Shionogi Study Title:	Phase-3 clinical study of duloxetine hydrochloride in children's and adolescent patients with depressive disorder: Superiority study versus placebo
Shionogi Study Number:	1701A3631
ClinicalTrials.gov Registration No.	NCT03315793
Study Document	Statistical Analysis Plan Version 1.1 17 Jan 2020

History of Statistical Analysis Plan Amendments

Version 1 (Original)	21 Nov 2017
Version 1.1	17 Jan 2020
Revisions to Version 1 and 1.1 are documented in Version 1.1	

Phase-3 clinical study of duloxetine hydrochloride in children's and adolescent patients with depressive disorder:

Superiority study versus placebo

Protocol No. 1701A3631 (NCT03315793)

Statistical Analysis Plan

Version 1.1

[redacted]

This is the translated version of the Statistical Analysis Plan (Version 1.1) written in Japanese

Signatures

Company name and roles	Approver	Date
[redacted]	[redacted] (Refer to the flag page)	Refer to the flag page
[redacted]	[redacted] (Refer to the flag page)	Refer to the flag page

Preparation and amendment history

Version	Date	Author, amender	Comments
1.0	November 21, 2017	[redacted]	New preparation
1.1	January 17, 2020		Addition of abbreviations; revision of handling of the tests at discontinuation and at the time of final analysis; addition of subgroup analysis; addition of apportionment conditions for tipping-point analysis; addition of CGI-S score category distribution analysis; addition of status of occurrence of adverse events and adverse drug reactions, classified by dosage at onset

Table of contents

List of abbreviations and terms

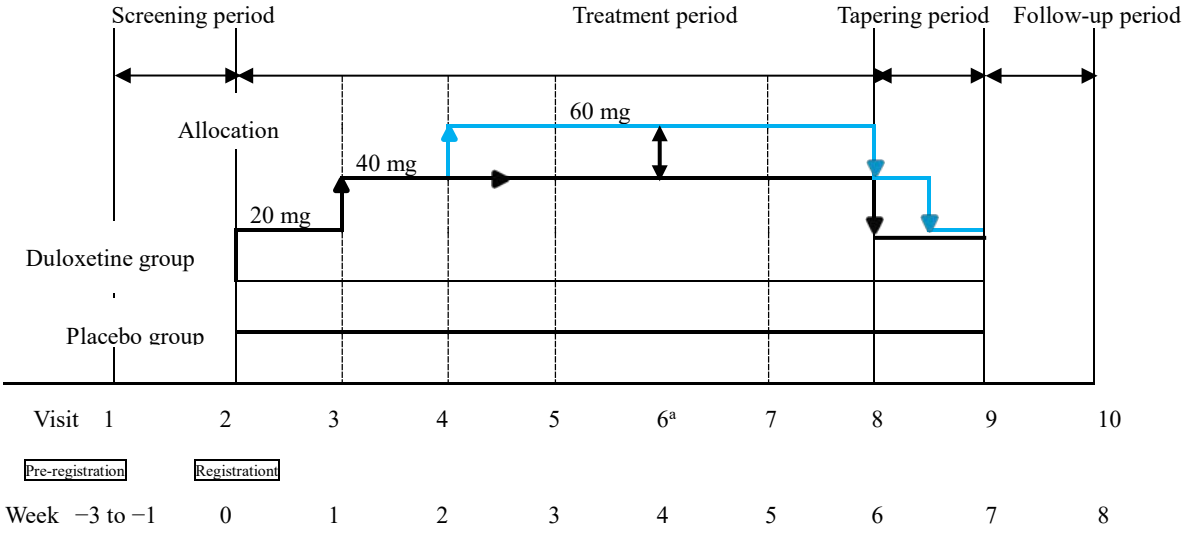
Abbreviation	Full term
BOCF	baseline observation carried forward
CDRS-R	Children’s Depression Rating Scale-Revised
CGI-S	Clinical Global Impression of Severity
C-SSRS	Columbia Suicide Severity Rating Scale
DSM-5	Diagnostic and Statistical Manual of Mental Disorders Fifth Edition
FAS	Full Analysis Set
FT3	free triiodothyronine
FT4	free thyroxine
LOCF	last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
MINI-KID	Mini International Neuropsychiatric Interview for children and adolescents
MMRM	mixed-effects model repeated measures
PPS	Per Protocol Set
SMQ	Standard MedDRA Queries

1. Introduction

This document presents in concrete terms the contents of Section 8 (Statistical analysis) of the study protocol (study no.: 1701A3631). This document will be fixed before the key break. With respect to the details of the output from the analysis results stated in the statistical analysis plan, the analysis figures and tables output plan (mock tables, listings and figures) will be prepared separately.

