Journal of Psychiatric Research 121 (2020) 118–125

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Early score fluctuation and placebo response in a study of major depressive disorder

Steven D. Targum[a,b,*], Beth R. Cameron[a], Ludvina Ferreira[a,c], I. David MacDonald[a]

[a] Methylation Sciences Inc, 15300 Croydon Drive, Suite 300, Surrey, BC, V3S 0Z5, USA
[b] Signant Health, Boston, MA, USA
[c] PRA Health Sciences, Vancouver, DC, USA

ARTICLE INFO

Keywords:
Placebo response
Major depressive disorder
Augmentation treatment
S-adenosylmethionine
Score fluctuation
Post-hoc analyses

ABSTRACT

Early score fluctuation in double-blind, placebo-controlled studies may affect the reliability of the baseline measurement and adversely affect the eventual study outcome. We examined the effect of early score fluctuation during a 2-week double-blind placebo lead-in period in a phase II, double-blind, placebo-controlled trial of adjunctive s-adenosyl methionine (MSI-195) in MDD subjects who had had an inadequate response to ongoing antidepressant treatment. The overall study failed to meet its specified endpoints.

We examined the score trajectories of all placebo-assigned subjects during the double-blind placebo lead-in period and subsequent 6-week treatment period. Placebo-assigned subjects with ≥20% $HamD_{17}$ or MADRS score fluctuations (improvement or worsening) during the double-blind placebo lead-in period (prior to randomization) had significantly higher rates of placebo response and remission at week 8 compared to subjects with < 20% response. A post-hoc analysis of evaluable subjects taken from the ITT population that excluded subjects with ≥20% early score response yielded higher effect sizes for both the $HamD_{17}$ and MADRS sub-groups and statistical significance for MSI-195 over placebo in the MADRS sub-group ($p = 0.012$) with an effect size of 0.404.

A reliable baseline measure is an asset for signal detection. These post-hoc findings suggest that study designs that anticipate and attempt to manage early response prior to randomization may yield more meaningful outcome data for trials of MDD and possibly other disorders as well.

1. Introduction

Signal detection in randomized, placebo-controlled clinical trials relies on the statistically significant differentiation of the candidate drug from placebo as a demonstration of a meaningful treatment effect. Placebo response rates in antidepressant (ADT) trials for Major Depressive Disorder (MDD) are often within the range of 35–40% (Furukawa et al., 2016). Approximately 50% of placebo-controlled trials of MDD fail, and it has been shown that placebo response rates > 40% generally fail to achieve signal detection (Laughren, 2001; Khan et al., 2002a,b, 2007; Iovieno and Papakostas, 2012). Many recent MDD trials yield higher than anticipated placebo response rates despite efforts to minimize it with restrictive eligibility criteria, site-independent review of subject selection, and innovative study designs (Khan et al., 2002a,b; Fava et al., 2003; Lee et al., 2004; Zimmerman et al., 2005; Kobak et al., 2007; Kirsch et al., 2008; Targum et al., 2008; Khin et al., 2011; Thase et al., 2011; Rutherford and Roose, 2013). In one recent analysis of over 100 depression trials, Whitlock and colleagues suggested that such efforts do not increase the treatment effect because they affect the treatment arm as well (Whitlock et al., 2019).

The experimental conditions of the clinical trial itself may facilitate a placebo response in many subjects (Fava et al., 2003; Kirsch et al., 2008; Kirsch, 2014; Targum et al., 2008; Thase et al., 2011). For instance, the informed consent process differentiates willing from unwilling subjects, offers special attention to their problems, and can foster expectation biases about treatment outcome. Subjects entering a MDD clinical trial must identify themselves as depressed at the screening visit and meet specific symptom severity thresholds to qualify for study participation. However, in the natural course of an acute major depressive episode, the magnitude of individual depressive symptoms may vary from visit to visit. Consequently, the stability of the baseline measurement may not always be reliable. The magnitude of individual symptoms may also be affected by inherent study-related factors as well as totally unrelated personal events (confounding

* Corresponding author. 2 Oliver Street, Suite 1003, Boston, MA, 02109, USA.
E-mail address: sdtargum@yahoo.com (S.D. Targum).

https://doi.org/10.1016/j.jpsychires.2019.11.014
Received 25 September 2019; Received in revised form 28 October 2019; Accepted 21 November 2019

factors) that may occur during the study. Further, some subjects may become familiarized with the assessment measures and respond to any randomly assigned treatment regimen simply by virtue of the repetitive study procedures (Fava et al., 2003; Targum et al., 2008; Kirsch, 2014).

Clearly, it is important to obtain a reliable baseline measurement in clinical trials. Scoring fluctuation prior to randomization may adversely affect the reliability of the baseline measurement. In the current study, a 2-week double-blind placebo lead-in period afforded the opportunity to examine the effect of early score fluctuation prior to randomization on the eventual treatment outcome.

We asked the following questions:

1. Does early score fluctuation affect the placebo response and remission at the study endpoint?
2. Does early score fluctuation prior to randomization affect signal detection?

We examined these questions in subjects participating in a phase II, double-blind, placebo-controlled trial of adjunctive s-adenosyl methionine (SAMe) in MDD subjects who had had an inadequate response to ongoing antidepressant treatment (Horizon trial). We found that early score fluctuation ≥20% during the 2-week double-blind placebo lead-in period was associated with a significantly higher placebo response and remission rate at the study endpoint compared to placebo-assigned subjects who had < 20% score fluctuation during that time interval. We also found that post-hoc exclusion of evaluable subjects with early score fluctuation ≥20% markedly affected the outcome between the candidate drug and placebo.

2. Methods and materials

Data for this analysis was derived from the Horizon study (Clinicaltrials.org NCT01912196), a multicenter, randomized, double-blind, placebo-controlled, Phase 2 adjunctive treatment study designed to compare the efficacy and safety of proprietary s-adenosyl methionine (SAMe) 800 mg (MSI-195) plus ongoing antidepressant (ADT) with that of placebo plus ongoing ADT in subjects with MDD who had experienced an inadequate response to their ongoing ADT (Targum et al., 2018). The study was conducted at 35 clinical trial sites in the United States between December 2013 and June 2015. The candidate drug, SAMe is a naturally produced molecule that is distributed in all body tissues and has been shown to have pharmacological activity in animal models of depression (Benelli et al., 1999; Czyrak et al., 1992) as well as clinical trials (Nguyen and Gregan, 2002; Papakostas et al., 2003, 2010). MSI-195 is a proprietary formulation of SAMe that increases bioavailability (2–3 fold) in human subjects over commercially available dietary supplements (Targum et al., 2018).

