

Original Research

Cite this article: Sliwa JK, Naranjo RR Jr, Turkoz I, Petrillo MP, Cabrera P, and Trivedi M (2024). Effects of esketamine nasal spray on depressive symptom severity in adults with treatment-resistant depression and associations between the Montgomery–Åsberg Depression Rating Scale and the 9-item Patient Health Questionnaire. *CNS Spectrums*
<https://doi.org/10.1017/S1092852924000105>

Received: 19 July 2023

Accepted: 26 February 2024

Keywords:

Depression; severity; esketamine nasal spray; oral antidepressant; treatment resistant

Corresponding author:

Jennifer Kern Sliwa;

Email: JKernsli@its.jnj.com

Effects of esketamine nasal spray on depressive symptom severity in adults with treatment-resistant depression and associations between the Montgomery–Åsberg Depression Rating Scale and the 9-item Patient Health Questionnaire

Jennifer Kern Sliwa¹ , Ronaldo R. Naranjo Jr², Ibrahim Turkoz³, Mary Pat Petrillo⁴, Patricia Cabrera⁵ and Madhukar Trivedi⁶ 

¹CNS Medical Information, Janssen Scientific Affairs, LLC, Titusville, NJ, USA, ²US Neuroscience Medical Affairs, Janssen Scientific Affairs, LLC, Titusville, NJ, USA, ³Statistics & Decision Sciences, Janssen Research & Development, LLC, a Johnson & Johnson company, Titusville, NJ, USA, ⁴Value & Evidence Scientific Engagement, Janssen Scientific Affairs, LLC, Titusville, NJ, USA, ⁵Neuroscience, Janssen Scientific Affairs, LLC, Titusville, NJ, USA and ⁶Center for Depression Research and Clinical Care, Peter O'Donnell Jr. Brain Institute and Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA

Abstract

Objective. To examine the effect of esketamine nasal spray (ESK) plus newly initiated oral antidepressant (OAD) versus OAD plus placebo nasal spray (PBO) on the association between Montgomery–Åsberg Depression Rating Scale (MADRS) and 9-item Patient Health Questionnaire (PHQ-9) scores in adults with treatment-resistant depression (TRD).

Methods. Data from TRANSFORM-1 and TRANSFORM-2 (two similarly designed, randomized, active-controlled TRD studies) and SUSTAIN-1 (relapse prevention study) were analyzed. Group differences for mean changes in PHQ-9 total score from baseline were compared using analysis of covariance. Associations between MADRS and PHQ-9 total scores from TRANSFORM-1/TRANSFORM-2 were assessed using simple parametric, nonparametric, and multiple regression models.

Results. In TRANSFORM-1/TRANSFORM-2 (ESK + OAD, $n = 343$; OAD + PBO, $n = 222$), baseline PHQ-9 mean scores were 20.4 for ESK + OAD and 20.6 for OAD + PBO (severe depression). At day 28, significant group differences were observed in least squares mean change (SE) in PHQ-9 scores from baseline ($-12.8 [0.46]$ vs $-10.3 [0.53]$, $P < .001$) and in clinically substantial change in PHQ-9 scores (≥ 6 points; 77.1% vs 64%, $P < .001$) in ESK + OAD and OAD + PBO groups, respectively. A nonlinear relationship between MADRS and PHQ-9 was observed; total scores demonstrated increased correlation over time. In SUSTAIN-1, 57.3% of patients receiving ESK + OAD ($n = 89$) versus 44.2% receiving OAD + PBO ($n = 86$) retained remission status (PHQ-9 score ≤ 4) at maintenance treatment end point ($P = .044$).

Conclusions. In adults with TRD, ESK + OAD significantly improved severity of depressive symptoms, and more patients achieved clinically meaningful changes in depressive symptoms based on PHQ-9, versus OAD + PBO. PHQ-9 outcomes were consistent with those of clinician-rated MADRS.

Trial registration. ClinicalTrials.gov: NCT02417064, NCT02418585, NCT02493868.

Introduction

Major depressive disorder (MDD) is a leading cause of disability and a major contributor of disease burden worldwide.^{1,2} Estimates of global prevalence among adults is 5.0%, which appears to increase with age.² In the United States, the 12-month prevalence among adults is more than double this number, at 10.4%, with many patients demonstrating moderate or severe symptoms.³ Much of the disability associated with MDD is a result of the substantial proportion of patients who do not respond adequately to treatment. Approximately one-third of patients with MDD have treatment-resistant depression (TRD),⁴ commonly defined as inadequate response to two or more oral antidepressants (OADs) of adequate dose and duration in the current episode.⁵

Patient-reported outcomes have increasingly demonstrated value in assessing therapeutic efficacy, effectiveness, and safety of MDD treatment strategies and have garnered interest from academic, industry, and regulatory stakeholders alike.^{6–10} Contemporary guidance on trial

© Janssen Scientific Affairs, LLC, 2024.

Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



design from the United States Food and Drug Administration and European Medicines Agency has focused on clinical outcome end points that assess how patients feel, function, and cope. A combination of patient-reported, observer-reported, clinician-reported, and performance outcome measures best assesses clinical benefits and risks of therapeutic strategies, patient experience, and disease progression. Patient experience data are valuable considerations when evaluating the efficacy and safety of new treatments. Patient self-rating scales are a valuable way to incorporate experience from the patient perspective when evaluating efficacy.

The Montgomery-Åsberg Depression Rating Scale (MADRS) is a 10-item clinician-reported questionnaire¹¹ used to measure the severity of MDD symptoms. Although it is accepted by regulatory authorities as an appropriate primary efficacy measure among patients diagnosed with TRD, the MADRS scale is most likely to be used in a clinical research setting and is not regularly implemented in routine clinical practice. The 9-item Patient Health Questionnaire (PHQ-9) is a patient-reported outcome measure that can provide the patient perspective of MDD and has been validated for use in real-world settings.¹²

The PHQ-9 can assess depressive symptom severity, monitor treatment effect, and provide additional information with which to evaluate depressive symptoms.^{13,14} Additionally, the PHQ-9 offers an effective cross-cultural measurement of invariance across genders, races, and ethnicities¹⁵ and is easy to use, understand, implement, and analyze because data can also be gathered via electronic medical records. Both assessments provide valuable insight into the efficacy of therapeutic strategies; however, there is little research evaluating relationships between MADRS and PHQ-9 scores.^{7,13} Understanding this relationship will allow clinicians to more precisely assess the severity of depression at each visit and the success of a treatment for reducing depressive symptoms.

Esketamine nasal spray (ESK) is a noncompetitive *N*-methyl-*D*-aspartate receptor antagonist approved by the United States Food and Drug Administration, in conjunction with an OAD, for the treatment of TRD in adults and for the treatment of depressive symptoms in adults with MDD with acute suicidal ideation or behavior.¹³

In the TRANSFORM-1 (NCT02417064)¹⁶ and TRANSFORM-2 (NCT02418585)¹⁷ phase 3 trials of ESK plus a newly initiated OAD (ESK + OAD) in adults with TRD, clinically meaningful reductions in MADRS total score were observed with ESK + OAD versus OAD + placebo nasal spray (OAD + PBO) after 4 weeks of treatment (primary end point). Eligible patients from TRANSFORM-1 or TRANSFORM-2 could enroll in SUSTAIN-1 (NCT02493868),¹⁸ a longer-term relapse-prevention study that evaluated ESK + OAD versus OAD + PBO among adults in stable remission following at least 16 weeks of ESK optimization. Improvements in mean MADRS score persisted with continuous ESK treatment in SUSTAIN-1.