2. Study summary

This study is a multicenter, randomized, placebo-controlled, double-blind, parallel-group, comparative study with children's and adolescent patients with depressive disorder, with the target number of subjects being 148. It consists of the following four periods: Screening period (1-3 weeks), treatment period (6 weeks), Tapering period (1 week), Follow-up period (1 week), (total: up to 11 weeks). Subjects whose eligibility has been confirmed after obtaining informed consent will be pre-registered (Visit 1), and, after completion of the 1- to 3-week screening period, their eligibility will be confirmed again, and they will be registered (Visit 2). Registered patients will be allocated randomly to either the duloxetine group or placebo group, in a 1:1 ratio, and will be orally administered duloxetine or placebo once daily after breakfast, under double-blind conditions. With the duloxetine group, administration will be initiated at 20 mg/day, and increased to 40 mg/day at Visit 3, 1 week later. From Visit 4, 1 week after Visit 3, the dose can be increased or decreased between 40 and 60 mg/day, in accordance with stipulations in the study protocol. Subjects who are to transfer to the extended long-term treatment study (study no.: 1702A3632) will do so after completion of the tapering period.



[a] Visit 6 is a visit that can be made to the hospital in connection with dose increase, decrease, etc.

3. Study objectives

3.1 Primary objective

To verify the superiority of duloxetine over placebo, using as the index change from baseline in total

children's depression rating scale - revised (CDRS-R) score.

3.2 Secondary objectives

- To evaluate the efficacy of duloxetine compared with placebo on the basis of the following indices:
  - 30% response rate: Proportion of subjects showing CDRS-R total total score decreases of 30% or more from baseline.
  - 50% response rate: Proportion of subjects showing CDRS-R total total score decreases of 50% or more from baseline.
  - Remission rate: Proportion of subjects with CDRS-R total scores of 28 or lower
  - Change from baseline in CDRS-R subscales and item 13 (suicidal ideation)
  - Change from baseline in clinical global impression: severity (CGI-S) score

- To evaluate the safety of duloxetine, with occurrence or non-occurrence and frequency of adverse events and adverse drug reactions as indices.

4. Study design

4.1 Observation schedule

A schedule table is presented in Appendix 1.

4.2 Allocation method

The subjects whose eligibility is confirmed after completion of the screening period will be allocated to either the duloxetine or placebo group in a ratio of 1:1, by the stochastic minimization method, using age (<12, ≥12) and study site as allocation factors.

Consideration should be given so that the maximum inter-group difference at each study site does not exceed two subjects.

4.3 Target number of subjects

The target number of subjects will be 148 (74 each in the duloxetine and placebo groups).

Rationale for the target sample size:

The number of subjects has been set with reference to the results obtained in two overseas, Phase-3 studies, the HMCL and HMCK studies, with the target population planned for the present study, that is, patients who were aged ≥9 when informed consent was obtained, and ≥7 at initial onset of depressive symptoms. In addition, the number of subjects in the present study has been set with the expectation that more accurate selection of institutions and subjects will enable minimization of the placebo response, and thus increase in the inter-group difference.

In this study, the inter-group difference (placebo - duloxetine) between the duloxetine group and the placebo group in CDSRS-R change to week 6 of the treatment period is assumed to be 5.9, the standard deviation in CDRS-R change is assumed to be 12.0 in both groups, and the effect size is predicted to be

0.49. It is thus estimated that 66 subjects per group (total: 132) will be required to achieve a detection power of 80% or higher in a two-sample t-test at a two-sided significance level of 0.05. Therefore, predicting the dropout rate to be 10%, the target total number of subjects has been set at 148 (74 per group).

5. Analysis sets

As efficacy analysis sets, a full-analysis set (FAS) and a per-protocol set (PPS) will be established as defined below. The FAS will be the primary analysis set, whereas the PPS will be used for sensitivity analysis in the primary analysis of the primary endpoint. In addition, for analysis of safety, a safety-analysis set will be established.

5.1 FAS

This comprises all randomized subjects to whom the study drug has been administered at least once, and with whom the CDRS-R total score has been determined at baseline and at least once after initiation of study drug administration. Even in the event of an incorrect prescription during the study, analysis of the FAS will be performed with the allocated group.

5.2 PPS

This comprises all randomized subjects who are included in the FAS, and to whom the below conditions do not apply. This set will be fixed before key break, on the basis of subject handling criteria that are prepared separately from the present document.

- Not meeting the inclusion criteria, and/or violating the exclusion criteria, in the protocol
- Insufficient compliance in study drug use.
- Violation of restrictions on concomitant treatment.

5.3 Safety-analysis set

This set comprises all randomized subjects to whom the study drug has been administered at least once. Analysis of this set will be performed with subjects actually administered the study drug, rather than allocated group.

6. Issues with statistical analysis, and data handling

6.1 Reporting of analysis results

The numbers of subjects, arithmetic means, standard deviations, minima, medians, and maxima will in principle be calculated as summary statistics for the endpoints measured as continuous data. Items observed as discrete values will be summarized as the number and proportion of subjects in each category. The software used for statistical analyses in this study will be SAS, Version 9.4.

