NCT02600507

Study ID: ITI-007-402

Title: A Phase 3, Randomized, Double-Blind, Placebo-Controlled, Multi-Center Study to Assess the Efficacy and Safety of ITI-007 Adjunctive to Lithium or Valproate in the Treatment of Patients With Major Depressive Episodes Associated With Bipolar I or II Disorder (Bipolar Depression)

Statistical Analysis Plan Date: August 11, 2020

STATISTICAL ANALYSIS PLAN

ITI-007-402

A PHASE 3, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTI-CENTER STUDY TO ASSESS THE EFFICACY AND SAFETY OF ITI-007 ADJUNCTIVE TO LITHIUM OR VALPROATE IN THE TREATMENT OF PATIENTS WITH MAJOR DEPRESSIVE EPISODES ASSOCIATED WITH BIPOLAR I OR BIPOLAR II DISORDER (BIPOLAR DEPRESSION)

AUTHOR: [redacted]

VERSION NUMBER AND DATE: V2.0, 11AUG2020

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V2.0 (Dated 11AUG2020) for Protocol ITI-007-402 V1.6.

[redacted]

Upon review of this document, the undersigned approves this version of the Statistical Analysis Plan, authorizing that the content is acceptable for the reporting of this study.

[redacted]

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
1.0	01NOV2019	[redacted]	Not Applicable – First Version
2.0	11AUG2020	[redacted]	Added additional Protocol deviation categories. Added exploratory efficacy analysis for MADRS total score. Added additional section/analyses to assess COVID-19 impact. Added additional text to ECG and Laboratory sections for data to include in summaries. Added additional estimand analysis in Appendix 3. Included primary and key secondary efficacy analyses specific for Non-US submissions in Appendix 4 by adding the hypotheses based on ITT patients enrolled from the rest of the world in addition to the hypotheses based on overall ITT patients (which includes US and the rest of the world patients).

TABLE OF CONTENTS

1. LIST OF ABBREVIATIONS

AE	Adverse Event
AIMS	Abnormal Involuntary Movement Scale
ANCOVA	Analysis of Covariance
AR(1)	Autoregressive(1)
ARH(1)	Heterogeneous autoregressive(1)
BARS	Barnes Akathisia Rating Scale
BLQ	Below the limit of quantification
BMI	Body mass index
BPI	Bipolarity Index
C-SSRS	Columbia – Suicide Severity Rating Scale
CGI-BP-S	Clinical Global Impression Scale, Bipolar Version – Severity
CI	Confidence interval
COVID-19	Coronavirus Disease 2019
DMC	Data Monitoring Committee
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ENR	All Patients Enrolled
FA0(q)	No diagonal factor analytic
HIV	human immunodeficiency virus
HR	Heart rate
IEC	Independent Ethics Committee
IRB	Institutional Review Board
ITCI	Intra-Cellular Therapies, Inc. (Sponsor)
ITT	Intent-to-Treat
LOCF	Last observation carried forward
LS	Least squares
MADRS	Montgomery-Åsberg Depression Rating Scale
MAR	Missing-at-random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed-Effect Model Repeated Measure
MNAR	Missing-not-at-random
NEO-FFI	Neuroticism, Extraversion and Openness to Experience - Five Factor Inventory
PK	Pharmacokinetic(s)
PR	PR interval of the electrocardiogram; time duration between the P and R waves
PT	Preferred Term
Q-LES-Q-SF	Quality of Life Enjoyment and Satisfaction Questionnaire – Short Form
QRS	QRS interval of the electrocardiogram; duration of the QRS complex
QT	QT interval of ECG, duration between the Q and T waves
QTcB	QT interval of ECG corrected for heart rate using Bazett's formula
QTcF	QT interval of ECG corrected for heart rate using Fridericia's formula
RND	All Patients Randomized
RR	Time duration between two consecutive R waves of the electrocardiogram
SAE	Serious Adverse Event
SAF	Safety Analysis
SAP	Statistical Analysis Plan
SAS	Simpson-Angus Scale
SAS®	Statistical Analysis Software

SD	Standard deviation
SDS	Sheehan Disability Scale
SMQs	Standard MedDRA Queries
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TOEPH	Heterogeneous Toeplitz structure
TOEP	Toeplitz structure
WHO	World Health Organization
YMRS	Young Mania Rating Scale

2. INTRODUCTION

This document describes the rules and conventions to be used in the presentation and analysis of efficacy, safety, and pharmacokinetic (PK) data for Protocol ITI-007-402. It describes the data to be summarized and analyzed, including specifics of the statistical analyses to be performed.

This statistical analysis plan (SAP) is based on protocol version 1.6, dated 23Mar2020. Exploratory analyses of biomarkers will be described in a separate analysis plan.

3. STUDY OBJECTIVES

3.1. PRIMARY EFFICACY OBJECTIVE

The primary objective of this study is to compare the efficacy of 2 doses of ITI-007 adjunctive to treatment with lithium or valproate, administered orally once daily, to that of placebo adjunctive to treatment with lithium or valproate as measured by mean change from baseline to Day 43 in the total score on the rater-administered Montgomery-Åsberg Depression Rating Scale (MADRS) in patients with bipolar dipression.

3.2. SECONDARY OBJECTIVES

3.2.1. KEY SECONDARY EFFICACY OBJECTIVE

The key secondary objective of this study is to compare the efficacy of 2 doses of ITI-007 adjunctive to treatment with lithium or valproate, administered orally once daily to that of placebo adjunctive to lithium or valproate as measured by mean change from baseline to Day 43 in Clinical Global Impression Scale-Bipolar Version-Severity (CGI-BP-S) Depression score in patients with bipolar depression.

[redacted]

3.3. SAFETY OBJECTIVES

The safety objectives of this study are to determine the safety and tolerability of 2 doses of ITI-007 adjunctive to treatment with lithium and valproate via:

- Incidence of adverse events (AEs);
- [redacted]

[redacted]

[redacted]

4. STUDY DESIGN

4.1. GENERAL DESCRIPTION

Study participation will last approximately 10-12 weeks (up to 9 visits) and will include the following periods: a *Screening Period,* lasting up to 2 weeks (unless medical monitor approves longer); an *On-Treatment Period* of 6 weeks of daily administration of study medication, and a *Safety Follow-up Period* of approximately 2 weeks after last dose of study medication (or within 1 week of early termination, if possible). The timing and assessments during each study period are described in the Schedule of Events (Table 6-1 of the protocol and Appendix 1 of this SAP).

Approximately 520 patients (173 patients/ treatment group) will be randomized in a 1:1:1 ratio to receive one of three study treatments: 40-mg ITI-007, 60-mg ITI-007, or placebo at approximately 93 global sites. The sample size estimation assumes a common dropout rate of 15% at Day 43, and a common within-patient correlation of 0.7 for the change from baseline in MADRS total scores at Days 8, 15, 22, 29, 36 and 43. This will provide approximately 90% power to detect an effect size of 0.4.

