STATISTICAL ANALYSIS PLAN

TOZ-CL05 Part A (double-blind phase)

A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study with an Open-Label Phase to Determine the Efficacy and Safety of Tozadenant as Adjunctive Therapy in Levodopa-Treated Patients with Parkinson's Disease Experiencing End-of-Dose "Wearing-Off"

Sponsor: Biotie Therapies, Inc.

Development Phase: 3

Preparation Date: 19-May-2015
Version number: FINAL v1.0
Analysis Stage: Final Analysis (Part A)
Prepared by:

SCiAN Services Inc.

SAP Version History:

Version Date	Modified By	Summary of Changes
20-Nov-2014	SCiAN	Initial Draft
11-Feb-2015	SCiAN	Updated draft based on Protocol revised Final Draft (11 February 2015).
26-Mar-2015	SCiAN	Updated draft to reflect change in sample size from N=882 to N=450. Added new section (Section 8) to describe unblinding at completion of Part A.
19-May-2015	SCiAN	Updated SAP to Final version 1.0. Revised Section 8, Unblinding Upon Completion of Part A to clarify that primary and secondary efficacy analyses will be conducted after the Part A data base is locked and will not be modified; and to clarify that exploratory analyses that do not involve the primary or secondary efficacy analyses are subject to modification.

Table of Contents

1 INTRODUCTION

TOZ-CL05 will include a 24-week double-blind phase (Part A) followed by a 52-week open-label phase (Part B). This statistical analysis plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under Part A (double-blind phase) of Biotie's Protocol TOZ-CL05. The SAP for Part B (open-label phase) will be submitted at a later date. The SAP was developed to support Biotie's Special Protocol Assessment (SPA).

This SAP should be read in conjunction with the study protocol and the SPA document. This version of the plan has been developed using final protocol Amendment No. 1, dated 19 May 2015.

Table, figure and listing (TFL) shells will be developed at a later date.

2 STUDY OBJECTIVES

2.1 Primary Efficacy Objective for Part A (double-blind phase)

1) To demonstrate the efficacy of the A2a receptor antagonist tozadenant in the treatment of levodopa-treated PD patients experiencing end-of-dose "wearing-off", based on the change from Baseline to Week 24 in the number of hours per day spent in the OFF state.

2.2 Key Secondary Efficacy Objectives for Part A (double-blind phase)

1) To evaluate the effect of tozadenant on good ON time (defined as the sum of ON time without dyskinesia and ON time with non-troublesome dyskinesia).
2) To evaluate the effect of tozadenant on Unified Parkinson's Disease Rating Scale (UPDRS) Parts II (Activities of Daily Living [ADL] subscale) + III (motor subscale) total scores.

2.3 Other Secondary Efficacy Objectives for Part A (double-blind phase)

1) To evaluate the effect of tozadenant on investigator global impressions of improvement (Clinical Global Impression of Improvement [CGI-I]).
2) To evaluate the effect of tozadenant on patient global impressions of improvement (Patient Global Impression of Improvement [PGI-I]).
3) To evaluate the effect of tozadenant on UPDRS Part III (motor subscale) scores in the ON state.
4) To evaluate the effect of tozadenant on investigator global impressions of severity of illness (Clinical Global Impression of Severity [CGI-S]).
5) To evaluate the effect of tozadenant on UPDRS Part II (ADL subscale) scores.
6) To evaluate the effect of tozadenant on the number of hours per day spent in the ON state without dyskinesia, with non-troublesome dyskinesia, or with troublesome dyskinesia.
7) To evaluate the effect of tozadenant on the number of hours per day spent in the ON state without dyskinesia.

8) To evaluate the effect of tozadenant on the number of hours per day spent in the ON state with any dyskinesia (troublesome or non-troublesome).

9) To evaluate the effect of tozadenant on the number of hours per day spent in the ON state with non-troublesome dyskinesia.

10) To evaluate the effect of tozadenant on PD-related quality of life as measured by Parkinson's Disease Quality of Life Questionnaire (PDQ-39) (total score and individual domain scores).

11) To evaluate the effect of tozadenant on UPDRS Part IV.

12) To evaluate the effect of tozadenant on UPDRS Part I.

2.4 Safety Objectives (Part A)

1) To evaluate the safety and tolerability of tozadenant in levodopa-treated PD patients experiencing end-of-dose "wearing-off", based on assessment of adverse events (AEs), vital signs, electrocardiograms (ECGs), physical and neurological exams, and clinical laboratory tests.

2) To evaluate the effects of tozadenant on the occurrence of daytime drowsiness (including episodes of sudden onset of sleep), impulsive behavior, and suicidality.

3 STUDY DESIGN

This is a Phase 3, international, multicenter, randomized, double-blind, placebo-controlled, parallel-group, 3-arm study with an open-label phase in levodopa-treated patients with PD experiencing end-of-dose "wearing-off". This study includes a Screening Period of up to 6 weeks that starts with a Screening Visit, a 24-week Double-Blind Treatment Period, a 52-week Open-Label Treatment Period, and a Safety Follow-Up Visit 4-weeks post-treatment. After providing written informed consent, patients will undergo screening evaluations. Patients must meet all inclusion criteria and none of the exclusion criteria, and they must successfully complete the diary training and achieve 80% overall diary concordance with the PD diary trainer/rater including at least 1 OFF interval to be considered for randomization. If preliminary eligibility is confirmed by the investigator and the Sponsor/Sponsor's designee, patients will then be scheduled for a Baseline Visit and will be required to return a valid set of Baseline diaries. Final eligibility will be determined at the Baseline Visit.

Eligible patients will be randomized in Part A to receive 1 of 2 tozadenant doses (60 mg BID or 120 mg BID) or placebo in a ratio of 1:1:1. Blinded IMP will be dispensed to patients at each study visit and will be taken BID (approximately 12 hours apart) for 24 consecutive weeks. Patients will return to the study site for evaluation at Weeks 2, 6, 12, 18, and 24. Patients will be contacted by telephone before the start of each 3-day PD diary completion period to be reminded to complete the diary for 3 consecutive days prior to their next scheduled visit. Unscheduled visits can be arranged as considered necessary by the investigator. In the event of Early Termination during Part A, patients will be asked to complete an Early Termination Visit as soon as possible and then return 28 ± 3 days after their last dose of IMP for a Post-Early Termination Safety Follow-Up Visit. Blinded safety data from the double-blind phase (Part A) will be provided to the Data and Safety Monitoring Board (DSMB), at a frequency specified by the DSMB charter, for ongoing monitoring of safety and detection of any trends. If requested by the DSMB, the DSMB statistician will provide unblinded safety data to the DSMB on a per needed basis, to enable a fully informed review.

The Schedule of Events/Evaluations for Part A is shown in **Table 3-1**.

Table 3-1: Part A – Schedule of Events/Evaluations

TOZ-CL05 PART A: DOUBLE-BLIND PHASE									
Study Period	**Screening** [a]	**Baseline Predose**	**Double-Blind Treatment** (24 weeks)					**Early Termination** [x]	**Fol**
Study Week	**−6 to −1**	**BL**	**2** (±3 d)	**6** (±3 d)	**12** (±3 d)	**18** (±3 d)	**24** (±7 d)		**28**
Assessments / Study Visit	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**A98**	
Written informed consent	X								
Patient to view study-provided video ("What to expect during this clinical trial")	X				X				
Demographic data, medical history, including neurological and PD history	X								
Recording of concomitant and anti-PD medications	X	X	X	X	X	X	X	X	
BP [b], pulse [b]	X [c]	X [c]	X	X	X	X	X	X	
Weight (include height at Screening)	X						X	X	
Physical and neurological examination	X						X	X	
Caffeine Intake Questionnaire		X							
PD diary training and diary concordance session	X								
12-lead ECG [e]	X	X	X	X	X	X	X	X	
Modified Hoehn and Yahr staging (observed ON; OFF estimated per history)	X								
UPDRS Parts I, II, III, and IV (in ON state) [f]	X	X	X	X	X	X	X	X [x]	
MMSE-II (in ON state)	X								
mMIDI [g]	X			X	X	X	X	X	
C-SSRS	X	X	X	X	X	X	X	X	
Sponsor eligibility review	X								
PD diary collection (phone call prior to start of 3 consecutive 24-hour diary completion periods)	X [h]	X [h]	X [i]	X [i]	X [i]	X [i]	X [i]	X [j, x]	

Table 3-1: Part A – Schedule of Events/Evaluations (continued)

