STATISTICAL ANALYSIS PLAN

Title	**SYN**120 a Dual 5-HT_6/5-HT_{2A} **A**ntagonist **P**roof of Concept Study to Evaluate its **S**afety, Tolerability and **E**fficacy in Parkinson's Disease Dementia (SYNAPSE)
Protocol Number	SYN120-CL03
Phase	2a
IND Number	77,435
Statistical Analysis Plan Version	1.0
Date of Statistical Analysis Plan	15 Nov 2017
Date of Current Protocol Version	Draft 15 April 2016
Sponsor	Acorda Therapeutics Inc.
Sponsor Study Director	[redacted]
Sponsor Chief Medical Officer	[redacted]

STATISTICAL ANALYSIS PLAN APPROVAL SIGNATURES

TABLE OF CONTENTS

1. INTRODUCTION

This Statistical Analysis Plan (SAP) defines the analysis samples, specifies the analysis, and guides the interpretation of results from the SYN120-CL03 or SYNAPSE clinical trial. The SAP supplements the clinical protocol. Readers are encouraged to refer to the clinical protocol for details on the rationale for the intervention, eligibility criteria, conduct of the trial, clinical assessments and the timing of their use in the trial, definitions and reporting of adverse events, data management conventions, and regulatory oversight and compliance procedures. The reader is referred to the Investigator's Brochure for additional details on physiochemistry and pre-clinical and clinical pharmacokinetics and safety of the investigational product, SYN120.

In case of discrepancies between the SAP and the clinical protocol concerning matters of analysis, the SAP is authoritative. On all other matters, the clinical protocol is authoritative.

2. STUDY DESIGN

2.1. Overview

This is a multicenter, double-blind, placebo-controlled, two-arm, parallel-group, phase 2a randomized controlled trial to evaluate the safety, tolerability, and preliminary efficacy of a 16-week course of 100 mg QD SYN120 in men or women with a diagnosis of Parkinson's disease dementia (PDD) who are taking a cholinesterase inhibitor. Trial participation includes a screening visit, randomization to SYN120 or placebo, 16 weeks of follow-up on study drug, and a final safety visit 2 weeks after last dose of study drug. The trial is registered at Clinicaltrials.gov as study NCT02258152 (see https://clinicaltrials.gov/ct2/show/NCT02258152). Participants may also separately consent to DNA testing for a pharmacogenetics substudy. The aims and planned analyses of the pharmacogenetics substudy are not described here.

2.2. Study Objectives

The primary efficacy objective of the SYN120-CL03 trial is to assess the effect of a fixed dosage of SYN120 on cognition as determined by the Cognitive Drug Research Computerized Cognition Battery (CDR) Continuity of Attention measure in patients with PDD treated with a stable dose of a cholinesterase inhibitor. The key secondary efficacy objective is to assess the effects of SYN120 in this population on CDR Quality of Episodic Memory. Other secondary and exploratory efficacy objectives include the effects of SYN120 on: (a) global impressions of change, (b) other measures of cognition, (c) activities of daily living, (d) nighttime sleep and daytime sleepiness, (e) positive symptoms, and (f) behavioral functioning.

The safety objectives of the SYN120-CL03 trial are to assess the safety and tolerability of a fixed dosage of SYN120 by assessing adverse events (AEs), vital signs (including orthostatic blood pressure measurements), laboratory assessments, electrocardiography (ECG), Unified Parkinson’s Disease Rating Scale (UPDRS) Parts I through IV, changes in intensity of symptomatic therapy, and Columbia Suicide Severity Rating Scale (C-SSRS).

The pharmacokinetic objectives of the SYN120-CL03 trial are to estimate population-level pharmacokinetic parameters for plasma levels of SYN120.

2.3. Study Population

Individuals eligible for trial participation are men or women at least 50 years old who have received a diagnosis of probable PDD according to the Movement Disorder Society Task Force clinical diagnostic criteria for dementia associated with Parkinson's disease (Emre et al. 2007) and who are taking a cholinesterase inhibitor. Detailed inclusion and exclusion criteria are specified in the clinical protocol.

Participants will be recruited from a total of 20 clinical sites located through the US.

2.4. Participant Flow

After providing informed consent and determining eligibility, approximately 80 participants will be randomized to receive SYN120 100 mg QD or placebo in a 1:1 ratio, stratified by clinical site. In the active arm, SYN120 is titrated as follows: 20 mg QD for the first 7 days, 50 mg QD for the next 7 days, and 100 mg QD for the remaining 14 weeks. Participants in the placebo arm receive a matching pseudo-titration of placebo. Participants who consent to participate in the pharmacogenetics substudy have a blood sample collected for DNA testing. Participants and caregivers return to the study site for evaluations at weeks 4, 8, and 16 and will be telephoned at weeks 2 and 12 to assess for AEs. Participants who discontinue study drug should remain on study following the normal schedule of assessments. Patients withdrawing consent will undergo an Early Termination Visit, ideally while still taking study drug. All participants are asked to return for a Safety Follow-up Visit approximately 2 weeks after their last dose administration. Detailed descriptions of study procedures and timing are specified in the clinical protocol.

2.5. Treatment Allocation

At the baseline visit, eligible participants are randomly allocated in equal proportion to one of two treatment groups, SYN120 100 mg QD or placebo, according to a permuted-block randomization schedule with a block size of 4, stratified by site. The randomization schedule was prepared by computer program by the unblinded study statistician. Stratification by site is achieved by distributing to sites allotments of study drug kits that consist of a full block from the permuted-block schedule.

2.6. Allocation Concealment

The randomization schedule is known only by the unblinded study statistician who generated the schedule and by the study drug distributor and the sponsor's quality control and bioanalytical staff and subcontractors as specified in the SYN120-CL03 Study Randomization Specification. Concealment of the true treatment allocation of specific participants is achieved by use of anonymous subject identifiers to link participants with specific study drug kits and by use of matching active and placebo tablets and titration schedules. Members of the SYNAPSE Steering Committee, site investigators and other site staff, clinical coordination and data management staff, the blinded study statistician, the medical monitor, and all participants are blinded to participant treatment allocations. The Safety Data Monitor (SDM) and members of the SDM Advisory Group are provided treatment-specific information in order to monitor the trial but

such information is masked by use of coded values to identify the treatment groups. The SDM may request the true treatment identities.

