MAVERIC Study Specific Documents	DATE OF ISSUE	DOCUMENT DESCRIPTION
	4/1/2016	This document describes the statistical analysis plan for the CSP590 [Li+] randomized clinical trial.

AUTHOR: Soe Soe Thwin, MS, PhD **SUBMITTED BY:** Soe Soe Thwin, MS, PhD

DOCUMENT/RECORD HISTORY:

DATE	VERSION	RECORD HISTORY
04/01/2016	1.0	Initial Version

AUTHORIZATION SIGNATURES

NAME	FUNCTION TITLE	SIGNATURE	DATE
Ira Katz MD, PhD	Study Chairman		
Matthew Liang, MD	Study Director		
Soe Soe Thwin, MS, PhD	Study Biostatistician		
Ryan Ferguson , MPH, ScD	Center Director		

Table of Contents

1. Introduction

This document outlines the statistical methods for the analysis of data collected in the Department of Veterans Affairs Cooperative Studies Programs (CSP) study #590 entitled “A Randomized double-blind placebo-controlled trial of lithium augmentation and enhanced suicide prevention management in preventing suicide re-attempts in individuals with depression and bipolar disorder.” The purpose of this document is to provide guidelines from which the analysis will proceed. Deviations from these guidelines will be documented and filed electronically in the study central file Sharepoint site.

The following documents were used in preparation of this statistical analysis plan (SAP):

- Clinical Study Protocol CSP #590
- Case Report Forms (CRF) from above entitled Protocol
- Global CSP SOP 2.9: “Developing and Conducting Statistical Analyses
- Local Work Instruction WI 201Statistical Analysis Plan and Biostatistical Research Data Processing Plan Creation and Amendment
- Local Job Aid No. 5: Guide to Writing Statistical Analysis Plans

2. Overview of the Study Design and Study Objectives

This is a double-blind, placebo-controlled randomized clinical trial that will assess the effectiveness of lithium augmentation to enhanced usual care for prevention of repeated fatal and non-fatal suicide attempts in participants with bipolar disorder or depression who have survived a recent episode of suicidal self-directed violence.

A total of 1862 subjects recruited from 29 Veterans Health Administration (VHA) centers over a 3 year period are expected to be randomized equally to lithium (Lithium ER) or placebo. Participants will be followed actively for 12 months from randomization. Adverse event monitoring for safety will be for one additional month beyond active follow up, i.e., up to 13 months.

The primary hypothesis is that lithium is superior to placebo for the prevention of episodes of suicidal self-directed violence over time. The primary outcome measure is time from randomization for treatment allocation to any first episode of suicide attempt, interrupted attempt, hospitalization specifically to prevent suicide, or death from suicide.

As secondary objectives of the study, lithium efficacy will be evaluated for prevention of subtypes of suicidal self-directed violence, for all suicidal self-directed violence events (even after the first recurrence), for prevention of repeated events in subgroups as well as in the entire sample, and to identify potential mediators.

The tertiary objectives of the study are to extend the follow up period using electronically available data to describe patterns of lithium use following active participation in the study participants, as well as to evaluate rates and determinants of suicide reattempts and all cause mortality over a longer follow up period.

3. Investigational Plan

3.1. Description of the Study Population

Patients are eligible to enroll if they satistfy all of the 5 inclusion conditions and patients are excluded if they satisfy any of the 11 exclusion conditions as outlined in Table 1 below.

Table 1. Inclusion and Exclusion Criteria.

Inclusion Criteria:
(1) Survival of an episode of suicidal self-directed violence (including suicide attempts and interrupted attempts) that occurred **within three months of initial consent to the study**, or they were admitted to a mental health inpatient unit specifically to prevent suicide within three months of initial consent to the study
(2) Diagnosis of Bipolar I Disorder, Bioplar II Disorder, or current or recurrent Major Depressive Disorder. Meeting the diagnosis for a current Major Depressive Disorder requires the presence or history of a Major Depressive Epidosde
(3) Concurrence from the patient's mental health provider and the provider responsible for the pateint's general medical care about inclusion/exclusion criteria, and confimration of the providers' willingness to work with the research team in managing the patient throughout the study
(4) Able to provide informed consent
(5) Patient identifies one or more contacts and gives permission for the clinical providers and/or the research team to contact them if the patient cannot be reached.