6.2 Statistical tests

The level of significance for the statistical tests will be 0.05 (two-sided), unless stated otherwise. The

primary endpoint in this study will be only change in CDRS-R total score from baseline, and no analyses will be performed taking into consideration multiplicity concerning multiple items and multiple time-points.

6.3 Assessments, observations and tests: Permitted timing and handling in statistical analyses

In principle, data obtained at the time-points shown in the case report forms, and collected in accordance with the performance schedule shown in the study protocol, will be used for the statistical analysis. In the case of data reported at the discontinuation visit, the time-points for use in analyses will be determined on the basis of the permitted ranges in the performance schedule. If there is duplication by data obtained at the scheduled visits and data obtained at the discontinuation tests, the data obtained closest to the scheduled date will be used. If this does not eliminate duplication, data from the scheduled visit will be used. Unscheduled visits will not be included in the time-points for analysis. For the permitted ranges for the observation, test, and evaluation time-points shown in the study protocol, refer to Appendix 2.

For the final evaluation, data obtained at week 7 or at discontinuation of the treatment or tapering period are used, except for CDRS-R and CGI-S scores, for which those obtained at week 6 or at discontinuation of the treatment period are used. However, if none of the above data are available, the latest data among reported data collected at scheduled visits are used for the final evaluation.

6.4 Handling of missing data

In principle, missing data will not be imputed. If missing data for CDRS-R total scores 6 weeks after initial administration are nevertheless imputed for the secondary efficacy analyses, this will be performed by the last observation carried forward (LOCF) method, baseline observation carried forward (BOCF) method, and modified BOCF method.

If missing data imputation is performed by the LOCF method, the measurements used as substitutes will be those obtained at the latest scheduled evaluation time-points (Visits 3, 4, 5, 6 and 7) before the missing-data time-points.

If missing data imputation is performed by the BOCF method, the baseline data will be used as substitutes. The modified BOCF method is defined as follows:

- With subjects withdrawn from the study due to adverse events and/or insufficient efficacy, imputation will be by the BOCF method.
- With subjects withdrawn from the study for any reason other than the above, imputation will be by the LOCF method.

6.5 Baseline

Data obtained before initiation of study drug administration (Visit 2) will be used as baseline data, unless otherwise specified. As with laboratory test results, if data from Visit 2 are missing, data from Visit 1 (screening period) will be used as baseline data.

7. Subject composition and backgrounds

7.1 Disposition of patients

Among registered subjects, the number who complete the treatment period and the number of withdrawals are calculated for each treatment group, and in the case of the withdrawals the disposition of the reasons for discontinuation is shown. Among subjects transferred to the tapering period, the number who complete the tapering period and the number of withdrawals are calculated, and the disposition of the reasons for discontinuation is summarized.

7.1.1 Patient composition in efficacy evaluation

With registered subjects, the numbers included in and excluded from the FAS will be calculated for each treatment group, and the disposition of the reasons for exclusion will be shown. Similar analysis will be performed for the PPS.

7.1.2 Patient composition in safety evaluation

With registered subjects, the numbers of subjects included in and excluded from the safety-analysis set will be calculated for each treatment group, and the disposition of the reasons for exclusion will be shown.

7.2 Analysis of demographic factors

With the FAS, for each treatment group, summary statistics will be determined for the following continuous data items; and the frequencies and proportions of subjects in each category will be determined for discrete values:

Data type	Item
Continuous	Age on date of obtaining legal representative's informed consent, age at initial onset of depression, height, body weight, CDRS-R total score baseline value, CDRS-R subscale (mood, somatic, subjective, behavioral) baseline values, CGI-S score baseline value
Discrete	Sex (male, female), ethnicity, race, age on date of obtaining legal representative's informed consent (<12, ≥12), depressive episode ordinal number (first time, second time, third to seventh time, eighth time or more, not known), duration of current depressive episode (0 to 2 weeks, 2 to 4 weeks, 4 to 8 weeks, more than 8 weeks, not known), CDRS-R total score baseline value (up to 60, 61 or more), disease class according to DSM-5 (depression, persistent depressive disorder), hospitalization status category (hospitalized, out-patient), medical and surgical history (present, absent), complications (present, absent), previous drug treatment (present, absent), previous therapy (present, absent)

The dispositions of medical and surgical history, complications, previous drug treatment, and previous therapy are shown for each treatment group. Drugs used previously will be coded for using the WHO Drug Dictionary, and summarized on the basis of consistently selected noncommercial names.

8. Treatment course and compliance

8.1 Compliance

8.1.1 Duration of administration and total dose

The following analyses will be performed with the safety-analysis set:

1) Summary statistics for study drug administration duration will be calculated for each treatment group. Administration duration is defined as follows:

   Administration duration (days) = (administration completion date) − (administration initiation date) + 1.

   In addition, times (days) will be converted to week units, and the distributions in the following categories will be summarized: less than 1 week; 1 to 2 weeks; 2 to 6 weeks; and more than 6 weeks.

2) Summary statistics will be calculated for the total dose of study drug in the duloxetine group.