The efficacy and safety of ESK + OAD for the treatment of TRD have been demonstrated in both short-term and long-term studies.^{16–18} Here, we report the findings of the impact of ESK + OAD versus OAD + PBO on TRD symptoms in the TRANSFORM-1 and TRANSFORM-2 short-term studies and the SUSTAIN-1 long-term study using clinician- (MADRS) and patient-reported (PHQ-9) assessment tools. We also evaluate the associations between clinician- and patient-reported outcomes using the MADRS and PHQ-9 among adult patients with TRD in TRANSFORM-1 and TRANSFORM-2.

Methods

Study design

Data from TRANSFORM-1, TRANSFORM-2, and SUSTAIN-1 were used in these analyses (Figure 1). TRANSFORM-1 and TRANSFORM-2 were two similarly designed, 4-week, randomized, controlled studies of ESK + OAD versus OAD + PBO in adults with TRD.^{16,17} SUSTAIN-1 was a longer term, phase 3, double-blind, active-controlled, randomized withdrawal, relapse prevention study evaluating ESK + OAD versus OAD + PBO in adults in stable remission following ESK optimization.¹⁸ Eligible patients were enrolled into SUSTAIN-1, either directly or after completing the double-blind phase of the TRANSFORM-1 or TRANSFORM-2 studies.¹⁸ The studies included in this analysis were registered at ClinicalTrials.gov (NCT02417064, NCT02418585, and NCT02493868) and approved by the local ethics committees, and written informed consent was obtained from all participating patients.

The MADRS

The MADRS is a 10-item clinician-rated instrument used to measure depression severity. The scale ranges from a score of 0 to 60, with higher scores indicating greater severity of depression.¹³ Each item is rated on a 7-point continuum (0 = no abnormality, 6 = severe depression).¹¹ In the TRANSFORM-1 and TRANSFORM-2 studies, the MADRS was administered at baseline and at days 2, 8, 15, 22, and 28. During SUSTAIN-1, the MADRS was administered at baseline; at days 8, 15, 22, and 28 during the induction phase; and weekly throughout the optimization and maintenance phases. In these analyses, a clinically substantial improvement in MADRS score was defined as a ≥ 12 -point change in the total score and a clinically meaningful improvement was defined as a ≥ 6 -point change in the total score.^{13,19} Response, defined as $\geq 50\%$ improvement in MADRS total score during double-blind induction phase, was also calculated.

The PHQ-9

The PHQ-9 is a self-rated assessment tool used to detect and evaluate the severity of depression. The score ranges from 0 to 27, with higher scores indicating greater severity of depression.²⁰ In the TRANSFORM-1 and TRANSFORM-2 studies, the PHQ-9 was administered at baseline, day 15, and day 28. During SUSTAIN-1, the PHQ-9 was administered at baseline, day 15, and day 28 during the induction phase and every other week throughout the optimization/maintenance phases. In this analysis, a clinically substantial improvement in PHQ-9 score was defined as a ≥ 6 -point change, with a clinically meaningful improvement defined as a ≥ 3 -point change.^{13,21} A PHQ-9 score ≤ 4 indicates “no depressive symptom status”²²; the patient was therefore considered to be in remission. Response, defined as $\geq 50\%$ improvement in PHQ-9 total score during double-blind induction phase, was also calculated.

Statistical analyses

Patients were pooled from the full analysis set of the TRANSFORM-1 and TRANSFORM-2 studies that includes all randomly assigned patients who received at least one dose of

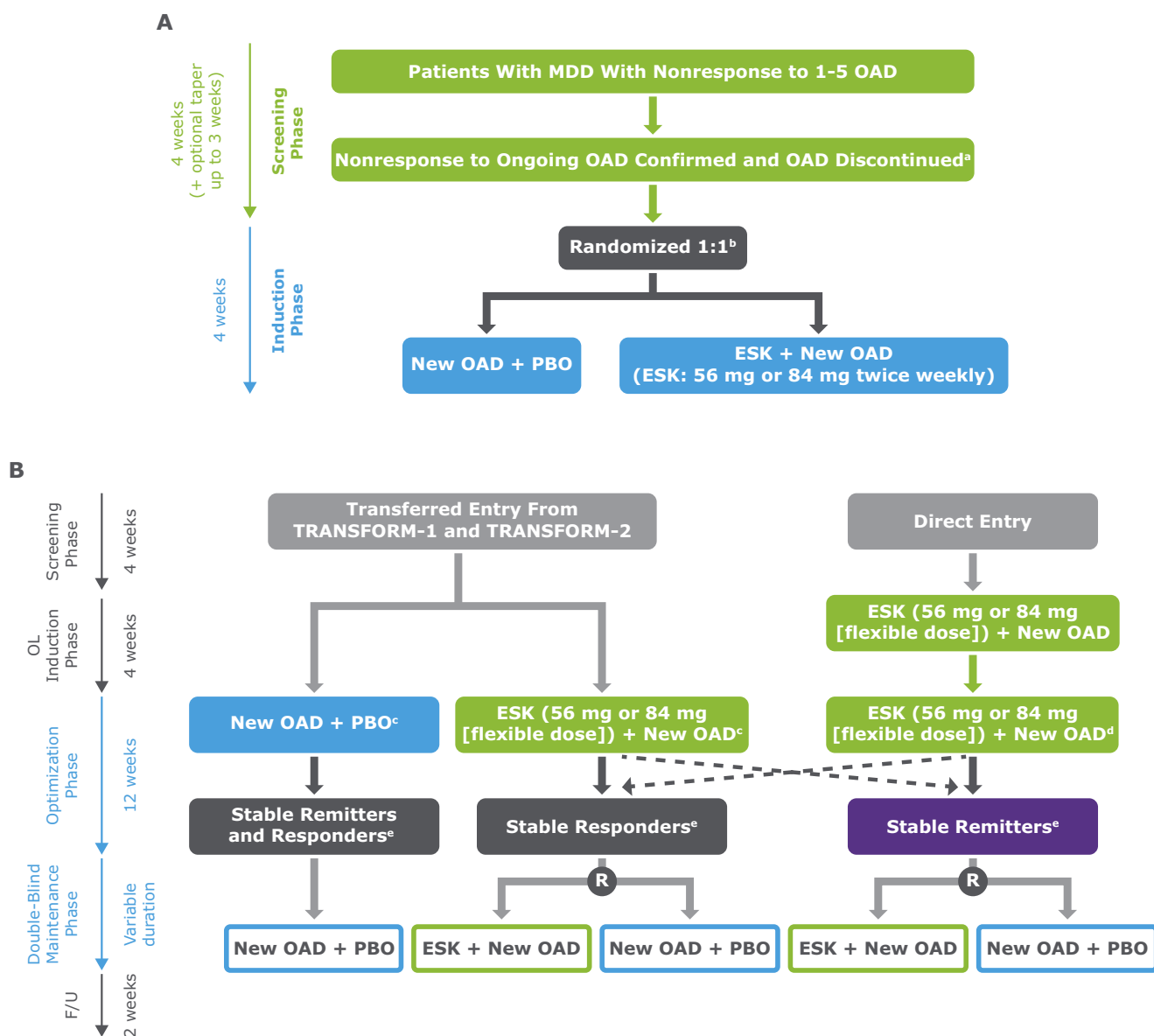


Figure 1. Study designs of (A) TRANSFORM-1¹⁶ and TRANSFORM-2,²¹ and (B) SUSTAIN-1.¹⁸ TRANSFORM-1 and TRANSFORM-2 had similar designs; however, flexible dosing was used in TRANSFORM-2 and fixed dosing was used in TRANSFORM-1.