Exclusion Criteria:
(1) Schizophrenia
(2) A history of psychotic symptoms in the absence of prominent mood symptoms attributable to Major Depressive or Manic Episodes, delirium, or substance use/withdrawal
(3) Six or more previous lifetime suicide attempts
(4) Enrollment in another randomized interventional clinical trial
(5) Lack of decision-making capacity to evaluate the risks versus the benefits of participation as determined by UBACC score <14
(6) Cognitive impairment defined by a score >10 on the OMC
(7) Current or recurrent (within the past 6 months) use of lithium, or a history of significant adverse effects of lithium
(8) Concurrent medications: a.Diuretics b.Angiotensin Converting Enzyme Inhibitors (ACE Inhibitors) c.Angiotensin II Receptor Antagonists d.Haloperidol e.Clozapine
(9) Unstable medical conditions or specific medical comorbidity: • Congestive heart failure by Framingham criteria (See form S12 for worksheet) • Estimated GFR < 60 mL/min/1.73m^2

QTc ≥450ms for men and ≥460ms for women
(10) Active substance abuse: • Active alcohol or opiate dependence requiring medically supervised withdrawal and stabilization • Active cocaine, methamphetamine, other stimulant, hallucinogen, or cannabis abuse requiring stabilization
(11)Any possibility of being pregnant, not on appropriate birth control, lactation and breastfeeding

3.2. Description of the Intervention Strategy

Eligible patients who consent to participate will be randomized to receive Lithium or placebo. Initital dosage will be 600 mg /day of lithium (or matching placebo), to be titrated to achieve serum lithium level of 0.6 to 0.8 meq/liter. Mock serum lithium level for placebo patients will be calculated using Theraputic Drug Monitoring method (Yukawa, 1993) based on age, weight, serum creatinine level, adherence, and side-effects. Titration visits will occur at intervals of 5-7 days over 3 months, with maximum allowed dosage not to exceed 1500mg.

Once the target serum lithium level has been reached the patient will be on steady state maintenance phase for the rest of the study with active participation and intervention terminating at 12 months after randomization.

3.3. Definition of Intention to Treat Primary Analytic Sample

All consented and randomized subjects will be accounted for and included in the primary intention to treat (ITT) analytic cohort. Those who revoked HIPAA authorization will be included in the analysis from randomization date to date of HIPAA revocation.

An analytic datafile consisting of only the “ITT” subjects as defined above will be created and maintained throughout the study. This file will be called the “CSP590_ITT” file.

3.4. Definition of per-Protocol sample

Adherence to allocated intervention will be ascertained for randomized subjects for identification of “per-Protocol” sub-group. Adherence will be determined by pill count for both lithium and placebo groups and additionally, by serum lithium level for the lithium group. These two methods are described below. Those in the lithium intervention group who are adherent to the protocol based on either of the two methods will be considered as adherent. All study subjects, who are adherent to the protocol, will be included in the “per-Protocol” analytic sub-sample.

Method 1: Based on pill count from Form T03 Study Drug Adherence Form data.

Percent adherence for each prescribed dosage will be computed using the following algorithm:

% adherence = (Total #tablets taken/Total #tablets prescribed)*100%

These percent adherences for each prescribed dosage will be averaged for each patient and those with >=80% overall adherence during the steady state follow up period will be considered adherent to the protocol.

Method 2: Based on target serum lithium level

Study participants in the lithium augmentation group who reach the target serum lithium level of 0.6-0.8 meq/liter during the titration period and maintain within target±0.1 meq/liter (i.e., 0.5-0.9 meq/liter) during the entire steady state follow up period will be considered adherent to the protocol.

4. Primary Analysis

4.1. Primary Objective

As the primary objective, this trial will test the hypothesis that lithium is superior to placebo for the prevention of suicidal self-directed violence. This will be accomplished by comparing event-free survival distribution in the two treatment groups.

4.2. Primary Endpoint

The primary outcome measure is time from randomization for treatment allocation to the first episode of suicidal self-directed violence assessed over one year follow up period. Episode of suicidal self-directed violence include (1) non-fatal suicide attempts, (2) interrupted attempts, (3) hospitalizations specifically to prevent suicide, and (4) deaths from suicide. All potential primary endpoints identified via CRFform data (mostly through the serious adverse event reporting or Columbia Follow Up form), will be evaluated by two independent adjudicators. If the adjudicators do not agree in deciding the event to be (or not to be) the primary endpoint event, then a third adjudicator will evaluate the same information and serve as a tie-breaker.