The primary analysis strategy will be based on overall ITT patients and will follow the study protocol.

An alternative analysis strategy specific for Non-US submissions will be considered, details on this approach can be found in Appendix 4 of this SAP.

4.2. [redacted]

[redacted]

Sections 7.5.4 and 7.5.6 of the protocol discuss adding the treatment-by-site (or pooled site)

Section 7.5 of the protocol discusses the pooled site algorithm, "If the primary efficacy analysis model presents convergence issues due to the too small number of patients per site, including pooled sites, the same site pooling algorithm will be applied again, but this time pooling sites with fewer than a pre-specified larger number of ITI-007 patients per site within each country or geographic region, as described in the SAP." Since there are efforts using different covariance structures if MMRM failed to converge and the failed to converge issues are often due to subject level data, not site level, the pooling algorithm will be considered final once there are 2 ITT patients per treatment group for each site. Section 7.2 of the protocol specifies the fixed sequence of testing for the primary and key secondary endpoints. The order in this plan has been updated to prioritize the 60 mg vs placebo analyses for the primary and key secondary endpoint first. If both of those endpoints are significant, testing will proceed to the 40 mg vs placebo analyses. More details can be found in section 7.2.

5. PLANNED ANALYSES

5.1. INTERIM ANALYSIS

No interim analyses will be performed.

5.2. FINAL ANALYSIS

The analyses detailed in this SAP will be performed by [redacted]

5.3. [redacted]

A blinded sample size re-estimation may be performed when 70% of randomized patients have [redacted]

6. ANALYSIS SETS

Analysis of efficacy and safety endpoints will be performed based on the analysis sets defined in this section and as specified for each endpoint throughout this SAP. The inclusion/exclusion of patients from each analysis set will be determined prior to the final analyses (and approved by ITCI) based on blinded data review.

Unless otherwise specified, the ITT Set will be used for efficacy endpoints and the SAF Set will be used for safety endpoints.

6.1. ALL PATIENTS ENROLLED [ENR] SET

The All Patients Enrolled (ENR) Set will contain all patients who provide informed consent for this study.

6.2. ALL PATIENTS RANDOMIZED [RND] SET

The All Patients Randomized (RND) Set will contain all patients who signed informed consent and were randomized to study medication.

For analyses and displays based on RND, patients will be classified according to randomized treatment.

6.3. INTENT-TO-TREAT [ITT] SET

The Intent-to-Treat (ITT) Set will contain all randomly assigned patients who received at least one dose of study medication and who had a valid (pre-dose) baseline measurement and at least one valid post-baseline measurement of rater-administered MADRS total score. All analyses using the ITT Set will classify patients according to randomized treatment, regardless of the treatment received during the course of the study.

6.4. SAFETY ANALYSIS [SAF] SET

The safety analysis (SAF) Set will contain all patients who received at least one dose of study medication. All analyses using the SAF Set will classify patients according to treatment actually received. In case of incorrect dosing, the treatment most often received will be taken as the actual treatment.

[redacted]

7. GENERAL CONSIDERATIONS

Relative Study Day will be calculated from the date of Day 1, which is the day of first treatment with double-blind study medication, and will be used to show the start and/or stop days of treatment, study procedures and assessments, prior, concomitant, and post-treatment medications, and adverse events.

- If the date of the treatment, procedure, or event is on or after Day 1 date then:

Relative Study Day = (date of variable of interest – Day 1 date) + 1.

- If the date of the treatment, procedure, or event is prior to the Day 1 date then:

Relative Study Day = (date of variable of interest – Day 1 date).

If the date of the treatment, assessment, or event is partial or missing, Relative Study Day, and any corresponding event durations will appear missing in the listings.

Analyses presented by visit or study day will be based on the protocol defined scheduled visits including scheduled, unscheduled and early discontinuation visits. Visit windows for unscheduled visits and early discontinuation visits are defined in Table A, which provides the mapping of relative day ranges to the scheduled target days and the study periods. If more than one assessment is available in the same 'Range of Relative Study Days' (window), the assessment closest to the Scheduled Target Day will be selected and assigned to the Scheduled Target Day. If two or more assessments are available in the same window and are equidistant from the Scheduled Target Day, the latest assessment will be selected. The assessment used in the by-visit summary may differ from the assessment to be carried forward for the LOCF methodology, described in Section 7.1, since the latest assessment per visit window will be carried forward for the LOCF analyses.

Listings will include scheduled, unscheduled, retest and early discontinuation data.

Table A: Mapping of Relative Day Ranges to Schedule Target Day

Study Phase	Study Period	Range of Relative Study Days	Scheduled Target Day	Scheduled Study Week
Pre-Treatment	Screening	1-14* days before Day 1 date	Day -14	-2
Pre-Treatment	Baseline	Day 1 date (definition of baseline varies by assessment – see descriptions below)	Day 1	0
Study Treatment	On-Treatment	2 to 11 days relative to Day 1 date	Day 8	1
Study Treatment		12 to 18 days relative to Day 1 date	Day 15	2
Study Treatment		19 to 25 days relative to Day 1 date	Day 22	3
Study Treatment		26 to 32 days relative to Day 1 date	Day 29	4
Study Treatment		33 to 39 days relative to Day 1 date	Day 36	5
Study Treatment		≥40 days relative to Day 1 date and after the last dose of treatment and before the start of the Safety Follow-up Period (one day after actual Day 43 visit)	Day 43	6
Post-treatment	Safety Follow-up	> 43 days relative to Day 1 date for treatment completers (42 day on-treatment) and after Day 43 post-treatment assessments or after early discontinuation visit (for patients who terminated early).	Day 57	8

*Not to exceed 28 days

First treatment with study medication and baseline assessments are scheduled for Visit 2 on Relative Study Day 1. For analysis purposes, baseline is defined as the last non-missing pre-treatment measurement.

Assessments will be considered baseline if the assessment date is before the date of the first treatment of the 6-week On Treatment Period or if the assessment was done on Study Day 1 and, according to the Study Schedule of Events, was supposed to be performed on Day 1, prior to treatment. Rater assessments (MADRS, [redacted] CGI-BP-S, [redacted] originally scheduled for Day -1 that are rescheduled for Day 1 and are confirmed to have been conducted prior to first dose will be considered baseline.