TOZ-CL05 PART A: DOUBLE-BLIND PHASE									
Study Period	**Screening** [a]	**Baseline Predose**	**Double-Blind Treatment** (24 weeks)					**Early Termination** [x]	**Fol**
Study Week	**−6 to −1**	**BL**	**2** (±3 d)	**6** (±3 d)	**12** (±3 d)	**18** (±3 d)	**24** (±7 d)		**28**
Assessments / Study Visit	**V1**	**V2**	**V3**	**V4**	**V5**	**V6**	**V7**	**A98**	
PD diary review	X	X	X	X	X	X	X	X [j]	
Final verification of inclusion and exclusion criteria		X							
Patient randomization		X							
CGI-I [k]			X	X	X	X	X	X [x]	
PGI-I [k]			X	X	X	X	X	X [x]	
CGI-S [k]		X	X	X	X	X	X	X [x]	
SCOPA-cog [k]		X					X	X	
PDQ-39 [k]		X			X	X	X	X	
EQ-5D-5L [k]		X					X	X	
TSQM-9 [k]				X			X	X	
Sudden onset of sleep		X	X	X	X	X	X	X	
ESS		X	X	X	X	X	X	X	
Fall questionnaire		X					X	X	
Healthcare Resource Utilization			X	X	X	X	X	X	
IMP dispensing and return [l]		X	X	X	X	X	X	X	
Recording of AEs	X [m]	X [m]	X	X	X	X	X	X	
Anti-PD medications		Patients continue stable dose of levodopa and other allowed anti-PD medication							
FSH test, females who are postmenopausal for < 2 years	X								
Thyroperoxidase antibody	X								
Laboratory tests: hematology [n], chemistry [o] (including thyroid [p])	X	X	X	X	X	X	X	X	
Urine pregnancy test, females of childbearing potential [q]	X	X	X	X	X	X	X	X	
Urinalysis [r]	X	X	X	X	X	X	X	X	
Tozadenant blood sampling [s]		X	X [t]	X	X	X	X [t]	X [t]	
Pramipexole blood sampling (applicable patients only)		X [u]	X [v]	X	X	X	X [v]	X [v]	
eCRF completion	X	X	X	X	X	X	X	X	

Table 3-1: Part A – Schedule of Events/Evaluations (continued)

TOZ-CL05 PART A: DOUBLE-BLIND PHASE
AE, adverse event; BP, blood pressure; CGI-I, Clinical Global Impression of Improvement; CGI-S, Clinician Global Impression of Severi Severity Rating Scale; d, day(s); ECG, electrocardiogram; eCRF, electronic case report form; EQ-5D-5L, EuroQol 5D-5L Health Questio Sleepiness Scale; ET, Early Termination; FSH, follicle stimulating hormone; IMP, investigational medicinal product; mMIDI, Modified M Interview; MMSE-II, Mini-Mental State Exam−Second Edition (MMSE-II); P, pulse (beats per minute); PD, Parkinson's disease; PDQ-3 of Life Questionnaire; PGI-I, Patient Global Impression of Improvement; SCOPA-cog, Scales for Outcomes in Parkinson's Disease - cog Satisfaction Questionnaire for Medication; UPDRS, Unified Parkinson's Disease Rating Scale; V, Visit.
Part A - Footnotes:
[a] Screening Period may not exceed 6 weeks.
[b] Obtain and record BP and pulse measurements after at least 5 minutes supine rest and again after standing for approximately 1 minute a
[c] At Screening and at Baseline (before dosing), obtain and record serial BP and pulse measurements after at least 5 minutes supine rest a approximately 1 and 3 minutes, on 3 occasions approximately 10 minutes apart.
[d] Optional assessments that may be performed for evaluation of AEs, at the investigator's discretion.
[e] Resting supine 12-lead ECGs will be collected after the patient has been in a supine position for a minimum of 5 minutes. ECGs should visit when the patient is not experiencing dyskinesia that would interfere with an adequate recording. At Baseline, obtain triplicate 12-l performed several minutes apart).
[f] UPDRS to be measured in ON state approximately 1 to 3 hours after patients have taken a scheduled dose of levodopa (preferably their Patients will be instructed to have already taken their normally scheduled dose of levodopa (and their IMP for visits following randomi 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 represent patient, not in the patient's "best" ON.
[g] At Screening, send patient for structured clinical interview if one or more positive mMIDI modules. If the structured clinical interview not have an ICD, he/she will not be considered ineligible on that basis. After Baseline, if patient has one or more positive modules on t participation in the study should be discussed with the Medical Monitor.
[h] The 3-day practice diary during the Screening Period and the 3-day Baseline diary must both be valid in order to randomize a patient. T patient before the scheduled start of the diary completion periods to remind him or her to keep the PD diary and to review completion i be reminded to send the completed practice diary to the trainer/rater and to bring their Baseline diary to the Baseline Visit. If the practic the patient may be retrained and complete another practice or Baseline diary within the 6-week window of the Screening Period, if the the study.
[i] PD diary collected over the 3 consecutive 24-hour periods before the day of the scheduled study visits on Week 2, 6, 12, 18 and 24. The the patient before the scheduled start of the 3-day PD diary completion period (at the latest, on the last working day before the schedule completion) and review completion instructions. The patient will also be reminded to bring their PD diary to the visit. The trainer/rater diary contains missing and/or invalid entries to reinforce instructions for appropriate completion.
[j] Done only if the ET date coincides with the scheduled diary collection return date.
[k] To be collected during patient's ON state.
[l] Patients will be instructed to take two (2) tablets of the dispensed blinded IMP by mouth twice daily (BID), in the morning and in the e time each day, for a total of four (4) tablets per day. The evening dose should be approximately 12 hours after the morning dose. Patient morning and evening doses at least 1 hour before or 2 hours after a meal and to refrain from eating for at least 1 hour after dosing.
[m] Pretreatment AEs.

Table 3-1: Part A – Schedule of Events/Evaluations (continued)

TOZ-CL05 PART A: DOUBLE-BLIND PHASE
[n] Hematology tests: Hemoglobin concentration, hematocrit, red blood cell count, total and differential white blood cell, thrombocyte (pla
[o] Blood chemistry (including liver function) tests: Aspartate amino transferase (AST), alanine amino transferase, (ALT), gamma-glutamy phosphatase (ALP), lactate dehydrogenase (LDH), total bilirubin (conjugated and unconjugated), albumin, creatinine, urea/blood urea n uric acid, total protein, sodium, chloride, potassium, calcium, phosphate, glucose, cholesterol, creatine phosphokinase (CK). Regarding Screening, refer to Exclusion Criterion #25 (Section 4.2).
[p] TSH, free T3, and free T4.
[q] Document method of contraception at Screening and verify continuation of (or any change to) contraceptive method at each visit. For f perform a urine pregnancy test.
[r] Urinalysis: Specific gravity, pH, ketones, blood, protein, glucose. If urine dipstick is positive for leukocytes, protein, or erythrocytes, a culture will be performed.
[s] PK blood samples will be collected for determination of tozadenant concentration levels at Baseline before dosing, at Weeks 2, 6, 12, 1 in the event of an Early Termination Visit unless the patient's last dose of IMP was taken more than 7 days before the visit. For Weeks 6 Termination, one (1) PK blood sample will be taken at the most convenient time at each visit. For Unscheduled Visits, one (1) PK blood investigator's discretion (e.g., for AEs thought to be study drug related). For each collection, the time of sampling and patient-reported the most recent IMP dosing prior to sampling will be recorded. See Section 8.1.
[t] Collection of two (2) tozadenant PK blood samples at Weeks 2 and 24 is required. The first sample will be taken on arrival and the seco with the time between the blood samples being at least 45 minutes or longer and no IMP dosing between the two blood draws. (For an l blood samples are not required if the patient's last dose of IMP was taken more than 7 days before the Early Termination Visit.) See Sec
[u] Only for patients concomitantly taking pramipexole: PK blood sample collected for tozadenant will also be analyzed for plasma pramip predose). A second PK blood sample for plasma pramipexole concentration will be taken at Baseline at least 45 minutes after the first P
[v] Only for patients concomitantly taking pramipexole: PK blood samples collected for tozadenant will also be analyzed for plasma prami both samples collected at Weeks 2 (Visit 3) and 24 (Visit 7) and Early Termination Visit during Part A. If a tozadenant sample is not be Termination visit (patient's last dose of IMP was taken more than 7 days before the visit) a PK sample for plasma pramipexole concent
[w] Only for patients with potential pramipexole-related AEs: During Part A, in the event of AEs thought to be pramipexole-related (e.g., d delusions, hypotension, somnolence, nausea, vomiting), either noted at the time of a scheduled study visit or an Unscheduled Visit, PK tozadenant will also be analyzed for plasma pramipexole, i.e., at Week 6 (Visit 4), Week 12 (Visit 5) or Week 18 (Visit 6). If a tozaden at an Unscheduled visit a PK sample for plasma pramipexole concentration will be collected. See Section 8.2.
[x] If patient has discontinued IMP, perform the Early Termination Visit as soon as possible after the last dose of IMP. Efficacy-related mea CGI-I, PGI-I, and CGI-S) will not be completed at an Early Termination Visit if patient stopped taking study drug more than 24 hours p measures during the Early Termination Visit. If patient took the last dose of IMP 28 or more days prior to the Early Termination Visit, a Follow-Up Visit (Section 6.2.24) is not required.