Efficacy measures included the Montgomery-Asberg Depression Rating Scale (MADRS) and Hamilton rating scale for depression ($HamD_{17}$) derived from a combined MADRS-$HamD_{28}$ rating instrument, the patient-rated Inventory of Depressive Symptomatology (IDS-SR_{30}), and the Clinical Global Impression of Severity (CGI-S) scale (Hamilton, 1960; Guy, 1976; Montgomery and Asberg, 1979; Rush et al., 2000; Sackheim et al., 2003).

The study was carried out in accordance with the Declaration of Helsinki, and the study design was reviewed and approved by an appropriate ethical committee (IRB). All potential study subjects agreed in writing to participate in the study after reading and reviewing the IRB-approved informed consent.

Rater training and certification was completed by 84 raters and included observation and scoring of 2 demonstration MADRS-$HamD_{28}$ video interviews based upon the structured combined interview format developed by Sackheim and colleagues (Sackheim et al., 2003). The intraclass correlation (0.872 and 0.854), Kendall concordance coefficients (0.876 and 0.848), and the weighted kappa (0.75 and 0.76) showed a high level of inter-rater agreement.

The overall results for this study have previously been published and revealed that the study did not meet its specified endpoints (Targum et al., 2018).

2.1. Subject selection and study design

Eligible subjects were men or women between the ages of 21 and 70 years who met the Diagnostic and Statistical Manual of Mental Disorder, 4th Edition, Text Revision (DSM IV-TR) criteria for recurrent MDD as confirmed the Mini-International Neuropsychiatric Interview (M.I.N.I.) at the screen visit (APA, 1994; Sheehan et al., 1998). Eligible subjects had failed 1-3 adequate ADT treatments in the current episode, received at least 6 weeks of an adequate ADT dose (stable dose for at least 3 weeks) prior to the baseline (week 0) visit, and reported < 50% response on the MGH antidepressant treatment questionnaire (MGH-ATRQ) at the screen visit (Chandler et al., 2010; Desseilles et al., 2013). The acceptable ongoing ADTs taken during the study included sertraline, citalopram, escitalopram, fluoxetine, bupropion, venlafaxine, desvenlafaxine, paroxetine, duloxetine, mirtazapine, and nortriptyline.

Eligible subjects required a total HAM-D_{17} score ≥16, individual HAM-D_{17} mood item score ≥2, and a patient self-rated IDS-SR_{30} score > 28 at the screen and presumptive baseline (week 0) visits. The original protocol excluded subjects with > 20% score fluctuation on the total HAM-D_{17} or IDS-SR_{30} between the screening and presumptive baseline (week 0) visits. These criteria were amended mid-study such that the total $HamD_{17}$ score fluctuation exclusion criterion was increased to > 25% and the IDS-SR_{30} criterion was removed. There were no MADRS eligibility criteria.

A description of additional inclusion and exclusion criteria are provided in a previous publication (Targum et al., 2018).

In addition to these protocol eligibility criteria, telephone interviews were conducted by staff from the Massachusetts General Hospital Clinical Trials Network Inc (MGH-CTNI) as part of a site-independent confirmation of subject eligibility (Targum et al., 2008; Desseilles et al., 2013). The telephone interviews included the $HamD_{17}$, MGH-ATRQ, and specific validity questions. Potential subjects were screen failed if the independent rater scored the $HamD_{17}$ < 16, did not confirm that the subject had < 50% treatment response on the ATRQ despite adequate treatment, or failed specific validity questions.

The study design is shown in Fig. 1. After screening, all eligible subjects were enrolled in a 2-week double-blind placebo lead-in treatment period and continued taking their ongoing, current antidepressant medication. A placebo lead-in design has been supported and used by other investigators to assess symptom stability prior to randomization, although not all investigators agree (Quitkin et al., 1984, 1987; Reimherr et al., 1989; Faries et al., 2001; Evans et al., 2004; Trivedi and Rush, 1995). Investigators, study site staff, and study subjects were blinded to the true baseline (randomization) visit that occurred at week 2.

At week 2, the blinded randomization criteria included a total $HamD_{17}$ score ≥16 and $HamD_{17}$ score improvement < 50% from week 0. There were no MADRS randomization criteria. Thus, subjects whose total $HamD_{17}$ or MADRS scores fluctuated by < 50% but remained above the minimum $HamD_{17}$ score criterion (≥16) at week 2 were randomized. An additional criterion for the IDS-SR_{30} > 28 was removed mid-study following an amendment.

Subjects who met the specified randomization criteria at the week 2 visit were randomly assigned (1:1 allocation) to placebo plus ongoing ADT or MSI-195 (800 mg) plus ongoing ADT for an additional 6 weeks of double-blind treatment. Subjects who did not meet randomization criteria at week 2 remained in the study as non-evaluable subjects and continued to receive placebo plus their ongoing ADT. Thus, non-evaluable subjects were managed identically to the randomized, evaluable subjects and trial sites presumed that this was an 8-week double-blind, placebo-controlled study. Following randomization at week 2 (the true baseline visit), all subjects returned to the clinic at weeks 4, 6, and 8 for