Unless otherwise specified, the following calculations will be used for change from baseline [redacted] of quantitative measurements:

Change from baseline will be calculated as:

- Value at Visit X – Baseline Value

[redacted]

All investigative sites with fewer than 2 ITT patients per treatment group will be pooled as follows: The largest site with fewer than 2 ITT patients per treatment group will be pooled with the smallest site with fewer than 2 ITT patients per treatment group within the same country or geographic region. If this results in a pooled site still having fewer than 2 ITT patients per treatment group, this site will be pooled together with the next smallest investigative site within the same country or geographic region, if one exists; otherwise, no further pooling is needed. Sites with the same number of ITT patients will be ordered in ascending order of their numerical site identification number. This will serve as a tie-breaker rule in case multiple sites have the same number of ITT patients. Should the primary efficacy analysis model present convergence issues, after testing the sequence of correlation structures, then the site effect will be reconsidered and may be dropped from the model. These pooled investigative sites, as determined based on the primary efficacy response variable, will be used for any analysis that has investigative site as a fixed effect in the model. The actual investigative site numbers will be included in the listings. For sensitivity analyses where site is a factor, additional pooling will be performed starting from the ITT pooled sites. Sites with patients who are in SEN Set but not ITT will be pooled, using the same algorithm as outlined above.

The default significance level for statistical tests will be 5%; confidence intervals will be 95% and all tests will be two-sided, unless otherwise specified in the description of the analyses. All analyses will be conducted using SAS® version 9.4 or higher.

7.1. MISSING DATA

The total and subscale scores of any assessment with more than one item, such as rater-administered MADRS, will be derived from individual items. Any individual missing item in any scale will not be imputed. If one or more items are missing at a visit, then the associated total

score or subscale score will be set to missing. A sensitivity analysis will be explored which imputes the missing MADRS total score when 2 or fewer individual items are missing. Specifically, if there are more than 2 items with missing scores, the MADRS total score will be set to missing; when <=2 items with missing score, the MADRS total score at a particular visit will be derived as follows:

MADRS total score = 10*(sum of non-missing item scores)/(number of non-missing items)

Change from baseline in the rater-administered MADRS total score will be evaluated for the ITT Set, using the same MMRM method specified for the primary analysis.

The main objective of the analyses of this study is to evaluate the effect of ITI-007 compared to placebo if the treatment is administered for the planned study duration. In order to evaluate this estimand in the presence of patients that may discontinue treatment prematurely, the primary efficacy analyses will be performed based on an assumption of data being missing at random (MAR). This implies that patients discontinuing from treatment early are considered to have unobserved values similar to the observed ones in their treatment group for patients who have similar history, i.e., the distribution of the unobserved future outcomes for patients who had discontinued treatment is the same as the distribution of the observed outcomes for those who continued treatment, conditional on the available data prior to discontinuation.

The Mixed-Effects Model for repeated Measures (MMRM) will be used for the analysis of the primary efficacy endpoint. The MMRM approach does not impute missing data but is based on all patients included in the analysis set using all available longitudinal data (either complete or partial). This approach is based on the MAR assumption as described above.

[redacted]

7.2. MULTIPLE COMPARISONS/ MULTIPLICITY

The study is designed to evaluate the primary and key secondary efficacy endpoints of two doses of ITI-007, 60 mg and 40 mg versus placebo, which involves multiple comparisons. A fixed sequence procedure will be used to control the overall Type I error rate.

The primary and key secondary efficacy endpoints will be tested in the following order:

1) ITI-007 60 mg vs placebo comparison change from baseline in MADRS total score at Day 43;

2) ITI-007 60 mg vs placebo comparison change from baseline in CGI-BP-S Depression score at Day 43;

3) ITI-007 40 mg vs placebo comparison change from baseline in MADRS total score at Day 43;

4) ITI-007 40 mg vs placebo comparison change from baseline in CGI-BP-S Depression score at Day 43

All tests will be performed at the 0.05 level. If at any step a hypothesis test does not show statistical significance, then all hypotheses at subsequent steps will not be tested for significance and their p-values will be reported for descriptive purposes only.

There will be no multiplicity adjustments for the comparison of other secondary or exploratory endpoints.

[redacted]

[redacted]

8. OUTPUT PRESENTATIONS

The tables, listings and figures (TLFs) shells provided with this SAP describe the presentations for this study and therefore the format and content of the summary TLFs to be provided by IQVIA Biostatistics.

Continuous variables will be summarized using descriptive statistics (number of patients [n], mean, standard deviation [SD], median, minimum, and maximum, unless otherwise stated). The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean and median will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database.

Categorical variables will be summarized using frequency counts and percentages. Unless otherwise stated, the calculation of percentage will be based on the number of patients in the analysis set of interest.

All p-values will be presented to four decimal places. P-values less than 0.0001 will be presented as "<0.0001".

Confidence intervals (CIs) will be presented to two more decimal places than the raw data.

Source data for summary tables and statistical analyses will be presented as patient data listings.

9. DISPOSITION AND WITHDRAWALS

Patient disposition and withdrawals, will be presented for the ENR Set.

Patient disposition and withdrawals will be presented by treatment group, when applicable, and overall. The number and percentage of patients who were screened, screen failed, randomized, completed or discontinued the 6-week On-Treatment Period, reasons for early

treatment discontinuation, completed or withdrew from the study, and reasons for study withdrawal will be presented. For patients who completed the treatment period, the number who enter safety follow-up will be summarized. [redacted]

[redacted]

A patient is defined to have completed the treatment period if the patient has completed the 6-week On-Treatment Period and procedures. Completion of treatment period is recorded on the End of Treatment CRF page.

A patient is defined to have completed the study if the patient has completed the 6-week On-Treatment Period and the End-of-Study assessments on Study Day 57 (±2). Completion of study is recorded on the Completion/Withdrawal from Study CRF page.

Table B: Reasons for Study Withdrawal and Study Medication Discontinuation Terminology

eCRF Terminology
Adverse event
Death
Lack of efficacy
Lost to follow-up
Protocol violation
Physician decision
Screen Failure
Study terminated by sponsor
Patient withdrew consent: Personal or family reasons Refused to provide a reason and refused all End of Study procedures Self-reported Adverse event

Self-reported Lack of efficacy
Other

Adverse event preferred terms associated with worsening of bipolar depression will be identified by the Medical Monitor review prior to database lock. The number and percentage of randomized patients who discontinued due to an adverse event associated with worsening of bipolar depression will be summarized. The number and percentage of randomized patients discontinued due to an adverse event not associated with worsening of bipolar depression will also be presented.

The time to treatment discontinuation due to any reason, adverse events (broken down by association with worsening of bipolar depression as well), lack of efficacy, withdrawn consent, or due to any other reason of special interest will be evaluated separately using the Kaplan-Meier method. Time to treatment discontinuation will be defined in days for those patients who are randomized and receive at least one dose of study medication but discontinue study medication prior to Day 43 as the date of discontinuation minus date of first dose of study medication plus 1. Patients who complete the On-Treatment Period or who discontinue for a reason other than the one being evaluated will be censored.

The number of patients randomized by site will be summarized by treatment group and overall for all randomized patients.

[redacted]

11. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographic data and other baseline characteristics will be presented for the ITT Set and SAF Set.