2.7. Schedule of Assessments

Study Period	Screening[a]	Baseline	16-Week Treatment Period (Double-blinded Dosing) (Days 1–112)					Safety Follow Up	Early Term.	Unscheduled
Study Week	−6 to −1	0	2	4	8	12	16	18		
Visit Window		±3 days	±3 days	±3 days	±3 days	±3 days	±3 days	±3 days		
Assessments	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**V8**	**V99**	**V98**
Onsite clinic visit	X	X		X	X		X	X	X	X
Obtain patient and caregiver informed consent	X									
Demographics; medical history including neurological and PD/PDD history	X									
Concomitant medications including anti-dementia and anti-Parkinson meds	X[b]	X		X	X		X	X	X	X
BP and pulse (supine and standing)[c]	XXX	X		X	X		X	X	X	X[d]
Weight (include height at Screening)	X						X		X	
Physical and neurological examination	X	X[d]		X	X		X		X	X[d]
MoCA	X						X		X	
Preliminary eligibility assessment by Investigator	X									
CDR Computerized Cognition Battery[e]	XX	X			X		X		X	
ADAS-cog[e]		X					X		X	
ADCS-CGIC[e]							X		X	
PDAQ (reported by caregiver)		X					X		X	
SCOPA-SLEEP		X					X		X	
SAPS-PD		X					X		X	
NPI (reported by caregiver)		X					X		X	
UPDRS Parts I–IV[f]		X		X	X		X	X	X	
Study drug accountability		X		X	X		X		X	X
Recording of AEs	X	X	X	X	X	X	X	X	X	X
C-SSRS	X	X		X	X		X	X	X	X[d]
Dispense IMP to patient/caregiver and provide dosing instructions		X		X	X					X[d]
FSH (♀ who are not surgically sterile and are postmenopausal, only)	X									
Urine hCG (♀ of childbearing potential, only)[g]	X	X		X	X		X		X	X[d]
Hematology[h], chemistry (including liver function)[i]	X	X		X	X		X		X	X[d]
TSH, freeT3 and free T4	X									
Vitamin B12	X									
Urinalysis[j]	X	X		X	X		X		X	X[d]
SYN120 blood sampling[k]		X		X[l]	X[l]		X		X[l]	X[d, l]
12-lead electrocardiogram[m]	X	X[n]		X	X		X		X	X[d]
Pharmacogenetic sample collection (optional)		X[o]		X[p]	X[p]		X[p]			

♀ = female; AEs, adverse events; ADAS-cog, Alzheimer's Disease Assessment Scale–cognitive subscale; ADCS-CGIC, Alzheimer's Disease Cooperative Study-Clinician's Global Impression of Change; BP, blood pressure; CDR, Cognitive Drug Research; C-SSRS, Columbia-Suicide Severity Risk Scale; ECG, electrocardiogram; FSH, follicle stimulating hormone; hCG, human chorionic gonadotropin; IMP, investigational medicinal product; MoCA, Montreal Cognitive Assessment; NPI, Neuropsychiatric Inventory; PD, Parkinson's disease; PDAQ-15, Penn Parkinson's Daily Activities Questionnaire-15; PDD, Parkinson's disease dementia; SAPS-PD, PD-adapted Scale for Assessment of Positive Symptoms; SCOPA-SLEEP, Scales for Outcomes in Parkinson's Disease-Sleep Scale; TSH, thyroid stimulating hormone; UPDRS, Unified Parkinson's Disease Rating Scale.

Footnotes:

[a] Screening period may not exceed 6 weeks. Note: For screen failures, document demographics and reason for ineligibility in source documents and eCRF.

[b] At Screening, obtain complete medication history including anti-dementia and anti-Parkinson medications (current and those received within the past year). Record the date and time of the most recent dose of each anti-dementia and anti-Parkinson medication taken prior to the Screening assessment.

[c] BP and pulse are to be measured after at least 5 minutes supine rest and again after standing for 1 and 3 minutes.

[d] Optional assessments that may be performed for evaluation of AEs, at the Investigator's discretion.

[e] Perform all cognitive and dementia-related assessments while patient is in ON state. Complete two training sessions during the screening period, as close to the Baseline visit as possible.

[f] UPDRS is to be measured in ON approximately 1 to 3 hours after patients have taken a scheduled dose of levodopa (preferably their morning dose of levodopa). Patients will be instructed to have already taken their normally scheduled dose of levodopa before arriving at the study site in order to have their UPDRS Part III evaluated in the ON state. UPDRS in the ON state will be measured at a time representative of the ON state in that patient, not in the patient's "best" ON.

[g] For female patients of childbearing potential, perform a urine hCG pregnancy test and document method of contraception at Screening and verify continuation of contraceptive method at each visit.

[h] Hematology: hemoglobin, hematocrit, red blood cell count, total and differential white blood cell, and platelet count.

[i] Chemistry: aspartate aminotransferase (AST/SGOT), alanine aminotransferase (ALT/SGPT), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), lactate dehydrogenase (LDH), total bilirubin (conjugated and unconjugated), albumin, creatinine, urea/BUN, bicarbonate, uric acid, total protein, sodium, chloride, potassium, calcium, phosphate, glucose, cholesterol, and creatine phosphokinase (CK).

[j] Urinalysis: specific gravity, pH, ketones, blood, protein, and glucose. If urine dipstick positive for leukocytes, protein or erythrocytes, microscopic evaluation will be performed.

[k] Blood sample for determination of SYN120 plasma concentration (and screening for presence of SYN120 metabolites) to be obtained at Baseline (Day 1) prior to initiation of dosing, Visit 4 (Week 4), Visit 5 (Week 8), Visit 7 (Week 16), Early Termination (if applicable), and Unscheduled Visit (optional, as per footnote d). Record date and time of SYN120 sample collection.

[l] For blood sample collections at Visit 4 (Week 4), Visit 5 (Week 8), Visit 7 (Week 16), and Early Termination or Unscheduled Visit (if applicable), record the date and time when the patient took the last dose of IMP.

[m] Resting supine 12-lead ECGs will be collected after the patient has been in a supine, or if unable, semi-supine (no more than 45 degrees) position for a minimum of 5 minutes. ECGs should be collected at a time during study visit when the patient is not experiencing dyskinesia that would interfere with an adequate recording.

[n] At Baseline, obtain triplicate 12-lead ECGs (3 serial readings performed several minutes apart).

[o] At Baseline, for patients who give consent for optional pharmacogenetic substudy, collect single blood sample (approximately 10 mL) for pharmacogenetic assessment. (If this sample is not collected at Baseline, it will be collected at the next possible in-clinic visit, i.e., Visit 4, 5 or 7.) See Section 7.13 of the clinical protocol.

[p] For patients who give consent for optional pharmacogenetic substudy but sample not collected at Baseline (or during previous visit), collect single blood sample (approximately 10 mL) for pharmacogenetic assessment. See Section 7.13 of the clinical protocol.

3. STATISTICAL METHODOLOGY

3.1. General Considerations

3.1.1. Statistical Software

All statistical analyses will be performed using SAS (SAS Institute, NC, USA) or R (R Foundation for Statistical Computing, Vienna, Austria).

3.1.2. Summary Statistics

Data will be summarized with respect to disposition, demographics, pre-treatment characteristics, safety outcomes, tolerability, and efficacy outcomes. Summary statistics for

continuous variables will include the number of subjects, the mean, median, standard deviation, and range. For categorical data, summaries will include counts and percentages.

3.1.3. Precision

Results will generally be reported to 3 significant figures. Percentages will generally be reported to 0.1 percentage points. P-values will be reported to two digits when greater than or equal to 0.10, to three digits when greater than or equal to 0.00095 and less than 0.10, and as <0.001 for all smaller values.

3.2. Study Endpoints

3.2.1. Efficacy Endpoints

The primary efficacy endpoint is the change from baseline to the week 16 visit in the CDR Continuity of Attention composite score captured in the ON state.

The key secondary efficacy endpoint is the change from baseline to the week 16 visit in the CDR Quality of Episodic Memory composite score captured in the ON state.