^aNonresponse to ongoing OAD at end of screening was defined as a $\leq 25\%$ improvement in MADRS total score from week 1 to week 4 and MADRS total scores ≥ 28 at weeks 2 and 4.

^bThe randomization ratio was 2:1 (ESK + OAD:OAD + PBO) for the TRANSFORM-1 study and 1:1 for the TRANSFORM-2 study.

^cOnly responders proceeded to the optimization phase. Response in TRANSFORM-1 and TRANSFORM-2 was defined as $\geq 50\%$ reduction from baseline in MADRS total score at day 28.

^dAfter optimization, patients receiving ESK + OAD who achieved stable remission or stable response were randomized to either continue ESK + OAD or be switched to OAD + PBO until relapse or study completion.

^eStable response was defined as $\geq 50\%$ reduction from baseline in MADRS total score in each of the last 2 weeks of the optimization phase in patients who did not meet criteria for stable remission; stable remission was defined as MADRS total score ≤ 12 for at least 3 of the last 4 weeks of the optimization phase, with up to 1 excursion (MADRS score > 12) or 1 missing MADRS assessment permitted at week 13 or 14 only.

intranasal study medication and one dose of OAD medication. Changes in MADRS and PHQ-9 total scores from baseline at days 15 and 28 between treatment groups were compared using analysis of covariance models with fixed effects for treatment, study ID, region, and class of OAD (serotonin and norepinephrine reuptake inhibitor [SNRI] or selective serotonin reuptake inhibitor [SSRI]), and baseline values as a covariate. Data for least squares mean change from baseline are shown only for days 15 and 28 to align with the timing of PHQ-9 assessments. Differences in proportions of patients attaining clinically

substantial improvement, achieving response, and having no depressive symptom status were examined using the Cochran–Mantel–Haenszel test controlling for region and class of OAD (SNRI or SSRI). Chi-square tests were used to compare the proportion of patients with a MADRS score ≤ 12 and who retained remission status²³ on the PHQ-9. Parametric and nonparametric simple and multiple regression models were used to explore the relationship between MADRS and PHQ-9 scores from baseline to day 15, day 28, and study end point as dependent variables for TRANSFORM-1 and TRANSFORM-

2 studies. Assuming potential deviations from linearity for sensitivity analysis purposes, proportional odds models were used to examine the relationship between MADRS and PHQ-9 scores at TRANSFORM-1 and TRANSFORM-2 study end point.

Adult patients who participated in the SUSTAIN-1 study were also included as part of the analysis. Differences between treatment groups in the proportion of patients achieving remission status at the end of the maintenance phase of SUSTAIN-1 were examined using a chi-square test.

Results

Baseline demographics and disease characteristics

A total of 565 patients from TRANSFORM-1 and TRANSFORM-2 (ESK + OAD, $n = 343$; OAD + PBO, $n = 222$) and 176 patients from SUSTAIN-1 (ESK + OAD, $N = 90$; OAD + PBO, $N = 86$) were included in the analysis. Patient baseline demographic and clinical characteristics were similar across all studies and treatment groups (Table 1). The mean age was 45.4–46.6 years, and most patients

were female (64.4%–68.6%) and White (86.2%–88.9%) with a mean duration of current episode of 110.5–175.7 weeks. The mean baseline MADRS (37.4–37.6) and PHQ-9 (19.2–20.6) scores were indicative of severe depression for both treatment groups in both studies, respectively.

Clinician-rated depressive symptoms MADRS outcome findings in TRANSFORM-1 and TRANSFORM-2

Least squares mean MADRS total scores for the pooled TRANSFORM-1 and TRANSFORM-2 patients were significantly lower in the ESK + OAD group compared with the OAD + PBO group starting from day 2 and persisting through day 28 (Figure 2A). MADRS total scores improved from baseline to each post-baseline time point in both treatment arms. However, the least squares mean change from baseline in MADRS total score was significantly greater in the ESK + OAD group compared with the OAD + PBO group on days 15 (-13.2 vs -10.1 , $P = .002$) and 28 (-21.6 vs -17.2 , $P < .001$ [Figure 2B]). A significantly greater proportion of patients who received

Table 1. Baseline demographics and disease characteristics in the TRANSFORM-1, TRANSFORM-2, and SUSTAIN-1 studies

	TRANSFORM-1 and TRANSFORM-2		SUSTAIN-1 (induction phase)	
	ESK + OAD $n = 343$	OAD + PBO $n = 222$	ESK + OAD $n = 90$	OAD + PBO $n = 86$
Mean age (SD), years	45.7 (11.6)	46.6 (11.2)	45.4 (12.1)	46.2 (11.2)
Sex, n (%)				
Female	235 (68.5)	144 (64.9)	58 (64.4)	59 (68.6)
Male	108 (31.5)	78 (35.1)	32 (35.6)	27 (31.4)
Race, n (%)				
White	282 (86.2)	188 (88.3)	80 (88.9)	76 (88.4)
Black	20 (6.1)	10 (4.7)	4 (4.4)	6 (7.0)
Other ^a	25 (7.7)	15 (7.0)	6 (6.7)	4 (4.7)
Mean MADRS total score (SD)	37.4 (5.4)	37.4 (5.9)	37.4 (5.2)	37.6 (4.7)
Mean PHQ-9 total score (SD)	20.4 (3.8)	20.6 (3.7)	19.2 (4.2)	19.8 (3.4)
Mean duration of current episode (SD), weeks	175.7 (261.3)	156.2 (232.2)	112.2 (171.3)	110.5 (147.4)
No. of previous antidepressant medications, n (%) ^b				
1 or 2	216 (63.3)	139 (62.6)	71 (78.9)	62 (73.8)
≥ 3	125 (36.7)	83 (37.4)	19 (21.1)	22 (26.2)
Class of new oral antipsychotic, n (%)				
SNRI	209 (60.9)	139 (62.6)	62 (68.9)	58 (67.4)
SSRI	134 (39.1)	83 (37.4)	28 (31.1)	28 (32.6)
New oral antipsychotic, n (%)				
Duloxetine	152 (44.3)	105 (47.3)	47 (52.2)	38 (44.2)
Escitalopram	70 (20.4)	41 (18.5)	13 (14.4)	14 (16.3)
Sertraline	64 (18.7)	41 (18.5)	15 (16.7)	14 (16.3)
Venlafaxine extended release	57 (16.6)	35 (15.8)	15 (16.7)	20 (23.3)

Abbreviations: ESK, esketamine nasal spray; MADRS, Montgomery-Åsberg Depression Rating Scale; OAD, oral antidepressant; PBO, placebo nasal spray; PHQ-9, 9-item Patient Health Questionnaire; SNRI, serotonin and norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor.

Data for SUSTAIN-1 reflect characteristics at the start of the induction phase for patients with stable remittance. In SUSTAIN-1, stable remission was defined as MADRS total score ≤ 12 for at least 3 of the last 4 weeks of the optimization phase, with up to 1 excursion (MADRS score > 12) or one missing MADRS assessment permitted at week 13 or 14 only.

^aOther includes patients who identified as American Indian or Alaskan native, more than 1 race, and those with race not reported.

^bTRANSFORM-1 and TRANSFORM-2 ESK + OAD $n = 341$, OAD + PBO $n = 222$, and SUSTAIN-1 ESK + OAD $n = 90$ and OAD + PBO $n = 84$.