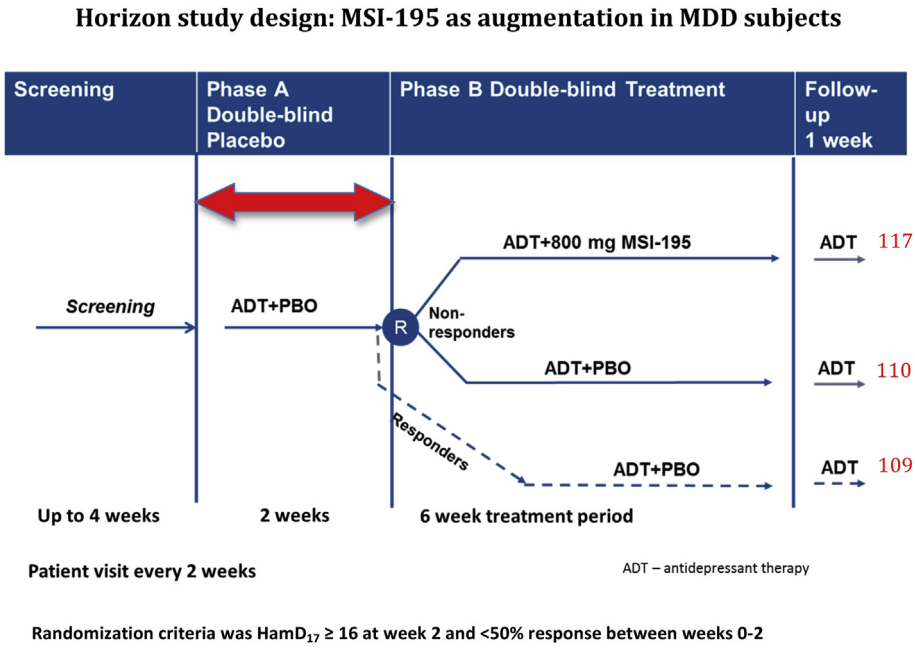

Fig. 1. Horizon study design: MSI-195 as augmentation in MDD subjects
Randomization criteria was HamD17 ≥ 16 at week 2 and < 50% response between weeks 0–2.

follow-up assessments. In addition, there were telephone safety visits at Weeks 1, 3, and 9.

2.2. Statistical analyses

The intent-to-treat (ITT) set was used for the primary analysis of all efficacy endpoints using only those subjects who had at least one assessment post randomization (week 2). For the primary analysis of study treatment outcome between the candidate drug (MSI-195) and the evaluable placebo group in the ITT population, the endpoint of change from randomization (Week 2) to the end of study (Week 8) of the total HAM-D_{17}, MADRS, IDS-SR_{30}, and CGI-S scores were analyzed using a mixed model, repeated measures (MMRM) procedure utilizing all post-randomization change scores. The full results of the primary analysis have been reported elsewhere (Targum et al., 2018).

In the current analysis, we focused on the $HamD_{17}$ and MADRS. Analyses of treatment outcome of the combined evaluable and non-evaluable placebo-assigned subjects used the 8-week treatment interval from the presumptive baseline (week 0) to week 8. Statistical analyses used the last observation carried forward (LOCF) method, analysis of variance, and Students *t*-test where appropriate. Treatment response for the placebo-assigned subjects was defined as an improvement of the total $HamD_{17}$ score or MADRS ≥50% for stated time intervals (weeks 0–2, 0–8, and 2–8). Remission was defined as a total $HamD_{17}$ score of ≤7 or total MADRS of ≤11.

Post-hoc analyses also compared a sub-group of subjects who had been randomized to either the candidate drug (MSI-195) or the evaluable placebo-assigned group. The post-hoc dataset included only subjects who had a total score fluctuation of < 20% between weeks 0–2 (the double-blind placebo lead-in period). This criterion was similar to the original screen to week 0 protocol eligibility criterion and is consistent with other reported studies (Altin et al., 2014; Targum and Catania, 2017). The $HamD_{17}$ and MADRS outcomes were examined in separate analyses.

3. Results

There were 615 subjects screened for this study of who 239 (38.9%) were screen failed prior to week 0. Of the remaining subjects, 336 subjects began the two-week double-blind placebo lead-in period and had at least one post-randomization visit beyond week 2. There were 227 subjects who met study eligibility criteria at week 2, qualified for the ITT set, and were randomly assigned (1:1 allocation) to either MSI-195 (n = 117) or the evaluable placebo-assigned group (n = 110) for the 6-week double-blind treatment phase. The subjects who did not meet randomization criteria were continued in the non-evaluable placebo-assigned group (n = 109). Within the non-evaluable group, 93 subjects (85.3%) were not randomized because their $HamD_{17}$ score was < 16 at week 2.

The full study results have been reported elsewhere (Targum et al., 2018) and revealed that MSI-195-assigned subjects did not achieve statistically significant separation from the evaluable placebo-assigned subjects on any of the clinical measures in the 6-week double-blind period that followed randomization. MSI-195 was well tolerated and the predominant treatment emergent adverse events (TEAEs) were mild and primarily related to the gastrointestinal tract.

3.1. Impact of early score fluctuation on eventual placebo response

There were 219 placebo-assigned subjects at week 2 that included 110 from the evaluable placebo group and 109 from the non-evaluable group. We examined:

- The total $HamD_{17}$ and MADRS score trajectories for all placebo-assigned subjects. We plotted the trajectories of the placebo-assigned subjects who had < 20% $HamD_{17}$ or MADRS score fluctuation, ≥ 20% but < 50% score fluctuation, and ≥50% score fluctuation (meeting the criterion for treatment response) between weeks 0 and week 2 (Figs. 2 and 3).
- The overall score changes of the mean $HamD_{17}$ and MADRS scores at week 8 relative to the $HamD_{17}$ and MADRS score fluctuation between week 0 (the presumptive baseline) and week 2 (the true baseline visit). We used < 20% and ≥20% score fluctuation between weeks 0–2 as the cut-off (Tables 1 and 2).