The other secondary efficacy endpoints are the change from baseline to the week 16 visit in the following measures:

- Alzheimer's Disease Cooperative Study-Clinician's Global Impression of Change (ADCS-CGIC),
- Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog) total score captured in the ON state by an evaluator blinded to the results of all cognitive assessments,
- CDR Power of Attention composite score captured in the ON state,
- CDR Speed of Memory Retrieval composite score captured in the ON state,
- PDAQ-15 total score (reported by caregiver),
- SCOPA-SLEEP nighttime sleep and daytime sleepiness scores,
- SAPS-PD total score,
- NPI total score, caregiver distress total score, and subscale total scores for delusions, hallucinations, agitation/aggression, dysphoria, anxiety, euphoria, apathy, disinhibition, irritability/lability, aberrant motor activity, night-time behavioral disturbances, and appetite and eating abnormalities (reported by caregiver), and
- MoCA total score.

Exploratory efficacy endpoints include the change from baseline to the week 8 visit in the CDR Continuity of Attention and Quality of Episodic Memory composite scores captured in the ON state and the change from baseline to week 8 and 16 visits in the CDR Quality of Working Memory, Cognitive Reaction Time, and Reaction Time Variability composite scores captured in the ON state, individual component tasks of the CDR, i.e., Immediate Word Recall (Words recalled, Errors), Simple Reaction Time (Median Reaction Time), Digit Vigilance (Mean Reaction Time, Accuracy, False Alarms), Choice Reaction Time (Median Reaction Time, Accuracy), Numeric Working Memory (Median Reaction Time), Spatial Working Memory

(Median Reaction Time, Sensitivity Index), Delayed Word Recall (Words recalled, Errors), Word Recognition (Median Reaction Time, Original Stimuli Accuracy, Novel Stimuli Accuracy), and Picture Recognition (Median Reaction Time, Original Stimuli Accuracy, Novel Stimuli Accuracy), and scores for individual tasks of the ADAS-cog.

3.2.2. Safety Endpoints

The following safety endpoints will be evaluated:

- Proportion of participants experiencing the following and event rates relative to time at risk of: (a) any treatment-emergent AE (TEAE), (b) any severe TEAE, (c) any serious TEAE, (d) any TEAE at least possibly related to study drug, (e) any TEAE leading to permanent study drug discontinuation or early termination from the trial, and (f) any TEAE resulting in death, summarized across all MedDRA terms
- Proportion of participants experiencing and number of unique events of all TEAEs classified by seriousness, severity, relatedness to study drug, action taken with study drug, and outcome, summarized across all MedDRA terms
- Proportion of participants experiencing and number of unique events of each type of TEAE and serious TEAE classified by MedDRA system organ class, high level term, and preferred term
- Proportion of participants experiencing treatment-emergent orthostatic hypotension
- Mean 16-week change from baseline in supine and standing SBP, supine and standing DBP, and differences from supine to standing SBP and DBP
- Mean 16-week change from baseline in weight
- Proportion of participants experiencing treatment-emergent clinically significant laboratory abnormalities classified by assay, whether the abnormality is below or above the normal range, and visit, with cross-classification of frequencies of abnormalities from baseline to each scheduled post-baseline visit at which safety labs are performed
- Mean 16-week change from baseline in laboratory assays
- Proportion of participants experiencing treatment-emergent clinically significant ECG abnormalities
- Mean 16-week change from baseline in ECG parameters
- Mean 16-week change from baseline in UPDRS Part I score, II score, III score, I through III total score, Part IV dyskinesia score, Part IV clinical fluctuations score, Part IV other complications score, and Part IV total score
- Mean 16-week change from baseline in levodopa daily dosage (LDD) and levodopa equivalent daily dosage (LEDD)
- Proportion of participants reporting any treatment-emergent suicidal ideation
- Maximum post-baseline suicidal ideation planning, frequency, duration, controllability, deterrents, and reasons

- Proportion of participants reporting any post-baseline suicidal behavior, actual attempts, non-suicidal self-injurious behavior, interrupted attempts, aborted attempts, preparatory acts or behaviors, or completed suicide
- Maximum post-baseline suicidal behavior lethality and potential lethality

Reported proportions will use as their denominator all participants in the Safety and Tolerability sample (see Section 3.5 below).

3.2.3. Tolerability Endpoint

Participants will be judged tolerant of study drug if they complete their week 16 visit without prior permanent discontinuation of study drug and without prior study drug suspension lasting longer than 14 days. Tolerability will be summarized as the proportion of participants in a treatment group who are tolerant of study drug.

3.3. Measurement Definitions

3.3.1. CDR Measures

The CDR (Wesnes et al. 1992) is a computerized neuropsychological test battery to assess cognitive tasks based on measures of choice reaction time, vigilance, and the sensitivity and speed of digit, word and picture recognition. The stimuli are presented on a computer screen and the subjects respond by pressing either a “Yes” or “No” on a response box. Detailed of the assessments are described in the clinical protocol.

Continuity of Attention (COA) measures the ability to sustain attention and avoid error. COA is calculated as (VIGACC*0.45) + (CRTACC*0.5) - VIGFA, where VIGACC is digit vigilance accuracy, CRTACC is choice reaction time accuracy, and VIGFA is digit vigilance false alarms. Higher COA scores represent greater sustained attention and avoidance of errors.

Quality of Episodic Memory (QEM) measures the ability to store, hold, and retrieve information of an episodic nature. QEM is calculated as (DRECOACC + DRECNACC – 100) + (DPICOACC + DPICNACC – 100) +((IRCL – IRCLERR)*100 / 12) + ((DRCL – DRCLERR)*100 / 12), where DRECOACC is word recognition original stimuli accuracy, DRECNACC is word recognition new stimuli accuracy, DPICOACC is picture recognition original stimuli accuracy, DPICNACC is picture recognition new stimuli accuracy, IRCL is immediate word recall words recalled, IRCLERR is immediate word recall errors, DRCL is delayed word recall words recalled, and DRCLERR is delayed word recall errors. Higher QEM scores represent greater ability to store, hold, and retrieve information of an episodic nature.

Quality of Working Memory (QWM) measures the ability to store and retrieve information in working memory. QWM is calculated as SPMSI + NWMSI, where SPMSI is the sensitivity index for spatial working memory and NWMSI is the sensitivity index for numeric working memory. Sensitivity index (Frey and Colliver 1973) measures the ability to identify previously presented stimuli and reject stimuli which were not previously presented. A score of 1 represents perfect discrimination; all of the previously presented stimuli are correctly identified and all distractor stimuli are correctly rejected as being novel. A score of -1 represents perfect reverse discrimination; none of the previously presented stimuli are correctly identified, and all distractor stimuli are selected as being novel. A score of 0 represents chance performance with no

discrimination between previously presented stimuli and distractor stimuli. Sensitivity index is calculated as (HIT - FA) / (2 (HIT + FA) - (HIT + FA)^2), where HIT is the proportion of previously presented items that were correctly identified and FA is the proportion of novel items that were incorrectly identified. If both HIT and FA are 0 or both HIT and FA are 1, leading to an indeterminant ratio as calculated above, then the sensitivity index is assigned a value of zero. Higher QWM scores represent greater ability to store and retrieve information in working memory.