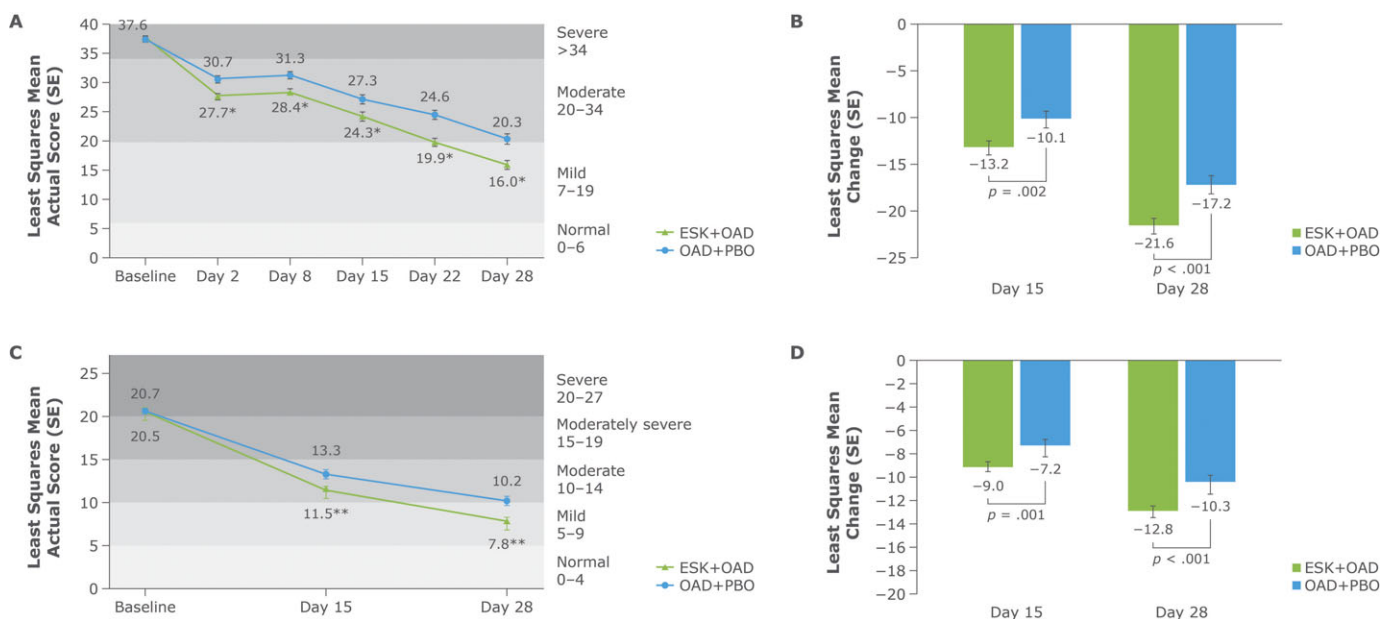


Figure 2. Total scores in TRANSFORM-1 and TRANSFORM-2. (A) MADRS least squares mean actual score (\pm SE) throughout the study. (B) MADRS least squares mean (\pm SE) change from baseline at days 15 and 28. (C) PHQ-9 least squares mean actual score (\pm SE) throughout the study. (D) PHQ-9 least squares mean (\pm SE) change from baseline at days 15 and 28. * $P \leq .01$. Between-group comparisons were based on an analysis of covariance model with fixed effects for treatment, study identification, region, and class of OAD, and baseline value as covariate. ** $P \leq .001$. Between-group comparisons were based on an analysis of covariance model with fixed effects for treatment, region, and class of OAD, and baseline value as covariate.

^aDepression severity based on PHQ-9 score was defined as normal (score of 0–4), mild (score of 5–9), moderate (score of 10–14), moderately severe (score of 15–19), or severe (score of 20–27).²⁰

ESK + OAD attained a clinically substantial improvement in MADRS (ie, ≥ 12 -point change from baseline) compared with patients who received OAD + PBO (46.5% vs 33.7%, $P = .002$, on day 15 and 69.0% vs 55.3%, $P < .001$, on day 28 [Figure 3A], respectively). A clinically meaningful improvement in MADRS (ie, ≥ 6 -point change from baseline) was attained by a significantly greater proportion of patients who received ESK + OAD compared with patients who received OAD + PBO on day 15 (69.3% vs 58.2%, $P = .008$) and day 28 (79.7% vs 68.8%, $P = .001$). A similar pattern was observed in the proportion of patients who attained response per MADRS total score on day 15 (26.3% vs 18.3%, $P = .023$) and day 28 (58.7% vs 45.2%, $P < .001$) (Figure 3C).

PHQ-9 patient-rated outcome findings in TRANSFORM-1 and TRANSFORM-2

The least squares mean PHQ-9 total scores were lower in the ESK + OAD versus OAD + PBO group at days 15 and 28 and were indicative of mild-to-moderate MDD (Figure 2C). Changes in PHQ-9 total score with ESK treatment shared directional changes with those observed for MADRS total score. The least squares mean change from baseline in PHQ-9 total score was significantly greater in the ESK + OAD group compared with the OAD + PBO group at days 15 and 28, -9.0 versus -7.2 , $P = .001$ and -12.8 versus -10.3 , $P < .001$, respectively (Figure 2D), representing a clinically meaningful improvement from baseline. A higher proportion of patients who received ESK + OAD attained a clinically substantial improvement in PHQ-9 total score (ie, ≥ 6 -point change from baseline) compared with patients who received OAD + PBO on both day 15 (64.4% vs 52.8%, respectively,

$P = .005$) and day 28 (77.1% vs 64.7%, respectively, $P < .001$) (Figure 3B). A clinically meaningful improvement in PHQ-9 (ie, ≥ 3 -point change from baseline) was attained by a significantly greater proportion of patients who received ESK + OAD compared with patients who received OAD + PBO on both day 15 (75.8% vs 68.5%, $P = .043$) and day 28 (85.8% vs 71.5%, $P < .001$). A similar pattern was observed in the proportion of patients who attained response per PHQ-9 total score on day 15 (43.8% vs 34.3%, $P = .027$) and day 28 (65.8% vs 52.9%, $P < .001$) (Figure 3D).

MADRS/PHQ-9 alignment results in TRANSFORM-1 and TRANSFORM-2

With a simple linear regression model, a 1-point increase in PHQ-9 total score at day 28 corresponded to a 1.5-point (standard error, 0.04) shift in the MADRS total score in TRANSFORM-1 and TRANSFORM-2 (Table 2). In this setting, the PHQ-9 total score can account for 73% of the MADRS total score variation. The alignment between the MADRS and the PHQ-9 total scores at day 28 is shown in Figure 4. The relationship between the MADRS and the PHQ-9 total scores tended to follow a nonlinear trend. This relationship was not constant over time, with the least alignment between MADRS and PHQ-9 total scores observed at baseline, with an increased correlation over time (Table 3). The relationship between MADRS and PHQ-9 scores at TRANSFORM-1 and TRANSFORM-2 study end point was examined, with predicted probabilities listed for widely used scoring categories for the PHQ-9 scores (Table 4). Different levels of severity in MADRS total score and PHQ-9 levels were not perfectly aligned.