4.3. Statistical Methods for Primary Efficacy Analysis

Primary efficacy analysis will include the modified intention to treat sample as defined in section 3.3 above.

The primary hypothesis can be stated as follows.

Under the null hypothesis: The 1-year event rate is 15% and the two treatment groups do not differ in their time to event hazard rates.

Under the alternative hypothesis: The 1-year event rate is 9.45% for patients treated with lithium and 15% for patients in the placebo group, resulting in an absolute reduction from 15% to 9.45% = 5.55% and a relative reduction of (15 – 9.45)/15 = 37%. The group on lithium augmentation has a lower hazard rate than the placebo augmentation group with a hazard ratio for lithium therapy compared to control therapy lower than or equal to 0.61.

We will test this hypothesis with a two sided log rank test at 90% power and 4.8% Type I error (i.e., 5% overall error allows for one interim test at 0.2%). Null hypothesis will be rejected if the hazard ratio exceeds 1/0.61= 1.64 or falls below 0.61.

A significant difference showing that the lithium augmentation compared to placebo decreases the hazard of suicidal self-violence will be regarded as a positive result, whereas a result that shows lithium augmentation does not significantly differ from placebo or significantly increases the hazard rate relative to placebo will be regarded as a negative result.

Log rank test will be followed by Cox proportional hazards regression modeling to provide the hazard rates, their ratio and the 95% confidence interval about the ratio.
We will test the proportional hazards assumption by including a time*treatment interaction term in the model and use graphic methods to study goodness of fit as described in Anderson et. al., 1991.

The SAS code for the primary analysis is as follows, where, tx is coded as {1=lithium, 0=placebo}, event is coded as {1=adjudicated outcome, 0=otherwise}, and FUdays is counting number of days from randomization date to the first occurrence of (adjudicated outcome, death, 12 month date, or last study visit/contact date).

Log-Rank Test:

```
title "Survival Estimates by Month";
proc lifetest data=endpt plots=(s); time FUdays*event(0); strata tx; run;
quit;
proc lifetest data=endpt method=life plots=(ls) intervals=30 90 120 150
180 210 240 270 300 330 365; time FUdays*event(0); strata tx; run; quit;

title "Survival Estimates by Month";
proc lifetest data=endpt method=life plots=s intervals=30 90 120 150 180
210 240 270 300 330 365; time FUdays*event(0); strata tx; run; quit;

title "Hazard Estimates by Month";
ods output lifetableestimates=life(keep=stratum lowertime uppertime
hazard);
proc lifetest data=endpt method=life intervals=30 90 120 150 180 210 240
270 300 330 365; time FUdays*event(0); strata tx; run; quit;
data life; set life; where hazard ne .;
time=lowertime+(uppertime-lowertime)/2; keep stratum time hazard; run;
proc export data=life outfile="life.csv" dbms=CSV replace; run;

title "Basic model - Exponential";
proc lifereg data=endpt; model FUdays*event(0)=tx5fu / dist=exp; run;
quit;
```

Cox-Proportional Model:

```
*CHECK PROPORTIONAL HAZARDS FOR TREATMENT*;
title 'Basic model - Cox model with treatment*time interaction';
ods select ParameterEstimates;
proc phreg data=endpt; where FUdays>0;
```

```
      model FUdays*event(0)=tx int/ rl;
      int=tx*log(FUdays);
      Proportionality: test int;
run; quit;

proc phreg data=endpt; where FUdays>0;
      model FUdays*event(0)=tx;
      output out=propcheck ressch=schres;
run; quit;
proc sgplot data=propcheck; loess x=FUdays y=schres / clm; run;

title "Basic model - Cox model";
proc phreg data=endpt; model FUdays*event(0)=tx / rl;
run; quit;
```

4.4. Evaluating Covariate Effects and Heterogeneity by Site

First, we will identify potential confounding factors that are not balanced between the intervention groups and evaluate the possibilities of treatment by covariate effects. This will be done by refiting the primary cox proportional hazard model with a vector of covariates included in the model.

