A Phase III Multicenter, Randomized, Double-blind, Placebo-controlled Study to Determine Efficacy and Safety of BXCL501 in Agitation Associated with Schizophrenia

NCT04268303

4/1/2020

BioXcel Therapeutics, Inc.

Protocol BXCL501-301
A Phase III Multicenter, Randomized, Double-blind, Placebo-controlled Study to Determine Efficacy and Safety of BXCL501 in Agitation Associated with Schizophrenia

Statistical analysis plan

Prepared by:

[redacted]

[redacted]

CONFIDENTIAL

Protocol BXCL501-301

Statistical analysis plan

Prepared by: [redacted]

Author: [redacted] Date: April 1, 2020

Approved by: BioXcel Therapeutics, Inc.

[redacted] Date: April 7, 2020

CONFIDENTIAL

Contents

Contents, continued

Contents, continued

Exhibits

Tables

Tables, continued

Abbreviations

ACES	Agitation and Calmness Evaluation Scale
AE	adverse event
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
DEX	dexmedetomidine
ECG	electrocardiogram
eCRF	electronic case report form
HR	heart rate
hr	hour
ITT	intent to treat
IUD	intrauterine device
IWRS	interactive web response system
MedDRA	Medical Dictionary for Regulatory Activities
MINI	Mini-Mental State Examination
MMRM	mixed model repeated measures
NCS	normalized cumulative sum
PANSS	Positive and Negative Syndrome Scale
PEC	Positive and Negative Syndrome Scale – Excited Component
PK	pharmacokinetic
PP	per protocol
SAF	safety population
SAP	statistical analysis plan
SBP	systolic blood pressure
SCI PANSS	Structured Clinical Interview Positive and Negative Syndrome Scale
THC	tetrahydrocannabinol
UDS	urine drug screen
WHO	World Health Organization

1. Introduction

This statistical analysis plan (SAP) describes the planned analyses for BioXcel's Protocol BXCL501-301, entitled "A Phase III Multicenter, Randomized, Double-blind, Placebo-controlled Study to Determine Efficacy and Safety of BXCL501 in Agitation Associated with Schizophrenia". This SAP is based on Version 3 of the protocol, dated January 31, 2020.

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.

2. Study objectives

2.1. Primary objective

The primary objective is to determine if a single dose of BXCL501 effectively reduces symptoms of acute agitation associated with schizophrenia, schizoaffective disorder, or schizophreniform disorder assessed using the Positive and Negative Syndrome Scale (PANSS) – Excited Component (PEC) change from baseline as compared to placebo.

The corresponding primary efficacy endpoint of the study is the absolute change from baseline in the PEC total score at 2 hours. This analysis will use the intent to treat (ITT) population, which consists of all patients who have received any study medication and who had both baseline and at least one efficacy assessment after dosing. Observations will continue to be recorded after use of rescue medication but will be censored (considered missing) in the analyses.

2.2. Key secondary objective

The key secondary objective is to determine the earliest time at which an effect on agitation is apparent as measured by change from baseline in PEC total score.

2.3. Exploratory objectives

Exploratory objectives are to further determine the efficacy, safety, tolerability, and pharmacokinetics (PK) of BXCL501 in patients with acute agitation associated with schizophrenia, schizoaffective disorder, or schizophreniform disorder, including to:

1. determine the overall clinical improvement after drug administration as measured by the Clinical Global Impression–Improvement Scale (CGI-I);
2. describe the duration of calming as measured by PEC and Agitation and Calmness Evaluation Scale (ACES);
3. determine the safety profile of BXCL501 as measured by reports of adverse events (AEs) and vital signs;
4. describe the overall tolerability in terms of treatment-emergent AE reports and local site (oral/sublingual) tolerability of oral film; and
5. describe the subject's opinion on taste, acceptability, and likability of study medication.

3. Study design and conduct

3.1. Study design

This is a randomized, double-blind, placebo-controlled Phase III study assessing the efficacy, safety, and tolerability of BXCL501 in adult (18–75 years old) males and females with acute agitation associated with schizophrenia, schizoaffective disorder, or schizophreniform disorder. This is an in-clinic study.

3.2. Study population

The study will enroll approximately 375 subjects randomized 1:1:1 to receive BXCL501 180µg, BXCL501 120µg, or placebo. Study randomization will be computer generated based on a permuted block design stratified by age (<65 versus ≥65 years old). At the time of the writing of this SAP, no subjects over the age of 65 have been enrolled in the trial.

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

Exhibit 1. Inclusion criteria

1. Male and female patients between the ages of 18 to 75 years, inclusive.
2. Patients who have met DSM-5 criteria for schizophrenia, schizoaffective, or schizophreniform disorder.
3. Patients who are judged to be clinically agitated at Screening and Baseline with a total score of ≥14 on the 5 items (poor impulse control, tension, hostility, uncooperativeness, and excitement) comprising the PEC.
4. Patients who have a score of ≥4 on at least 1 of the 5 items on the PEC at Baseline.
5. Patients who read, understand, and provide written informed consent.
6. Patients who are in good general health prior to study participation as determined by a detailed medical history, physical examination, 12-lead electrocardiogram (ECG) with rhythm strip, blood chemistry profile, hematology, urinalysis, and in the opinion of the Principal Investigator.
7. 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 alone.

Exhibit 2. Exclusion criteria

1. Patients with agitation caused by acute intoxication, including positive identification of alcohol by breathalyzer or drugs of abuse (with the exception of THC) during urine screening.
2. Use of benzodiazepines or other hypnotics or antipsychotic drugs in the 4 hours before study treatment.
3. Treatment with alpha-1 noradrenergic blockers (terazosin, doxazosin, tamsulosin, alfuzosin, or prazosin) or other prohibited medications.
4. Patients judged to be at serious risk of suicide must be excluded.
5. Female patients who have a positive pregnancy test at Screening or are breastfeeding.
6. Patients who have hydrocephalus, seizure disorder, or history of significant head trauma, stroke, transient ischemic attack, subarachnoid bleeding, brain tumor, encephalopathy, meningitis, Parkinson's disease or focal neurological findings.
7. History of syncope or other syncopal attacks, current evidence of hypovolemia, orthostatic hypotension (average of 1, 3, and 5 minute measurements), a baseline heart rate of <55 beats per minute or systolic blood pressure <110 mmHg or diastolic blood pressure <70 mmHg.