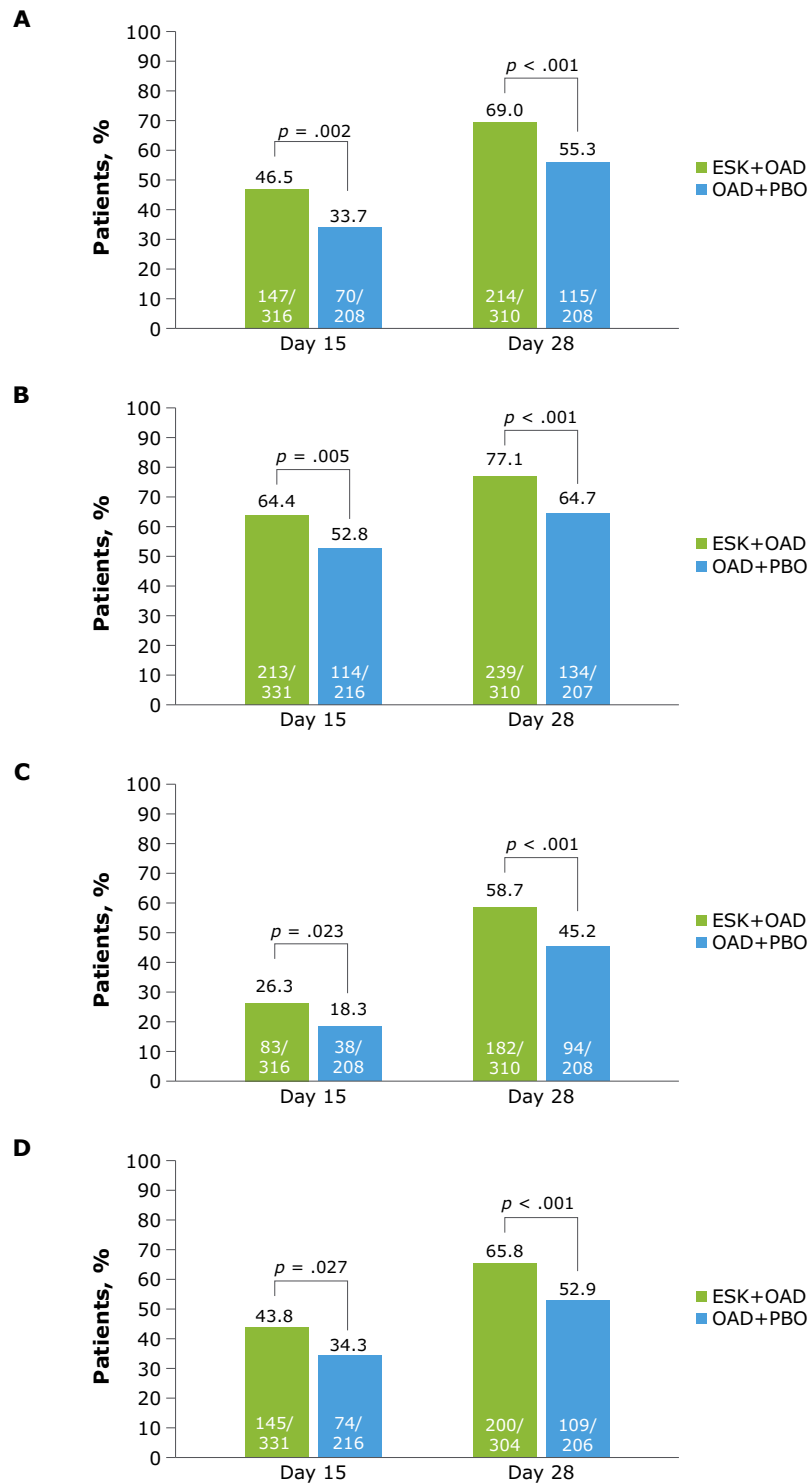


Figure 3. Proportions of patients with a clinically substantial improvement. (A) ≥ 12 -point improvement in MADRS total score. (B) ≥ 6 -point improvement from baseline in PHQ-9 total score in TRANSFORM-1 and TRANSFORM-2 and proportions of patients with response. (C) $\geq 50\%$ improvement in MADRS total score. (D) $\geq 50\%$ improvement in PHQ-9 total score in TRANSFORM-1 and TRANSFORM-2.

Table 2. MADRS total score over time associated with a 1-point shift in PHQ-9 total score during TRANSFORM-1/TRANSFORM-2

	Estimate (SE)	95% CI	r ²
Baseline	0.7 (0.06)	0.6–0.8	0.228
Day 15	1.2 (0.05)	1.1–1.3	0.556
Day 28	1.5 (0.04)	1.4–1.6	0.732
End point	1.5 (0.04)	1.4–1.6	0.724

Abbreviations: MADRS, Montgomery–Åsberg Depression Rating Scale; PHQ-9, 9-item Patient Health Questionnaire. PHQ-9 total score was set as a predictor.

Clinician-rated depressive symptoms MADRS outcome findings in SUSTAIN-1

In SUSTAIN-1, the median (range) duration of exposure in the maintenance phase was 17.7 (0–83) weeks in the ESK + OAD group and 10.2 (0–76) weeks in the OAD + PBO group. The mean MADRS total scores decreased substantially in patients treated with ESK + OAD during the optimization phase and this decrease in total score was sustained throughout the maintenance phase (Figure 5A). A significantly higher percentage of patients treated with ESK + OAD than OAD + PBO retained a MADRS score of ≤12 (mild to no depressive symptoms) at maintenance treatment end point (57.6% vs 35.2%, *P* < .001) (Figure 5B).

PHQ-9 total scores in SUSTAIN-1

Mean PHQ-9 total scores decreased substantially in patients treated with ESK + OAD during the optimization phase, with lower PHQ-9 scores observed in the ESK + OAD group than in the OAD + PBO group throughout the maintenance phase (Figure 5C). More patients with stable remission who entered the double-blind maintenance phase and received ESK + OAD

(*N* = 89) versus OAD + PBO (*N* = 86) retained remission status (PHQ-9 score ≤4) at maintenance treatment end point (57.3% vs 44.2%, *P* = .044) (Figure 5D). Significantly more patients in the OAD + PBO group versus than in the ESK + OAD group had a clinically substantial worsening in PHQ-9 total score (ie, an increase of ≥6 points) at week 2 (16.7% vs 3.5%, *P* = .004) and end point (38.4% vs 21.4%, *P* = .014) of the double-blind, randomized maintenance phase. Clinically meaningful worsening (ie, an increase of ≥3 points) was significantly greater in OAD + PBO patients compared with ESK + OAD patients at week 2 (33.3% vs 12.8%, *P* = .001) and numerically greater at the maintenance phase end point (53.5% vs 39.3%, *P* = .060).

Discussion

This post hoc analysis of three phase 3 studies of ESK + OAD versus OAD + PBO afforded an opportunity to highlight the associations between the MADRS and PHQ-9 and compare therapeutic interventions among patients who are severely ill with TRD. Regardless of the study length, ESK + OAD demonstrated greater efficacy than OAD + PBO for the treatment and management of TRD symptoms among adults. Furthermore, the short-term TRANSFORM-1 and TRANSFORM-2 studies demonstrated associations between clinician- (MADRS) and patient-reported (PHQ-9) outcomes, favoring treatment with ESK + OAD versus OAD + PBO, while the SUSTAIN-1 study provided an opportunity to demonstrate the long-term, sustained efficacy of ESK + OAD treatment compared to OAD + PBO. This analysis demonstrates the value of including patient-reported outcome assessments when evaluating the effectiveness of TRD treatment strategies.