No formal statistical testing will be carried out for comparing demographic or other baseline characteristics between treatment groups.

The following demographic and other baseline characteristics will be reported for this study:

Demographics

- Age (years) calculated as (Date of Informed Consent – Date of Birth)/365.25;
- Gender;

- Race;
- Ethnicity;

Other Baseline Characteristics

- Waist circumference (cm), if collected in inches then Waist circumference (cm) = Waist Circumference (in) * 2.54 ;
- Weight (kg), if collected in lbs then Weight (kg) = Weight (lbs) * 0.4536;
- Height (cm), if collected in inches then Height (cm) = Height (in) * 2.54;
- Body Mass Index (BMI) (kg/m^2), where BMI (kg/ m^2) = weight (kg) / height $(m)^2$

[redacted]

Bipolar Disorder Baseline Efficacy

- Baseline efficacy parameters, including rater-administered MADRS total score, CGI-BP-S Depression score, [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

15. EFFICACY OUTCOMES

15.1. PRIMARY EFFICACY

15.1.1. PRIMARY EFFICACY ENDPOINT AND DERIVATION

The primary efficacy variable is change from baseline in rater-administered MADRS total score to Week 6 (Day 43). The MADRS is a 10-item checklist designed to measure the overall severity of depressive symptoms. The MADRS individual items are rated by a qualified site rater at each site on a scale of 0 to 6 in which a score of 6 represents the most severe symptoms for each item assessed. The MADRS total score is the sum of all 10 items and ranges from 0 to 60. If one or more items are missing at a visit, the total score will also be set to missing for that visit.

In addition to the rater-administered MADRS assessment, a self-administered MADRS assessment will be conducted. At screening, the self-administered MADRS assessment will be conducted prior to the rater-administered one. At all other visits, the rater-administered MADRS assessment will be conducted first. Self-administered MADRS scores will be compared to the ones from the rater-administered MADRS assessment on an ongoing basis as part of a remote quality assurance program.

15.1.2. MISSING DATA METHODS FOR PRIMARY EFFICACY ENDPOINT

The primary analysis method for evaluating the primary efficacy endpoint is the likelihood-based analysis of repeated measures, MMRM, described in details in Section 15.1.3. The use of MMRM inherently implies that the treatment effect on the change from baseline in the rater-administered MADRS total score will be similar for the patients who withdraw and for those who complete the study in their respective treatment groups, conditional on the outcomes observed prior to withdrawal (MAR assumption). To assess the robustness of the MAR assumption, sensitivity analyses which utilize multiple imputations and a different assumption about unobserved outcomes will be performed, as detailed in Section 15.1.4.

15.1.3. PRIMARY ANALYSIS OF PRIMARY EFFICACY ENDPOINT

The treatment effect on the primary efficacy endpoint will be evaluated using an MMRM model for the ITT Set. The model will include the change from baseline at each pre-specified timepoint as the response variable and study visit, the adjunctive treatment and bipolar disorder stratification variables, baseline MADRS total score, site (or pooled site), baseline MADRS total score-by-study visit interaction, treatment (ITI-007 40 mg, ITI-007 60 mg, placebo), and

treatment-by-study visit interaction. An unstructured covariance matrix will be used to model the correlation among repeated measurements within patient. In the event the convergence cannot be attained with the unstructured correlation matrix, the following alternative structures will be attempted in the specified order: heterogeneous Toeplitz structure (TOEPH), heterogeneous autoregressive(1) (ARH(1)), heterogeneous compound symmetry (CSH), No Diagonal Factor Analytic (FA0(q), with q equal to the number of time points), Toeplitz structure (TOEP), autoregressive(1) (AR(1)), and compound symmetry (CS). The Kenward-Rogers method will be used to estimate the denominator degrees of freedom. The treatment and treatment-by-time interaction terms allow for comparisons of the treatment groups at each of the following time points: Weeks 1, 2, 3, 4, 5, and 6 (Days 8, 15, 22, 29, 36, and 43). Treatment differences will be evaluated via contrasts for the time-by-treatment factor. This methodology will be used to compare the treatment groups versus placebo for change from baseline at each time point.

Estimates of model parameters will be presented, as well as least-squares mean (LSM) estimates for change from baseline in MADRS score, standard errors and 95% confidence intervals (CI) for LSMs will be presented by treatment group and time point. Contrast estimates (LSMs) for between-group comparisons (ITI-007 60 mg vs. placebo and ITI-007 40 mg vs. placebo), the corresponding standard errors, 95% CIs, effect sizes, and p-values will be presented for each visit. Effect size will be calculated for each ITI-007 dose group (vs. placebo) as $\frac{\text{LS Mean Difference}}{\text{Pooled estimate of within patient error standard deviation}}$. For the primary and key secondary efficacy endpoints, the effect size will be obtained.

The empirical cumulative distribution function of the percent change from baseline in MADRS total score at Day 43 will be provided. The horizontal axis corresponds to increasing levels of improvement, with a vertical line anchoring an improvement of 50%. The vertical axis corresponds to the percent of patients achieving each level of response. Patients discontinuing the study are considered non-responders and represented by the left-most portion of the graph. This responder analysis corresponds to the utility estimand (Permutt, 2016) as it incorporates both patients discontinuation status as well as the improvements in MADRS total score.

[redacted]

15.2. KEY SECONDARY EFFICACY

The key secondary endpoint of the study is the change from baseline to Day 43 in CGI-BP-S Depression score. The CGI-BP-S depression score ranges from 1= "Normal, not ill" to 7= "Very severely ill". It will be used in this study at screening (as a criterion for inclusion or exclusion), at baseline, and throughout the study, at Days 8, 15, 22, 29, 36, and 43 as a measure of efficacy, and at Day 57 as a safety measure. A higher score is associated with greater illness severity.

The treatment effect on the key secondary efficacy endpoint will be evaluated using an MMRM model. The model will include the change from baseline at each pre-specified timepoint as the response variable and study visit, the adjunctive treatment and bipolar disorder stratification variables, baseline CGI-BP-S depression score, site (or pooled site), baseline CGI-BP-S depression score-by-study visit interaction, treatment (ITI-007 40 mg, ITI-007 60 mg, placebo), and treatment-by-study visit interaction. An unstructured covariance matrix will be used to model the correlation among repeated measurements within patient.

The key secondary efficacy analysis will be performed for the ITT Set.

