RESEARCH

Assessing the meaningful change threshold of Quality of Life in Depression Scale using data from two phase 3 studies of esketamine nasal spray

Heather Rozjabek[1*], Nan Li[2], Holger Hartmann[3], Dong Jing Fu[4], Carla Canuso[4] and Carol Jamieson[5]

Abstract

Background: Major depressive disorder (MDD) directly impacts patients' lives including symptoms, functioning and health-related quality-of-life (HRQoL). Patient-reported outcomes can capture these impacts, however interpretation of clinical meaningfulness of these measurements are often not readily available. Meaningful change thresholds (MCTs) can be derived for clinical outcome assessments to quantify the change in symptoms that is meaningful to the patient following pharmacologic treatment or other intervention**s**. The objective of this analysis was to determine the within-patient MCT of the self-reported Quality-of-Life in Depression Scale (QLDS) among patients with MDD and active suicidal ideation with intent (MDSI) using an anchor-based approach.

Methods: Data from 2 randomized phase-3 trials of esketamine nasal spray (ASPIRE I and ASPIRE II) were analyzed. The Montgomery–Åsberg Depression Rating Scale (MADRS) was the primary anchor with three different severity criteria. Other anchor variables utilized were Clinical Global Impression of Severity of Suicidality-revised version, Clinical Global Impression of Imminent Suicide Risk, and EuroQol Visual Analog Scale [EQ-VAS]. Spearman correlation coefficients between the change in QLDS and anchor variables were calculated. The mean change in QLDS score at Day 25 from baseline was calculated based on the categorical change in the anchor. Coefficient yield from linear regression of the mean changes in EQ-VAS and QLDS, and distribution-based approach with ½ SD of change in QLDS were considered.

Results: In ASPIRE I, mean (SD) improvement in QLDS score among patients with one category improvement in MADRS from baseline to Day 25 was − 8.22 (8.87), − 8.30 (9.01), and − 8.20 (8.92) using severity criteria #1, #2, and #3, respectively. Patients who achieved a 7-point improvement (MCT) in EQ-VAS yielded a mean − 9.69-point improvement in QLDS at Day 25. The ½ SD of change in QLDS was 5.63. Similar results were obtained for ASPIRE II. The MCTs identified using multiple anchors across both trials ranged from − 11.4 to − 6.7 and had an overall mean of − 7.90 (ASPIRE I) and − 7.92 (ASPIRE II). Thus, an 8-point change was recommended as the MCT for QLDS.

Conclusion: The recommended MCT will help quantify within-person changes in HRQoL using patient-reported QLDS and determine meaningful treatment benefit in an MDD patient population with acute suicidal ideation or behavior.

*Correspondence: hrozjabe@its.jnj.com

[1] Janssen Global Services, LLC, Raritan, NJ, USA
Full list of author information is available at the end of the article

Introduction

Major depressive disorder (MDD) is a common mental health condition affecting over 163 million people worldwide [1] and over 21 million people in the United States [2]. The major symptoms of MDD include low mood, lack of energy, insomnia, sadness, and inability to enjoy life [3]. MDD also affects normal functioning as it reduces health-related quality-of-life (HRQoL) [4] and productivity at work [5]. MDD is also associated with suicide ideation [2]; and the prevalence of suicide attempts in patients with a lifetime MDD episode in the United States is reported to be 13.6% [6].

Despite availability of multiple treatments, first-line agents fail to adequately treat MDD in many patients [7], which raises the question whether existing methods of evaluating depression and its treatment in clinical trials are sufficiently comprehensive. Gathering patients' perspectives provide valuable insights on potential depression symptoms, day to day problems, health-related quality of life and treatment satisfaction and is a promising approach to address currently unmet needs [8]. Optimizing patient-reported outcomes (PRO) in clinical trials and improving their interpretation is therefore gaining momentum [9]. Kamenov et al. [10] analyzed 247 interventional studies for depression and reported that 80% of the areas covered by the employed PROs and clinician-rated tools represented clinical symptomatology, while other areas of functioning such as daily routine, work activities, and social participation were insufficiently or hardly covered. Since patients seem to prioritize functional outcomes over symptomatic outcomes [11], there is increasing emphasis on incorporating functional outcomes into clinical trials [12].

Patient-reported outcome measures can effectively capture the impact of a disease on the patient's daily life, including symptoms, functioning, and overall HRQoL [13–15]. The Quality of Life in Depression Scale (QLDS) is a disease-specific patient-reported questionnaire designed to assess the impact of depression on the HRQoL of patients [16]. The QLDS has been shown to have high reliability and construct validity [17], has been translated into multiple languages [18], and has been used in clinical trials and observational studies for the past 2 decades [19–27]. However, it is important to understand whether the results from a PRO are meaningful from the patients' perspective. Critical Path Institute's PRO Consortium and the Consensus Panel for Outcomes Measurement and Psychometrics: Advancing the Scientific Standards (COMPASS) recommends the use of the term meaningful change threshold (MCT), which describes the threshold of change at which the change becomes meaningful to the patient [28, 29]. According to these guidelines, the preferred method for deriving MCTs in a clinical trial and regulatory setting is to use anchor-based approaches. The approach is used to estimate mean change in a PRO measure in patients who reported a magnitude of change consistent with a clinically relevant change in the global impression of disease severity or health measure, with supportive cumulative distribution and probability density function curves [14, 15, 30, 31].

Establishing MCTs for clinical outcome assessments are essential for interpreting treatment effects; therefore, the objective of this analysis was to determine the within-person MCT for QLDS among patients with MDD who have active suicidal ideation with intent (MDSI), using data from two phase 3 studies of esketamine nasal spray (ESK).

Methods

Study selection

Data for this secondary analysis were obtained from two identically designed, randomized, double-blind, placebo-controlled, multicenter, phase 3 studies (ASPIRE I [NCT03039192] and ASPIRE II [NCT03097133]) that evaluated efficacy and safety of ESK+standard of care (SOC) versus placebo+SOC. Methods and primary data of these studies have been reported earlier [32, 33].

In brief, these two studies included men and women (aged 18–64 years) who were diagnosed with MDD (as per Diagnostic and Statistical Manual of Mental Disorders, 5th Edition) [34] without psychotic features as confirmed by the Mini International Neuropsychiatric Interview [35]; had current suicidal ideation with intent within 24 h prior to randomization, as confirmed by responding 'Yes' to the questions "Think about suicide?" and "Intend to act on thoughts of killing yourself?"; were in need of acute psychiatric hospitalization due to imminent risk of suicide; and had a Montgomery–Åsberg

Rozjabek *et al. Journal of Patient-Reported Outcomes*   (2022) 6:74

Page 3 of 11

Depression Rating Scale (MADRS) [36] total score > 28 pre-dose on day 1.