We will also fit a random effects or frailty model to investigate potential heterogeneity of effect due to site variability. This will be accomplished by refitting the primary cox proportional hazard model with additional classification and random statements.

```
title "Random Effects - Cox Model";
proc phreg data=endpt;
class site tx;
model FUdays*event(0)=tx / rl;
random site;
hazardratio "Frailty Model Analysis' site;
run; quit;
```

4.5. Sensitivity Analysis

The primary efficacy analysis will be repeated under 3 different scenarios as follows.

(1) Randomization was stratified on two factors: (1) Bipolar or Major Depressive Disorder diagnosis, and (2) prior history of suicide attempt. Cox Regression Model as described in section **4.3** will be refitted with the inclusion of diagnosis variable and suicide history variable in the model.
(2) Primary efficacy analysis will be repeated on the "per-Protocol" sample with adherence as defined in section **3.4** to evaluate efficacy among those adherent to the protocol.
(3) Primary efficacy analysis will be repeated on the modified ITT analytic cohort that excludes all randomized subjects who terminated prior to the first titration visit.

4.6. Handling of Missing Data in Primary Analysis

Up to 40% of subjects have been estimated to drop out over the course of follow up prior to end of study at 12 months after randomization. To account for complete data loss from attrition, evaluable sample size has been inflated by 20% accordingly. Distribution of subjects lost to follow-up and missing data on key variables across the treatment groups will be monitored throughout the study.

All those who terminated early will be censored at the date of last visit or last contact and their time in the study from randomization to termination date will be considered in the primary analysis.

The primary analysis assumes that data are missing at random. This assumption will be tested by evaluating the distribution of missing data across the two treatment groups. Additionally, a Baysan sensitivity analysis utilizing Pattern Mixture Modelling approach will be performed to determine potential bias in inference from all possible non-random patterns of missing data. (Rybin et. al., 2015)

5. Analysis Plan for Secondary Objectives

5.1. Secondary Objective

This trial has 5 secondary objects, which are described below. Analyses for these objectives are exploratory in nature and as such, they are not statistically powered for formal hypothesis testing. These objectives are to:

1) explore the effectiveness of lithium augmentation of enhanced suicide prevention management in preventing separately:
    a. all-cause mortality
    b. impulsive suicide reattempts
    c. non-impulsive suicide reattempt
    d. non-suicidal self-directed violence
2) explore heterogeneity of response to treatment subgroups including:
    a. categories of age, race/ethnicity, and gender
    b. diagnosis of depression vs. bipolar disorder
    c. impulsive index attempt vs. non-impulsive index attempt
    d. substance use vs. absence of substance use
    e. individuals with and without trait impulsivity
    f. individuals with and without trait aggression
    g. subjects who have vs. have not received evidence based psychotherapies
3) explore the effect of lithium augmentation of enhanced suicide prevention management on total occurrence of all event types and cumulative occurrence of individual event type as follows:
    a. all suicide attempts
    b. all hospitalizations for prevention of suicide attempts
    c. all self-directed violence
    d. all emergency department visits

4) explore longitudinally the effect of lithium augmentation of enhanced suicide prevention management on symptoms of bipolar disorder and depression and suicidal ideation

5) explore mediators of the effect of lithium augmentation on prevention of suicide reattempt, specifically improved control of symptoms and lithium plasma levels

5.2. Secondary Endpoints

There are multiple secondary endpoints being considered for the study where endpoint events for secondary objectives #1, 2, and 5 will be assessed as time to event outcomes, endpoint events for secondary objective # 3 will be assessed as count data, and endpoint events for Secondary Objective #4 will be evaluated as a binary outcome.

For Secondary Objective #1, Time from randomization to death from any cause will be analyzed. Additionally, self-directed violence as described in section **4.2** will be classified further as either impulsive or non-impulsive attempt based on Beck Suicide Intent Scale, (Question#15 on CRF-Form S8), and will be analyzed separately. Lastly, non-suicidal self-directed violence behavior defined as the primary endpoint wihout the death from suicide will be considered as a separate outcome.

For Secondary Objective #2 which evaluates efficacy in different subgroups, the primary endpoint as described in section **4.2** will be analyzed.

The endpoint for Secondary Objective #3 considers all components of self-directed violence as described in section **4.2**. Therefore, a combined total of all episodes of all components of the primary outcome will be an endpoint, and total counts of each component, in addition to all emergency department visits will be separate endpoints.