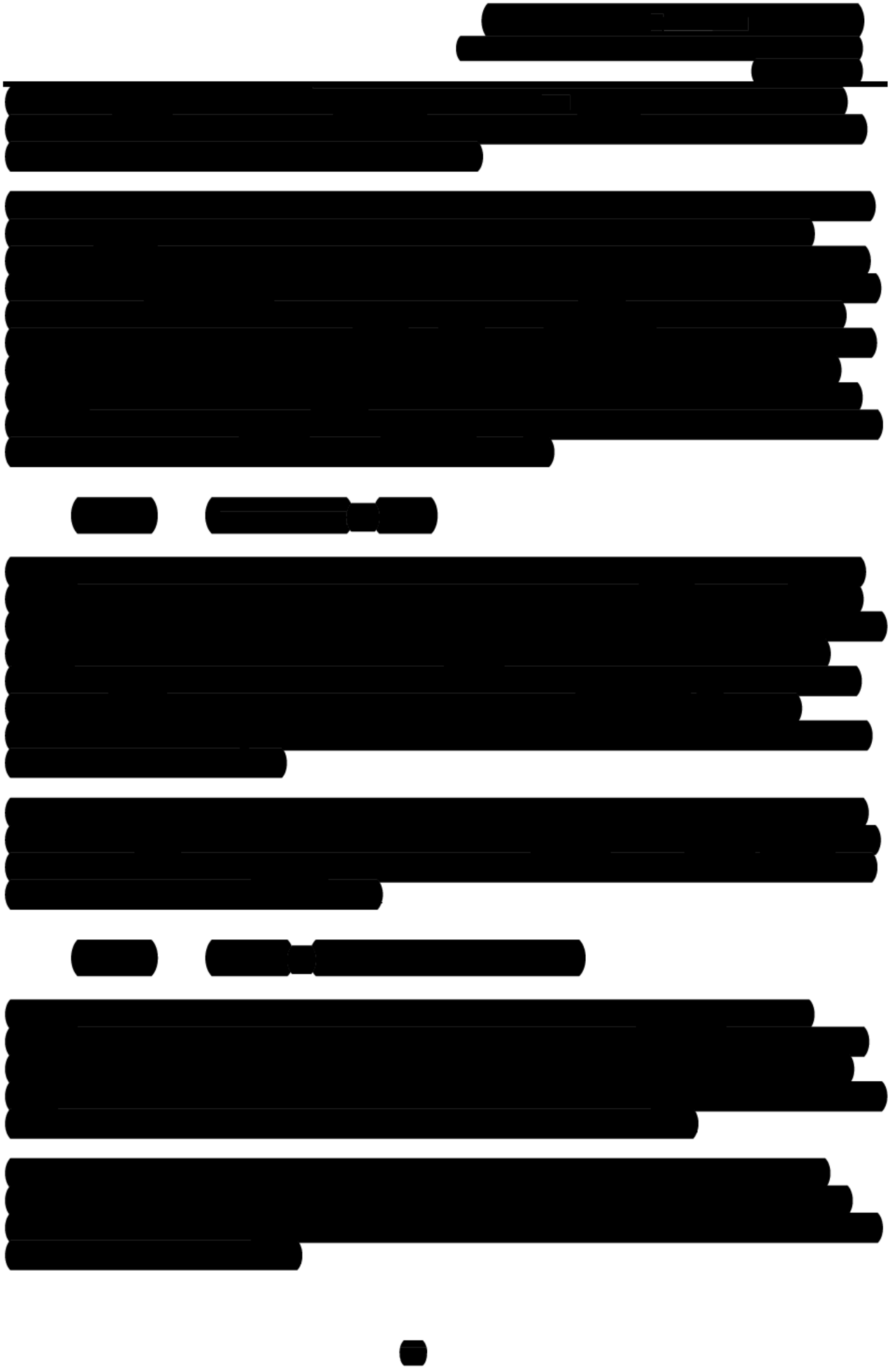
A fixed sequence of testing, as described in section 7.2, will be implemented to control the overall type 1 error rate.

The change from baseline to visits other than Week 6 (Day 43) will be analyzed as part of the secondary analyses and will be performed for the ITT Set. Additionally, as a supportive analysis, the change from baseline to Day 43 in CGI-BP-S Depression score will be evaluated using analysis of covariance (ANCOVA) model with last observation carried forward (LOCF) based on the ITT Set. The model will contain the main effects of treatment group, site (or pooled site), baseline CGI-BP-S Depression score, and the adjunctive treatment and bipolar disorder stratification variables. LSMs for each treatment, LSM difference between treatment groups, the associated standard errors and 2-sided 95% Cis for the differences between treatment groups, and the p-values for between-treatment tests of differences will be presented.

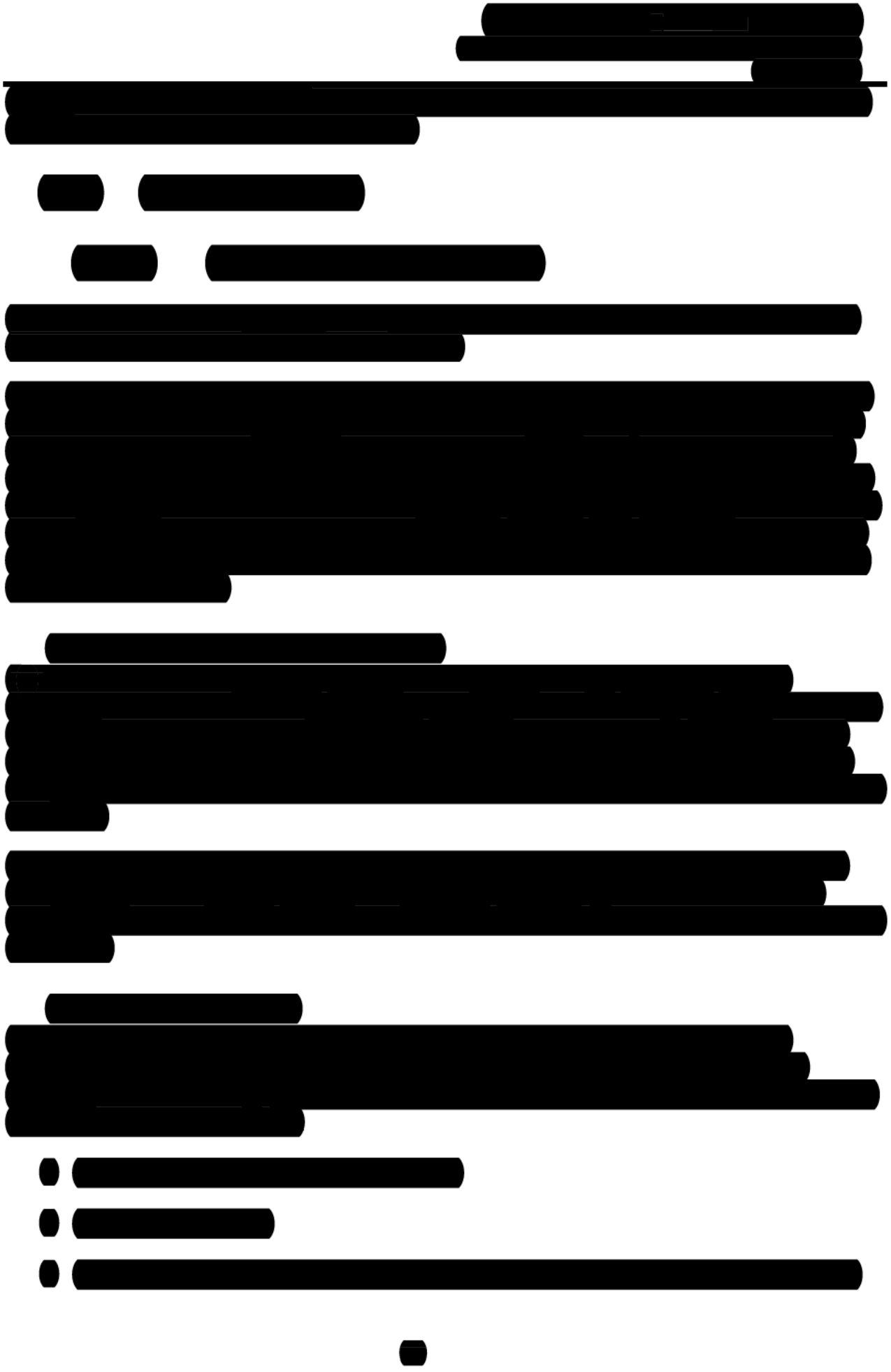
If at least one patient (or any site personnel) are accidentally unblinded during the study, a