(page 1 of 2)

Exhibit 2. Exclusion criteria, continued

8. Patients with laboratory or ECG abnormalities considered clinically significant by the investigator or qualified designee (advanced heart block [second-degree or above atrioventricular block without pacemaker], diagnosis of sick sinus syndrome) that would have clinical implications for the patient's participation in the study.
9. Patients with serious or unstable medical illnesses. These include current hepatic (moderate–severe hepatic impairment), renal, gastroenterologic, respiratory, cardiovascular (including ischemic heart disease, congestive heart failure), endocrinologic, or hematologic disease.
10. Patients who have received an investigational drug within 30 days prior to the current agitation episode.
11. Patients who are considered by the investigator, for any reason, to be an unsuitable candidate for receiving DEX; e.g. patients with a history of allergic reactions to DEX.

(page 2 of 2)

3.3. Study treatment

BXCL501 is a thin film formulation of dexmedetomidine (DEX) for sublingual administration. Dosing delivers 180µg or 120µg of DEX sublingually. The duration of treatment with BXCL501 or placebo is one day.

3.4. Schedule of events

Exhibit 3 displays the schedule of events and assessments.

Exhibit 3. Schedule of events

	Screening	Day 1 treatment evaluation											Day 2 (+1) follow-up	Day 3 discharge	Day 7 (+2) end of study
		Pre-dose[1]	Post-dose[1]												
Timepoint	Pre-treatment	-1 hr to time 0	10 min	20 min	30 min	45 min	1 hr	1.5 hr	2 hr	4 hr	6 hr	8 hr	24 hr (-9/+12 hr)		
Informed consent	×														
Medical history	×														
Demographics	×														
Weight	×												×		
Height	×														
BMI	×														
Alcohol breathalyzer	×														
MINI	×														
Physical exam	×												×		
Safety labs[2]	×													×	×
ECG with rhythm strip[3]	×	×							×				×		
Pulse oximetry		×			×		×		×	×		×			

1. Pre-dose assessments will have a window of 60 minutes prior to dose, with the exception of PEC and ACES which will be performed within 15 minutes of dosing (15 to 0 minutes). All post-dose assessments will have a window of -5/+15 minutes through the 1.5 hour assessments, -5/+25 minutes for the 2 hour assessments (with the exception of the PEC, which will have a +/-3 minute window) and ±30 minutes for the 4, 6, and 8 hour assessments, and full PANSS can be performed at any time.
2. Safety labs will include chemistry, hematology, urinalysis, UDS (local lab; only conducted at Screening), alcohol breathalyzer (only conducted at Screening), and urine pregnancy (only conducted at Screening). Screening/enrollment labs: local labs drawn within 7 days prior to Screening may suffice with the exception of urine drug screen. If results are not available on the same day, a 'desktop' or non-CLIA test may be performed; to confirm, results from a CLIA-certified laboratory should be recorded once available. Central Labs should be performed on Screening, Day 3 and Day 7.
3. ECG for pre-dose does not need to be repeated if Screening ECG is conducted on the day of dosing. ECGs collected following treatment are to be performed prior to PK assessments.

(page 1 of 3)

Exhibit 3. Schedule of events, continued

	Screening	Day 1 treatment evaluation											Day 2 (+1) follow-up	Day 3 discharge	Day 7 (+2) end of study
		Pre-dose[1]	Post-dose[1]												
Timepoint	Pre-treatment	-1 hr to time 0	10 min	20 min	30 min	45 min	1 hr	1.5 hr	2 hr	4 hr	6 hr	8 hr	24 hr (-9/+12 hr)		
Resting vital signs[4]	×	×			×		×		×	×	×	×	×	×	×
Orthostatic vital signs[4]	×	×							×	×		×	×	×	×
Admit to unit	×														
Inclusion/ exclusion criteria	×	×													
Randomization		×													
Study drug administration[10]		×													
PANSS[9]		×									×		×		
PCRS[5]	×	×							×				×		
PEC[5]	×	×	×	×	×	×	×	×	×	×	×	×	×		
ACES[5]		×							×	×		×			

4. Resting (recumbent) vital signs (SBP, DBP, and HR) will be taken upon having the subject lay down for 5 min at Screening, pre-dose, and at 30 minutes, 1, 2, 4, 6, 8, and 24 hours post-dose. Triplicate measurements to be performed in case of systolic BP <90 mmHg, diastolic BP <60 mmHg or pulse <60 bpm. Orthostatic measurements (SBP, DBP, HR, respiratory rate) will be taken upon having the subject stand, with measurements taken after 1, 3, and 5 minutes and temperature will be taken at Screening, Pre-dose, 2, 4, 8, and 24 hours post–first dose, as well as Day 3 and Day 7.
5. PEC will be performed at Screening, pre-dose (within 15 min prior to first dose), and at 10, 20, 30, 45 minutes; 1, 1.5, 2, 4, 6, 8, and 24 hours post-dose. The PCRS must be performed prior to PEC rating, when required. At 6 and 24 hours, the PEC rating must be performed before the PANSS interview. ACES will be performed at pre-dose (within 15 minutesof dose), 2, 4, and 8 hours post-dose.

