REVIEW ARTICLE

Baseline reliability and early response in clinical trials of major depressive disorder

Author
Steven D. Targum
Scientific Director, Signant Health (Boston, MA)
Email: sdtargum@yahoo.com

Abstract

Many double-blind, placebo-controlled, antidepressant clinical trials conducted in patients with major depressive disorder (MDD) fail to separate the candidate drug from placebo. Multiple factors can affect trial outcomes including the patient's expectations and motivation, the clinician's ratings competency and reliability, and the design of the trial itself. Two factors affecting treatment outcome associated with trial design are 1) the reliability of the baseline measure, and 2) an early, indiscriminate symptomatic response following randomization that can occur regardless of treatment assignment.

The motivation to participate in the trial itself can influence the baseline measurement as well as the early symptomatic response that follows the baseline. An unreliable baseline measure will affect all subsequent symptomatic assessments during the clinical trial and may affect the interpretation of results.

The baseline measure is usually the primary contingency variable used to evaluate treatment outcome in clinical trials and is typically a single point in time measurement obtained sometime on the baseline day. It is well known that the symptoms of MDD naturally fluctuate during the day (diurnal variation) and from day to day as well. Consequently, it is unrealistic to presume that a single point in time measurement can accurately and reliably capture the true symptom severity of all MDD patients at baseline.

Several MDD trials have revealed an early symptomatic response that occurs shortly after randomization regardless of treatment assignment. It has been shown that the early symptomatic response may be sustained throughout the trial, includes patients assigned to placebo, and may impede signal detection at the end of the study.

This review explores the importance of baseline reliability and the influence of early symptomatic response in some clinical trials of MDD patients and considers an alternative assessment method (ecological momentary assessment) as an innovative strategy to improve the reliability of the baseline measurement.

Keywords: Major depressive disorder; baseline reliability; early symptomatic response; ecological momentary assessment

1.0 Introduction

Major depressive disorder (MDD) is a highly prevalent, often under-treated illness that affects nearly 340 million people worldwide and is the second leading cause of disability in the world (1-3). In the United States, it is estimated that the social, economic, and medical costs of MDD are between $17-$44 billion and result in over 200 million lost workdays each year (1). Although many antidepressant treatments (ADT) have been approved, only 30-40% of MDD patients achieve remission after their first antidepressant treatment, and some patients do not respond at all (4). Therefore, there is still a critical need for more effective antidepressant drugs for MDD.

The route to new drug approval is achieved through an extensive clinical trial process that necessarily requires the demonstration of both drug efficacy and safety. Many of the double-blind, placebo-controlled, ADT clinical trials conducted in patients with MDD fail to separate the candidate drug from placebo. Clinical trials are experiments that require voluntary patient consent and cooperation with rigorously defined study procedures and time-consuming study designs that include placebo-control groups. The informed consent form provides details about the chances of getting placebo and a list of possible adverse events that may occur as well. It is understandable that many depressed patients are not willing to participate in a clinical trial and that the patients who do consent may represent a unique population whose hopes and expectations may influence their treatment response (5-6). In fact, some authors argue that the mere participation in a clinical trial can generate a placebo response that can subsume a large part of the overall response (7-11).

The baseline measurement is a fundamental key to clinical trial success because it is the primary contingency variable for all subsequent measurements. The assessment is generally obtained as a single point in time symptomatic measurement administered sometime on the baseline day. Although assessed at a single point in time, the instruments typically used are the Hamilton rating scale for depression (HamD$_{17}$) or Montgomery-Asberg depression rating scale (MADRS) that inquire about the presence and severity of symptoms in the past week (12-13). The reliability of the baseline measurement has been challenged by several investigators (5-6,14-16). Symptom fluctuation is one reason for baseline unreliability. It is well known that the mood symptoms of MDD patients can fluctuate during the day (diurnal variation) as well as from day to day (17-19). Consequently, it is unrealistic to presume that a single point in time measurement obtained on the baseline day can reliably capture the true clinical status of all MDD patients entering clinical trials. Obviously, a reliable baseline is critical to the reliability of all subsequent assessments of symptomatic change (6, 14, 20).

It is not uncommon that a moderate, early symptomatic improvement will occur within the first week or two following the baseline day regardless of the randomized treatment assignment (14, 16, 21-29). Expectation biases, inflated scores, and other non-specific factors related to study participation may influence this early response (5-7, 11, 28). The early symptomatic response is often sustained throughout the duration of the clinical trial regardless of whether patients are assigned to the drug candidate or placebo (14, 16, 23, 28). Nearly 50% of placebo controlled MDD drug studies fail to separate drug from placebo due, in part to a higher than anticipated placebo response at the end of the study (22-23, 30-32).