sensitivity analysis will be presented, excluding the patient(s) CGI-BP-S Depression score and repeating the analysis.

[redacted]



[redacted]



[redacted]

16. SAFETY OUTCOMES

All outputs for safety outcomes will be based on the Safety Set.

For by visit summaries of change from baseline summaries, the last value on treatment (last assessment on or before the last dose of study medication plus one day) will be presented and labeled as 'Day 43/ET'.

16.1. ADVERSE EVENTS

Adverse Events (AEs) will be coded using MedDRA central coding dictionary currently in effect at the time of final database lock.

Treatment emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the date of the first dose of study medication and on or before the date of last dose of study medication plus one day. Consecutive AEs with the same Preferred term and differing severity will be assessed. AEs are considered to be consecutive if the end date of the initial AE and the start date of the following AE are at most 1 day apart. If the severity of later AE stays the same or decreases after the first dose of study medication, then it will not be considered treatment emergent.

All AEs with an onset date after the last dose of study medication plus one day will be listed in the AE data listing and labelled as 'Follow-up Adverse Event'.

See Appendix 2 for handling of partial or completely missing dates for AEs. In the case where it is not possible to define an AE as treatment emergent or not, the AE will be classified as the worst case, i.e., treatment emergent.

An overall summary of the number of patients within each of the categories described in the sub-sections 16.1.1 and 16.1.2 below will be provided as specified in the TLF shells.

Listings will include all adverse events, TEAEs and Non-TEAEs.

16.1.1. AEs AND TEAEs

The incidence of TEAEs will be presented by SOC and PT and also broken down further by maximum severity and relationship to study medication.

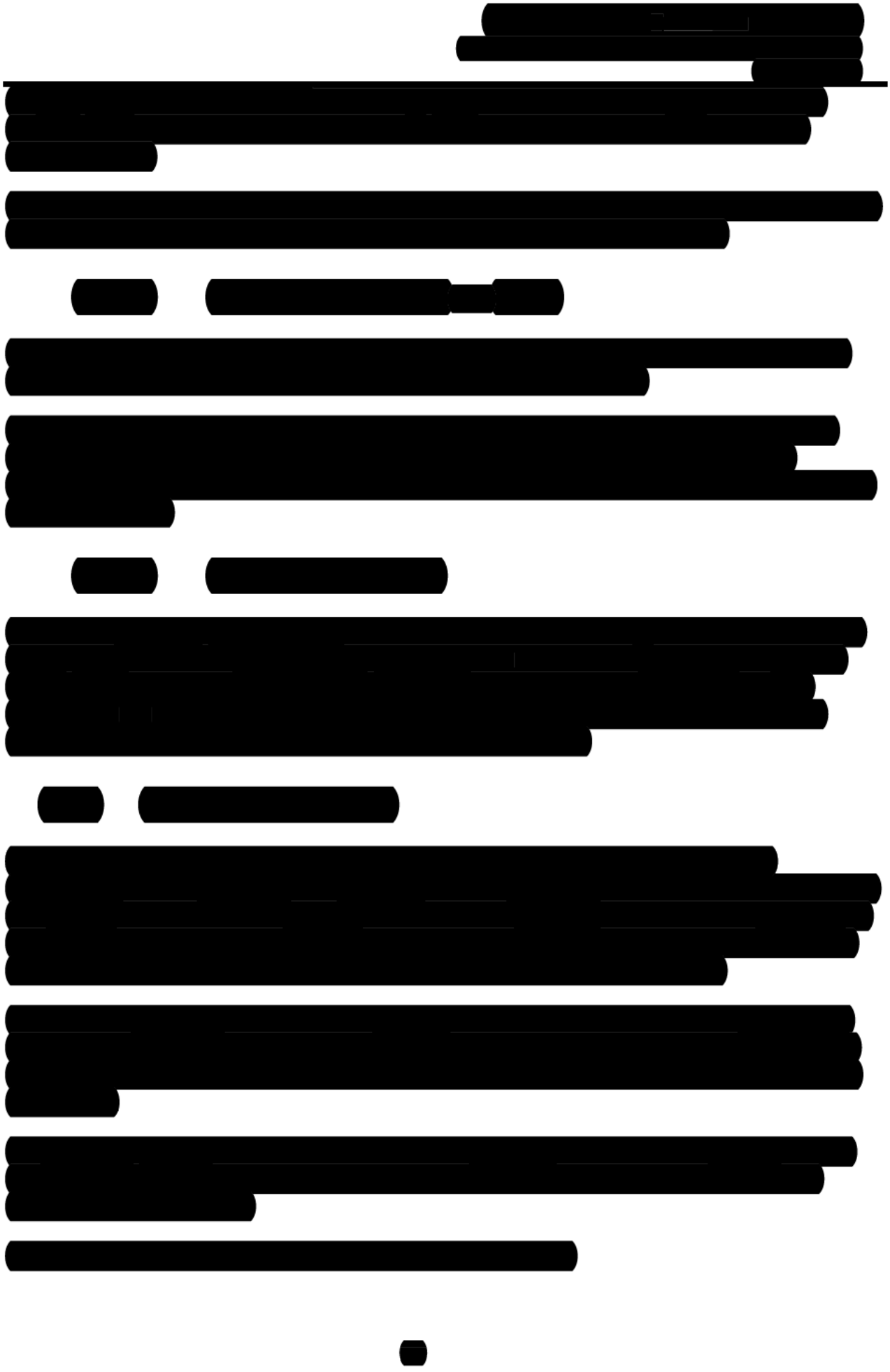
The number and percentage of patients with at least one TEAE and total number of patients having events for each PT and SOC will be summarized. A summary of TEAEs will be provided with only PTs and a separate summary for only SOCs. An additional summary of TEAEs will be provided for PTs occurring in at least 5% of patients in any treatment group (ITI-007 60 mg, ITI-007 40 mg, or placebo). This summary will be repeated for TEAEs related (possibly, probably, or definitely) to study drug. Within each patient, multiple reports of events that map to a common MedDRA PT and/or SOC will be condensed into a single AE for incidence counts. Summaries will be presented by treatment group and in decreasing frequency by decreasing dose group (ITI-007 60 mg, ITI-007 40 mg) and placebo group. AEs which occur during the follow-up period will be summarized by PT and SOC and treatment group.

