Depression is a highly prevalent mental disorder, with the World Health Organization (WHO) estimating that more than 264 million people of all ages suffer from depression. Although randomized clinical trials (RCTs) have demonstrated the efficacy of antidepressants, their acute and long-term efficacy is limited, and one-third of patients develop treatment-resistant depression (TRD). Therefore, new treatments, including novel antidepressants with high efficacy, are needed.

On the other hand, there are major problems in the development of antidepressants, one of which is the high placebo response rate. High placebo response rates have been reported not only in mood disorders, Parkinson’s disease, and pain, but also in various disorders such as schizophrenia, substance use disorders, and postoperatively. In particular, clinical trials of antidepressants have reported that the response rate of antidepressants is about 50%, while the response rate of placebo is also high, averaging 31–45%. Approval of an antidepressant requires a statistically significant difference in efficacy to placebo in a double-blind randomized trial. Therefore, the inability to distinguish between antidepressant efficacy and placebo response is a major obstacle to the development of new depression drugs, and many pharmaceutical companies have withdrawn or scaled back their operations in the field of drugs for depression and other mental disorders.

Within this context, it is important to understand why placebo responses occur and whether there is a qualitative difference, such as different underlying mechanisms, between placebo responses and the effects of antidepressants. The mechanism of the placebo response is assumed to be the patient’s expected effect or regression to the mean. On the other hand, studies examining the relationship between clinical symptoms and placebo response have reported that the higher the severity of depression at baseline, the greater the advantage of the active drug over placebo. Technically, the low probability of being assigned to the placebo group and the large number of study sites are also thought to increase the placebo response. There is also the issue of how to evaluate antidepressant effects. In clinical trials of antidepressants, total scores on symptom assessments such as the Hamilton Depression Rating Scale (HAM-D) and the Montgomery-Asberg Depression Rating Scale (MADRS) are used as primary endpoints, and changes in subscales are often not examined. However, there is compelling evidence that depression is not a single construct and that symptoms of major depressive disorder are organized into clusters of two to five. Different types and doses of antidepressants have been reported to improve depressive symptoms in each cluster to different degrees. However, whether there is a difference in the change in each symptom of depression (differences in

Aims: The aim of this study is to determine whether there is difference in the change in each symptom of depression and in symptomatic improvement pattern between placebo and antidepressant responses.

Methods: Using data from a randomized, double-blind (DB), placebo-controlled trial of esketamine (ESK) in patients with treatment-resistant depression (TRD), we conducted exploratory analyses. To determine differences in the change in each depressive symptom on the MADRS subscale between placebo and antidepressant responses, a two-way factorial analysis was conducted using the amount of change on Day 2 and 28 of treatment. In addition, exploratory and confirmatory factor analyses were conducted on the MADRS subtotal variables on Day 2 and 28 of treatment to determine symptomatic improvement pattern between placebo response and antidepressant responses.

Results: We found that as well as MADRS total score, each subscale of MADRS score did not significantly differ between esketamine and placebo at Day 2 and 28. On the other hand, factor analysis revealed that the factor structure of the response was different between esketamine and placebo at the 2nd day. There was no difference in the factor structure between esketamine and placebo in response on Day 28 of treatment.

Conclusion: Factor analysis revealed different patterns of symptom improvement in the early phase of the intervention between esketamine and placebo. This finding suggests that a data driven approach may provide detailed efficacy information in clinical trials for antidepressants.

Clinical Trial Registration: ClinicalTrials.gov Identifier: NCT02918318. Registered: 28 September 2016.

Keywords: clinical trial, depression, factor analysis, ketamine, placebo response.
subscapes) between the response to placebo and the response to antidepressants, and whether there is a difference in the symptom clusters of responsiveness to placebo and antidepressants has remained to be clarified.

In this study, we conducted an exploratory analysis using data from a clinical trial of nasal spray of esketamine (ESK) in patients with treatment-resistant depression (TRD). The purpose of this study is to determine the following points.

1. Is there a difference in the change in each depressive symptom measured by MADRS subscales between placebo responses and responses to antidepressants?
2. Is there a difference between pattern of clusters of symptom improvement caused by antidepressant and clusters of symptom improvement caused by placebo responses?

**Methods**

We conducted a randomized trial of ESK vs. placebo in patients with treatment-resistant depression. Details of the study have previously been published. The primary efficacy endpoint was change from baseline in the MADRS total score to the end of 4-week double blind induction phase. However, there was no significant difference in MADRS improvement between ESK treatment arms and a placebo arm. The main reason for the lack of significant difference between the two groups was inferred to be the high placebo response in the placebo group.