This brief review will examine the effects of symptom fluctuation on baseline reliability, explore the influence of early symptomatic improvement regardless of treatment assignment on treatment outcome, and consider an alternative assessment method (ecological momentary assessment) as an innovative strategy to improve the reliability of the baseline measurement to optimize signal detection.

2.0 The enigma of baseline reliability

Clinical investigators have long understood the importance of a reliable baseline measurement to achieve meaningful clinical trial outcomes. Among the multiple factors that influence trial outcomes are the patient's expectations and motivation, and the clinician's ratings competency and reliability. The inclusion criteria for most clinical trials require that a potential study candidate meet a minimum symptom severity threshold criterion at the screen and baseline visits in order to qualify for randomization into the clinical trial. Awareness of the severity criterion may influence the endorsement and ratings of symptoms by both the patients seeking enrollment and by the clinicians who rate the potential study candidate. There may be a temptation to exaggerate symptoms and inflate the baseline score to meet the eligibility threshold. The consequence of baseline score inflation is an unreliable baseline measurement that can affect all subsequent assessments.

Ratings reliability is another factor that may influence every assessment in a clinical trial, and particularly the baseline severity score. The symptomatic questionnaires used in MDD trials include many subjective items that require personal perspective that can differ from individual to individual. In fact, the scores recorded by two trained clinician raters about the same patient scored at the same time may be discordant and can often differ from the patient's self-ratings (29, 33). It is not possible to affirm the accuracy of any singular measure as a "true" reflection of a patient's current clinical condition. However, it is possible to affirm the reliability of paired ratings made by two different raters of the same assessment are reasonably close. The recent availability of remote surveillance methods that offer site-independent scoring confirmation of recorded site-based interviews has increased confidence about ratings reliability at all study visits (29). Several studies have demonstrated a high inter-rater scoring correlation on depression rating instruments between paired clinician-clinician ratings and between paired clinician-patient self-ratings (29, 33).

2.1 Baseline reliability and symptom fluctuation

Another factor affecting the reliability of the baseline measurement is symptom fluctuation, an inherent clinical characteristic observed in many patients with MDD (17, 19, 34). It is well known that the severity of the symptoms associated with major depressive disorder may fluctuate during the day (diurnal variation) and change from day to day as well (17-18). Daily symptom fluctuation may affect the stability and reliability of the baseline measure.

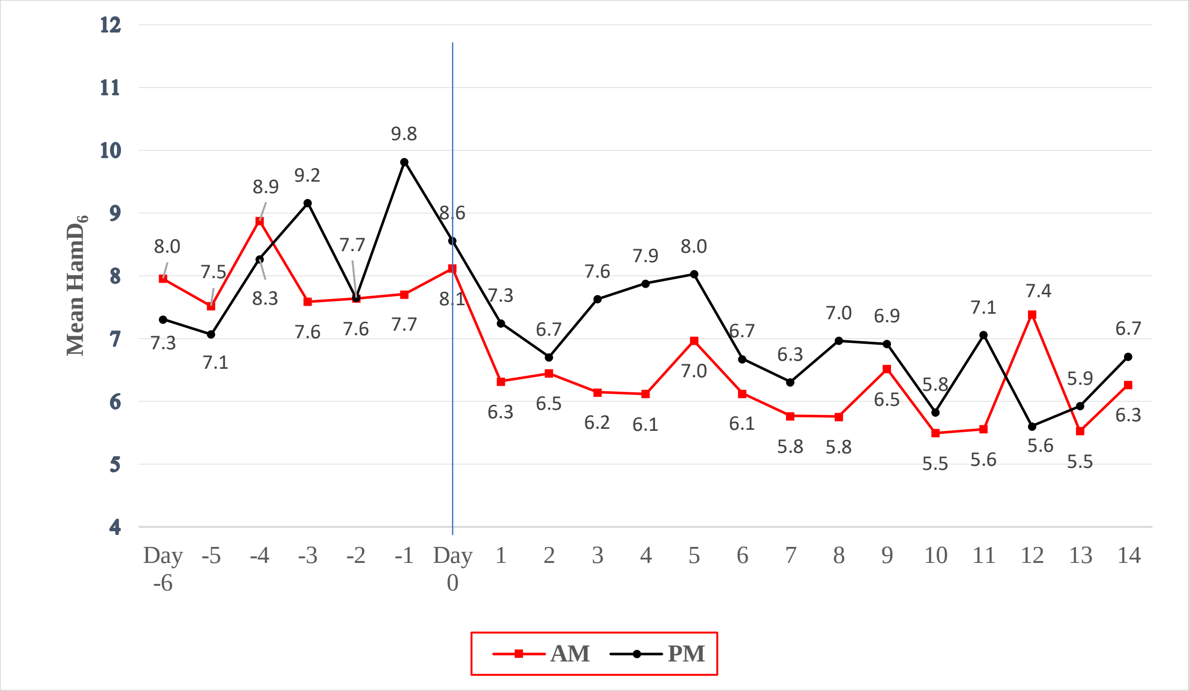
Evans and colleagues (15) suggested that an unreliable baseline measure is a possible source of unsuccessful trial outcomes. They examined placebo response and treatment outcome relative to the extent of symptomatic change (fluctuation) that occurred between the screen and baseline visits on the total score of the HamD$_{17}$. They merged data from four different 6-8 week, randomized, double-blind, placebo-controlled ADT studies of MDD patients (n=557 patients). The study inclusion criteria for each of these 4 ADT studies differed from most MDD trials in that all eligible patients required a minimum score of 22 on the