The relative risk of at least one TEAE and for each PT and SOC will also be presented along with 95% CIs and p-values obtained by the Chi-square test for association.

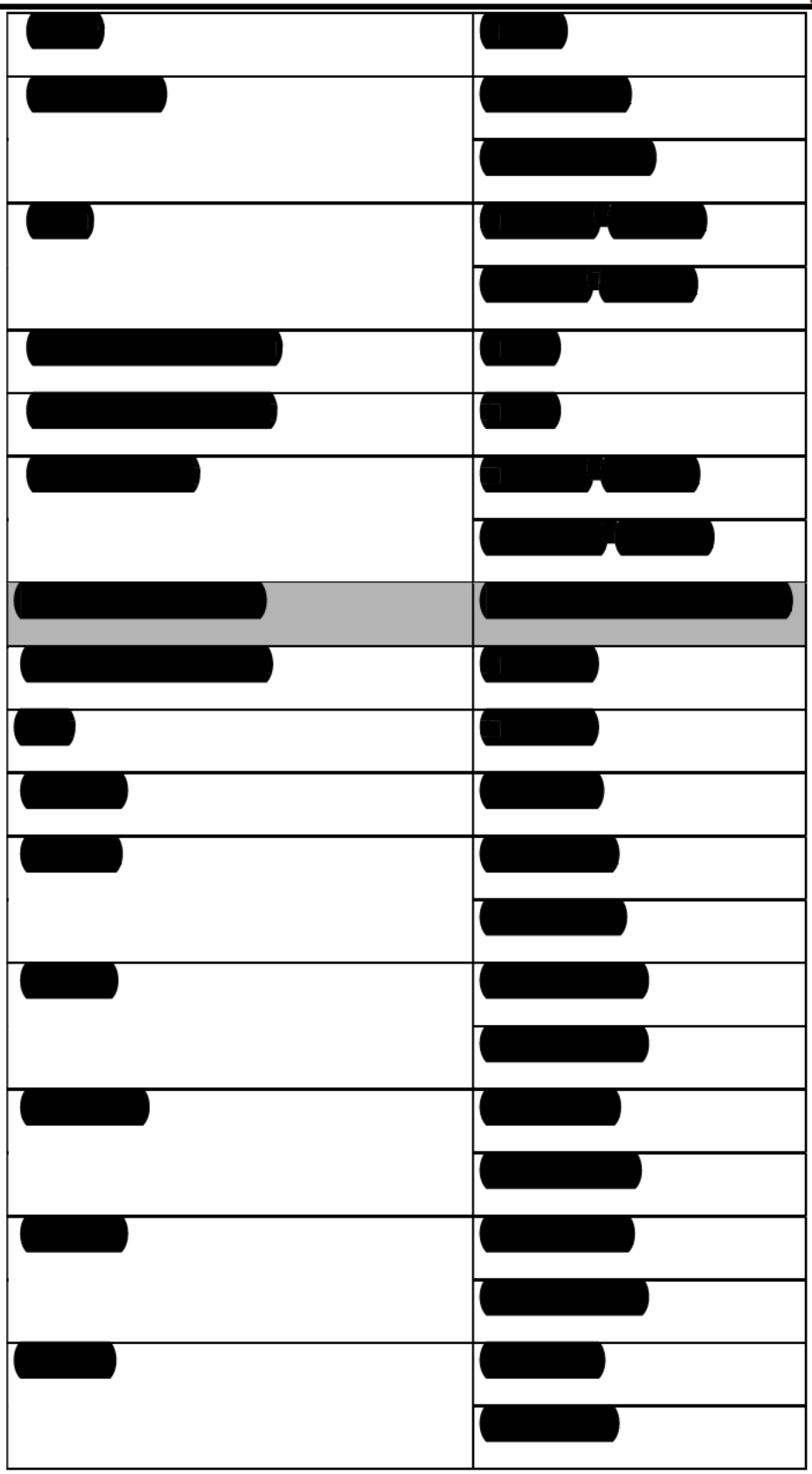
Intensity is classed as “mild”, “moderate”, “severe”, or “life-threatening” (increasing severity). AEs and TEAEs with a missing severity will be classified as “not specified”. If a patient reports a TEAE more than once within the same PT and SOC, the event with the worst case severity will be used in the corresponding severity summaries.

Relationship, as indicated by the Investigator, is classed as “unrelated”, “unlikely to be related”, “possibly related”, “probably related”, or “definitely related” (increasing severity of relationship). A “related” TEAE is defined as a TEAE with a relationship to study medication as “possibly related”, “probably related”, or “definitely related” to study medication. TEAEs with a missing relationship to study medication will be classified as “related”. If a patient reports the same AE more than once within the same PT and SOC, the AE with the worst case relationship to study medication will be used in the corresponding relationship summaries. A summary of related TEAEs by SOC and PT will be presented.

AEs leading to early withdrawal or discontinuation of study medication will be identified by using the Adverse Events page of the eCRF, where item ‘Action taken with study treatment’ indicates permanent discontinuation of study medication, i.e., “Drug withdrawal”. AEs leading