The patients went through an initial screening phase conducted within 48 h prior to Day 1 dose, followed by a 4-week double-blind treatment phase wherein patients were randomized (1:1) to receive ESK (84 mg) + SOC or placebo + SOC twice weekly followed by a 9-week follow-up phase. In both ASPIRE I and II, significant improvement in MADRS total score from Baseline at 24 h (primary endpoint) was observed in patients receiving ESK + SOC versus placebo + SOC (least-squares mean difference [SE]: ASPIRE I, − 3.8 [1.39], 95% CI: − 6.56, − 1.09; $p = 0.006$; ASPIRE II, − 3.9 [1.39], 95% CI: − 6.60, − 1.11; $p = 0.006$) [37, 38].

Study measures
QLDS
The QLDS is a disease-specific patient-reported outcome designed to assess HRQoL in patients with MDD [16, 17]. The instrument has a recall period of "at the moment," contains 34-items with "true"/ "not true" response options and takes approximately 5–10 min to complete. The score range is from 0 (good quality of life) to 34 (very poor quality of life) [17]. It has been shown to have acceptable psychometric properties and sensitivity to change [17, 18]. The QLDS was self-administered in the participant's native language and in electronic format using a computer tablet in both ASPIRE trials.

MADRS
The MADRS is a 10-item clinician-rated scale designed to measure depression severity [36, 39], with each item scored from 0 (item not present or normal) to 6 (severe or continuous presence of symptoms) for a total possible score range of 0 to 60. The MADRS evaluates apparent sadness, reported sadness, inner tension, sleep, appetite, concentration, lassitude, interest level, pessimistic thoughts, and suicidal thoughts. Since there is no consensus on how to categorize MADRS for different severity levels of depression, three sets of cut-offs were adopted in this analysis (Table 1) [40–42].

CGI-SS-r
The CGI-SS-r is included in Module 7 of the Suicide Ideation and Behavior Assessment Tool (SIBAT) [43]. The SIBAT is a computerized assessment tool with five patient-reported and three clinician-reported modules that has been designed to systematically and comprehensively capture suicidal ideation and behavior (SIB), and measure rapid changes in SIB. After collection of relevant background information using Modules 1–5, clinicians conduct a semi-structured interview (Module 6) and finalize ratings for four outcome measures in Module 7, followed by documentation in Module 8 [43]. The CGI-SS-r in Module 7 summarizes the clinician's overall impression of severity of suicidality on a 7-point scale (0-normal, 1-questionably, 2-mildly, 3-moderately, 4-markedly, 5-severely, and 6-most extremely suicidal) based on the totality of information available to the clinician, including information from the completed modules of the SIBAT. The category ratings in the CGI-SS-r are directly interpretable as different levels of suicidality and a 1-point change in a CGI scale is consistent with a clinically observable change [44, 45].

CGI-SR-I
The CGI-SR-I (included in Module 7 of the SIBAT) summarizes the clinician's best assessment of the likelihood that a patient will attempt suicide in the next 7 days on a 7-point scale (0-no imminent suicide risk to 6-extreme imminent suicide risk) [43].

EQ-VAS
The EQ-VAS is a part of the European Quality of Life Group-5 Dimension 5-Level questionnaire, a measure for HRQoL designed for self-completion by the respondent [46]. Patients self-rated their overall health status from 0 (worst health) to 100 (best health). Changes in EQ-VAS on the order of 7 to 10 were recognized as a threshold for meaningful change for an individual patient [47].

Assessment of MCT
Various methods, including anchor-based and distribution-based approaches, can be used to determine a meaningful change threshold. The anchor-based

Table 1 MADRS severity criteria

Criteria #1 and cut-off values [42]	Criteria #2 and cut-off values [41]	Criteria #3 and cut-off values [40]
No depression (0–12)	No depression (0–6)	No depression (0–12)
Slight depression (13–21)	Slight depression (7–19)	Slight depression (13–17)
Moderate depression (22–28)	Moderate depression (20–34)	Moderate depression (18–34)
Severe depression (> 28–60)	Severe depression (> 34–60)	Severe depression (> 34–60)

MADRS Montgomery–Åsberg Depression Rating Scale

Rozjabek *et al. Journal of Patient-Reported Outcomes*   (2022) 6:74

Page 4 of 11

approach compares the change in the measure under evaluation to another measure of change, considered an anchor or external criterion, while the distribution-based approach compares the change in the measure under evaluation to a measure of variability (e.g. standard error of measurement or standard deviation) [48]. While both methods can be used to measure meaningful change, anchor-based methods are preferred to derive an MCT and distribution-based thresholds can be supportive [30]. The choice of an MCT is often through a multifaceted triangulation of analytic approaches which may include feedback from experts, including clinicians and patients, which cannot be prescribed in an analysis plan [49].

Full efficacy analysis datasets were utilized for the MCT analysis, which had 224 patients for ASPIRE I and 227 patients for the ASPIRE II study. Treatment groups for individual studies were combined for the current analysis. The MCT was assessed using primarily an anchor-based approach with multiple anchors. The MADRS was used as the primary anchor. Three variable sets of score criteria to determine severity levels for depression on the MADRS are reported in the literature; therefore, all 3 scoring criteria were evaluated in this analysis to evaluate consistency. The three severity criteria are as follows: severity criteria #1 [42]– no depression (score 0–12), slight depression (score 13–21), moderate depression (score 22–28), and severe depression (score > 28–60); severity criteria #2 [41]– no depression (score 0–6), mild depression (score 7–19), moderate depression (score 20–34), and severe depression (score > 34–60); and severity criteria #3 [40]– no depression (score 0–12), mild depression (score 13–17), moderate depression (score 18–34), and severe depression (score > 34–60) (Table 1). Other potential anchor variables were the Clinical Global Impression of Severity of Suicidality-Revised (CGI-SS-r) [43], the Clinical Global Impression of Imminent Suicide Risk (CGI-SR-I) [43], and EuroQol Visual Analog Scale (EQ-VAS) [46] were also explored. Distribution based MCT was calculated using 1/2 standard deviation (SD) of change in QLDS score.

