



OPEN ACCESS

EDITED BY

Sherry-Anne Muscat,
Alberta Hospital Edmonton, Canada

REVIEWED BY

William Lenderking,
Evidera, United States
Mariusz Stanisław Wigiłusz,
Medical University of Gdansk, Poland

*CORRESPONDENCE

Mark Opler
✉ mark.opler@panss.org

RECEIVED 06 September 2023

ACCEPTED 16 April 2024

PUBLISHED 01 May 2024

CITATION

Capodilupo G, Blattner R, Must A, Navarro SG and Opler M (2024) A qualitative investigation of the Montgomery–Åsberg depression rating scale: discrepancies in rater perceptions and data trends in remote assessments of rapid-acting antidepressants in treatment resistant depression. *Front. Psychiatry* 15:1289630. doi: 10.3389/fpsy.2024.1289630

COPYRIGHT

© 2024 Capodilupo, Blattner, Must, Navarro and Opler. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A qualitative investigation of the Montgomery–Åsberg depression rating scale: discrepancies in rater perceptions and data trends in remote assessments of rapid-acting antidepressants in treatment resistant depression

Gianna Capodilupo¹, Raymond Blattner^{1,2}, Anita Must^{1,3},
Silvia Gamazo Navarro^{1,4} and Mark Opler^{1,5*}

¹WCG Clinical Endpoint Solutions, Princeton, NJ, United States, ²Seton Hall University, South Orange, NJ, United States, ³Department of Psychiatry, Whanganui District Health Board, Whanganui, New Zealand, ⁴Chelsea and Westminster Hospital NHS Foundation Trust, London, United Kingdom, ⁵The PANSS Institute, New York, NY, United States

Introduction: Despite the development of many successful pharmaceutical interventions, a significant subset of patients experience treatment-resistant depression (TRD). Ketamine and its derivatives constitute a novel therapeutic approach to treat TRD; however, standard tools, such as the Montgomery–Åsberg Depression Rating Scale (MADRS) are still being used to measure symptoms and track changes.

Methods: The aim of this study was to review item-level differences between rate of data change (MADRS score) and rater-weighted perception of the most useful items for assessing change in symptoms while remotely conducting the 10-item version of the MADRS in TRD in a clinical trial of rapid-acting antidepressants. Two studies of rapid-acting antidepressants in the treatment of TRD were used to identify item-scoring trends when MADRS is administered remotely and repeatedly (733 subjects across 10 visits). Scoring trends were evaluated in tandem to a rater survey completed by 75 raters. This was completed to gain insight on MADRS items' perceived level of helpfulness when assessing change of symptoms in rapid-acting antidepressant trials.

Results: MADRS items 'Reduced sleep', 'Apparent sadness', and 'Pessimistic thoughts' were found to have the greatest average data change by visit, while raters ranked 'Reported sadness', 'Lassitude' and 'Apparent sadness' as the most helpful items when assessing symptom change.

Discussion: The diversion between rate of data-change ranking and rater perception of helpfulness could be related to difficulty in assessing specific items, to the novel treatment itself, and/or to the sensitivity to symptom change to which raters are accustomed in traditional antidepressant treatments.

KEYWORDS

depression, MADRS = Montgomery-Asberg depression rating scale, rater perception, rapid acting antidepressants, structured interview guide for the MADRS (SIGMA)

1 Introduction

In addressing the challenge of treatment-resistant depression (TRD), this manuscript focuses on the evolving landscape of depression treatment and the role of rapid-acting antidepressants (RAAD), such as ketamine and its derivative, esketamine. These novel treatments, approved by the FDA in 2019 for TRD (1), constitute a significant advance in the pharmacotherapy of depression, but pose new questions about the effectiveness of traditional assessment tools like the Montgomery-Åsberg Depression Rating Scale (MADRS) in evaluating rapid symptom changes. The 10-item MADRS, a standard in clinical research and practice, is used to measure symptom severity and changes in depressive syndromes, yet its suitability for novel, rapid-acting treatments remains under-explored, particularly in the context of remote administration (2, 3).

