Toward a Definition of "No Meaningful Benefit" From Antidepressant Treatment:

An Equipercentile Analysis With Cross-Trial Validation Across Multiple Rating Scales

Carl Zhang[a]; Sanya Virani, MD[b]; Taryn Mayes, MS[c]; Thomas Carmody, PhD[d]; Paul E. Croarkin, DO[e]; Richard Weinshilboum, MD[a]; A. John Rush, MD[f,g]; Madhukar Trivedi, MD[c]; Arjun P. Athreya, PhD[a]; and William V. Bobo, MD[h,*]

ABSTRACT

Background: Many patients with major depressive disorder (MDD) who experience no meaningful benefit (NMB) from antidepressive treatment go undetected. However, there is a lack of consensus on the definition of NMB from antidepressants.

Methods: Equipercentile linking was used to identify a threshold for percent change in 17-item Hamilton Depression Rating Scale (HDRS-17) scores that equated with a Clinical Global Impressions-Improvement (CGI-I) score of 3 (minimally improved), a proxy for NMB, after 4 and 8 weeks of citalopram or escitalopram treatment, using data from the Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS). The NMB threshold for the HDRS-17 was validated by equating a CGI-I rating of 3 with percent change values from the clinician- and patient-rated versions of the Quick Inventory of Depressive Symptomatology (QIDS-C and QIDS-SR) using data from PGRN-AMPS and phase 1 of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial. This study was conducted between June 2021 and September 2021.

Results: In PGRN-AMPS, a 30% improvement in HDRS-17 score corresponded to a CGI-I rating of 3 at 4 and 8 weeks. The 30% improvement threshold was also observed for QIDS-C and QIDS-SR scores in both PGRN-AMPS and STAR*D. Similar results were observed for percent change in HDRS-17 and QIDS-based measures in lower- and higher-severity groups based on a median split of baseline total scores.

Conclusions: Improvement in depressive severity of ≤ 30%, as assessed using the HDRS-17, QIDS-C, and QIDS-SR, may validly define NMB from antidepressants during short-term treatment.

J Clin Psychiatry 2022;83(4):21m14239

To cite: Zhang C, Virani S, Mayes T, et al. Toward a definition of "no meaningful benefit" from antidepressant treatment: an equipercentile analysis with cross-trial validation across multiple rating scales. *J Clin Psychiatry.* 2022;83(4):21m14239.

To share: https://doi.org/10.4088/JCP.21m14239

[a]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota

[b]Department of Psychiatry and Human Behavior, Alpert School of Medicine, Brown University, Providence, Rhode Island

[c]Peter O'Donnell Jr. Brain Institute and Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas

[d]Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, Texas

[e]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota

[f]Department of Psychiatry and Behavioral Sciences, Department of Medicine, Duke University, Durham, North Carolina

[g]Duke-National University of Singapore, Singapore

[h]Department of Psychiatry and Psychology, Mayo Clinic, Jacksonville, Florida

**Corresponding author:* William V. Bobo, MD, MPH, 4500 San Pablo Rd, Jacksonville, FL 32224 (bobo.william@mayo.edu).

Major depressive disorder (MDD) affects over 264 million people worldwide,[1] making it one of the most prevalent of illnesses in medicine and the leading global cause of disability from chronic diseases.[2,3] Antidepressant medications and evidence-based psychotherapy are standard treatments for MDD. Unfortunately, over one-third of depressed patients fail to respond to adequate antidepressant treatment trials,[4] and multiple sequential trials of antidepressive treatments are usually needed before depressive symptoms are effectively managed.[5,6] The process of managing poor response to treatment with antidepressants can be substantially improved with the systematic measurement of depressive symptoms (measurement-based care), which has been associated with improved clinical outcomes as compared with treatment as usual, thus narrowing the gap between clinical research and real-world practice.[5,7,8]

For managing an antidepressant response that falls short of remission, the choice of a specific next-step treatment depends on the accurate characterization of response to the existing treatment. If there is meaningful clinical benefit without remission (eg, response), therapeutic options include changing the dose or adding a second agent or psychotherapy if dose optimization is ineffective or poorly tolerated.[9] Conversely, if there is no meaningful benefit (NMB) after an adequate period of observation, switching to an alternative treatment is needed.[10] Therefore, a validated threshold for defining NMB would considerably benefit clinicians and patients by providing an alert to the need to switch from an ineffective treatment to a potentially more effective one, thus reducing the duration of active depressive symptoms and their negative impacts on quality of life and functioning.[11] There is a fair consensus about the definition of a minimally acceptable response (eg, a 50% reduction in depression symptom scores[12]). Among the approximately 30%–50% of antidepressant-treated patients who do not respond, it would be worthwhile to better characterize the subset who obtain NMB.[13] However,

Clinical Points

- Many patients with major depressive disorder who experience no meaningful benefit (NMB) from antidepressive treatment go undetected. However, there is a lack of consensus on the definition of NMB from antidepressants.
- Equipercentile analyses demonstrated that an improvement in depressive severity of 30% or less, as assessed using the HDRS-17, QIDS-C, and QIDS-SR, may validly define NMB from antidepressants during acute-phase treatment.
- The early detection of NMB from a given antidepressive treatment may prompt a switch to a potentially more effective mode of treatment, thus lessening the time needed for an individual patient to achieve a clinically meaningful response to treatment if more than one therapeutic trial is needed.

to our knowledge, there is no validated definition of NMB from an antidepressant.

This study aimed to develop and validate a categorical definition of NMB by equating percent change in scores on 3 common depression rating scales with scores on the Clinical Global Impression (CGI) scale,[14] a validated measure of clinician impressions of a patient's symptoms and functioning after initiation of treatment.[15,16] Equipercentile linking has been used to validate the 50% improvement threshold for antidepressive response using a CGI-Improvement subscale (CGI-I) score ≤ 2 (much improved) as a proxy for clinically significant change[17–19] and to identify thresholds of clinically significant change in depression severity by linking CGI-Severity subscale ratings and quality of life measures.[20] Here, we link changes in 17-item Hamilton Depression Rating Scale[21] (HDRS-17) and CGI-I scores to derive an NMB threshold after 4 and 8 weeks of treatment with citalopram or escitalopram. We validated the NMB threshold using clinician- and subject-rated versions of the Quick Inventory of Depressive Symptomatology[22] (QIDS-C and QIDS-SR) and CGI-I scores in a separate dataset.