Power of Attention (POA) measures the intensity of concentration at a particular moment. POA is calculated as SRTM + VIGRT + CRTM, where SRTM is median simple reaction time, VIGRT is mean digit vigilance reaction time, and CRTM is median choice reaction time. Higher POA scores represent less intensity of concentration at a particular moment.

Speed of Memory Retrieval (SMR) measures the time it takes to retrieve a memory. SMR is calculated as SPMRTM + NWMRTM + DRECRTM + DPICRTM, where SPMRTM is median spatial working memory reaction time, NWMRTM is median numeric working memory reaction time, DRECRTM is median word recognition reaction time, and DPICRTM is median picture recognition reaction time. Higher SMR scores represent slower speed in retrieving memories.

Immediate Word Recall is measured by the CDR variables IRCL (number of words correctly recalled) and IRCLERR (number of words recalled that were not presented during current testing session).

Simple Reaction Time is measured by the CDR variable SRTM (median speed of individual correct responses).

Digit Vigilance is measured by the CDR variables VIGRT (mean speed of individual responses to targets within 1.5 sec window), VIGACC (percentage of targets responded to within 1.5 sec window), and VIGFA (number of responses falling outside of 1.5 sec window).

Choice Reaction Time is measured by the CDR variables CRTM (median speed of individual correct responses) and CRTACC (percentage of stimuli correctly identified).

Numeric Working Memory is measured by the CDR variables NWMRTM (median speed of individual correct responses to all stimuli) and NWMSI (Frey and Colliver sensitivity index).

Spatial Working Memory is measured by the CDR variables SPMRTM (median speed of individual correct responses to all stimuli) and SPMSI (Frey and Colliver sensitivity index).

Delayed Word Recall is measured by the CDR variables DRCL (number of words correctly recalled) and DRCLERR (number of words recalled that were not presented during current testing session).

Word Recognition is measured by the CDR variables DRECRTM (median speed of individual correct responses to all stimuli), DRECOACC (percentage of original stimuli correctly identified), and DRECNACC (percentage of novel stimuli correctly identified).

Picture Recognition is measured by the CDR variables DPICRTM (median speed of individual correct responses to all stimuli), DPICOACC (percentage of original stimuli correctly identified), and DPICNACC (percentage of novel stimuli correctly identified).

3.3.2. CDR Datasets

Two datasets will be obtained from Bracket: (1) raw data, and (2) adjudicated data. The adjudicated dataset, with modifications suggested by Dr. Wesnes that in his expert opinion would best reflect the cognitive status of participants at the time of testing, will be used for primary analysis. All changes from the raw data resulting from algorithmic flipping of data and from Dr. Wesnes' adjudication will be documented together with their justification.

3.3.3. ADCS-CGIC

The ADCS-CGIC (Schneider et al. 1997) is a single-question clinician-completed scale for assessing clinically relevant global change in patients with dementia based on an organized but unstructured assessment of the participants's function and mental status, a caregiver interview, and a standardized set of questions. The ADCS CGIC evaluation will be performed by an Investigator who is blinded to other cognitive evaluations. The ADCS-CGIC score takes values from 1 to 7, with higher scores representing greater worsening, centered on a score of 4 for no change from baseline.

3.3.4. ADAS-cog

The ADAS-cog is the 11-item cognitive subscale of the Alzheimer's Disease Assessment Scale (ADAS; Rosen et al. Am J Psychiatry 1984) designed to measure the severity of the most prominent symptoms of dementia in patients with Alzheimer's disease (AD). The ADAS-cog subscale measures alterations of memory, language, praxis, attention and other cognitive abilities which comprise the core symptoms of AD and PDD through a series of clinician-administered tests. This study utilizes an early revision of the ADAS-cog (unpublished manual Mohs, 1994; see: http://adni.loni.usc.edu/data-samples/data-faq/) and does not include the supplemental delayed word recall or number cancellation tasks (cf. Petersen et al. 2005) nor the maze task (cf. Mohs et al. 1997). Item-level scores from the 11 tasks are summed for a total score (range 0 to 70) with higher scores representing worse cognitive function.

3.3.5. PDAQ-15

The PDAQ-15 (Brennan et al. 2016) is a 15-item scale completed by a knowledgeable informant and designed to measure cognitive instrumental activities of daily living. Each item is rated from 0 (Cannot Do) to 4 (None [=no difficulty doing]). All items are summed for a total score (range 0 to 60) with higher scores indicating greater ability to perform instrumental activities of daily living.

3.3.6. SAPS-PD

The SAPS-PD (Voss et al. 2013) is a 9-item clinician-rated scale consisting of a subset of items (H1 Auditory Hallucinations, H3 Voices Conversing, H4 Somatic or Tactile Hallucinations, H6 Visual Hallucinations, H7 Global Rating of Severity of Hallucinations, D1 Persecutory Delusions, D2 Delusions of Jealousy, D7 Ideas and Delusions of Reference, D13 Global Rating of Severity of Delusions) selected on the basis of their relevance to PDD from the full 20-item SAPS (Andreasen, 1984) and designed to measure positive symptoms in patients with schizophrenia. Each item is rated from 0 (absent) to 5 (severe). All items are summed for a total score (range 0 to 45) with higher scores indicating more severe positive symptoms.

3.3.7. SCOPA-SLEEP

The SCOPA-SLEEP (Marinus et al. 2003) is a 13-item self-reported scale designed to evaluate nighttime sleep and daytime sleepiness among patients with PD. The SCOPA-SLEEP consists of four sections: (a) a single question (range 0 to 3) addressing the frequency of using sleeping tablets; (b) 5 questions addressing specific elements of nighttime sleep quality; (c) one question evaluating overall nighttime sleep quality over the previous month, and (d) 6 questions addressing daytime sleepiness. The 5 nighttime sleep questions rate the perceived extent of problems with sleep fragmentation, sleep efficiency, sleep duration, and early wakening, each rated from 0 (not at all) to 3 (a lot). They are summed for a total score (range 0 to 15) with higher scores indicating greater problems with nighttime sleep. The overall sleep quality question is rated on a 7-point scale (range 0 to 6) with higher scores indicating worse sleep quality. The 6 daytime sleepiness questions rate the frequency of experiences indicative of sleepiness, each rates from 0 (never) to 3 (often). They are summed for a total score (range 0 to 18) with higher scores indicating greater daytime sleepiness.

3.3.8. NPI

The NPI (Cummings et al. 1994, Cummings 1997, Kaufer et al. 1998) is a caregiver-completed scale designed to evaluate 12 subdomains of behavioral functioning: delusions, hallucinations, agitation, dysphoria, anxiety, apathy, irritability, euphoria, disinhibition, aberrant motor behavior, night-time behavior disturbances, and appetite and eating abnormalities. Each domain has a screening question that assesses the presence of the specific disturbance. If present, a series of 7 to 9 more detailed Yes/No subquestions are asked, the frequency is rated on a scale from 1 (occasionally, less than once per week) to 4 (very frequently, once or more per day or continuously), the severity is rated on a scale from 1 (mild) to 3 (marked), and caregiver emotional distress is rated on a scale from 0 (not at all [distressed]) to 5 (very severely or extremely [distressed]). The total score for each domain (range 0 to 12) is calculated by multiplying the frequency by the severity or given a value of zero if presence of the disturbance for that domain was not endorsed. A total NPI score is calculated by adding all domain-specific total scores together (range 0 to 144) with higher scores indicating worse neuropsychiatric disturbance. A total caregiver distress score (NPI-D, range 0 to 60) is calculated by adding all domain-specific caregiver distress scores.