3) With the duloxetine group, for the maximum and final doses during the treatment period, distribution in the following categories will be summarized: 20 mg, 40 mg, and 60 mg.

8.1.2 Compliance rate

In the safety-analysis set, summary statistics for compliance rate during the treatment period will be calculated for each treatment group. Distribution in the following categories will be summarized: less than 70%, and 70% or more. Compliance rate is defined as follows:

$$\text{Compliance rate [\%]} = \frac{\{Total\ Prescription - 20 \times (Total\ number\ of\ unfilled\ capsules) - Total\ mis\ prescription\}}{(Total\ Prescription)} \times 100$$

8.2 Concomitant treatment

The following analyses will be performed with the safety-analysis set: Concomitant drugs will be coded for using the WHO Drug Dictionary, and then summarized on the basis of consistently selected noncommercial names.

1) The number and proportion of subjects who have taken any concomitant medication at least once during the study period will be calculated. In addition, the disposition of the concomitant medication(s) will be shown. If use of the same concomitant medication has been reported more than once with the same subject, this will be taken to be one concomitant medication.

2) The number and proportion of subjects who have had any concomitant therapy at least once during the study period will be calculated. In addition, the disposition will be shown by the name of concomitant therapy. If use of the same concomitant therapy has been reported more than once with the same subject, this will be taken to be one concomitant therapy.

9. Efficacy evaluation

Efficacy endpoints and the relevant analytical methods are summarized in Table 1. For the primary

endpoint, CDRS-R total total score, the primary analysis will be by the mixed-effects model repeated measures (MMRM) method, and the secondary analysis will be by analysis of covariance. With the FAS, each endpoint will be analyzed. In addition, with the PPS, the primary analysis of the primary endpoint, change in CDRS-R total total score, will be performed.

Table 1. Summary of endpoints and analysis methods

Endpoints		Analysis method[a]	Output contents[b]	Analysis sets[c]	
				FAS	PPS
Primary	Change in CDRS-R total score	1, 2, 4	1, 2, 4	○	○[d]
Secondary	CDRS-R total score response rate, remission rate	3	1, 3	○	−
	Change in CDRS-R subscales and item 13 (suicidal ideation)	1	1, 2	○	−
	Change from baseline in CGI-S score	1	1, 2	○	−

a Analysis method 1: MMRM method
Analysis method 2: Analysis of covariance
Analysis method 3. Cochran-Mantel-Haenszel test
Analysis method 4. Tipping-point analysis

b Output contents 1: Summary statistics
Output contents 2: Adjusted mean, inter-group difference in adjusted mean and 95% confidence interval, p-value
Output contents 3: Inter-group difference and 95% confidence interval, p-value
Output contents 4: Time-course graph for adjusted mean and standard error

c In the table: ○: performed; −: not performed

d In the case of the primary analysis of the primary endpoint only, analysis is performed with the PPS.

9.1 Primary endpoint

CDRS-R total score is the primary endpoint, and its change from baseline is the primary evaluation index.

9.1.1 Primary analysis of primary efficacy endpoint

As the primary analysis, the superiority of duloxetine over placebo will be evaluated in terms of change in CDRS-R total score by week 6 of administration. In detail, all usable data obtained at stipulated time-points after study drug administration will be used, and a linear model (MMRM) will be applied, with change in CDRS-R total score from baseline taken to be a response variable; treatment group, observation time-points, and interaction between treatment group and observation time-points taken to be fixed effects; baseline CDRS-R total score and age (<12, ≥12) taken to be covariates; and an unstructured covariance structure for error variance. In this step, the degree of freedom will be adjusted using the approximation of Kenward and Roger. On the basis of this model, the estimated difference between the duloxetine and placebo groups at week 6 of administration, and its 95% confidence interval, and the p-value, will be calculated.

Examples of SAS codes for applying the above model are shown below.

```
proc mixed data=CDRSR order=internal;
  class SUBJID TRT VISIT AGEGR;
  model CHG=CDRSR_BASE AGEGR TRT VISIT TRT*VISIT / ddfm=kr;
```

```
  repeated VISIT / type=un subject=SUBJID;
  lsmeans TRT*VISIT;
  estimate 'DLX vs PLB' TRT 1 -1 TRT*VISIT 0 0 0 0 1  0 0 0 0 -1 / cl;
run;
```

SUBJID: Subject identification number
TRT: Group (DLX = duloxetine group; PLB = placebo group)
VISIT: Stipulated observation time-points (Visits 3, 4, 5, 7 and 8)
AGEGR: Age category (<12, ≥12)
CDRSR_BASE: CDRS-R total score baseline value
CHG: Change in CDRS-R total score from baseline

If the above model does not converge, a covariance structure will be selected in the following order, and the MMRM method will be applied: (i) heterogeneous AR (1); and (ii) heterogeneous compound symmetry. In addition, time-course graphs of adjusted mean (± standard error) of change in total CERS-R score, estimated by the MMRM method, will be prepared for each treatment group.