METHODS

Depressive Symptom Measures

The HDRS-17 is a clinician-rated measure consisting of 17 items that rate the severity of depressive symptoms, 8 of which are rated on a 5-point scale and 9 of which are rated on a 3-point scale (total score ranges from 0 to 52). The QIDS-C and QIDS-SR consist of 16 items that rate the severity of depressive symptoms on a 4-point scale with some individual items combined (total score ranges from 0 to 27). For all 3 scales, higher scores indicate greater depressive symptom severity. A percent (%) change in depression severity at follow-up visits (ie, 2, 4, or 8 weeks) was defined as

$$\frac{\text{Severity at Follow-up Visits} - \text{Baseline Severity}}{\text{Baseline Severity}} \times 100$$

Sources of Data, Sample, and Treatment

Table 1 summarizes the demographic and clinical characteristics of studies used in this work.

Development dataset. The Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study[23] (PGRN-AMPS, NCT00613470) was designed to examine genetic factors associated with clinical response after 8 weeks of open-label treatment with citalopram or escitalopram in adults with MDD. All PGRN-AMPS subjects provided written informed consent, and the study protocol was approved by the institutional review board of Mayo Clinic. The PGRN-AMPS sample consisted of 922 adults (aged 18–84 years) with Structured Clinical Interview for *DSM-IV* (SCID)–confirmed diagnoses of MDD and an HDRS-17 total score ≥ 14 at trial entry. PGRN-AMPS participants received open-label treatment with citalopram (starting at 20 mg/d) or escitalopram (starting at 10 mg/d), with postbaseline study visits occurring at weeks 4 and 8. The doses of study medications could be increased at week 4 (to 40 mg/d of citalopram or 20 mg/d of escitalopram) if the QIDS-C total score was ≥ 9. The dataset for PGRN-AMPS served as the development dataset for deriving a definition of NMB from antidepressants using HDRS-17 scores, as described further below.

Validation dataset. Data from the first phase of the Sequenced Treatment Alternatives to Relieve Depression[5,22] (STAR*D; NCT00021528) trial were used to validate the NMB threshold derived from PGRN-AMPS data. Participants were adults with MDD who received open-label treatment with citalopram for up to 14 weeks. All STAR*D subjects provided written informed consent. The STAR*D protocol was approved by institutional review boards at the national coordinating center, the data coordinating center, 14 regional centers, and individual clinical sites. The STAR*D samples for this work consisted of 1,866 citalopram-treated adults (aged 18–75 years) with *DSM-IV*–defined MDD and an HDRS-17 total score ≥ 14 who had complete data at baseline, week 4, and week 8. A total of 1,636 (of the 1,866 STAR*D phase 1 participants with 4- and 8-week data) also had complete data from the week 2 visit. Citalopram was started at a dose of 20 mg/d. Subsequent dose titrations were governed by a pre-established plan, up to a maximum dose of 60 mg/d.

Depressive symptoms were measured at baseline, week 4, and week 8 using the HDRS-17, the QIDS-C, and the QIDS-SR in the PGRN-AMPS trial—and at weeks 2, 4, 6, 8, 12, and 14 using the QIDS-C and QIDS-SR in the STAR*D phase 1 trial (only data from weeks 2, 4, and 8 were used for the analyses in this study)—by trained clinical raters.

Measure of Global Clinical State

For this study, scores on the HDRS-17, QIDS-C, and QIDS-SR were linked with CGI-I ratings. In the PGRN-AMPS and STAR*D trials, the CGI-I was rated by experienced clinicians at postbaseline follow-up visits using the following 7-point scale: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse,

Table 1. Sample Characteristics

Variable	PGRN-AMPS	STAR*D[a]			
Total N	927	1,636	1,866	1,581	1,804
Age, mean (SD), y	39.10 (14.27)	42.14 (13.14)	42.03 (13.14)	41.97 (13.18)	41.98 (13.02)
Sex	M: 352; F: 570	M: 639; F: 997	M: 709; F: 1,157	M: 621; F: 960	M: 697; F: 1,107
Drug exposure	Citalopram/escitalopram	Citalopram	Citalopram	Citalopram	Citalopram
Rating scale(s)	QIDS-C, QIDS-SR, HDRS- 17	QIDS-C	QIDS-C	QIDS-SR	QIDS-SR
Timepoints for assessing treatment response	4 and 8 weeks	2, 4, and 8 weeks	4 and 8 weeks	2, 4, and 8 weeks	4 and 8 weeks
Race					
White	848	1,198	1,338	1,158	1,296
Black	17	221	275	205	252
Hispanic	0	173	206	173	208
Asian	12	25	27	27	29
Hawaiian	1	9	10	9	10
American Indian	2	10	10	9	9
Other	47	0	0	0	0

[a]For the STAR*D sample, there were 1,198 subjects with complete data for QIDS-C at 2, 4, and 8 weeks; 1,338 subjects with complete data for QIDS-C at 4 and 8 weeks; 1,158 subjects with complete data for QIDS-SR at 2, 4, and 8 weeks; and 1,296 subjects with complete data for QIDS-SR at 4 and 8 weeks.

Abbreviations: F = female, HDRS = 17-item Hamilton Depression Rating Scale, M = male, PGRN-AMPS = Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study, QIDS-C = clinician-rated version of the Quick Inventory of Depressive Symptomatology, QIDS-SR = subject-rated version of the Quick Inventory of Depressive Symptomatology, STAR*D = Sequenced Treatment Alternatives to Relieve Depression.

6 = much worse, and 7 = very much worse. For this study, a CGI-I score of 3 or higher was chosen as the cutoff point for defining NMB because a CGI-I score ≤ 2 is an accepted threshold for defining clinically meaningful improvement in depressive symptoms[18] and because defining NMB based on the absence of improvement in depressive symptoms (a CGI-I score of 4) does not account for minimal but non-meaningful levels of improvement.

