A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, EFFICACY AND SAFETY STUDY OF PRN DOSING OF BXCL501 OVER A 12 WEEK PERIOD IN SUBJECTS WITH AGITATION ASSOCIATED WITH DEMENTIA

NCT05271552

SAP Addendum: May 17, 2023

SAP: November 4, 2022

BioXcel Therapeutics, Inc.
Protocol BXCL501-303
Statistical analysis plan addendum

This addendum to the BXCL501-303 BioXcel Therapeutics, Inc. Protocol BXCL501-303 statistical analysis plan (SAP) dated November 4, 2022, describes the following changes and additions to be made to the SAP:

- Study baseline is defined as the last observation prior to initiation of study treatment. For each episode the endpoints are defined based on using the pre-dose value for the episode as baseline.
- In Section 3.2 of the SAP the second paragraph should include the sentence in italics:

  "Upon confirmation of eligibility following Screening and Pre-dose assessments, subjects will be randomized 1:1:1 to receive BXCL501 40 μg, BXCL501 60 μg, or matching placebo film. Randomization will be stratified based on antipsychotic use in the past month. *If subjects are misstratified, then they will be analyzed according to the stratum to which they were assigned at randomization*."
- In Table 1 "Estimand attributes for the primary efficacy analysis" and in Table 2 "Estimand attributes for the key secondary analyses of change from pre-dose in total PEC score at 1 hour and at 30 minutes for the first treated episode of agitation", the description of the attribute for the handling of intercurrent events should specify that only anti-agitation medications are considered to be rescue treatments.
- The following sensitivity analysis for the primary analysis may be included in addition to those described in Section 6.4.1 of the SAP: the MMRM model will be fit with site, baseline PEC, treatment group, antipsychotic use (stratification factor), visit number, and a treatment group by visit number interaction term as covariates.
- In Section 4.3.3, the following endpoints will be added:
  - Percentage of all treated episodes of agitation satisfying the definition of PEC responder, an episode with at least a 40% reduction in PEC total score from pre-dose **by** 2 hours post-dose of study treatment.
  - Percentage of all treated episodes of agitation for which the PEC change from pre-dose **by** 2 hours post-dose of study treatment is a reduction of at least 5 points (measures continued efficacy).

The analysis approach outlined in Section 6.4.3 of the SAP will be used to analyze these endpoints.

- Section 7.1 (Adverse events) is expanded to include the following paragraph:

  "Additional tables will present AEs that emerge in the first 24 hours after administration of BXCL501. These tables will present this information for the first administration and summarized across all administrations of BXCL501."

- Graphical displays may be revised from what is specified in the SAP based on a detailed review of proposed TLFs as specified in Section 6.4.5 of the SAP.

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I am the author of this document
Signing Time: 17-May-2023 | 9:19:08 AM EDT
48E277A86C9E43F8A90E9125A9243E90

Date: 17-May-2023

[redacted], PhD
Executive Director, Biostatistics, WCG Clinical, Inc.

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I have reviewed this document
Signing Time: 17-May-2023 | 7:19:42 AM PDT
[redacted]B3006CB434898D1752EACA46C70

Date: 17-May-2023

[redacted], MD
VP, Clinical Development, BioXcel Therapeutics, Inc.

E:\Proj\Bioxcel\2. SAP\Phase 3 dementia\BXCL501-303 SAP addendum V1.docx\LW

BioXcel Therapeutics, Inc.

Protocol BXCL501-303
A Multicenter, Randomized, Double-blind, Placebo-controlled, Efficacy and Safety Study of PRN Dosing of BXCL501 Over a 12 Week Period in Subjects with Agitation Associated with Dementia

Statistical analysis plan
Version 1.0
November 4, 2022

Prepared by:

WCG Statistics Collaborative

CONFIDENTIAL

Protocol BXCL501-303

Statistical analysis plan

Prepared by: WCG Statistics Collaborative

Authors: DocuSigned by: [redacted] Signer Name: [redacted] Signing Reason: I am the author of this document Signing Time: 09-Nov-2022 | 8:15:35 AM EST [redacted], PhD 48E277A86C9E43F8A90E9125A9243E90 Executive Director, Biostatistics Date: 09-Nov-2022

Approved by: BioXcel Therapeutics, Inc.

DocuSigned by: [redacted] Signer Name: [redacted] Signing Reason: I have reviewed this document Signing Time: 08-Nov-2022 | 5:51:48 PM PST [redacted], MD 252C5B3006CB434898D1752EACA46C70 Chief Medical Officer Date: 08-Nov-2022

Contents

Contents, continued

Contents, continued

Exhibits

Tables

Abbreviations

ACES	Agitation and Calmness Evaluation Scale
AD	Alzheimer's Disease
ADAS-Cog	Alzheimer's Disease Assessment Scale - Cognitive Subscale 12
AE	adverse event
ANCOVA	analysis of covariance
ATC	anatomical therapeutic chemical
BDR	blind data review
BMI	body mass index
BP	blood pressure
CGI-I	Clinical Global Impression - Improvement
CGI-S	Clinical Global Impression - Severity
CS	cumulative sum
CSR	clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
DBP	diastolic blood pressure
ECG	electrocardiogram
EOS	end of study
FAS	full analysis set
ICF	informed consent form
IUD	intrauterine device
LAR	legally authorized representative
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model repeated measures
MMSE	Mini-Mental State Examination
NCS	normalized cumulative sum
NIA-AA	National Institute on Aging and Alzheimer's Association
PAS	Pittsburgh Agitation Scale
PEC	Positive and Negative Syndrome Scale – Excited Component
PK	pharmacokinetic
PRN	as needed (pro re nata)
SAP	statistical analysis plan
SBP	systolic blood pressure
WHO	World Health Organization

1. Introduction

This statistical analysis plan (SAP) describes the planned analyses for BioXcel's Protocol BXCL501-303, entitled "A Multicenter, Randomized, Double-blind, Placebo-controlled, Efficacy and Safety Study of PRN Dosing of BXCL501 Over a 12 Week Period in Subjects with Agitation Associated with Dementia". This SAP is based on Version 2 Amendment 1 of the protocol, dated May 20, 2022.

This SAP is to be interpreted in conjunction with the protocol. Should the SAP and protocol be inconsistent with respect to the planned analyses, the language of the SAP is governing. If the final clinical study report (CSR) contains changes to any planned statistical analyses, the justification for any such differences will be fully documented in the CSR.

The statistical principles applied in the design and planned analyses of this study are consistent with the International Conference on Harmonisation (ICH) guidelines E9 (Statistical Principles for Clinical Trials).

2. Study objectives

2.1. Primary objective

To determine the safety and efficacy of BXCL501 dosing for episodes of agitation associated with dementia when they occur (given as needed [PRN]), for a maximum of 28 doses within a 12-week treatment period.

2.2. Key secondary objective

The key secondary objectives are:

- To determine the earliest time at which an effect on agitation is apparent, as measured by change from baseline in Positive and Negative Syndrome Scale – Excited Component (PEC) total score.

3. Study design and conduct

3.1. Study design

This is a randomized, double-blind, placebo-controlled, parallel group, 3-arm study assessing efficacy, safety, and tolerability of up to 28 doses of BXCL501 in male and female subjects (65 years and older) with acute psychomotor agitation. BioXcel Therapeutics intends to conduct this study in subjects who have a diagnosis of probable Alzheimer's Disease (AD) based on the National Institute on Aging and Alzheimer's Association (NIA-AA) criteria (2018) and explore the efficacy and safety of an "as needed" BXCL501 dosing regimen on agitation in these subjects. Subjects will be dosed PRN with a maximum of 28 doses over a 12-week period. Subjects may only be dosed once per day; Day 84 is the last day a subject may receive a dose of study drug. Once a subject has received 28 doses of BXCL501, they will continue to be followed for the remainder of the 12-week study period.

It is expected that not all episodes of agitation will be able to be assessed due to timing of the episodes and other factors. Every attempt will be made to capture and treat the majority of episodes.

Subjects must reside in a care facility where all study-related procedures and study drug dosing will be performed by trained research staff under the supervision of the Principal Investigator.

3.2. Study population

Individuals with probable AD dementia are eligible for enrollment in this study. Subjects must require at least minimal assistance with activities of daily living (e.g., bathing, dressing and toileting) and reside in a facility where such assistance is available. Subjects, or their legally authorized representative (LAR), if necessary, will sign an informed consent form (ICF) before any study-related procedures are performed.

Upon confirmation of eligibility following Screening and Pre-dose assessments, subjects will be randomized 1:1:1 to receive BXCL501 40 μg, BXCL501 60 μg, or matching placebo film. Randomization will be stratified based on antipsychotic use in the past month.

Exhibit 1 and Exhibit 2 display the study's inclusion and exclusion criteria, respectively.

Exhibit 1. Inclusion criteria

1. Male and female subjects 65 years and older.
2. All subjects must have a diagnosis of Alzheimer's clinical syndrome based on NIA-AA criteria (Jack et al., 2018).
3. Episodes of psychomotor agitation (e.g., kick, bite, flailing) to the point that it impairs social activities, requires staffing, or medical intervention, or impairs ability for functional activities of daily living should be observed if possible 2 weeks prior to Screening.
4. Subjects exhibit behaviors that are congruent with the International Psychogeriatric Association criterion for agitation representing a change from the subject's usual behavior.
5. Subjects who have a score of 15 to 23 on the Mini-Mental State Exam (MMSE) at Screening and at Pre-dose and require minimal assistance with activities of daily living (e.g., bathing, dressing, and toileting).
6. Subjects who read, understand, and provide written informed consent, or who have a LAR to provide consent on their behalf.
7. Subjects who are deemed to be medically appropriate for study participation by the principal investigator supported by a detailed medical history, physical examination, 12-lead ECG, blood chemistry profile, hematology, and urinalysis.
8. Female participants, if of child-bearing potential and sexually active, and male participants, if sexually active with a partner of child-bearing potential, who agree to use a medically acceptable and effective birth control method throughout the study and for one week following the end of the study. Medically acceptable methods of contraception that may be used by the participant and/or his/her partner include abstinence, birth control pills or patches, diaphragm with spermicide, intrauterine device (IUD), condom with foam or spermicide, vaginal spermicidal suppository, surgical sterilization, and progestin implant or injection. Prohibited methods include the rhythm method, withdrawal, condoms alone, or diaphragm.
9. Subjects who are at their current location for at least 14 days before Screening and plan to remain at the same location for the duration of the study.
10. Subjects must be able to self-administer the investigational product.
11. Subjects who are on a stable concomitant medications regimen for the treatment (including off-label agents for the prevention of agitation) of any concurrent conditions for at least 2 weeks prior to the Screening Visit.