3.1 Analysis sets

If a subject receives at first dose a dose other than the dose randomized to, the table, figures and listings (TFLs) based on the safety analysis set will display a footnote describing the dosing error.

3.1.1 Intent-to-Treat Set (ITT)

The ITT according to the intention-to-treat principle will consist of all randomized patients. Patients will be accounted for in the treatment group to which they were originally randomized. Summaries of demographic and baseline data will be done in the ITT.

3.1.2 Safety Set (SS)

The SS will consist of all randomized patients who received at least one dose of IMP. Patients will be accounted for in the treatment group based on IMP actually received at first dose. Summaries and/or analyses of safety data will be done in the SS.

3.1.3 Modified Intent-to-Treat Set (mITT)

The mITT will consist of all randomized patients who took at least 1 dose of study drug and had valid diaries at baseline and had valid diaries on at least 1 post-baseline visit. Patients will be accounted for in the treatment group to which they were originally randomized.

Summaries and/or analyses of all efficacy data will be done in the mITT.

3.1.4 Per Protocol Set (PPS)

The PPS will consist of all randomized patients who have an efficacy evaluation (number of hours per day spent in the OFF state while awake) at Week 24 and have no major protocol deviations/violations. In addition, subjects randomized to tozadenant must have measureable plasma concentrations at each post-baseline visit for the subject to be considered in the PPS (i.e. not have any below-the-limit-of-quantification result from the bioanalysis at any post-baseline visit). This will not exclude subjects who had missing samples or values, for any reason, from the analysis. (For example, instances where the analytical lab was unable to obtain a result from a sample they received, or no sample was received, or the bioanalytical results were considered invalid.)

Patients will be accounted for in the treatment group to which they were originally randomized.

Summaries and/or analyses of the primary endpoint and four key secondary endpoints will be done in the PPS.

3.2 Sample Size Considerations

The primary efficacy endpoint is the change from Baseline to Week 24 in the number of hours per day spent in the OFF state. A sample size of 150 in each group will provide 85% power to detect a difference in mean response of 0.9 hours between tozadenant and placebo assuming the common SD is 2.6 hours and using a two group t-test with a 0.050 two-sided significance level. Across the 3 treatment groups, the total number of patients to be randomized is 450.

Although the sample size calculation is based on the primary endpoint, OFF time, for this study the total sample size of 450 patients is anticipated to yield power to detect treatment differences

in the two key secondary endpoints as summarized in the table below. The assumed treatment differences and standard deviations are based on the findings of the SYN115-CL02 study.

Estimated Power for Key Secondary Efficacy Endpoints

Secondary Endpoint 120 mg BID vs. placebo	**SYN115-CL02**		**Power**
	Treatment Difference	**Standard Deviation**	
Good ON time: Number of hours spent in the ON state without troublesome dyskinesia	1.19	2.64	97%
UPDRS Parts II + III total score	3.00	8.38	87%

Approximately 645 patients will be screened, assuming a 30% screen failure rate, to randomize 450 patients.

3.3 Randomization

During Part A, an IXRS will be used for assigning eligible patients to a treatment regimen based on a predetermined randomization schedule. The final randomization schedule will be created and sent to the IXRS provider. The IXRS is responsible for assigning blinded IMP kits to randomized patients, as appropriate, according to the randomization and the visit schedule. Patients will be randomly allocated in equal proportion (1:1:1) to 1 of 3 double-blind treatment groups:

- Group A: tozadenant 60 mg BID (one 60 mg active tozadenant tablet plus one placebo tablet, BID)
- Group B: tozadenant 120 mg BID (two 60 mg active tozadenant tablets, BID)
- Group C: placebo BID (two placebo tablets, BID)

4 STUDY VARIABLES AND COVARIATES

4.1 Primary Efficacy Endpoint (Part A)

The primary efficacy endpoint will be the change from Baseline to Week 24 in the number of hours per day spent in the OFF state, as assessed by patient-completed PD diaries and averaged over 3 consecutive days.

A valid diary (see **Section 5.3**) will not have more than 2 hours (4 invalid entries) over a given 24-hour period. An invalid diary entry is defined as more than one entry recorded in a given half hour interval, an unreadable entry, or the absence of an entry in a given half hour interval. The average diary information from 3 valid diaries (if available) for each visit will be used to calculate diary-based efficacy endpoints. If there are only 2 valid diaries for a visit then the average information from the 2 valid diaries will be used. If only one diary is valid, information from the single valid diary will be used. If no valid diaries are available for a patient visit then the diary information is considered missing.

4.2 Secondary Endpoints

4.2.1 Key Secondary Efficacy (Part A)

Key secondary efficacy endpoints include:

1. Change from Baseline to Week 24 in number of hours per day spent in good ON time, defined as the sum of ON time without dyskinesia and ON time with non-troublesome dyskinesia.
2. Change from Baseline to Week 24 on UPDRS Parts II (ADL subscale) + III (motor subscale) total scores.

4.2.2 Other Secondary Efficacy (Part A)

The other secondary efficacy endpoints include the change from Baseline to Week 24 (where applicable), for the following:

1. CGI-I at Week 24.
2. PGI-I at Week 24.
3. UPDRS Part III (motor subscale) score in the ON state.
4. CGI-S.
5. UPDRS Part II (ADL subscale) score.
6. Number of hours per day spent in the ON state (without dyskinesia, with non-troublesome dyskinesia, or with troublesome dyskinesia).
7. Number of hours per day spent in the ON state without dyskinesia.
8. Number of hours per day spent in the ON state with any dyskinesia (troublesome or non-troublesome).
9. Number of hours per day spent in the ON state with non-troublesome dyskinesia.
10. PDQ 39 (total score and individual domain scores).
11. UPDRS Part IV score.
12. UPDRS Part I score.

4.2.3 Exploratory (Part A)

The exploratory endpoints in Part A include the change from Baseline to Week 24 (where applicable), for the following:

1. Number of hours per day spent in the asleep state.
2. Dyskinesia as measured by questions 32, 33 and 34 of UPDRS Part IV.
3. Falls as measured by question 13 on UPDRS Part II.
4. Responder analysis of number (%) of patients with ≥ 1.0 h improvement in OFF time (evaluated at Week 24).
5. Responder analysis of number (%) of patients with ≥ 2.0 h improvement in OFF time (evaluated at Week 24).
6. Percent change in total levodopa equivalent dose.

7. Percent change in the number of hours per day spent in OFF state.
8. Percent change in the number of hours per day spent in ON state without troublesome dyskinesia.
9. Percent change in the number of hours per day spent in ON state with troublesome dyskinesia.
10. Scales for Outcomes in Parkinson's Disease–cognition (SCOPA-cog) score.
11. Fall questionnaire score.
12. Number of hours per day spent in the OFF state in the sub-population of subjects taking levodopa and only one additional concomitant anti-PD medication.
13. Number of hours per day spent in the OFF state in the sub-population of subjects taking levodopa and two additional concomitant anti-PD medications.
14. Number of hours per day spent in the OFF state in the sub-population of subjects taking levodopa and three additional concomitant anti-PD medications.
15. EuroQol 5D-5L Health Questionnaire (EQ-5D-5L).
16. Treatment Satisfaction Questionnaire for Medication (TSQM 9) (evaluated at Week 76).

4.3 Safety Endpoints

The safety and tolerability of tozadenant will be evaluated in this study using the following measures:

- Treatment-emergent adverse events (TEAE).
- Physical and neurological examination.
- Supine and standing pulse and blood pressure.
- Standard 12-lead electrocardiogram (ECG): RR, PR, QRS, QT and QTcF.
- Laboratory parameters: hematology, chemistry, thyroid function (thyroid stimulating hormone [TSH], free T3, and free T4), and urinalysis.
- Columbia-Suicide Severity Rating Scale (C-SSRS).
- Daytime sleepiness as measured by the Epworth Sleepiness Scale (ESS), including assessment of episodes of sudden onset of sleep.
- Modified Minnesota Impulsive Disorders Interview (mMIDI).

4.4 Pharmacokinetic Variables

- Tozadenant concentrations and time point of IMP intake and blood sampling

4.5 Predetermined Covariates and Prognostic Factors

The primary, secondary and exploratory analyses will be based on a mixed model repeated measures (MMRM) ANCOVA, where the covariate in the analysis is the baseline value of the respective response variable. For example, if the number of hours per day spent in the OFF state is being analyzed, the covariate will be the baseline number of hours per day spent in the OFF state. More details are provided in **Sections 9.6.1** to **9.6.3** of this SAP.