Statistical Analysis

Descriptive statistics were used to summarize the demographic and clinical characteristics of the PGRN-AMPS and STAR*D phase 1 study participants (see Table 1). Equipercentile linking was used to equate HDRS-17 (development dataset), QIDS-C and QIDS-SR (development and validation datasets), and CGI-I measures. Equipercentile linking is a nonparametric statistical process that is used to find equivalent points on separate but correlated scales, accounting for possible measurement error for each of the scales under consideration in this study. Correlations between the CGI-I and percent change measures from each depression rating scale were assessed using Spearman rank correlation coefficients, testing the coefficients versus no correlation using an F test at a statistical significance threshold of $P < .05$. After establishing correlation, equipercentile linking was performed by calculating the empirical distribution functions for the CGI-I and each of the depression scale percent change values (as percentiles for all measures) and then matching the percentiles between the two measures. Therefore, for a given score on the CGI-I rating, a corresponding percent change in score from baseline for a given depression scale with the same percentile rank was identified. The resulting pairs of scores were plotted with each point on graphs representing equivalent (linked) CGI-I scores and percent changes in total scores for the HDRS-17 and QIDS-C/-SR. These points were connected by a smooth curve, thus displaying the equipercentile relationship between CGI-I and percent change measures for each depression scale across the entire range of values at each follow-up time point. To account for confounding by baseline depression severity, we repeated the equipercentile analyses within strata based on a median split of baseline HDRS-17, QIDS-C, and QIDS-SR scores, thus creating higher-severity (baseline scores equal to or higher than the median value) and lower-severity groups.

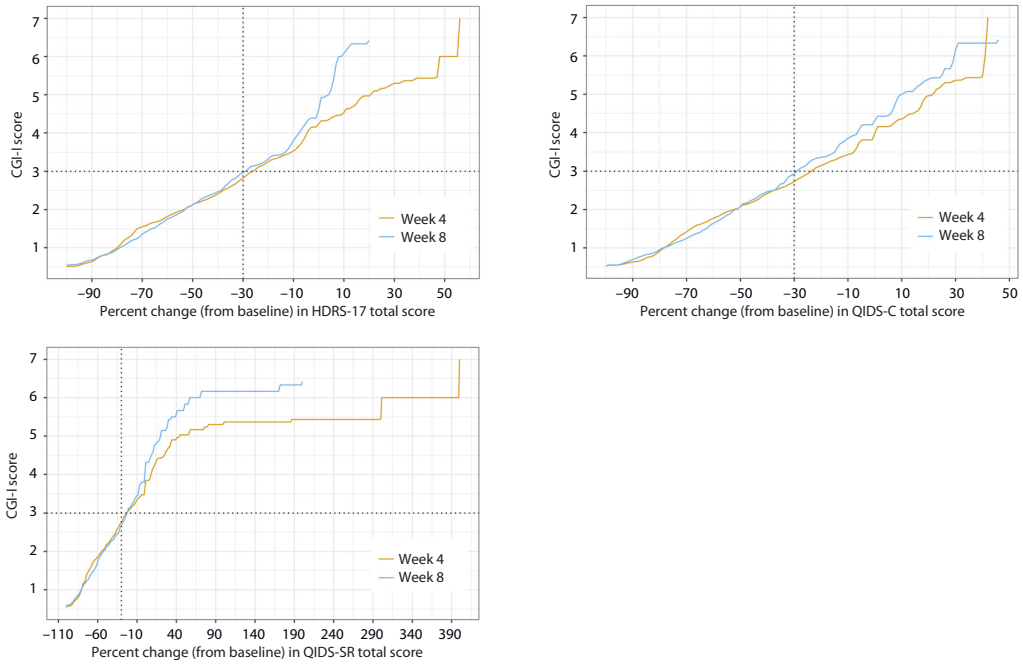
Table 2. Correlation Between HDRS-17, QIDS-C, QIDS-SR, and CGI-Based Measures

Study and linking variables	Timepoint	Spearman correlation coefficient	*P* value
PGRN-AMPS			
% Change in HDRS-17 from baseline and CGI-I	Week 4	0.726	2.50E-111
	Week 8	0.729	1.20E-100
% Change in QIDS-C from baseline and CGI-I	Week 4	0.69	4.30E-96
	Week 8	0.682	3.00E-83
% Change in QIDS-SR from baseline and CGI-I	Week 4	0.675	1.40E-90
	Week 8	0.656	2.20E-16
STAR*D (with complete data for baseline and 2, 4, and 8 weeks)			
% Change in QIDS-C from baseline and CGI-I	Week 4	0.802	~0
	Week 8	0.836	~0
% Change in QIDS-SR from baseline and CGI-I	Week 4	0.617	5.60E-190
	Week 8	0.697	5.20E-263
STAR*D (with complete data for baseline and 2, 4, and 8 weeks)			
% Change in QIDS-C from baseline and CGI-I	Week 2	0.749	1.50E-294
	Week 4	0.799	~0
	Week 8	0.835	~0
% Change in QIDS-SR from baseline and CGI-I	Week 2	0.504	2.20E-102
	Week 4	0.612	7.60E-193
	Week 8	0.688	7.50E-222

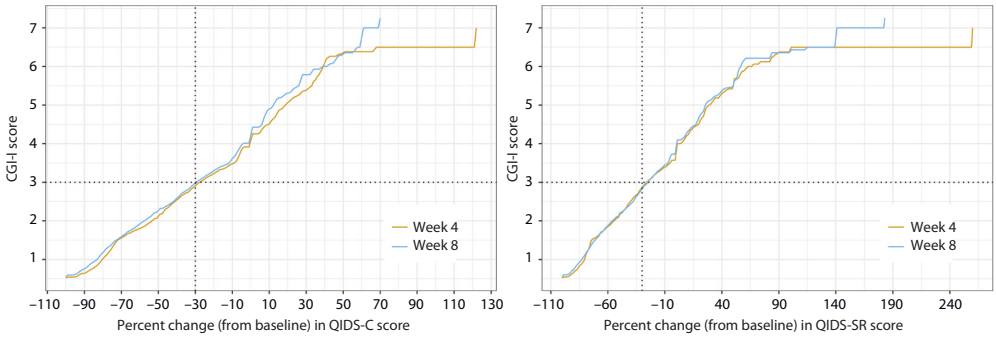
Abbreviations: CGI-I = Clinical Global Impressions-Improvement subscale, HDRS-17 = 17-item Hamilton Depression Rating Scale, PGRN-AMPS = Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study, QIDS-C = clinician-rated version of the Quick Inventory of Depressive Symptomatology, QIDS-SR = subject-rated version of the Quick Inventory of Depressive Symptomatology, STAR*D = Sequenced Treatment Alternatives to Relieve Depression study.