(page 1 of 1)

Exhibit 2. Exclusion criteria

1. Subjects with dementia or other memory impairment not due to probable AD, such as mixed or vascular dementia, dementia with Lewy bodies, Parkinson's disease dementia, frontotemporal dementia, substance-induced dementia, HIV-dementia, traumatic brain injury, normal pressure hydrocephalus, or any other specific non-Alzheimer's-type dementia.
2. Clinical diagnosis of probable AD should not be applied when there is evidence of a cerebrovascular incident temporally related to the worsening of cognitive function.
3. Subjects with agitation caused by acute intoxication must be excluded. Positive identification of non-prescription drugs during urine screening excludes the subject.
4. Subjects with significant risk of suicide or homicide per the investigator's assessment, or any patient with an answer of "yes" to item 4 or 5 on the Columbia-Suicide Severity Rating Scale (C-SSRS) must be excluded.
5. Subjects who have hydrocephalus, seizure disorder, or history of significant head trauma, subarachnoid bleeding, brain tumor, encephalopathy, meningitis, or focal neurological findings, with a recent (1 year) large (non-microvascular) stroke who may be considered medically unstable or in recovery must be excluded.
   Note: Subjects with a remote (>5 years) history of stroke may be included, regardless of size/location.
6. Subjects with laboratory or ECG abnormalities (e.g., advanced heart block [second-degree or above atrioventricular block without pacemaker], diagnosis of sick sinus syndrome) considered clinically significant by the investigator or qualified designee and that would have clinical implications for the subject's participation in the study must be excluded.
7. Subjects with serious, unstable, or uncontrolled medical illnesses must be excluded. These include current moderate to severe hepatic impairment, or renal, gastro-enterologic, respiratory, cardiovascular (including ischemic heart disease, congestive heart failure), endocrinologic, or hematologic disease.
8. Subjects who have received an investigational drug within 30 days prior to Screening must be excluded.
9. Subjects who are considered by the investigator, for any reason, to be an unsuitable candidate for receiving dexmedetomidine or who are unable to use the sublingual film must be excluded, e.g., subjects with a history of allergic reactions to dexmedetomidine.
10. Subjects whose agitation is attributed to pain or infection, concomitant medication, environmental conditions, or psychiatric condition other than dementia as determined by the investigator.
11. Subjects with any other condition, which in the judgment of the investigator would prevent them from entering or completing the study, such as recent clinical weight loss, chronic dehydration, or a recent clinically significant infection are excluded.
12. Subjects who are currently suffering from substance abuse.
13. Subjects who have had surgery within 30 days prior to Screening or scheduled surgery during the study period.
14. Subjects who are pregnant or breast feeding.
15. Patients with a potential cause for delirium (relatively recent onset agitation and dementia).

3.3. Study treatment

BXCL501 is a sublingual film comprised of dexmedetomidine, the active pharmaceutical ingredient and the following inactive ingredients: polyethylene oxide, hydroxypropyl cellulose, sucralose, peppermint oil, Emerald Green colorant, and Food, Drug, and Cosmetic Act Blue #1 colorant.

3.4. Schedule of events

Exhibit 3 and Exhibit 4 display the schedule of events and assessments.

Exhibit 3. Schedule of events – screening, pre-dose, and post-dose assessments

Activity	Timepoints									
			Post Dose							
	Screening[1] (≤ 45 days)	Pre-dose[2] -90 min to time 0	30 min (+10 min)	1 hr (+15 min)	2 hr (+30 min)	4 hr (±1 hr)	*Optional*[3] 8 hr (±1 hr)	*Optional*[3] 12 hr (±1 hr)	Day 2 24 hr	Day 3 (+1 day)
Informed consent	X									
Medical history	X									
Inclusion/Exclusion	X	X[4]								
Demographics	X									
Weight	X									X[5]
Height	X									
Mini-Mental State Exam	X	X[4]								X[5]
ADAS-Cog 12	X									X[5]
Physical exam	X									X[5]
Safety laboratory assessments[6]	X									X[5]
Urine pregnancy test	X									
UDS[7]	X									
UTI[8]	X									
Johns Hopkins Fall Risk Assessment	X									
ECG with rhythm strip[9]	X	X			X	X	X	X	X	X[5]
Pulse oximetry		X			X	X	X	X	X	
Resting vital signs[10]	X	X	X	X	X	X	X	X	X	X[5]
Orthostatic vital signs[11]	X	X		X	X	X	X		X	X[5]
Randomization		X								
Agitation behaviors		X			X					
Study drug administration		X								
PAS	X	X[2]			X	X	X	X	X	
PEC	X	X[2]	X	X	X	X	X	X	X	

Exhibit 3. Schedule of events – screening, pre-dose, and post-dose assessments, continued

Activity	Timepoints									
			Post Dose							
	Screening[1] (≤ 45 days)	Pre-dose[2] -90 min to time 0	30 min (+10 min)	1 hr (+15 min)	2 hr (+30 min)	4 hr (±1 hr)	*Optional*[3] 8 hr (±1 hr)	*Optional*[3] 12 hr (±1 hr)	Day 2 24 hr	Day 3 (+1 day)
ACES		X[2]		X	X	X	X	X	X	
CGI-Severity (agitation)	X	X[2]			X	X				
CGI-Improvement (change in agitation)			X	X	X	X	X	X	X	
C-SSRS	X									X[5]
Sublingual (SL) assessment		X	X		X	X			X	X[5]
PK sampling[12]		X		X	X	X	X			
Drug Likability Scales						X				
Likability Questionnaire						X				
Concomitant medications	X	X	X	X	X	X	X	X	X	X and X[5]
Adverse events			X	X	X	X	X	X	X	X and X[5]

Abbreviations: ACES = Agitation-Calmness Evaluation Scale; ADAS-Cog 12 = Alzheimer's Disease Assessment Scale-Cognitive Subscale 12; CGI-I = Clinical Global Impression-Improvement; CGI-S = Clinical Global Impression-Severity; CLIA = Clinical Laboratory Improvement Amendments; C-SSRS = Columbia-Suicide Severity Rating Scale; ECG = electrocardiogram; hr = hour(s); min = minutes; MMSE = Mini-Mental State Exam; PAS = Pittsburgh Agitation Scale; PEC = Positive and Negative Syndrome Scale – Excited Component; PK = pharmacokinetic; SL = sublingual; UDS = urinary drug screen; UTI = urinary tract infection

Notes to the Schedule of Events:

[1] Screening assessments to be performed within 45 days before the first dose of study drug and may be conducted over more than one day. Screening assessments to be performed within 45 days before the first dose of study drug; the assessments may be conducted over more than one day during the Screening period. If a subject does not become agitated within the 45 day window, the subject is considered a screen failure. However, that subject can be rescreened once at the discretion of the investigator. Please note that some screening assessments do not need to be repeated if assessments were conducted on the day of dosing. All Screening and Pre-dose assessments should be completed before the study drug is administered. Antihypertensives or other medications may be held as per protocol on the day of study drug administration at the discretion of the investigator.

[2] Pre-dose assessments will have a window of 90 minutes prior to administration of study drug. If possible, PEC, PAS, ACEs, and CGI-S should be performed within 15 minutes prior to dosing. All post-dose efficacy assessments should be conducted prior to any other assessments at each time point.

[3] The 8 hour and 12 hour post-dose assessments are not to be performed if the subject is asleep for the night.

[4] Review of Inclusion and Exclusion criteria and administration of the MMSE, which is only performed before the first dose of study drug.

[5] Only complete if Day 84 is the last dose of study drug.

[6] Safety laboratory assessments will include clinical chemistry, hematology, and urinalysis. Laboratory samples drawn within 45 days prior to dosing may suffice, with the exception of UDS.

Exhibit 3. Schedule of events – screening, pre-dose, and post-dose assessments, continued

[7] UDS is required at Screening. UDS will be re-collected prior to dosing if more than 21 days have passed since the Screening visit. Urine drug screen will be analyzed by a central laboratory.

[8] In cases where a repeat laboratory tests may be required to confirm eligibility prior to dosing, a local laboratory may be used to obtain repeat test results. .

[9] 12-lead ECGs at Pre-Dose need to be collected, but if unable to be acquired due to agitation it will not constitute a protocol deviation. ECGs collected following treatment are to be performed prior to PK sampling. Triplicate ECGs are to be obtained approximately 1-3 minutes apart.

[10] Resting vital signs (SBP, DBP and HR) will be taken upon having the subject recumbent for 5 min at Screening, at Pre-Dose, and at 1, 2, and 4 hours post-dose. Triplicate measurements to be performed in case of SBP <90 mmHg, DBP <60 mmHg or pulse <60 bpm. Vital signs are to be done prior to drawing PK blood samples.

[11] Orthostatic measurements (SBP, DBP, HR) will be taken upon having the subject stand, with measurements taken upon standing after 1, 3, and 5 minutes, per Centers for Disease Control and Prevention guidelines for elderly at Screening, at Pre-dose, and at 1, 2, 4, 8, and 24 (Day 2) hours post-dose, and on Day 3 post-dose. A change from supine to standing of 20 mmHg SBP or 10 mmHg DBP must be repeated. Temperature and respiratory rate will be recorded when orthostatic measurement is indicated and are not required to be measured at resting vital sign timepoints. Vital signs are to be done prior to drawing PK blood samples.

[12] PK blood samples are to be collected at timepoints including pre-dose, and 1, 2, 4 and 8 hours after the first dose and fifth dose of study drug only. However, if the investigator documents it is not appropriate to collect a sample from the subject (e.g., inability to access a venous site to collect a blood sample due to psychomotor agitation, subject refusal, subject's current physical condition) at any of the scheduled timepoints, this will not result in the ineligibility of the subject's participation, should not result in early termination nor will this be considered a protocol deviation. All PK collections will have a window of ± 10 minutes. All PK sampling will occur only after all other assessments at that timepoint are conducted.