$HamD_{17}$ at the screen visit but had no symptom severity threshold criterion at the baseline visit. Thus, severity scores could improve but the patient was still eligible to enter the study. The analysis showed a mean change in the total $HamD_{17}$ score between the screen and baseline visits of only 1 point in the combined study population, but that individual patient total $HamD_{17}$ scores changed widely from 13 points of improvement to 8 points of worsening. Furthermore, many patients revealed substantial individual HamD item score volatility (symptom fluctuation) as well. The authors' found that the patients who had total $HamD_{17}$ score improvement or worsening of 2 points or greater between screen and baseline had higher placebo responses and poorer drug–placebo separation at the end of the study than the patients who had less score fluctuation between the screen and baseline visits (15). These findings suggest that symptom fluctuation between the screen and baseline visits could affect treatment outcome and reinforce the concern about the reliability of the baseline measure.

The challenge of establishing a reliable baseline measure is compounded by the conventional method of measurement that relies on a single point-in time assessment at baseline. As shown in the analyses of Evans and colleagues (15), changes of 2 points or greater between screen and baseline on the $HamD_{17}$ can impede signal detection. The assessment method in their 4 studies as well as the assessments used in most clinical trials is based upon the patient's recall of their symptoms over the past week. It is probably unrealistic to expect that depressed patients will have an accurate retrospective recall of their symptoms over the past week (35). Further, the diurnal variation of depressive symptoms affects the reliability of any single point in time measurement. An alternative strategy for measurement is offered by the availability of ecological momentary assessment (EMA) tools (36-38).

EMA (also called experience sampling methods) was introduced as a method to sample the daily life experience of patients in real-time (18, 36-38). We used EMA via smartphone to assess patient self-ratings of the $HamD_6$, a subscale of the $HamD_{17}$ in depressed patients meeting criteria for MDD (39-40). Patients were asked to assess their symptoms twice daily during the week prior to an open-label initiation of antidepressant treatment and on each subsequent treatment day. Figure 1 displays the mean patient rated $HamD_6$ scores for 34 treated patients from day -6 through the first 14 days of treatment. As shown, the mean $HamD_6$ scores fluctuated from AM to PM as well as from day to day. The mean AM and PM baseline values differed from the mean $HamD_6$ scores on day -1 and day 1. Some patients revealed marked symptom fluctuation from assessment to assessment. These findings demonstrate the presence of symptom fluctuation as an inherent clinical characteristic of MDD as described in the DSM-5 (17).

Figure 1: Daily symptom fluctuation recorded in patients with major depressive disorder



NOTE: Displays patient self-ratings of the 6 item Hamilton rating scale for depression (HamD$_6$) obtained via smartphone twice daily before and after initiation of open-label antidepressant on day 0 (n=34)

Symptom fluctuation is particularly important for studies of rapidly acting antidepressants when an endpoint may be 24 hours. In one recent ADT study of MDD patients who had not responded to ongoing ADT, the MADRS was administered on a daily basis to hospitalized MDD patients before and after randomization by remote raters who were blind to the visit day, timing of randomization, or any treatment emergent adverse events (41). The primary study endpoint for this study was the MADRS at 24 hours after randomization. The study plan was to randomize only those MDD patients who demonstrated relative symptom score stability (<25% total score fluctuation) and maintained a total MADRS severity score of at least 21 for the 4 days preceding the baseline visit. The findings showed that the total MADRS score fluctuated daily in many patients. In this study, 31 of 73 potential study candidates continued to meet eligibility criteria and were ultimately randomized at baseline. Among the screen failures, 14 potential study candidates were excluded because of marked symptom fluctuation or MADRS scores that fell below 21 prior to randomization.

It is clear that the inherent clinical characteristic of daily symptom fluctuation experienced by many MDD patients challenges the stability and consequently the reliability of a single point-in-time baseline measurement.