Figure 1. Equipercentile Linking of Scores on the CGI-I to Percent Change (From Baseline) in Total Depression Severity Scores in (A) PGRN-AMPS and (B) STAR*D Participants With Complete Data at Baseline and 4 and 8 Weeks and in (C) STAR*D Participants With Complete Data at Baseline and 2, 4, and 8 Weeks

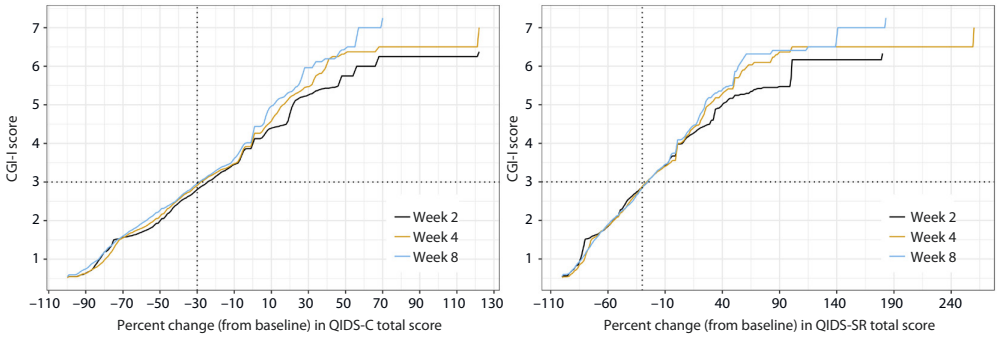
A. PGRN-AMPS participants with data at baseline and 4 and 8 weeks



B. STAR*D participants with data at baseline and 4 and 8 weeks



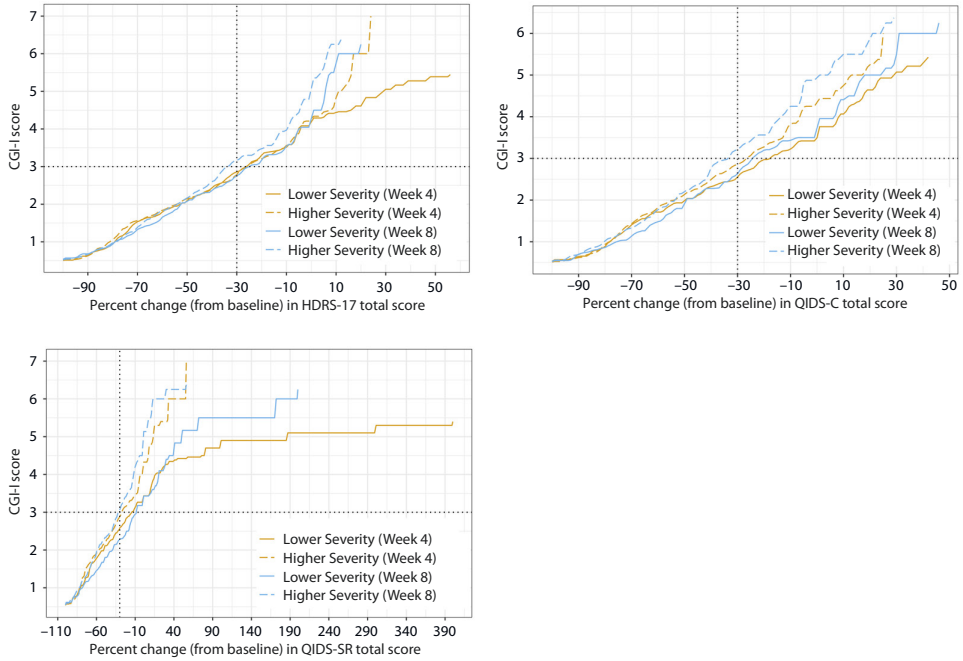
C. STAR*D participants with data at baseline and 2, 4, and 8 weeks



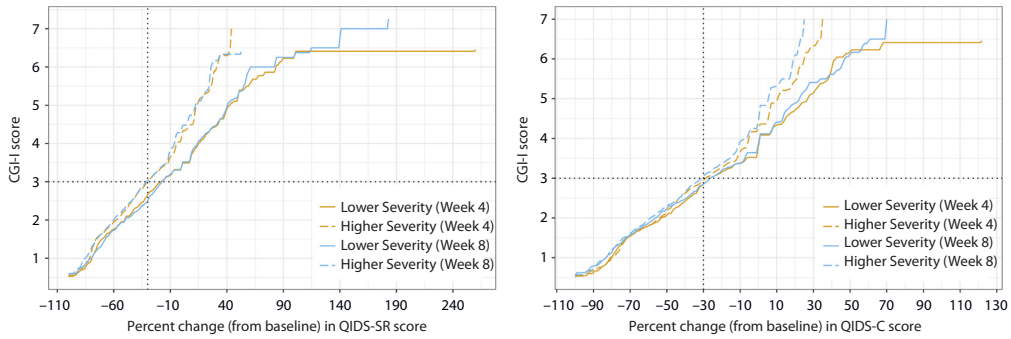
Abbreviations: CGI-I = Clinical Global Impression-Improvement, PGRN-AMPS = Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study, QIDS-C = clinician-rated version of the Quick Inventory of Depressive Symptomatology, QIDS-SR = subject-rated version of the Quick Inventory of Depressive Symptomatology, STAR*D = Sequenced Treatment Alternatives to Relieve Depression.