Exhibit 4. Schedule of events – weekly assessments

Activity	Timepoint[1]											
	Day 7	Day 14	Day 21	Day 28	Day 35	Day 42	Day 49	Day 56	Day 63	Day 70	Day 77	Day 84 (EOS)[1]/ET
ECG with rhythm strip[2]	X		X		X		X		X		X	X
Resting vital signs[3]	X	X	X	X	X	X	X	X	X	X	X	X
Orthostatic vital signs[4]	X		X		X		X		X		X	X
ADAS-Cog 12 Living												X
MMSE				X								X
C-SSRS	X	X	X	X	X	X	X	X	X	X	X	X
Physical exam (including body weight)				X				X				X
Safety laboratory assessments	X		X		X		X		X		X	X
Sublingual (SL) assessment	X		X		X		X		X		X	X
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X
Adverse events	X	X	X	X	X	X	X	X	X	X	X	X

Abbreviations: ADAS-Cog 12 = Alzheimer's Disease Assessment Scale-Cognitive Subscale 12; C-SSRS = Columbia-Suicide Severity Rating Scale; ECG = electrocardiogram; EOS = end of study; ET = early termination; MMSE = Mini-Mental State Exam; SL = sublingual

Notes to the Schedule of Events:

[1] All visits have a window of ± 2 days. If a dose is administered on the same day or within the window of a scheduled weekly assessment, all dosing assessments as noted in Table 3.1 will be performed in lieu of the scheduled weekly assessment. If a 24 hour post-dose assessment coincides with a weekly assessment, then the post-dose 24 hour assessment will take precedence. Safety laboratory assessments should be conducted if they are scheduled as part of the weekly visit even though they are not part of the 24 hour post-dose assessments.

[2] If unable to be acquire an ECG due to agitation it will not constitute a protocol deviation. Triplicate ECGs are to be obtained approximately 1-3 minutes apart.

[3] Resting vital signs (SBP, DBP and HR) will be taken upon having the subject recumbent for 5 minutes. Triplicate measurements to be performed in case of SBP <90 mmHg, DBP <60 mmHg, or pulse <60 bpm. Vital signs are to be measured prior to drawing PK blood samples.

[4] Orthostatic measurements (SBP, DBP, HR) will be taken upon having the subject stand, with measurements taken upon standing after 1, 3, and 5 minutes, per Centers for Disease Control and Prevention guidelines for elderly. A change from supine to standing of 20 mmHg SBP or 10 mmHg DBP must be repeated. Temperature and respiratory rate will be recorded when orthostatic measurement is indicated and are not required to be measured at resting vital sign timepoints.

4. Outcome variable definitions

4.1. Screening and baseline characteristics

Screening characteristics: Inclusion/exclusion criteria, demographic characteristics (age, sex, race, ethnicity), height, weight, medical history, prior and concomitant medications, physical examination, resting and orthostatic vital signs, pregnancy test, 12-lead electrocardiogram (ECG), clinical laboratory results, Johns Hopkins Fall Risk Assessment, urine drug screen, Mini-Mental State Examination (MMSE), ADAS-Cog 12, Pittsburgh Agitation Scale (PAS), PEC, Clinical Global Impression–Severity (CGI-S), and Columbia-Suicide Severity Rating Scale (C-SSRS) will be collected at Screening.

Baseline characteristics: Eligibility, resting and orthostatic vital signs, concomitant medications, sublingual assessment, CGI-S, PEC, Agitation and Calmness Evaluation Scale (ACES), and pharmacokinetic sampling will be collected prior to dose administration.

4.2. Efficacy assessments

The effect of study drug will be evaluated using several validated instruments as described below.

4.2.1. Positive and Negative Syndrome Scale – Excited Component (PEC)

Assessment of drug effect on acute agitation will be done using the PEC. The PEC comprises 5 items associated with agitation, each scored from 1 (minimum) to 7 (maximum): poor impulse control, tension, hostility, uncooperativeness, and excitement. The PEC is the sum of these 5 subscales and thus ranges from 5 to 35.

4.2.2. Clinical Global Impression–Severity (CGI-S) and Clinical Global Impression–Improvement (CGI-I)

The CGI-S will be based on the severity of agitation at Screening and pre-dose (immediately prior to start of dosing). The CGI-S scores will assess the severity of illness based on the following scale:

- 0 = not assessed
- 1 = not at all ill
- 2 = borderline mentally ill
- 3 = mildly ill
- 4 = moderately ill
- 5 = markedly ill
- 6 = severely ill
- 7 = among the most extremely ill subjects

Drug response on agitation will be evaluated by the CGI-I. It will be performed at 30 minutes, 1 hour, 2 hours, and 4 hours post-dose. The CGI-I scores range from 1 to 7:

- 0 = not assessed (missing)
- 1 = very much improved
- 2 = much improved
- 3 = minimally improved
- 4 = no change
- 5 = minimally worse
- 6 = much worse
- 7 = very much worse

Both CGI-I and CGI-S will be focused on the severity of agitation rather than the severity of the overall illness of dementia.

4.2.3. Agitation-Calmness Evaluation Scale (ACES)

The ACES is a single-item measure rating overall agitation and sedation from 1 to 9:

- 1 = marked agitation
- 2 = moderate agitation
- 3 = mild agitation
- 4 = normal behavior
- 5 = mild calmness
- 6 = moderate calmness
- 7 = marked calmness
- 8 = deep sleep
- 9 = unarousable

4.2.4. Likert scales

After dosing with the study drug, subjects will assess their preference of the study medication by answering the statements "I like the taste of the medication" and "The medication is acceptable" using a five-level Likert scale:

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Agree strongly

4.2.5. Alzheimer's Disease Assessment Scale-Cognitive Subscale 12

The Alzheimer's Disease Assessment Scale - Cognitive Subscale 12 (ADAS-Cog12) was developed in the 1980s to assess the level of cognitive dysfunction in AD. Although the ADAS-Cog12 was designed for people with AD, the ADAS-Cog12 has also been used as an outcome measure for studies of interventions in people with MCI. It is also used for assessing the efficacy of antidementia treatments. The cognitive subscale of the original ADAS-Cog12

includes tasks that include both subject-completed tests and rater-based assessments. Together these tasks assess the cognitive domains of memory, language, orientation, and praxis.

The ADAS-Cog12 includes the following: Word Recall, Naming Objects and Fingers, Following Commands, Delayed Word Recall, Constructional Praxis, Ideational Praxis, Orientation, Word Recognition, Remembering Test Directions, Spoken Language, Comprehension, and Word Finding Difficulty.

4.2.6. Drug likability

Subjects will also respond to open-ended questions regarding their experience. Additional comments about aftertaste, smell, dissolve time, etc. will be asked as Yes/No questions, with Yes responses prompting an explanation field.

4.2.7. Pittsburgh agitation scale

The Pittsburgh Agitation Scale (PAS) is an instrument based on direct observations of the patient that is developed to monitor the severity of agitation associated with dementia. There are 4 behavior groups observed (using a 0 to 4-point scale) in the patient, Aberrant Vocalization, Motor Agitation, Aggressiveness, Resting Care (Rosen et al, 1994).

4.3. Efficacy endpoints

4.3.1. Primary endpoint

The change from pre-dose in PEC total score at 2 hours post-dose for the first treated episode of agitation constitutes the primary endpoint for this study.

4.3.2. Key secondary endpoints

The key secondary efficacy endpoints are:

- PEC change from pre-dose at 1 hour post-dose of study treatment for the first treated episode of agitation (measures initial efficacy).
- PEC change from pre-dose at 30 minutes post-dose of study treatment for the first treated episode of agitation (measures initial efficacy).

4.3.3. Continued efficacy endpoints

BioXcel is proposing a collection of endpoints that can be considered as part of the continued efficacy analysis. These endpoints can be viewed as follows: 1) those focusing on a single treated episode following the first treated episode of agitation which include the following: PEC change from pre-dose at 2 hours post-dose of study treatment for the last treated episode of agitation, the second treated episode of agitation, and the third treated episode of agitation; 2) summary measures of efficacy across all administrations of BXCL501 which include the following: average PEC change from pre-dose at 2 hours post-dose of study treatment over all episodes of agitation, the percentage of all treated episodes of agitation satisfying the definition of PEC responder, frequency of all treated episodes of agitation, and the percentage of all treated episodes of agitation for which the PEC change from pre-dose at 2 hours post-dose of study treatment is a reduction of at least 5 points; 3) multivariate outcomes of PEC change from pre-dose at 2 hours post-dose of study treatment for all episodes of agitation and severity of agitation pre-dose for all episodes of agitation, and 4) the comparison of efficacy at the last treated episode of agitation and the first treated episode of agitation. These endpoints are discussed in detail below.

For the full group of endpoints, BioXcel is anticipating that there may be no difference between the BXCL501 groups and placebo for some of the endpoints; recognizing that failure to reject the null hypothesis does not presume that it is true. There will also be a group of endpoints where there may be a difference between groups.

For the following group of continued efficacy endpoints, BioXcel is expecting to see a difference between the BXCL501 dose groups and placebo, recognizing that significance levels will be nominal and that trends across all endpoints will be described across the presentations. These continued efficacy endpoints are as follows:

- PEC change from pre-dose at 2 hours post-dose of study treatment for the last treated episode of agitation.
- PEC change from pre-dose at 2 hours post-dose of study treatment for all treated episodes of agitation. This is a multivariate endpoint consisting of the PEC change

from pre-dose at 2 hours post-dose of study treatment for all treated episodes of agitation.

- Percentage of all treated episodes of agitation satisfying the definition of PEC responder, an episode with at least a 40% reduction in PEC total score from pre-dose at 2 hours post-dose of study treatment.
- Percentage of all treated episodes of agitation for which the PEC change from pre-dose at 2 hours post-dose of study treatment is a reduction of at least 5 points (measures continued efficacy).
- Average PEC change from pre-dose at 2 hours post-dose of study treatment over all treated episodes of agitation. This is a single endpoint which is computed based on averaging the change from pre-dose at 2 hours post dose of study treatment across all treated episodes.

Continued efficacy endpoints based on subgroup analyses are proposed for an assessment of efficacy at the second and third doses of BXCL501. These subgroups and endpoints are as follows:

- Subgroup of subjects with at least two treated episodes of agitation with an endpoint of the PEC change from pre-dose at 2 hours post-dose of study treatment for the second episode of agitation.
- Subgroup of subjects with at least three episodes of treated agitation with an endpoint of the PEC change from pre-dose at 2 hours post-dose of study treatment for the third episode of agitation.