9.1.2 Secondary analysis of primary endpoint

1) Analysis of the PPS by the MMRM method

Change in CDRS-R total score will be analyzed with the PPS, in the same manner as the primary analysis.

2) Analysis of change at week 6 of administration (analysis of covariance)

A) An analysis of covariance will be performed with change in CDRS-R total score from baseline to week 6 of administration as the response variable; treatment group as the fixed effect; and baseline CDRS-R total score, and age (<12, ≥12) as covariates. The estimated inter-group difference and 95% confidence interval, and p-value, will thus be calculated. If the CDRS-R total score at week 6 of administration is not available, the missing data will be imputed by the LOCF method.

B) On the basis of missing data imputation by the BOCF method, analysis of covariance will be performed in the same manner as above.

C) Likewise, analysis of covariance will be performed on the basis of missing data imputation by the modified BOCF method.

3) Supplemental analysis of missing data imputation methods:

With respect to change in CDRS-R total score, with multiple imputation with missing at random hypothesized, solely in the case of the duloxetine group the k-multiple of the inter-group difference (k = 0.1, 0.2, ...) for the treatment-group primary effect in the primary analysis MMRM method will be subtracted from the data obtained by missing data imputation, and analysis will be performed in a similar manner to the primary analysis. Despite differences in the time from baseline to the point at which data are missing, the subtracted values will be set the same. In the case of multiple imputation, taking the imputation to be performed 100 times, the seed value is set at 3,631. With respect to the analysis results (adjusted mean inter-group difference and standard error) with

imputation performed this number of times, these will be integrated as single results using proc mianalyze.

If the upper limit of the 95% confidence interval for adjusted mean inter-group difference is below 0, the k-value that overturns the results of the primary analysis is above 0, and this is taken to be the tipping point. The results of the primary analysis with the tipping point and the k-value up to reaching the tipping point will be reported. On the other hand, if the upper limit of the 95% confidence interval for adjusted mean inter-group difference in the primary analysis is above 0, a similar analysis is performed, but with a k-value below 0 taken to be the tipping point.

4) Summary statistics for changes

With respect to CDRS-R total score at each evaluation time-point, and changes from baseline, summary statistics will be calculated for each treatment group. On the basis of the model used for the primary analysis, the estimated inter-group difference and 95% confidence interval, and the p-value, will be calculated for each evaluation time-point.

9.1.3 Primary endpoint subgroup analysis

Subgroup analyses will be performed with the below demographic factors. For each subgroup, an MMRM model will be applied, taking treatment group, observation time-points, and interaction between treatment group and observation time-points to be fixed effects; the baseline CDRS-R total score to be a covariate; and the covariance structure for error variance to be unstructured. The estimated difference between the duloxetine and placebo groups 6 weeks after initiation of administration, and the p-value, will thus be calculated. The degree of freedom will be adjusted using the approximation of Kenward and Roger.

- Age category (<12, ≥12)
- Sex (male, female)
- Baseline CDRS-R total score (up to 60, 61 or more)
- Baseline body weight (below the median for the entire FAS, above the median for the entire FAS)
- Depressive episode ordinal number (first time, second time or more)

9.2 Secondary endpoints:

The secondary endpoints will be as follows:

1) CDRS-R total score response rate, remission rate
2) Change from baseline in CDRS-R subscales and item 13 (suicidal ideation):

   The CDRS-R total score will be calculated by adding together the scores for answers to all questions in the 17 items. The subscales are defined as mood, somatic, subjective and behavioral (refer to Appendix 3).
3) Clinical Global Impression of Severity (CGI-S):

   Disease severity will be evaluated on the basis of seven grades, from 1 (no disease or abnormality) to 7 (extremely severe disease). The analysis of CGI-S will be based on the score.

9.2.1 Analysis of secondary efficacy endpoints

1) Analysis of response rate and remission rate

For each of the indices defined below, the number of responders (number of subjects showing remission) and response rate (remission rate) will be obtained for each treatment group, at weeks 1 to 3, 5, and 6 of administration, and at the final evaluation (week 6 of administration, or at discontinuation of the treatment period), and the Cochran-Mantel-Haenszel test, with stratification by the allocation factor age (<12, ≥12), will be used to compare the duloxetine and placebo groups. In addition, the inter-group difference in response rate (remission rate), and the 95% confidence interval, will be calculated.

Index	Definition
30% response	Reduction of 30% or more in CDRS-R total score compared with baseline
50% response	Reduction of 50% or more in CDRS-R total score compared with baseline
Remission	CDRS-R total score of 28 or less

An analysis with withdrawn subjects taken to be non-responders (non-remission subjects) will also be performed.