Figure 2. Equipercentile Linking of Scores on the CGI-I to Percent Change (From Baseline) in Total Depression Severity Scores of Patients Stratified by Baseline Depression Severity (Based on a Median Split of Total Depression Scores at Baseline) in (A) PGRN-AMPS and (B) STAR*D Participants With Complete Data at Baseline and 4 and 8 Weeks and in (C) STAR*D Participants With Complete Data at Baseline and 2, 4, and 8 Weeks

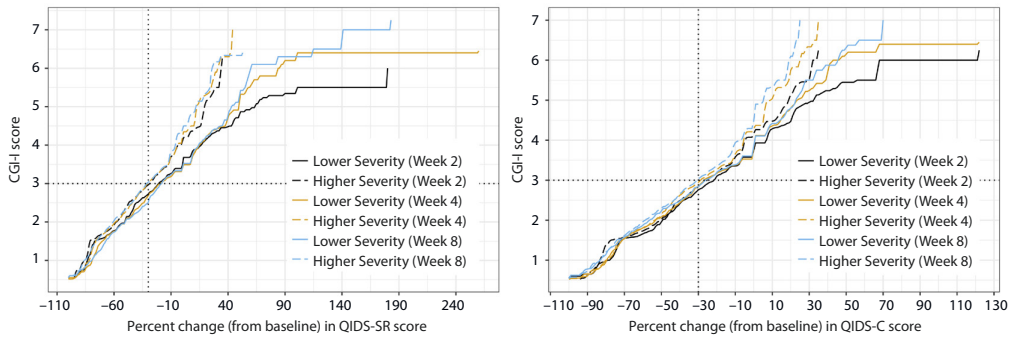
A. PGRN-AMPS participants with data at baseline and 4 and 8 weeks



B. STAR*D participants with data at baseline and 4 and 8 weeks



C. STAR*D participants with data at baseline and 2, 4, and 8 weeks



Abbreviations: CGI-I = Clinical Global Impression-Improvement, PGRN-AMPS = Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study, QIDS-C = clinician-rated version of the Quick Inventory of Depressive Symptomatology, QIDS-SR = subject-rated version of the Quick Inventory of Depressive Symptomatology, STAR*D = Sequenced Treatment Alternatives to Relieve Depression.

Data for this study were analyzed using the Equate package in R. The current study, conducted between June 2021 and September 2021, was considered exempt by the local institutional review board.

RESULTS

Correlation Between HDRS-17, QIDS-C, QIDS-SR, and CGI-Based Measures

Spearman correlations between CGI-I and percent change from baseline in depression scale scores at each follow-up time point are presented in Table 2. Strong correlations were observed for all pairs of CGI-I scores and percent change values for each depression scale, thus allowing equipercentile linkage for these measures.

Development of No Meaningful Benefit Definition Using PGRN-AMPS Data

Figure 1A illustrates the results of equipercentile linking of percent changes (from baseline) in HDRS-17 total scores with CGI-I scores at weeks 4 and 8 for PGRN-AMPS participants. At week 4 and week 8, an improvement of 30% or less in HDRS-17 total scores equated to a CGI-I score of 3 or higher. A change threshold of ≤ 30% from baseline in HDRS-17 total scores was thus chosen to define NMB among PGRN-AMPS participants. At week 4 and week 8, an improvement of 30% or less in QIDS-C and QIDS-SR total scores also equated to a CGI-I score of 3 or higher, thus providing cross-scale replication. An improvement in HDRS-17 scores of 0%–8% at week 4 and 0%–13% at week 8 mapped to a CGI-I score of 4 (no change), which indirectly validated the equipercentile link between these measures.

Validation of NMB Definition Using STAR*D Data

The results of equipercentile linking of CGI-I scores and QIDS-C and QIDS-SR percent change values (from baseline) are shown in Figures 1B and 1C for STAR*D subjects who had complete data at baseline and 4 and 8 weeks and at baseline and 2, 4, and 8 weeks, respectively. Each of these equipercentile analyses were used to validate an NMB threshold of ≤ 30% improvement from baseline in depression scale scores. For each of these analyses, an improvement of 30% of less on both QIDS-based measures in each dataset linked to a CGI-I score of 3 and above, thus providing validation of the 30% NMB threshold in 2 separate datasets.

Stratified Analyses by Baseline Depression Severity

Figure 2 and Table 3 summarize the results of equipercentile linking of CGI-I ratings and percent change values (from baseline) for HDRS-17 (PGRN-AMPS only) and QIDS-C/-SR scores in higher- and lower-severity groups (based on a median split of baseline total scores) at weeks 2 (STAR*D only), 4, and 8 (both PGRN-AMPS and STAR*D). For these analyses, percent change in the HDRS-17, QIDS-C, and QIDS-SR scores from baseline in PGRN-AMPS and STAR*D subjects that corresponded to a CGI-I score of 3 ranged from 22%–39% in the higher-severity groups and 12%–24% in lower-severity groups, as illustrated in Table 3.

Table 3. Range of Percent Change in Depression Severity Mapping to CGI-I Score of 3 in Patients, Stratified by Depression Severity

Study and timepoint	Depression rating scale	Range of % change in depression severity mapping to CGI-I = 3	
		Lower severity	Higher severity
PGRN-AMPS			
Week 4	HDRS-17	22%–32%	22%–39%
Week 8	HDRS-17	14%–35%	23%–41%
Week 4	QIDS-C	11%–24%	24%–32%
Week 8	QIDS-C	15%–32%	27%–41%
Week 4	QIDS-SR	9%–23%	22%–32%
Week 8	QIDS-SR	0%–21%	25%–39%
STAR*D (with complete data for baseline and 4 and 8 weeks)			
Week 4	QIDS-C	22% –29%	26%–30%
Week 8	QIDS-C	21%–31%	29%–34%
Week 4	QIDS-SR	15%–24%	28%–32%
Week 8	QIDS-SR	12%–22%	27%–34%
STAR*D (with complete data for baseline and 2, 4, and 8 weeks)			
Week 2	QIDS-C	21%–25%	24%–28%
Week 4	QIDS-C	22%–29%	26%–31%
Week 8	QIDS-C	21%–32%	27%–34%
Week 2	QIDS-SR	18%–23%	27%–31%
Week 4	QIDS-SR	15%–24%	28% –32%
Week 8	QIDS-SR	14%–23%	27%–34%

Abbreviations: CGI-I = Clinical Global Impressions improvement subscale, HDRS-17 = 17-item Hamilton Depression Rating Scale, PGRN-AMPS = the Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study, STAR*D = the Sequenced Treatment Alternatives to Relieve Depression study, QIDS-C = clinician-rated version of the Quick Inventory of Depressive Symptomatology, QIDS-SR = subject-rated version of the Quick Inventory of Depressive Symptomatology.