3.1.1. Early $HamD_{17}$ score fluctuation and eventual outcome

There were 336 subjects (including all 219 placebo-assigned subjects and 117 subjects later assigned to MSI-195) who entered the

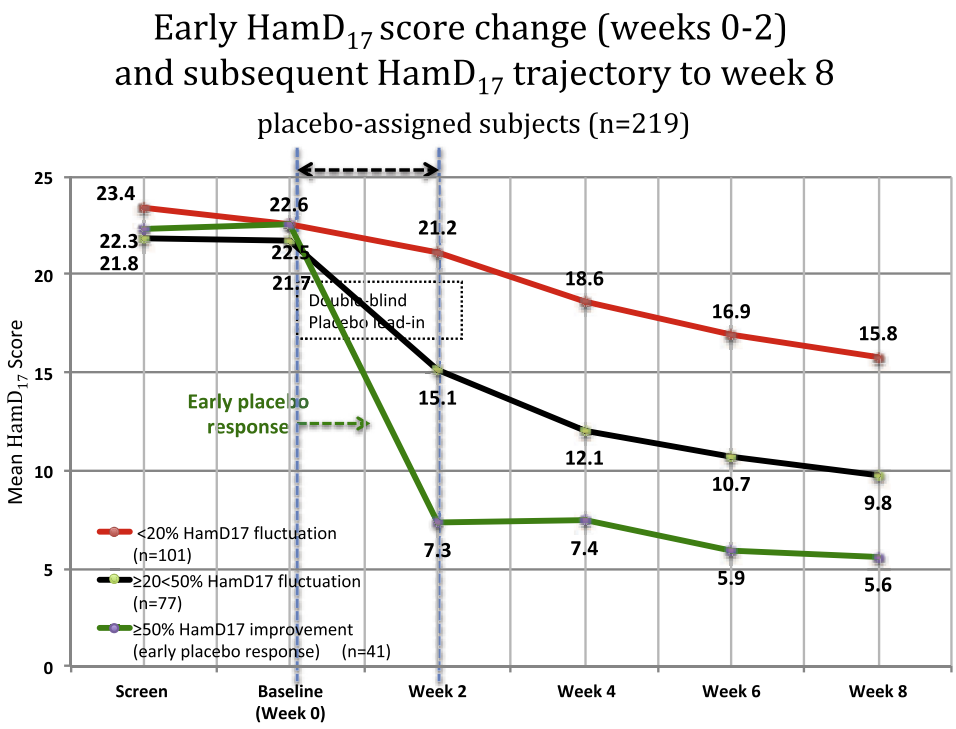

Fig. 2. Early $HamD_{17}$ score change (weeks 0–2) and subsequent $HamD_{17}$ trajectory to week 8 placebo-assigned subjects (n = 219).

double-blind placebo lead-in at week 0. In this study, 41 of all 336 subjects enrolled at week 0 (12.2%) met criteria for a $HamD_{17}$ placebo response by week 2.

Fig. 2 graphically depicts the total $HamD_{17}$ score trajectories for all 219 placebo-assigned subjects during the 8 weeks of double-blind treatment. An early $HamD_{17}$ score fluctuation during the double-blind placebo lead-in was associated with sustained response over 8 weeks.

Table 1 examines the effect of the early $HamD_{17}$ score changes between weeks 0–2 on the eventual study outcome in the placebo-treated subjects. The table compares the eventual 8-week (endpoint) placebo response and remission rates for placebo-assigned subjects who had < 20% $HamD_{17}$ score fluctuation, ≥20% score fluctuation, as well

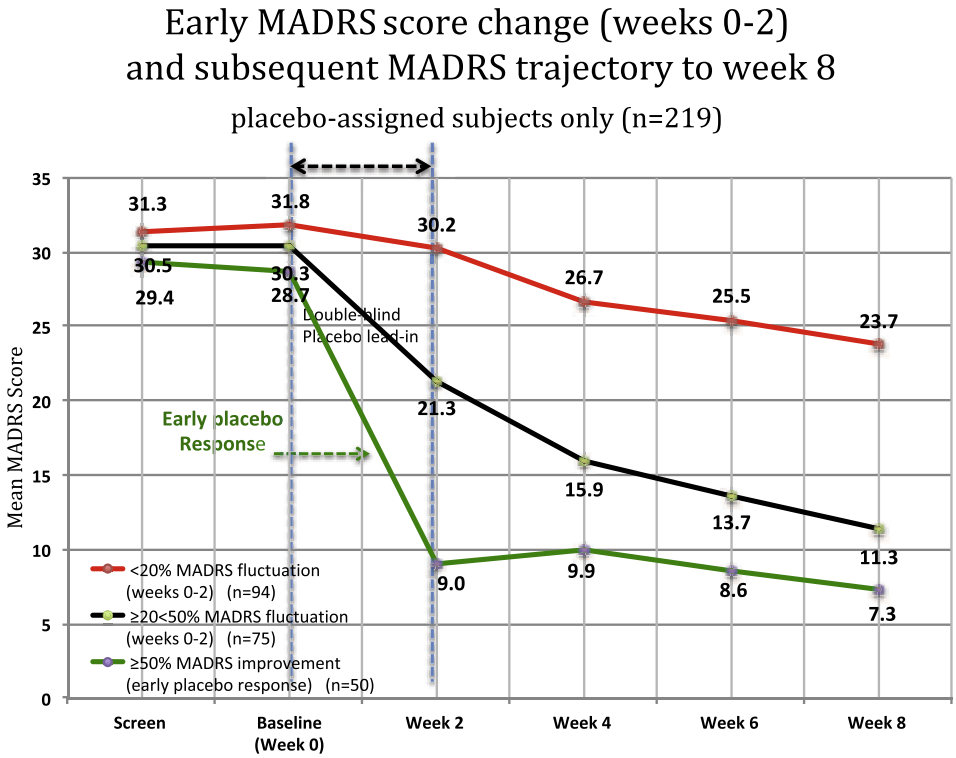

Fig. 3. Early MADRS score change (weeks 0–2) and subsequent MADRS trajectory to week 8 placebo-assigned subjects only (n = 219).

Table 1
Early $HamD_{17}$ score changes and study outcome in placebo-treated subjects $HamD_{17}$.