Several studies of MDD and/or TRD have sought to assess the relationship between clinician-rated and patient-rated scales to characterize the trajectory of the course of disease and treatment effects.^{7,8,13,19,24} Establishing quantitative relationships and

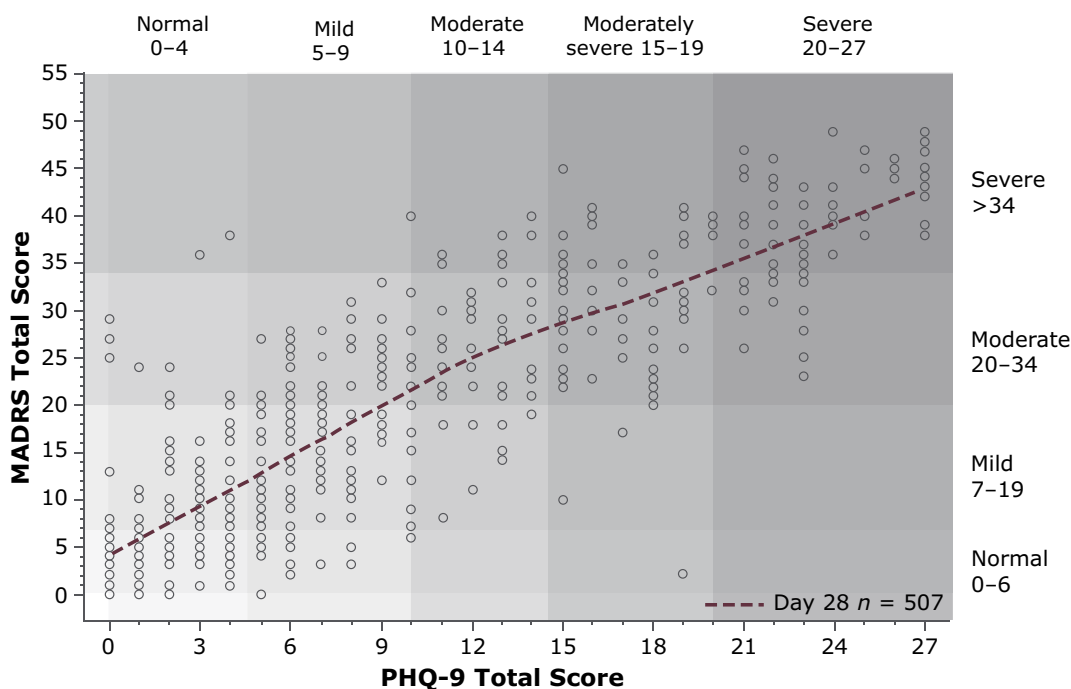


Figure 4. Relationship between MADRS total scores and PHQ-9 total scores at day 28. Dotted line denotes the locally weighted smoothing regression line (LOESS).

Table 3. Correlation coefficients of PHQ-9 and MADRS total score by visit during TRANSFORM-1 and TRANSFORM-2

	Correlation coefficient	P-value
Baseline		
Pearson	0.478	<.001
Spearman	0.473	<.001
Day 15		
Pearson	0.746	<.001
Spearman	0.732	<.001
Day 28		
Pearson	0.856	<.001
Spearman	0.849	<.001
End point		
Pearson	0.851	<.001
Spearman	0.851	<.001

Abbreviations: MADRS, Montgomery–Åsberg Depression Rating Scale; PHQ-9, 9-item Patient Health Questionnaire.

clinically meaningful score changes provide clinicians with tools to facilitate the translation of clinical trial outcomes into meaningful approaches in clinical practice. Efforts to estimate clinically meaningful changes in the Clinical Global Impression-Severity scale with changes in the MADRS, Sheehan Disability Scale, and PHQ-9 have laid the foundation for linking score changes between clinician- and patient-rated assessment strategies.¹⁹ One study conducted regression analyses to generate a predictive equation between depression scale pairs (PHQ-9 and MADRS, PHQ-9 and Beck Depression Inventory II, Zung Self-Rated Scale [SRS] and MADRS, and SRS and PHQ-9) to aid in translating scores between scales.²⁴ Results of the analyses showed conversion equations for depression scores were precise when applied to averages but were less useful at the idiographic level.²⁴ Others have equated the Hospital Anxiety and Depression Scale (HADS) depression (HADS-D) and anxiety (HADS-A) subscales to the PHQ-9 and Generalized Anxiety Disorder Scale, providing tables for converting raw scores between instruments.⁸ Recent emphasis has considered identification of meaningful assessment of complementary symptom changes assessed by the MADRS and PHQ-9.^{7,13} The current analysis builds

upon these studies, demonstrating the increased agreement between the MADRS and PHQ-9 outcomes throughout the pooled TRANSFORM-1 and TRANSFORM-2 patient populations.

Although the clinician-rated assessment of response is an important strategy for evaluating efficacy, the patient perspective provides a quantitative measure of treatment response and disease burden. Unlike other common long-term disorders, there are no laboratory or physical diagnostic tests to quantify how patients respond to TRD treatments. Often, clinician-rated and patient-rated assessments of symptom status differ,¹⁰ with patient-rated assessments conferring the real-world value of a therapeutic intervention and capturing the presence and burden of symptoms from the patient perspective. However, MADRS and PHQ-9 are similar in that the PHQ-9 measures the frequency of the same DSM-5 depressive symptoms that MADRS rates for severity. This analysis suggests that there is consistency between the MADRS and PHQ-9 in identifying symptom improvement. Further investigation is warranted to understand the role of perspective in the recognition and timing of minor symptom changes. The patient perspective may serve as a more timely and sensitive measure than the clinician perspective in recognizing early subtle symptom changes. The relationship established between the MADRS and PHQ-9 in this analysis facilitates the translation of a frequently used clinical trial measure of depressive symptoms with a patient-reported tool commonly used in clinical practice. Providing clinicians with greater confidence of the validity of the PHQ-9 in assessing MDD symptoms confers access to a reliable tool that is succinct and simple to administer in any healthcare setting.

Limitations

Data from TRANSFORM-1 and TRANSFORM-2 were pooled in this analysis; however, these studies were not identically designed in that TRANSFORM-1 involved fixed dosing and TRANSFORM-2 involved flexible dosing.^{16,17} Regardless of the dosing strategy, a significant improvement from baseline was observed among patients treated with ESK + OAD compared with those treated with OAD + PBO (TRANSFORM-1 and TRANSFORM-2). The studies included in this post hoc analysis were not specifically designed to investigate the associations between the MADRS and PHQ-9 assessments. However, the non-linear associations identified between the scales may contribute to more comprehensive evaluations in clinical decision-making. It is

Table 4. Distribution of MADRS total scores by PHQ-9 total score at TRANSFORM-1 and TRANSFORM-2 study end point

PHQ-9 total score, n (%)	Observed	Expected	MADRS total score, n (%)					
			Predicted probabilities					
			Normal	Mild	Moderate	Severe		
			0–6	7–12	13–19	20–34	>34	
Normal	0–4	162 (29.2)	162 (29.2)	92 (79.3)^a	39 (46.4)	20 (24.1)	9 (5.3)	2 (2.0)
Mild	5–9	155 (28.0)	152 (27.4)	22 (19.0)	39 (46.4)	50 (60.2)	42 (24.6)	2 (2.0)
Moderate	10–14	74 (13.4)	75 (13.5)	1 (0.9)	5 (6.0)	11 (13.3)	49 (28.7)	8 (8.0)
Moderately severe	15–19	77 (13.9)	79 (14.2)	1 (0.9)	1 (1.2)	1 (1.2)	50 (29.2)	24 (24.0)
Severe	20–27	86 (15.5)	87 (15.7)	0 (0)	0 (0)	1 (1.2)	21 (12.3)	64 (64.0)
Total	–	554	554	116 (20.9)	84 (15.2)	83 (15.0)	171 (30.9)	100 (18.1)

Abbreviations: MADRS, Montgomery–Åsberg Depression Rating Scale; PHQ-9, 9-item Patient Health Questionnaire.

Predicted probabilities are listed as percentages from proportional odds models. Probabilities greater than ≥20% are bolded and indicate the relationship between the PHQ-9 and MADRS total scores at end point.