2.2 The effect of early symptomatic response on treatment outcome

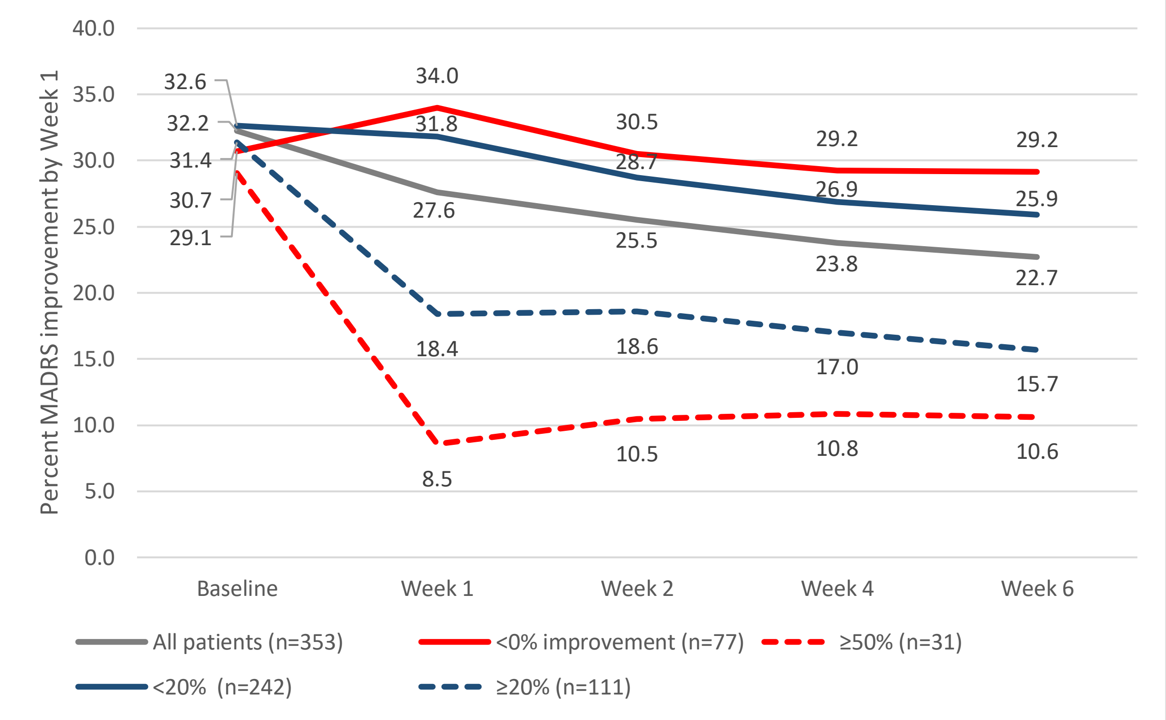
As noted above, patient and clinician motivations may affect the reliability of the baseline measurement and generate score inflation in some instances. Therefore, it is not surprising that many depressed patients show substantial symptomatic improvement immediately following the randomization (baseline) visit regardless of treatment assignment (14, 16, 21-29). Baseline score inflation followed by indiscriminate symptomatic improvement regardless of treatment assignment will increase the placebo response and impede signal detection at the end of the study.

In a meta-analysis of over 6500 MDD patients, Szegedi and colleagues (26) reported that early symptomatic improvement within the first 2 weeks following randomization was a predictor of treatment outcome. Similarly, in an analysis of 8 double-blind MDD trials, Altin and colleagues (42) reported that a ≥20% improvement of the total $HamD_{17}$ score within 2 weeks post-randomization yielded higher response and remission rates at the end of the study in both the duloxetine and placebo treated groups than in the patients who had <20% $HamD_{17}$ score improvement.

In a study of MDD patients who had an inadequate response to their ongoing ADT, Targum (28) found an early symptomatic response of ≥20% improvement one week after randomization in 111 of 353 randomized patients (31.4%) regardless of treatment assignment on the total MADRS score. As shown in Figure 2, a higher percentage of MADRS score improvement within the first week following double-blind randomization was sustained at the study endpoint (week 6). Early response yielded significantly higher response and remission rates at the study endpoint than patients who had less improvement in the first week regardless of treatment assignment (28). The 77 patients who had symptomatic worsening at week 1 (<0% MADRS score improvement from baseline) had significantly less response and remission than the other 276 patients in the study ($p < 0.0001$ for both). Similarly, patients with <20% MADRS score improvement at week 1 achieved a 21.9% response and 12.8% remission rate at week 6 in contrast to 55.0% and 42.3% in patients who had ≥20% improvement at week 1 (both $p < 0.0001$). Finally, 24 of the 31 patients (74.4%) who achieved a conventionally defined full response by week 1 (≥50% MADRS improvement) sustained the response at week 6 in contrast to 46 of 322 patients (39.4%) with <50% improvement at week 1 who eventually became responders (X^2= 66.98; $p < 0.0001$), and 22 early responding subjects (71.0%) achieved remission by week 6 in contrast to 56 patients (17.4%) with <50% improvement at week 1 (X^2= 44.09; $p < 0.0001$).