5 DEFINITIONS

5.1 Treatment Group Labels

The following treatment group labels will be used in the tables, listings and figures:

- Placebo
- 60 mg BID Tozadenant
- 120 mg BID Tozadenant
- All Tozadenant (select tables only)
- All (select tables only)

5.2 Age

The following SAS® code is used to calculate subject age (years):

Age = floor((intck('month', birth date, IC date) - (day(IC date) < day(birth date))) / 12),

where intck is a SAS® function counting integer days, date of birth is database variable BRTHDT, and informed consent (IC) date is database variable INFCDT.

5.3 Valid Diary

A valid diary will not have more than 2 hours (4 invalid entries) over a given 24-hour period. A valid diary record will not have more than 4 invalid entries (double or missed entries) over one 24-hour period. An invalid diary entry is defined as more than 1 entry recorded in a given half-hour interval, an unreadable entry, or the absence of an entry in a given half-hour interval.

5.4 Baseline for Diary Data

Baseline is defined as the average of a maximum of 3 days diary preceding the first dose.

5.5 Baseline Complete UPDRS Assessment

A valid UPDRS assessment is one in which UPDRS Part III has been completed in the ON state. The UPDRS assessment utilized as the baseline will be the later of the Screening or Baseline visit assessment that is valid.

The selected assessment will be used to determine the Baseline UPDRS Parts I – IV subscores and the UPDRS I-III total score. If a valid UPDRS is not available at Screening or Baseline visit, then the subject will be excluded from "change from baseline" UPDRS summary statistics and analyses.

5.6 Baseline for Blood Pressure and Pulse Rate

Baseline is defined as the average of the 3 pre-dose measurements for the respective supine or standing measurements that are taken following 5 minutes supine rest and following 3 minutes standing, respectively and at least 10 minutes apart prior to first dose on Day 1. Should one of the 3 measurements be missing, the baseline will be the average of the available pre-dose measurements and if two or more of the 3 measurements are missing, the baseline will be missing.

5.7 Baseline Non-Blood Pressure, Non-Pulse Rate and Non-Diary Data

Baseline is defined as the last measurement prior to first dose of study medication at study Day 1. If for any parameter the study Day 1 pre-dose value is not done or missing, then the last value obtained at a Screening visit is used as Baseline.

5.8 Change from Baseline

Change from baseline will be the calculated as the later data value minus baseline.

5.9 Concomitant and Prior Medication other than Anti-PD Medications

Concomitant medications other than anti-PD medications are defined as any medications ongoing at the start of treatment or with a start date and time on or after the date of first study medication dose at study Day 1. Prior medications are defined as a medication with a stop date prior to first dose of study drug.

For the classification of medications as prior or concomitant, medications with missing or partial stop dates will be considered as follows:

- If the stop date is completely missing, the medication will be regarded as concomitant.
- If the day and month are missing but the year is given and the year is prior to the year of first dose, the medication will be classified as prior medication. Otherwise such medications will be considered as concomitant.
- If the day is missing but the month and year are given and the month and year are prior to the day of first dose, the medication will be classified as prior medication. Otherwise such medications will be considered as concomitant.

5.10 Concomitant Anti-PD Medications

During Part A, patients are required to be on a stable regimen of permitted anti-PD medications for at least 4 weeks prior to Screening (see inclusion and exclusion criteria). The stable anti-PD medication regimen must include at least four doses of a levodopa-containing medication per day, and at least one other concomitant anti-PD medication (dopamine agonists, MAO-B inhibitors, anticholinergic agents, amantadine or entacapone).

All efforts should be made to maintain patients on the same stable dose and frequency of their anti-PD medications throughout Part A. After randomization, the addition of any new anti-PD medications or an increase in the dose of any anti-PD medications is not permitted. Likewise, changes to the frequency (number of doses taken per day) or to the intervals between doses (duration of time between doses on a given day) of a patient's anti-PD medication(s) are not permitted. A PRN (as needed) dose of levodopa or other anti-PD medication would be considered a change in frequency as it is not a regular part of the daily regimen.

However, at the investigator's discretion, a decrease in the total daily dose of concomitant anti-PD medication because of medication-related AEs is permitted:

- A decrease in total daily dose should be done by lowering the number of tablets taken and must not be done by changing (lowering) the frequency (# of times per day the dose is taken).
- Following a decrease in total daily dose, the dose may be increased again but cannot exceed the dosage at randomization.

5.10.1 Baseline Total Daily Levodopa

The total daily levodopa dose at Baseline will be calculated as dose x # of tablets x frequency per day.

5.10.2 Baseline Total Daily Levodopa Equivalents

The Levodopa Equivalents (LED) of each concomitant anti-PD drug being taken at Baseline will be calculated as the total daily dose of the medication multiplied by a conversion factor to obtain the LED for the medication. The total daily LED will be the sum of each LED medication. Conversion factors for each medication will be based on **Tomlinson et al, 2010**.

5.11 Informed Consent Date

Informed consent date is the date the subject signs the informed consent and this date is recorded in the database.

5.12 Missing Adverse Event Relationship or Severity

AEs with missing relationship will be considered as having a relationship of possible to study drug. AEs with missing or unknown severity will be considered severe. A subject with missing relationship or severity for an AE will be footnoted as having missing data in the corresponding tables and listings for the AE with missing data.

5.13 Treatment-Emergent Adverse Events (TEAEs)

TEAEs are those which first occur or increase in severity or relationship to study drug after the first dose of study drug through the Post-dose Follow-up (28 ± 3 days after end of treatment). In reality all AEs which change in severity or relationship to study drug are assigned a new start date and captured as a new record. Should the day of the AE be “unknown” and the AE in question occurs in the same month or a following month as the date of first dose, the AE will be considered treatment emergent. Should the month of the AE be “unknown” and the AE in question occurs in the same year or a following year as the date of first dose, the AE will be considered treatment emergent. Should the start date be missing then the AE will be considered treatment emergent.

6 INTERIM ANALYSES

There is no planned interim analysis for this study.

7 DATA REVIEW

7.1 Data Handling and Transfer

Data are entered electronically into a clinical database built by <<Provider>> and exported as SAS® datasets. SDTM and ADaM datasets will be generated following the Clinical Data Interchange Standards consortium (CDISC) conventions. Data analyses including summary tables, figures, and data listings are produced using SAS®.

Medical conditions and AEs are coded using the Medical Dictionary for Regulatory Activities (MedDRA) version <<xx>> or higher, to assign a system organ class (SOC) and preferred term (PT) to each AE. Prior and concomitant medications are coded to preferred drug names using the World Health Organization Drug Dictionary (WHODRUG) version <<version>> or higher.

Clinical laboratory results and normal ranges will be provided as SAS datasets from the central clinical laboratory vendor. Protocol deviations and violations (PDVs) will be identified by the Sponsor prior to database lock and flagged for exclusion from the ES population.

Randomization data will be imported into the dataset after database lock at the time of unblinding.

7.2 Data Screening

There will be one database freeze during which the data will be exported and the TLFs generated based on this data export. The TFLs will be reviewed by the Sponsor in a blinded data review meeting to identify any final data issues, determine major/minor violations and evaluable subjects analysis set and seek corrections prior to database lock. The Sponsor must approve database lock.

8 UNBLINDING UPON COMPLETION OF PART A

Immediately after Part A (double-blind phase) of the study completes and the database is cleaned and locked, the responsible CRO(s) will follow their SOPs to unblind the Part A. . A blinded data review will be performed (per ICH E9) before the database is locked and before the study groups are unblinded. Primary and secondary efficacy analyses are final and will not be modified. Exploratory analyses that do not involve the primary or secondary efficacy analyses are subject to modification.

9 STATISTICAL METHODS

The software used for all summary statistics and statistical analyses will be SAS® (SAS Institute, Inc.)[1] Version 9.2 or later.

Most continuous data will be summarized with the following descriptive statistics unless otherwise noted: number of observations, mean, SD, median, minimum, and maximum; interquartile ranges will be provided as appropriate. Categorical data will be summarized with frequencies and percentages.

Missing data will not be imputed unless otherwise documented in this SAP.

Unless stated otherwise, statistical tests conducted will be two-sided at an α level of 0.05. No adjustment for multiplicity will be made unless otherwise stated in the following sections.

For presentation of data, the mean and median will be presented to 1 decimal place greater than the original data, SD will be to 2 decimal places greater than the original data, and the minimum and maximum will have the same number of decimal places as the original data. P-values, if any, will be reported to three decimal places; p-values less than 0.001 will be reported as p<0.001. The format for minimum and maximum will be “Min, Max”. Standard deviation will be abbreviated as “SD”.

If a subject is incorrectly dosed, the tables, figures and listings based on the ITT or mITT analysis sets will include a footnote indicating the dosing error.