**Study design**

A Phase 2b, randomized, double-blind (DB), placebo-controlled, multicenter study was conducted. It consisted of the following five phases: (i) a 4-week screening phase, (ii) a 6-week open-label prospective lead-in phase, (iii) a 4-week DB induction phase, (iv) a late treatment phase of up to 24 weeks including an optional 4-week open-label induction phase, and (v) a 4-week follow-up phase.

During the prospective lead-in phase, patients received a new oral antidepressant therapy and non-responders to this oral antidepressant before randomization and at randomization were considered to meet the definition of TRD and entered the 4-week double blind induction phase. The definition of non-responder at prospective lead-in phase was patients who achieved ≤25% improvement in the MADRS total score during the prospective lead-in phase and at randomization and patients with a MADRS total score of ≥28 at 2 weeks.

During the double-blind induction phase, patients were randomized to placebo or fixed dose ESK nasal spray (28 mg, 56 mg, or 84 mg) (2:1:1:1; hereafter ESK 28, ESK56, or ESK84, respectively) in addition to oral antidepressant therapy, which continued unchanged from the prospective lead-in phase. Regarding the late treatment phase and the follow-up phase, details were described in a previous report. The study protocol and all amendments were approved by the institutional review board (IRB) at each study site (representative of IRB: National Center Hospital, National Center of Neurology and Psychiatry). The study was conducted in compliance with the ethical principles of the Declaration of Helsinki, International Council for Harmonization (ICH) Good Clinical Practice (GCP), and applicable regulatory requirements. Informed consent was obtained from all the participants.

**Statistical analyses**

We analyzed 183 out 202 patients for whom a final evaluation was obtained.

1. Two-way factorial Analysis of Variance (ANOVA)

A two-way factorial ANOVA was conducted to identify differences in the change in each depressive symptom between the placebo and antidepressant responses.

Patients were divided following four categories.

1. Responders in the placebo arm.
2. Responders in the active drug arms.
3. Non-responders in the placebo arm.
4. Non-responders in the active drug arms.

Responders were defined as those who achieved ≥50% reduction from baseline in the MADRS total score at the end of DB induction phase (28 days after DB induction), and others were defined as non-responders. Therefore, the definition of non-responders at the end of the DB induction phase was different from the definition of non-responders in the lead-in phase. In the active drug arms, we treated as one group regardless of active drug doses, because we could not find a relationship between dose and efficacy. Two-way factorial

| Table 1. Patients’ demographics | All cases | Placebo | Esketamine | ANOVA |
|--------------------------------|----------|---------|------------|-------|
|                                | 183      | 45      | 27         | 70    | 41    | P     |
| Gender                         |          |         |            |       |       |       |
| Age                            | 43.24 ± 10.36 | 44.20 ± 11.67 | 41.19 ± 9.77 | 42.24 ± 9.58 | 45.24 ± 10.41 | 0.306 |
| MADRS subscale                 |          |         |            |       |       |       |
| Reported Sadness               | 4.24 ± 0.92 | 4.44 ± 0.92 | 4.30 ± 1.07 | 4.06 ± 0.88 | 4.29 ± 0.84 | 0.155 |
| Apparent Sadness               | 4.43 ± 0.91 | 4.49 ± 0.87 | 4.44 ± 1.01 | 4.37 ± 0.90 | 4.46 ± 0.92 | 0.912 |
| Inner Tension                  | 3.73 ± 0.82 | 3.58 ± 0.78 | 3.93 ± 0.83 | 3.76 ± 0.79 | 3.73 ± 0.90 | 0.367 |
| Reduced Sleep                  | 4.11 ± 1.26 | 3.71 ± 1.32 | 4.37 ± 0.88 | 4.27 ± 1.21 | 4.10 ± 1.41 | 0.078 |
| Reduced Appetite               | 3.63 ± 1.20 | 3.67 ± 1.21 | 3.93 ± 1.24 | 3.36 ± 1.18 | 3.85 ± 1.13 | 0.079 |
| Concentration Difficulties     | 3.95 ± 0.89 | 4.00 ± 0.95 | 4.19 ± 0.92 | 3.71 ± 0.87 | 4.15 ± 0.76 | 0.029 |
| Lassitude                      | 3.77 ± 0.87 | 3.91 ± 0.79 | 3.78 ± 0.89 | 3.60 ± 0.94 | 3.88 ± 0.81 | 0.218 |
| Inability to Feel              | 4.04 ± 0.92 | 4.11 ± 0.78 | 3.81 ± 1.00 | 4.04 ± 0.92 | 4.12 ± 1.00 | 0.532 |
| Pessimistic Thoughts           | 3.32 ± 0.90 | 3.27 ± 0.94 | 3.41 ± 0.75 | 3.27 ± 0.93 | 3.41 ± 0.89 | 0.785 |
| Suicidal Thoughts              | 2.30 ± 1.31 | 2.22 ± 1.43 | 2.15 ± 1.26 | 2.39 ± 1.42 | 2.34 ± 1.04 | 0.838 |
| MADRS total score              | 37.52 ± 5.77 | 37.40 ± 5.58 | 38.30 ± 6.15 | 36.83 ± 5.78 | 38.34 ± 5.76 | 0.508 |
Table 2. Result of ANOVA