Statistical analysis

Descriptive statistics of the mean QLDS total scores and mean change from Baseline to Day 25 (end point of the double-blind treatment phase) are presented. Spearman correlation coefficients between the change in QLDS score and the change in anchor variables from Baseline to Day 25 were calculated to confirm the relevance of the anchor. A correlation coefficient of ≥ 0.40 was needed for the anchor to be used in this analysis [50]. Change of QLDS score between Baseline and Day 25 was calculated in patients within change groups defined by the anchor variable. Mean (SD) within-patient change from Baseline,

along with the median, 95% confidence intervals, standardized response mean, and standardized effect size were calculated for each of the categories of change in the anchor variable from Baseline. Coefficients yielded from linear regression of EQ-VAS and QLDS scores were used to provide additional data to evaluate the MCT of QLDS. Cumulative distribution function (CDF) curves were also generated of the change in QLDS score from Baseline to Day 25 stratified by MADRS change category at Day 25. The CDF curves provide a graphic representation of the full distribution of the QLDS change scores between MADRS change anchor categories, which provide additional support for the selection of the anchor category where the change in QLDS is observed.

Results

ASPIRE I results

The mean (SD) age of the patients at Baseline was 39.3 (12.9) years and most of the patients were women (61.6%). Majority of the patients (88.8%) were considered by clinicians to be moderately to extremely suicidal while 84.8% of the patients were considered to have moderate to extreme imminent risk for suicide within the next seven days (Table 2). The mean (SD) MADRS score at Baseline was 41.1 (6.07), which improved to 17.9 (12.09) at Day 25 for a mean change of -23.2 (13.03). The mean (SD) QLDS score decreased from 27.24 (6.40) at Baseline to 14.79 (11.33) at Day 25; mean (SD) change of − 12.46 (11.26).

The correlation between change in QLDS score and change in MADRS total score at Day 25 was 0.66 ($p < 0.0001$). The mean (SD) improvement in QLDS score from Baseline to Day 25 among patients who improved one, two, and three MADRS severity categories (as previously defined) was slightly different using alternative MADRS criteria: − 8.22 (8.87), − 14.70 (10.85), and − 18.61 (10.15), respectively, using severity criteria #1; − 8.30 (9.01), − 16.76 (10.28), and − 23.34 (8.42), respectively, using severity criteria #2; and − 8.20 (8.92), − 15.12 (10.80), and − 20.38 (9.41), respectively, using severity criteria #3 (Fig. 1; Table S1 in Additional file 1).

The correlation between change in QLDS score and CGI-SS-r score at Day 25 was 0.43 ($p < 0.0001$). The mean (SD) improvement in QLDS score from Baseline to Day 25 among patients who improved one, two, and three points in CGI-SS-r was − 8.20 (9.13), − 9.74 (10.54), and − 12.33 (11.75), respectively. The correlation between change in QLDS score and CGI-SR-I score at Day 25 was 0.44 ($p < 0.0001$). The mean (SD) improvement in QLDS score from Baseline to Day 25 among patients who improved one, two, and three

Rozjabek *et al. Journal of Patient-Reported Outcomes*   (2022) 6:74

Page 5 of 11

Table 2 Summary of demographics and anchor variables at Baseline

| Parameter	ASPIRE I (n = 224)	ASPIRE II (n = 227)
Age, years		
Mean (SD)	39.3 (12.91)	40.8 (13.07)
Range	18–64	18–64
Sex, n (%)		
Women	138 (61.6)	136 (59.9)
Men	86 (38.4)	91 (40.1)
QLDS		
Mean (SD)	27.24 (6.40)	26.77 (5.64)
Range	0–34	5–34
MADRS score		
Mean (SD)	41.1 (6.07)	39.7 (5.48)
Range	29–58	29–54
CGI-SS-r, n (%)		
Normal, not at all suicidal	0	0
Questionably suicidal	8 (3.6)	4 (1.8)
Mildly suicidal	17 (7.6)	16 (7.0)
Moderately suicidal	57 (25.6)	68 (30.0)
Markedly suicidal	80 (35.9)	90 (39.6)
Severely suicidal	56 (25.1)	45 (19.8)
Among the most extremely suicidal patients	5 (2.2)	4 (1.8)
CGI-SR-I, n (%)		
No imminent suicide risk	5 (2.2)	3 (1.3)
Minimal imminent suicide risk	12 (5.4)	8 (3.5)
Mild imminent suicide risk	17 (7.6)	18 (7.9)
Moderate imminent suicide risk	57 (25.6)	63 (27.8)
Marked imminent suicide risk	75 (33.6)	80 (35.2)
Severely imminent suicide risk	53 (23.8)	49 (21.6)
Extreme imminent suicide risk	4 (1.8)	6 (2.6)
EQ-VAS		
Mean (SD)	42.0 (24.44)	39.5 (22.62)
Range	0–100	0–95

CGI-SR-I Clinical Global Impression of Imminent Suicide Risk, *CGI-SS-r* Clinical Global Impression of Severity of Suicidality-Revised, *EQ-VAS* EuroQol Visual Analog Scale, *MADRS* Montgomery–Åsberg Depression Rating Scale, *QLDS* Quality of Life Depression Scale, *SD* Standard Deviation

points in CGI-SR-I was -6.06 (8.33), -10.84 (9.81), and -12.55 (11.20), respectively.

The mean (SD) Baseline EQ-VAS score was 42.0 (24.44), which improved to 62.7 (21.76) at Day 25; mean change of 20.7 (25.15). The correlation between change in QLDS score and change in EQ-VAS score (0–100) between Baseline and Day 25 was -0.53 ($p < 0.0001$). Based on the linear regression of QLDS on EQ-VAS ($\beta_0 = -8.00$, $p < 0.0001$; $\beta_1 = -0.24$, $p < 0.0001$), a 7-point improvement in EQ-VAS score—which is the MCT for EQ-VAS—led to a -9.69-point improvement in QLDS score. The MCT for QLDS using the distribution-based approach was 5.63 with 1/2 SD of change in QLDS score between Baseline and Day 25.