9.1 Subject Disposition

Subject disposition data will be summarized for all screened subjects. A table will be presented by treatment group, all tozadenant and overall and will include:

- Number of subjects screened
- Number of subjects in ITT

- Number (%) of subjects in SS
- Number (%) of subjects in mITT
- Number (%) of subjects in PPS
- Number (%) of subjects complete 24 weeks of treatment
- Number (%) of subjects who prematurely withdrew along with reasons for withdrawal

A listing of all subjects screened along with data on their disposition, will be provided, including reasons for non-participation for screen failures.

A table will be presented that will summarize number and percent of subjects enrolled at each site by treatment group for the ITT.

9.2 Protocol Deviations and Violations

Protocol deviations and violations (PDVs) will be identified by the Sponsor prior to database lock and flagged for exclusion from the PPS population. The protocol deviations and violations are categorized and designated major or minor, and finalized prior to database lock and unblinding. Categories include but are not limited to the following: inclusion criteria, exclusion criteria, study drug, assessment - safety, assessment - efficacy, visit window, lab/ endpoint data, visit window, informed consent, anti-PD medication (increased or additional/new), non-anti PD medication (increased or additional/new), PD diary violation and other.

If there is a significant number of major protocol deviations and violations, they will be summarized by deviation/violation category and treatment group. Otherwise, all PDVs will be documented in a data listing only.

9.3 Treatments

9.3.1 Extent of Study Drug Exposure

The number of days of exposure to study drug will be summarized by treatment group in the SS. Descriptive statistics (n, mean, standard deviation, median, minimum, and maximum) for the total number of tablets taken, number of tablets returned and number of tablets lost will be displayed. Compliance with study drug will be calculated as the number of doses taken divided by the scheduled number of doses taken, expressed as a percentage. Compliance will be summarized overall and by scheduled weeks (Week 2, 6, 18 and 24) for each treatment group for the safety analysis set. The reason for not returning tablets is captured on the CRF and will be presented in the data listings.

9.3.2 Concomitant and Prior Medications other than Anti-PD Medications

Prior and concomitant medications other than anti-PD medications will be coded to a preferred name according to the WHODRUG coding dictionary version <<version>> or higher. Prior medications are defined as those medications that had the last dose taken prior to the first dose of study drug given, whereas, concomitant medications are defined as medication ongoing at baseline, or medication initiated on or after the first dose day of study drug administration. The number and percent of subjects will be summarized for prior and concomitant medications for each preferred term (PT) by treatment group in the SS. All prior and concomitant medications taken during the study will be listed for each subject including dosage and indication.

9.3.3 Concomitant and Prior Anti-PD Medications

Prior and concomitant anti-PD medications will be coded to a preferred name according to the WHODRUG coding dictionary version <<version>> or higher. Prior anti-PD medications are defined as those medications that were taken for the disease and stopped or were modified since the time of diagnosis. Concomitant anti-PD medications are defined as those medications that were taken for the disease on or after the day of first dose of study drug. The number and percent of subjects will be summarized for prior and concomitant anti-PD medications for each preferred term (PT) by treatment group in the SS. All prior and concomitant anti-PD medications will be listed separately for each subject.

Any changes to anti-PD regimen during or after the study treatment period will be presented in a frequency table by treatment group and as a subject data listing. The summary will be based on the safety analysis set and percentages will be based on the number of subjects in the safety analysis set.

The last dose of anti-PD medication taken the evening before the day when UPDRS part III test are done in defined “ON“ state at Screening, Baseline (Pre-dose Day 1), Weeks 2, 6, 12, 18 and 24 will be documented in a subject data listing along with the date and time of UPDRS Part III performed before anti-PD medications. The subject data listing will include medication/treatment, dose, number of tablets or capsules taken, frequency, date of last dose and time of last dose.

9.3.3.1 Total Levodopa Dose at Baseline

The total levodopa dose at baseline will be calculated (**Section 5.10.1**) and summarized by treatment group.

9.3.3.2 Total Levodopa Equivalents at Baseline

The total levodopa equivalent dose at baseline will be calculated (**Section 5.10.2**) and summarized by treatment group.

9.4 Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized descriptively for the ITT, SS, mITT and PPS by treatment group and total tozadenant. These variables include:

- Gender
- Race and ethnicity
- Age (years)
- Body weight (kg)
- Height (cm)
- Body Mass Index (BMI) (kg/m^2)
- Caffeine Intake (mg/day)
- Duration of PD (years)
- Hoehn and Yahr Staging of PD in ON state and OFF state (best estimate)

An analysis of variance (ANOVA) will be performed independently for age, body weight, height, body mass index, duration of PD, duration of levodopa treatment, with treatment group as a class variable. Age will also be classified into 3 categories (≥ 30 to 54, ≥ 55 to 68 and ≥ 69 to 80 years). Within each ANOVA, each tozadenant treatment group will be compared to placebo

through "Estimate" statements. Chi-square tests will be performed for age (categorized), gender, race, ethnicity, caffeine intake and Hoehn and Yahr Staging of PD in ON state and OFF state (best estimate) comparing each tozadenant treatment group to placebo. Additionally, an ANOVA or Chi-square test with region/country as a factor will be performed on the each variable listed above, to determine if there are any country differences.

9.5 Medical History events

Medical conditions resolved at randomization and medical conditions ongoing at randomization will be summarized by MedDRA system organ class (SOC) and PT for the SS by treatment group. All medical conditions collected on the eCRF will be included in the listings.

9.6 Efficacy Analyses

All analyses of the primary and four key secondary endpoints will focus on the comparison of tozadenant 120 mg BID vs. placebo followed by a comparison of tozadenant 60 mg BID vs. placebo. Adjustment for multiplicity for p-values from the analyses of the four key secondary endpoints will be performed per **Section 9.6.6**. All other p-values derived from analyses of the remaining efficacy endpoints will be reported as nominal without adjustment for multiplicity.

To explore treatment differences by geographical location, a country/region term will be included in the analysis model. If there are insufficient number of patients represented in a country, patients by be pooled using a specific criterion that will be determined prior to database lock.

9.6.1 Primary Variable

The primary variable, change from baseline to Week 24 in the 3-day average of total awake time in hours per day spent in the OFF state, will be summarized and analyzed in the mITT and repeated in the PPS.

Descriptive statistics (n, mean, standard deviation, median, minimum, and maximum) will be used to summarize results for OFF time by treatment group and week. Only valid diaries (see **Section 5.3**) post baseline are included in the analyses.

At Week 24, a plot of the cumulative distribution (cumulative of subjects) of average change from baseline in OFF state in hours will be displayed.

The null and alternative hypotheses are as follows:

H_0: $\mu_a - \mu_p = 0$

H_1: $\mu_a - \mu_p \neq 0$

where μ_a and μ_p are the mean changes from baseline to Week 24 in the 3-day average of total awake time in hours per day spent in the OFF state for the tozadenant and placebo groups, respectively.

The primary analysis will be performed on the change from baseline OFF hours using a mixed model repeated measures ANCOVA that includes fixed terms for country/region, treatment group, week (categorical: 2, 6 12, 18 and 24), interaction between treatment group and week, baseline number of hours of OFF time (covariate) and random effect for subject. This model will allow for estimation of treatment effects (each dose regimen of tozadenant vs placebo) separately at each week as well as averaged across weeks.

Restricted Maximum Likelihood (ReML) estimation will be used with an unstructured covariance structure to model the within-patient error and the Kenward-Roger approximation to estimate the

degrees of freedom. If the model does not converge using the unstructured covariance structure, the AR(1) structure may be considered.

MMRM model assumptions including constant variance, normality of errors and normality of random intercepts will be assessed by residual plots.

A test of the null hypothesis of the equality of the adjusted group means in the tozadenant and placebo groups will be performed for each dose regimen of tozadenant using a significance level of 5% (two tailed). Treatment differences will be summarized by least squares means (standard errors) and corresponding 95% confidence intervals.

9.6.2 Imputation Rules using Multiple Imputation for Sensitivity Analysis

This section describes two methods for multiple imputation of missing data. With each imputation method, the analysis described in **Section 9.6.1** will be repeated using the mITT and PPS.

In addition, for each imputation method, a univariate ANCOVA will be performed at Week 24 which includes the model terms: country/region, treatment group and baseline value.

9.6.2.1 LOCF

The LOCF approach will be applied to the change from baseline in the 3-day average of total awake time in hours per day spent in the OFF state. Intermittent missing data between weeks will not be imputed.

9.6.2.2 Multiple Imputation Method

The sensitivity analysis described in this section is based on a particular assumption of missing not at random (MNAR), or non-ignorable missingness mechanism. For subjects who withdraw from one of the tozadenant treatment groups, the primary endpoint for missing weeks will be derived using multiple imputation based on the mean in the placebo group. Withdrawals from the placebo group will also be imputed based on the placebo data. Intermittent missing data, that is, data that is missing at one week but present at a later week, will be assumed to be missing at random (MAR) and will be imputed using a method appropriate for that assumption. When the missing data has been imputed, the primary endpoint data will be analyzed as a sensitivity analysis using the same method as the primary analysis.