For Secondary Objective #4, dependent variables being considered longitudinally include (1) presence or absence of symptoms, and (2) presence or absence of suicidal ideation. Presence or absence of symptoms will be derived by the Internal State Scale (CRF-Form B8), and presence or absence of suicidal ideation will be derived from the Columbia Follow Up Questionnaire (CRF-Form B6).

Lastly, for Secondary Objective #5, the endpoint event will be the same as the primary endpoint described in section **4.2**.

5.3. Statistical Methods for Secondary Analyses

Secondary objective#1 will use a time to event, survival analysis to determine whether there is a treatment difference in event rates for each of the defined endpoints. The null hypothesis for each endpoint is that the two treatment groups do not differ in their hazard

rates. The alternative hypothesis is that the intervention group on lithium has a lower hazard rate than the placebo group. As with the primary analysis, log-rank test and Cox proportional hazards models will be used.

For secondary objective#2, the null hypothesis for each of the factors is that treatment effect does not differ over levels of the factor. The alternative hypothesis is that there is differential treatment effect over levels of the factor. Effect modification will be evaluated through regression methods. Cox proportional hazards regression model will include the treatment term, a term for the target factor, and the interaction term of the two. If the null hypothesis is rejected at p=0.05 level, i.e., the interaction term in the model is statisticially significant, separate regression models will be refitted for each level of the factor. A forest plot will be used to display the results.

For secondary objective#3, the null hypothesis is that total event rates do not differ by treatment assignment, and the alternative hypothesis is that there is a difference in total event rates across treatment groups. We will use Generalize Linear Models (glm), namely Poisson regression, to model cumulative total of endpoint events with an offset for follow up time.

For secondary objective#4, the null hypothesis is that change in symptoms and suicidal ideation in the lithium group is similar to that of the placebo group. The alternative hypothesis is that there is a difference in change across treatment groups over time. We will use Generalize Linear Models (glm), namely Logistic regression, to model presence or absence of symptoms and suicidal ideation at each follow up visit,while allowing for random effects of variation in symptoms of bipolar disorder and depression at baseline.

Secondary objective#5 will be evaluated by Cox Proportional Hazard regression analysis method comparing regression coefficients for treatment effect in the models with and without the mediator (Sobel, 1982). Additionally, we will explore bootstrapping methods to evaluate mediators as described in MacKinnon et al.(2002), Preacher & Hayes (2008), and MacKinnon (2003).

6. Analysis Plan for Tertiary Objectives

6.1. Tertiary Objective

Active follow up period for study subjects is 12 months after randomization. As a tertiary objective the study will obtain electronically available data up to the end of study recruitment and follow up period, which is 4 years from start of enrollment for the study. Data sources for this objective will be the Suicide Prevention and Application Network (SPAN) and VHA Corporate Data Warehouse (CDW).

Analyiss of this passively obtained data will be to:
1) describe the pattern of Lithium use following active participation in the study participants,
2) evaluate suicide reattempt and all cause mortality over a longer follow up period, and
3) evaluate all cause mortality over a longer follow up period.

6.2. Tertiary Endpoints

For Teriary Objective #1 continued use (yes/no) of Lithium is the outcome event of interest.

For Tertiary Objectives #2 the endpoint event is the same as the primary endpoint as described in section **4.2**.

For Tertiary Objective #3, death from any cause is the endpoint event.

6.3. Statistical Methods for Tertiary Analyses

Tertiary objective#1 will be evaluated by descriptive analysis on duration of lithium use, and by regression methods on any lithium use within one month following active study participation. The latter endpoint (continued use yes/no) will be modelled as a binary outcome to identify determinants and evaluate their association.

Tertiary objectives #2 and #3 will be evaluated following the same analytic strategy as the primary analysis described in section **4.3**.

7. General Considerations

7.1. Univariate and Bivariate Distributions

In general, the number of observations, mean, median, standard deviation, minimum, and maximum will be calculated for continuous variables. The number of decimals places will be up to plus two decimal places. Frequencies and percentages will be calculated for categorical data. Distribution of continuous variables and proportions of categorical variables will be tabulated by treatment group, and t-test and chi-square tests will be performed to evaluate if these variables are balanced across the 2 intervention groups.
Baseline is defined as the randomization visit.

7.2. Description of Baseline Data

Demographic details (e.g., age, sex, race, ethnicity, marital status, military service), medical history (e.g., co-morbidities), information on index attempt, psychometric measures, and baseline labs will be evaluated overall, by intervention group, by diagnosis of Bipolar Disorder or Major Depressive Disorder, and by enrollment site.