Depression, a leading contributor to global disability, affects about 322 million people worldwide, with an 18.4% increase in prevalence between 2005 and 2015 (4). Traditional treatments have evolved significantly since the 1950s, from tricyclic antidepressants and monoamine oxidase inhibitors to second-generation antidepressants like SSRIs and SNRIs. However, for a substantial subset of patients, these treatments fail to provide full or partial remission, leading to the classification of TRD (5–7). The complexity of defining and measuring treatment resistance, as highlighted by studies like the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial, underscores the need for effective treatment options and reliable assessment methods (8–10). Rapid-acting antidepressants, a departure from the monoaminergic focus of traditional treatments, offer a promising avenue for treating TRD. Ketamine, antagonizing the NMDA receptor, induces a glutamate surge, fostering brain adaptability and pathway creation (11, 12). Its rapid symptom relief, observed within 24 to 72 hours post-administration, marks a stark contrast to the gradual effects of conventional antidepressants (13–15). However, the long-term effects and optimal dosing of ketamine remain areas for further research (16–18).

The MADRS, designed to sensitively capture treatment-induced symptom changes in depression, has historically been administered in-person but is increasingly used in remote settings. This shift raises concerns, particularly for items like ‘Apparent sadness,’ which rely on observational assessment (19–24). Despite its widespread use and proven interrater reliability across various languages, the appropriateness of MADRS for rapid-acting antidepressants, especially in remote settings, warrants examination (25–35). In addition to that RAADs can generate responses within hours or days, rather than weeks or months, while our rating instruments were designed to assess mood symptoms over a 7-day time frame, typically. Consequently, adaptive approaches are required for existing scales to meet the need of adjustment. Yavorsky et al. provide an excellent summary on adaptations to the standard rating instruments allowing to reflect short-term changes in which RAADs act, as well as implementing novel rating measures. Critically, they also discuss limitations and challenges to the currently used rating measures including any conceptual biases of raters (36).

This publication, therefore, seeks to analyze the item-level differences in MADRS scores and the raters’ perception of their helpfulness in determining symptom severity change in TRD during clinical trials of rapid-acting antidepressants. The study hypothesizes that while some MADRS items may align with rater perceptions, discrepancies are expected due to the unique nature of rapid-acting antidepressant treatment and the challenges of remote assessment. This investigation is crucial for ensuring that depression assessment tools remain relevant and effective in the rapidly evolving landscape of antidepressant therapy.

2 Materials and methods

2.1 Data change on MADRS items in two protocols of a phase 3 study

We have established the rate of change of all items of MADRS for the entirety of two separate protocols of a Phase 3 study of rapid-acting antidepressants using a similar design, but varying in terms visit frequency (between 24 hours and 7 days). We used deidentified datasets only containing the visit date, but no other information that would be considered protected health information, only to perform a qualitative, exploratory analysis looking at individual MADRS items and changes in their score across visits. Rate of change was calculated by first determining between-visit MADRS data change by item; we then divided the per-item change by the number of days that had passed between an individual visit and the previous visit to establish change per day. We rank-ordered the per-visit change for a given item. The data was solely used to create the rankings discussed, and no other quantitative analysis has been completed for the purpose of this study.

2.2 Remote rater-experience survey

We conducted a survey of clinical research professionals who have participated in the above-mentioned Phase 3 programs of rapid-acting antidepressants utilizing new modalities for the assessment of depressive symptoms, including remote evaluation technologies (telepsychiatry) and versions of MADRS that have been adapted for use over shorter recall periods, e.g. last 24 hours. A total of 75 experienced raters from 13 countries were recruited for this survey study, all of whom participated in at least one TRD clinical trial for rapid-acting antidepressants. All survey participants were polled about which MADRS items they considered most and least helpful for assessing changes in symptoms in two protocols of a Phase 3 study with the same rapid-acting agent to treat TRD. Many raters participated in these studies on almost a daily basis, with a rotation of subjects. Items considered as most useful as endorsed by raters for assessing TRD related symptom changes were ranked as most important, second most important, and so on.

We compared the data change on MADRS items and the raters’ rankings to gain additional insight into a.) the rater experience of each item’s helpfulness in determining depression severity and b.) the rate of change in MADRS scores in a clinical trial for rapid-acting antidepressants in TRD.