DISCUSSION

The effective management of patients with MDD relies on the ability to efficiently detect patients who are responding poorly to a current course of treatment and thus require a change.[8] In this study, we developed and validated a definition of NMB from antidepressants using multiple depression scales and antidepressant trial datasets. An improvement in HDRS-17, QIDS-C, and QIDS-SR of 30% or less from baseline mapped to CGI-I scores of 3 and above, a range of CGI-I scores that represents, at best, a nonmeaningful level of change in depressive symptoms.

The clinical importance of the findings from this study rests on the facts that measurement-based treatment of MDD has become the clinical standard in research settings and among a growing number of clinicians[24] and that a very large number of patients require multiple therapeutic trials of antidepressive treatment before achieving a positive treatment outcome.[4,24] The early detection of NMB from a given antidepressive treatment may prompt a switch to a potentially more effective mode of treatment, thus lessening the time needed for an individual patient to achieve a clinically meaningful response if more than 1 therapeutic trial is needed.[25–27] The importance of having a practical and valid definition of NMB is further highlighted by the fact that failure to detect patients who poorly respond to antidepressive

treatment is common in clinical practice[28]—a problem that is preventable with the systematic use of depression rating scales to monitor the effects of antidepressive treatments.[8]

Despite its importance, there is still no consensus in the field on how to define NMB from antidepressants.[24] The lack of a consensus definition complicates the ability to systematize knowledge about poor response to antidepressants, with important implications for clinical research and practice.[29] Selecting antidepressants for depressed patients entails a "try-and-try-again" approach, with several weeks of watchful waiting at each step.[30] The large number of depressed patients who do not achieve meaningful benefit from a given treatment and require multiple therapeutic trials highlights the importance of characterizing those who are at risk of NMB. Developing prediction models for NMB in research settings and making decisions about treatment in clinical settings both require a valid categorical definition. For a large number of depressed patients, a reliable and valid definition of NMB would be instrumental toward the goal of reducing the exposure to treatments that are unlikely to work.[27]

From a practical viewpoint, our goal was to validate a NMB definition that replicates across depression scales and is simple enough for clinical use. The NMB threshold of ≤ 30% improvement in depressive symptoms from baseline appears to be reliable given its remarkable consistency across 3 depression scales in 2 independent datasets. In terms of practical use, calculating a percent reduction in symptoms is relatively simple and at least partially accounts for variation in baseline depression severity. Indeed, after rerunning the equipercentile analyses in higher- and lower-severity groups, analogous to the approach taken in prior studies,[17–19] the robustness of our main findings was supported.

A variety of statistical approaches have been previously used to link information from continuous depression rating scales and proxy measures for clinical significance, with the goal of deriving clinically meaningful levels of improvement in depressive symptoms.[17–20,24,31] To our knowledge, this report describes the first use of equipercentile linking to develop and test a categorical definition of NMB from antidepressants. Equipercentile linking was chosen to equate values derived from depression scales and CGI-I ratings in this study given that it is nonparametric (a specific type of distribution of measured values is not required) and accounts for possible measurement error for the scales used in each of the antidepressant trials datasets.[32,33] A CGI-I score of 3 was selected a priori for defining and validating a definition for NMB because it is greater than a CGI-I score of 2, which denotes a clinically meaningful improvement in depressive symptoms (ie, a positive antidepressive response) and because it is less than a CGI-I score of 4, which denotes the complete absence of improvement in depressive symptoms. Complete absence of improvement is too stringent for developing an ecologically valid definition of NMB because it disallows the possibility of minimal, nonmeaningful levels of improvement that are commonly encountered in clinical practice and would still necessitate a change in therapy.[34,35] Our results show that antidepressive response and NMB are not mere inverses of one another, an observation that is consistent with the results of a large, 5-year clinical registry study that included 328 patients with treatment-resistant MDD who received vagal nerve stimulation.[36] In that study, Quality of Life Enjoyment and Satisfaction Questionnaire scores improved to a clinically significant degree in patients who experienced as little as a 35% improvement in depression severity.

Our study had several strengths, including large numbers of depressed patients in the PGRN-AMPS and STAR*D datasets, the use of certified clinical raters in both trials, and the strong correlations between CGI-I and percent change values for each rating scale at all follow-up time points. By linking scores on the CGI-I and the self-reported version of the QIDS, validation of the NMB threshold in this study did not rely exclusively on clinician ratings, a common criticism for studies of clinical interventions for MDD and beyond.[37,38] And finally, our NMB definition was validated at 3 different time points, including as early as 2 weeks—an important clinical consideration given data suggesting that a lack of meaningful benefit from an antidepressant at 2 weeks could indicate the need for an early change in treatment.[39]