Exhibit 3. Schedule of events, continued

	Screening	Day 1 treatment evaluation											Day 2 (+1) follow-up	Day 3 discharge	Day 7 (+2) end of study
		Pre-dose[1]	Post-dose[1]												
Timepoint	Pre-treatment	-1 hr to time 0	10 min	20 min	30 min	45 min	1 hr	1.5 hr	2 hr	4 hr	6 hr	8 hr	24 hr (-9/+12 hr)		
CGI-S[6]	×	×													
CGI-I[6]					×		×		×	×					
C-SSRS	×	×											×	×	
Buccal (SL) assessment for local irritation[7]					×				×	×			×		
Likert scales				×											
Likability questions				×											
PK sampling[8]							×			×		×			
Concomitant medications	×	×	×					×					×	×	×
Adverse events	×	×	×					×					×	×	×

6. CGI-S will be performed at Screening and pre-dose. CGI-I will be performed at 30 minutes, 1, 2, and 4 hours post-dose.
7. Buccal exam at 30 minutes, 2, 4, and 24 hours post-dose for local irritation performed by blinded staff.
8. PK blood samples will be collected at 1, 4, and 8 hours (while awake) after dose. A sample may not be collected if the physician indicates in source documents that the patient is in a mental state that is not conducive to PK sample collection. Non-compliance or refusal of all or any PK draw will not be exclusionary nor result in ET. Vital signs are to be done prior to PK sample draws, when performed at the same timepoints.
9. Pre-dose PANSS may be administered at any time prior to dosing on the day of dosing and 6 and 24 hours (-1/+2 hours) post-dose. At 6 and 24 hours PANSS interview must be performed after PEC rating. The 6 hour and 24 hour PANSS is conducted with reference to the time of dosing.
10. The investigator may choose to re-dose the patient with half of a film after the 2 hour post-dose assessments are performed if the PEC change from baseline is $\leq$40%. Patients can be re-dosed after completing the 2 hour post first dose assessments. Repeat dosing administers half of a film. Patients can be re-dosed twice in the 12 hour period post first dose. All assessments listed in this Schedule of Events at the 2 hour post first dose timepoint should be repeated at 2 hours post every re-dose. Assessments at 4, 6, or 8 hour post–first dose that occur within 1 hour of a post re-dose assessment are not required to be performed.

4. Outcome variable definitions

4.1. Screening and baseline characteristics

Screening characteristics: Inclusion/exclusion criteria, demographic characteristics (age, sex, race, ethnicity), height, weight, body mass index (BMI), medical history, prior and concomitant medications, physical examination, resting and orthostatic vital signs, pregnancy test, 12-lead ECG, clinical laboratory results, urine drug screen, alcohol screen, Mini-Mental State Exam (MINI), 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, adverse events, concomitant medications, CGI-S, PEC, ACES, 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 schizophrenia.

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. The Positive and Negative Syndrome Scale (PANSS)

The Structured Clinical Interview (SCI) PANSS is a 30-item scale that measures the severity of 30 individual schizophrenia symptoms based on interview as well as reports of family members or primary care hospital workers. It contains four basic domains (Positive, Negative, General Psychopathology, and Composite Scales); the Composite Scale is the difference between the Positive and Negative Scale scores. Each of the 30 items is rated from 1 to 7:

- 1 = absent
- 2 = minimal
- 3 = mild
- 4 = moderate
- 5 = moderate severe
- 6 = severe
- 7 = extreme

4.2.5. 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

4.2.6. 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.

4.3. Efficacy endpoints

4.3.1. Primary endpoint

The primary efficacy endpoint for this study is the absolute change in the PEC total score from baseline to two hours following dose administration with no rescue medication administered during the two-hour period.

4.3.2. Key secondary endpoints

If the primary endpoint is found to be statistically significant, key secondary analyses will be evaluated using the fixed-sequence method, with the analyses ordered in a hierarchy in which testing is stopped once a significance level within a given dose is greater than 0.025. Thus, if the significance level associated with the analysis of a key secondary endpoint is greater than 0.025, then all significance levels associated with the remaining key secondary endpoints for that dose will be nominal. The key secondary efficacy endpoints, which are measured following dose administration with no rescue medication administered during the two-hour period, will be tested in the following order:

1. Change from baseline in the PEC score at 90 minutes
2. Change from baseline in the PEC score at 60 minutes
3. Change from baseline in the PEC score at 45 minutes
4. Change from baseline in the PEC score at 30 minutes
5. Change from baseline in the PEC score at 20 minutes
6. Change from baseline in the PEC score at 10 minutes

4.3.3. Exploratory endpoints

The exploratory endpoints for this study are:

1. Overall clinical improvement after drug administration as measured by the CGI-I score at 2 hours after dose administration.
2. ACES scores at 2, 4, and 8 hours after dose administration.
3. Change from baseline in total PEC score over time measured from 10 minutes through 24 hours after dosing.
4. PEC responders and CGI-I responders at 2 hours following dose of BXCL501 compared with placebo:
    a. PEC responders will be defined as those who achieve at least a 40% reduction in PEC total score from baseline at or before 2 hours post-dose.
    b. CGI-I responders will be defined as subjects with a score of 1 or 2 on the CGI-I scale (the CGI-I non-responders will be defined as subjects with scores from 3 to 7 at 2 hours).
5. Time to rescue medication during the entire 24-hour post-treatment evaluation period for subjects receiving BXCL501 compared to placebo.
6. Proportion of subjects per treatment group who received rescue medication by 4 hours and within 24 hours after dosing.
7. Duration of calming effect as described by the change from baseline in PEC total score and ACES score at 2, 4, and 8 hours after dosing.
8. Describe the effect on overall psychotic symptoms and subscales (PANSS total, positive, negative, and general psychopathology subscales).
9. Determine the safety profile of BXCL501 as measured by vital signs and treatment-emergent AE reports.
10. Describe the overall tolerability in terms of AE reports and local site (oral/sublingual) tolerability of oral film.
11. Descriptive PK of BXCL501 in the patient population.

12. Determine patient acceptability, taste, and likability of study medication using Likert scales to capture subject's acceptability, opinion on taste, and questions regarding likability.

4.4. Safety assessments

Safety will be assessed during the study by the monitoring and recording of 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 4 presents the laboratory tests that will be collected at the timepoints specified in the schedule of events (Exhibit 3).