3.3.9. MoCA

The MoCA (Nasreddine et al. 2005) consists of 8 clinician-administered cognitive tasks designed to screen for mild cognitive impairment. The MoCA assesses attention and concentration, executive functions, memory, language, visuoconstructional skills, conceptual thinking, calculations, and orientation. One point is awarded for correct completion of each item of the visuospatial/executive function task (5 items), naming task (3 items), digit vigilance and tapping items of the attention task (3 items), the sentence repetition items of the language task (2 items), abstraction task (2 items), delayed recall task (5 items), and orientation task (6 items). One point is awarded for naming 11 or more words during the fluency item of the language task. Zero (none correct) to 3 (4 or more correct) points are awarded based on the number of correct subtractions by 7 starting at 100 in the attention task. One point is awarded if the participant has 12 years or less of education unless the score is already 30. Scores for each task are summed for a total score (range 0 to 30) with higher scores indicating greater cognitive capacity.

3.3.10. UPDRS

The UPDRS (Fahn et al. 1987) is a 4-part, clinician-completed scale: Part I, Mentation, Behavior and Mood; Part II, Activities of Daily Living; Part III, Motor Examination; Part IV, Complications of Therapy. Each item in Parts I (4 items), II (13 items), and III (27 items) is rated on a scale from 0 (normal, none, or condition absent; text varies by item) to 4 (the most severe state of a given condition; text varies by item). The items are summed separately by Part and across all three parts for four total scores (range Part I, 0 to 16; Part II, 0 to 52; Part III, 0 to 108; Parts I-III, 0 to 176) with higher scores indicating more severe impairments. Part IV contains 11 items divided across three subsections: A, Dyskinesias (4 items); B, Clinical Fluctuations (4 items); and C, Other Complications (3 items). Part IV A consists of 3 items rated on a scale from 0 (no experience of a given condition; text varies by item) to 4 (the most severe state of a given condition; text varies by item) and 1 item rated on a scale from 0 (no; condition not present) to 1 (yes; condition present). Part IV B consists of 1 item rated on a scale from 0 (none) to 4 (76 to 100% of the day) and 3 items rated on a scale from 0 (no; condition not present) to 1 (yes; condition present). Part IV C consists of 3 items rated on a scale from 0 (no; condition not present) to 1 (yes; condition present). The items are summed separately by subsection and across all Part IV items for four total scores (range Part IV A, 0 to 13; Part IV B, 0 to 7; Part IV C, 0 to 3; Part IV, 0 to 23) with higher scores indicating more severe complications.

3.3.11. Orthostasis

Orthostatic hypotension will be defined as a maximum decrease of 20 mm Hg or greater in systolic blood pressure (SBP) or a maximum decrease of 10 mm Hg or greater in diastolic blood pressure (DBP) when moving from a supine to a standing position, with measurements completed 1 min and 3 min after standing.

3.3.12. Levodopa Equivalent Daily Dosage

The levodopa equivalent daily dosage (LEDD) will be calculated using data from the concomitant medications log. The conversion from dopamergic drugs other than carbidopa-levodopa will follow the recommendations by Tomlinson et al. (2010) with the following additions for more recently approved drugs.

- Extended release formations of carbidopa-levodopa (Rytary™) will be converted at 60% of their actual levodopa content (which is 5 mg greater than their labeled content) based on mean post-baseline daily dosages reported in Hauser et al. (2013);
- Extended release formations of ropinorole (RequipXL™) will be converted at 20 mg levodopa equivalent per mg ropinirole based on equivalent maximum recommended daily dosage.

The full calculation is as follows:

1. At each time any oral levodopa is taken per day, calculate the immediate release levodopa (e.g., Sinemet, Parcopa) dosage (mg) based on the number of tablets taken per dose and the dosage of levodopa in each tablet
2. At each time any oral levodopa is taken per day, calculate the controlled release levodopa (e.g., Sinemet CR) dosage (mg) and multiply by 0.75 to account for loss of bioavailability

3. At each time any oral levodopa is taken per day, calculate the extended release levodopa (e.g., Rytary) dosage (mg), noting that the listed dosage per tablet is 5 mg less than the actual dosage, and multiply by 0.60 to account for loss of bioavailability
4. At each time any oral levodopa is taken per day, calculate the total daily oral levodopa dosage by summing together the oral levodopa dosage from immediate release, controlled release, and extended release formulations (quantities 1 through 3)
5. At each time any oral levodopa is taken per day, if entacapone but not tolcapone is taken at the same time, either separately or as a combination drug (e.g., Stalevo), then multiply the total levodopa dosage at that time (quantity 4) by 0.33
6. At each time any oral levodopa is taken per day, if tolcapone (e.g., Tasmar) is taken at the same time, then multiply the total levodopa dosage at that time (quantity 4) by 0.50
7. Calculated the COMT-adjusted total daily levodopa equivalent dosage associated with oral levodopa as the sum over all time points of quantities 4 through 6
8. If enteral levodopa is taken without any COMT inhibitor, then multiply the total enteral levodopa (e.g., Duopa) daily dosage (mg) by 1.11
9. If enteral levodopa is taken with entacapone but not tolcapone, then multiply the total daily enteral levodopa dosage (mg) by 1.48
10. If enteral levodopa is taken with tolcapone, then multiply the total daily enteral levodopa dosage (mg) by 1.67
11. If immediate release or modified release pramipexole (e.g., Mirapex) is taken, multiply the daily dosage (mg/day of the dihydrochloride monohydrate salt) by 100
12. If immediate release or extended release ropinirole (e.g., Requip, Requip XL) is taken, multiply the dosage (mg/day) by 20
13. If a rotigotine patch (e.g., Neupro) is used, multiply the dosage (mg/day) by 30
14. If oral selegiline (e.g., Eldepryl) is taken, multiply the dosage (mg/day) by 10
15. If sublingual selegiline (e.g., Zelapar) is taken, multiply the dosage (mg/day) by 80
16. If oral rasagiline (e.g., Azilect) is taken, multiply the dosage (mg/day) by 100
17. If oral amantadine (e.g., Symmetrel) is taken, multiply the dosage (mg/day) by 1
18. If an injection or infusion of apomorphine (e.g., Apokyn) is taken, multiply the dosage (mg/day) by 10
19. Calculate total LEDD as the sum of quantities 7 through 18.

3.4. Determination of Sample Size

The primary efficacy measure is the change from baseline to the week 16 visit in the CDR COA. Based on a two-tailed test at alpha = 0.05 and assuming up to 15% loss to follow up, a total sample size of 80 randomized participants provides at least 80% power for a true treatment-dependent difference in 16-week change in CDR COA equal to an effect size of 0.69. Additional power will be obtained from analysis using a shared-baseline repeated-measures analysis of

variance. Effect sizes of similar magnitude were observed for several CDR measures in a 24-week trial of memantine in patients with PDD and dementia with Lewy bodies (Wesnes et al. 2014).