There are two types of missing data that will be imputed:

1. Data that is missing at a specific week but present at a later week (intermittent missing data)
2. Data that is missing after a subject drops out (monotonic missing data)

The following process will be used to impute all types of missing values for the primary analysis:

Step 1:

- For intermittent missing data, Markov Chain Monte Carlo (MCMC) simulation will be used to impute values.

The MCMC imputation process will follow the general pre-planned guidelines described below. However the ability of the process to converge can be dependent on the amount of data that is missing. Convergence of the process will be assessed after unblinding.

The number of burn in iterations will be set at 200, or higher if deemed necessary. The thinning option will be set to select one out of every 100 iterations to insure sufficient separation from previous iterations. Ten imputations will be the initial setting. Trace plots will be checked for

convergence. Factors included in the MCMC algorithm used to impute the missing values will be randomized treatment and the baseline value of the outcome variable.

Step 2:

3. For monotone missing data, a regression model will be used to impute values as follows:
4. Subjects that do not have a Week 2 value or are in the Placebo arm are selected.
5. SAS PROC MI will be used to impute missing values using the Baseline time point.
6. The imputed data will be combined with the non-missing tozadenant data to achieve a database with data present for all subjects in the analysis set.

Step 3:

7. Step 2 will be repeated for each week iteratively adding a single Week at each step (i.e. the second imputation will have Baseline, Week 2 in the model, the next will also include Week 6).

Step 4:

8. The final outcome dataset (which will have no missing data) will be analyzed using the primary MMRM analysis.

Estimates will be derived using the SAS PROC MIANALYZE procedure.

Because intermittent missing values may not occur, and there may not be patients in the combined tozadenant dose group dropping out at a designated week, not all the steps above will necessarily be implemented. Assuming there will be subjects in some of the tozadenant groups dropping out from the study, the total number of imputations will be set at 10 if there are intermittent missing values, and 20 otherwise.

The total number of imputations is controlled as follows. Set K = 10 if there are intermittent missing values, and K = 20 otherwise. Starting from Step 2, select the week where the combined tozadenant group had the maximum total number of dropouts among all weeks. Only at this week will the number of imputations be set to K using the regression method at that step. If more than one week has the same maximum number of tozadenant dropouts, then the earliest among such weeks will be selected to have K imputations. Only one imputation will be made for all other weeks.

The seed used for the primary side for Step 1 will be 32192. All subsequent seeds generated will be stored for reproduction of the results if necessary.

9.6.3 Key Secondary Variables

All secondary variables will be analyzed by the same model as described for the primary outcome variable, with the baseline covariate value being the baseline outcome variable of interest included in the statistical model instead of the baseline number of hours of OFF time in the mITT and PPS. The sensitivity analyses described in **Section 9.6.2** will also be performed for the key secondary outcomes.

The following key secondary efficacy endpoints will be summarized and analyzed:

- Change from baseline to Week 24 in the number of hours per day spent in good ON time, defined as the sum of ON time without dyskinesia and ON time with non-troublesome dyskinesia.
- Change from baseline to Week 24 in UPDRS Parts II (ADL subscale) + III (motor subscale) total scores.

For the UPDRS Part I, II, III, sum of Parts II+III and I – III and Part IV (Part A, B and C and total) if a question making up the total score is missing, the total score for that Part will not be determined for the subject for the given visit, as well as the sum of Parts I - III. Subjects with missing total score will be footnoted.

9.6.4 Other Secondary Variables

All other secondary variables will be analyzed by the same model as described for the primary outcome variable, with the covariate value being the baseline outcome variable of interest included in the statistical model instead of the baseline number of hours of OFF time. The analyses will only be done in the mITT. No sensitivity analyses are planned and there will be no adjustment for multiplicity.

The following other secondary efficacy variables to be summarized and analyzed include: the change from Baseline to Week 24 (where applicable) for the following:

1. CGI-I at Week 24
2. PGI-I at Week 24.
3. UPDRS Part III (motor subscale) score in the ON state.
4. CGI-S.
5. UPDRS Part II (ADL subscale) score.
6. Number of hours per day spent in the ON state (without dyskinesia, with non-troublesome dyskinesia, or with troublesome dyskinesia).
7. Number of hours per day spent in the ON state without dyskinesia.
8. Number of hours per day spent in the ON state with any dyskinesia (troublesome or non-troublesome).
9. Number of hours per day spent in the ON state with non-troublesome dyskinesia.
10. PD-related quality of life as measured by PDQ-39 (total score and individual domain scores).
11. UPDRS Part IV score.
12. UPDRS Part I score.

For the UPDRS Part I, II, III, sum of Parts I – III and Part IV (Part A, B and C and total) if a question making up the total score is missing, the total score for that Part will not be determined for the subject for the given visit, as well as the sum of Parts I - III. Subjects with missing total score will be footnoted.

For the PDQ-39 questionnaire, if a question is not answered, then the total score or the dimension score for the PDQ-39 will not be determined for the subject at the visit. Missing data will be footnoted.

Dimensions for the PDQ-39 will be derived by the sum of scores of each question in the dimension divided by (maximum score per question multiplied by number of questions in the dimension). The PDQ-39 single index score of the 8 dimensions will be derived as the sum of all 8 dimension scores divided by 8. The mobility dimension consists of questions 1 to 10 of the PDQ-39, activities of daily living dimension consists of questions 11 to 16, emotional well-being dimension consists of questions 17 to 22, stigma dimension consists of questions 23 to 26, social support dimension consists of questions 27 to 29, cognitive impairment dimension consists of questions 30 to 33, communication dimension consists of questions 34 to 36 and bodily discomfort dimension consists of questions 37 to 39.

9.6.5 Exploratory Variables

The exploratory variables will be summarized by descriptive statistics by treatment group in the mITT. No statistical analyses are planned. The list of variables includes the change from Baseline to Week 24 (where applicable) for the following:

1. Number of hours per day spent in the asleep state.
2. Dyskinesia as measured by questions 32, 33 and 34 of UPDRS Part IV.
3. Falls as measured by question 13 on UPDRS Part II.
4. Responder analysis of number (%) of patients with ≥ 1.0 h improvement in OFF time (evaluated at Week 24).
5. Responder analysis of number (%) of patients with ≥ 2.0 h improvement in OFF time (evaluated at Week 24).
6. Percent change in total levodopa equivalent dose.
7. Percent change in the number of hours per day spent in OFF state.
8. Percent change in the number of hours per day spent in ON state without troublesome dyskinesia.
9. Percent change in the number of hours per day spent in ON state with troublesome dyskinesia.
10. Scales for Outcomes in Parkinson's Disease–cognition (SCOPA-cog) score.
11. Fall questionnaire score.
12. Number of hours per day spent in the OFF state in the sub-population of subjects taking levodopa and only one additional concomitant anti-PD medication.
13. Number of hours per day spent in the OFF state in the sub-population of subjects taking levodopa and two additional concomitant anti-PD medications.
14. Number of hours per day spent in the OFF state in the sub-population of subjects taking levodopa and three additional concomitant anti-PD medications.

9.6.6 Adjustment for Multiplicity

9.6.6.1 Tozadenant 120 mg BID vs. Placebo and tozadenant 60 mg BID vs. Placebo

All analyses of the primary and two key secondary endpoints will focus on the comparison of tozadenant 120 mg BID vs. placebo followed by tozadenant 60 mg BID vs. placebo. The following six comparisons will be conducted using sequential testing with a fixed sequence, which controls the family-wise error for multiple comparisons at an alpha level of 0.05 (two-tailed):

1. Comparison of tozadenant 120 mg BID vs. placebo using the primary endpoint (change from baseline to Week 24 in the number of hours per day spent in OFF state).
2. Comparison of tozadenant 120 mg BID vs. placebo using the first key secondary endpoint (Change from baseline to Week 24 in the number of hours per day spent in good ON time, defined as the sum of ON time without dyskinesia and ON time with non-troublesome dyskinesia).
3. Comparison of tozadenant 120 mg BID vs. placebo using the second key secondary endpoint (Change from baseline to Week 24 in UPDRS Parts II + III total score).

4. Comparison of tozadenant 60 mg BID vs. placebo using the primary endpoint (change from baseline to Week 24 in the number of hours per day spent in OFF state).
5. Comparison of tozadenant 60 mg BID vs. placebo using the first key secondary endpoint (Change from baseline to Week 24 in the number of hours per day spent in good ON time, defined as the sum of ON time without dyskinesia and ON time with non-troublesome dyskinesia).
6. Comparison of tozadenant 60 mg BID vs. placebo using the second key secondary endpoint (Change from baseline to Week 24 in UPDRS Parts II + III total score).