$HamD_{17}$	< 20% $HamD_{17}$ fluctuation between weeks 0–2	≥ 20% $HamD_{17}$ fluctuation between weeks 0–2	p < 20% vs. ≥ 20%	Early placebo response (at week 2)	Early remission $HamD_{17}$ ≤ 7 (at week 2)
n	101	118		41[a]	20
Δ week 0–2	−1.4 ± 2.2	−9.6 ± 5.7	F = 184.7; p < 0.0001	−15.2 ± 4.3	−17.4 ± 4.6
Δ week 0–8	−6.8 ± 6.5	−13.7 ± 6.8	F = 57.4; p < 0.0001	−17.0 ± 6.1	−17.3 ± 6.9
Δ week 2–8	−5.4 ± 6.4	−4.1 ± 6.1	F = 2.3; p < 0.13	−1.8 ± 4.8	+0.05 ± 4.8
Placebo response at 8 weeks	31 (30.4%)	82 (69.5%)	λ^2 = 31.3; p < 0.0001	35 (85.4%)	17 (85.0%)
Remission at 8 weeks	15 (17.1%)	71 (60.2%)	λ^2 = 45.0; p < 0.0001	28 (68.3%)	13 (65.0%)

[a] 41 of 336 enrolled subjects (12,2%) were $HamD_{17}$ placebo responders between weeks 0–2.

as those with early placebo or remission responses at week 2. As shown, the 101 subjects with < 20% early $HamD_{17}$ score fluctuation (46.1% of all placebo-assigned subjects) had significantly less overall improvement by week 8 relative to subjects who had ≥20% fluctuation between weeks 0 and 2 ($p < 0.0001$).

As shown in Table 1, 35 of the 41 subjects (85.4%) who met the $HamD_{17}$ placebo response criterion at week 2 were still placebo responders at week 8 in contrast to 31 of the 101 placebo-assigned subjects (30.4%) with < 20% $HamD_{17}$ fluctuation between weeks 0–2 ($X^2 = 32.9$; df = 1; $p < 0.0001$).

The early $HamD_{17}$ response and remission rates were largely sustained at 8 weeks. Between weeks 2–8 (the true double-blind treatment period), the subjects with ≥20% $HamD_{17}$ score fluctuation between weeks 0–2 actually had less score improvement than subjects with < 20% score fluctuation ($p = 0.13$).

Only 5 of the 118 placebo-assigned subjects with ≥20% $HamD_{17}$ score fluctuation got worse between weeks 0–2. Four of these 5 subjects became placebo responders and 3 were in remission at week 8.

3.1.2. Early MADRS score fluctuation and eventual outcome

In this study, 50 of the 336 subjects entering the double-blind placebo lead-in at week 0 (14.9%) met criteria for a MADRS placebo response by week 2.

Fig. 3 graphically depicts the total MADRS score trajectories for all placebo-assigned subjects during the 8 weeks of double-blind treatment. Similar to the $HamD_{17}$ response trajectory, an early MADRS score fluctuation (≥20%) during the double-blind placebo lead-in was associated with sustained response at 8 weeks.

Table 2 examines the effect of early MADRS score changes on the eventual study outcome in the placebo-treated subjects. Similar to the $HamD_{17}$ response, the 94 subjects with < 20% early MADRS score fluctuation (42.9% of all placebo-assigned subjects) had significantly less overall improvement by week 8 relative to subjects who had ≥20% MADRS fluctuation between weeks 0 and 2 ($p < 0.0001$).

Similar to the $HamD_{17}$ response, 41 of the 50 subjects (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 fluctuation between weeks 0–2 ($X^2 = 52.8$; df = 1; $p < 0.0001$).

Between weeks 2 and 8 (the true double-blind treatment period), the MADRS score improvement was identical between placebo-assigned subjects who had < 20% or ≥20% MADRS fluctuation between weeks 0 and 2 (p = ns).

Only 3 of the 125 placebo-assigned subjects with ≥20% MADRS score fluctuation got worse between weeks 0–2. Two of these 3 subjects became placebo responders and 1 was in remission at week 8.

3.2. Post-hoc analyses of evaluable subjects with < 20% early score fluctuation

We conducted a post-hoc analysis of the 217 evaluable ITT subjects using a randomization criterion of < 20% score fluctuation on either the total $HamD_{17}$ or MADRS between weeks 0–2 (the double-blind placebo lead-in period). The resulting sub-groups had 162 subjects in the $HamD_{17}$ analysis and 144 subjects in the MADRS analysis (Figs. 4 and 5).

As shown in Figs. 4 and 5, the effect sizes (ES) favoring MSI-195 over placebo improved in both the $HamD_{17}$ and MADRS sub-group analyses relative to the ITT population. In the MADRS sub-group analysis, the ES improved from 0.125 in the ITT group to 0.404, and the MSI-195 sub-group revealed a statistically significant benefit over placebo ($F = 6.39$; df = 1; $p = 0.012$). A reduced placebo response observed during weeks 2–8 in subjects with < 20% score fluctuation facilitated the enhanced ES. Between weeks 2–8, only 15 of the 74 evaluable subjects (20.2%) with < 20% MADRS score fluctuation between weeks 0–2 were placebo responders in contrast to 35 of the 110

Table 2
Early MADRS score changes and study outcome. in placebo-treated subjects MADRS.

MADRS	< 20% MADRS fluctuation between weeks 0–2	≥20% MADRS fluctuation between weeks 0–2	p < 20% vs. ≥20%	Early placebo response (at week 2)	Early remission MADRS ≤11 (at week 2)
n	94	125		50[a]	35
Δ week 0–2	−1.9 ± 2.8	−13.3 ± 7.3	F = 203.4; p < 0.0001	19.6 ± 5.7	−20.5 ± 6.3
Δ week 0–8	−8.6 ± 9.4	−20.0 ± 9.8	F = 75.0; p < 0.0001	−21.4 ± 10.2	−21.6 ± 9.5
Δ week 2–8	−6.6 ± 9.3	−6.7 ± 9.6	F = 0.0; p = ns	−1.7 ± 8.2	−1.1 ± 6.7
Placebo response at 8 weeks	17 (23.0%)	96 (76.8%)	λ^2 = 71.7; p < 0.0001	41 (82.0%)	29 (82.9%)
Remission at 8 weeks	15 (20.3%)	76 (60.8%)	λ^2 = 42.6; p < 0.0001	36 (72.0%)	25 (71.5%)

[a] 50 of 336 enrolled subjects (14.9%) were MADRS placebo responders between weeks 0–2.

subjects (31.8%) in the original placebo-assigned ITT population (X^2 = 2.43; df = 1; p = 0.12).