2) Analysis of CDRS-R total score subscales and item 13 (suicidal ideation), and CGI-S score

All usable data obtained at stipulated time-points after study drug administration will be used, and a linear model (MMRM) will be applied, with change from baseline taken to be a response variable; treatment group, observation time-points, and interaction between treatment group and observation time-points taken to be fixed effects; baseline value, and age (<12, ≥12) taken to be covariates, and the covariance structure for error variance taken to be unstructured. In this step, the degree of freedom will be adjusted using the approximation of Kenward and Roger. On the basis of this model, the estimated difference between the duloxetine and placebo groups at each time-point after administration initiation, and the 95% confidence interval, and the p-value, will be calculated.

3) CGI-S score category distribution

The CGI-S score category distribution will be summarized for each group at baseline and at the final evaluation (week 6 of administration or discontinuation of the treatment period).

10. Safety evaluation

Safety analyses will be performed with the safety-analysis set.

10.1 Adverse events and adverse drug reactions

Adverse events occurring at or after the initial administration of the study drug will be evaluated. Reported adverse events will be converted to Medical Dictionary for Regulatory Activities (MedDRA) terms, and tabulated by system organ class and preferred term.

1) For each treatment group, the number of subjects with a given adverse event and the number of cases will be determined, and the incidence rate and 95% confidence interval will be calculated. The incidence rate will be obtained as the proportion of subjects in the analysis set with a specific adverse event, and the confidence interval for the incidence rate will be determined by the Clopper-Pearson method. In addition, the incidence rates will be compared between treatment groups using Fisher's exact test. Totaling will be performed in a similar manner to above with deaths, other serious adverse events, adverse events leading to administration discontinuation, adverse events leading to dose reduction, adverse drug reactions, deaths as adverse drug reactions, other serious adverse drug reactions, adverse drug reactions leading to administration discontinuation, and adverse drug reactions leading to dose reduction. The relevant definitions are shown below.

Death	Adverse event with "Death" as the outcome
Other serious adverse event	Adverse event in the seriousness category "Serious", except for death
Adverse event leading to administration discontinuation	Adverse event for which the study drug action is "Administration discontinuation"
Adverse event leading to dose reduction	Adverse event for which the study drug action is "Dose reduction"
Adverse drug reaction	Adverse event with the causal relationship with the study drug judged to be other than "Not related"
Death as adverse drug reaction	Death, with the causal relationship with the study drug judged to be other than "Not related"
Other serious adverse drug reaction	Other serious adverse event with the causal relationship with the study drug judged to be other than "Not related"
Adverse drug reaction leading to administration discontinuation	Adverse event leading to administration discontinuation, with the causal relationship with the study drug judged to be other than "Not related"
Adverse drug reaction leading to dose reduction	Adverse event leading to dose reduction, with the causal relationship with the study drug judged to be other than "Not related"

2) For each treatment group, the number of subjects with a given adverse event, the number of cases, and the incidence rate will be determined by adverse event system organ class and preferred term. However, when calculating the number of subjects with a given adverse event and the incidence rate, if the same adverse event occurs more than once with the same subject, these events will be taken to be have affected one subject. Adverse drug reactions will be tabulated similarly.

3) With respect to adverse events with an incidence of 2% or more in at least one treatment group, the number of subjects, number of cases, and incidence rate will be calculated for each adverse event preferred term. Adverse drug reactions will be tabulated similarly.

4) For each treatment group, the number of subjects and incidence rate will be calculated for each severity and outcome category, by adverse event system organ class and preferred term. If the same adverse event occurs more than once in different categories with the same subject, it will be taken to have occurred once, in the category with the highest of the below priority ranks. Adverse drug reactions will be tabulated similarly.

	Survey item category	
Priority rank	Severity	Outcome
1	Severe	Death
2	Moderate	Recovery with sequelae
3	Mild	No improvement
4		Partial recovery
5		Recovery
6		Unknown

5) For each treatment group, the number of subjects and the incidence will be calculated for each onset timing category (weeks 0 to 1; weeks 1 to 2; weeks 2 to 3; weeks 3 to 6; later than week 6) for adverse events during the treatment period, by system organ class and preferred term. The denominator of the incidence rate will be the number of subjects with administration duration no shorter than the lower limit of each onset time category. The onset timing will be calculated as follows:

Onset timing [day] = (onset date) − (administration initiation date) + 1

When calculating the number of subjects and the incidence rate, if the same adverse event occurs in different categories more than once in the same subject, it will be taken to have occurred with one subject in each of the relevant categories. Adverse drug reactions will be tabulated similarly.

6) With subjects in the safety analysis set who are transferred to the tapering period, adverse events that occur during the tapering period will be summarized. For each treatment group, the number of subjects with a given adverse event, the number of cases, and the incidence rate will be determined by adverse event system organ class and preferred term. However, when calculating the number of subjects with a given adverse event and the incidence rate, if the same adverse event occurs more than once with the same subject, this event will be taken to be have affected one subject. Adverse drug reactions will be tabulated similarly.

7) For each treatment group, with respect to adverse events for which the below standard MedDRA queries (SMQ) are applicable, the number of subjects, number of cases, and incidence rate will be determined by system organ class and preferred term. However, when calculating the number of subjects with a given adverse event, and the incidence rate, if the same adverse event occurs more than once with the same subject, this event will be taken to be have affected one subject.