In another MDD study that used patient self-ratings, Targum and Catania (29) reported that early improvement on the Quick Inventory of Depressive Symptoms ($QIDS-SR_{16}$) scored 2 weeks after randomization was associated with significantly higher response rates regardless of treatment assignment at the end of a 6-week double-blind, placebo-controlled clinical trial. In this study, 26% of the acutely depressed MDD patients met criteria for a full treatment response (≥50% improvement from baseline) on the QIDS- SR_{16} at week 2 regardless of treatment assignment. By week 6, 87.5% of these early $QIDS-SR_{16}$ responders were still responders. A post-hoc analysis that excluded the patients who had demonstrated >50% $QIDS-SR_{16}$ improvement from the baseline score during the first 2 weeks enhanced the effect size favouring the experimental treatment from 0.33 to 0.64 for the $QIDS-SR_{16}$ despite reducing the sample size (29).

Figure 2: Effect of early symptomatic response on eventual total MADRS score (includes all patients regardless of treatment assignment)



Adapted from Targum, J Psychiatric Research 95: 276–281.2017

2.3 The effect of placebo lead-in designs on trial outcome

The concern about possible symptom exaggeration and/or score inflation affecting the baseline score contributed to the application of single-blind placebo lead-in designs that sought to identify and exclude patients who exaggerated their symptoms or whose severity scores simply improved (fell below the minimum eligibility threshold) at the baseline visit. In the single-blind placebo lead-in design, the patient is unaware (blind) to the existence of the placebo lead-in period or the timing of the true randomization visit. Unfortunately, single-blind placebo lead-in designs have not resolved the issue of high placebo response rates. Trivedi and Rush (43) conducted metanalyses of 101 depression studies that had been done at the time of their review and concluded that a single-blind placebo lead-in design did not lower the placebo response, did not improve drug response, and did not improve signal detection.

One limitation of the single-blind design is that is does not control for clinician rater biases or their possible motivation to inflate their clinician rated scores to enroll a patient. Consequently, some ADT clinical trials have employed a double-blind placebo lead-in design to ensure that neither the patient nor the clinician rater knows the true timing of the randomization visit.

Faries and colleagues (14) compared the single-blind and double-blind paradigms and found that the double-blind, placebo lead-in design generated 5 times as many early responders as the single-blind placebo lead-in design. In 2 large MDD studies that used a 1-week single-blind, placebo lead-in design, 33 of 627 MDD patients (5.3%) improved by 25% or more on the total score of the $HamD_{17}$ during the lead-in period (14). The investigators conducted 2 additional MDD studies that employed a double-blind variable placebo lead-in design that hid the true randomization visit from everyone. In this double-blind variable design, approximately 50% of the eligible patients were randomized to drug or placebo at the baseline visit (week 0) and the other 50% were given double-blind placebo for 1 week prior to a randomization visit at week 1. Both patients and clinical trial staff were blinded to the placebo lead-in design and the variable timing of the randomization visits. To maintain the blind after randomization, all of the patients continued in the study regardless of their early responses. In these two double-blind placebo lead-in studies, 36 of the 128 MDD patients (28.1%) met the authors' criterion for early response (≥25 % improvement from baseline) during the 1-week double-blind placebo lead-in period in contrast to only 5.3% as noted in the abovementioned single-blind studies conducted by the same investigators. It is noteworthy that 22 patients in the double-blind placebo lead-in design (13.9%) improved by ≥50% from baseline (the conventionally defined placebo response) during 1-week placebo lead-in period (14).

Faries and colleagues (14) reported that the double-blind placebo lead-in responders sustained their early response at the study endpoint. After 8 weeks of continued double-blind placebo treatment, the early placebo lead-in responders had a significantly lower endpoint total $HamD_{17}$ score than the patients who were not placebo lead-in responders ($HamD_{17}$ = 7.6 vs. 16.0; p = 0.001). The sustained placebo response that followed the early response affected signal detection at the end of the study. As part of their prospective analysis, these investigators removed the ≥25 % improvement placebo lead-in responders and found an increase in the mean endpoint $HamD_{17}$ placebo group severity scores. This adjustment for early responders resulted in an increased drug–placebo treatment difference in one of the two studies (14).

The 2 double-blind placebo lead-in studies conducted by Faries and colleagues (14) placed only half of the enrolled subjects in the 1-week placebo lead-in period. In a more recent study, Targum and colleagues (16) included all of the eligible depressed patients in a two-week double-blind placebo lead-in period that preceded a 6-week ADT trial of MDD patients who had an inadequate response to their ongoing ADT. Similar to Faries and colleagues, both the patients and clinician raters were blinded to the 2-week placebo-lead-in period and presumed that the trial was a conventional 8-week double-blind, placebo-controlled study. The study included assessments of both the MADRS and the $HamD_{17}$ scales. At the end of the 2-week double-blind placebo period, 14.9% of the depressed patients were MADRS placebo responders (≥50% total score improvement from baseline) and 12.2% were $HamD_{17}$ placebo responders. These results are very similar to the 13.9% $HamD_{17}$ placebo response rate reported by Faries and colleagues during the 1-week double blind placebo lead-in phase of their studies (14).