ASPIRE II results

The mean (SD) age of the patients at Baseline was 40.8 (13.1) years and most of the patients were women (59.9%). Majority of the patients (91.2%) were considered by clinicians to be moderately to extremely suicidal while 87.2% of the patients were considered to have moderate to extreme imminent risk for suicide within the next seven days (Table 2). The mean (SD) MADRS score at Baseline was 39.7 (5.48), which improved to 17.3 (11.73) at Day 25 for a mean change of -22.33 (12.65). The mean (SD) QLDS score decreased from 26.77 (5.64) at Baseline to 14.07 (11.02) at Day 25; mean (SD) change of -12.66 (11.20).
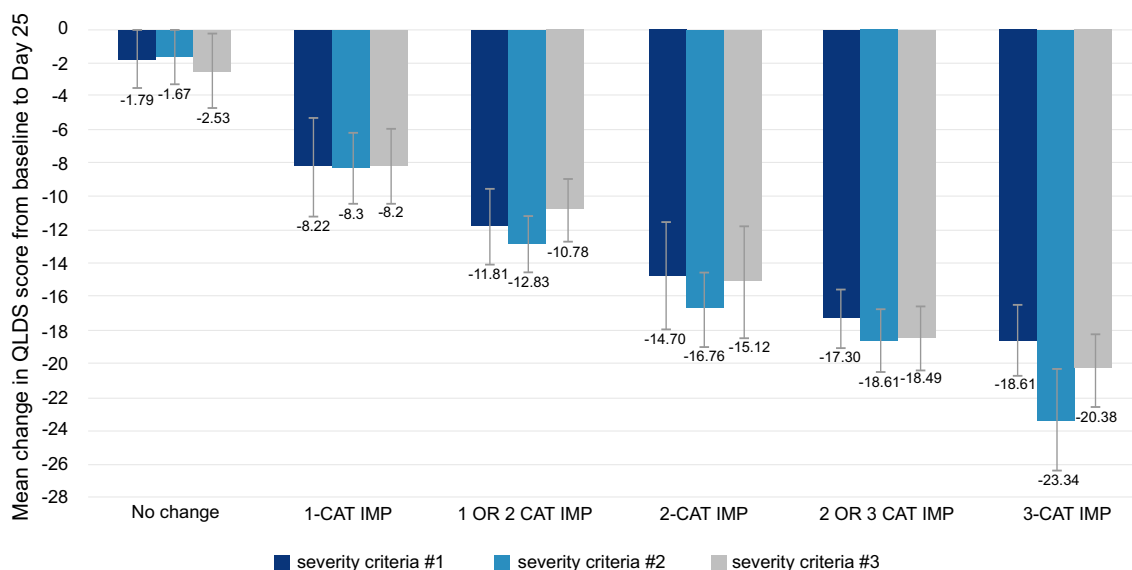
Fig. 1 ASPIRE I–mean change in QLDS score from baseline to Day 25 using MADRS as an anchor. Values are represented as mean ± 95% CI. MADRS was categorized as: severity criteria #1—no depression (score 0–12), slight depression (score 13–21), moderate depression (score 22–28), and severe depression (score > 28–60); severity criteria #2—no depression (score 0–6), mild depression (score 7–19), moderate depression (score 20–34), and severe depression (score > 34–60); and severity criteria #3—no depression (score 0–12), mild depression (score 13–17), moderate depression (score 18–34), and severe depression (score > 34–60). *CAT* Category, *CI* Confidence Interval, *IMP* Improvement, *MADRS* Montgomery–Åsberg Depression Rating Scale, *QLDS* Quality of Life in Depression Scale

The correlation between the change in QLDS score and change in MADRS category at Day 25 was 0.70 ($p < 0.0001$). The mean (SD) improvement in QLDS score from Baseline to Day 25 among patients who improved one, two, and three categories was slightly different using alternative MADRS criteria: -6.71 (8.33), -12.04 (9.26), and -20.31 (8.92), respectively, using severity criteria #1; -7.97 (8.10), -16.43 (10.31), and -23.58 (6.32), respectively, using severity criteria #2; and -7.94 (8.21), -14.28 (10.24), and -22.15 (7.67), respectively, using severity criteria #3 (Fig. 2; Table S2 in Additional file 1).

The correlation between change in QLDS score and CGI-SS-r at Day 25 was 0.43 ($p < 0.0001$). The mean (SD) improvement in QLDS score from Baseline to Day 25 among patients who improved one, two and three categories in CGI-SS-r was -4.81 (7.27), -11.28 (10.13), and -14.65 (11.14), respectively. The correlation between change in QLDS score and CGI-SR-I score at Day 25 was 0.41 ($p < 0.0001$). The mean (SD) improvement in QLDS score from Baseline to Day 25 among patients who improved one, two, and three categories in CGI-SR-I was -10.27 (11.18), -9.08 (10.23), and -13.57 (10.26), respectively.