Additional continued efficacy endpoints are proposed as part of the assessment of continued efficacy where it is anticipated that there may be no difference between the BXCL501 treated groups and placebo. These endpoints are focused on the severity of agitation over time as

measured by the PEC, the frequency of episodes in the three groups, and the difference in efficacy between the last and first treated episode of agitation:

- Severity of each treated episode of agitation at pre-dose, as measured by PEC (measures continued efficacy). This is a multivariate endpoint consisting of the pre-dose PEC value at all treated episodes of agitation.
- Frequency of treated episodes of agitation (measures continued efficacy). This is a single outcome defined as the number of treated episodes of agitation for each subject.
- PEC change from pre-dose at 2 hours post-dose of study treatment at the last treated episode of agitation – PEC change from pre-dose at 2 hours post-dose of study treatment at the first treated episode of agitation. This is a single outcome.

4.3.4. Secondary efficacy endpoints

The secondary efficacy endpoints are as follows:

- Number of PEC responders (patients who achieve at least a 40% reduction in PEC total score from pre-dose at 2 hours post-dose of study treatment) for the first treated episode of agitation (measures initial efficacy).
- PAS change from pre-dose at 2 hours post-dose of study treatment for the first treated episode of agitation (measures initial efficacy).
- Clinical Global Impression - Severity (CGI-S) change from pre-dose at 2 hours post-dose of study treatment for the first treated episode of agitation (measures initial efficacy).
- Clinical Global Impression - Improvement (CGI-I) at 2 hours post-dose of study treatment for the first treated episode of agitation (measures initial efficacy).
- Number of CGI-I responders (patients who achieve a CGI-I score of 1 or 2 at 2 hours post-dose of study treatment) for the first treated episode of agitation (measures initial efficacy).
- Average PAS change from pre-dose at 2 hours post-dose of study treatment over all treated episodes of agitation (measures continued efficacy).

- Average CGI-S change from pre-dose at 2 hours post-dose of study treatment over all treated episodes of agitation (measures continued efficacy).
- Average CGI-I at 2 hours post-dose of study treatment over all treated episodes of agitation (measures continued efficacy).
- Percentage of all treated episodes of agitation satisfying the definition of CGI-I responder, an episode with a CGI-I score of 1 or 2 at 2 hours post-dose of study treatment (measures continued efficacy).
- Average PEC change from pre-dose at 1-hour post-dose of study treatment over all treated episodes of agitation (measures continued efficacy).
- Average PEC change from pre-dose at 30 minutes post-dose of study treatment over all treated episodes of agitation (measures continued efficacy).
- Change from screening in ADAS-Cog 12 at day 84.
- Change from screening in MMSE at day 84.
- Duration of calming effect as described by the change from pre-dose in Agitation Calmness Evaluation Scale (ACES) score at 2, 4, and 8-hours post-dose of study treatment for the first treated episode of agitation (measures initial efficacy).

The PK analyses will be described in a separate document.

4.4. Safety assessments

Safety will be assessed during the study by the monitoring and recording of adverse events (AEs), clinical laboratory test results (hematology, biochemistry, and urinalysis), vital sign measurements (systolic and diastolic blood pressures, heart rate measured as pulse, respiratory rate, and temperature), ECG, and physical examination findings.

4.4.1. Adverse events

An AE is defined as any untoward medical occurrence in a subject or clinical investigation patient administered a pharmaceutical product that does not necessarily have a causal relation with the product. An AE can therefore be any unfavorable and unintended sign (including a

new, clinically important abnormal laboratory finding), symptom, or disease temporally associated with the product, whether or not it is related to the product.

4.4.2. Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a suicidal ideation rating scale that identifies behaviors and thoughts that are associated with an increased risk of suicidal actions in the future. The C-SSRS Baseline/Screening version will be conducted at Screening. The C-SSRS Since Last Visit version will be conducted 24 hours post-dosing and at discharge.

4.4.3. Laboratory safety assessments

Exhibit 5 presents the laboratory tests that will be collected at the timepoints specified in the schedule of events (Exhibit 3).

Exhibit 5. Laboratory safety assessments

Hematology	Complete blood count (hemoglobin, hematocrit, white blood cell count with differential, red blood cell count, and platelet count)
Serum chemistry	Includes blood urea nitrogen, creatinine, total bilirubin, alkaline phosphatase, aspartate aminotransferase (serum glutamic-oxaloacetic transaminase), alanine aminotransferase (serum glutamic pyruvic transaminase), glucose, albumin, total protein, and electrolytes (sodium, chloride, potassium, and bicarbonate)
Urinalysis	Includes pH, specific gravity, protein, glucose, ketones, bilirubin, blood, nitrites, leukocytes, urobilinogen, microscopic urine analysis if dipstick positive
Urine pregnancy test*	Women of child-bearing potential only
Urine drug screen*	Cocaine, amphetamine, phencyclidine, benzodiazepines, marijuana. (Note: marijuana positive is allowed provided subject is not moderately to severely dependent, benzodiazepine positive are allowed if prescribed.)

*Conducted at Screening only

(page 1 of 1)

4.4.4. Vital signs

Resting vital signs including systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate (measured as pulse) will be measured after the subject has been in a recumbent position for at least five minutes at the timepoints specified in the schedule of events (Exhibit 3). Measurements should be made at least one minute apart using the same arm at each visit.

At indicated timepoints, orthostatic measurement of SBP, DBP, and heart rate will be measured after the subject has been standing for a total of five minutes. Temperature and respiratory rate will be recorded when orthostatic measurement is indicated in the schedule of events and are not required to be measured at resting vital sign timepoints.

If the first measurement of vital signs (SBP, DBP, and pulse) shows SBP <90 mmHg, DBP <60 mmHg, or pulse <60 bpm, vital signs will be measured again in triplicate (same arm, separated by at least 1 minute).

4.4.5. Electrocardiogram (ECG)

A 12-lead ECG with rhythm strip will be performed at Screening, pre-dose (not required if Screening ECG is conducted on the day of dosing), 2 hours, 4 hours, and 24 hours post-dose.

4.4.6. Physical examinations

A standard physical examination will be performed at Screening. The examination will include assessment of skin, head, ears, eyes, nose, throat, neck, thyroid, lungs, heart, cardiovascular, abdomen, lymph nodes, and musculoskeletal system/extremities. Interim physical examinations will be performed at the investigator's discretion, if necessary, to evaluate AEs or clinical laboratory abnormalities.

Height and weight will be measured at Screening, and weight will be measured again on the day of discharge.

4.4.7. Concomitant medications

Concomitant medications will be reviewed and documented during the study.

5. Database lock and unblinding

5.1. Blind review of selected data prior to final database lock

The clinical database will undergo a "soft lock" when (1) all patients have completed their Day 84 End of Study visit and (2) the database is nearly clean. After the soft data lock and prior to accessing the randomized treatment assignments, the following series of tables, listings, and figures will be generated without treatment assignments:

- A listing of all efficacy outcome values
- A listing of all subjects that indicates their inclusion in, or the reasons for their exclusion from, each analysis population
- A listing of unanticipated cases of study drug exposure (e.g., placebo for patients who were randomized to BXCL501)
- Resolution of other last-minute ambiguities related to the primary and secondary efficacy outcomes

These blind data review (BDR) documents will provide the basis for a data review in order to finalize the primary and secondary outcomes and the analysis populations. BioXcel will document approval of the BDR minutes prior to the final database lock.

5.2. Final database lock

After completion of all BDR procedures, validation of the project databases, and BioXcel's approval of the BDR, the clinical database will be locked. After the database lock and the authorization for unblinding, the treatment codes will be merged to the analysis datasets. Any change to the clinical database after this time will require written authorization, with explanation, by BioXcel.

5.3. Authorization for unblinding

After the database lock and upon receipt of written authorization from BioXcel, the (previously blinded) study team will receive the actual treatment codes. This team will generate top-line results and provide them to BioXcel. BioXcel will not have direct access to the randomized treatment codes until they have been provided the top-line results and datasets.

6. Statistical analyses

Statistical analyses will be performed using SAS® software version 9.4 or higher.

6.1. Statistical methodology

6.1.1. Sample size determination

Approximately 150 subjects are anticipated to be enrolled. Assuming a standard deviation of 4 in change from baseline in the PEC at two hours post-dose and a two-sided two-sample t-test, 150 subjects provide in excess of 90% power to detect a difference in mean PEC score of 2.9 units or more, at the 0.025 alpha-level (adjusted for multiplicity of doses) for either dose versus placebo.

6.1.2. Blinded sample size recalculation

A blinded sample size recalculation is currently planned when approximately 50% of the PEC total score data at 2 hours for the first treated episode are available. The standard deviation of the change from baseline in PEC total score will be computed at the 30 minute, 1 hour, and 2 hour timepoints. The standard deviations will be computed in the full dataset and not by treatment group to ensure that the blinding is maintained. The sample size will then be recalculated with the potential to adjust the sample size accordingly.

A small committee consisting of select BioXcel leadership and an independent blinded statistician will evaluate the need for an increase in sample size. The independent blinded statistician will use the standard deviation values estimated from the full dataset to recalculate

the sample size. In addition to these pooled standard deviation values, the committee will also assess data completeness for the primary endpoint.

The committee will meet and document its decision in a memorandum. This memorandum will include the pooled standard deviations, the updated power/sample size computations, and the final decision regarding any increase in sample size.

6.1.3. Analysis populations

The following analysis populations are planned:

- Safety Population: All subjects receiving at least 1 dose of study drug, with subjects classified according to the drug actually received.
- Full Analysis Set: All patients who take any study medication and who had both baseline and at least one efficacy assessment after dosing.

6.1.4. General considerations

In general, summary statistics (n, mean, standard deviation, median, minimum, and maximum values for continuous variables, and number and percentage of subjects in each category for categorical variables) will be provided by treatment group for evaluated variables.

All statistical tests and confidence intervals, unless stated otherwise, are 2-sided and will be set at α=0.05.

6.1.5. Procedures for handling missing data

Multiple imputation will be used to handle missing data. Details are provided in Section 11.3.

6.2. Demographics and baseline characteristics

The following demographics and baseline characteristics will be presented in summary tables by treatment for the safety population: demographic characteristics (age, sex, race, ethnicity), weight, height, body mass index (BMI), medical history, prior medications, laboratory examinations, vital signs, and ECG.

