ELEKT–D: Electroconvulsive therapy (ECT) vs. ketamine in patients with treatment resistant depression (TRD)

Statistical Analysis Plan

Version 2.0

07/01/2022

ELEKT-D is an open label, multi-center clinical trial to compare electroconvulsive therapy (ECT) with ketamine (KET) in patients with treatment resistant depression (TRD). The specific study aims include:

- Aim 1: to investigate the comparative effectiveness of ECT and ketamine on measures of depression.
- Aim 2: To investigate the relative impact of ECT and ketamine on measures of memory and cognitive function.
- Aim 3: To investigate the relative impact of ECT and ketamine on patient reported QoL measures after acute treatment and at follow-up over six months.

Table 1 below lists the primary, secondary and safety endpoints measured from the study participants. More details and the measurement schedules of these endpoints are described in the study protocol.

Analysis of baseline characteristics. The analysis will be based on the randomized population (i.e., all participants randomized). Participant characteristics at baseline will be summarized as descriptive statistics (mean, standard deviation [SD], median, inter-quartile range, or range for continuous data, and count and percentage for categorical data). The baseline characteristics will be compared between the two study groups using the t-test, the Wilcoxon rank sum test, or the chi-squared test as appropriate.

Analysis of primary endpoint. The primary endpoint of the trial is treatment response, which is defined as a ≥ 50% decrease in QIDS-SR-16 scores from the baseline visit/visit 1 to the end of treatment visit in the acute phase. The study hypothesis is that the response rate of ketamine is non-inferior to ECT with a margin of -10%, that is,

$$H_0: P_k - P_e \leq -10\% \text{ vs. } H_1: P_k - P_e > -10\%$$

where P_k represents the response rates in the ketamine group, and P_e represents the response rate in the ECT grou; H_0 and H_1 denote the null and alternative hypotheses, respectively. The primary endpoint of treatment response will be summarized as proportions, along with 95% confidence intervals, for each treatment group. The Farrington-Manning score test will be then used to compare the response rates between ECT and KET at an inferiority margin of -10% (ketamine minus ECT) and a one-sided type I error of 0.025. The difference of the two response rates will be also reported with the 95% confidence interval.

The principal analysis of the primary outcome will be based on the modified intention to treatment (mITT) population, which include the participants who receive at least one treatment and have at least one QIDS-SR-16 measurement during the acute phase. Multiple *sensitivity analyses* will be performed by imputing the missing primary endpoints. (1) Multiple imputations based on the chained-equation approach will be applied to the dataset, including baseline demographic and illness characteristics, the QIDS-SR-16 and MADRS scores at baseline, and their percent changes at the EOT visit. The imputed percent changes of QIDS-SR16 are then dichotomized to derive the imputed response outcomes. The Farrington-Manning score test will be applied to each of the five imputed datasets, and the results need to be summarized using Rubin's rule. (2) The worst-best and best-worst imputation methods will be used to examine the results in the extreme cases of the missing data. The former method will impute all missing outcomes in ECT as responders and all missing outcomes in KET as non-responders, and the latter imputes the missing outcomes in ECT as non-responders and all missing outcomes in KET as responders. The results from these two methods can be considered as the range of uncertainty due to the missing data.

Analysis of secondary endpoints. The secondary outcome of MADRS response will be tested for noninferiority of KET using the same approach as the primary outcome. MADRS response is defined as 50% or greater decline in the MADRS total score from baseline to the EOT visit. Similar as the primary outcomes, the MADRS response rates will be compared between the two groups to test the non-inferiority of KET at a margin of -10% and one-sided alpha of 0.025.

The other dichotomous secondary outcomes such as remission will be summarized as percentages by group and compared using the chi-squared test.

Since the QIDS-SR16 score itself is repeatedly measured over time, a linear mixed-effects model will be applied to estimate its trajectories of the two groups during the acute treatment phase. The model will include measurement time, treatment group, their interaction, and site as fixed effects, and a Gaussian random intercept at the patient level. Baseline patient characteristics with statistically significant differences between the two groups may be also included as fixed effects if deemed necessary. The covariance matrix of the repeated measurements within each patient will be selected from the candidates of compound symmetric, first-order autoregressive (AR1), Toeplitz, and unstructured structures using the information criterion. Least-square estimates will be derived to estimate the QIDS-SR16 mean at each visit by group, and corresponding linear contrasts can be used for group comparison at each visit. Since the responders will be further followed at months 1, 3 and 6 after the acute phase, another linear

mixed-effects model with the same formation will be applied to the subgroup of responders for their longitudinal QIDS-SR-16 data until the end of the follow up phase. The model for the change of QIDS-SR-16 score will include the score changes as the dependent variable and further include the baseline QIDS-SR-16 score as a covariate. For the other continuous secondary endpoints measured longitudinally, their analyses will follow the analysis of the QIDS-SR-16 score. These endpoints include various patient and clinician rated scales for depression, suicidality, and cognitive function (e.g., MADRS and YMRS).