Testing will begin with tozadenant 120 mg BID vs. placebo for change from baseline to Week 24 in the number of hours per day spent in OFF state. If the p-value is less than or equal to an α=0.05, then the result is considered statistically significant and testing will proceed to the next comparison, and so forth. Testing will stop with the first adjusted p-value > 0.05 and that comparison, as well as any comparison later in the sequence will be considered not statistically significant.

The adjusted p-values $\acute{p}_i$ for i = 1 to 6 will be calculated by taking the maximum of the raw p-value (p_i) from the statistical test and the adjusted p-value ($\acute{p}_{i-1}$) from the prior test, as shown below:

$$\acute{p}_1=p_1$$
$$\acute{p}_2=\max(\acute{p}_1,p_2)$$
$$\acute{p}_3=\max(\acute{p}_2,p_3)$$
$$\ldots$$
$$\acute{p}_6=\max(\acute{p}_5,p_6)$$

p-values from the analyses of all other efficacy outcomes will be reported as nominal, with no adjustment for multiplicity.

9.6.7 Tozadenant Plasma Concentrations

During Part A of the study, blood samples will be collected for determination of plasma tozadenant concentrations at baseline before dosing, at weeks 2, 6 12, 18 and 24 (end of dosing). In addition, samples may be taken at an unscheduled visit and at an early termination visit.

Plasma concentration data will be presented for all subjects in the safety population by treatment group, including placebo. In all calculations, including placebo-group subjects, zero (0 ng/mL) will be substituted for concentration below the quantification limit (BLQ; equivalent to <5 ng/mL) of the assay.

Elapsed time between dosing and plasma sampling will not be tightly controlled and is expected to vary considerably between patients. Summary statistics for tozadenant blood concentrations (number of sample analyzed, number of BLQ values, minimum, and maximum) will be presented by sampling occasion. For treatment visits when 2 consecutive plasma samples are analyzed (weeks 2 and 24), the 2 values will be presented separately.

All data collected on the tozadenant blood concentrations will be included in the data listings.

Separate population analyses on the data from this study alone or combined with data from other studies as deemed appropriate may be performed and reported separately.

9.6.8 Pramipexole Plasma Concentrations

During Part A of the study, blood samples will be collected for determination of plasma pramipexole concentrations at baseline before tozadenant dosing, and at weeks 2 and 24 (end of tozadenant dosing). In addition, samples may be taken at an unscheduled visit and at an early termination visit.

Plasma concentration data will be presented for all subjects in the safety population by treatment group, including placebo. In all calculations, zero (0 ng/mL) will be substituted for concentration below the quantification limit (BLQ; equivalent to <5 ng/mL) of the assay.

Elapsed time between dosing and plasma sampling will not be tightly controlled and is expected to vary considerably between patients. Summary statistics for pramipexole blood concentrations (number of sample analyzed, number of BLQ values, minimum, and maximum) will be presented by sampling occasion. Since 2 consecutive plasma samples will be analyzed on each sampling day, the 2 values will be presented separately.

All data collected on the pramipexole blood concentrations will be included in the data listings.

Separate population analyses on the data from this study alone or combined with data from other studies as deemed appropriate may be performed and reported separately.

9.7 Safety Analyses

9.7.1 Adverse Events

A summary of TEAEs including the number of events reported, the number and percentage of subjects reporting at least 1 AE, the number and percentage of subjects discontinuing due to an AE, the number and percentage of subjects with at least 1 serious adverse event, and the number and percentage of deaths will be presented by treatment group and combined tozadenant group.

A breakdown of the number and percentage of subjects reporting each TEAE, categorized by SOC and PT coded according to the MedDRA dictionary, by treatment group and combined tozadenant group. Note that counting will be by subject, not event and subjects are only counted once within each SOC or PT.

A further tabulation of these data will be presented by treatment group and combined tozadenant group for TEAEs that are related to study drug, as defined by those AEs with relationship to study drug recorded on the eCRF as Possible or Probable. Events that are unrelated or remotely/Unlikely related will not be considered “related” for the Causality assessment. An AE with a missing relationship will be considered as having a relationship to study drug of possible. Subjects with multiple events within a particular SOC or PT will be counted under the category of their most drug-related event within that SOC or PT.

A summary of events reported, categorized by severity (mild, moderate, severe and very severe) will also be provided by treatment group and combined tozadenant group. AEs with missing or unknown severity will be considered severe. Subjects with multiple events within a particular SOC or PT will be counted under the category of their most severe event within that SOC or PT.

A summary of adverse events leading to discontinuation will be provided, grouped by system organ class and preferred term by treatment group and combined tozadenant group.

All AEs (including non-treatment-emergent events) recorded on the eCRF will be listed along with other information such as duration, severity, action taken, and perceived relationship to study drug.

9.7.2 Deaths and Serious Adverse Events

Serious TEAEs will be summarized by the number and percentage of subjects reporting each AE, categorized by SOC and PT coded according to the MedDRA dictionary, by treatment group and combined tozadenant group. Note that counting will be by subject, not event and subjects are only counted once within each SOC or PT.

9.7.3 Laboratory Data

Clinical laboratory values will be evaluated by subject for each laboratory parameter. Laboratory test results will be reported and they will be assigned a LNH classification by each laboratory according to whether the value was below (L), within (N), or above (H) the laboratory reference range. Reference (normal) ranges for laboratory parameters provided by each of the central labs will be included in the clinical study report for this study and will be used for shift tables. Summary tabulations of the continuous clinical laboratory (hematology, chemistry, thyroid function (TSH, free T3 and free T4), and urinalysis) data (n, mean, standard deviation, median, minimum, and maximum) for baseline, each scheduled visit (defined in Table 1), and change from baseline for laboratory data will be provided for the safety analysis set by treatment group. The baseline measurement will be defined as the last assessment prior to the first dose of study drug, including any repeated or unscheduled evaluations. Quantitative urinalysis data will only be listed.

An ANOVA will be performed on change from baseline of last on treatment observation for all continuous clinical laboratory tests, with treatment group as a class variable and within the ANOVA, each tozadenant treatment group will be compared to placebo through estimate statements. The treatment difference from placebo and corresponding p-value will be reported.

Shift from baseline to each scheduled visit will be summarized for the safety analysis set by treatment group for hematology, chemistry, thyroid, and urinalysis variables that are assigned a LNH classification. In addition, shift from baseline to worst post-baseline for the same laboratory variables will be presented based on the LNH classification.

The continuous laboratory hematology, thyroid function (TSH, free T3 and free T4), and urinalysis will have the worst post-baseline value plotted versus the baseline visit for the safety analysis set. The minimum lower limit of normal range and maximum upper limit of normal range amongst laboratories will be drawn in on the x and y axis as vertical lines. The reference normal range will be determined from the normal ranges of all the central laboratories.

All laboratory data will be included in the data listings.

9.7.4 Systolic and Diastolic Blood Pressure and Pulse Rate

Systolic and diastolic blood pressure and pulse rate will be measured in both the supine and standing positions. The scheduled visits that will be summarized are Weeks 2, 6, 12, 18, 24 and early termination. In addition, on Day 1, there are to be 3 measurements for both supine and standing, respectively, at least 10 minutes apart. Orthostatic change (difference between standing and supine) in systolic and diastolic blood pressure and pulse rate will be derived. The baseline value for the supine and standing blood pressure and pulse rate will be the average of the 3 measurements taken at least 10 minutes apart prior to first dose on Day 1, as defined in **Section 5**.

Summary tabulations of the systolic and diastolic blood pressure and pulse rate in both standing and supine positions as well as the orthostatic change (n, mean, standard deviation, median, minimum, and maximum) for baseline, each scheduled visit, and change from baseline will be

provided for the safety analysis set by treatment group. In addition, the placebo corrected mean change from baseline (tozadenant treatment group mean change from baseline minus placebo mean change from baseline) for systolic and diastolic blood pressure and pulse rate in both standing and supine positions as well as the orthostatic change will be summarized (n, mean, standard deviation, median, minimum, and maximum) for the safety analysis set by treatment group.

At each scheduled post-baseline visit and time point (Weeks 2, 6, 12, 18, 24 and early termination), the number and percent of subjects meeting each abnormal criteria displayed below will be summarized by treatment group.

- Supine and standing systolic blood pressure
  - ≥160 mmHg
  - < 100 mmHg
  - ≥20 mmHg increase from baseline
  - ≥20 mmHg decrease from baseline
  - ≥40 mmHg increase from baseline
  - ≥40 mmHg decrease from baseline
- Supine and standing diastolic blood pressure
  - ≥100 mmHg
  - <50 mmHg
  - ≥10 mmHg increase from baseline
  - ≥10 mmHg decrease from baseline
  - ≥20 mmHg increase from baseline
  - ≥20 mmHg decrease from baseline
- Supine and Standing Pulse rate
  - ≥120 bpm
  - <48 bpm
  - ≥30 bpm increase from baseline
  - ≥30 bpm decrease from baseline
- Orthostatic change ≥20 mmHg increase from baseline for systolic blood pressure and diastolic blood pressure
- Orthostatic change ≥20 mmHg decrease from baseline for systolic blood pressure and diastolic blood pressure

For Weeks 2, 6, 12, 18 and 24, plots of the mean change from baseline and placebo corrected mean change from baseline at each timepoint will be displayed for blood pressure and pulse rate in supine and standing position and orthostatic change for the safety analysis set.