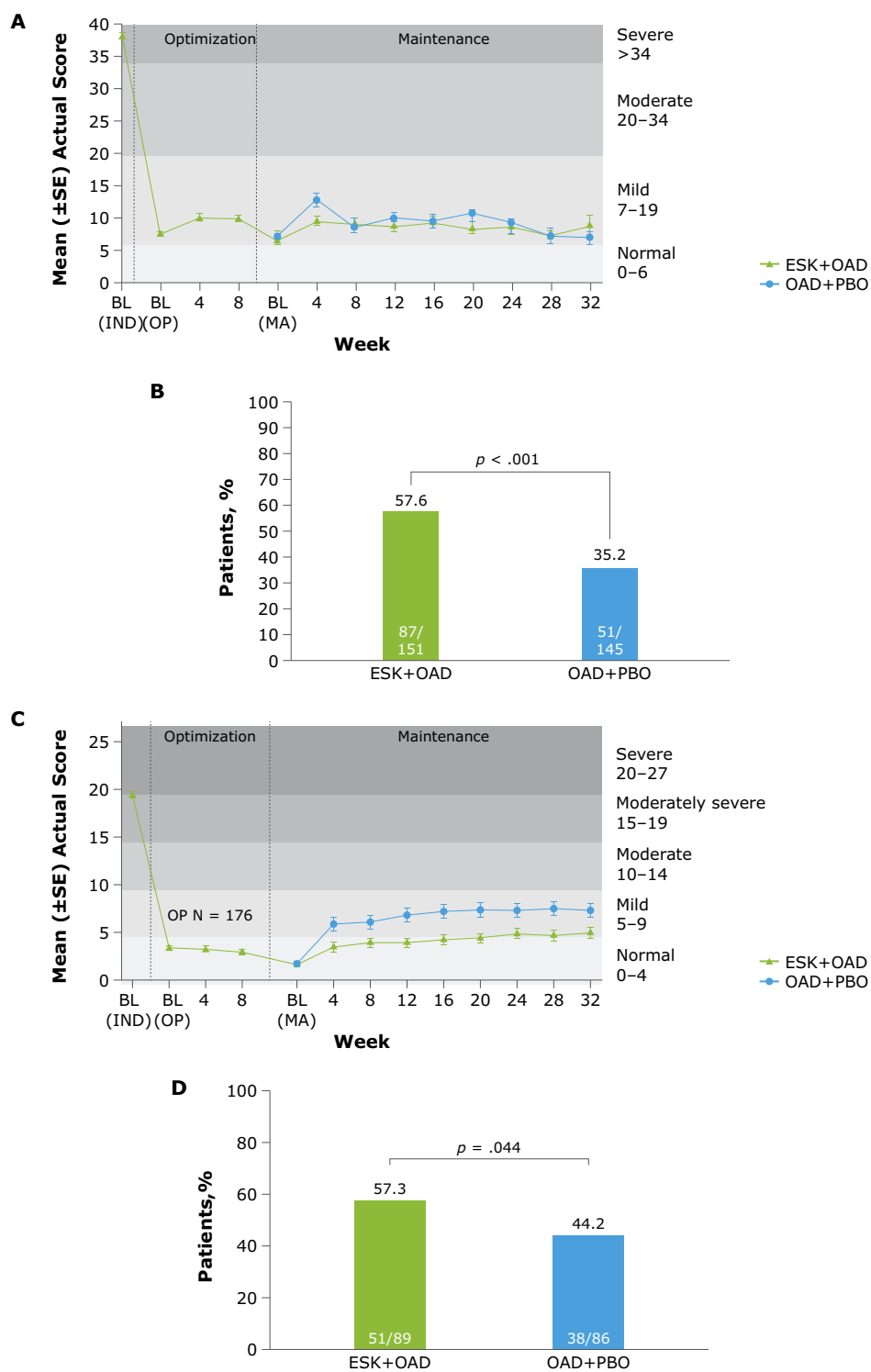


Figure 5. MADRS and PHQ-9 total scores in SUSTAIN-1. (A) MADRS total score (±SE) throughout the study. (B) Proportion of patients who retained a MADRS score of ≤12 at end point of the maintenance phase. (C) PHQ-9 mean (±SE) actual score throughout the study.^a (D) proportion of patients who retained remission status on the PHQ-9 (≤4) at the end point of the maintenance phase.^b

^aDepression severity based on PHQ-9 score was defined as normal (score of 0–4), mild (score of 5–9), moderate (score of 10–14), moderately severe (score of 15–19), or severe (score of 20–27).²⁰

^bPatients who retained remission status had a PHQ-9 score of ≤4 at end point.²⁷

also important to note that data from clinical trial participants may not be generalizable to real-world populations of patients with TRD. In addition, the PHQ-9 is a patient-reported outcome that is dependent upon patient recall over time. Furthermore, because

clinician- and patient-reported assessments were performed every 2 weeks (ie, baseline, day 15, and day 28), fluctuations in response to treatment may not have been captured. Approaches such as ecological momentary assessment, which involves collecting

patient data on repeated occasions in real time and in the context of daily life (eg, via a smartphone), is a potential tool for more frequent assessments from a patient perspective.²⁵ Although it is recognized that measurement-based care may enhance the quality of care and improve clinical outcomes, informed clinician judgment is also necessary to guide treatment decisions in real-world practice.

With respect to SUSTAIN-1 outcomes, it should be noted that the same PHQ-9 thresholds (≥ 6 -point change and ≥ 3 -point change) were used to measure both clinical improvement and clinical worsening in depression severity. Additional studies are needed to determine if these PHQ-9 thresholds are appropriate to assess both improvement and worsening of depression severity in patients with TRD. Correlations evaluated in this study grew stronger at post-baseline visits, indicating a potential learning curve that brings both patients and clinicians into increasingly analogous frames of reference over time. It also suggests progression in integrating the perspective of the clinician and patient when focus is directed to similar aspects of the disease. This phenomenon has also been observed in other disease states with different clinician- and patient-reported outcomes.²⁶

An additional limitation is the small number of patients historically underrepresented in clinical trials, for example, patients of color who participated in the studies. Responses from a more diverse patient population could have provided additional inter-racial insights on the use of the PHQ-9 and how the assessment relates to MADRS total scores across different races and ethnicities. Further research in this area is warranted.

Conclusions

This post hoc analysis of treatment with ESK + OAD versus OAD + PBO resulted in significant benefit among patients severely ill with TRD as measured by both clinician- and patient-rated evaluation of depression symptoms. The PHQ-9, a measure of depression severity that is relevant and easily administered in routine clinical practice, produced results that are consistent with those observed for the clinician-rated MADRS assessment in adults with TRD. Compared to the MADRS, more clinically meaningful improvements in depressive symptoms were observed as measured by the patient-rated PHQ-9 assessment with ESK + OAD versus OAD + PBO treatment.

Acknowledgments. The authors thank Audrey Shor, PhD (ApotheCom, Yardley, PA, USA), for editorial and writing assistance, which was funded by Janssen Scientific Affairs, LLC.

Author contribution. J.K.S. and M.T.: Conceptualization, writing, reviewing, and editing. R.R.N. Jr, M.P.P., and P.C.: Writing, reviewing, and editing. I.T.: Conceptualization, methodology, software, validation, formal analysis, data curation, writing, reviewing, editing, and visualization.

Financial support. This study was funded by Janssen Scientific Affairs, LLC, Titusville, NJ, USA. The study sponsor was involved in the design and conduct of the study and in the collection, management, analysis, and interpretation of data. As detailed in the authorship statement, all authors made significant contributions to the development of the manuscript in accordance with ICMJE authorship criteria. All authors provided direction and comments on the manuscript, reviewed and approved the final version prior to submission, made the final decision about where to publish these data, and approved submission to this journal. All authors had full access to the study data and take responsibility for data integrity and the accuracy of the analyses.