The mean (SD) Baseline EQ-VAS score was 39.5 (22.62), which improved to 62.12 (23.18) at Day 25; mean change of 22.7 (26.73). The correlation between change
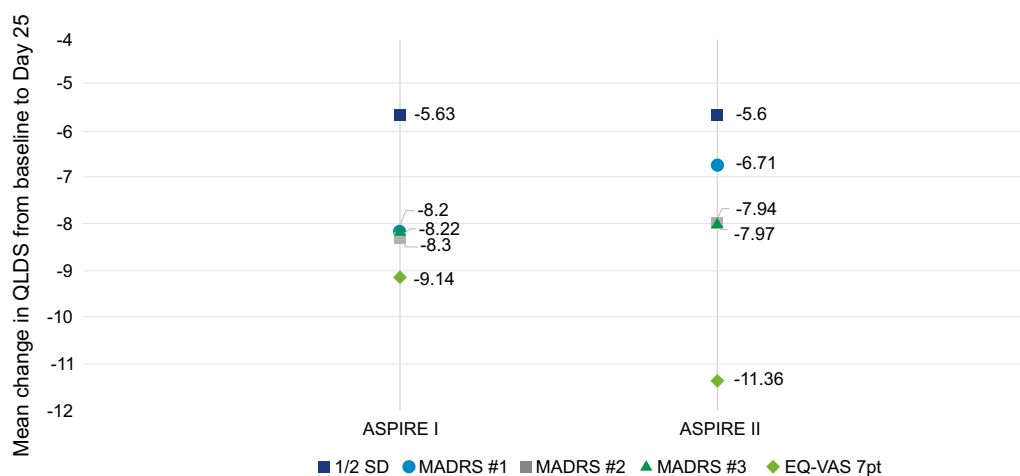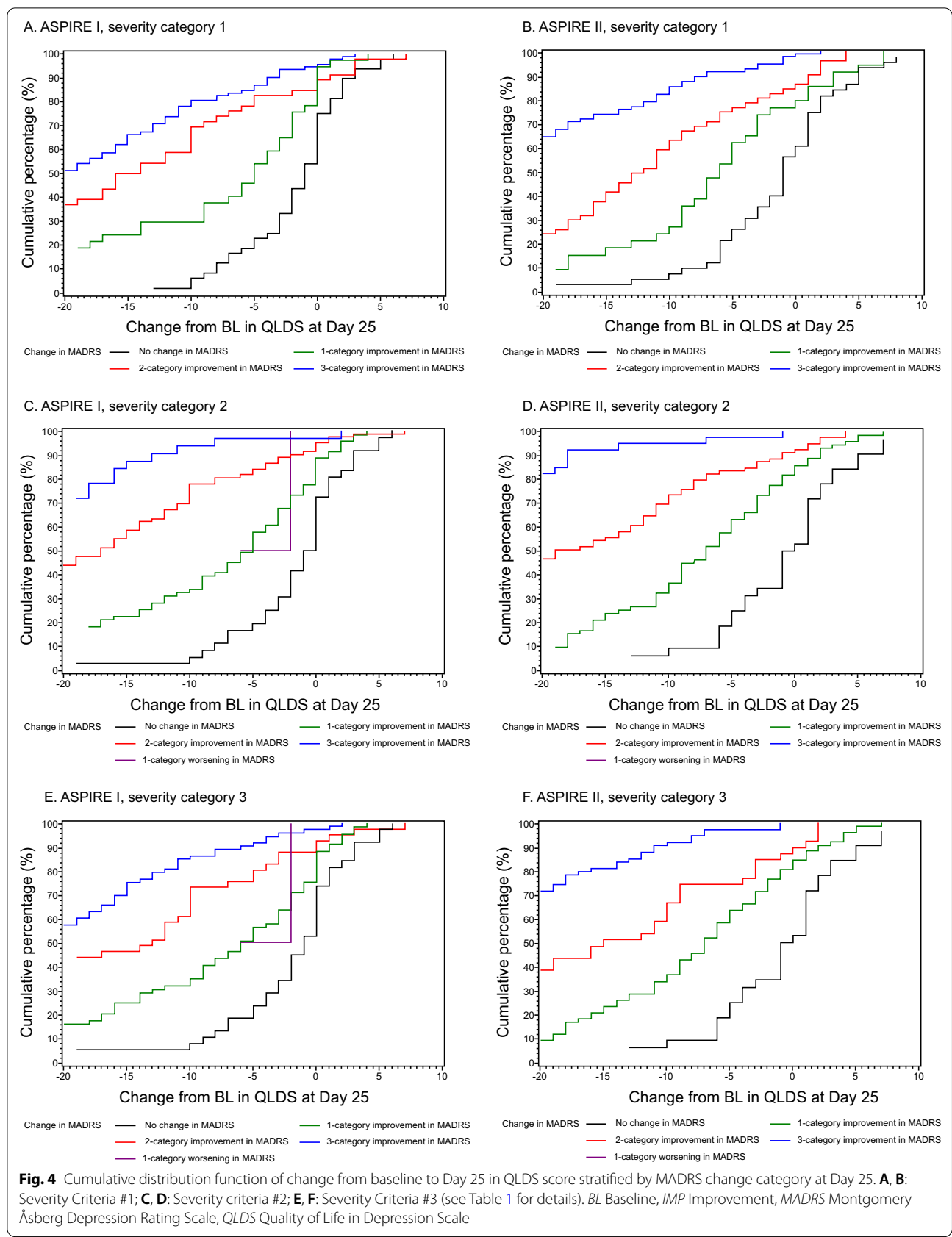
in QLDS score and change in EQ-VAS score between Baseline and Day 25 was -0.62 ($p < 0.0001$). A 7-point improvement in EQ-VAS score led to a -9.66-point improvement in QLDS score in the regression model ($\beta_0 = -7.84$, $p < 0.0001$; $\beta_1 = -0.26$, $p < 0.0001$). The MCT of QLDS using the distribution-based approach was -5.60 with 1/2 SD of change in QLDS score between Baseline and Day 25.

MCT for QLDS

The MADRS and the EQ-VAS anchors had the strongest correlations with the change in QLDS score at Day 25 compared to the CGI-SS-r and CGI-SR-I. Therefore, the MADRS and EQ-VAS anchors were considered more predominantly in the determination of the MCT. The MADRS was the primary anchor variable with the EQ-VAS, CGI-SS-r and CGI-SR-I used as supporting anchor measures. The results obtained using the supporting anchors and the distribution results were consistent with the MADRS anchor-based results. Considering the MCTs for the QLDS score identified using the MADRS and EQ-VAS, the mean improvement in QLDS score was -7.90 for the ASPIRE I study and -7.92 for the ASPIRE II study. The MCTs obtained from the MADRS and EQ-VAS anchor-based analysis and the distribution-based approach for both ASPIRE
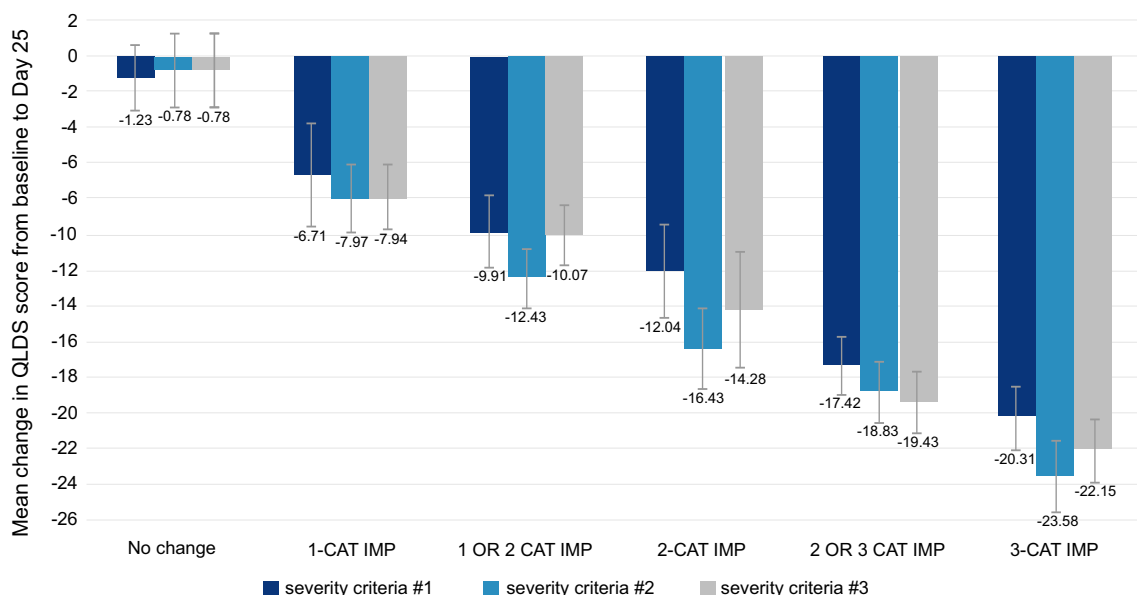
Fig. 2 ASPIRE II–mean change in QLDS score from baseline to Day 25 using MADRS as an anchor. Values are represented as mean ± 95% CI. MADRS was categorized as: severity criteria #1—no depression (score 0–12), slight depression (score 13–21), moderate depression (score 22–28), and severe depression (score > 28–60); severity criteria #2—no depression (score 0–6), mild depression (score 7–19), moderate depression (score 20–34), and severe depression (score > 34–60); and severity criteria #3—no depression (score 0–12), mild depression (score 13–17), moderate depression (score 18–34), and severe depression (score > 34–60). *CAT* Category, *CI* Confidence Interval, *IMP* Improvement, *MADRS* Montgomery–Åsberg Depression Rating Scale, *QLDS* Quality of Life in Depression Scale