|                          | Result of ANOVA (Day 2 of treatment) | Result of ANOVA (Day 28 of treatment) |
|--------------------------|--------------------------------------|---------------------------------------|
|                          | Sum of Squares | Mean Squares | F value | P value | Sum of Squares | Mean Squares | F value | P value |
| Reported Sadness (n = 182) |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 0.32          | 0.323        | 0.209   | 0.648   | 2.63          | 2.632        | 2.082   | 0.151   |
| Treatment response       | 11.86         | 11.859       | 7.681   | 0.006   | 250.2         | 250.2        | 197.935 | <0.001  |
| Interaction              | 0.21          | 0.212        | 0.137   | 0.711   | 0.11          | 0.115        | 0.091   | 0.763   |
| Residual                 | 274.81        | 1.544        |         |         | 226.27        | 1.264        |         |         |
| Apparent Sadness (n = 182)|                         |             |         |         |                         |             |         |         |
| Treatment arm            | 2.82          | 2.82         | 1.256   | 0.264   | 0.83          | 0.833        | 0.543   | 0.462   |
| Treatment response       | 9.93          | 9.928        | 4.423   | 0.037   | 239.24        | 239.238      | 155.84  | <0.001  |
| Interaction              | 3.36          | 3.362        | 1.498   | 0.223   | 0.37          | 0.369        | 0.27    | 0.604   |
| Residual                 | 399.58        | 2.245        |         |         | 274.79        | 1.535        |         |         |
| Inner Tension (n = 182)  |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 4.75          | 4.748        | 2.683   | 0.103   | 0.01          | 0.007        | 0.005   | 0.942   |
| Treatment response       | 22.26         | 22.259       | 12.579  | <0.001  | 187.09        | 187.09       | 137.221 | <0.001  |
| Interaction              | 1.82          | 1.815        | 1.026   | 0.312   | 0.37          | 0.369        | 0.27    | 0.604   |
| Residual                 | 314.98        | 1.77         |         |         | 244.05        | 1.363        |         |         |
| Reduced Sleep (n = 182)  |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 5.71          | 5.714        | 2.025   | 0.156   | 0.03          | 0.026        | 0.011   | 0.915   |
| Treatment response       | 10.73         | 10.727       | 7.188   | 0.008   | 163.94        | 163.938      | 98.585  | <0.001  |
| Interaction              | 1.14          | 1.139        | 0.763   | 0.384   | 0.24          | 0.242        | 0.146   | 0.703   |
| Residual                 | 265.62        | 1.492        |         |         | 297.66        | 1.663        |         |         |
| Concentration Difficulties (n = 182) |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 1.04          | 1.039        | 0.665   | 0.416   | 0.18          | 0.18         | 0.117   | 0.732   |
| Treatment response       | 20.66         | 20.659       | 13.22   | <0.001  | 145.31        | 145.309      | 94.841  | <0.001  |
| Interaction              | 0.05          | 0.055        | 0.035   | 0.852   | 0.01          | 0.006        | 0.004   | 0.948   |
| Residual                 | 278.16        | 1.563        |         |         | 274.25        | 1.532        |         |         |
| Inability to Feel (n = 182) |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 1.31          | 1.306        | 0.873   | 0.351   | 0.4           | 0.403        | 0.281   | 0.597   |
| Treatment response       | 6.62          | 6.616        | 4.423   | 0.037   | 161.85        | 161.849      | 112.857 | <0.001  |
| Interaction              | 0.25          | 0.251        | 0.168   | 0.683   | 0.79          | 0.791        | 0.552   | 0.459   |
| Residual                 | 266.25        | 1.496        |         |         | 256.71        | 1.434        |         |         |
| Pessimistic Thoughts (n = 182) |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 5.96          | 5.956        | 3.316   | 0.07    | 0.44          | 0.435        | 0.338   | 0.562   |
| Treatment response       | 9.81          | 9.812        | 5.463   | 0.021   | 135.63        | 135.629      | 105.373 | <0.001  |
| Interaction              | 0.63          | 0.626        | 0.349   | 0.556   | 0.01          | 0.01         | 0.008   | 0.931   |
| Residual                 | 319.7         | 1.796        |         |         | 230.4         | 1.287        |         |         |
| Suicidal Thoughts (n = 182) |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 2.24          | 2.243        | 1.41    | 0.237   | 0.63          | 0.634        | 0.455   | 0.501   |
| Treatment response       | 0.53          | 0.529        | 0.333   | 0.565   | 56.71         | 56.709       | 40.73   | <0.001  |
| Interaction              | 0.22          | 0.223        | 0.14    | 0.709   | 0.02          | 0.019        | 0.013   | 0.908   |
| Residual                 | 283.23        | 1.591        |         |         | 249.23        | 1.392        |         |         |
| Total (n = 182)          |                         |             |         |         |                         |             |         |         |
| Treatment arm            | 142.58        | 142.58       | 2.114   | 0.148   | 7.47          | 7.473        | 0.149   | 0.7     |
| Treatment response       | 992.09        | 992.093      | 14.706  | <0.001  | 16480.53      | 16480.53     | 329.631 | <0.001  |
ANOVA were formed to analyze the effect of type of treatment groups (placebo or ESK) and treatment response (responder or non-responder) on each MADRS subscale. Analyses were performed independently on data from Day 2 and 28 of treatment.