SMQ	Range
Suicide/self-injury	Broad search
Hostility/aggression	Broad search

8) With respect to the duloxetine group, the number of subjects and incidence rate will be calculated for each duloxetine dose category (20 mg, 40 mg, 60 mg) at the adverse event onset date, by system organ class and preferred term. When calculating the number of subjects and the incidence rate, if the same adverse event occurs in different categories more than once in the same subject, it will be taken to have occurred with one subject in each of the relevant categories. Adverse drug reactions will be tabulated similarly.

9) Subgroup analyses will be performed for the below demographic factors. For each treatment group, the number of subjects and the incidence rate will be calculated by adverse event system organ class and preferred term.
   - Age (<12, ≥12)
   - Sex (male, female)

10.2 Analysis of vital signs

With respect to blood pressure (systolic and diastolic), pulse rate, and body weight, summary statistics of observed values at each evaluation time-point, and changes from baseline, will be calculated for each treatment group. The evaluation time-points are taken to be the screening period, before administration initiation (baseline), weeks 1 to 6 after administration initiation, and the final evaluation time (week 7, or at discontinuation of the treatment period or tapering period).

10.3 Analysis of laboratory test results

The below laboratory test results will be analyzed. The time-points for evaluation of laboratory test results are taken to be the screening period (baseline), weeks 3 and 6 after administration initiation, and the final evaluation time (week 7, or at discontinuation of the treatment period or tapering period). However, for HbA_{1c}, TSH, FT3 and FT4, evaluation is solely during the screening period.

Classification	Parameter
Hematology tests	Leukocyte count, erythrocyte count, hemoglobin, hematocrit, leukocyte fractions (eosinophils, basophils, neutrophils, monocytes, lymphocytes), platelet count
Blood chemistry	AST, ALT, LDH, γ-GTP, ALP, creatine kinase, total bilirubin, total protein, blood

Classification	Parameter
	urea nitrogen, serum creatinine, uric acid, total cholesterol, triglycerides, Na, K, Cl, Ca, blood glucose, HbA1c, TSH, FT3, FT4
Urinalysis (qualitative)	Urinary protein, urinary glucose, urobilinogen, urinary occult blood

1) With respect to laboratory test results obtained as continuous values, summary statistics of observed values at each evaluation time-point, and changes from baseline, will be calculated for each treatment group.
2) With respect to urinalysis parameters obtained as qualitative values, the number and proportion of subjects in each category at each evaluation time-point will be determined for each treatment group.
3) As evaluation of liver function, the numbers and proportions of subjects meeting the following criteria at each evaluation time-point, and the numbers and proportions meeting these criteria at least one evaluation time-point after administration initiation, will be determined for each treatment group:
   - AST >5 times the upper limit of the reference values.
   - ALT >5 times the upper limit of the reference values.
   - AST or ALT >3 times the upper limit of the reference values, and total bilirubin >2 times the upper limit of the reference values
   - AST >3 times the upper limit of the reference values.
   - ALT >3 times the upper limit of the reference values.

10.4 Analysis of ECG

With respect to the presence or absence of ECG abnormalities, the frequency of each category before and after administration initiation will be summarized in the form of a shift table for each treatment group. With respect to categories after administration initiation, events will be taken to have occurred once, in the category with the highest of the below priority ranks.

	Survey item category
Priority rank	Judgment
1	Abnormal, clinically significant
2	Abnormal, not clinically significant
3	Normal
4	Not observed

10.5 Analysis of Columbia suicide severity rating scale

The evaluation time-points for the Columbia suicide severity rating scale will be the screening period (Visit 1), before administration initiation (Visit 2), weeks 1 to 7 of administration, and discontinuation, but the week-4 time-point will only be for subjects who visit hospital as required. The distribution of categories (yes, no) on the below evaluation scale will be summarized in the form of a shift table before and after

administration initiation, for each treatment group. For categories before administration initiation, subjects with at least one “yes” assessment at any one evaluation time-point in the screening period or before administration initiation will be classified as ”yes”, and other subjects will be classified as “no”. For categories after administration initiation, subjects with at least one “yes” assessment at any one evaluation time-point after administration initiation will be classified as “yes”, and other subjects will be classified as “no”. Both before and after administration initiation, subjects missing observed values at all evaluation time-points will be classified as “non-assessed subjects”.

Columbia suicide severity rating scale	
Suicidal ideation	1. Wish to be dead 2. Non-specific Active suicidal thoughts.
Suicidal behavior	1. Actual attempt 2. Non-Suicidal Self-Injurious Behavior 3. Self-Injurious Behavior, intent unknown 4. Interrupted attempt 5. Aborted attempt 6. Preparatory Act or Behavior 7. Suicidal behavior 8. Completed suicide

11. Performance of interim analysis

In this study, no interim analysis will be performed.