6.3. Subject disposition

Subject disposition will include the number of subjects who enroll in the study and the number and percentage of subjects included in each analysis population by treatment. The frequency and percentage of subjects who withdraw or discontinue from the study, along with the reason for withdrawal or discontinuation, will be summarized by treatment.

6.4. Efficacy analyses

Efficacy analyses will use data from the full analysis set (FAS) as the target population. Change from pre-dose in the PEC at 2 hours post-dose for the first treated agitation episode will be considered primary.

Efficacy analyses for measures assessed at specific scheduled time points will evaluate the change from baseline using a mixed model repeated measures (MMRM) approach.

6.4.1. Primary analysis

The change from pre-dose in PEC total score at 2 hours post-dose for the first treated episode of agitation constitutes the primary endpoint for this study. Table 1 provides the information for the four estimand attributes needed for the primary efficacy analysis.

Table 1. Estimand attributes for the primary efficacy analysis

Estimand attribute	Description of attribute
Target population	FAS population - All patients who take any study medication and who had both baseline and at least one efficacy assessment after dosing.
Primary endpoint	Change from pre-dose in PEC total score at 2 hours post-dose for the first treated episode of agitation.
Handling of intercurrent events	All values collected after the use of rescue treatment and withdrawal from study will be used in the analysis (treatment policy strategy).
Population-level summary	Difference between dose-specific treatment arm and placebo arm in mean change from pre-dose in the total PEC score at 2 hours using all measurements from baseline through 4 hours.

(page 1 of 1)

The null and alternative hypotheses to be tested for the primary endpoint are stated as H_{01}: $\Delta_{BXCL501\_60} = \Delta_{PBO}$ and H_{A1}: $\Delta_{BXCL501\_60} \neq \Delta_{PBO}$ and H_{02}: $\Delta_{BXCL501\_40} = \Delta_{PBO}$ and H_{A2}: $\Delta_{BXCL501\_40} \neq \Delta_{PBO}$, where $\Delta_{BXCL501\_60}$ denotes the change from pre-dose in the PEC at 2 hours post-dose in the BXCL501 60 μg group, $\Delta_{BXCL501\_40}$ denotes the change from pre-dose in the PEC at 2 hours post-dose in the BXCL501 40 μg group, and Δ_{PBO} denotes the change from pre-dose in the PEC at 2 hours post-dose in the placebo group. These hypotheses will be tested using a MMRM model. To account for the testing of two hypotheses, the two-sided significance level for each test will be determined using the Bonferroni correction with each hypothesis tested at a significance level of 0.025.

The MMRM will include the change from pre-dose in the PEC at 30 minutes, 1 hour, 2 hours, and 4 hours post-dose as the outcome and the baseline PEC, treatment group, antipsychotic use (stratification factor), visit number, and a treatment group by visit number interaction term as covariates. In all prior studies of BXCL501 there have been few missing data points in the two-hour assessment window; however, methods to address missing data are proposed as part of the sensitivity analyses.

The model will be fit using an unstructured covariance matrix with the Kenward-Roger adjustment for the degrees of freedom. As the change from pre-dose in PEC score is only measured at 30 minutes, 1 hour, 2 hours, and 4 hours the covariance structure includes a total of 10 parameters making the likelihood that the model fails to converge low.

In the event that the model fails to converge with an unstructured covariance matrix, the following covariance structures will be applied, in order, until convergence is reached: heterogenous Toeplitz structure, Toeplitz structure, auto-regressive of order 1 (AR(1)), and compound symmetry structure.

1. Heterogenous Toeplitz covariance structure (assuming different variances at each time point and that measurements taken closer together in time are more highly correlated than those taken farther apart).

2. Toeplitz covariance structure (assuming measurements taken closer together in time are more highly correlated than those taken farther apart).
3. First order auto-regressive (AR[1]) covariance structure (assuming measurements taken closer together in time are more highly correlated than those taken farther apart, but more constrained than the Toeplitz structure).
4. Compound symmetry covariance structure (assuming equal correlation for measurements from a patient, regardless of how far apart in time when they were taken).

For the alternative covariance structures (heterogeneous Toeplitz, Toeplitz, first order auto-regressive, and compound symmetry) the sandwich estimator will be used to obtain the estimates of the standard error and the Kenward-Roger adjustment for degrees of freedom will not be applicable. If convergence proves to be a problem and an alternative covariance structure is needed there may be an inflation of type I error. To address this, the methods proposed in Lu and Mehrotra (2009) will be used as well with the Fisher scoring algorithm being used to obtain the initial values of the covariance parameters.

The estimated mean difference between the groups treated with either 40 μg BXCL501 or 60 μg BXCL501 and the placebo group in change from pre-dose in the PEC score will be estimated at 2 hours. The details of the SAS code for this analysis are provided in the Appendix of this SAP (Section 11). The SAS code presented in this SAP may be modified based on test data from the trial when a separate set of programming specifications will be developed. Results will be presented in a table and a figure.

There was little missing data in prior Phase 3 studies of BXCL501, and the expectation is that missing data will likely not be an issue in this study. However, methods to account for missing

data may be needed in this study, so a plan for the handling of missing data is pre-specified in this SAP using multiple imputation. The following sensitivity analyses will be performed:

- Missing at random multiple imputation, in which imputed values are based on the treatment group to which the subject is randomized will be used to estimate the treatment effect.
- Control-based imputation, in which imputed values are obtained from the placebo data rather than the treatment group to which the subject is randomized will be used to estimate the treatment effect.
- MMRM with change from pre-dose at 30 minutes, 1 hour, and 2 hours as the outcome will be fit with an unstructured covariance as part of the sensitivity analyses. The model will include the same covariates and approach as that for the primary analysis; baseline PEC, treatment group, antipsychotic use (stratification factor), visit number, and a treatment group by visit number interaction term.

Monotone missing imputation will be used as the method of choice with 100 imputed datasets derived for the missing at random approach and missing not at random imputation will be used for the control-based approach. All results will be summarized in tables and a forest plot. Details are provided in the Appendix of this SAP (Section 11.3) to be finalized before the blinded data review. This code will need to be vetted on a blinded test set of data from the trial during the development of the programming specifications for this SAP.

6.4.2. Key secondary analyses

The key secondary efficacy endpoints are:

- PEC change from pre-dose at 1 hour post-dose of study treatment for the first treated episode of agitation (measures initial efficacy)
- PEC change from pre-dose at 30 minutes post-dose of study treatment for the first treated episode of agitation (measures initial efficacy)

The attributes for the estimands of the key secondary endpoints of PEC change from pre-dose at 1 hour post-dose for the first episode of agitation and the PEC change from pre-dose at 30 minutes post-dose for the first episode of agitation are provided in Table 2.

Table 2. Estimand attributes for the key secondary analyses of change from pre-dose in total PEC score at 1 hour and at 30 minutes for the first treated episode of agitation

Estimand attribute	Description of attribute
Target population	FAS population - All patients who take any study medication and who had both baseline and at least one dose of study drug.
Primary endpoint	PEC change from pre-dose at 1 hour or at 30 minutes.
Handling of intercurrent events	All values collected after the use of rescue treatment and withdrawal from study will be used in the analysis (treatment policy strategy).
Population-level summary	Difference between dose-specific treatment arm and placebo arm in mean change from pre-dose in the total PEC score at 1 hour or at 30 minutes using all measurements from baseline through 4 hours.

(page 1 of 1)

The null and alternative hypotheses to be tested for the first two key secondary endpoints listed above are similar to those for the primary endpoint, with the timepoint for testing being 1 hour post-dose for the first key secondary endpoint and 30 minutes post-dose for the second key secondary endpoint. These hypotheses will be tested using the MMRM model described for the primary analysis.

The results for the key secondary endpoints will be presented in tabular form and line plots for each group over time. The endpoints, in the order to be tested, are as follows:

1. PEC change from baseline to 1 hour for the first treated episode of agitation.
2. PEC change from baseline to 30 minutes for the first treated episode of agitation.

The fixed-sequence method will be used to adjust for multiplicity and applied to each dose separately. If at any point in the testing, the significance level is greater than 0.025 for the

comparison of interest, then all testing ceases for that dose level and the remaining p-values are reported as nominal levels.

6.4.3. Continued efficacy analyses

The continued efficacy endpoints and a description of the proposed analysis approach are outlined below.

- PEC change from pre-dose at 2 hours post-dose of study treatment for the last treated episode of agitation. This endpoint will be analyzed using the MMRM model that is used for the primary analysis.
- PEC change from pre-dose at 2 hours post-dose of study treatment for all treated episodes of agitation. This is a multivariate endpoint consisting of the PEC change from pre-dose at 2 hours post-dose of study treatment for all treated episodes of agitation. The null hypothesis being tested is that there is no decline in treatment effect over all episodes of agitation. The model has been parametrized with the PEC change from pre-dose at 2 hours post-dose for each treated episode of agitation as the outcome and the following covariates: pre-dose PEC value for the treated episode, antipsychotic stratification variable, number of episodes, time (days) of the treated episode, trt_40 which is an indicator variable that is equal to 1 for all subjects in the BXCL 40 μg group and 0 otherwise, trt_60 which is an indicator variable that is equal to 1 for all subjects in the BXCL 60 μg group and 0 otherwise, and interaction term of time and trt_40, and an interaction term of time and trt_60. The coefficient associated with each interaction term will be tested separately. The estimates of the slope for each group are obtained from the corresponding estimate statements in Table 3. The code will be vetted on blinded data and modified as needed. Note that it is anticipated that the slopes may not be different from zero making this a supporting analysis.

  Additional mixed models will be fit to the data that include a quadratic term for time.