Targum et al (16) examined the effect of the 2-week double-blind placebo lead-in period on the eventual treatment outcome after 6 weeks of randomized, double-blind treatment. There were 227 patients who met week 2 randomization criteria and were allocated to the study drug (n = 117) or placebo (n = 110) in addition to their ongoing ADT. To maintain the blind, the 109 non-evaluable subjects who failed to meet the randomization criteria at week 2 continued on placebo for the full 8-week double-blind treatment. The response trajectory of the study drug treated group, the evaluable, and the non-evaluable placebo-assigned subjects was examined. Similar to other studies, most of the double-blind placebo lead-in responders sustained their response to the 8-week study endpoint. Over 80% of the subjects who had ≥50% $HamD_{17}$ or MADRS score improvement during the 2-week double-blind placebo lead-in period (weeks 0–2) sustained that response at week 8. For instance, 41 of the 50 patients (82.0%) who were MADRS placebo responders at week 2 sustained the placebo response by week 8 in contrast to 17 of 94 subjects (23.0%) with<20% MADRS score change between weeks 0–2 ($X^2 = 52.8$; df = 1; $p < 0.0001$). Further, placebo-assigned subjects with ≥20% but <50% $HamD_{17}$ or MADRS total score fluctuation (improvement or worsening) during the double-blind placebo lead-in period had significantly higher rates of placebo response and remission at week 8 compared to patients with <20% score changes during that period as well (16). Hence, the early score improvement prior to the true randomization visit at week 2 had a marked effect on the endpoint.

This study failed to separate the experimental drug treatment from placebo in its primary analysis. However, a post-hoc analysis that excluded the patients who had ≥20% score change during the double-blind placebo lead-in period enhanced the effect size (ES) favoring the experimental treatment over placebo on both the MADRS and $HamD_{17}$ sub-group post-hoc analyses relative to the ITT population. For instance, in the MADRS sub-group analysis, the ES improved from 0.125 in the ITT group to 0.404, and the experimental treatment now revealed a statistically significant benefit over placebo ($F = 6.39$; df = 1; $p = 0.012$).

The early symptomatic response as reported in these studies are consistent with the reports of other investigators and serve to document an inherent phenomena of early response seen in randomized, placebo-controlled clinical trials with MDD patients (14, 22-25, 28-29, 42).

2.4 Ecological momentary assessment

The challenge of obtaining meaningful clinical trial outcomes is compounded by the conventional clinician rated method of measurement that relies on a single point-in time assessment at baseline. An alternative strategy for symptom severity measurement is offered by the availability of ecological momentary assessment (EMA) tools. EMA (also called experience sampling methods) have been introduced as methods to sample the daily life experience of patients in real-time (18, 36-38). EMA can assess symptoms, activity, cognitive functioning, and biology in the moment as frequently during the day as desired and obviates any concerns about retrospective recall bias (37). EMA may be a better approach to establishing a reliable baseline because it is based on recency of experience rather than a one week recall of symptoms. Clearly, retrospective recall of symptoms or recent behavior may be inaccurate because memories may be affected by immediate concurrent events or poorly recalled, particularly in depressed patients (35). Further, the diurnal variation of depressive symptoms may affect the reliability of any single point in time measurement. The collection and averaging of multiple measures over a shorter time frame may yield a more stable, clinically reliable baseline score.

We conducted a small pilot study to examine the utility of EMA to track mood symptoms in a clinical trial in acutely depressed MDD patients (40). In this small study, patient self-ratings of the $HamD_6$ by EMA were highly correlated with the corresponding clinician ratings of the $HamD_6$ done at screen, baseline, and weeks 2, 4, and 6 of treatment.

Further, daily patient self-rated EMA $HamD_6$ scores done between the bi-weekly clinic visits identified some of the treatment responders 1-2 weeks before the clinician ratings caught up. These preliminary findings are interesting but require additional, much larger studies to explore the usefulness of EMA in clinical trials.

EMA offers the opportunity to collect frequent patient self-ratings of their own symptoms. Patient rated outcomes are important in MDD. Depression is primarily a subjective experience such that the assessment of patient-perceived depressive symptoms may contribute more than clinician rated measures to predict pharmacological treatment outcome and offer important clinical information not accessible through the conventional clinician rating scales. The patient's self-assessment of his or her own depressive symptoms may be an essential companion to traditional clinician ratings to fully evaluate treatment outcomes in clinical trials.