Exhibit 4. 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, and total protein
Urinalysis	Includes pH, specific gravity, protein, glucose, ketones, bilirubin, blood, nitrites, leukocytes, urobilinogen, microscopic urine analysis if dipstick positive
Urine pregnancy test*	Conducted for females of childbearing potential only (local only)
Urine drug screen*	Cocaine, amphetamine, phencyclidine, benzodiazepines, marijuana. (Note: Marijuana positive is allowed provided subject is not moderately or severely dependent. Benzodiazepine positive is allowed if prescribed.)
Alcohol breathalyzer*	

*Conducted at Screening only

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 sitting or supine 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, and 24 hours post-dose.

4.4.6. Physical examinations

A standard physical examination will be performed at Screening and 24 hours post-dose. 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 each day 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 7 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. In addition, beginning at the time of the database lock, an audit trail will be maintained for all versions of the analysis datasets that may result from refinements of the algorithms for derived variables in the course of the analysis.

5.3. Authorization for unblinding

After the database lock and upon receipt of written authorization from BioXcel, a blinded study team will receive the actual treatment codes directly from the group maintaining the interactive web response system (IWRS). 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.

6.1. Statistical methodology

6.1.1. Sample size determination

The sample size for this study was estimated using information from BXCL501-102 (the Phase Ib study of BXCL501) regarding the standard deviation of the change from baseline in

PEC score and the expected magnitude of this difference. As two dose groups (120μg and 180μg) will be compared to the placebo group, a two-sided significance level of 0.025 will be used for each comparison, resulting in an overall significance of 0.05. Assuming a two-sided significance level of 0.025, a two-sample t-test, a randomization ratio of 1:1:1, an assumed power of 0.90, and standard deviations of 4.0, 4.2, 4.4, and 4.6, the estimated sample size per group is provided in Table 1 below. Table 2 provides the estimated sample size per group for a power of 0.85. The observed difference between the BXCL501 and placebo groups in the 180μg group in the Phase Ib study was 0.7, 3.5, 5.0, 6.0, and 6.2 at 30, 45, 60, 90, and 120 minutes, respectively. With a sample size of at least 375 subjects (125 per group), this study is well powered to detect differences at 45 minutes or later.

Table 1. Estimated sample size per group for power of 0.90, a two-sided significance level of 0.025, varying mean difference in change from baseline in PEC score, and varying standard deviations of the change from baseline in PEC score

	Standard deviation			
Mean difference	4.0	4.2	4.4	4.6
2.0	101	111	122	133
1.9	112	123	135	147
1.8	124	137	150	164
1.7	139	153	168	184

Table 2. Estimated sample size per group for power of 0.85, a two-sided significance level of 0.025, varying mean difference in change from baseline in PEC score, and varying standard deviations of the change from baseline in PEC score

	Standard deviation			
Mean difference	4.0	4.2	4.4	4.6
2.0	88	97	106	115
1.9	97	107	117	128
1.8	108	119	130	142
1.7	121	133	146	159

6.1.2. Analysis populations

The following populations are planned:

- Safety population (SAF): All subjects who receive study drug
- Intent to treat (ITT) population: All subjects in the safety population who have a PEC score post-dose
- Per protocol (PP) population: All subjects in the ITT population with no major protocol deviations

6.1.3. General considerations

The protocol-specified time windows for a visit and time of a measurement determine the visit assignment for each observation. Listings will present, by patient number and measurement time, all data collected on electronic case report forms (eCRFs), results analyzed by the central laboratory, parameters calculated from activity monitors, plasma concentrations, and other biomarker results.

The same number of decimal places as in the raw data will be presented when reporting minimum and maximum. One more decimal place than in the raw data will be presented when reporting mean and standard deviation.

Continuous variables will be summarized by treatment using descriptive statistics (n, mean, median, standard deviation, minimum, and maximum). For categorical variables, frequencies and percentages will be presented by treatment. Baseline is defined as the last observation prior to initiation of study medication.

All statistical testing will be based on a two-sided significance level of 0.025 unless otherwise stated.

Data listings will be provided for all subjects up to the point of withdrawal.

Baseline will be defined as the value collected at the pre-dose timepoint. Should this value be unavailable, the Screening value will be used. Instead of relying solely on visit labels in the

clinical database for post-baseline values, results collected at a particular time will be attributed to a specific time point by calculating the time relative to randomization. Listings will present all measurements.

For visits (or events) that occur on or after randomization, time is defined as:

time = time of assessment – pre-dose time (time 0)

For listings (such as for AEs), the quantity "time since first dose" is defined as:

time since first dose = event time – time of first dose (time 0)

6.1.4. Procedures for handling missing data

Multiple imputation will be used to handle missing data. Details are provided in Section 6.4.1.

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, 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

The efficacy analysis will be conducted within the estimand framework. A summary of the attributes of the estimand is provided in Table 3.

Table 3. Estimand attributes for the primary efficacy analysis

Estimand attribute	Description of attribute
Target population	ITT population—All patients who take any study medication and who had both baseline and at least one efficacy assessment after dosing
Primary endpoint	Absolute change from baseline in the total PEC total score at 120 minutes
Handling of intercurrent events	All values collected after the use of rescue treatment and withdrawal from study will be set to missing.
Population-level summary	Difference between dose-specific treatment arm and placebo arm in mean change from baseline in the total PEC score using all measurements from baseline through 120 min

The trial product estimands are defined as:

- The difference in the mean change from baseline through 120 minutes in the total PEC scores between the BXCL501 180μg arm and the placebo arm in the ITT population for subjects who did not use rescue medication.
- The difference in the mean change from baseline through 120 minutes in the total PEC scores between the BXCL501 120μg arm and the placebo arm in the ITT population for subjects who did not use rescue medication.