Table 3. SAS code for longitudinal assessment of change in PEC

```
proc mixed data=new method=reml covtest asycov asycorr ic;
 class PatientID;
 model chg= base no episode time dose trt_40 trt_60
 time*trt_40 time*trt_60 anti_psy_strata
 /ddfm=kr solution cl;
 random intercept/subject=patientid type=un gcorr;

 estimate "40 µg" INTERCEPT 0 TIME 1 TRT_40 0 TRT_60 0 TIME*TRT_40
       1 TIME*TRT_60 0 /cl;
 estimate "60 µg" INTERCEPT 0 TIME 1 TRT_40 0 TRT_60 0 TIME*TRT_40
       0 TIME*TRT_60 1 /cl;
 estimate "placebo" INTERCEPT 0 TIME 1 TRT_40 0 TRT_60 0
       TIME*TRT_40 0 TIME*TRTG_60 0/cl;
run;
```

- Percentage of all treated episodes of agitation satisfying the definition of PEC responder, an episode with at least a 40% reduction in PEC total score from pre-dose at 2 hours post-dose of study treatment will be analyzed using an ANCOVA model with percentage of episodes meeting the definition of PEC responder as the outcome, and treatment group and antipsychotic use (stratification factor) as covariates. A weighted least squares approach may also be used to account for the differing number of episodes across subjects.
- Percentage of all treated episodes of agitation for which the PEC change from pre-dose at 2 hours post-dose of study treatment is a reduction of at least 5 points (measures continued efficacy) will be analyzed using an ANCOVA model with percentage of episodes meeting the definition of PEC responder as the outcome, and treatment group and antipsychotic use (stratification factor) as covariates. A weighted least squares approach may also be used to account for the differing number of episodes across subjects.
- Average PEC change from pre-dose at 2 hours post-dose of study treatment over all treated episodes of agitation. This is a single endpoint which is computed based on the average of the change from pre-dose at 2 hours post dose of study treatment

across all treated episodes. An ANCOVA model will be used for this analysis. The model will include the average change from pre-dose in PEC score at 2 hours over all episodes as the outcome, and treatment group and antipsychotic use (stratification factor) as covariates. A weighted least squares approach may also be used to account for the differing number of episodes across subjects.

- Subgroup of subjects with at least two treated episodes of agitation with an endpoint of the PEC change from pre-dose at 2 hours post-dose of study treatment for the second episode of agitation. For this analysis, the MMRM model described for the primary analysis will be used.
- Subgroup of subjects with at least three episodes of treated agitation with an endpoint of the PEC change from pre-dose at 2 hours post-dose of study treatment for the third episode of agitation. For this analysis, the MMRM model described for the primary analysis will be used.
- Severity of each treated episode of agitation at pre-dose, as measured by PEC (measures continued efficacy). This is a multivariate endpoint and will be analyzed using a longitudinal mixed model with the outcome being the pre-dose PEC value at each episode. See Section 11.6 for SAS code. Time is defined as the time of the episode in study day (ranging from 1 to 84 days). The model will include an intercept as a random effect, treatment group as a class variable, antipsychotic use (stratification factor), and a time by treatment group interaction term as fixed effects. The model will be fit with no fixed effect intercept term. Under this model the null hypothesis that the slope measuring the rate of change of the pre-dose PEC score across all treated episodes is zero will be tested. This hypothesis is stated as H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ where β_1, β_2, and β_3 denote the coefficients associated with the time by treatment group interaction term for the placebo, BXCL501 40 μg, and BXCL501 60 μg groups, respectively. The significance level for this test is obtained from the type III test of the fixed effect of the interaction term of time by treatment group. Should this hypothesis be rejected, then the significance level associated with each of the parameters, β_1, β_2, and β_3 would be used to test the null hypothesis that the slope

for a single parameter is zero. If the significance level is greater than 0.05, then the null hypothesis will fail to be rejected supporting no change over time in the pre-dose PEC scores.

- Frequency of treated episodes of agitation (measures continued efficacy). This is a single outcome defined as the number of treated episodes of agitation for each subject. It will be assessed with a Poisson model that includes treatment group and antipsychotic use (stratification factor) as covariates. The results will be summarized as an incidence rate ratio and the significance level for the test of the null hypothesis of no difference between each dose level and placebo will be obtained from the significance level for the test of each dose level versus placebo. The pseudo SAS code for this analysis is presented in Table 12 in the Appendix.
- PEC change from pre-dose at 2 hours post-dose of study treatment at the last treated episode of agitation –PEC change from pre-dose at 2 hours post-dose of study treatment at the first treated episode of agitation. This is a single outcome. The goal of the proposed analysis is to determine if there is a change in treatment effect when comparing the first and last treated episodes. It is anticipated that there will not be a large difference between the treatment effect at the first and last treated episode for the groups receiving either BXCL501 40 μg or BXCL501 60 μg. The same result is anticipated for the placebo group as well. Thus, it is likely that this will provide a larger significance level with a result of failure to reject the null hypothesis. While this result does not demonstrate that the outcome is the same at the first and last treated episodes, this analysis will contribute to the body of evidence supporting continued efficacy.

  The proposed analysis is based on a multiple linear regression model using the following endpoint

  ΔPEC at the last treated episode - ΔPEC at the first treated episode

where ΔPEC denotes the change from baseline in PEC at hour 2. The multiple linear regression model also includes the following covariates:

- Baseline PEC value at the first treated episode of agitation
- Anti_psy_strata – stratification factor for antipsychotic use
- number of treated episodes
- treatment group (placebo, BXCL501 40 μg, BXCL501 60 μg)
- time (days) of the last treated episode.

The following three hypotheses will be tested:

H_{01}:(ΔPEC at the last treated episode - ΔPEC at the first treated episode)$_{BXCL501\_40,}$

H_{02}:(ΔPEC at the last treated episode - ΔPEC at the first treated episode)$_{BXCL501\_P60,}$

and

H_{03}:(ΔPEC at the last treated episode - ΔPEC at the first treated episode)$_{BXCL501\_PBO.}$

Here (ΔPEC at the last treated episode - ΔPEC at the first treated episode)$_{BXCL501\_60}$ denotes the difference in ΔPEC between the last and first treated episodes in the BXCL 60 μg group, (ΔPEC at the last treated episode - ΔPEC at the first treated episode)$_{BXCL501\_40}$ denotes the difference in ΔPEC between last and first treated episodes in the BXCL 40 μg group, and (ΔPEC at the last treated episode - ΔPEC at the first treated episode)$_{BXCL501\_PBO}$ denotes the difference in ΔPEC between the last and first treated episode in the placebo group.

Note that this model does not assume a linear trajectory in ΔPEC over time and focuses only on the difference in ΔPEC at the first and last treated episodes. Additionally, the model is a multiple linear regression model and not a longitudinal mixed model. The only inclusion of a time element is the covariate of time (days) of the last treated episode of agitation.

Table 4 provides example SAS code for this comparison. In this code trt1 and trt2 are indicator variables for the BXCL501 40 µg and BXCL501 60 µg groups, respectively. The variable trt_40 is defined as 1 if the subject is in the BXCL501 40 µg group and 0 otherwise, while trt_60 is defined as 1 if the subject is in the BXCL501 60 µg group and 0 otherwise. The variable base is the baseline PEC score at the first treated episode, no_episode denotes the number of treated episodes, and time denotes the time (days) of the last treated episode. The estimate statements correspond to the statistical hypotheses described above.

Table 4. Example SAS code for comparison of first and last treated episode

```
proc glm data=new2;
      model diff= anti_psy_strata trt_40 trt_60 no_episode base time;
      estimate '40 µg' intercept 1 trt_40 1 trt_60 0;
      estimate '60 µg' intercept 1 trt_40 0 trt_60 1;
      estimate 'Placebo' intercept 1 trt_40 0 trt_60 0;
run;
```

6.4.4. Secondary efficacy analyses

The methods that will be used to analyze each outcome will be as follows:

- The number of PEC responders for the first treated episode of agitation will be analyzed using a stratified Mantel-Haenszel test stratified by antipsychotic use (stratification factor).
- The change from baseline in PAS and CGI-S at 2 hours and the CGI-I at 2 hours post-dose for the first episode of agitation will be analyzed using an ANCOVA model with the change from baseline at 2 hours post-dose as the outcome, and baseline value of the outcome (used only for PAS and CGI-S), antipsychotic use (stratification factor), and treatment group as covariates.
- The number of CGI-I responders (patients who achieve a CGI-I score of 1 or 2 at 2 hours post-dose of study treatment) for the first episode of agitation will be

analyzed using a stratified Mantel-Haenszel test, stratifying according to the stratification factor of antipsychotic use.

- The average change in PAS and CGI-S at 2 hours post-dose over all treated episodes of agitation will be analyzed using an ANCOVA model with the average values of change in PAS or CGI-S across all episodes as the outcome, and treatment group and antipsychotic use (stratification factor) as covariates.
- The average CGI-I at 2 hours post-dose over all treated episodes of agitation will be analyzed using an ANCOVA model with the average CGI-I across all episodes as the outcome, treatment group and antipsychotic use (stratification factor) as covariates.
- The percentage of all treated episodes of agitation satisfying the definition of CGI-I responder will be analyzed using an ANCOVA model with treatment group and antipsychotic use (stratification factor) as covariates.
- Average PEC change from pre-dose at 1-hour post-dose of study treatment over all treated episodes of agitation will be analyzed using an ANCOVA model with treatment group and antipsychotic use (stratification factor) as covariates.
- Average PEC change from pre-dose at 30 minutes post-dose of study treatment over all treated episodes of agitation will be analyzed using an ANCOVA model with treatment group and antipsychotic use (stratification factor) as covariates.
- Change from screening in ADAS-Cog 12 at day 84 will be analyzed using a two-sample t-test.
- The change from baseline in ACES scores at 2, 4, and 8 hours post-dose for the first episode of agitation, will be analyzed using an ANCOVA model with the change from baseline in ACES score as the outcome, and the baseline value of the outcome, antipsychotic use (stratification factor), and treatment group as covariates.
- Change from screening in MMSE at Day 28 and at Day 84 will be analyzed using a two-sample t-test.

6.4.5. Additional analyses

Figures will be created to supplement the primary, key secondary, secondary, and exploratory analyses. These may include plots of mean values over time either in the form of boxplots or mean values. Empirical distribution functions of the change from baseline to 2 hours in the PEC score may be presented for categories of the CGI-I and CGI-S. As appropriate, additional figures may be created to supplement any of the analyses.

7. Safety and Tolerability analyses

Safety data analysis will be conducted on all subjects receiving at least 1 dose of study drug. The number and percentage of subjects experiencing 1 or more AEs will be summarized by treatment, relationship to study drug, and severity. AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. Laboratory parameters will be summarized by treatment using descriptive statistics and data listings of clinically significant abnormalities. Vital signs and ECG data will be summarized by changes from baseline values using descriptive statistics.