1 Exploratory / Confirmatory factor analysis

A set of factor analyses was conducted sequentially to determine if there was a qualitative difference in symptom change between placebo response and the response to antidepressants. First, exploratory factor analyses (EFA) were performed to determine the number of factors in the ESK and placebo groups, using the change in the MADRS subscales as variables on Day 2 and 28 of treatment. Analyses were performed for each treatment arm regardless of treatment response (responder or non-responder). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was used to evaluate the sufficiency of conducting factor analysis with our data. To determine the number of factors, we used Parallel analysis. In parallel analysis, random data having numbers of the same sample size as the actual data are generated, and extracted. Then, the eigenvalues were plotted in the same way as the eigenvalues from the actual data. Factors above the intersection of the simulation line (a plot of the eigenvalues of the random correlation matrix) and the trend of the eigenvalues of the actual data were determined to be “meaningful factors”. To increase the readability of results, oblimin rotation was performed. Second, confirmatory factor analysis (CFA) was performed to evaluate the factor structures suggested by the previous EFA in the ESK and placebo groups. The model fit with the data was evaluated by utilizing Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA). A model showing CFI>0.9, TLI>0.9 and RMSEA <0.08 is considered to be a good fit to the data. All statistical analyses were performed with R version 3.5.3. (The R Foundation for Statistical Computing, Vienna, Austria). Alpha level was set at 0.05.

Results

Study population and demographic data

Table 2 shows the demographic data of the cases included in the analysis. Of the 183 patients analyzed, 72 in the placebo group and 111 in the ESK group, there were 27 responders in the placebo group and 41 responders in the ESK group, respectively. There were no significant differences in gender, age, and total MADRS score at baseline among the four groups. There was also no significant difference in scores of subscales of the MADRS at baseline except subscale of concentration difficulties among the four groups.

Results of analysis of variance (ANOVA) for changes in MADRS subscales

One patient had missing data for MADRS subscales on the second day of treatment; therefore, ANOVA was performed using the data of 182 patients.