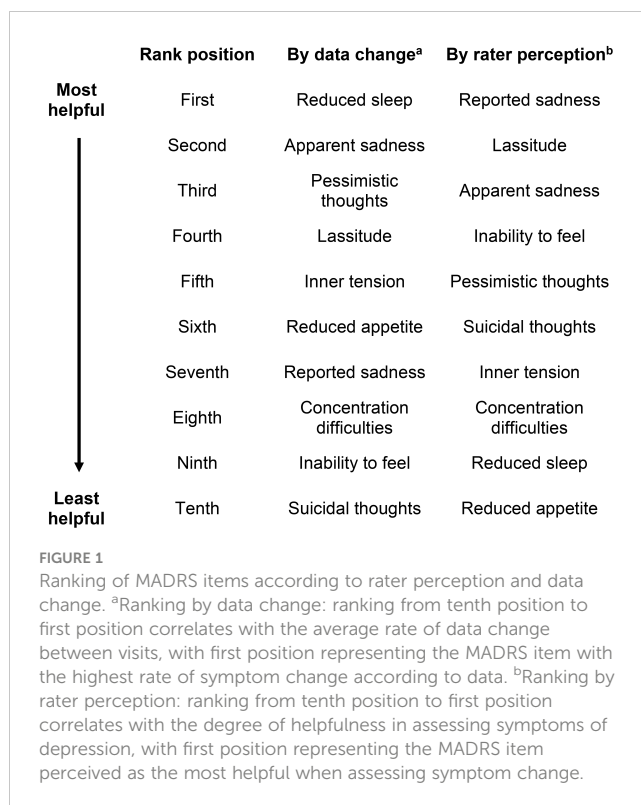
3 Results

Surveyed raters conducted assessments via telephone to ensure the integrity of blinding in these trials. Raters were selected from studies with more than 700 subjects combined (N = 733) across 20 countries and 16 different languages. In the two examined protocols of a Phase 3 study with the same rapid-acting agent to treat TRD, 733 subjects were included in the dataset across 10 visits (excluding early termination; 12 in visits in total). MADRS total scores ranged from 0 to 53 across visits for each protocol. Mean total MADRS score was 30.39. Individual item scores ranged from 0 to 6 for each item. Ranking of items was determined by rater perception of helpfulness in assessing symptom change, and by average MADRS score change in between-visit data. Rater ranking order (from 10 to 1) correlates with an increasing level of perceived helpfulness in determining symptom change, with 1 being the most helpful. Data ranking of MADRS items based on the deidentified dataset (from 10 to 1) correlates with a greater average rate of MADRS item-level score change between visits, with 1 representing the item with the greatest average between-visit MADRS item score change. Helpfulness of MADRS items in assessing symptom change, ranked according to average rate of data change and by rater perception of helpfulness is shown in Figure 1, with a ranking in the first position representing the MADRS item with either the highest rate of symptom change according to greatest average between-visit MADRS data change ('By data change') or as being perceived as the most helpful by raters ('By rater perception') when assessing symptoms.

'Reduced sleep', 'Apparent sadness', and 'Pessimistic thoughts' were found to have the highest rate of data change by visit, however raters ranked 'Reduced sleep' in the ninth position of helpfulness, with 'Apparent sadness' and 'Pessimistic thoughts' placed third and fifth in terms of perceived helpfulness. In contrast, 'Reported sadness' was perceived as being the most helpful item for raters in assessing symptoms of depression but ranked seventh according to rate of data change. 'Reduced appetite' was ranked as the least useful item by rater perception, while rate of data change places this in the sixth position of helpfulness. 'Suicidal thoughts' was ranked sixth according to rater perception but placed tenth by rate of data change. The low ranking according to data change was caused by the smallest average change between visits for this item. 'Inability to feel' is ranked as the fourth most helpful item in terms of rater perception and shows the second lowest average change between visits in the data, ranking ninth according to data change.

4 Discussion

When examining the first item of the MADRS, 'Reported sadness', the account of symptoms and depression by subjects is the most important evidence one can receive. However, several factors including age of subject, age at onset of first depressive episode, education, length of illness, etc. could influence how sadness is being reported by subjects (37), and these may be due to differences in how depression has been conceptualized (38).