7.3. Description of Follow-up Assessment Data

Details on titration process, concomitant medications, pattern of service use, clinical events such as hospital re-admissions, emergency department admissions, suicide reattempts, psychometric measures, will be evaluated overall and by intervention group.

7.4. Safety Data

Safety information to be collected for this study include description of the event, distinction between serious and non-serious adverse event, severity and expectedness of the event, relatedness to the study intervention, and outcome of the event. In addition, data on all events will be classified using the Medical Dictionary for Regulatory Activities (MedDRA) (Version 9.0) coding dictionary.

Safety data will be collected for all consented subjects from the date the first consent was signed up to 13 months post randomization for related non-serious adverse events, and all serious adverse events. If the subject terminates from the study prior to 13 months post randomization date, collection of safety data will cease on termination date.

For reporting to the DMC and inclusion in the final report safety data will be aggregated for the ITT analytic sample only as described in **3.3**. Data will be summarized as follows.

- Frequency and percentages of related non-serious and all serious adverse events overall and by intervention group
- Frequency and percentages of all unique subjects with non-serious and serious adverse events overall and by intervention group
- Rates of non-serious and serious adverse events overall, by intervention group, and by center, calculated as # events/ person-time in years
- Tabulation of event type, MEDRA classification, severity, expectedness, relatedness and outcome of all events overall and by intervention group
- Overall hospitalization, suicide reattempt, and mortality rates at 1year overall and by intervention group

8. Data Monitoring Committee Reports

Data and study progress will be monitored by the study executive committee and by the Data Monitoring Committee (DMC). The DMC will review the study progress and safety semi-annually with additional meetings and communications as needed. All reports will be generated using SQL and SAS.

8.1. Analytic Sample for DMC Reports

All subjects randomized more than six weeks prior to the DMC meeting date will be included in the analytic cohort for the upcoming DMC report, and all data collected up to four weeks prior to the DMC meeting date will be analyzed. This allows for two weeks lag time for submission of data from sites.

8.2. Algorithm

Shell tables with annotated algorthim for creating the tables to be included in the DMC report are kept on the CSP590 Study Central file on Sharepoint in the Statistics and Data Management folder. These tables are updated as the study progresses and revisions are requested by the DMC members.

8.3. Outline of DMC Reports

The report is divided into four sections to cover subject disposition, baseline assessment, follow-up assessment, and safety assessment. The “Open Session” report includes summary tables that are not aggregated by treatment group (i.e., blinded data), and the “Closed Session” report includes summary tables for safety assessment that are aggregated by treatment groups (i.e., unblinded data). Following is the list of tables to be included within each of these 4 sections. Revision to this list, if any, will be discussed at the first DMC meeting to be held within six months of the initiation of subject enrollment in the VA sites.

8.3.1. Section A: Subject Disposition (OPEN SESSION ONLY)

Figure A1. CONSORT Diagram for CSP590
Table A2. Enrollment by site
Figure A3. Enrollment by site
Figure A4. Projected Enrollment
Table A5. Form Completion, by Form Type

Table A6. Form Completion, by Site
Table A7. Protocol Deviations
Table A8. Early Study Terminations

8.3.2. Section B: Baseline Assessment (OPEN SESSION ONLY)

Table B1. Demographics
Table B10. Study Medication Adherence
Table B11. Side Effects

8.3.3. Section C: Follow-up Assessment (OPEN SESSION ONLY)

Table C1. Study Outcomes
Table C2. Columbia Follow-up – Repeat Attempts in Detail

8.3.4. Section D: Safety Assessment (OPEN SESSION ONLY)

Table D1. Rates of Non-Serious Related Adverse Events (NAEs), by site
Table D2. Rates of Serious Adverse Events (SAEs), by site
Table D3. Summary of Related Non-Serious Adverse Events
Table D4. Summary of Serious Adverse Events (Form B15)

8.3.5. Section D: Safety Assessment (CLOSED SESSION ONLY)

Table D1. Rates of Non-Serious Related Adverse Events (NAEs), by Site and Treatment
Table D2. Rates of Serious Adverse Events (SAEs), by Site and Treatment
Table D3. **Reported in OPEN SESSION only**
Table D4. Summary of Serious Adverse Events (Form B15), by Treatment
Table D5. Related Non-Serious Adverse Events by System Organ Class, by Treatment
Table D6. Serious Adverse Events by System Organ Class, by Treatment
Table D7. Deaths Reported, by Treatment

9. Analytic Tables not included in the Data Monitoring Committee Reports

Following analytic tables will be generated annually and made available to the study team, Upon request of the DMC at the intial meeting these tables will not be included in the DMC report.