Limitations. Both PGRN-AMPS and phase 1 of STAR*D used an open design, and, although assessments of inter-rater reliability for depression symptom ratings were performed in both trials, periodic assessments of inter-rater reliability of CGI scores were not conducted. PGRN-AMPS trial procedures did not prevent study clinicians from accessing HDRS-17 or QIDS-C/-SR scores, which could have influenced CGI-I scores during follow-up. More importantly, the accuracy and value of the NMB definition derived in this work may be compromised by limited recall accuracy of baseline depression severity and inherent inter- and intrarater variability in depression symptom assessments, primarily by clinicians rather than patients.[40] Across the trials that provided data for this study, QIDS-SR scores were generally lower than QIDS-C scores at weeks 4 and 8, with correspondingly lower thresholds for NMB (24%–25%); thus, use of equipercentile analyses for identifying NMB thresholds may be sensitive to whether instruments used to measure depressive symptom change are patient- or clinician-rated. In both trials, the impacts of study attrition on the relationship between CGI-I ratings and percent change values for the depression scales are unknown. The generalizability of our findings is limited to adults with non–treatment-resistant MDD who were given 2 related antidepressants, citalopram and escitalopram. PGRN-AMPS consisted of a predominantly Caucasian sample enrolled at a single site, although the STAR*D validation sample was derived from multiple clinical sites and was more diverse in terms of racial makeup. Generalizability of our findings may also be somewhat limited using citalopram doses as high as 60 mg/d in the STAR*D sample, which now exceeds the FDA-recommended limit. Our NMB definition was not validated against patient-rated measures of clinical global

state or measures of quality of life or functioning, both of which are important aspects of recovery from depression that are not always highly correlated with symptom improvement.[41,42] We were also unable to ascertain how adverse effect burden influenced or intersected with NMB in this research. Similarly, we were unable to validate our NMB definition against biological measures that may reflect underlying mechanisms of depression or responses to treatment.[12,40,43–45] Finally, although we tested our NMB definition at 3 different time points, none extended beyond 8 weeks of treatment. Future work should focus on testing the value of the categorical NMB ≤ 30% depressive symptoms score reduction on additional scales and measures of functioning, such as the Sheehan Disability Scale and widely used Patient Health Questionnaire-9.

CONCLUSIONS

In summary, our findings suggest that improvement of depressive symptoms of 30% or less from baseline, as measured by the HDRS-17 and the clinician- and subject-rated versions of the QIDS, validly defines NMB during short-term treatment with citalopram or escitalopram. Future studies are needed to establish its predictive validity and its applicability to other antidepressive treatments and to patients with treatment-resistant depression.

Submitted: September 4, 2021; accepted January 10, 2022.

Published online: June 27, 2022.

Author contributions: Drs Bobo and Athreya served as overall co-principal investigators and, in this capacity, planned and oversaw the analysis of study data. Drs Bobo and Athreya and Mr Zhang wrote the initial draft of the manuscript. All authors contributed to the interpretation of findings and subsequent drafts of the submitted manuscript and approved the final manuscript. Drs Bobo and Athreya had access to all data and take responsibility for the accuracy of the data analysis and the writing of the manuscript.

Relevant financial relationships: **Dr Carmody** has provided consulting services to Alkermes. **Dr Croarkin** has received research grant support from Neuronetics, NeoSync, and Pfizer and has received grant in-kind (equipment and supply support for research studies) from Assurex Health, Neuronetics, and MagVenture and has served as a consultant for Engrail Therapeutics, Myriad Neuroscience, Procter & Gamble, and Sunovion. **Dr Weinshilboum** is a co-founder and stockholder in OneOme, LLC, a pharmacogenomic clinical decision support company. **Dr Rush** has received consulting fees from Akili, Brain Resource, Compass, Curbstone Consultant LLC, Emmes Corp., Johnson & Johnson (Janssen), Liva-Nova, Mind Linc, and Sunovion; speaking fees from Janssen and Liva-Nova; and royalties from Guilford Press and the University of Texas Southwestern Medical Center, Dallas (for the Inventory of Depressive Symptomatology and its derivatives). He is also named co-inventor on 2 patents: US Patent No. 7,795,033: Methods to Predict the Outcome of Treatment with Antidepressant Medication and US Patent No. 7,906,283: Methods to Identify Patients at Risk of Developing Adverse Events During Treatment with Antidepressant Medication. **Dr Trivedi** has provided consulting services to Acadia Pharmaceuticals, Alkermes, Alto Neuroscience, Axsome Therapeutics, GH Research Limited, GreenLight VitalSign6, Janssen, Merck Sharp & Dohme, Mind Medicine (MindMed), Neurocrine Biosciences, Orexo US, Otsuka, SAGE Therapeutics, Signant Health, and Titan Pharmaceuticals, Inc, and he has received grant/research funding from NIMH, National Institute on Drug Abuse, Patient-Centered Outcomes Research Institute, and Cancer Prevention Research Institute of Texas. Additionally, he has received editorial compensation from Oxford University Press. **Dr Athreya** receives funding support from National Science Foundation and National Institutes of Health. **Dr Bobo's** research has been supported by the National Institute for Mental Health (NIMH), Agency for Healthcare Research and Quality, National Science Foundation, Mayo Foundation for Medical Education and Research, and the Myocarditis Foundation, and he has contributed chapters to UpToDate on the treatment of bipolar disorders. The other authors declared no competing interests for this work.

Funding/support: This study was supported, in part, by funds from the Mayo Foundation for Medical Education and Research, Mayo Clinic Summer Undergraduate Research Fellowship, National Institute of General Medical Sciences under grants GM61388 and GM28157, National Science Foundation under award 2041339, and National Institute of Mental Health under grant MH90003. The STAR*D trial was supported by Bristol-Myers Squibb, Forest Pharmaceuticals, GlaxoSmithKline, King Pharmaceuticals, Organon, Pfizer, and Wyeth-Ayerst Laboratories, who provided study medications.

Role of the sponsor: The sources of funding for this study had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