Therefore, relying heavily on clinical background can be imperative for the clinician when administering and scoring the MADRS. 'Apparent sadness' also appears to be of critical interest, as it would have been presumed as a difficult item to assess in remote assessments. Nevertheless, both earlier studies by Kobak et al. (32) and a recent study by Sumiyoshi et al. conducted in Japan (39) found excellent reliability for the MADRS in remotely interviewing patients with MDD showing high consistency between remote and in-person interviews. These studies also emphasize the key importance of well-trained raters, which might be critical for accurately assigning ratings for more challenging items such as 'apparent sadness'.

Appetite changes are not shared symptoms across all subjects with depression, and there can be marked increases or decreases in appetite in patients diagnosed with depression. Simmons et al. (40) report that only 48% of subjects experience reduced appetite, leaving 52% unaccounted for. Raters may be noticing this disconnect, with the focus of appetite in MADRS relying on reduction as opposed to bidirectional change and weighting the importance of this item accordingly.

We believe the discrepancy between rater-perceived ranking and ranking based on rate of data change for 'Reduced sleep' is due to the variability in the data itself. Sleep as a construct may be moving too quickly to be meaningful to raters more used to conducting traditional assessments. Additionally, sleep can move in more than one direction (i.e., it can both increase and decrease), and can be attributed to multiple factors not including symptoms of depression or drug treatment effects (environmental, pain, change in caffeine consumption, etc.), potentially deprioritizing sleep to raters and encouraging them to rely more heavily on other items.

Ranking for raters on ‘Suicidal thoughts’ (6th on perceived helpfulness, but last on data change order) could demonstrate the importance of this item when it is reported, although the frequency of report may be low, and overall severity in most cases may also be low. When patients report suicidal thoughts or behaviors during MADRS interviews, the weight of this item may increase for raters, although the frequency with which it is seen in our dataset was also low.

In assessing depression, ‘Inability to feel’, involving changes in emotions, may become one of the most useful items to assess overall depression when reported, however, it may again be an item that infrequently shows large variability over shorter assessment periods.

The performance of traditional assessments to obtain ratings/scores for studies of novel treatments for depression is a critically important matter. It is imperative to reliably measure changes in symptoms—particularly for treatments that may carry unique side-effect profiles and safety risks. Researchers have used assessments on the widely agreed upon core symptoms of depression for decades. MADRS, though widely accepted by regulatory agencies and used by clinicians and researchers, may have a different value and clinical significance in the context of rapid-acting antidepressants. Using MADRS in short-interval, remote evaluations, with repeated assessments performed within 24 hours, might present challenges in accurately capturing symptom change. Certainly, not all depressed individuals have the same depressive symptoms at baseline and the dynamic nature of a therapeutic response to RAADs could potentially result in a rapid alleviation of certain symptoms, e.g. improvements in subjective mood, while leaving some other functional aspects less improved, resulting in a heterogeneity of symptom resolution and a sense of uncertainty in raters (36).

As research increasingly supports the use of rapid-acting antidepressants, and as their market approval increases, it is incumbent on clinicians to review and refine assessment processes. The rapid change in symptoms presents a challenge for clinicians, especially if the assessment is administered remotely. This study has compared how the rate of between-visit data changes for MADRS items relate to those that raters have identified as being more helpful or more challenging. The diversion between rate of data-change and rater perception of helpfulness could be related to several factors, one being difficulty in assessing specific items. A depression rater training study by Sajatovic et al. (41) showed no significant difference between raters based on country, level of experience with diagnosis, or previous training in terms of the items they identified as the most difficult to rate, namely ‘Apparent sadness’, ‘Inner tension’, ‘Concentration difficulties’, ‘Lassitude’, and ‘Inability to feel’. Similar results have been shown when comparing rater training using MADRS to other mood rating scales, such as the Hamilton Depression Rating Scale (HAM-D) and Young Mania Rating Scale (YMRS) in a bipolar disorder trial (42). We see the same MADRS items (listed above) identified as most difficult to rate with no significant difference for raters based on country, experience, diagnosis, or previous training, suggesting the items themselves present difficulty to raters (43). MADRS has also been noted as a more difficult scale to utilize when compared to other commonly used depressive symptom rating scales (HAM-D

and YMRS), thus it could benefit from further insight and qualitative analysis (41).