Table X1. Military Service
Table X2. Medical and Psychiatric History
Table X3. MINI Module Assessment
Table X4. Baseline Physical Assessments
Table X5. Baseline Laboratory Results
Table X6. Baseline Psychiatric Measures
Table X7. Baseline Substance Use (SCID-SUD)
Table X8. Concomitant Medications

10. Sample Size and Power

CSP590 is targeted to enroll a total of 1862 subjects, and approximately 40% of the enrolled sample is expected to terminate early (i.e., 20% complete data loss), thereby resulting in an analytic sample of 1490 evaluable subjects. With the evaluable study sample equally allocated to the 2 treatment arms, the study is powered to detect at least a 37% reduction in the one year outcome event rate.

Expected event rates used in sample size estimation to achieve 90% statistical power for hypothesis testing were 15% re-attempt rate in the placebo group and 9.45% re-attempt rate in the lithium group. Justification for selecting these rates to estimate sample size and power are provided in the protocol.

The results for the primary outcome measure will be analyzed by a two-sided log-rank test to detect either a hazards ratio that exceeds 1/0.61=1.64 or is less than 0.61, i.e., the test has 90% power to detect a hazard ratio of 1.64 or larger or 0.61 or less.

These estimations are adjusted for an O'Brien-Fleming stopping rule with one interim and one final test of hypothesis, and allow for an overall two-sided type I error of 5% for the hypothesis being tested. Hence, the interim primary hypothesis will be tested at p=0.002 level (allowing for 0.2% Type I error), and the final primary hypothesis will be tested at p=0.048 level (allowing for 4.8% Type I error).

Tables 2 below consider three scenarios for possible shift in statistical power, specifically (a) in the event that an intervention effect different from the projected 37% reduction is observed, (b) in the event that the sample size for the study falls either above or below the targeted 1490 evaluable subject mark, and (c) in the event that the event rate is below the expected 15% in the placebo group. This study will have greater than 80% statistical power at 5% overall error if the observed effect is as low as 33% reduction, if the final evaluable sample is as low as 1200 subjects, or if we observe an event rate as low as 12% in the placebo group.

Table 2: Evaluation of variability in Statistical Power assuming 5% Overall Type I error.

(a) Change in statistical power for fixed sample size and a range of effect sizes

		Event Rates				
Evaluable N	N per arm	Placebo	Active	#Events	Effect	Power
1490	745	15%	11%	190	30%	73%
1490	745	15%	10%	188	31%	77%
1490	745	15%	10%	187	33%	81%
1490	745	15%	10%	185	35%	85%
*1490	745	15%	9.5%	182	37%	90%
1490	745	15%	9%	179	40%	94%

(b) Change in statistical power for fixed effect size and a range of sample sizes

	Event Rates					
Effect	Placebo	Active	Evaluable N	N per arm	#Events	Power
37%	15%	9.5%	1700	850	208	93%
*37%	15%	9.5%	1490	745	182	90%
37%	15%	9.5%	1400	700	171	88%
37%	15%	9.5%	1200	600	147	83%
37%	15%	9.5%	1100	550	134	79%
37%	15%	9.5%	1000	500	122	75%

(c) Change in statistical power for fixed effect and sample size and a range of event rates

		Event Rates				
Total N	N per arm	Placebo	Active	#Events	Effect	Power
*1490	745	15%	9.5%	182	37%	90%
1490	745	14%	8.8%	170	37%	88%
1490	745	13%	8.2%	158	37%	85%
1490	745	12%	7.6%	146	37%	82%
1490	745	11%	6.9%	134	37%	78%
1490	745	10%	6.3%	121	37%	73%

*Highlight indicates parameters proposed in CSP590

10.1. Interim Analysis

We will conduct one interim analysis after half of the participants have completed their 12 month follow up. This analysis is projected to take place approximately 2.5 years after initiation of enrollment (assuming it will take 1.5 years to enroll half the sample and allow 12 months of follow-up). After adjusting for attrition, a sample of 746 evaluable subjects, (373 in each group), is expected to be included in the analysis.