The null and alternative hypotheses to be tested for these estimands are H_{01}: $\Delta_{BXCL501\_180} = \Delta_{PBO}$ and H_{A1}: $\Delta_{BXCL501\_180} \neq \Delta_{PBO}$ and H_{02}: $\Delta_{BXCL501\_120} = \Delta_{PBO}$ and H_{A2}: $\Delta_{BXCL501\_120} \neq \Delta_{PBO}$, where $\Delta_{BXCL501\_180}$ denotes the change from baseline in the PEC score at 2 hours post-dose in the BXCL501 180μg group, $\Delta_{BXCL501\_120}$ denotes the change from baseline in the PEC score at 2 hours post-dose in the BXCL501 120μg group, and Δ_{PBO} denotes the change from baseline in the PEC at 2 hours post-dose in the placebo group. These hypotheses, H_{01} and H_{02}, will be tested using a mixed model repeated measures (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 and set at 0.025. For the analysis of the key secondary endpoints, a Bonferroni correction will be applied with testing for each hypothesis using a significance level

of 0.025. A fixed sequence approach will then be used to adjust for the multiple testing at each timepoint as described in Section 4.3.2.

6.4.1. Primary analysis

The primary efficacy endpoint is the absolute change from baseline in the PEC total score at 2 hours in the ITT population as described in Table 3. The population-level summary is the difference between the BXCL501 180µg and placebo arms and the difference between the BXCL501 120µg and placebo arms in the mean absolute change from baseline in the PEC total score at 2 hours. These differences will be estimated using a MMRM as described below.

The MMRM will be fit with change from baseline as the outcome at each of the following timepoints: 10, 20, 30, 45, 60, 90, and 120 minutes. The covariates included in the model are as follows: baseline PEC score, age stratum, study site, visit, treatment group (BXCL501 180µg, BXCL501 120µg, placebo), baseline PEC score by visit interaction term, and treatment group by visit interaction term. The model will be fit using restricted maximum likelihood with an unstructured covariance matrix and the Kenward-Roger approximation for the computation of the degrees of freedom. If the MMRM with an unstructured covariance matrix does not converge, the following covariance structures will be substituted, in the order listed below. Each subsequent covariance structure will be used only if each previous covariance structure was used and no previous model converged.

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).

There are two tests of interest: the comparison of the BXCL501 180µg group to placebo and the comparison of the BXCL501 120µg group to placebo. The estimated mean difference between the group treated with either BXCL501 180µg or 120µg and the placebo group in change from baseline 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 9.1). 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.

There was little missing data in the Phase 2 study of BXCL501, and the expectation is that missing data will 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. Sensitivity analyses 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 based on control-based imputation, in which imputed values are obtained from the placebo data rather than the treatment group to which the subject is randomized. 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 9.2) 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

In addition to the primary efficacy endpoint, PEC scores for each of the comparisons of BXCL501 180µg or 120µg to placebo will be summarized over the course of the study using the MMRM described in Section 6.4.1. The null hypothesis of no difference in the change from

baseline between groups will be tested at each of the time points listed below. This analysis is part of the primary efficacy analysis and the individual test results will be obtained from the least squares means analysis for the primary efficacy analysis. The results 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. Change from baseline in the PEC score at 90 minutes
2. Change from baseline in the PEC score at 60 minutes
3. Change from baseline in the PEC score at 45 minutes
4. Change from baseline in the PEC score at 30 minutes
5. Change from baseline in the PEC score at 20 minutes
6. Change from baseline in the PEC score at 10 minutes

The fixed-sequence method will be used to adjust for multiplicity. 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.

The estimand framework for this analysis is described in Table 4.

Table 4. Estimand attributes for the key secondary efficacy analysis

Estimand attribute	Description of attribute
Target population	ITT population – all patients who take any study medication and who had both baseline and at least one efficacy assessment after dosing
Key secondary endpoints	Absolute change from baseline in the total PEC total score at 90, 60, 45, 30, 20, and 10 min
Handling of intercurrent events	All values collected after the use of rescue treatment and withdrawal from study will be set to missing.
Population-level summary	Difference between dose-specific treatment arm and placebo arm in mean change from baseline in the total PEC score using all measurements from baseline through 90, 60, 45, 30, 20, and 10 minutes

The trial product estimands are defined at each timepoint (90, 60, 45, 30, 20 and 10 minutes) as:

- The difference in the mean change from baseline through the timepoint of analysis in the total PEC scores between the BXCL501 180µg arm and the placebo arm in the ITT population for subjects who did not use rescue medication.
- The difference in the mean change from baseline through the timepoint of analysis in the total PEC scores between the BSCL501 120µg arm and the placebo arm in the ITT population for subjects who did not use rescue medication.

6.4.3. Exploratory analyses

The exploratory endpoints for this study and a description of the analysis approach for each endpoint are provided below:

1. Overall clinical improvement after drug administration as measured by the CGI-I score: This will be summarized for each timepoint with an ANCOVA, which will be used to compare each of the dose groups to placebo at 30 minutes, 1 hour, 2 hours, and 4 hours. The model will include treatment group, study site, and age stratum as covariates. Empirical cumulative distribution function curves of the change in CGI-I will be presented for the change at each time point. The plots will include five curves representing the following five groups: ≥2 category decrease, 1 category decrease, no change, 1 category increase, and ≥2 category increase. An empirical cumulative distribution function curve of the change in CGI-I will also be presented for each treatment group versus placebo comparison and timepoint.
2. ACES scores at 2, 4, and 8 hours after dose administration: This will be evaluated using ANCOVA with baseline ACES, age stratum, study site, and treatment group as covariates at each timepoint. Empirical cumulative distribution function curves of the change in ACES will be presented for the change at each timepoint. The plots will include five curves representing the following seven groups: ≥3 category decrease, 2 category decrease, 1 category decrease, no change, 1 category increase, 2 category increase, and ≥3 category increase. An empirical cumulative distribution function

curve of the change in ACES will be presented for each treatment group versus placebo comparison and timepoint.