3.5. Analysis Samples

The following analysis samples will be used for testing safety, tolerability, efficacy, and pharmacokinetic endpoints:

- Safety and Tolerability (ST) Sample: Participants who are eligible, randomized, and take at least one dose of study drug, classified according to the actual treatment received.
- Efficacy Modified Intent-to-treat (mITT) Sample: Participants who are eligible, randomized, have at least one valid pre-treatment and at least one valid post-treatment CDR COA measurement, and take at least one dose of study drug, classified according to their randomized treatment assignment. This sample will include observations made after discontinuation of study drug should any such participants remain on study.
- Efficacy Per-protocol (PP) Sample: Participants who are eligible, randomized, have at least one valid pre-treatment and one valid post-treatment CDR COA measurement, take at least one dose of study drug, and meet all other major protocol compliance criteria specified prior to unblinding, classified according to the actual treatment received. Observations made after permanent discontinuation of study drug (whether due to intolerance, early termination, loss to follow-up, or death) will be omitted, but completion of follow-up on study drug will not be required. Any other individual observations affected by important protocol deviations may also be excluded, as specified prior to unblinding.
- Pharmacokinetic (PK) Sample: Participants in the ST sample who have at least one valid determination of their plasma SYN120 concentration.

3.6. Baseline Characterization

Each analysis sample will be summarized overall and by treatment group for the following characteristics: randomization site; age, gender, race, and ethnicity; education level, years of education; years since PD diagnosis, years since PDD diagnosis, UPDRS Part I, II, III, I-III, IV A, IV B, and IV C total scores, and Hoehn and Yahr Stage; CDR COA, QEM, QWM, POA, and SWM composite scores; ADAS-cog total score; MoCA total score; PDAQ-15 total score; SAPS-PD total score; SCOPA-SLEEP nighttime sleep, overall sleep quality, and daytime sleepiness; NPI domain-specific and overall score and overall caregiver distress score; weight and body mass index; orthostasis and orthostatic changes in systolic and diastolic blood pressure; QTc interval by Fridericia's equation; baseline specific cholinesterase inhibitor; baseline memantine use, and baseline LDD and LEDD. The magnitude of differences between treatment groups will be summarized using Fisher's exact tests for nominal variables, exact Cochran-Armitage trend tests for ordinal variables, and two-sample t-tests for approximately continuous variables.

3.7. Interim Analysis

No interim analysis for futility, efficacy, or sample size re-estimation is planned. The independent Safety Data Monitor and the Safety Data Monitor Advisory Group review safety data quarterly and may request specific analyses at any time.

3.8. Efficacy Analysis

3.8.1. Primary Analysis of the Primary Efficacy Endpoint

The primary efficacy analysis will be performed on the mITT sample using the adjudicated CDR data and will use a shared-baseline, repeated-measures analysis of variance that includes fixed effects for visit (3 levels: Baseline, Week 8, and Week 16) and the interaction between treatment group (2 levels: SYN120 and placebo) and post-baseline visit (2 levels: Weeks 8 and 16), random center-specific intercepts, and unstructured within-person covariance. The interaction between treatment group and visit will be restricted to post-baseline visits by including a numeric indicator variable (0 pre-treatment, 1 post-treatment) in the interaction. Use of a shared baseline reflects the true state of the population sampled prior to randomization and has the advantage of adjusting for any chance differences at baseline in a manner similar to analysis of covariance (Liang and Zeger, 2000) with potentially greater efficiency with more than one follow-up visit and drop-out. Data from Early Termination visits will be assigned to the next scheduled post-baseline visit. The following SAS code specifies the model:

```
proc mixed data=xxx method=reml;
  class site id trtrnd visit;
  model Value = visit post*trtrnd*visit;
  random intercept / subject=site type=vc;
  repeated visit / subject=id(site) type=un;
```

Treatment-dependent differences in the change from baseline to the week 16 visit on the CDR COA will be estimated by a linear contrast and tested using a two-tailed Wald-test at alpha = 0.05. The following SAS code specifies the linear contrasts used to estimate treatment- and visit-specific means, treatment- and visit-specific changes from baseline, and treatment-dependent visit-specific changes from baseline (with the sort order for treatment group being SYN120 first and placebo second):

```
  estimate "0|Active |Wk  00" intercept 1 visit  1 0 0 post*trtrnd*visit 0 0 0  0 0 0 / cl;
  estimate "0|Active |Wk  08" intercept 1 visit  0 1 0 post*trtrnd*visit 0 1 0  0 0 0 / cl;
  estimate "0|Active |Wk  16" intercept 1 visit  0 0 1 post*trtrnd*visit 0 0 1  0 0 0 / cl;
  estimate "1|Active |dWk 08" intercept 0 visit -1 1 0 post*trtrnd*visit 0 1 0  0 0 0 / cl;
  estimate "1|Active |dWk 16" intercept 0 visit -1 0 1 post*trtrnd*visit 0 0 1  0 0 0 / cl;
  estimate "0|Placebo|Wk  00" intercept 1 visit  1 0 0 post*trtrnd*visit 0 0 0  0 0 0 / cl;
  estimate "0|Placebo|Wk  08" intercept 1 visit  0 1 0 post*trtrnd*visit 0 0 0  0 1 0 / cl;
  estimate "0|Placebo|Wk  16" intercept 1 visit  0 0 1 post*trtrnd*visit 0 0 0  0 0 1 / cl;
  estimate "1|Placebo|dWk 08" intercept 0 visit -1 1 0 post*trtrnd*visit 0 0 0  0 1 0 / cl;
  estimate "1|Placebo|dWk 16" intercept 0 visit -1 0 1 post*trtrnd*visit 0 0 0  0 0 1 / cl;
  estimate "2|ActvPlb|dWk 08" intercept 0 visit  0 0 0 post*trtrnd*visit 0 1 0  0 -1 0 / cl;
  estimate "2|ActvPlb|dWk 16" intercept 0 visit  0 0 0 post*trtrnd*visit 0 0 1  0 0 -1 / cl;
```

The estimate labeled "2|ActvPlb|dWk 16" specifies the primary hypothesis. A significant mean improvement over 16 weeks in CDR COA among participants randomized to SYN120 relative to mean changes among participants randomized to placebo would be considered evidence of therapeutic benefit from SYN120.

3.8.2. Secondary Analyses of the Primary Efficacy Endpoint

Several secondary analyses will be investigated to assess sensitivity of our estimates of treatment effect to alternative modeling assumptions and alternative handling of the CDR data.