Fig. 3 Meaningful change thresholds of QLDS using anchor- and distribution-based approach. Values represent MCT for various categories or points of improvement in QLDS score from Baseline to Day 25. For all MADRS severity criteria—one-category improvement and for EQ-VAS—7-point improvement. *EQ-VAS* EuroQol Visual Analog Scale, *MADRS* Montgomery–Åsberg Depression Rating Scale, *MCT* meaningful change threshold; *QLDS* Quality of Life in Depression Scale, *SD* Standard Deviation

I and ASPIRE II are summarized in Fig. 3. In both trials, a separation was observed on CDF curves between different MADRS change category groups with all three MADRS severity criteria (Fig. 4).

Discussion

This study aimed to determine the within-person MCT for QLDS among patients with MDD and acute suicidal ideation or behavior using data from two separate phase-3 trials of esketamine nasal spray. Using

Fig. 4 Cumulative distribution function of change from baseline to Day 25 in QLDS score stratified by MADRS change category at Day 25. **A**, **B**: Severity Criteria #1; **C**, **D**: Severity criteria #2; **E**, **F**: Severity Criteria #3 (see Table 1 for details). *BL* Baseline, *IMP* Improvement, *MADRS* Montgomery–Åsberg Depression Rating Scale, *QLDS* Quality of Life in Depression Scale

Rozjabek *et al. Journal of Patient-Reported Outcomes*   (2022) 6:74

Page 9 of 11

multiple anchors and distribution-based approaches, our findings suggest that within-patient MCT of QLDS among patients with MDD and acute suicidal ideation or behavior ranged from -11.4 to -6.7. The median of MCTs obtained using various approaches were -8.09 (mean $= -7.9$). Therefore, we recommend using an 8-point change as an MCT for the QLDS.

Evidence for an MCT of -8 points on the QLDS was supported using multiple anchors, including both patient-reported and clinician-reported outcomes measures, and distribution-based approaches. Although the MCT can be determined using either an anchor-based or distribution-based approach [48], we chose to utilize both methods for validation. Results from other anchor-based analyses such as the CGI-SS-r and CGI-SR-I provided additional information for the selection of the most appropriate MCT. For both the CGI-SR-I and the CGI-SS-r in ASPIRE II, the number of patients who improved only one category was small (26 and 21 respectively), which may have contributed to the variability in the results. The median QLDS score change for a one category change in CGI-SR-I and CGI-SS-r was -9.0 and -5.0, respectively. These findings are further supported by the separation of CDF curves between the MADRS change categories. CDF curves demonstrated a clear separation in the distribution of QLDS score between no change and a 1, 2 or 3-category improvement in MADRS scores. Although the distribution-based methods have been previously utilized in MCT analyses, these methods are sensitive to the homogeneity of the distribution and can result in MCTs which may be too small of a change to be considered meaningful. Use of both approaches in this analysis allowed for a range of MCT values to help interpret meaningful within-patient change. Studies have reported the inadequate coverage of outcomes integral from the patient's and clinician's point of view in research, as symptomatic outcomes are often the only outcome of interest in clinical trials [10, 51]. Lack of quantitative measures in patient-reported outcomes instruments to distinguish between treatment responders and non-responders, as well as inadequate implementation of such instruments, could be a reason for the aforementioned bias towards symptomatic outcomes [51]. However, the concept of MCT has been increasingly utilized to evaluate pharmacological interventions across therapy areas [52, 53]. MCTs have also been developed for many depression rating scales such as the MADRS, 9-item Patient Health Questionnaire, Beck Depression Inventory and others [54–57]. Our study adds a quantitative aspect to the QLDS and will assist researchers and clinicians in assessing the benefit of a therapy. The reported MCT values are generalizable to the study population in which they were derived, which is adult patients with MDD requiring acute psychiatric hospitalization due to an imminent suicide risk.

Limitations

A potential limitation of this study is that a patient-reported global impression of change or severity in depression was not available for use as an anchor for deriving the MCT. Instead, MADRS and other clinician global impression anchors were used to assess symptoms and behavior of the patient. The EQ-VAS, which is patient-reported, can assess a patient's overall assessment of their health; however, it would still be important to confirm the MCT in future studies where a depression specific patient global impression anchor is available. Although we obtained the MCT using data from two phase-3 trials among patients with MDSI, further research should be conducted to confirm these results in a similar MDSI population and evaluate an MCT value on the QLDS within a general MDD population.

Conclusion

Analysis of data from two double-blind, randomized, phase-3 trials among patients with MDD and acute suicidal ideation or behavior resulted in an MCT of 8-point improvement for the QLDS. This information is potentially useful to researchers and clinicians in interpreting changes in the HRQoL as measured by the QLDS among those with MDD and acute suicidal ideation or behavior.

Supplementary Information

> **Additional file 1. Table 1.** Results – MADRS as anchor in ASPIRE I. **Table 2.** Results – MADRS as anchor in ASPIRE II.

Rozjabek *et al. Journal of Patient-Reported Outcomes*   (2022) 6:74

Page 10 of 11

Author details

[1]Janssen Global Services, LLC, Raritan, NJ, USA. [2]Janssen Research and Development, LLC, Horsham, PA, USA. [3]Health Economics and Market Access Research, Janssen-Cilag, Neuss, North Rhine-Westphalia, Germany. [4]Janssen Research and Development, LLC, Titusville, NJ, USA. [5]Janssen Research and Development, LLC, Milpitas, CA, USA.