12. Data handling in analyses

12.1 Number of decimal places for calculated values

In principle, the number of decimal places to which calculated summary statistics and proportions are expressed will be as shown below. Minima, medians and maxima will be rounded off to the first decimal place.

Index	Rules for data expression
p-Value	Rounded off to the fourth decimal place. If below 0.0001, expressed as “<.0001”.
Mean, standard deviation, median	Rounded off to one significant figure of the raw data.
Adjusted mean, standard error	Rounded off to the second decimal place.
Maximum, minimum	The same number of significant figures as the raw data.
Proportion (%)	Rounded off to the first decimal place.

Appendix 1. Schedule

	Screening period	Treatment period								Tapering period	Follow-up period
Visit	1	2	3	4	5	6[b]	7	8	discontinuation	9/discontinuation	10
Evaluation week	−3 to −1	0	1	2	3	4	5	6	–	7	8
Obtaining informed consent or assent	X										
Demographic factors	X										
MINI-KID	X										
Inclusion/exclusion criteria:	X	X									
Physical Examination	X	X	X	X	X	(X)	X	X	X	X	X[f]
Registration	X Pre-registration	X Registration									
CDRS-R	X	X	X	X	X		X	X	X		
CGI-S	X	X	X	X	X	(X)	X	X	X		
Psycho-education	X[c]	X[c]			X[c]			X[c]	X[c]		
C-SSRS	X	X	X	X	X	(X)	X	X	X	X	
Study drug administered		X	X	X	X	(X)	X	X[d]	X[d]		
Adherence of drug administration			X	X	X	(X)	X	X	X	X	
Randomization		X									
Laboratory tests[a]	X				X			X	X	X	
Blood pressure and heart rate	X	X	X	X	X	(X)	X	X	X	X	
ECG	X							X	X	X	
Body weight		X						X	X	X	
Urinary drug screening		X									
Pregnancy test (post-menarchic females only)	X								X[e]	X	
Adverse events	X										

a Hematology tests, blood chemistry tests, and urinalysis. HbA1c, TSH, FT3 and FT4 measurements are only performed at Visit 1.

b Visit 6 is a visit that can be made to the hospital in connection with dose increase, decrease, etc. If Visit 6 is made, observations and tests are performed as far as possible.

c Psycho-education is provided after evaluation of CDRS-R and CGI-S.

d Tapering-period prescription.

e Performed at discontinuation if not performed during the tapering period.

f Adverse events occurring up to 7 days after the final administration of the study drug will be investigated. The subject is checked by telephone or other appropriate method, even if he/she fails to make a visit.

Appendix 2. Permitted range of timing of examinations, observations and tests

	Visit	Week	Stipulated date of visit	Permitted range (day)
Screening period	1	−3 to −1	−21 to −7	–
Treatment period	2	0	0	0
	3	1	7	±3
	4	2	14	±3
	5	3	21	±3
	6[a]	4	28	±3
	7	5	35	±3
	8	6	42	−3 to +1
	At discontinuation	–	Date of final dosing of treatment period	0 to +3
Tapering period	9	7	On day 7, taking Visit 8 (or the time of discontinuation during the treatment period) to be day 0.	0 to +3
	At discontinuation	–	Date of final dosing of tapering period	0 to +3
Follow-up period	10	8	On day 7 or later, taking final administration date to be day 0.	–

a: Visit 6 is a visit that can be made to the hospital in connection with dose increase, decrease, etc.

Appendix 3. Endpoint Scoring

1. Calculation methods for CDRS-R total score and subscale scores

The scores for items 1 to 17 will be totaled, and the CDRS-R total score calculated.

Endpoints	Score range	Endpoints	Score range
1. Impaired Schoolwork	1 to 7	11. Depressed Feelings	1 to 7
2. Difficulty Having Fun	1 to 7	12. Morbid Ideation	1 to 7
3. Social Withdrawal	1 to 7	13. Suicidal Ideation	1 to 7
4. Sleep Disturbance	1 to 5	14. Excessive Weeping	1 to 7
5. Appetite Disturbance	1 to 5	15. Depressed Facial Affect	1 to 7
6. Excessive Fatigue	1 to 7	16. Listless Speech	1 to 5
7. Physical Complaints	1 to 7	17. Hypoactivity	1 to 7
8. Irritability	1 to 7		
9. Excessive Guilt	1 to 7		
10. Low Self-Esteem	1 to 7	Total	17 to 113

The CDRS-R subscale scores will be calculated as shown below.

Subscale	Calculation method
Mood	Total of items 8, 11, 14 and 15
Somatic	Total of items 4, 5, 6, 7, 16 and 17
Subjective	Total of items 9, 10, 12 and 13
Behavioral	Total of items 1, 2 and 3

[redacted]

Reason for signing: Approved	Name: [redacted] Role: Approver Date of signature: 21-Jan-2020 02:29:19 GMT+0000

Reason for signing: Approved	Name: [redacted] Role: Approver Date of signature: 21-Jan-2020 02:58:40 GMT+0000

[redacted]