- Random-slopes model: The repeated-measures covariance will be replaced with random participant-level intercepts and slopes with unstructured covariance. Data from Early Termination visits will be placed at their observed follow-up time. The following SAS code

specifies the model and the linear contrasts used to estimate treatment-specific and treatment-dependent changes over 16 weeks:

```
proc mixed data=xxx method=reml;
  class site id trtrnd;
  model Value = week trtrnd*week;
  random intercept / subject=site type=vc;
  random intercept week / subject=id(site) type=un;
  estimate "1|Active |Chg 16" intercept 0 week 16 trtrnd*week 16   0 / cl;
  estimate "1|Placebo|Chg 16" intercept 0 week 16 trtrnd*week  0  16 / cl;
  estimate "2|ActvPbl|Chg 16" intercept 0 week  0 trtrnd*week 16 -16 / cl;
```

- Shared-baseline, repeated-measures ANOVA model with pMI for missing data: The shared-baseline repeated-measures ANOVA model described above for the primary analysis will be applied to datasets generated by placebo multiple imputation (Ayele et al. 2014). Fifty imputed datasets will be generated for a given endpoint.
- Cochran-Mantel-Haenszel row mean score LOCF: 16-week change scores will be compared by mean score test using modified ridit scores stratified by clinical site. The last observed CDR COA value will be carried forward when CDR COA values are not available from the Week 16 visit. The following SAS code specifies the model:

```
proc freq data=xxx;
  table site * trtrnd* Value / cmh scores=modridit;
```

- Cochran-Mantel-Haenszel row mean score WO: 16-week change scores will be compared by mean score test using modified ridit scores stratified by clinical site. The worst observed CDR COA 16-week change from baseline will be imputed when CDR COA values are not available from the Week 16 visit.
- PP sample: The primary efficacy model and the secondary random-slopes model will be applied to the PP sample.
- Alternative CDR data: The primary efficacy model and the secondary random-slopes model will be applied to mITT and PP samples using the raw CDR data.

3.8.3. Secondary Efficacy Endpoints

The key secondary efficacy endpoint will all be analyzed using the same analysis samples and models as specified for the primary efficacy endpoint in Sections 3.8.1 and 3.8.2. Non-key secondary efficacy endpoints and exploratory efficacy endpoints will all be analyzed using the mITT sample and the primary efficacy analysis described in Section 3.8.1. Variables that are strictly positive and strongly right-skewed (e.g., CDR POA, CDR SWM, and other CDR measures of reaction time) will be analyzed on their original scale and after log-transformation. Estimates from analysis of log-transformed data will be back-transformed for reporting.

3.8.4. Subgroup Analyses

Differences in treatment efficacy will be explored in several pre-defined subgroups: site, age (median split), screening MoCA total score (median split), baseline LEDD (median split), and baseline cholinesterase inhibitor (donepezil vs. rivastigmine). Subgroup specific estimates will be obtained by including subgroup, subgroup × visit, and subgroup × treatment × post-baseline visit terms to models for primary and key secondary efficacy outcomes based on the mITT sample.

3.8.5. Multiplicity Adjustments

A single primary efficacy outcome is specified with a single treatment comparison and a pre-specified primary analysis with no interim analysis. Therefore, no multiplicity exists for the primary analysis, and interference will be based on two-tailed testing at alpha = 0.05. If the primary analysis is significant, then testing the key secondary outcome at alpha = 0.05 maintains an overall type I error rate at 5% under a closed testing sequential analysis. When the primary analysis is not significant, we limit the overall type I error rate to 10% or less (depending on the magnitude of correlation between the primary and key secondary outcomes) when significance of the key secondary outcome alone is accepted as evidence of a therapeutic benefit from SYN120. Results from analysis of non-key secondary efficacy endpoints, exploratory efficacy endpoints, and subgroup analyses will report nominal, comparison-wise p-values, recognizing that the totality of results will be evaluated in judging the potential of SYN120 as a therapeutic agent for PDD.

3.8.6. Missing Data

Baseline values for efficacy endpoints will be determined from the last non-missing data collected prior to the first dose of study medication. The planned repeated-measures model yields estimates that are unbiased conditional on the observed scores under a missing at random assumption. In addition, a secondary analysis of the primary endpoint will use placebo-based multiple imputation of missing data. The Cochran-Mantel-Haenszel row mean score LOCF and WO analyses provide estimates of treatment effect when normality assumptions are violated. Additional sensitivity analyses may be pursued to impute missing values or otherwise construct models for unobserved outcomes if more than 20% of participants are missing follow-up data for any reason.

3.8.7. Clinical Interpretation

Unadjusted summary statistics by treatment group and visit and estimates from the shared-baseline, mixed model analyses for all CDR measures will be provided to Dr. Keith Wesnes by the study statistician. Dr Wesnes will provide a clinical interpretation of the effects of SYN120 based on treatment-dependent effects on the CDR measures. This clinical interpretation will supplement the numerical and statistical results.

3.9. Safety Analysis

3.9.1. Treatment-emergent Adverse Events

The incidence of TEAEs will be summarized by the number of events of a given classification experienced by participants in each treatment group and by the number and proportion of participants experiencing such an event in each treatment group in the ST sample. TEAEs will be summarized in aggregate across all MedDRA terms and separately by MedDRA term.

Aggregate summaries of TEAE incidence will include: (a) any TEAE, (b) any severe TEAE, (c) any serious TEAE, (d) any TEAE at least possibly related to study drug, (e) any TEAE leading to permanent study drug discontinuation or early termination from the trial, and (f) any TEAE resulting in death. Aggregate summaries will be presented by treatments for event rates and proportion of participants.

Aggregate summaries of TEAE grade will include characteristics of: (a) seriousness, (b) severity, (c) relatedness to study drug, (d) action taken with study drug, and (e) outcome. For each level of a given TEAE characteristic, summaries will include the number of events of a given classification and by the number and proportion of participants for which that level of a characteristic was the worst they experienced (treating any unknown characteristic as not worst).

TEAEs will also be summarized separately by MedDRA term at three levels: (a) System Organ Class, (b) High Level Term, and (c) Preferred Term. At each level, the number of events of a given classification and the number and proportion of participants experiencing such an event will be summarized for each treatment group.

3.9.2. Suicidality

The proportion of participants who report any post-baseline suicidal ideation or any post-baseline suicidal behavior will be summarized by treatment group. Suicidal behaviors will include: actual attempts, non-suicidal self-injurious behavior, interrupted attempts, aborted attempts, preparatory acts or behaviors, and completed suicide. The most severe ideation, maximal frequency, maximal duration, minimal controllability, minimal deterrents, maximal reasons for ideation, maximal actual lethality or medical damage, and maximal potential lethality will be summarized as means, standard deviations, medians, and ranges by treatment group.

3.9.3. Safety Labs

The absolute level and the absolute change from baseline for each safety laboratory assay will be summarized as means, standard deviations, medians, and ranges at each visit by treatment group. The proportion of participants with safety lab levels below the lower limit of normal or above the upper limit of normal will be summarized by treatment group by visit and at any post-baseline visit. Shift tables will be used to summarize changes in laboratory assays from baseline to the week 4, 8, and 16 visits when classified as below the lower limit of normal, within normal limits, and above the upper limit of normal separately by treatment group.

3.9.4. Vital Signs

The absolute level and the absolute change from baseline for vital signs will be summarized as means, standard deviations, medians, and ranges at each visit by treatment group.

3.9.5. Blood Pressure

The absolute level and the absolute change from baseline of supine and standing SBP, supine and standing DBP, and differences from supine to standing SBP and DBP will be summarized as means, standard deviations, medians, and ranges at each visit by treatment group. The proportion of participants who experience orthostatic hypotension (as defined in Section 3.3.11) at the week 4, 8, and 16 visits and at any post-baseline visit will be summarized by treatment group.