For the cross-sectional endpoints such as the MoCA rating assessed at the end of treatment visit, linear regression will be used to analyze and compare the data between the two groups. Generalized linear regression (e.g., logistic regression or ordinal logistic regression) will be applied to binary or multi-categorical outcomes, depending on the variable type of the outcome.

Analysis of safety endpoints. Mortality will be summarized as percentage by group, and the chi-squared test will be used to compare the mortality rates between the two groups. If the expected number of deaths is small, the Fisher's exact test will be used instead. The other dichotomous safety outcomes such as the unexpected adverse events, adverse events causing treatment discontinuation and serious adverse events will be analyzed in the same manner.

For the vital signs including BMI, blood pressure and heart rate, the linear mixed-effect models will be applied to estimate their least-square means over time by group.

Subgroup analyses. Multiple *subgroup analyses* will be performed in order to determine whether there is heterogeneity of treatment effect due to demographic and clinical risk factors. The predefined subgroups include: psychiatric comorbidity, depression subtype, gender, race, age, treatment resistance (Stage IV-V vs other) and prior history of ECT treatment. For each subgroup analysis, we will test the interaction using the Breslow-Day test. Recognizing that the power to detect significant interactions will be low, all subgroup analyses are thus considered to be hypothesis generating.

Missing data. The missing primary endpoints will be imputed as described in the section of analysis of primary endpoint. Regarding the missing data in the secondary endpoints, the mixed-effects models or the generalization regression models implicitly assume data missing at random. We will also perform sensitivity analyses if needed. The first sensitivity analysis will be imputation.

If the endpoints are longitudinal and have mixed variable types, the set of longitudinal data will be arranged chronologically. The results are then summarized using Rubin's rule. Another sensitivity analysis will be using the pattern mixed modeling if the missingness is considered non-ignorable (i.e., missing not at random).

Sample size justification. For the outcome measure of QIDS-SR16 in TRD patients in standard anti-depressant analyses, an overall response rate of 47% was reported in the STAR*D study [1]. Zarate et al. showed that a 71% response rate of ketamine on HAMD-21 was associated with patients with major depression [2]. Murrough et al. reported a 64% response rate of ketamine on MADRS in patients with TRD [3]. Other studies revealed response rates of ketamine ranging from 50% to 90% in patients with TRD [4-14]. Janicak et al. reported a 56% response rate of ECT on HAMD-24 in patients of major depression [15]. Our preliminary internal data showed 45.8% patients as responders in the ketamine group, and 45.1% patients as responders in the ECT group, using clinician rated MADRS scales. The patients in the ketamine group were resistant to ECT treatment and still had a comparable response rate as ECT. For the ELEKT-D study, patients do not have to be ECT resistant, and hence the ketamine response rate is expected to be even higher. Based on the above data from the previous studies and our pilot study, we assume a 50% response rate for ECT patients. At a non-inferiority margin of 10% (i.e. response rate of ketamine is at most 10% lower than ECT) and a one-sided alpha value of 0.025, a total sample of 346 is needed to provide 80% power using the Farrington-Manning score test of risk difference if the actual difference is +5% (i.e. response rate of ketamine is 5% higher than ECT). If the actual difference is +10%, the statistical power would increase to 96.4% with the same sample size; if the non-inferiority margin is reduced to 5%, then the power will be 80.3%. Assuming an attrition rate of 15%, a total number of 400 subjects (200 per treatment group) are then required.

References

[1] Warden D, Rush AJ, Trivedi MH, et al. The STAR*D project results: a comprehensive review of finding. Current Psychiatry Reports. 2007;9(6):449-459.

[2] Zarate CA Jr., Singh JB, Carlson PJ, et al. A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. Arch Gen Psychiatry. 2006; 63(8):856–864.

[3] Murrough JW, Iosifescu DV, Chang LC, et al. Antidepressant efficacy of ketamine in treatment-resistant major depression: a two-site randomized controlled trial. Am J Psychiatry. 2013;170(10):1134-1142.

[4] Diazgranados N, Ibrahim L, Brutsche NE, et al. A randomized add-on trial of an N-methyl-Daspartate antagonist in treatment-resistant bipolar depression. Arch Gen Psychiatry. 2010; 67(8):793–802.

[5] Valentine GW, Mason GF, Gomez R, et al. The antidepressant effect of ketamine is not associated with changes in occipital amino acid neurotransmitter content as measured by [(1)H]-MRS. Psychiatry Res. 2011; 191(2):122–127.