3. Change from baseline in total PEC score over time measured from 10 minutes through 24 hours after dosing: These results will be obtained from the MMRM used for the primary efficacy analysis with additional outcomes added in the model to include the 10-minute through 24-hour timeline. Empirical cumulative distribution function curves will also be presented for each treatment group versus placebo comparison and timepoint. Descriptive changes will also be summarized by study site. This will include the estimated mean change from baseline in total PEC score by time and treatment group for each study site to describe potential variation across sites. Forest plots will be constructed for important timepoints to summarize the results.
4. PEC responders and CGI-I responders at 2 hours following dose of BXCL501 compared with placebo:
    a. PEC responders will be defined as those who achieve at least a 40% reduction in PEC total score from baseline at or before 2 hours post-dose and will be evaluated using a Fisher's exact test for each dose group comparison.
    b. CGI-I responders will be defined as subjects with a score of 1 or 2 on the CGI-I scale (the CGI-I non-responders will be defined as subjects with scores from 3–7 at 2 hours) and will be evaluated using a Fisher's exact test for each dose group comparison.
5. Time to rescue medication during the entire 24-hour post–treatment evaluation period for subjects receiving BXCL501 180μg or 120μg compared to placebo: Descriptive statistics will be used to evaluate this endpoint, with a Kaplan-Meier curve describing time to rescue medication estimated for each of the three groups (BXCL501 180μg, BXCL501 120μg, and placebo). A log rank test will be used to compare each of the two active treatment groups to placebo.

6. Number and proportion of subjects per treatment group who received rescue medication by 4 hours and within 24 hours after dosing: This will be evaluated using a Fisher's exact test for each dose group comparison.
7. Duration of calming effect as described by the change from baseline in PEC total score, and ACES score at 2, 4, and 8 hours after dosing: This will be evaluated using the following definitions:
    - For the PEC total score, calming is defined as achieving at least a 40% reduction in PEC score from baseline at or before 2 hours post-dose.
    - For the ACES score, calming is defined as an improvement of 1 or more points in the ACES score relative to baseline.

    The proportion of subjects experiencing calmness at each timepoint will be described using plots and shift tables.
8. Overall psychotic symptoms and subscales (PANSS total, positive, negative, and general psychopathology subscales): Descriptive methods will be used for each of these endpoints, with an MMRM model that includes baseline subscale value, study site, time, treatment group, a baseline subscale by time interaction term, and a treatment group by time interaction term. The results will be summarized in tabular and graphical form. An empirical cumulative distribution function curve of the change in PANSS subscale will be presented by treatment group for each timepoint.
9. Determine the safety profile of BXCL501 as measured by vital signs and treatment-emergent AE reports: Descriptive methods including frequency distributions, box plots, shift tables and listings will be used to present the safety data. All safety data will be presented by treatment group.
10. Describe the overall tolerability in terms of AE reports and local site (oral/sublingual) tolerability of oral film: Descriptive methods, including frequency distributions and listings, will be used to describe tolerability.
11. Descriptive PK of BXCL501 180μg and 120μg in the patient population: This will be presented in a separate PK report. Tables of PK parameters will be presented.

12. Determine patient acceptability, taste and likability of study medication using Likert scales to capture subject's acceptability, opinion on taste, and questions regarding likability: An ANCOVA will be used to describe any differences between the BXCL501 dose groups and placebo groups.

An additional exploratory efficacy endpoint, time to resolution of agitation as measured by achieving an ACES score of at least 4, will also be analyzed. A Fisher's exact test for difference in proportions of subjects with resolution will be conducted at each time point for each dose group comparison. The results will be presented in tabular form and graphically.

7. Safety analyses

All safety analyses will be performed using the safety population. Safety and tolerability will be assessed by clinical review of all safety parameters including AEs, laboratory values, and vital signs. The safety analyses will include all results collected from randomization through the end of the study. All safety presentations will be presented by treatment group (BXCL501 180μg, BXCL501 120μg, and placebo).

7.1. Adverse events

AEs will be coded using the Medical Dictionary for Regulatory Affairs (MedDRA) version 22.0 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

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 for each treatment at Screening, discharge, and end of study. Boxplots of laboratory values will be provided.

7.3. Physical examination

A listing of physical examination findings will be provided by subject.

7.4. Vital signs

Each resting vital sign observed value and change from baseline (when appropriate) will be summarized for each treatment group at Screening, pre-dose, 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 8 hours, and 24 hours post–dose administration. Each orthostatic vital sign observed value and change from baseline (when appropriate) will be summarized for each treatment group at Screening, pre-dose, 2 hours, 4 hours, 8 hours, and 24 hours post–dose administration.

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, and 24 hours post–dose administration. In addition, frequency tabulation of the overall ECG results (normal, abnormal normalized cumulative sum [NCS], and abnormal cumulative sum [CS]) will be summarized (to include any emergent arrhythmias and determination of resolution). Conduction intervals including PR, QRS, QT (>450, >500ms) and QTc (>450, >500 ms) will be summarized and tabulated. Shift tables of clinically significant findings will be reported for corresponding QT and QTc interval changes.

7.6. Concomitant medications

Concomitant medications will be summarized (n and %) by anatomical therapeutic chemical (ATC) class and preferred term (coded by WHO Drug coding dictionary March 2019) for each treatment group. This table will also include an overall total column.

7.7. Buccal exam for local irritation

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

- Negative reaction to the sub-lingual 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 and figure will be presented for these parameters for each treatment group at each timepoint.

8. References

1. 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.

9. 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.

9.1. MMRM analysis

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

- DOSE – treatment group assignment with dose of 120µg and 180µg
- VISITNUM – denotes time of measurement (10, 20, 30, 45, 60, 90, and 120 minutes)
- AGE_STRATA – indicates age stratum for randomization
- SITE – study site
- CHG_PEC – change from baseline in PEC total score
- VISITNUM*DOSE – interaction of visit and dose group
- BASE_PEC*VISITNUM – interaction of visit and baseline PEC measure