A recent factor analysis conducted on two esketamine trials has taken an interesting approach to explore potential symptomatic clusters grouped around the rate of symptom change as detected by MADRS items (44). Three factors were identified labelled as affective/anhedonic symptoms (apparent sadness, reported sadness, lassitude, inability to feel); anxiety and vegetative symptoms (inner tension, reduced sleep, reduced appetite, concentration difficulties); and hopelessness (pessimistic thoughts, suicidal thoughts). Strikingly, our results on rater perception of items follows exactly these clusters with the affective/anhedonic factor listed as the most helpful for raters, followed by the hopelessness factor, and lastly the anxiety and vegetative factor symptom group, which is probably prone to the highest variability and heterogeneity overall.

Thus, another aspect potentially explaining the diversion between MADRS score changes and raters’ perceptions could be the novel treatment itself or the sensitivity to the consequent symptom change to which raters are accustomed. An example for this is the fluctuation of sleep on a day-to-day basis that can affect daily ratings when assessing change in symptoms with the MADRS. Research indicates that MADRS item responses - those related to sleep, most specifically - change in relation to patient experience and when compared to the prior administered assessment. Sleep, however, also influences memory and emotional memory in subjects with depression (45) and this can present additional challenges when ratings are conducted based on subject’s report every day or every other day, as is common in clinical trials of rapid-acting antidepressants.

There are inherent complications in the assessment of rapid change in TRD over short periods of time and rapid (i.e., between-visit) symptom changes at the item level can pose a challenge to ranking and assessing for severity. Singh et al. (14) examined the efficacy of both full and abbreviated MADRS scores in evaluating the response to intravenous esketamine. They conducted assessments of depressive symptoms at 2-, 4-, and 24-hours post-infusion, opting to omit the sleep and appetite items for the shorter 2 and 4-hour assessments. Johnson and colleagues also investigated the MADRS’s suitability within a 24-hour recall period, finding comparable content validity and high internal consistency and test-retest reliability (46). While most participants deemed a 24-hour recall period sufficient for assessing meaningful changes in depression symptoms, reduced sleep and appetite were noted exceptions, echoing Singh et al.’s decision to exclude these items. Yavorsky et al. analyzed esketamine trial data, revealing the MADRS’s sensitivity to change over short periods, albeit with limited responsiveness in the sleep and suicide items (47). These findings underscore the challenge of effectively capturing clinical change within a 24-hour timeframe, particularly for traditional depressive symptoms like mood, appetite, and sleep disturbances. Novel treatments for TRD have yet to yield novel assessments that are sensitive to change over such short periods of time.

Each item discussed and present in the MADRS is important in assessing overall depression and symptomology of the diagnoses, and clinical experience brings clarity to differences in items and

scores, such as the importance of suicidal thoughts and sleep deprivation.

5 Conclusion

Rapid-acting antidepressants, such as ketamine and its derivatives, appear to induce more rapid changes in depression symptoms, which presents a challenge in the accurate capture of symptom change when using conventional rating scales. Our findings indicate a diversion between the rate of data change as measured by MADRS scores vs the raters' perception of helpfulness of specific MADRS items in determining clinical improvement and depression severity change. While each MADRS item remains important in an assessment of depression symptomatology, it may be beneficial to refine raters' sensitivity to changes in depressive symptoms over short periods of time, and to the specific side effects associated with novel treatment approaches. Further to this, gaining experience in the use of rapid-acting antidepressants, and in the ability to measure short and long-term effects of these novel treatments, might influence our notion of defining criteria for treatment resistance in depression.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study of human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