[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

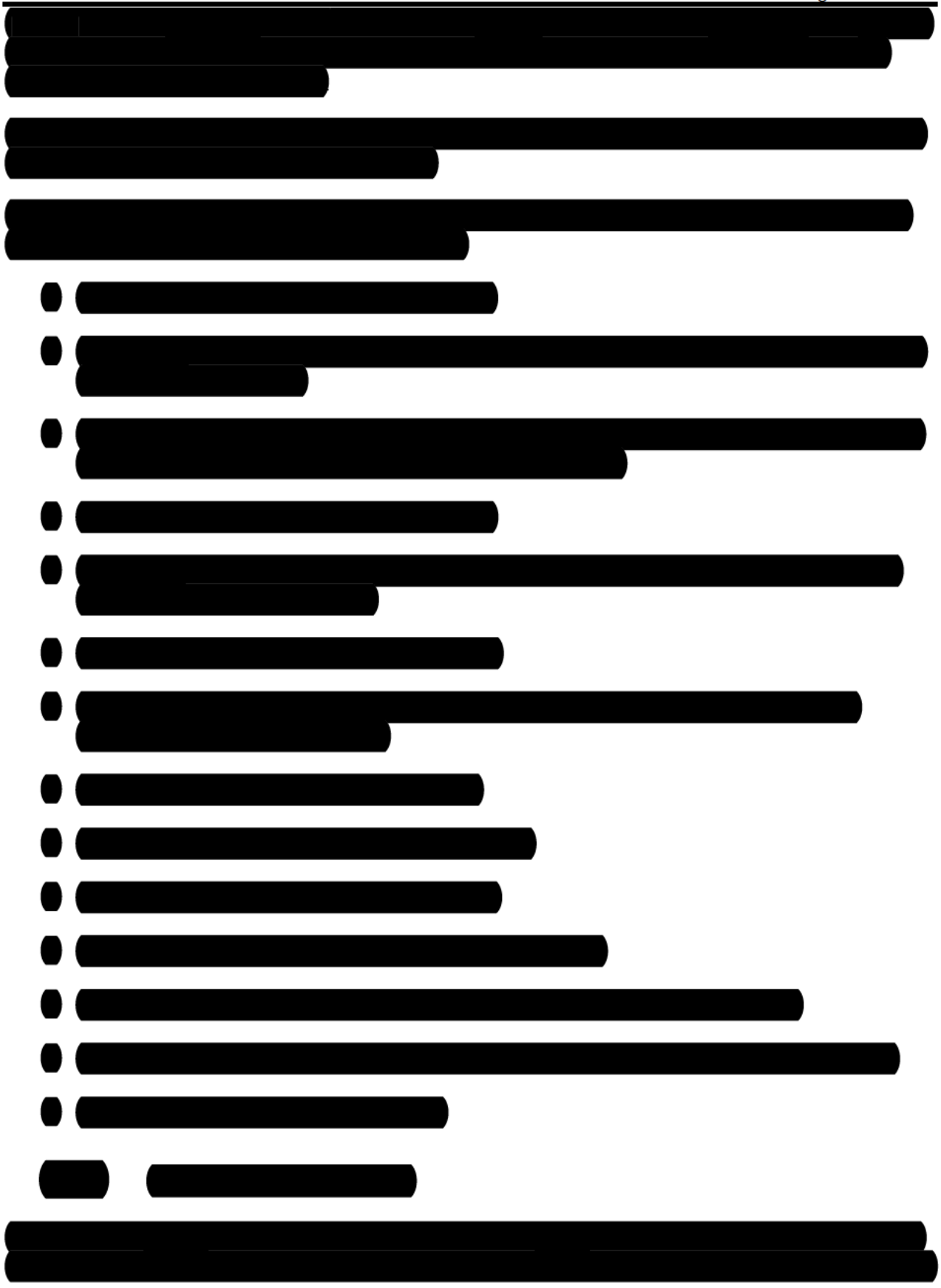
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

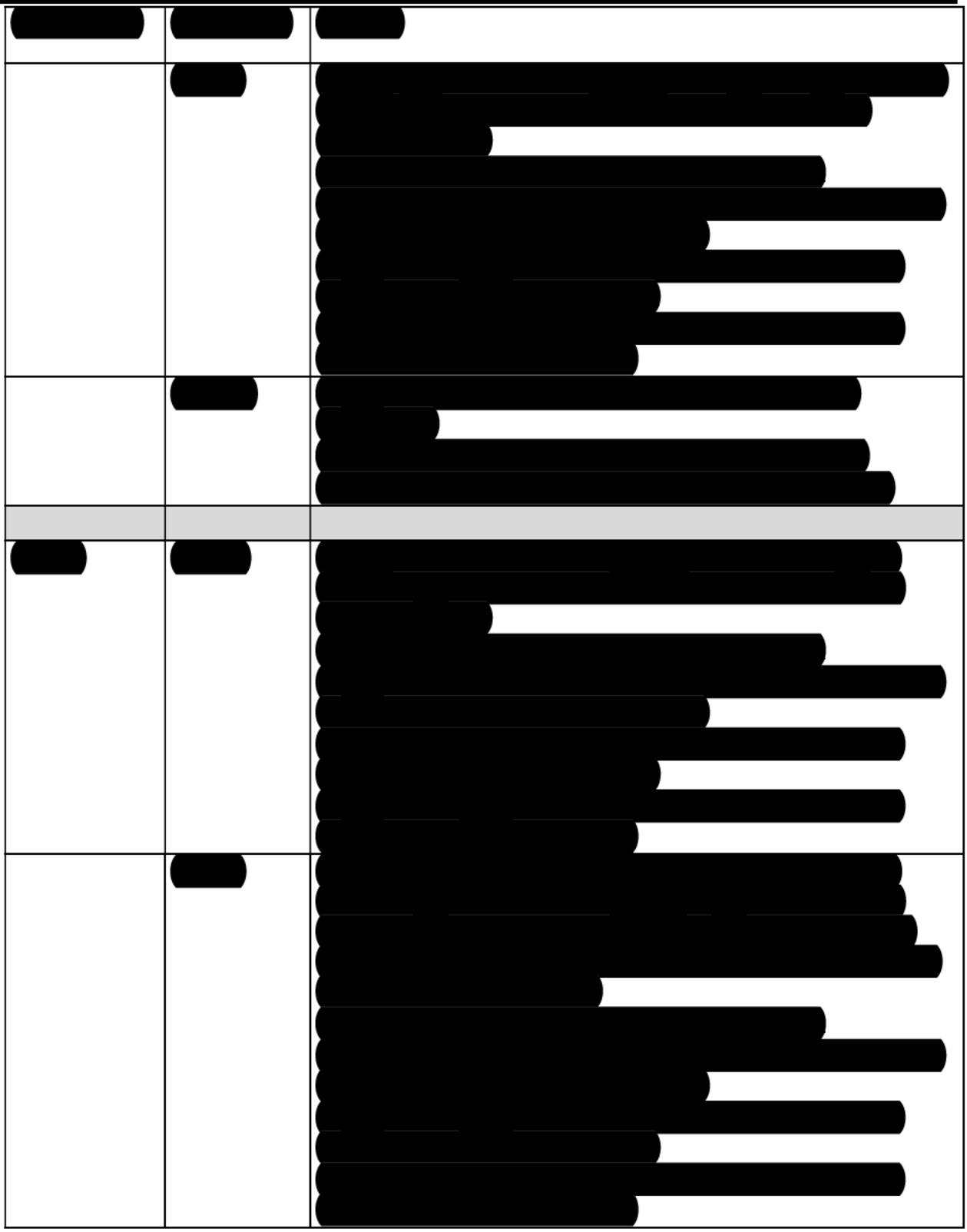
[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]



[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
	[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

==========

[redacted]