REFERENCES

1. James SL, Abate D, Abate H; GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789–1858.
2. Bromet E, Andrade LH, Hwang I, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med*. 2011;9(1):90.
3. Smith K. Mental health: a world of depression. *Nature*. 2014;515(7526):180–181.
4. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163(11):1905–1917.
5. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
6. Carvalho AF, Berk M, Hyphantis TN, et al. The integrative management of treatment-resistant depression: a comprehensive review and perspectives. *Psychother Psychosom*. 2014;83(2):70–88.
7. Guo T, Xiang YT, Xiao L, et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry*. 2015;172(10):1004–1013.
8. Fortney JC, Unützer J, Wrenn G, et al. A tipping point for measurement-based care. *Psychiatr Serv*. 2017;68(2):179–188.
9. Kudlow PA, McIntyre RS, Lam RW. Early switching strategies in antidepressant non-responders: current evidence and future research directions. *CNS Drugs*. 2014;28(7):601–609.
10. Quitkin FM, McGrath PJ, Stewart JW, et al. Chronological milestones to guide drug change: when should clinicians switch antidepressants? *Arch Gen Psychiatry*. 1996;53(9):785–792.
11. Pae CU, Wang SM, Lee SY, et al. Early switch strategy in patients with major depressive disorder. *Expert Rev Neurother*. 2012;12(10):1185–1188.
12. Nierenberg AA, DeCecco LM. Definitions of antidepressant treatment response, remission, nonresponse, partial response, and other relevant outcomes: a focus on treatment-resistant depression. *J Clin Psychiatry*. 2001;62(suppl 16):5–9.
13. Fava M, Davidson KG. Definition and epidemiology of treatment-resistant depression. *Psychiatr Clin North Am*. 1996;19(2):179–200.
14. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. US Department of Health, Education, and Welfare, Public Health Service; 1976.
15. Busner J, Targum SD. The Clinical Global Impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. 2007;4(7):28–37.
16. Spielmans GI, McFall JP. A comparative meta-analysis of Clinical Global Impressions change in antidepressant trials. *J Nerv Ment Dis*. 2006;194(11):845–852.
17. Bobo WV, Angleró GC, Jenkins G, et al. Validation of the 17-item Hamilton Depression Rating Scale definition of response for adults with major depressive disorder using equipercentile linking to Clinical Global Impression scale ratings: analysis of Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) data. *Hum Psychopharmacol*. 2016;31(3):185–192.
18. Leucht S, Fennema H, Engel R, et al. What does the HAMD mean? *J Affect Disord*. 2013;148(2-3):243–248.
19. Leucht S, Fennema H, Engel RR, et al. What does the MADRS mean? equipercentile linking with the CGI using a company database of mirtazapine studies. *J Affect Disord*.

2017;210:287–293.
20. Turkoz I, Alphs L, Singh J, et al. Clinically meaningful changes on depressive symptom measures and patient-reported outcomes in patients with treatment-resistant depression. *Acta Psychiatr Scand*. 2021;143(3):253–263.
21. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23(1):56–62.
22. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), Clinician rating (QIDS-C), and Self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573–583.
23. Mrazek DA, Biernacka JM, O'Kane DJ, et al. CYP2C19 variation and citalopram response. *Pharmacogenet Genomics*. 2011;21(1):1–9.
24. Rush AJ, South C, Jain S, et al. Clinically significant changes in the 17- and 6-Item Hamilton Rating Scales for Depression: A STAR*D Report. *Neuropsychiatr Dis Treat*. 2021;17:2333–2345.
25. Dimidjian S, Hollon SD. How would we know if psychotherapy were harmful? *Am Psychol*. 2010;65(1):21–33.
26. Greden JF. The burden of recurrent depression: causes, consequences, and future prospects. *J Clin Psychiatry*. 2001;62(suppl 22):5–9.
27. Li J, Kuk AY, Rush AJ. A practical approach to the early identification of antidepressant medication non-responders. *Psychol Med*. 2012;42(2):309–316.
28. Henke RM, Zaslavsky AM, McGuire TG, et al. Clinical inertia in depression treatment. *Med Care*. 2009;47(9):959–967.
29. Gloster AT, Rinner MTB, Ioannou M, et al. Treating treatment non-responders: a meta-analysis of randomized controlled psychotherapy trials. *Clin Psychol Rev*. 2020;75:101810.
30. Leuchter AF, Cook IA, Hunter AM, et al. A new paradigm for the prediction of antidepressant treatment response. *Dialogues Clin Neurosci*. 2009;11(4):435–446.
31. Furukawa TA, Akechi T, Azuma H, et al. Evidence-based guidelines for interpretation of the Hamilton Rating Scale for Depression. *J Clin Psychopharmacol*. 2007;27(5):531–534.
32. Kolen MJ, Brennan RL. *Test Equating, Scaling, and Linking*. Springer Link; 2004.
33. Kolen MJ, Brennan RL. *Test Equating, Scaling, and Linking: Methods and Practices*. 3rd ed. Springer Science & Business Media; 2014.
34. Areán PA, Alvidrez J. Treating depressive disorders: who responds, who does not respond, and who do we need to study? *J Fam Pract*. 2001;50(6):E2.
35. Corey-Lisle PK, Nash R, Stang P, et al. Response, partial response, and nonresponse in primary care treatment of depression. *Arch Intern Med*. 2004;164(11):1197–1204.
36. Conway CR, Kumar A, Xiong W, et al. Chronic vagus nerve stimulation significantly improves quality of life in treatment-resistant major depression. *J Clin Psychiatry*. 2018;79(5):18m12178.
37. Bridges JF, Jones C. Patient-based health technology assessment: a vision of the future. *Int J Technol Assess Health Care*. 2007;23(1):30–35.
38. Sartorius N. Patient-reported outcomes in psychiatry. *Dialogues Clin Neurosci*. 2014;16(2):123–124.
39. Szegedi A, Jansen WT, van Willigenburg AP, et al. 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.
40. Kobak KA, Brown B, Sharp I, et al. Sources of unreliability in depression ratings. *J Clin Psychopharmacol*. 2009;29(1):82–85.
41. McKnight PE, Kashdan TB. The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. *Clin Psychol Rev*. 2009;29(3):243–259.
42. Zimmerman M, Posternak M, Friedman M, et al. Which factors influence psychiatrists' selection of antidepressants? *Am J Psychiatry*. 2004;161(7):1285–1289.
43. Husain SF, Yu R, Tang TB, et al. Validating a functional near-infrared spectroscopy diagnostic paradigm for major depressive disorder. *Sci Rep*. 2020;10(1):9740.
44. Athreya AP, Brückl T, Binder EB, et al. Prediction of short-term antidepressant response using probabilistic graphical models with replication across multiple drugs and treatment settings. *Neuropsychopharmacology*. 2021;46(7):1272–1282.
45. Athreya AP, Neavin D, Carrillo-Roa T, et al. Pharmacogenomics-driven prediction of antidepressant treatment outcomes: a machine-learning approach with multi-trial replication. *Clin Pharmacol Ther*. 2019;106(4):855–865.