References

- Food & Drug Administration. FDA approves new nasal spray medication for treatment-resistant depression; available only at a certified doctor's office or clinic. (2019).
- Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. (1979) 134:382–29. doi: 10.1192/bjp.134.4.382
- First MB. Diagnostic and statistical manual of mental disorders, 5th edition, and clinical utility. *J Nerv Ment Dis*. (2013) 201(9):727–9. doi: 10.1097/NMD.0b013e3182a2168a
- World Health Organization. *Depression and other common mental disorders: global health estimates* (No. WHO/MSD/MER/2017.2) (2017). Available online at: <https://apps.who.int/iris/handle/10665/254610> (Accessed September 5, 2023).
- Nestler EJ, Barrot M, DiLeone RJ, Eisch AJ, Gold SJ, Monteggia LM. Neurobiology of depression. *Neuron*. (2002) 34:13–25. doi: 10.1016/S0896-6273(02)00653-0
- Fava M, Davidson KG. Definition and epidemiology of treatment-resistant depression. *Psychiatr Clin North Am*. (1996) 19:179–200. doi: 10.1016/S0193-953X(05)70283-5
- Sackeim HA. The definition and meaning of treatment-resistant depression. *J Clin Psychiatry*. (2001) 62 Suppl 16:10–7.
- Souery D, Papakostas GI, Trivedi MH. Treatment-resistant depression. *J Clin Psychiatry*. (2006) 67 Suppl 6:16–22.
- Rush AJ. STAR*D: what have we learned? *Am J Psychiatry*. (2007) 164:201–4. doi: 10.1176/ajp.2007.164.2.201
- Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. (2006) 163:1905–17. doi: 10.1176/ajp.2006.163.11.1905
- Abdallah CG, Adams TG, Kelmendi B, Esterlis I, Sanacora G, Krystal JH. Ketamine's mechanism of action: a path to rapid-acting antidepressants. *Depress Anxiety*. (2016) 33:689–97. doi: 10.1002/da.22501
- Jelen LA, Young AH, Stone JM. Ketamine: A tale of two enantiomers. *J Psychopharmacol*. (2021) 35:109–23. doi: 10.1177/0269881120959644
- Aan Het Rot M, Zarate CA Jr, Charney DS, Mathew SJ. Ketamine for depression: where do we go from here? *Biol Psychiatry*. (2012) 72:537–47. doi: 10.1016/j.biopsych.2012.05.003
- Singh JB, Fedgchin M, Daly EJ, De Boer P, Cooper K, Lim P, et al. A double-blind, randomized, placebo-controlled, dose-frequency study of intravenous ketamine in patients with treatment-resistant depression. *Am J Psychiatry*. (2016) 173:816–26. doi: 10.1176/appi.ajp.2016.16010037
- Murrough JW, Iosifescu DV, Chang LC, Al Jurdi RK, Green CE, Perez AM, et al. Antidepressant efficacy of ketamine in treatment-resistant major depression: a two-site

Author contributions

GC: Writing – review & editing, Writing – original draft. RB: Writing – review & editing, Writing – original draft. AM: Writing – review & editing, Writing – original draft. SN: Writing – review & editing, Writing – original draft. MO: Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

GC, RB and SN are employed by WCG Clinical Endpoint Solutions. AM is a contractor of WCG Clinical Endpoint Solutions and a part-time employee of Whanganui District Health Board, Whanganui, NZ. MO is employed by WCG Clinical Endpoint Solutions; has received royalties from Pearson Inc. from sales of the PANSS Manual; and has received past grant funding from NIH, Brain & Behavior Foundation, the Stanley Medical Research Institute, and the Qatar National Research Fund.