[6] aan het Rot M, Collins KA, Murrough JW, et al. Safety and efficacy of repeated-dose intravenous ketamine for treatment-resistant depression. Biol Psychiatry. 2010; 67(2):139–145.

[7] DiazGranados N, Ibrahim LA, Brutsche NE, et al. Rapid resolution of suicidal ideation after a single infusion of an N-methyl-D-aspartate antagonist in patients with treatment-resistant major depressive disorder. J Clin Psychiatry. 2010; 71(12):1605–1611.

[8] Ibrahim L, Diazgranados N, Luckenbaugh DA, et al. Rapid decrease in depressive symptoms with an N-methyl-d-aspartate antagonist in ECT-resistant major depression. Prog Neuropsychopharmacol Biol Psychiatry. 2011; 35(4):1155–1159.

[9] Machado-Vieira R, Yuan P, Brutsche N, et al. Brain-derived neurotrophic factor and initial

antidepressant response to an N-methyl-D-aspartate antagonist. J Clin Psychiatry. 2009; 70(12):1662–1666.

[10] Mathew SJ, Murrough JW, aan het Rot M, et al. Riluzole for relapse prevention following intravenous ketamine in treatment-resistant depression: a pilot randomized, placebo-controlled continuation trial. Int J Neuropsychopharmacol. 2010; 13(1):71–82.

[11] Phelps LE, Brutsche N, Moral JR, et al. Family history of alcohol dependence and initial antidepressant response to an N-methyl-D-aspartate antagonist. Biol Psychiatry. 2009;65(2):181–184.

[12] Price RB, Nock MK, Charney DS, et al. Effects of intravenous ketamine on explicit and implicit measures of suicidality in treatment-resistant depression. Biol Psychiatry. 2009; 66(5):522–526.

[13] Salvadore G, Cornwell BR, Colon-Rosario V, et al. Increased anterior cingulate cortical activity in response to fearful faces: a neurophysiological biomarker that predicts rapid antidepressant response to ketamine. Biol Psychiatry. 2009; 65(4):289–295.

[14] Salvadore G, Cornwell BR, Sambataro F, et al. Anterior cingulate desynchronization and functional connectivity with the amygdala during a working memory task predict rapid antidepressant response to ketamine. Neuropsychopharmacology. 2010; 35(7):1415–1422.

[15] Janicak PG, Dowd SM, Martis B, et al. Repetitive transcranial magnetic stimulation versus electroconvulsive therapy for major depression: preliminary results of a randomized trial. Biol Psychiatry. 2002;51(8):659–667

Table 1. Primary, secondary and safety endpoints in ELEKT-D

Study Endpoint	Definition	Measure Type
Primary	Response (≥50% reduction in QIDS-SR16 scores from the Baseline/Visit 1 to End of Treatment visit)	Patient rated scale
Secondary	Remission (QIDS-SR16<5 at two consecutive visits)	Patient rated scale
	Summary score of MINI 7.0.2 questions	Diagnostic interview
	Summary score of QIDS-SR-16	Patient rated scale
	Summary score of GSE-My	Patient rated scale
	Summary score of SMCQ	Patient rated scale
	Summary score of PGI-S and PGI-I	Patient rated scale
	Summary score of QOLS	Patient rated scale
	Summary score of PRISE	Patient rated scale
	Summary score of Treatment Preference	Patient rated scale
	Summary score of MADRS	Clinician rated scale
	Summary score of CSSRS	Clinician rated scale
	Summary score of CADSS	Clinician rated scale
	Summary score of YMRS	Clinician rated scale
	Summary score of BPRS	Clinician rated scale
	Summary score of CGI-S and CGI-I	Clinician rated scale
	Summary score of MoCA	Cognitive testing
	Summary score of COWAT	Cognitive testing
	Summary score of HVLT-R	Cognitive testing
	Summary score of Stroop	Cognitive testing
	Summary score of NAART	Cognitive testing
	Summary score of CPFQ	Cognitive testing
Safety	Mortality	
	Unexpected adverse event	
	Study treatment discontinuation adverse event	
	Serious adverse event	
	Vital signs (BMI, BP, heart rate)	

Summary of the Changes in the Statistical Analysis Plan (SAP)

The two SAPs generally have the same content. The first version of the SAP was drafted for the original grant application, and the second version was drafted at the end of the study. The differences between the two versions are summarized below.

1. Section 3.2 in version 1, the analysis plan for the secondary measures of cognitive assessment, was revised in version 2. In version 2, since these measures were longitudinal data, their analyses were proposed to follow the analyses of the longitudinal QIDS-SR-16 or MADRS scores ("Analysis of secondary endpoints" in version 2).
2. Section 4 in version 1, data analysis for follow-up period, was revised in version 2. In version 2, separate linear mixed models were proposed for responders and non-responders for the follow-up data since these two groups had different follow-up schedules after the acute phase.