Orthostatic assessments will follow the Centers for Disease Control and Prevention guidelines for the elderly (e.g., BP upon standing for 1, 3, and 5 minutes) (https://www.cdc.gov/steadi/pdf/Measuring_Orthostatic_Blood_Pressure-print.pdf).

Results will be presented for first dose of study drug and for all doses of study drug.

7.1. Adverse events

AEs will be coded using the MedDRA version 22.0 or later coding system. Frequency tables will be presented by treatment groups summarizing:

- All treatment-emergent AEs
- All treatment-emergent AEs by severity
- All treatment-emergent treatment-related AEs
- All serious treatment-emergent AEs
- All AEs leading to discontinuation

A set of AE tables will be created for reporting of adverse events related to the first administration of BXCL501. A separate set of AE tables will provide information for all adverse events recorded across all administrations of BXCL501.

Additional tables specific to the reporting of the following adverse events will also be created: orthostatic hypotension, hypotension, bradycardia, sinus bradycardia, and somnolence (rated moderate by the investigator). These tables will present information for the first administration and summarized across all administrations of BXCL501.

7.2. Clinical laboratory evaluations

Each laboratory value and change from baseline (when appropriate) will be summarized in tables and figures for hematology, blood chemistry, and urinalysis at Day 7, Day 21, Day 35, Day 49, Day 63, Day 77, and Day 84. All abnormal clinical laboratory evaluations will be reported in separate presentations for the hematology, serum chemistries and urinalysis results.

7.3. Physical examination

A listing of physical examination findings will be provided by subject for Day 28, Day 56, and Day 84.

7.4. Vital signs

Each resting vital sign observed value and change from baseline (when appropriate) will be summarized for the first dose and for each treatment group at pre-dose, 30 minutes, 1 hour, 2 hours, and 4 hours. Optional measurements may be made at 8 hours and 12 hours. Orthostatic vital signs (as outlined in the footnotes of Exhibit 3) will be summarized for each treatment group at pre-dose, 1 hour, 4 hours, Day 2, and Day 3. Additional measurements of vital signs are collected weekly starting at Day 7 through Day 84. Orthostatic vital signs are also collected at Day 7, Day 21, Day 35, Day 49, Day 63, Day 77, and Day 84. The observed value and change from baseline (when appropriate) will be summarized for each treatment group as collected. Optional measurements of vital signs will be provided in a listing.

Results will be summarized across all doses of study treatment using descriptive statistics. Abnormal results will be presented in an additional table with rates presented.

7.5. 12-lead electrocardiogram

Each 12-lead ECG observed value and change from baseline (PR, QRS, QT) will be summarized for each treatment group at Screening, pre-dose (not required if Screening ECG is conducted on the day of dosing), 2 hours, 4 hours, Day 2, and Day 3 for the first dose and at Day 7, Day 21, Day 35, Day 49, Day 63, Day 77, and Day 84. Results will also be summarized across all doses of study drug. Optional measures may be taken at 8 hours and 12 hours and will be provided in a listing. Frequency tabulation of the overall ECG results (normal, abnormal not clinically significant [NCS], and abnormal clinically significant[CS]) will be summarized (to include any emergent arrhythmias and determination of resolution). These results will be summarized for the first dose and across all doses. Conduction intervals including PR, QRS, QT (>450, >500ms) and QTc (>450, >500 ms) will be summarized and tabulated.

7.6. Concomitant medications

Concomitant medications will be summarized (n and %) by anatomical therapeutic chemical (ATC) class and preferred term (coded by World Health Organization [WHO] Drug coding dictionary March 2019 or later version) for each treatment group. Use of concomitant medications will be summarized separately for the first dose of study drug and all doses of study drug.

Rescue medications will be summarized in separate presentations using the same approach as that used for the concomitant medications.

7.7. Sublingual exam for local irritation

The local irritation will be assessed by sublingual exam at 30 minutes, 2 hours, 4 hours, and 24 hours post-dose based on following parameters:

- Negative reaction to the sublingual film in the examiner's opinion (Yes/No): assessed at all above time points.
- Time taken for medication to dissolve (1–30 seconds, 31–59 seconds, 1–2 minutes, 3+ minutes): only assessed at 30 minutes.

A frequency summary will be presented for these parameters for each treatment group at each timepoint for the first dose. A summary of negative reactions and time taken for medication to dissolve will also be presented across all doses of study drug.

8. Additional assessments

The drug likability scale and likability questionnaire will be summarized descriptively for the first dose of BXCL501. Overall summary measures will be provided across all doses of BXCL501 will also be computed.

The C-SSRS will be summarized at each time point including the responses to the individual items belonging to the suicidal ideation and suicidal behavior subscales.

Agitation behaviors, which are collected pre-dose and at 2 hours post-dose, will be summarized for each dose administration.

The MMSE will be administered at Screening, at the time of the first dose of BXCL501 or placebo, at Day 28, and at Day 84. The results will be summarized over time including the mean values at each timepoint and the change from baseline.

The ADAS-Cog12 is administered at Screening and at Day 84. The results will be summarized descriptively at Screening and at Day 84. The change from baseline to Day 84 will also be summarized descriptively.

9. Pharmacokinetic analyses

The pharmacokinetic analyses will be described in a separate SAP.

10. References

1. Jack CR, Bennett DA, Blennow K, et al. 2018 National Institute on Aging-Alzheimer's Association (NIA-AA) Research Framework NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease: Alzheimer's & Dementia 14 (2018) 535-562. https://doi.org/10.1016/j.jalz.2018.02.018

2. Oquendo MA, Halberstam B, Mann JJ. Colombia Suicide Severity Rating Scale (C-SSRS) – Risk Factors for Suicidal Behavior: The Utility and Limitations of Research Instruments. In: First MB, ed. *Standardized Evaluation in Clinical Practice*. Washington, DC: American Psychiatric Publishing, 103–131, 2003.

3. Rosen J, Burgio L, Kollar M, Cain M, Allison M, Fogleman M, Michael M, Zubenko GS. The Pittsburgh Agitation Scale: A User-Friendly Instrument for Rating Agitation in Dementia Patients, The American Journal of Geriatric Psychiatry, Volume 2, Issue 1, 1994, Pages 52-59.

4. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Statistical principles for clinical trials (E9). International Conference on Harmonisation; 1998.

11. Appendix

This appendix contains details for programming the MMRM and the techniques to be applied to account for missing data. This code is pseudocode and will need to be tested on a blinded data transfer before being finalized. The code presented here will be finalized before the final database lock as part of the programming specifications but may be subject to change depending on the stability of the imputation and/or analysis models.

11.1. MMRM analysis

Table 5 contains pseudo SAS code with the variables defined as follows:

- DOSE – treatment group assignment with dose of 40µg and 60µg
- VISITNUM – denotes time of measurement (30, 60, 120, and 240 minutes)
- ANTI_PSY_STRATA – indicates antipsychotic use in the past month stratum for randomization
- CHG_PEC – change from baseline in PEC total score
- VISITNUM*DOSE – interaction of visit and dose group
- BASE_PEC – baseline value of the PEC total score

The LSMEANS statement will provide information on all possible differences for VISITNUM*DOSE. The ESTIMATE statements provide information for the specific comparison and should be checked against the LSMEANS output to ensure that the correct values were extracted.

The code in Table 5 will be used for all MMRM analyses proposed in this SAP with appropriate modifications made for the variable names and number of repeated measures.

Table 5. SAS code for MMRM analysis

```
proc mixed data=OUT2 method=reml alpha=0.025 covtest;
   class  VISITNUM(ref=LAST)
          DOSE(ref = "0") PATNO ANTI_PSY_STRATA;
   model CHG_PEC = BASE_PEC ANTI_PSY_STRTA DOSE VISITNUM
                   VISITNUM*DOSE
          / solution ddfm=kr;
   repeated VISITNUM / subject=PATNO type=un r;
   lsmeans VISITNUM*DOSE / diff om cl e;
   estimate '60 - PBO at 120 minutes' DOSE 0 1 -1
   VISITNUM*DOSE 0 0 0 0 0 0 0 1 -1 0 0 0
         / CL;
   estimate '40 - PBO at 120 minutes' DOSE 1 0 -1
   VISITNUM*DOSE 0 0 0 0 0 0 1 0 -1 0 0 0
         / CL;
   ods output estimates=SI_ESTIMATES;
run;
```

(page 1 of 1)

11.2. ANCOVA analysis

Table 6 contains pseudo SAS code for the ANCOVA analyses proposed in this SAP with appropriate modifications made for the variable names.

Table 6. SAS code for ANCOVA analysis

```
proc glm data=OUT2;
   class  DOSE ANTI_PSY_STRAT;
   model PCT = DOSE ANTI_PSY_STRTA;
   lsmeans DOSE / stderr;
run;
```

(page 1 of 1)

11.3. Methods for missing data

Missing data is less of an issue in this trial due to the nature of the data collection; however, this SAP prespecifies the use of multiple imputation to account for the potential for missing data. Multiple imputation will be performed based on missing at random multiple imputation, in which imputed values are based on the treatment group to which the subject is randomized, and on control-based imputation, in which imputed values are obtained from the placebo group rather than the treatment group to which the subject is randomized. The imputation

process will assume that the data are monotone missing with 100 imputed datasets derived for each of the two scenarios. All results of this sensitivity analysis will be summarized in tables and a forest plot.

The first step in the imputation process consists of the steps outlined below:

1. Use multiple imputation to create 100 datasets that satisfy the assumption of monotone missingness (see Table 7).
2. Using the datasets from step 1, use multiple imputation to create 100 complete datasets (see Table 8 for code for missing at random imputation and Table 11 for code for control-based imputation).
3. Fit the MMRM model to each of the 100 complete datasets (see Table 9).
4. Use Rubin's rule (PROC MIANALYZE) to obtain the combined estimates and final result (see Table 10).