Table 2 shows the results of ANOVA for changes in MADRS subscales on Day 2 and 28 of treatment. Two-way analysis of variance of Day 2 data revealed no statistically significant interaction between the effects of type of treatment groups (placebo or ESK) and treatment response (non-response or response) on all MADRS subscales (Table 2). In a simple main effects analysis, the type of treatment group had no statistically significant effect on all MADRS subscales. On the other hand, a simple main effect analysis showed that treatment response had a statistically significant effect on MADRS subscales other than reduced sleep and suicidal thoughts (Reported Sadness =0.006, Apparent Sadness; P = 0.037, Inner Tension; P<0.001,
Such factors may contribute to this, and these were dis-

ting the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the samples were factorable (ESK; MSA = 0.81, placebo: MSA = 0.83). In the ESK treatment group, three factors were extracted in EFA of the response to treatment by using parallel analysis on data of Day 2 of treatment (Supplementary Fig. 1a). The CFA with three factor solution showed a good fit to the data (CFI = 0.956, TLI = 0.938, RMSEA = 0.065). Two items, reported sadness and apparent sadness, loaded onto Factor 1 (Fig. 1a, Table 3). This factor was labeled, “Sadness”. Five items load onto a second factor related to vegetative symptoms of depression, such as reduced sleep and reduced appetite, and amotivational symptoms, such as concentration difficulties, lassitude, and inability to feel. This factor was labeled, “amotivational and vegetative” (Fig. 1a, Table 3). The three items that load onto Factor 3 relate to negative thoughts, such as pessimistic thoughts and suicidal thoughts, and inner tension. This factor was labeled, “negative emotion and thoughts” (Fig. 1a, Table 3). On the other hand, in the placebo group, parallel analysis revealed one factor to explain response to treatment at Day 28 in the ESK group and placebo group (Fig. 1b, Supplementary Fig. 1c,1d).

**Results of exploratory/confirmatory factor analysis**

**Results of treatment response at Day 2**

The examinations of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the samples were factorable (ESK; MSA = 0.94, placebo: MSA = 0.88). The parallel analysis revealed one factor to explain response to intervention at Day 28 in the ESK and placebo group. Reduced sleep; $P = 0.343$, Reduced Appetite; $P = 0.008$, Concentration Difficulties; $P < 0.001$, Lassitude; $P < 0.001$, Inability to Feel; $P = 0.037$, Pessimistic Thoughts; $P = 0.021$, Suicidal thoughts; $P = 0.565$). Two-way analysis of variance of Day 28 data revealed no statistically significant interaction between the effects of type of treatment groups (placebo or ESK) and treatment response (non-response or response) on all MADRS subscales (Table 2). In a simple main effects analysis, the type of treatment group had no statistically significant effect on all MADRS subscales. On the other hand, a simple main effect analysis showed that treatment response had a statistically significant effect on all MADRS subscales (all subscales: $P < 0.001$).

**Results of treatment response at Day 28**

The examinations of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the samples were factorable (ESK; MSA = 0.94, placebo: MSA = 0.88). The parallel analysis revealed one factor to explain response to treatment at Day 28 (Supplementary Fig. 1b). The CFA with one factor solution also showed a good fit to the data (CFI = 0.955, TLI = 0.942, RMSEA = 0.063).

**Table 3. Results of confirmatory factor analysis of response to Esketamine (Day 2 of treatment)**

| MADRS subscale | Factor 1  | Factor 2  | Factor 3  | $h^2$ | $c^2$ | comp. |
|----------------|-----------|-----------|-----------|------|------|-------|
| Concentration diff. | 0.071     | 0.639     | 0.040     | 0.490 | 0.510 | 1.03  |
| Reduced sleep     | −0.133    | 0.644     | −0.024    | 0.338 | 0.662 | 1.09  |
| Lassitude         | 0.096     | 0.481     | 0.104     | 0.360 | 0.640 | 1.18  |
| Pessimistic thoughts | −0.022   | 0.074     | 0.799     | 0.690 | 0.310 | 1.02  |
| Reduced appetite  | 0.018     | 0.450     | 0.089     | 0.264 | 0.736 | 1.08  |
| Inner tension     | 0.247     | 0.231     | 0.369     | 0.491 | 0.509 | 2.50  |
| Inability to feel | 0.345     | 0.401     | −0.030    | 0.389 | 0.611 | 1.97  |
| Suicidal thoughts | 0.038     | −0.108    | 0.655     | 0.386 | 0.614 | 1.06  |
| Apparent sadness  | 0.800     | 0.004     | 0.124     | 0.759 | 0.241 | 1.05  |
| Reported sadness  | 0.911     | 0.001     | −0.064    | 0.775 | 0.225 | 1.01  |
| SS loadings       | 1.857     | 1.676     | 1.409     |      |      |       |
| Proportion Var    | 0.186     | 0.168     | 0.141     |      |      |       |
| Cumulative Var    | 0.186     | 0.353     | 0.494     |      |      |       |
| Proportion Explained | 0.376   | 0.339     | 0.285     |      |      |       |
| Cumulative Proportion | 0.376    | 0.715     | 1.000     |      |      |       |