The LSMEANS statement will be 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 of the 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 AGE_STRATA SITE;
   model CHG_PEC = BASE_PEC AGE_STRTA SITE DOSE VISITNUM
                   VISITNUM*DOSE BASE_PEC*VISITNUM
         / solution ddfm=kr;
   repeated VISITNUM / subject=PATNO type=un r;
   lsmeans VISITNUM*DOSE / diff om cl e;
   estimate '180 - PBO at 120 minutes' DOSE 0 1 -1
   VISITNUM*TRTGRP 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 -1
         / CL;
   estimate '120 - PBO at 120 minutes' DOSE 1 0 -1
   VISITNUM*TRTGRP 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1
         / CL;
   ods output estimates=SI_ESTIMATES;
run;
```

9.2. Methods for missing data

Missing data is less of an issue in this trial due to the in-clinic 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 6).
2. Using the datasets from step 1, use multiple imputation to create 100 complete datasets (see Table 7 for code for missing at random imputation and Table 10 for code for control-based imputation).
3. Fit the MMRM model to each of the 100 complete datasets (see Table 8).

4. Use Rubin's rule (PROC MIANALYZE) to obtain the combined estimates and final result (see Table 9).

Table 6. SAS code to create datasets satisfying monotone missing condition

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

proc transpose data= DATA_IMP out=DATA_IMP;
      by PATNO DOSE AGE_STRAT SITE 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) AGE_STRAT SITE
            DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);
      mcmc  chain=multiple impute=monotone;
run;

proc mi data=DATA_IMP nimpute=100 round=0.1 seed=
out=OUT_IMP_CONT_DOSE120;
      where DOSE='120';
      var   BASE_PEC(baseline) AGE_STRAT SITE
            DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);
      mcmc  chain=multiple impute=monotone;
run;
```

(page 1 of 2)

Table 6. SAS code to create datasets satisfying monotone missing condition, continued

```
proc mi data=DATA_IMP nimpute=100 round=0.1 seed=
out=OUT_IMP_CONT_DOSE180;
      where DOSE='180';
      var   BASE_PEC(baseline) AGE_STRAT SITE
            DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);
      mcmc chain=multiple impute=monotone;
run;

data OUT_IMP_CONT;
      set OUT_IMP_CONT_DOSE0 OUT_IMP_CONT_DOSE120
      OUT_IMP_CONT_DOSE180;
run;
```

(page 2 of 2)

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.

9.2.1. Missing at random imputation

The code presented in Table 7 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 7. 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 AGE_STRAT SITE;
      var   BASE_PEC(baseline) AGE_STRAT SITE
            DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);
      monotone reg (DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                DELT_PEC120(120 min))/details);
run;

proc mi data=OUT_IMP_CONT (where= (DOSE="120")) nimpute=1 round=0.1
      seed= out=IMPDAT_DOSE120;
  by _IMPUTATION_;
      class AGE_STRAT SITE;
      var   BASE_PEC(baseline) AGE_STRAT SITE
            DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);
      monotone reg (DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                DELT_PEC120(120 min))/details);
run;

proc mi data=OUT_IMP_CONT (where= (DOSE="180")) nimpute=1 round=0.1
      seed= out=IMPDAT_DOSE180;
      by _IMPUTATION_;
      class AGE_STRAT SITE;
      var   BASE_PEC(baseline) AGE_STRAT SITE
            DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);
      monotone reg (DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min))/details;
run;
```

(page 1 of 2)

Table 7. SAS code for missing at random imputation, continued

```
data RANDOM_IMP;
      set IMPDAT_DOSE0 IMPDAT_DOSE120 IMPDAT_DOSE180;
run;
```

(page 2 of 2)

The output dataset is then used for the analysis. This dataset must be transposed for the MMRM analysis. For the SAS code in Table 8, it is assumed that the dataset COMBINE includes the 100 datasets.

Table 8. 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 AGE_STRAT SITE;
      model DELT_PEC = BASE_PEC AGE_STRAT DOSE SITE
         VISITNUM VISITNUM*DOSE BASE_PEC*VISITNUM
                / solution ddfm=kr;
      repeated VISITNUM / subject=PATNO type=un r;
      estimate '180 - PBO at 120 minutes' DOSE 0 1 -1
      VISITNUM*TRTGRP 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 -1
            / CL;
      estimate '120 - PBO at 120 minutes' DOSE 1 0 -1
      VISITNUM*TRTGRP 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 -1
            / CL;
      ods output estimates=SI_ESTIMATES;
run;
```

This analysis will produce 100 estimates of the difference between the dose group of interest (BXCL501 180µg or BXCL501 120µg) and placebo in change from baseline in PEC at 120 minutes. These values are contained in SI_ESTIMATES. Table 9 provides the code for the implementation of Rubin's rules to obtain the combined estimates of treatment effect at the two dose levels.

Table 9. SAS code for Rubin's rule

```
proc mianalyze data=LSM alpha=0.025;
 by TRT VISITNUM;
      modeleffects estimate;
      stderr stderr;
      ods output ParameterEstimates=Week_48_MMRM_LSM;
run;
```

9.2.2. 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 7 is replaced with the code in Table 10. 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 10.

Table 10. 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 AGE_STRAT DOSE SITE;
      var   BASE_PEC(baseline) AGE_STRAT SITE
            DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);