References

1. GBD Results Tool (2017) Retrieved 05 Oct 2020, from http://ghdx.healthdata.org/gbd-results-tool
2. Key substance use and mental health indicators in the United States: results from the 2018 National Survey on Drug Use and Health (HHS Publication No. PEP19–5068, NSDUH Series H-54). (2019). Substance Abuse and Mental Health Services Administration, Rockville, MD
3. Cui R (2015) Editorial: a systematic review of depression. Curr Neuropharmacol 13(4):480. https://doi.org/10.2174/1570159x1304150831123535
4. IsHak WW, Mirocha J, James D, Tobia G, Vilhauer J, Fakhry H, Pi S, Hanson E, Nashawati R, Peselow ED, Cohen RM (2015) Quality of life in major depressive disorder before/after multiple steps of treatment and one-year follow-up. Acta Psychiatr Scand 131(1):51–60. https://doi.org/10.1111/acps.12301
5. Beck A, Crain AL, Solberg LI, Unutzer J, Glasgow RE, Maciosek MV, Whitebird R (2011) Severity of depression and magnitude of productivity loss. Ann Fam Med 9(4):305–311. https://doi.org/10.1370/afm.1260
6. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, Grant BF (2018) Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiat 75(4):336–346. https://doi.org/10.1001/jamapsychiatry.2017.4602
7. Fife D, Reps J, Cepeda MS, Stang P, Blacketer M, Singh J (2018) Treatment resistant depression incidence estimates from studies of health insurance databases depend strongly on the details of the operating definition. Heliyon 4(7):e00707. https://doi.org/10.1016/j.heliyon.2018.e00707
8. Cuijpers P (2011) The patient perspective in research on major depression. BMC Psychiatry 11:89. https://doi.org/10.1186/1471-244X-11-89
9. Crossnohere NL, Brundage M, Calvert MJ, King M, Reeve BB, Thorner E, Wu AW, Snyder C (2021) International guidance on the selection of patient-reported outcome measures in clinical trials: a review. Qual Life Res 30(1):21–40. https://doi.org/10.1007/s11136-020-02625-z
10. Kamenov K, Cabello M, Coenen M, Ayuso-Mateos JL (2015) How much do we know about the functional effectiveness of interventions for depression? A systematic review. J Affect Disord 188:89–96. https://doi.org/10.1016/j.jad.2015.08.035
11. Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Attiullah N, Boerescu D (2006) How should remission from depression be defined? The depressed patient's perspective. Am J Psychiatry 163(1):148–150. https://doi.org/10.1176/appi.ajp.163.1.148
12. Lam RW, Parikh SV, Michalak EE, Dewa CS, Kennedy SH (2015) Canadian Network for Mood and Anxiety Treatments (CANMAT) consensus recommendations for functional outcomes in major depressive disorder. Ann Clin Psychiatry 27(2):142–149
13. FDA (2009) Guidance for industry on patient reported outcome measures: use in medical product development to support labeling claims; availability. Fed Reg 74(235):65132–65133
14. Patient-focused drug development guidance: methods to identify what is important to patients and select, develop or modify fit-for-purpose clinical outcome assessments (2018)
15. Public workshop on patient-focused drug development: guidance 4—incorporating clinical outcome assessments into endpoints for regulatory decision making (2019)
16. Hunt SM, McKenna SP (1992) The QLDS: a scale for the measurement of quality of life in depression. Health Policy 22(3):307–319. https://doi.org/10.1016/0168-8510(92)90004-u
17. McKenna SP, Hunt SM (1992) A new measure of quality of life in depression: testing the reliability and construct validity of the QLDS. Health Policy 22(3):321–330. https://doi.org/10.1016/0168-8510(92)90005-v
18. McKenna SP, Doward LC, Kohlmann T, Mercier C, Niero M, Paes M, Patrick D, Ramirez N, Thorsen H, Whalley D (2001) International development of the Quality of Life in Depression Scale (QLDS). J Affect Disord 63(1–3):189–199. https://doi.org/10.1016/s0165-0327(00)00184-1
19. Dunner DL, Kwong WJ, Houser TL, Richard NE, Donahue RM, Khan ZM (2001) Improved health-related quality of life and reduced productivity loss after treatment with bupropion sustained release: a study in patients with major depression. Prim Care Companion J Clin Psychiatry 3(1):10–16. https://doi.org/10.4088/pcc.v03n0103
20. Trick L, Stanley N, Rigney U, Hindmarch I (2004) A double-blind, randomized, 26-week study comparing the cognitive and psychomotor effects and efficacy of 75 mg (37.5 mg b.i.d.) venlafaxine and 75 mg (25 mg mane, 50 mg nocte) dothiepin in elderly patients with moderate major depression being treated in general practice. J Psychopharmacol 18(2):205–214. https://doi.org/10.1177/0269881104042622
21. Burt VK, Wohlreich MM, Mallinckrodt CH, Detke MJ, Watkin JG, Stewart DE (2005) Duloxetine for the treatment of major depressive disorder in women ages 40 to 55 years. Psychosomatics 46(4):345–354. https://doi.org/10.1176/appi.psy.46.4.345
22. Zajecka J, Schatzberg A, Stahl S, Shah A, Caputo A, Post A (2010) Efficacy and safety of agomelatine in the treatment of major depressive disorder: a multicenter, randomized, double-blind, placebo-controlled trial. J Clin Psychopharmacol 30(2):135–144. https://doi.org/10.1097/JCP.0b013e3181d420a7
23. Callaghan P, Khalil E, Morres I, Carter T (2011) Pragmatic randomised controlled trial of preferred intensity exercise in women living with depression. BMC Public Health 11:465. https://doi.org/10.1186/1471-2458-11-465
24. Kornstein SG, Wohlreich MM, Mallinckrodt CH, Watkin JG, Stewart DE (2006) Duloxetine efficacy for major depressive disorder in male vs. female patients: data from 7 randomized, double-blind, placebo-controlled trials. J Clin Psychiatry 67(5):761–770. https://doi.org/10.4088/jcp.v67n0510
25. Shi L, Namjoshi MA, Swindle R, Yu X, Risser R, Baker RW, Tohen M (2004) Effects of olanzapine alone and olanzapine/fluoxetine combination on health-related quality of life in patients with bipolar depression: secondary analyses of a double-blind, placebo-controlled, randomized clinical trial. Clin Ther 26(1):125–134. https://doi.org/10.1016/s0149-2918(04)90013-6
26. Guelfi JD, Ansseau M, Timmerman L, Korsgaard S, Mirtazapine-Venlafaxine Study, G (2001) Mirtazapine versus venlafaxine in hospitalized severely depressed patients with melancholic features. J Clin Psychopharmacol 21(4):425–431. https://doi.org/10.1097/00004714-200108000-00010

27. Gorwood P, Benichou J, Moore N, Wattez M, Secouard MC, Desobry X, Picarel-Blanchot F, de Bodinat C (2020) Agomelatine in standard medical practice in depressed patients: results of a 1-year multicentre observational study in France. Clin Drug Investig 40(11):1009–1020. https://doi.org/10.1007/s40261-020-00957-9