In this study, we performed ad-hoc analyses of data from failed clinical trial for developing an antidepressant for TRD. The primary efficacy endpoint was change from baseline in the MADRS total score at Day 28 in double blind phase. However, there was no significant difference of change of MADRS total score between active drug and placebo. Several factors may contribute to this, and these were discussed in the previous manuscript.

In this analysis, we found that as well as MADRS total score, each subscale of MADRS score did not differ between ESK and placebo at early response (Day 2 of intervention) and late phase of response (Day 28 of intervention). On the other hand, factor analysis revealed that the factor structure of the response was different between ESK and placebo at the early phase of the intervention (Day 2). There was no difference in the factor structure between ESK and placebo in response on Day 28 of the intervention. These results suggest that placebo and ESK should not fully share the same underlying mechanism or effect-mediated architecture for improving depressive symptoms, even if there was no statistical difference in the change in depressive symptoms as measured by the MADRS score.

Neuroscience studies using positron emission tomography have reported that placebo-induced endogenous opioid release in brain regions involved in the pathogenesis of depression, such as the anterior cingulate cortex below the knee, nucleus accumbent, thalamic core, and amygdala, is associated with symptom improvement. Such data indicates that the placebo response is actually caused by changes in disease-modifying neurotransmitters in brain regions closely related to depressive symptoms, such as emotion and motivation. In this context, neurobiological mechanisms behind antidepressants effect and placebo response are essentially the same, therefore, it seems to be difficult to distinguish placebo response from effect of antidepressant. On the other hand, our factor analysis revealed that early response to ESK has a different factor structure from that of placebo response. If the placebo response is mediated by the same mechanism and architecture as the drug-induced symptom improvement, then the factor structure should be the same for both the placebo response and the drug-induced response. Therefore, we speculate that placebo and ESK would not fully share the same underlying mechanism or effect-
mediated architecture for improving depressive symptoms. Mayberg et al. studied the functional neuroanatomy of placebo and antidepressant (fluoxetine) responses using positron emission tomography (PET) and reported that placebo and antidepressant responses share common neural mechanisms (architecture), however, they also found fluoxetine specific changes in subcortical and limbic areas. They suggested that such additional neural mechanisms of fluoxetine may contribute to additional benefits in maintaining long-term clinical efficacy and adherence. We speculate that such effects may be proven by combining factor analysis of antidepressant response with evaluation of neural activity by PET or functional magnetic resonance imaging, and evaluating neural correlations with factor scores of each cluster of symptomatic improvement caused by antidepressant.

The MADRS as the most extensively used clinician rated measure of depressive severity in clinical research. The use of the MADRS to monitor changes in depression over the course of treatment requires that the psychometric properties of the instrument should be robust over time. However, stability of the factor structure during the treatment period is questioned by evidence that the structure of MADRS, which was multiple factors at the start of treatment, changes to a single factor after more than a month of treatment. In this analysis, we conducted factor analysis by using changes of MADRS subscales at Day 2 (early phase) and Day 28 rather than MADRS subscales of each time point. Despite the differences of methodology, the current analysis also showed that ESK treatment changed the factor structure over the treatment period, similar to previous studies. On the other hand, the placebo response showed a single factor both on Day 2 and Day 28 of treatment. A previous study revealed a hierarchical model of depressive syndromes in which the domains of sadness, negative thoughts, detachment and neurovegetative domains of the depressive syndrome loaded onto a second order overall depression factor. Similar to the previous study, the present analysis revealed that the early effects of ESK on depressive symptoms were divided into following three factors: ‘sadness’, ‘amotivational and vegetative’, and ‘negative emotion and thoughts’. On the other hand, both early and late placebo response can be explained by a single factor, and it indicates that placebo response may be a simple structure that each variable loads highly onto one and only one factor. Another possibility is due to the study design. In this study, patients received other oral antidepressants from the induction phase until the end of the randomization period, regardless of whether they were assigned to the ESK or placebo group. Thus, the placebo group had received more than one month of oral antidepressant treatment at 2 days after randomization. As mentioned above, it has been reported that the structure of MADRS, which was multifactorial at the start of treatment, changes to a single factor after more than one month of treatment. We believe that the results may be due to the study design. In conclusion, we found a different pattern of clusters of symptoms evaluated by 16-item self-report Quick Inventory of Depressive Symptomatology (QIDS-SR) checklist and Hamilton Depression (HAM-D) rating scale. Furthermore, for each symptom cluster, they found a significant difference in the therapeutic effect between antidepressants. Although by using a different approach, Sakurai et al. demonstrated that there are some differences in trajectories of individual symptoms over time between treatments for depression. These suggest that selecting the best drug for a particular cluster may provide a benefit greater than that obtained by using the active compound compared to placebo. Clinical trials for developing antidepressants have used total scores on questionnaires (such as MADRS and HAM-D) as the endpoint that include a diverse range of symptoms despite evidence that depression is not a unitary construct. Including this study, several studies suggest that, in addition to traditional statistical approach, data driven approach such as a factor analysis and a clustering may provide detailed efficacy information in clinical trials for antidepressants. In the development of new drugs for neuropsychiatric disorders, we believe that the efficiency of clinical trials can be improved by targeting symptom groups that match the symptom improvement patterns of drugs, or by selecting patient populations based on symptom clusters, rather than just specific diseases based on operational diagnostic criteria. In clinical practice, such an approach would allow us to characterize the symptomatic efficacy of each antidepressant and to select the appropriate drug for each symptom cluster in depressed patients.