      mnar model(DELT_PEC10(10 min) DELT_PEC20(20 min)
                  DELT_PEC30(30 min) DELT_PEC45(45 min)
            DELT_PEC60(60 min) DELT_PEC90(90 min)
                  DELT_PEC120(120 min);
 / modelobs=(DOSE='0'));
run;
```

The output dataset CONT_BASED_CONTROL must be transposed, then the code provided in Table 8 and Table 9 is used to obtain the final result.

BioXcel Therapeutics, Inc.
Protocol BXCL501-301
Statistical analysis plan addendum

This addendum to the BXCL501-301 BioXcel Therapeutics, Inc. Protocol BXCL501-301 statistical analysis plan (SAP) dated April 1, 2020, describes the following changes to be made to the SAP:

- To ensure consistency with the availability of assessments per the schedule of events, nominal timepoints, rather than protocol-specified time windows as specified in Section 6.1.3 of the SAP, will be used to determine the timepoint assigned to each observation.
- Exploratory endpoints that were specified to be assessed using ANOVA or ANCOVA and are available at multiple timepoints will instead use the MMRM approach specified for the primary analysis, using baseline as a covariate if a baseline assessment is available for that measure. These include CGI-I (using CGI-S as a baseline covariate, with a dependent variable of CGI-I – not change) and ACES (with a dependent variable of change from baseline).
- As a point of clarification, the exploratory endpoint of PEC response by 2 hours (Section 4.3.3, 4.a.) will include all subjects who meet the definition of response (i.e., :'. 40% reduction from baseline) on or before 2 hours, irrespective of whether or not that reduction is maintained until the 2-hour timepoint.
- Graphical displays may be revised from what is specified in the SAP based on a detailed review of proposed TLFs.

[redacted]

Date: May 20, 2020

[redacted]

Date: May 20, 2020

BioXcel Therapeutics, Inc.
Protocol BXCL501-301
Statistical analysis plan addendum

This addendum to the BioXcel Therapeutics, Inc. Protocol BXCL501-301 statistical analysis plan (SAP) dated June 8, 2020, describes the following changes to be made to address concerns that FDA raised concerning the clarity of the estimands and the handling of intercurrent events. Their specific concern is as follows:

> "Your description of the primary estimand is not clear. You did not provide a rationale either, so it is difficult for us to evaluate it. Is the use of rescue treatment the only relevant intercurrent event needing to handle? You need to first list all intercurrent events, then clarify the strategy for handling each listed intercurrent event with rationale. For instance, what is the rationale for excluding from the primary analysis model the observations collected while on rescue medication (that is, considering the values missing)? Also, please provide the anticipated percentage of patients who may need rescue medication during the two hours. You should plan a sensitivity analysis by including all observations whether on rescue medication or not."

The primary estimands are defined as:

- The difference in the mean change from baseline through 120 minutes in the total PEC scores between the BXCL501 180μg arm and the placebo arm in the ITT population for subjects who did not use rescue medication
- The difference in the mean change from baseline through 120 minutes in the total PEC scores between the BXCL501 120μg arm and the placebo arm in the ITT population for subjects who did not use rescue medication

The two intercurrent events are the use of rescue medication and withdrawal from study. Given the short treatment period of 120 minutes and the fact that rescue medication was not allowed during this time interval, we anticipated that there would be few intercurrent events. This thinking was further supported by the results of the phase 2 study (BXCL501-102) where there were no missing data in the first 120 minutes and no rescue medication in the first 120

minutes in the cohorts where it was prohibited. The SAP should have provided this information to clarify our thinking around the handling of intercurrent events.

As both the BXCL501-301 and -302 studies are now complete, we can confirm that there was no use of rescue medication in the first 120 minutes and that there are also no missing values in this time interval. The SAP provides details for the use of missing data methods if missing data were a problem.

Considering this information, the section describing the estimands is changed as follows:

6.4 Efficacy analyses

The efficacy analysis will be conducted within the estimand framework. A summary of the attributes of the estimand is provided in Table 1.

Table 1. Estimand attributes for the primary efficacy analysis

Estimand attribute	Description of attribute
Target population	ITT population—All patients who take any study medication and who had both baseline and at least one efficacy assessment after dosing
Primary endpoint	Absolute change from baseline in the total PEC total score at 120 minutes
Handling of intercurrent events	No intercurrent events, including missing values or use of rescue treatment, were observed from baseline through 120 minutes
Population-level summary	Difference between dose-specific treatment arm and placebo arm in mean change from baseline in the total PEC score using all measurements from baseline through 120 minutes

The trial product estimands are defined as:

- The difference in the mean change from baseline through 120 minutes in the total PEC scores between the BXCL501 180μg arm and the placebo arm in the ITT population

- The difference in the mean change from baseline through 120 minutes in the total PEC scores between the BXCL501 120µg arm and the placebo arm in the ITT population

The null and alternative hypotheses to be tested for these estimands are H_{01}: $\Delta_{BXCL501\_180} = \Delta_{PBO}$, H_{A1}: $\Delta_{BXCL501\_180} \neq \Delta_{PBO}$, H_{02}: $\Delta_{BXCL501\_120} = \Delta_{PBO}$, and H_{A2}: $\Delta_{BXCL501\_120} \neq \Delta_{PBO}$, where $\Delta_{BXCL501\_180}$ denotes the change from baseline in the PEC score at 2 hours post-dose in the BXCL501 180µg group, $\Delta_{BXCL501\_120}$ denotes the change from baseline in the PEC score at 2 hours post-dose in the BXCL501 120µg group, and Δ_{PBO} denotes the change from baseline in the PEC at 2 hours post-dose in the placebo group. The null hypotheses, H_{01} and H_{02}, will be tested using a mixed model repeated measures (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 and set at 0.025. For the analysis of the key secondary endpoints, a Bonferroni correction will be applied, and each hypothesis will be tested using a significance level of 0.025. A fixed sequence approach will then be used to adjust for the multiple testing at each timepoint.

[redacted]

Date: June 8, 2020

[redacted]

Date: [redacted]

BioXcel Therapeutics, Inc.
Protocol BXCL501-301
Statistical analysis plan addendum to evaluate clinically meaningful change

This addendum to the BioXcel Therapeutics, Inc. Protocol BXCLS0l-301 statistical analysis plan (SAP) dated January 20, 2021, includes additional analyses to address clinically meaningful within patient change in PEC scores using anchor-based methods and cumulative distribution function curves. The following analyses will beconducted:

1. An analysis of covariance (ANCOVA) will be perfor_med (PROC GLM) using the CGI-1rating at 2 hours post administration controlling for age, sex and baseline PECScore. The individual response options of the CGH will be used as individual groups to examine the PEC change score but may also be collapsed pending sample size of each response group.

2. Empirical cumulative distribution function (eCDF) curves will be programmed to evaluate the raw change score of the PEC at 2 hours post-dose on the ITT population as follows:
   a. By treatment group assignment
   b. By CGI-Improvement rating

3. Probability density function (PDF) plots will be programmed to evaluate the raw change score of the PEC at 2 hours post-dose on the ITT population as follows:
   a. By treatment group assignment
   b. By CGI-Improvement rating

4. Mean change score and effect sizes will becalculated for the overall sample, by treatment assignment, and by CGI-improvement group to provide distribution-based support for the above analyses.

[redacted]

Date: ZO-:Jin-20-2-..J

[redacted]

Date: [redacted]