3.0 Conclusion

This review has explored the importance of baseline reliability and examined the influence of early symptomatic response on treatment outcome from some recent clinical trials of MDD patients. These different studies consistently reflect the challenge of establishing a reliable baseline measure and the consequences of early, indiscriminate response on treatment outcome. Meaningful clinical trial outcomes may be impeded by the conventional method of clinician rated measurements that rely on a single point-in time assessment at baseline. It is suggested that ecological momentary assessment tools may offer a better alternative approach to establishing a reliable baseline because it is based on the recency of experience rather than a recall of symptoms and may in conjunction with clinical tools improve the precision of symptomatic measurement during clinical trials.

Declaration of competing interests

Steven D. Targum is an employee of Signant Health and has received vendor grants or consulted with Acadia Pharmaceuticals, Alkermes Inc., BioXcel, EMA Wellness LLC., Epiodyne, Functional Neuromodulation, Intra Cellular Therapies, Johnson and Johnson PRD, Karuna Therapeutics, Merck Inc., Methylation Sciences Inc., Navitor Pharmaceuticals Inc., Neurocrine Inc., and Sunovion Inc. during the past 3 years.

References

1. Stewart WF, Ricci JA, Chee E, Morganstein D. Lost productive work time costs from health conditions in the United States: results from the American productivity audit. J Occup Environ Med 2003, 45(12):1234-46.

2. Kessler RC, Demler O, Frank RG, Olfson M, Pincus HA, Walters EE, Wang P, Wells KB, Zaslavsky AM. Prevalence and Treatment of Mental Disorders, 1990 to 2003. N Engl J Med 2005, 352:2515-2523.

3. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, Ferrari AJ, Hyman S, Laxminarayan R, Levin C, Lund C, Medina Mora ME, Petersen I, Scott, J, Shidhaye R, Vijayakumar L, Thornicroft G, Whiteford H and Group, DMA. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. Lancet 2016, 387: 1672-85.

4. Cipriani A, Furukawa T., Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, Leucht S, Ruhe H G, Turner EH, Higgins, JPT, Egger M, Takeshima N, Hayasaka Y, Imai H, Shinohara K, Tajika A, Ioannidis JPA, Geddes JR. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018, 391, 1357-1366.

5. Fava, M., Evins, A., Dorer, D., Schoenfeld, D. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother. Psychosom. 2003, 72 (3): 115-127.

6. Targum, S.D., Pollack, M.H., Fava, M., Re-defining affective disorders: relevance for drug development. CNS Neurosci. Ther. 2008, 14: 2-9.

7. Lambert MJ. Handbook of Psychotherapy Integration. Basic Books, New York,
1992, pp. 94e129.

8. Kirsch I, Deacon BJ, Huedo-Medina, TB, Scoboria A, Moore TJ, Johnson BT.
Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med. 2008, 5, e45. http://dx.doi.org/10.1371/journal.pmed.0050045.

9. Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, et al. Components of the placebo effect: a randomized controlled trial in irritable bowel syndrome. BMJ 2008, 336: 999–1003.

10. Kaptchuk TJ, Friedlander E, Kelley JM, Sanchez MN, Kokkotou E, et al. Placebos without deception: a randomized controlled trial in irritable bowel syndrome. PLoS One 2010; 5(12):e15591.
http://dx.doi.org/10.1371/journal.pone.0015591.

11. Kirsch I. Antidepressants and the Placebo Effect. Zeitschrift fur Psychologie
2014, 222(3):128–34.

12. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960, 23:56-62.

13. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Brit J Psychiat., 1979, 134:382-389.

14. Faries DF, Heiligenstein JH, Tollefson GD, Potter WZ. The double blind variable placebo lead-in period: results from two antidepressant clinical trials. J. Clin. Psychopharmacol. 2001, 6: 561-568.

15. Evans KR, Sills T, Wunderlich GR, McDonald HP. Worsening of depressive symptoms prior to randomization in clinical trials: a possible screen for placebo responders?
J Psychiatric Res., 2004, 38:437–44.

16. Targum SD, Cameron BR, Ferreira L, MacDonald ID. Early score fluctuation and placebo response in a study of major depressive disorder. J Psychiatr. Res. 2020, 121: 118-125.

17. American Psychiatric Association. Diagnostic and Statistical Manual of Mental

Disorders, Fifth edition, 2013. Arlington VA: American Psychiatric Press.

18. Peeters F, Berkhof J, Delespaul P, Rottenberg J, Nicolson NA. Diurnal mood variation in major depressive disorder. Emotion 2006, 6 (3): 383-391.

19. Morris DW, Rush AJ, Jain S, Fava M, Wisniewski SR, Balasubramani GK, Khan AY, Trivedi MH. Diurnal mood variation in outpatients with major depressive disorder: implications for DSM-V from an analysis of the Sequenced Treatment Alternatives to Relieve Depression Study data. J Clin Psychiatry 2007, 68 (9): 1339-1347.