The reviewer WL declared a past co-authorship with the author MO to the handling editor.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- randomized controlled trial. *Am J Psychiatry*. (2013) 170:1134–42. doi: 10.1176/appi.ajp.2013.13030392
16. Lent JK, Arredondo A, Pugh MA, Austin PN. Ketamine and treatment-resistant depression. *AANA J*. (2019) 87:411–9.
17. Katalinic N, Lai R, Somogyi A, Mitchell PB, Glue P, Loo CK. Ketamine as a new treatment for depression: a review of its efficacy and adverse effects. *Aust N Z J Psychiatry*. (2013) 47:710–27. doi: 10.1177/0004867413486842
18. Short B, Fong J, Galvez V, Shelker W, Loo CK. Side-effects associated with ketamine use in depression: a systematic review. *Lancet Psychiatry*. (2018) 5:65–78. doi: 10.1016/S2215-0366(17)30272-9
19. Asberg M, Schalling D. Construction of a new psychiatric rating instrument, the Comprehensive Psychopathological Rating Scale (CPRS). *Prog Neuropsychopharmacol*. (1979) 3:405–12. doi: 10.1016/0364-7722(79)90055-9
20. Iannuzzo RW, Jaeger J, Goldberg JF, Kafantaris V, Sublette ME. Development and reliability of the HAM-D/MADRS interview: an integrated depression symptom rating scale. *Psychiatry Res*. (2006) 145:21–37. doi: 10.1016/j.psychres.2005.10.009
21. Hermens ML, Adèr HJ, van Hout HP, Terluin B, van Dyck R, de Haan M. Administering the MADRS by telephone or face-to-face: a validity study. *Ann Gen Psychiatry*. (2006) 5:3. doi: 10.1186/1744-859X-5-3
22. Maier W, Philipp M, Heuser I, Schlegel S, Buller R, Wetzel H. Improving depression severity assessment—I. Reliability, internal validity and sensitivity to change of three observer depression scales. *J Psychiatr Res*. (1988) 22:3–12. doi: 10.1016/0022-3956(88)90022-2
23. Maust D, Cristancho M, Gray L, Rushing S, Tjoa C, Thase ME. Psychiatric rating scales. In: *Handbook of Clinical Neurology*, vol. 106. Elsevier (2012). p. 227–37.
24. Svanborg P, Asberg M. A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Asberg Depression Rating Scale (MADRS). *J Affect Disord*. (2001) 64:203–16. doi: 10.1016/S0165-0327(00)00242-1
25. Schmidtke A, Fleckenstein P, Moises W, Beckmann H. Studies of the reliability and validity of the German version of the Montgomery-Asberg Depression Rating Scale (MADRS). *Schweiz Arch Neurol Psychiatr*. (1988) 139:51–65.
26. Corruble E, Purper D, Payan C, Guelfi J. Inter-rater reliability of two depression rating scales, MADRS and DRRS, based on videotape records of structured interviews. *Eur Psychiatry*. (1998) 13:264–6. doi: 10.1016/S0924-9338(98)80032-1
27. Ozer SK, Demir B, Tugal O, et al. Montgomery-Asberg depression rating scale: inter-rater reliability and validity study. *Turk Psikiyatri Derg*. (2022) 33(3):187–95. doi: 10.5080/u25844
28. Takahashi N, Tomita K, Higuchi T, Inada T. The inter-rater reliability of the Japanese version of the Montgomery-Asberg depression rating scale (MADRS) using a structured interview guide for MADRS (SIGMA). *Hum Psychopharmacol*. (2004) 19:187–92. doi: 10.1002/hup.576
29. Satthapit S, Posayaanuwat N, Sasaluksananont C, Kaewpornawan T, Singhakun S. The comparison of Montgomery and Asberg Depression Rating Scale (MADRS thai) to diagnostic and statistical manual of mental disorders (DSM) and to Hamilton Rating Scale for Depression (HRSD): validity and reliability. *J Med Assoc Thai*. (2007) 90(3):524–31.
30. Zhong B, Wang Y, Chen H, et al. Reliability, validity and sensitivity of Montgomery-Asberg depression rating scale for patients with current major depression disorder. *Chin J Behav Med Brain Sci*. (2011) 20:85–7.
31. Ahmadpanah M, Sheikhabaei M, Haghghi M, Roham F, Jahangard L, Akhondi A, et al. Validity and test-retest reliability of the Persian version of the Montgomery-Asberg Depression Rating Scale. *Neuropsychiatr Dis Treat*. (2016) 12:603–7. doi: 10.2147/NDT.S103869