All vital sign data collected on the eCRF will be included in the data listings.

9.7.5 Body Weight

Body weight will be summarized at Baseline (Screening), Day 168 and early termination. Baseline will be defined as the last assessment prior to the first dose.

Summary tabulations of weight (n, mean, standard deviation, median, minimum, and maximum) for baseline, each scheduled visit, and change from baseline will be provided for the safety analysis set by treatment group.

9.7.6 Columbia-Suicide Severity Rating Scale

The C-SSRS is a semi-structured interview that captures the occurrence, severity, and frequency of suicide-related thoughts and behaviors during the assessment period. The interview includes definitions and suggested questions to solicit the type of information needed to determine if a suicide-related thought or behavior occurred.

The interview measures presence of suicidality, specifically suicidal ideation (including intensity of ideation) and suicidal behavior.

The C-SSRS must be performed by an individual who is trained in its use and certified to perform the assessment.

Two versions of the C-SSRS are used in this study: the “Baseline/Screening” version, and the “Since Last Visit” version.

The C-SSRS “Baseline” version is completed at Screening (Visit 1), and C-SSRS “Since Last Visit” version is completed at all subsequent visits.

All C-SSRS data collected on the eCRF will be included in the data listings. The number and percent of subjects reporting "Yes" for each of the 5 questions to suicidal ideation and 6 questions about suicidal behavior will be summarized by treatment group and visit. In addition, for suicidal behavior, the total number of actual attempts, total number of interrupted attempts and total number of aborted attempts will be summarized (n, mean, standard deviation, median, minimum and maximum) by treatment group and visit.

9.7.7 Epworth Sleepiness Scale

There are eight situations to the ESS. All data from the ESS eCRF will be included in the data listings. The total score from the ESS will be summarized (n, mean, standard deviation, median, minimum and maximum) by treatment group and visit and change from baseline. The total score is derived as the sum of eight situations. If one or more of the eight situations are missing at a given visit for a subject, the total score will not be derived. All data collected on the ESS will be included in the data listings.

9.7.8 Episodes of Sudden Onset of Sleep

The sudden onset of sleep will be recorded at baseline and Weeks 0, 2, 6, 12, 18 and 24 as well as Early Termination. The responses to "did the subject experience any sudden onset of sleep" will be summarized by count and percentages by treatment group and visit. Shift from baseline to each scheduled visit will be summarized for the safety analysis set by treatment group for each sudden onset of sleep questions. All data collected on the episodes of sudden onset of sleep will be included in the data listings.

9.7.9 Modified Minnesota Impulsive Disorders Interview

mMIDI is reported at Weeks 6, 12, 18, 24 and early termination. For the mMIDI module scores and total score, if a question is not answered with a module, then that given module score and the total score will not be determined for the subject at the visit.

The mMIDI modules are Buying Disorder, Compulsive Gambling, Compulsive Sexual Behavior, Compulsive Eating, and Punding Behavior. The number and percent of subjects in each module with a positive/negative indication for the module will be reported. In addition, shift from baseline to each scheduled visit will be summarized for the safety analysis set by treatment group

All data collected on the mMIDI will be included in the data listings.

9.7.10 ECG

ECG data will be summarized for the following visits: Screening, Weeks 0, 2, 6, 12, 18, 24 and early termination. Descriptive statistics (n, mean, standard deviation, median, minimum and maximum) of each parameter (HR, PR, RR, QRS, QT, QTcB and QTcF) by treatment group will be displayed for the observed value and change from baseline at each scheduled listed above for the safety analysis set.

The number and percentage of subjects with maximum post-dose QT/QTcB/QTcF interval within the following ranges will be summarized by treatment group:

- ≤450 msec
- >450 msec and ≤480 msec
- >480 msec and ≤ 500 msec
- >500 msec.

In addition the number and percentage of subjects with maximum change from Baseline in QT/QTcB/QTcF interval within the following ranges will be summarized by treatment group:

- ≤30 msec
- >30 msec
- ≤60 msec, and >60 msec.

ECG findings will be classified as normal or abnormal, with abnormal values further classified according to clinical significance (yes/no). The percentage of subjects with clinically significant findings and clinically significant treatment-emergent findings will be summarized by treatment group for each visit.

All ECG data collected will be included in the data listings

9.7.11 Physical and Neurological Examinations

Physical and Neurological examinations are performed at Screening, week 24 and early termination. The number and percentage of subjects with normal/abnormal findings will be summarized by body system/neurological area, treatment group and visit. All abnormalities will be included in the data listings.

9.8 Other Exploratory Analyses

9.8.1 Efficacy Sub-group Analyses

The change from baseline to Week 24 in the number of hours per day spent in the OFF state while awake will be assessed in the following sub-groups:

- Caffeine intake at baseline: no caffeine vs. 1 – 500 mg/day caffeine vs. > 500 mg/day caffeine.
- Levodopa dose at baseline
- Levodopa equivalent dose at baseline

The MMRM model described in **Section 9.6.1** will be used in all sub-group analyses with the specific subgroup included as an additional covariate. The analysis will be performed in the mITT analysis set.

The assumption of the absence of a treatment by sub-group interaction will be tested using a liberal significance level of 15%. If an interaction is suggested, the results regarding treatment effects will be summarized descriptively for each sub-group and treatment to explore the source of the interaction.

10 VALIDATION

- 100% of unique tables will be independently reprogrammed by a second programmer for numeric results. Statisticians will be involved in the process of programming and validating tables that include inferential statistical results.
- Figures will be checked for consistency against the corresponding tables and listings
- Listings will be checked for consistency against corresponding tables, figures, and derived datasets.

The entire set of TFL will be checked for completeness and consistency prior to its delivery to the client by the lead statistician/Project Manager and by the Project Director.

11 REFERENCES

1) SAS® Version 9.2, SAS Institute, Cary, NC 27513.
2) Westfall PH, Tobias RD, Rom D., Wolfinger RD and Hochberg (1999). Multiple Comparisons and Multiple Tests: Using the SAS System; SAS Inst. Books by Users, Cary NC, Pages 35-36, Section 2.5.8, Sequential Testing with Fixed Sequences.
3) Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE (2010). Systematic review of levodopa equivalency reporting in Parkinson's disease. *Mov Disord*. **25**(15):2649-53.

APPENDIX 1 GLOSSARY OF ABBREVIATIONS

ADaM	Analysis Dataset Model
AE	Adverse Event
ANCOVA	Analysis of Covariance
BMI	Body Mass Index
BID	Twice a Day
CDISC	Clinical Data Interchange Standards Consortium
CGI-S	Clinical Global Impression-Severity
CGI-I	Clinical Global Impression-Improvement
CI	Confidence Interval
CRO	Contract Research Organization
C-SSRS	Columbia-Suicide Severity Rating Scale
CTMS	Clinical Trial Management System
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
ES	Evaluable Subjects
ESS	Epworth Sleepiness Scale
ITT	Intent-To-Treat
LLQ	Lower Level of Quantitation
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
mMIDI	Modified Minnesota Impulsive Disorders Interview
MAR	Missing At Random
mITT	Modified Intent-to-Treat Set
MNAR	Missing Not At Random
MMRM	Mixed Models Repeated Measures
PD	Parkinson's Disease
PDQ-39	Parkinson's Disease Questionnaire
PDV	Protocol Deviation and Violation
PGI-I	Patient Global Impression-Improvement
PK	Pharmacokinetics
PPS	Per-Protocol Set
PT	Preferred Term
SAP	Statistical Analysis Plan
SAE	Serious Adverse Event
SD	Standard Deviation
SDTM	Study Data Tabulation Model
SOC	System Organ Class
SS	Safety Set
TEAE	Treatment Emergent Adverse Event
TFL	Table, Figures and Listings
UPDRS	Unified Parkinson’s Disease Rating Scale
WHODRUG	World Health Organization Drug Dictionary

APPENDIX 2 LIST OF IN-TEXT TABLES, FIGURES, AND LISTINGS

In-Text tables, figures and listings will be determined at a later date.

APPENDIX 3 LIST OF POST-TEXT TABLES, FIGURES, LISTINGS, AND SUPPORTIVE SAS OUTPUT APPENDICES

The list of post-text listings will be contained in a separate document.

APPENDIX 4 SHELLS FOR POST-TEXT TABLES, FIGURES, AND LISTINGS