20. Kobak KA, Kane JM, Thase, ME, Nierenberg AA. Why do clinical trials Fail? the problem of measurement error in clinical trials: time to test new paradigms? J. Clin. Psychopharmacol. 2007, 27 (1): 1-5.

21. Quitkin FM, Rabkin JG, McGrath PJ, Stewart JW, Harrison W, Ross DC, et al. Heterogeneity of clinical response during placebo treatment. Am J Psychiatry 1991, 148:193–196.

22. Khan A, Cohen S, Dager S, Avery, DH, Dunner DL. Onset of response in relation to outcome in depressed outpatients with placebo and imipramine. J. Affect. Disord. 1989, 17 (1): 33-38.

23. Khan A, Detke M, Khan SR, Mallinckrodt C. Placebo response and antidepressant clinical trial outcome. J. Nerv. Ment. Dis. 2003, 191: 211-218.

24. Cusin C, Fava M, Amsterdam JD, Quitkin FM, Reimherr FW, Beasley Jr CM, Rosenbaum JF, Perlis RH, Early symptomatic worsening during treatment with fluoxetine in major depressive disorder: prevalence and implications. J. Clin. Psychiatry 2007, 68 (1): 52-57.

25. Katz, MM, Meyers AL., Prakash A, Gaynor PJ, Houston JP. Early symptom change: prediction of remission in depression treatment. Psychopharmacol.
Bull. 2009, 42: 94-107.

26. Szegedi, A., Jansen, W.T., van Willigenburg, A.P., van der Meulen, E., Stassen, H.H.,
Thase, M.E., Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6562 patients. J. Clin. Psychiatry 2009, 70 (3): 344–353.

27. Thase ME, Larsen KG, Kennedy SH. Assessing the 'true' effect of active antidepressant therapy v. placebo in major depressive disorder. Br. J. Psychiatry 2011, 199: 501-507.

28. Targum, S.D., Early symptomatic improvement affects treatment outcome in a study of major depressive disorder. J. Psychiatr. Res. 2017, 95: 276–281.

29. Targum, S.D., Catania, C.J., Early treatment response affects signal detection in a placebo-controlled depression study. Pers. Med. Psychiatry 2017, 4–6: 19–24.

30. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression variable, substantial, and growing. JAMA 2002, 287 (14): 1840-1847.

31. Khan A, Leventhal RM, Khan, SR, Brown, WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. J. Clin. Psychopharmacol. 2002, 22 (1): 40–45.

32. Iovieno N, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. J. Clin. Psychiatry 2012, 73 (10), 1300–1306.

33. Targum SD, Wedel PC, Bleicher LS, Busner J, Daniel DS, Robinson J, Rauh P, Barlow C. A comparative analysis of centralized, site-based, and patient ratings in a clinical trial of Major Depressive Disorder. J. Psychiatr. Res. 2013, 47: 944-954.

34. Wirz-Justice A. Diurnal variation of depressive symptoms. Dialogues in Clin Neuroscience 2008, 10 (3): 337-353.

35. Ben-Zeev D, Young MA, Madsen JW. Retrospective recall of affect in clinically depressed individuals and controls. Cognition and Emotion 2009, 23: 1021-1040.

36. Armey MF, Schatten HT, Haradhvala N, Miller IW. Ecological momentary assessment (EMA) of depression-related phenomena. Current Opin Psychol. 2015, 1 (4): 21-25.

37. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess. 2009, 21 (4): 463-475.

38. Aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. Clin Psychol. Review. 2012, 32 (6): 510-523.

39. Bech P. Rating scales in depression: limitations and pitfalls. Dialogues in Clin Neuroscience 2006, 8 (2): 207-215.

40. Targum SD, Sauder C, Evan M, Saber JN, Harvey PD. Ecological momentary assessment as a measurement tool in depression trials. Submitted for publication.

41. Targum SD, Leventer S, Hughes TE, Owen JR, Vlasuk GP. NV-5138 a Novel, Direct Activator of the Mechanistic Target of Rapamycin Complex 1 (mTORC1): A Phase 1b Randomized, Double-Blind, Placebo-Controlled Single Oral Dose Study in Subjects with Treatment-Resistant Depression (TRD). 2019: December 9; Presented at ACNP, Orlando Florida.

42. Altin M, Harada E, Schacht A, Berggren L, Walker D, Dueñas H. Does early improvement in anxiety symptoms in patients with major depressive disorder affect remission rates? a post-hoc analysis of pooled duloxetine clinic trials. Open J Depression 2014, 3:112–23.

43. Trivedi M, Rush J. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? Neuropsychopharmacology 1995, 11: 33-43.