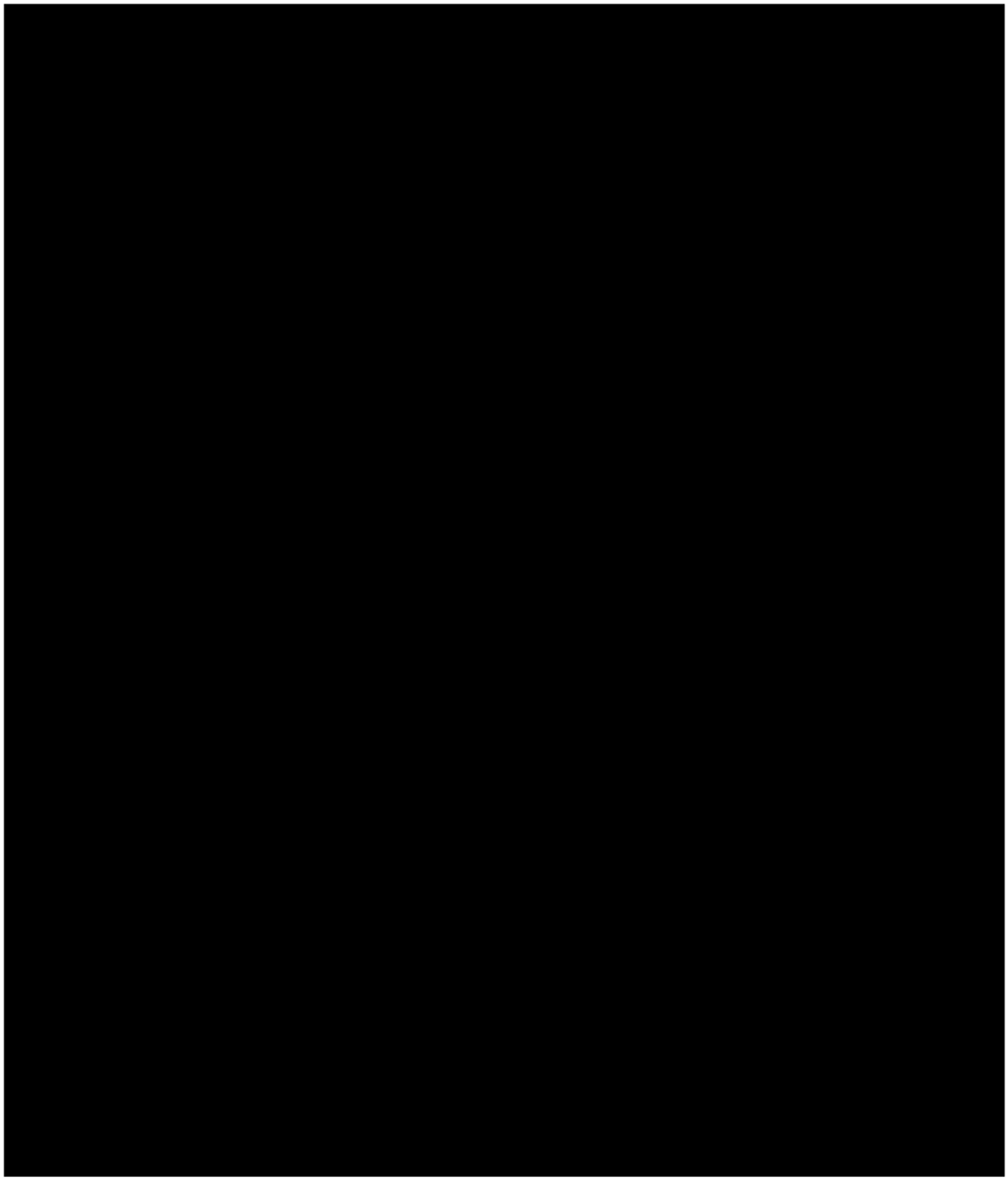
No other interim analyses for efficacy were conducted.

The Roche IMC had also conducted three interim reviews of safety data on

- 2nd February 2017 (after 10 patients had completed at least one week of treat),
- 7th March 2017 (after n=25 patients had completed at least one week of treatment)
- 20th April 2017 (after n=40 patients had completed at least one week of treatment).

After each meeting the IMC recommended the study to proceed.

5. APPENDIX 1: IMPUTATION FOR MISSING SUBTESTS IN THE MCCB



[redacted] explaining how imputed values will be created.

MCCB has three levels: These are the

(i) The overall MCCB composite t-score (OCT) and the neurocognitive composite T score (NCT).

(ii) The 7 domain scores (DS) that are "in some way" used to derive OCT, six thereof are used to derive NCT.

(iii) Each DS again is made up of the results of 1-2 individual "tests". Values can be missing at the "test" level (c3). As I understood, these are never imputed.

If there were a missing value for a test, and no imputation were done, the corresponding DS the test feeds into would eventually be missing - and again also the resulting OCT or NCT relying on this DS is missing.

Imputation replaces the missing value for the DS by a number. Therefore, in the file it is ok to mark these with "IMPUTED" in the comment column of the file. Likewise, the imputed DS is now used to derive the OCT/NCT, and hence also that OCT/NCT value needs to be marked "IMPUTED". We do not need the original result in case of imputation. First, such an original result would not exist anyway (imputation is otherwise not needed). Second, even in case where there were a result which is deemed invalid, we'd discard it anyway.

6. APPENDIX 2: DETAILS ON SELECTED QUESTIONNAIRES

PANSS Negative Symptoms, Disorganized Thoughts and Positive Symptoms Factors

Factor	Original PANSS item number	PANSS item name
Negative symptoms	N1	Blunted affect
	N2	Emotional withdrawal
	N3	Poor rapport
	N4	Passive apathetic social withdrawal
	N6	Lack of spontaneity and flow of conversation
	G7	Motor retardation
	G16	Active social avoidance
Disorganized thought/cognition	P2	Conceptual disorganization
	N5	Difficulty in abstract thinking
	G5	Mannerisms and posturing
	G10	Disorientation
	G11	Poor attention
	G13	Disturbance of volition
	G15	Preoccupation
Positive symptoms	P1	Delusions
	P3	Hallucinatory behavior
	P5	Grandiosity
	P6	Suspiciousness
	N7	Stereotyped thinking
	G1	Somatic concern
	G9	Unusual thought content
	G12	Lack of judgment and insight

PANSS factors based on "Marder" PANSS factor analysis published in J Clin Psycl 58:12, December 1997 p 538.