Projecting the same event rate in this sub-sample we expect to observe a 15% or 112 events (56 events in each group) under the interim null hypothesis that event rates are equal in the treatment groups. Under the interim alternative hypothesis, the absolute reduction from 15% to 5.3% is (15-5.3) = 9.7%, the relative reduction is (15-5.3)/15 = 65%, and the hazard ratio (treatment hazard rate/control hazard rate) is 0.34. The interim log-rank test is two-sided, and the interim null hypothesis will be rejected if the hazard ratio exceeds 1/0.34= 2.94 or falls below 0.34. Using the O'Brien-Fleming procedure, (Turnbull & Jennison, 2000), this analysis will have a Type I error of 0.2%.

We will confer with the Data Monitoring Committee members and the program leadership for potential stopping guidelines based on results from interim efficacy analysis.

10.2. Futility Analysis

As a part of the interim futility analysis we will estimate the conditional probability (CP) of observing a significant final result based on assumptions about the distribution of future event rates. Assumptions for additional data to be collected in the second half of the study are:

(1) "no-change" scenario where future event rates will occur at the same rate as currently observed rates,
(2) "expected" scenario where future event rates occur at the rates expected at the beginning of the trial (i.e., as proposed in the protocol),
(3) "extreme" scenario where future event rates occur at the currently observed overall event rate, but all new events occur in the placebo group so as to set the best scenario for achieving a significant difference.

For conditional probability estimation, we will use the estimation method developed by Proschan, Lan, and Wittes,2006 and further described by Jitlal et. al, 2012, where

$$CP \text{ at a specific time} = 1 - \phi[(Z\alpha/2 - E[B(1)|B(t)])/\sqrt{1-t}],$$

$E[B(1)|B(t)]$ is the expected value of $B(t)$ at the end of the trial (when $t=1$), given the data observed until point t.

The information fraction, t, is the number of observed events so far, expressed as a proportion of the planned number of events.

11. Final Statistical Report

The final statistical report for CSP590 will be the submitted and accepted primary manuscript which includes results from the primary ITT analysis as outlined in section **4** and any addiitonal analyses to be conducted based on the journal's peer review process.
At minimum, the final statistical report will include the CONSORT diagram (section **8.3.1),** descriptive characteristics of the ITT cohort (section **8.3.2),** medical and psychiatric history of the cohort (sectiion **9**), and the primary analysis results (sectiion **4.3**).

12. References

Anderson K, Borgan O, Gill RD, and Keiding N. Statisticial Models based on Counting Processes. Section VII.3.1., Springer-Verlag New York, 1991

Jitlal M, Khan I, Lee SM, Hackshaw A (2012) Stopping clinical trials early for futility: retrospective analysis of several randomised clinical studies. British Journal of Cancer (2012) 107, 910–917

MacKinnon, D. P.; Lockwood, C. M. Hoffman, J. M., West S. G., Sheets V. A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. Psychological Methods 7: 83–104, 2002

Preacher, K. J.; Hayes, A. F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behaviour Research Methods 40: 879–891, 2008

Proschan MA, Lan KKG, Wittes JT (2006) Statistical Monitoring of Clinical Trials: A Unified Approach. 1st edn. Springer: USA

Rybin D, Doros G, Rosenheck R, Lew R. The impact of missing data on the results of a schizophrenia study.Pharmaceutical Statistics. Vol 14 (1) pages 4-10. Feb 2015

Sobel, M. E. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. Sociological Methodology 13: 290–312, 1982

Tein JY, MacKinnon DP. Estimating mediated effects with survival data. In: Yanai H, Rikkyo AO, Shigemasu K, Kano Y, Meulman JJ, editors. New Developments in Psychometrics: Psychometric Society Proceedings. Tokyo: Springer-Verlag; pp. 405–12, 2003

Turnbull B., Jennison C. Group Sequential Methods with Applications to Clinical Trials, Chapman & Hall/CRC Press LLC Interdisciplinary Statistics. Ch2, pp29-39, 2000

Yukawa E, Nomiyama N, Higuchi S, Aoyama T. Lithium population pharmacokinetics from routine cliinical data: role of patient characteristics for estiimating dosing regimens. Therapeutic Drug Monitoring 15: 75-82, 1993