Competing interest. J.K.S., R.R.N. Jr, M.P.P., and P.C. are employees of Janssen Scientific Affairs, LLC, and hold stock in Johnson & Johnson, Inc. I.T. is an employee of Janssen Research & Development, LLC, and holds stock in Johnson & Johnson, Inc. M.T. has provided consulting services to Alkermes Inc., Axsome Therapeutics, Biogen MA Inc., Cerebral Inc., Circular Genomics Inc., Compass Pathfinder Limited, GH Research Limited, Heading Health Inc., Janssen, Legion Health Inc., Merck Sharp & Dohme Corp., Mind Medicine (MindMed) Inc., Naki Health, Ltd., Neurocrine Biosciences Inc., Noema Pharma AG, Orexo US Inc., Otsuka American Pharmaceutical Inc., Otsuka Canada Pharmaceutical Inc., Otsuka Pharmaceutical Development & Commercialization Inc., Praxis Precision Medicines Inc., SAGE Therapeutics, Sparian Biosciences Inc., Takeda Pharmaceutical Company Ltd., and WebMD; is a member of the scientific advisory board of Alto Neuroscience Inc., Cerebral Inc., COMPASS Pathfinder Ltd., Heading Health, GreenLight VitalSign6 Inc., Legion Health Inc., Merck Sharp & Dohme Corp., Orexo US Inc., and Signant Health; has received editorial compensation from the American Psychiatric Association and Oxford University Press; and holds stock in Alto Neuroscience Inc., Cerebral Inc., Circular Genomics Inc., GreenLight VitalSign6 Inc., and Legion Health Inc.

References

- Baldessarini RJ, Forte A, Selle V, et al. Morbidity in depressive disorders. *Psychother Psychosom.* 2017;**86**(2):65–72.
- World Health Organization. Depression. <https://www.who.int/news-room/fact-sheets/detail/depression>. Accessed June 12, 2020.
- Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry.* 2018;**75**(4):336–346.
- Rush AJ, Trivedi MH, Wisniewski SR et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry.* 2006;**163**(11):1905–1917.
- Gaynes BN, Asher G, Gartlehner G, et al. *Definition of Treatment-Resistant Depression in the Medicare Population.* Rockville, MD: Agency for Healthcare Research and Quality; 2018.
- Zimmerman M, Walsh E, Friedman M, Boerescu DA, Attiullah N. Are self-report scales as effective as clinician rating scales in measuring treatment response in routine clinical practice? *J Affect Disord.* 2018; **225**:449–452.
- Floden L, Hudgens S, Jamieson C, et al. Evaluation of individual items of the Patient Health Questionnaire (PHQ-9) and Montgomery–Åsberg Depression Rating Scale (MADRS) in adults with treatment-resistant depression treated with esketamine nasal spray combined with a new oral antidepressant. *CNS Drugs.* 2022;**36**(6):649–658.
- Huang XJ, Ma HY, Wang XM, Zhong J, Sheng DF, Xu MZ. Equating the PHQ-9 and GAD-7 to the HADS depression and anxiety subscales in patients with major depressive disorder. *J Affect Disord.* 2022;**311**: 327–335.
- Bushnell DM, McCarrier KP, Bush EN, et al. Symptoms of major depressive disorder scale: performance of a novel patient-reported symptom measure. *Value Health.* 2019;**22**(8):906–915.
- Pandina GJ, Revicki DA, Kleinman L, et al. Patient-rated troubling symptoms of depression instrument results correlate with traditional clinician- and patient-rated measures: a secondary analysis of a randomized, double-blind, placebo-controlled trial. *J Affect Disord.* 2009;**118**(1–3):139–146.
- Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979;**134**:382–389.
- Costantini L, Pasquarella C, Odone A, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord.* 2021;**279**:473–483.
- Hudgens S, Floden L, Blackowicz M, et al. Meaningful change in depression symptoms assessed with the Patient Health Questionnaire (PHQ-9) and Montgomery–Åsberg Depression Rating Scale (MADRS) among patients with treatment resistant depression in two, randomized, double-blind, active-controlled trials of esketamine nasal spray combined with a new oral antidepressant. *J Affect Disord.* 2021;**281**:767–775.
- Mitchell AJ. Clinical utility of screening for clinical depression and bipolar disorder. *Curr Opin Psychiatry.* 2012;**25**(1):24–31.

15. Patel JS, Oh Y, Rand KL, et al. Measurement invariance of the Patient Health Questionnaire-9 (PHQ-9) depression screener in U.S. adults across sex, race/ethnicity, and education level: NHANES 2005-2016. *Depress Anxiety*. 2019;**36**(9):813–823.
16. Fedgchin M, Trivedi M, Daly EJ, et al. Efficacy and safety of fixed-dose esketamine nasal spray combined with a new oral antidepressant in treatment-resistant depression: results of a randomized, double-blind, active-controlled study (TRANSFORM-1). *Int J Neuropsychopharmacol*. 2019;**22**(10):616–630.
17. Popova V, Daly EJ, Trivedi M, et al. Efficacy and safety of flexibly dosed esketamine nasal spray combined with a newly initiated oral antidepressant in treatment-resistant depression: a randomized double-blind active-controlled study. *Am J Psychiatry*. 2019;**176**(6):428–438.
18. Daly EJ, Trivedi MH, Janik A, et al. Efficacy of esketamine nasal spray plus oral antidepressant treatment for relapse prevention in patients with treatment-resistant depression: a randomized clinical trial. *JAMA Psychiatry*. 2019;**76**(9):893–903.
19. Turkoz I, Alphs L, Singh J, et al. Clinically meaningful changes on depressive symptom measures and patient-reported outcomes in patients with treatment-resistant depression. *Acta Psychiatr Scand*. 2021;**143**(3):253–263.
20. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;**16**(9):606–613.
21. Turkoz I, Daly E, Singh J, et al. Treatment response with esketamine nasal spray plus an oral antidepressant in patients with treatment-resistant depression without evidence of early response: a pooled post hoc analysis of the TRANSFORM studies. *J Clin Psychiatry*. 2021;**82**(4):20m13800.
22. Zimmerman M. Using the 9-item Patient Health Questionnaire to screen for and monitor depression. *JAMA*. 2019;**322**(21):2125–2126.
23. Daly EJ, Turkoz I, Salvatore G, et al. The effect of esketamine in patients with treatment-resistant depression with and without comorbid anxiety symptoms or disorder. *Depress Anxiety*. 2021;**38**(11):1120–1130.
24. Hawley CJ, Gale TM, Smith PS, et al. Equations for converting scores between depression scales (MÁDRS, SRS, PHQ-9 and BDI-II): good statistical, but weak idiographic, validity. *Hum Psychopharmacol*. 2013;**28**(6):544–551.
25. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. *Psychiatr Serv*. 2017;**68**(11):1172–1181.
26. Askanase A, Tang W, Zuraw Q, Gordon R, Brotherton B, Merrill J. Evaluation of the LFA-REAL clinician-reported outcome (ClinRO) and patient-reported outcome (PRO): prespecified analysis of the phase III ustekinumab trial in patients with SLE. *Lupus Sci Med*. 2023;**10**(1):e000875.
27. American College of Physicians. Depression remission at twelve months. American College of Physicians. <https://www.acponline.org/clinical-information/performance-measures/depression-remission-at-twelve-months>. Accessed February 7, 2023.