3.9.6. Additional Continuous Safety Outcomes

The additional continuous safety outcomes of weight, ECG parameters, UPDRS scores (Part I score, Part II score, Part III score, Parts I through III total score, Part IV dyskinesia score, Part IV clinical fluctuations score, Part IV other complications score, and Part IV total score), LDD, and LEDD and their absolute change from baseline will be summarized as means, standard

deviations, medians, and ranges at each visit by treatment group. Treatment-dependent differences in 16-week change scores will be estimated in the ST sample for UPDRS scores using the shared-baseline, repeated-measures analysis of variance described for the primary efficacy analysis in Section 3.8.1.

3.10. Other Analyses

3.10.1. Participant Disposition

The number of patients who were screened, randomized, completed scheduled follow up, and prematurely withdrew study participation will be summarized overall and by treatment group. Reasons for screen failure and for withdrawal from study will be presented.

3.10.2. Study Drug Compliance and Tolerance

The number of days of exposure to study drug will be summarized by treatment group. Compliance with study drug will be calculated as the number of doses taken divided by the scheduled number of doses taken prior to permanent discontinuation, expressed as a percentage. Compliance will be summarized overall and by 4-week interval for each treatment group. Time to discontinuation of study drug will be estimated using Kaplan-Meier product-limit estimates. Treatment-dependent differences will be tested by log-rank test.

3.10.3. Prior and Concomitant Medication Use

All prior and concomitant medications taken during the study period will be listed for each patient, including dosage and indication. Each medication will be coded using the World Health Organization Drug Dictionary Enhanced and classified as a past medication (last dose taken prior to the first dose of study drug), a concomitant medication ongoing at baseline, or a concomitant medication initiated after baseline. The percentage of patients taking each medication (or class of medications) will be summarized overall and by treatment group.

3.10.4. Pharmacokinetic Samples

Results of the plasma assays of SYN120 will be descriptively analyzed by treatment group and visit. Population estimates of C_{max}, t_{max}, and half-life will be estimated based on self-reported time since last dose of study drug. Separate population analysis on the data from this study alone or combined with data from other studies as deemed appropriate will be performed and reported separately.

3.11. Handling of Protocol Deviations

Important protocol deviations are deviations from the protocol which could have a meaningful impact on either the primary efficacy or safety endpoints. The criteria for identifying important protocol deviations and the classification of important protocol deviations will be defined prior to unblinding. To the extent feasible, rules for identifying protocol deviations will be defined without review of the data and without consideration of the frequency of occurrence of such deviations. Whenever possible, criteria for identifying important protocol deviations will be implemented algorithmically to ensure consistency in the classification of important protocol deviations across all patients.

4. REFERENCES

Andreasen N. (1984) Scale for the assessment of positive symptoms (SAPS). Iowa City, IA: University of Iowa.

Ayele BT, Lipkovich I, Molenberghs G, Mallinckrodt CH. A multiple-imputation-based approach to sensitivity analyses and effectiveness assessments in longitudinal clinical trials. J Biopharm Stat. 2014;24(2):211-28.

Brennan L, Siderowf A, Rubright JD, Rick J, Dahodwala N, Duda JE, Hurtig H, Stern M, Xie SX, Rennert L, Karlawish J, Shea JA, Trojanowski JQ, Weintraub D. The Penn Parkinson's Daily Activities Questionnaire-15: Psychometric properties of a brief assessment of cognitive instrumental activities of daily living in Parkinson's disease. Parkinsonism Relat Disord. 2016 Apr;25:21-6.

Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology. 1994 Dec;44(12):2308-14.

Cummings JL. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. Neurology. 1997 May;48(5 Suppl 6):S10-6.

Emre M, Aarsland D, Brown R, Burn DJ, Duyckaerts C, Mizuno Y, Broe GA, Cummings J, Dickson DW, Gauthier S, Goldman J, Goetz C, Korczyn A, Lees A, Levy R, Litvan I, McKeith I, Olanow W, Poewe W, Quinn N, Sampaio C, Tolosa E, Dubois B. Clinical diagnostic criteria for dementia associated with Parkinson's disease. Mov Disord. 2007 Sep 15;22(12):1689-707

Fahn S, Elton RL, Members of the UPDRS Development Committee (1987). The Unified Parkinson's Disease Rating Scale. In Fahn S, Marsden CD, Calne DB, Goldstein M, eds. Recent developments in Parkinson's disease, vol 2:153-163, 293-304. (Florham Park, NJ: Macmillan Health Care Information).

Frey PW, Colliver JA. Sensitivity and responsivity measures for discrimination learning. Learn Motiv. 1973 Aug;4(3):327-42.

Hauser RA1, Hsu A, Kell S, Espay AJ, Sethi K, Stacy M, Ondo W, O'Connell M, Gupta S; IPX066 ADVANCE-PD investigators. Extended-release carbidopa-levodopa (IPX066) compared with immediate-release carbidopa-levodopa in patients with Parkinson's disease and motor fluctuations: a phase 3 randomised, double-blind trial. Lancet Neurol. 2013 Apr;12(4):346-56.

Kaufer DI, Cummings JL, Christine D, Bray T, Castellon S, Masterman D, MacMillan A, Ketchel P, DeKosky ST. Assessing the impact of neuropsychiatric symptoms in Alzheimer's disease: the Neuropsychiatric Inventory Caregiver Distress Scale. J Am Geriatr Soc. 1998 Feb;46(2):210-5.

Liang KY, Zeger S (2000). Longitudinal data analysis of continuous and discrete responses for pre–post designs. *Sankhyā Ser B*. **62**(1):134–148.

Marinus J, Visser M, van Hilten JJ, Lammers GJ, Stiggelbout AM (2003). Assessment of sleep and sleepiness in Parkinson disease. *Sleep*. **26**(8):1049-54.

Mohs R, Knopman D, Petersen R, Ferris S, Ernesto, C, Grundman M, et al. (1997). Development of Cognitive Instruments for Use in Clinical Trials of Antidementia Drugs: Additions to the Alzheimer's Disease Assessment Scale That Broadens its Scope. *Alzheimer's Dis Assoc Disord.* **11**(S2):S13-S21.

Mohs, 1994 unpublished manual

Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. (2005). The Montreal Cognitive Assessment (MoCA): A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. **53**:695-699.

Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, Galasko D, Jin S, Kaye J, Levey A, Pfeiffer E, Sano M, van Dyck CH, Thal LJ; Alzheimer's Disease Cooperative Study Group. Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med. 2005 Jun 9;352(23):2379-88.

Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am J Psychiatry. 1984 Nov;141(11):1356-64.

Schneider LS, Olin JT, Doody RS, Clark CM, Morris JC, Reisberg B, et al. (1997). Validity and reliability of the Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord*. **11(**Suppl 2):S22-32.

Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. Mov Disord. 2010 Nov 15;25(15):2649-53.

Voss T, Bahr D, Cummings J, Mills R, Ravina B, Williams H (2013). Performance of a shortened Scale for Assessment of Positive Symptoms for Parkinson's disease psychosis. *Parkinsonism Relat Disord.* **19**(3):295-299.

Wesnes KA, Simpson PM, White L, Pinker S (1992). The Cognitive Drug Research Computerized Assessment Systems for elderly, AAMI & demented patients. *J Psychopharmacol.* **6**(1):108.

References will be provided upon request.