32. Kobak KA, Williams JB, Jeglic E, Salvucci D, Sharp IR. Face-to-face versus remote administration of the Montgomery-Asberg Depression Rating Scale using videoconference and telephone. *Depress Anxiety*. (2008) 25(11):913–9. doi: 10.1002/da.v25:11
33. Holländare F, Andersson G, Engström I. A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. *J Med Internet Res*. (2010) 12:e49. doi: 10.2196/jmir.1392
34. Targum SD, Daly E, Fedgchin M, Cooper K, Singh JB. Comparability of blinded remote and site-based assessments of response to adjunctive esketamine or placebo nasal spray in patients with treatment resistant depression. *J Psychiatr Res*. (2019) 111:68–73. doi: 10.1016/j.jpsychires.2019.01.017
35. Targum SD, Catania CJ. Audio-digital recordings for surveillance in clinical trials of major depressive disorder. *Contemp Clin Trials Commun*. (2019) 14:100317. doi: 10.1016/j.conctc.2019.100317
36. Yavorsky C, Ballard E, Opler M, Sedway J, Targum SD, Lenderking W. Recommendations for selection and adaptation of rating scales for clinical studies of rapid-acting antidepressants. *Front Psychiatry*. (2023) 14:1135828. doi: 10.3389/fpsy.2023.1135828
37. Rush AJ, Carmody TJ, Ibrahim HM, Trivedi MH, Biggs MM, Shores-Wilson K, et al. Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. *Psychiatr Serv*. (2006) 57:829–37. doi: 10.1176/ps.2006.57.6.829
38. Carter JD, Frampton CM, Mulder RT, Luty SE, Joyce PR. The relationship of demographic, clinical, cognitive and personality variables to the discrepancy between self and clinician rated depression. *J Affect Disord*. (2010) 124:202–6. doi: 10.1016/j.jad.2009.11.011
39. Sumiyoshi T, Morio Y, Kawashima T, Tachimori H, Hongo S, Kishimoto T, et al. Feasibility of remote interviews in assessing disease severity in patients with major depressive disorder: A pilot study. *Neuropsychopharmacol Rep*. (2024) 44(1):149–57. doi: 10.1002/npr.212411
40. Simmons WK, Burrows K, Avery JA, Kerr KL, Bodurka J, Savage CR, et al. Depression-related increases and decreases in appetite: dissociable patterns of aberrant activity in reward and interoceptive neurocircuitry. *Am J Psychiatry*. (2016) 173:418–28. doi: 10.1176/appi.ajp.2015.15020162
41. Sajatovic M, Gaur R, Tatsuoka C, De Santi S, Lee N, Laredo J, et al. Rater training for a multi-site, international clinical trial: what mood symptoms may be most difficult to rate? *Psychopharmacol Bull*. (2011) 44:5–14.
42. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. (1960) 23:56–62. doi: 10.1136/jnnp.23.1.56
43. Gaur R, Sajatovic M, Lee N, Nath G, Ramirez L, Kaviani H, et al. *Rater Training on HAMD, MADRS and YMRS – what were the difficult items to rate? Poster session presented at New Clinical Drug Evaluation Unit (NCDEU) 50th Anniversary Meeting*, Boca Raton, Florida, June 14–17, 2010.
44. Borentain S, Gogate J, Williamson D, Carmody T, Trivedi M, Jamieson C, et al. Montgomery-Asberg Depression Rating Scale factors in treatment-resistant depression at onset of treatment: Derivation, replication, and change over time during treatment with esketamine. *Int J Methods Psychiatr Res*. (2022) 31:e1927. doi: 10.1002/mpr.1927
45. Harrington MO, Nedberge KM, Durrant SJ. The effect of sleep deprivation on emotional memory consolidation in participants reporting depressive symptoms. *Neurobiol Learn Mem*. (2018) 152:10–9. doi: 10.1016/j.nlm.2018.04.013
46. Johnson K, Devine J, Ho K, Howard K, Saretsky T, Jamieson C, et al. Evidence to support Montgomery-Asberg depression rating scale administration every 24 hours to assess rapid onset of treatment response. *J Clin Psychiatry*. (2016) 77:1681–6. doi: 10.4088/JCP.15m10253
47. Yavorsky C, Singh J, Engelhardt N. Can the MADRS measure rapid changes in depressive symptoms in response to esketamine treatment? *Neuropsychopharmacol*. (2017) 42:S476–652. doi: 10.3389/fpsy.2023.1135828