The major strength of this method is that it is a rigorous randomized trial. However, there are some limitations to this study. These are as follows: (i) the relatively small sample size (ii) no independent sample to confirm EFA, and (iii) the sample of participants is treatment-resistant depression, which may not reflect the continuum of patients with mood disorders and healthy individuals. In conclusion, we found a different pattern of clusters of symptomatic improvement in the early phase of the intervention between ESK and placebo, even if there is no significant difference in the effects of ESK and placebo on MADRS total scores and each MADRS subscale. Such difference may be related to the characteristics of the effects of interventions, such as antidepressants and placebos, on symptom improvement. It suggests that, in addition to a traditional statistical approach, a data driven approach such as a factor analysis may provide detailed efficacy information in clinical trials for antidepressants.

Availability of Data and Materials
The data sharing policy of Janssen Pharmaceutical Companies of Johnson & Johnson is available at https://www.janssen.com/clinical-trials/transparency. As noted on this site, requests for access to the study data can be submitted through Yale Open Data Access (YODA) Project site at http://yoda.yale.edu.
Acknowledgment
No acknowledgement

Disclosure statement
This clinical trial was conducted by Janssen Pharmaceutical K.K. of Johnson & Johnson in Japan. All authors contributed to interpretation of the results, developed the draft of the manuscript, participated in subsequent revisions, and read and approved the final manuscript. All authors have disclosed that they are full-time employees of Janssen Pharmaceutical K.K. of Johnson & Johnson in Japan.

References
1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the global burden of disease study 2017 [published correction appears in Lancet. 2019 Jun 22;393(10190):e44]. Lancet 2018; 392: 1789–1858.

2. Köhler-Forschberg O, Cusin C, Nierenberg AA. Evolving issues in the measurement of placebo: A systematic review of migraine prophylaxis. JAMA Internal Medicine 2013; 170 1335–1344 [PubMed] [Google Scholar].

3. Stolk P, Ten Berg MJ, Hemels ME, Einarson TR. Meta-analysis of placebo rates in major depressive disorder trials. The Annals of Pharmacotherapy 2003; 37: 1891–1899 [PubMed] [Google Scholar].

4. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: Variable, substantial, and growing. JAMA 2002; 287: 1840–1847 [PubMed] [Google Scholar].

5. Khan A, Bhat A, Kolts R, Thase ME, Brown W. Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades? CNS Neurosci. Ther. 2010; 16: 217–226.

6. Kesler SE, Mauro AD, Smith TL, Warren JS, Coccaro EF. Placebo effect in the treatment of depression and anxiety. Front. Psychiatry 2019; 10: 407. Published 2019 Jun 13. doi: 10.3389/fpsyt.2019.00407.

7. Fournier JC, DeRubeis RJ, Hollon SD et al. Antidepressant drug effects and depression severity: A patient-level meta-analysis. JAMA 2010; 303: 47–53.

8. Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G. Reevaluating the efficacy and predictability of antidepressant treatments: A symptom clustering approach. JAMA Psychiat. 2017; 74: 370–378 PMID: 28241180; PMCID: PMC5863470.

9. Takahashi N, Yamada A, Shiraiishi A, Shimizu H, Goto R, Tomimagaya Y. Efficacy and safety of fixed doses of intranasal esketamine as an add-on therapy to Oral antidepressants in Japanese patients with treatment-resistant depression: A phase 2b randomized clinical study. BMC Psychiatry 2021; 21: 526 PMID: 34696742.

10. Peciña M, Bohnert AS, Sikora M et al. Association between placebo-activated neural systems and antidepressant response: Neurochemistry of placebo effects in major depression. JAMA Psychiat. 2015; 72: 1087–1094.

11. Mayberg HS, Silva JA, Brannan SK et al. The functional neuroanatomy of the placebo effect. Am. J. Psychiatry 2002; 159: 726–737 PMID: 11986125.

12. Santor DA, Gregus M, Welch A. FOCUS ARTICLE: Eight decades of measurement in depression. Measurement: Interdisciplinary Research and Perspectives 2006; 4: 135–155.

13. Galinowski A, Lehert P. Structural validity of MADRS during antidepressant treatment. Int. Clin. Psychopharmacol. 1995; 10: 157–161 PMID: 8675968.

14. Rocca P, Fonzo V, Ravizza L et al. A comparison of paroxetine and amisulpride in the treatment of dysthymic disorder. J. Affect. Disord. 2002; 70: 313–317 PMID: 12128243.

15. Quilty LC, Robinson JI, Rolland JP, Fryt FD, Rouillon F, Bagby RM. The structure of the Montgomery-Asberg depression rating scale over the course of treatment for depression. Int. J. Methods Psychiatr. Res. 2013; 22: 175–184.

16. Ballard ED, Yarrington JS, Farmer CA et al. Parsing the heterogeneity of depression: An exploratory factor analysis across commonly used depression rating scales. J. Affect. Disord. 2018; 231: 51–57.

17. Hashimoto K. Rapid-acting antidepressant ketamine, its metabolites and other candidates: A historical overview and future perspective. Psychiatry Clin Neurosci. 2019; 73: 613–627. Epub 2019 Jul 11. PMID: 31215725; PMCID: PMC6851782.

18. Hasler G, Suler S, Shoretanis G, Mihov Y. Sustained improvement of negative self-schema after a single ketamine infusion: An open-label study. Front. Neuosci. 2020; 1: 687 PMID: 32714138; PMCID: PMC7341975.

19. Rush AJ, Trivedi MH, Ibrahim HM et al. The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003; 54: 573–583 Epub in: Biol Psychiatry 2003 Sep 1;54(5):573, PMID: 12946886.

20. Hamilton M. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 1960; 23: 56–62 PMID: 14399272; PMCID: PMC495331.

21. Sakurai H, Uchida H, Abe T et al. Trajectories of individual symptoms in remitters versus non-remitters with depression. J. Affect. Disord. 2013; 151: 506–513 Epub 2013 Jul 22. PMID: 23886402.

22. Sakurai H, Uribe S, Cirillo P, Fuertes-Saiz A, Camprodon JA, Barbour T. Residual symptoms after achieving remission with repetitive transcranial magnetic stimulation in depression. J. Affect. Disord. 2022; 15: 154–161 Epub 2022 Jan 5. PMID: 34998805.

Supporting information
Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

Supplementary Fig. 1a A scree plot of Day 2 treatment response of esketamine (Parallel analysis).

The solid line represents a scree plot with actual data, and the dotted line represents a simulation with random data. In the parallel analysis, a scree plot of actual data and random data is created, and the meaningful factors are determined from the intersection of the two. In this figure, the actual data and random data intersect at factor 4, therefore, three factors are extracted as meaningful factors.

Supplementary Fig. 1b A scree plot of Day 2 treatment response of placebo (Parallel analysis). In this figure, one factor is extracted as a meaningful factor because the actual data and the random data intersect at factor 2.

Supplementary Fig. 1c A scree plot of Day 28 treatment response of esketamine (Parallel analysis) In this figure, one factor is extracted as a meaningful factor because the real data and the random data intersect at factor 2.

Supplementary Fig. 1d A scree plot of Day 28 Placebo response (Parallel analysis) In this figure, one factor is extracted as a meaningful factor because the real data and the random data intersect at factor 2.