Table 7. SAS code to create datasets satisfying monotone missing condition

```
proc sort data= DATA_IMP(observed data set);
      by PATNO DOSE ANTI_PSY_STRATA BASE_PEC;
run;

proc transpose data= DATA_IMP out=DATA_IMP;
      by PATNO DOSE ANTI_PSY_STRATA BASE_PEC;
      ID VISITNUM;
      VAR CHG_PEC;
run;

proc sort data= DATA_IMP;
      by PATNO;
run;

proc mi data=DATA_IMP nimpute=100 round=0.1 seed=
out=OUT_IMP_CONT_DOSE0;
      where DOSE='0';
      var   BASE_PEC(baseline) ANTI_PSY _STRATA
            DELT_PEC30(30 min) DELT_PEC60(60 min)
            DELT_PEC120(120 min) DELT_PEC240(240 min);
      mcmc  chain=multiple impute=monotone;
run;

proc mi data=DATA_IMP nimpute=100 round=0.1 seed=
out=OUT_IMP_CONT_DOSE40;
      where DOSE='40';
      var   BASE_PEC(baseline) ANTI_PSY_STRATA
                  DELT_PEC30(30 min) DELT_PEC60(60 min)
                  DELT_PEC120(120 min) DELT_PEC240(240 min);
      mcmc  chain=multiple impute=monotone;
run;

proc mi data=DATA_IMP nimpute=100 round=0.1 seed=
out=OUT_IMP_CONT_DOSE60;
      where DOSE='60';
      var   BASE_PEC(baseline) ANTI_PSY_STRATA
            DELT_PEC30(30 min) DELT_PEC60(60 min)
            DELT_PEC120(120 min) DELT_PEC240(240 min);
      mcmc chain=multiple impute=monotone;
run;

data OUT_IMP_CONT;
      set OUT_IMP_CONT_DOSE0 OUT_IMP_CONT_DOSE40
      OUT_IMP_CONT_DOSE60;
run;
```

The 100 datasets satisfying the condition of monotone missingness will be contained in the dataset OUT_IMP_CONT. The next step in the process is to fill in any remaining values using multiple imputation.

11.3.1. Missing at random imputation

The code presented in Table 8 will fill in any remaining missing values using missing at random imputation, where subjects' data are imputed based on the treatment group to which they were randomized.

Table 8. SAS code for missing at random imputation

```
proc sort data= DATA_IMP_CONT;
      by PATNO;
run;

proc mi data=OUT_IMP_CONT (where= (DOSE="0")) nimpute=1 round=0.1
      seed= out=IMPDAT_DOSE0;
  by _IMPUTATION_;
      class ANTI_PSY_STRATA;
      var   BASE_PEC(baseline) ANTI_PSY_STRATA
            DELT_PEC30(30 min) DELT_PEC60(60 min)
            DELT_PEC120(120 min) DELT_PEC240(240 min);
      monotone reg (DELT_PEC30(30 min) DELT_PEC60(60 min)
                    DELT_PEC120(120 min)
                    DELT_PEC240(240 min))/details);
run;

proc mi data=OUT_IMP_CONT (where= (DOSE="40")) nimpute=1 round=0.1
      seed= out=IMPDAT_DOSE40;
  by _IMPUTATION_;
      class ANTI_PSY_STRATA;
      var   BASE_PEC(baseline) ANTI_PSY_STRATA
            DELT_PEC30(30 min) DELT_PEC60(60 min)
                 DELT_PEC120(120 min) DELT_PEC240(240 min);
      monotone reg (DELT_PEC30(30 min) DELT_PEC60(60 min)
                    DELT_PEC120(120 min)
                    DELT_PEC240(240 min))/details);
run;

proc mi data=OUT_IMP_CONT (where= (DOSE="60")) nimpute=1 round=0.1
      seed= out=IMPDAT_DOSE60;
      by _IMPUTATION_;
      class ANTI_PSY_STRATA;
      var   BASE_PEC(baseline) ANTI_PSY_STRATA
            DELT_PEC30(30 min) DELT_PEC60(60 min)
              DELT_PEC120(120 min) DELT_PEC240(240 min);
      monotone reg (DELT_PEC30(30 min) DELT_PEC60(60 min)
                    DELT_PEC120(120 min)
                   DELT_PEC240(240 min))/details;
run;

data RANDOM_IMP;
      set IMPDAT_DOSE0 IMPDAT_DOSE40 IMPDAT_DOSE60;
run;
```

(page 1 of 1)

The output dataset is then used for the analysis. This dataset must be transposed for the MMRM analysis. For the SAS code in Table 9, it is assumed that the dataset COMBINE includes the 100 datasets.

Table 9. SAS code for MMRM analysis as part of the imputation process

```
proc mixed data=COMBINE method=reml alpha=0.025 covtest;
      by _IMPUTATION_;
      class  VISITNUM(ref=LAST) DOSE PATNO ANTI_PSY_STRATA;
      model DELT_PEC = BASE_PEC ANTI_PSY_STRATA DOSE
        VISITNUM VISITNUM*DOSE
                / solution ddfm=kr;
      repeated VISITNUM / subject=PATNO type=un r;
      estimate '60 - PBO at 120 minutes' DOSE 0 1 -1
      VISITNUM*DOSE 0 0 0 0 0 0 0 1 -1 0 0 0
            / CL;
      estimate '40 - PBO at 120 minutes' DOSE 1 0 -1
      VISITNUM*DOSE 0 0 0 0 0 0 1 0 -1 0 0 0
            / CL;
      ods output estimates=SI_ESTIMATES;
run;
```

(page 1 of 1)

This analysis will produce 100 estimates of the difference between the dose group of interest (BXCL501 60μg or BXCL501 40μg) and placebo in change from baseline in PEC at 120 minutes. These values are contained in SI_ESTIMATES. Table 10 provides the code for the implementation of Rubin's rules to obtain the combined estimates of treatment effect at the two dose levels.

Table 10. SAS code for Rubin's rule

```
proc mianalyze data=LSM alpha=0.025;
 by TRT VISITNUM;
      modeleffects estimate;
      stderr stderr;
      ods output ParameterEstimates=Week_12_MMRM_LSM;
run;
```

(page 1 of 1)

11.3.2. Control-based imputation

The second part of the sensitivity analysis includes multiple imputation where all missing data values are imputed as if the subject were in the placebo group. The steps are the same as those for the standard imputation outlined above, but the code provided in Table 8 is replaced with

the code in Table 11. The input dataset for this analysis is OUT_IMP_CONT, which is the dataset that satisfies the condition of monotone missingness and is created from the code provided in Table 7.

Table 11. SAS code for control-based imputation

```
proc sort data= OUT_IMP_CONT;
   by _IMPUTATION_ PATNO;
run;

proc mi data=OUT_IMP_CONT nimpute=1 round=0.1 seed=
out=CONT_BASED_CONTROL;
 by _IMPUTATION_;
      class ANTI_PSY_STRATA DOSE;
      var   BASE_PEC(baseline) ANTI_PSY_STRATA
            DELT_PEC30(30 min) DELT_PEC60(60 min) DELT_PEC120(120
            min) DELT_PEC240(240 min);

      mnar model(DELT_PEC30(30 min)DELT_PEC60(60 min)
            DELT_PEC120(120 min) DELT_PEC240(240 min);
            / modelobs=(DOSE='0'));
run;
```

(page 1 of 1)

The output dataset CONT_BASED_CONTROL must be transposed, then the code provided in Table 9 and Table 10 is used to obtain the final result.

11.4. SAS code for Poisson regression modeling frequency of episodes

Table 12 provides the SAS code for the Poisson regression model used to test for differences in the frequency of treated agitation episodes between each dose group and the placebo group as described in Section 6.4.4. The model is fit in PROC GENMOD and the significance level for the test of the null hypothesis of no difference between a specific dose group and the placebo group is obtained from the significance level associated with the test of the coefficient in the model. Each coefficient is interpreted as an incidence rate for a given group.

Table 12. SAS code for Poisson regression

****SAS code for Poisson regression

Note that placebo is the reference group for dose;

```
proc genmod data=new;
      class dose (ref='placebo');
      model no_episode=dose anti_psy_strata/dist=poisson link=log;
      assess link/resample=10000;
run;
```

11.5. SAS code for the PEC change from pre-dose at 2 hours post-dose of study treatment for all treated episodes of agitation (key secondary endpoint)

The PEC change from pre-dose at 2 hours post-dose of study treatment for all treated episodes of agitation (measures continued efficacy) will be assessed using a longitudinal mixed model (see Section 6.4.4). The variables in the model are defined as follows:

- Time – time, as measured in days, of the treated agitation episode
- Base – pre-dose PEC value at the beginning of the treated agitation episode
- Dose – represents the two dose groups and placebo
- Chg – change from pre-dose at 2 hours post-dose of study treatment for each treated episode
- Time * dose – interaction of time and dose level
- Anti_psy_strata – stratification factor

Note that the model includes the time of each treated episode, the pre-dose PEC value for each treated episode, the dose group, stratification factor, and the time*dose group interaction term as covariates. The outcome is the change from pre-dose at 2 hours post-dose of the study treatment for each treated episode. The covariance structure is set to be unstructured for the single random effect of the intercept term.

In this model time is included as a continuous variable and the coefficient associated with dose provides a test of the difference between placebo and dose groups.

Table 13. SAS code for longitudinal assessment of change in PEC

```
proc mixed data=new method=reml covtest asycov asycorr ic;
      class PatientID dose(ref='placebo';
      model chg= base time dose time*dose anti_psy_strata
      /ddfm=kr solution cl;
      random intercept/subject=patno type=un gcorr;
run;
```

11.6. SAS code for pre-dose severity of each treated episode

Severity of each agitation episode at pre-dose, as measured by PEC (measures continued efficacy) will be analyzed using a longitudinal mixed model (see Section 6.4.4). The variables to be included in the model are as follows:

- Base – pre-dose value of the PEC at each treated agitation episode
- Time - time, as measured in days, of the treated agitation episode
- Dose – represents the two dose groups and placebo
- Time * dose – represents the interaction term between time and dose
- Anti_psy_strata – stratification factor

The null hypothesis that the slope measuring the rate of change of the pre-dose PEC score across all treated episodes is zero will be tested. This hypothesis is stated as H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ where β_1, β_2, and β_3 denote the coefficients associated with the time by treatment group interaction term and the significance level for this test is obtained from the type III test of the fixed effect of the interaction term of time by treatment group. Should this hypothesis be rejected, then the significance level associated with each of the parameters, β_1, β_2, and β_3 would be used to test the null hypothesis that the slope for a single parameter is zero. Note that the options for the model statement include "noint" so that the model is fit without an intercept term.

Table 14. Code for assessing severity of each agitation episode at pre-dose

```
proc mixed data=new method=reml covtest asycov asycorr ic;
      class PatientID dose;
      model base= dose time*dose/noint ddfm=kr solution cl;
      random intercept/subject=patientid type=un gcorr;
run;
```