4. Discussion

The Horizon study compared MSI-195 (a proprietary formulation of SAMe) with placebo in MDD subjects with a documented inadequate response to their ongoing ADT. The overall study failed to meet its pre-specified clinical endpoints (Targum et al., 2018). As part of the study design, we included a two-week double-blind placebo lead-in phase to identify and exclude early placebo responders. The clinical trial sites were blind to the 2-week placebo-lead-in period, blind to the week 2 randomization criteria, and consequently presumed that the trial was a conventional 8-week double-blind, placebo-controlled study.

The 227 subjects who met the week 2 randomization criteria were allocated in 1:1 fashion to MSI-195 (n = 117) or placebo (n = 110) in addition to their ongoing ADT. To maintain the blind, the 109 non-evaluable subjects who failed the week 2 randomization criteria continued on placebo plus their ongoing ADT for the full 8-week double-blind treatment.

We examined the response trajectory of the combined group of 219 evaluable and non-evaluable placebo-assigned subjects. Placebo-assigned subjects with ≥20% $HamD_{17}$ 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 (Tables 1 and 2). These findings of a sustained placebo response following an early response (within 2 weeks of randomization) are consistent with several other studies that have analyzed different MDD populations (Faries et al., 2001; Szegedi et al., 2009; Altin et al., 2014; Targum and Catania, 2017; Targum, 2017; Targum et al., 2018).

In this study, 12.2% of subjects entering week 0 had ≥50% $HamD_{17}$ improvement (the defined placebo response) by week 2 and 14.9% were MADRS placebo responders. 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. Hence, the early score improvement prior to randomization at week 2 had a marked effect on the endpoint.

In this study, a ≥20% total score fluctuation (improvement or worsening) of either the $HamD_{17}$ or MADRS during the 2-week double-blind placebo lead-in period adversely affected the eventual treatment outcome at week 8 (endpoint). As shown in Figs. 4 and 5, a post-hoc analysis of evaluable subjects (ITT population) that excluded subjects

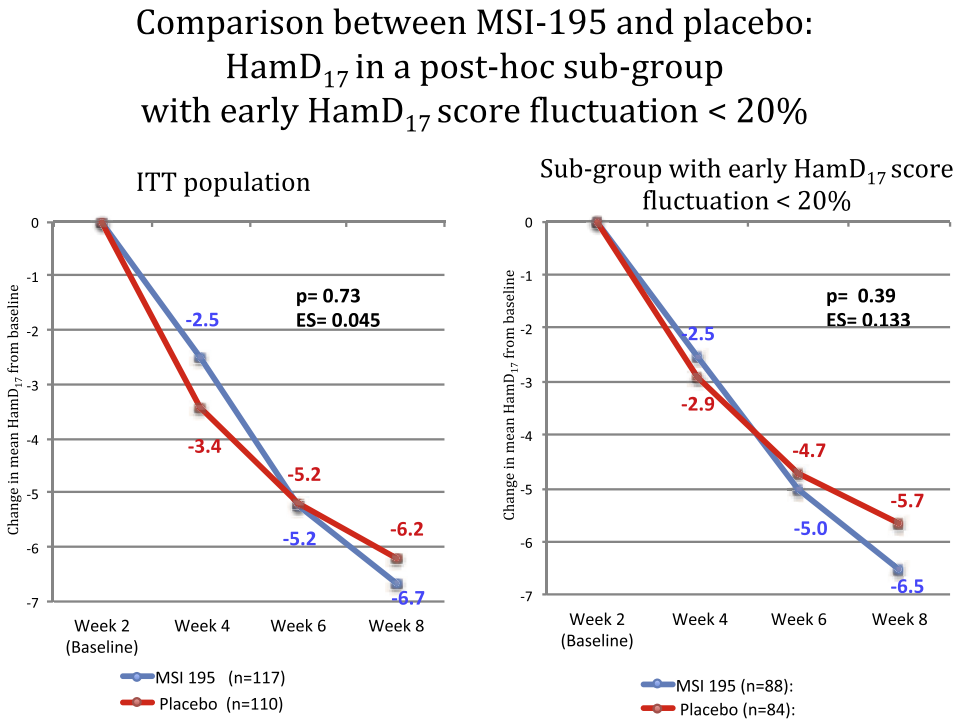

Fig. 4. Comparison between MSI-195 and placebo: $HamD_{17}$ in a post-hoc sub-group with early $HamD_{17}$ score fluctuation < 20%.

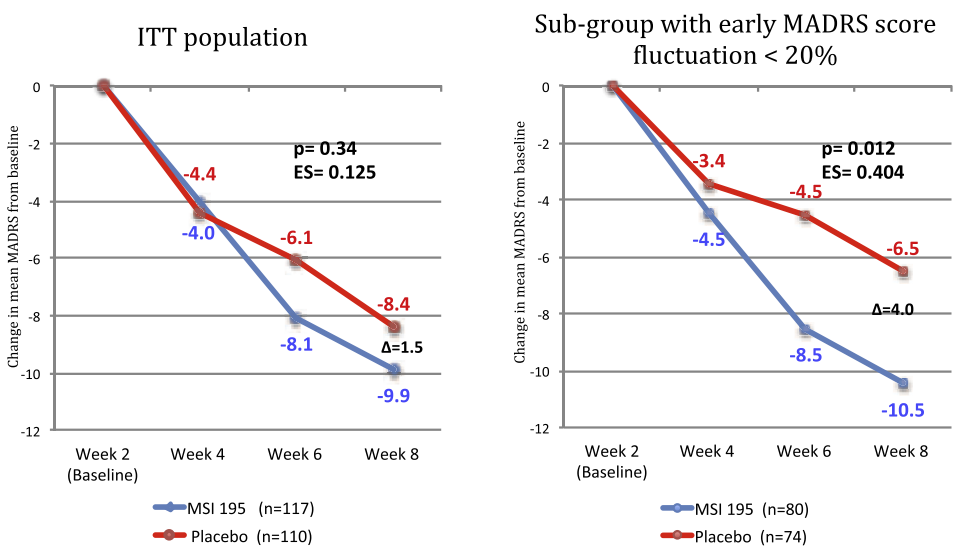

Fig. 5. Comparison between MSI-195 and placebo: MADRS in a post-hoc sub-group with early MADRS score fluctuation < 20%.

with ≥20% early score fluctuation yielded higher effect sizes (ES) for both the $HamD_{17}$ and MADRS sub-groups, and yielded statistical significance for MSI-195 over placebo in the MADRS sub-group ($p = 0.012$). The ES in the MADRS sub-group favoring MSI-195 improved from 0.125 to 0.404. An ES ≥ 0.40 is considered clinically significant in antidepressant trials (Faries et al., 2000; Bech, 2017). The marked difference of the ES in this post-hoc analysis was largely determined by a reduced placebo response between weeks 2–8 (Table 2). We have shown a similar enhancement of the ES following post-hoc adjustments for early score fluctuation in other MDD studies as well (Targum, 2018; Targum and Catania, 2018).