28. Capelleri JC, Coon C, Wyrwich K (2016) PRO consortium webinar: methods for determining clinically meaningful change. Retrieved 7 Oct 2020, from https://c-path.org/methods-for-determining-clinically-meaningful-change/

29. Hudgens S, Symonds T, McLeod L, Coon C (2016) Moving the science forward: psychometric considerations and study designs for understanding meaningful change and conducting mixed methods research, 2016 ISPOR 19th annual european congress. International Society for Pharmacoeconomics and Outcomes Research, Vienna, Austria

30. Coon CD, Cook KF (2018) Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res 27(1):33–40. https://doi.org/10.1007/s11136-017-1616-3

31. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S, Industry Advisory Committee of International Society for Quality of Life, R (2013) Methods for interpreting change over time in patient-reported outcome measures. Qual Life Res 22(3):475–483. https://doi.org/10.1007/s11136-012-0175-x

32. Fu DJ, Ionescu DF, Li X, Lane R, Lim P, Sanacora G, Hough D, Manji H, Drevets WC, Canuso CM (2020) Esketamine nasal spray for rapid reduction of major depressive disorder symptoms in patients who have active suicidal ideation with intent: double-blind, randomized study (ASPIRE I). J Clin Psychiatry. https://doi.org/10.4088/JCP.19m13191

33. Ionescu DF, Fu DJ, Qiu X, Lane R, Lim P, Kasper S, Hough D, Drevets WC, Manji H, Canuso CM (2021) Esketamine nasal spray for rapid reduction of depressive symptoms in patients with major depressive disorder who have active suicide ideation with intent: results of a phase 3, double-blind, randomized study (ASPIRE II). Int J Neuropsychopharmacol 24(1):22–31. https://doi.org/10.1093/ijnp/pyaa068

34. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (DSM–5), 5th edn. American Psychiatric Association, Washington, DC

35. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC (1998) The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 59(Suppl 20):22–33 (**quiz 34–57**)

36. Montgomery SA, Åsberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatry 134(4):382–389

37. Fu D-J, Canuso C, Ionescu DF, Li X, Lane R, Lim P, Hough D, Drevets W, Manji H (2019) ASPIRE-1: a phase 3 randomized study of esketamine nasal spray for rapid reduction of major depressive disorder symptoms in adult patients at imminent risk for suicide, 2019 IASR/AFSP international summit on suicide research. Miami, FL

38. Ionescu DF, Canuso CM, Qiu X, Lane R, Lim P, Hough D, Drevets W, Manji H, Fu D-J (2019) ASPIRE-2: a phase 3 randomized study of esketamine nasal spray for rapid reduction of major depressive disorder symptoms in adult patients at imminent risk for suicide, 2019 IASR/AFSP international summit on suicide research. Miami, FL

39. Montgomery SA (1994) Clinically relevant effect sizes in depression. Eur Neuropsychopharmacol 4(3):283–284

40. Muller MJ, Szegedi A, Wetzel H, Benkert O (2000) Moderate and severe depression. Gradations for the Montgomery–Asberg Depression Rating Scale. J Affect Disord 60(2):137–140. https://doi.org/10.1016/s0165-0327(99)00162-7

41. Muller-Thomsen T, Arlt S, Mann U, Mass R, Ganzer S (2005) Detecting depression in Alzheimer's disease: evaluation of four different scales. Arch Clin Neuropsychol 20(2):271–276. https://doi.org/10.1016/j.acn.2004.03.010

42. Weyer G (2005) International Scales for Psychiatry, 5th edn. Hogrefe Verlag, Göttingen, Germany

43. Alphs L, Fu DJ, Williamson D, Turkoz I, Jamieson C, Revicki D, Canuso CM (2020) Suicide Ideation and Behavior Assessment Tool (SIBAT): evaluation of intra- and inter-rater reliability, validity, and mapping to Columbia classification algorithm of suicide assessment. Psychiatry Res 294:113495. https://doi.org/10.1016/j.psychres.2020.113495

44. Busner J, Targum SD (2007) The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont) 4(7):28–37

45. Wyrwich K, Reeve B, Coon CD (2017) What's the score? Moving from items to scores—methods, considerations, and case examples, eighth annual patient-reported outcome consortium workshop, Bethesda, MD

46. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 20(10):1727–1736. https://doi.org/10.1007/s11136-011-9903-x

47. Pickard AS, Neary MP, Cella D (2007) Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 5:70. https://doi.org/10.1186/1477-7525-5-70

48. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC (2007) Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 7(5):541–546. https://doi.org/10.1016/j.spinee.2007.01.008

49. Revicki D, Hays RD, Cella D, Sloan J (2008) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 61(2):102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012

50. Jones IA, Togashi R, Heckmann N, Vangsness CT Jr (2020) Minimal clinically important difference (MCID) for patient-reported shoulder outcomes. J Shoulder Elbow Surg 29(7):1484–1492. https://doi.org/10.1016/j.jse.2019.12.033

51. Kamenov K, Cabello M, Nieto M, Bernard R, Kohls E, Rummel-Kluge C, Ayuso-Mateos JL (2017) Research recommendations for improving measurement of treatment effectiveness in depression. Front Psychol 8:356. https://doi.org/10.3389/fpsyg.2017.00356

52. Burback D, Molnar FJ, St John P, Man-Son-Hing M (1999) Key methodological features of randomized controlled trials of Alzheimer's disease therapy. Minimal clinically important difference, sample size and trial duration. Dement Geriatr Cogn Disord 10(6):534–540. https://doi.org/10.1159/000017201

53. Meyer RJ (2005) U.S. regulatory perspective on the minimal clinically important difference in chronic obstructive pulmonary disease. COPD 2(1):47–49. https://doi.org/10.1081/copd-200050660

54. Duru G, Fantino B (2008) The clinical relevance of changes in the Montgomery–Asberg Depression Rating Scale using the minimum clinically important difference approach. Curr Med Res Opin 24(5):1329–1335. https://doi.org/10.1185/030079908X291958

55. Hudgens S, Floden L, Blacowicz M, Jamieson C, Popova V, Fedgchin M, Drevets W, Cooper K, Lane R, Singh J (2020) 175 Determining meaningful change in depression symptoms assessed with PHQ-9 and SDS in treatment-resistant depression trials of esketamine nasal spray. CNS Spectr 25(2):311–312

56. Masson SC, Tejani AM (2013) Minimum clinically important differences identified for commonly used depression rating scales. J Clin Epidemiol 66(7):805–807. https://doi.org/10.1016/j.jclinepi.2013.01.010

57. Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, Ades AE, Lewis G (2015) Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. Psychol Med 45(15):3269–3279. https://doi.org/10.1017/S0033291715001270

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.