Some authors believe that a placebo lead-in period does not improve the assessment of treatment effect (Trivedi and Rush, 1995; Whitlock et al., 2019). We believe that the establishment of a more reliable baseline score may offset some of the inherent limitations of rating scale measurement in MDD (Bech, 2017). Study designs that identify and exclude potential subjects who present with fluctuating symptoms prior to randomization may facilitate a more stable baseline measurement. The optimization of a reliable baseline measure is clearly an asset for signal detection.

We do not presume that a study design strategy that excludes subjects with early score fluctuation will necessarily be effective in other depression studies. However, the results do highlight the importance of subject selection and study design on the ultimate success of any clinical trial. Clearly, a validated neurobiological marker that identifies likely placebo responders would be desirable and might ameliorate some of the inherent challenges of subject selection (Trivedi et al., 2016). For instance, recent brain activation studies using resting-state fMRI during emotional conflict task paradigms have generated models that may differentiate antidepressant versus placebo-specific response patterns that could be prognostic markers (Goldstein-Piekarski et al., 2018; Zilcha-Mano et al., 2019; Fonzo et al., 2019). Meanwhile, appropriate subject selection remains a challenge in every clinical trial.

The findings reported in this post-hoc analysis of the Horizon study demonstrate that subjects who reveal early score fluctuation regardless of treatment assignment can adversely affect the assessment of the eventual treatment effect. Our findings are based on this one study and require prospective confirmation from other studies. However, the findings suggest that study designs that anticipate and attempt to manage early response prior to randomization may yield more meaningful outcome data for trials of MDD, and possibly other disorders as well.

Contributors

Dr. Targum, Beth R. Cameron, Ludvina Ferreira, and I. David Macdonald all participated in the design, implementation, and analysis of the original study and conceived, analyzed, and wrote the current post-hoc analysis reported in this manuscript.

Declaration of competing interest

Dr. Targum was chief medical officer at Methylation Sciences Inc. at the time of this study. He is currently an employee of Signant Health. He has received consultation fees or vendor grants from Acadia Pharmaceuticals, Alkermes Inc., BrainCells Inc., Functional Neuromodulation Inc., Intracellular Therapies, Inc., Karuna Pharmaceuticals, Johnson and Johnson PRD, Methylation Sciences Inc., Navitor, Neurim Pharmaceuticals, Perception Neuroscience, Pfizer Inc., and Sunovion Pharmaceuticals.

Beth Cameron was an employee of Methylation Sciences Inc. at the time this study was conducted. Ludvina Ferreira was an employee of Methylation Sciences Inc. at the time this study was conducted. She is currently a project manager at PRA Health Sciences. David Macdonald is an employee of Methylation Sciences Inc.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychires.2019.11.014.

Source of funding/Role of the sponsor

This research was funded by Methylation Sciences Inc. (Vancouver, British Columbia).

References

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry (59 Suppl. l), 34–57.

Altin, M., Harada, E., Schacht, A., Berggren, L., Walker, D., Dueñas, H., 2014. 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. Depress. 3, 112–123.

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), fourth ed. American Psychiatric Association, Washington DC.

Bech, P., 2017. Rating scales in depression: limitations and pitfalls. Dialogues Clin. Neurosci. 8 (2), 207–215.

Benelli, A., Filaferro, M., Bertolini, A., Genedani, S., 1999. Influence of S-adenosylmethionine on chronic mild stress-induced anhedonia in castrated rats. Br. J. Pharmacol. 127 (3), 645–654.

Chandler, G.M., Iosifescu, D., Pollack, M., Targum, S.D., Fava, M., 2010. Validation of the Massachusetts general hospital antidepressant treatment history questionnaire (ATRQ). CNS Neurosci. Ther. 16 (5), 322–325.

Czyrak, A., Rogoz, Z., Skuza, G., Maj, J., 1992. Antidepressant activity of S-adenosylmethionine in mice and rats. J. Basic Clin. Physiol. Pharmacol. 3 (1), 1–17.

Desseilles, M., Witte, J., Chang, T.E., Iovieno, N., Dording, C., Ashih, H., Nyer, M., Freeman, M.P., Fava, M., Mischoulon, D., 2013. Massachusetts General Hospital SAFER criteria for clinical trials and research. Harv. Rev. Psychiatry 21 (5), 269–274.

Evans, K.R., Sills, T., Wunderlich, G.R., McDonald, H.P., 2004. Worsening of depressive symptoms prior to randomization in clinical trials: a possible screen for placebo responders? J. Psychiatr. Res. 38, 437–444.

Faries, D.F., Herrera, J., Rayamajhi, J., DeBrota, D., Demitrack, M., Potter, W.Z., 2000. The responsiveness of the Hamilton depression rating scale. J. Psychiatr. Res. 34, 3–10.

Faries, D.F., Heiligenstein, J.H., Tollefson, G.D., Potter, W.Z., 2001. The double blind variable placebo lead-in period: results from two antidepressant clinical trials. J. Clin. Psychopharmacol. 6, 561–568.

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

Fonzo, G.A., Etkin, A., Zhang, Y., Wu, W., Cooper, C., Chin-Fatt, C., Jha, M.K., Trombello, J., Deckersbach, T., Adams, P., McInnis, M., McGrath, P.J., Weissman, M.M., Fava, M., Trivedi, M.H., 2019. Brain regulation of emotional conflict predicts antidepressant treatment response for depression. Nat. Human Behav. https://doi.org/10.1038/s41562-019-0732-1.

Furukawa, T.A., Cipriani, A., Atkinson, L.Z., Leucht, S., Ogawa, Y., Takeshima, N., Hayasaka, Y., Chaimani, A., Salanti, G., 2016. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomized controlled studies. Lancet Psychiatry 3, 1059–1066.

Goldstein-Piekarski, A.N., Ball, T.M., Staveland, B.R., Yesavage, J., Korgaonkar, M.S., Williams, L.M., 2018. Intrinsic functional connectivity predicts remission on antidepressants: a randomized controlled trial to identify clinically applicable imaging biomarkers. Transl. Psychiatry 8 (57), 1–11. https://doi.org/10.1038/s41398-018-0100-3.

Guy, W., 1976. EDCEU Assessment Manual for Psychopharmacology – Revised (DHEW Publ No ADM 76-338). Rockville, MD, U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration. NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs, pp. 534–537.

Hamilton, M., 1960. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389.

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

Khan, A., Leventhal, R.M., Khan, S.R., Brown, W.A., 2002a. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. J. Clin. Psychopharmacol. 22 (1), 40–45.

Khan, A., Khan, S., Bowen, W.A., 2002b. Are placebo controls necessary to test new antidepressants and anxiolytics? Int. J. Neuropsychopharmacol. 5 (3), 193–197.

Khin, N.A., Chen, Y.F., Yang, Y., Yang, P., Laughren, T.P., 2011. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. J. Clin. Psychiatry 72, 464–472.

Kirsch, I., 2014. Antidepressants and the placebo effect. Zeitschrift fur Psychologie 222 (3), 128–134.

Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., Johnson, B.T., 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med. 5, e45. https://doi.org/10.1371/journal.pmed.0050045.

Kobak, K.A., Kane, J.M., Thase, M.E., Nierenberg, A.A., 2007. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? J. Clin. Psychopharmacol. 27, 1–5.

Laughren, T.P., 2001. The scientific basis and ethical basis for placebo-controlled trials in depression and schizophrenia: an FDA perspective. Eur. Psychiatry 16 (7), 418–423.

Lee, S., Walker, J.R., Jakul, L., Sexton, K., 2004. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. Depress. Anxiety 19, 10–19.

Montgomery, S.A., Asberg, M.A., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389.

Nguyen, M., Gregan, A., 2002. S-adenosylmethionine and depression. Aust. Fam. Physician 31 (4), 339–343.

Papakostas, G.I., Alpert, J.E., Fava, M., 2003. S-adenosyl-methionine in depression: a comprehensive review of the literature. Curr. Psychiatr. Rep. 5 (6), 460–466.

Papakostas, G.I., Mischoulon, D., Shyu, I., Alpert, J.E., Fava, M., 2010. S-adenosyl methionine (SAMe) augmentation of serotonin reuptake inhibitors for antidepressant nonresponders with major depressive disorder: a double-blind, randomized clinical trial. Am. J. Psychiatry 167 (8), 942–948.

Quitkin, F.M., Rabkin, J.G., Ross, D., Stewart, J.W., 1984. Identification of true drug response to antidepressants. Arch. Gen. Psychiatr. 41, 782–786.

Quitkin, F.M., Rabkin, J.G., Markowitz, J.M., Stewart, J.W., McGrath, P.J., Harrison, W., 1987. Use of pattern analysis to identify true drug response: a replication. Arch. Gen. Psychiatr. 44, 259–264.

Reimherr, F.W., Ward, M.F., Byerley, W.F., 1989. The introductory placebo washout: a retrospective evaluation. Psychol. Res. 30, 191–199.

Rush, A.J., Carmody, T., Reimitz, P.E., 2000. The inventory of depressive Symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. Int. J. Methods Psychiatr. Res. 9 (2), 45–59.

Rutherford, B.R., Roose, S.P., 2013. A model of placebo response in antidepressant clinical trials. Am. J. Psychiatry 170 (7), 723–733.

Sackheim, H.A., Kiersky, J.E., Demitrack, M.A., 2003. In: A Structured Interview for the Combined Hamilton Rating Scale for Depression (HRSD) and Montgomery-Asberg Depression Rating Scale (MADRS) Version 4, vol. 3. pp. 11–19.

Szegedi, A., Jansen, W.T., van Willigenburg, A.P., van der Meulen, E., Stassen, H.H., Thase, M.E., 2009. 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 70 (3), 344–353.

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

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

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

Targum, S.D., Cameron, B.R., Ferreira, L., MacDonald, I.D., 2018. An augmentation study of MSI-195 (S-adenosylmethionine) in major depressive disorder. J. Psychiatr. Res. 107, 86–96.

Thase, M.E., Larsen, K.G., Kennedy, S.H., 2011. Assessing the 'true' effect of active antidepressant therapy v. placebo in major depressive disorder. Br. J. Psychiatry 199, 501–507.

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

Trivedi, M.H., McGrath, P.J., Fava, M., Parsey, R.V., Kurian, B.T., Phillips, M.L., Oquendo, M.A., Bruder, G., Pizzagalli, D., Toups, M., Cooper, C., Adams, P., Weyandt, S., Morris, D.W., Grannemann, B.D., Ogden, R.T., Buckner, R., McInnis, M., Kraemer, H.C., Petkova, E., Carmody, T.J., Weissman, M.M., 2016. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. J. Psychiatr. Res. 78, 11–23.

Whitlock, M.E., Woodward, P.W., Alexander, R.C., 2019. Is high placebo response really a problem in depression trials? Innov. Clin. Neurosci. 16 (7), 12–17.

Zilcha-Mano, S., Wang, Z., Peterson, B.S., Wall, M.M., Chen, Y., Wager, T.D., Brown, P.J., Roose, S.P., Rutherford, B.R., 2019. Neural mechanisms of expectancy-based placebo effects in antidepressant clinical trials. J. Psychiatr. Res. 116, 19–25. https://doi.org/10.1016/j.jpsychires.2019.05.023.

Zimmerman, M., Chelminski, I., Posternak, M.A., 2005. Generalizability of antidepressant efficacy trials: differences between depressed psychiatric outpatients who would or would not qualify for an efficacy trial. Am. J. Psychiatry 162, 1370–1372.