Early Response to Antidepressant Medications in Adults With Major Depressive Disorder

A Naturalistic Study and Odds of Remission at 14 Weeks

Heather G. Belanger, PhD,[1,2] Christine Lee, PhD,[1] Zachary Poliacoff, MD,[2] Carina T. Gupta, MA,[1] and Mirène Winsberg, MD[1]

Abstract:

Purpose/Background: Early response after 2 to 4 weeks of antidepressant therapy has been shown to predict remission by 8 to 12 weeks. Most of the work to date on early response has been done using data from randomized controlled trials.

Methods/Procedures: This naturalistic study uses archival data from a national tele-mental health company. The positive and negative predictive values as well as sensitivity and specificity were calculated using different drops in baseline Patient Health Questionnaire 9 scores at various periods. Demographic and clinical characteristics were compared between early responders versus those lacking early response. Binary logistic regression analyses determined if early response was predictive of remission, response, and greater than minimal improvement at 14 weeks. For those who do not show early improvement, treatments were investigated using binary logistic regression to see if changes predicted later outcomes.

Findings/Results: Positive predictive values for all endpoints improved with the strength of early response but did not improve much with the time allowed for that response to occur. In contrast, negative predictive values increased substantially with time. Using a definition of 30% drop in Patient Health Questionnaire 9 score at week 4, 56.5% of patients were early responders. Early responders were ~3.2 times more likely to achieve remission than those lacking early response. Of nonresponders by week 4, those prescribed atypical antipsychotics (+SSRI) had significantly reduced odds of response at week 14, whereas those prescribed a norepinephrine and dopamine reuptake inhibitor had increased odds.

Implications/Conclusions: Early response may be associated with better outcomes at 14 weeks. In those with lack of response by week 4, patients prescribed a norepinephrine and dopamine reuptake inhibitor may achieve superior outcomes.

Key Words: early response, telehealth, depression, psychopharmacology

Major depressive disorder is one of the most prevalent,[1] high impact, and prevalent health disorders in the country. Whereas treatment with a single antidepressant (ATD) is a recommended first-line treatment,[2] ATD monotherapy produces an approximately 60% response rate and remission rates of around only 40%.[3]

From the [1]Brightside Health Inc, Oakland, CA; [2]Department of Psychiatry and Behavioral Neurosciences, University of South Florida, Tampa, FL.

Switching medications before 4 to 8 weeks has typically not been recommended, based on the idea that antidepressant response is often delayed. However, the optimal time to switch antidepressants is unclear and may vary depending on the ATD in question. For instance, patients often respond faster to tricyclic antidepressants (TCAs) and mirtazapine than selective serotonin reuptake inhibitors (SSRIs).[4,5]

Early response (ie, a drop of ≥20% from baseline depression severity score after 2 to 4 weeks of medication therapy) has been shown to predict remission at 8 to 12 weeks in recent meta-analytic summaries of the literature.[5,6] A review by Kudlow et al[7] revealed that roughly 1 in 5 patients with a lack of improvement by 4 weeks will have a response by 8 weeks and those without improvement in the first 2 to 4 weeks might need a change in treatment plan. Certain meta-analyses have found statistical separation between response to ATDs and placebo after 1 week.[8,9] Most meta-analyses have therefore focused on the predictive value of response after 2 weeks of treatment, with contradictory conclusions. Some have determined that lack of at least 20% improvement by 2 weeks is sufficient reason to switch.[10] However, efforts to calculate the positive predictive value (PPV) and negative predictive value (NPV) of at least 20% improvement by 2 weeks for response and remission over 12 weeks have indicated that predictive reliability varies substantially between agents.[5] Uher et al[4] concluded that, because of this variability in response patterns, it is only possible to effectively predict 12-week response after 8 weeks.

Delaying a switch until 6 to 12 weeks of treatment in nonresponders may be a mistake. Those who fail to respond early may be more likely to become treatment resistant.[11] In general, patients who fail to respond to an initial ATD are 12% to 38% less likely to respond to an alternative agent.[12–15] It is currently unclear if that is because certain individuals are more difficult to treat and/or depression evolves under the influence of failed treatment. If the latter is true, both the first-line treatment and subsequent iterations much be made more carefully and efficiently.

Most of the work to date on early response has been done using data from randomized controlled trials (RCTs), which may have limited utility for more heterogenous patients seen in clinic. A meta-analysis of naturalistic studies by Olgiati et al[6] summarized 9 studies revealing a significant association between early response and later response and remission. In those studies, typically, a single ATD was studied and/or there was a relatively small sample. The odds ratios (ORs) were estimated to be 3.3 for response and 2.1 for remission. These authors also looked at those who did not respond after 2 to 4 weeks and found that an early switching strategy at week 4 was inferior to a control scenario of maintenance. Inconsistent findings suggest the need for a large naturalistic study with multiple ATDs.

The overarching goal of the current study was to (1) explore the association between varying definitions of early improvement within the first several weeks of antidepressant treatment and ultimate remission, response, and greater than minimal improvement

(GTMI) during the initial 12 to 14 weeks of treatment in a naturalistic sample. Secondary goals were to (2) explore if any factors differentiate early improvers from nonearly improvers, (3) determine the relationship between early improvement and ultimate outcome, (4) provide clinicians with classification statistics for early response for different medication classes, and (5) determine if there are any ATDs (or combination of ATDs) that predict remission in the absence of early response. Finally, the early response literature to date, with one exception,[16] has relied upon on either the Hamilton Depression Rating Scale or the Montgomery-Asberg Depression Rating Scale.[6] The Patient Health Questionnaire 9 (PHQ-9) is another commonly used measure of depression symptom severity.[17] As such, another secondary goal of the current study was to examine the PPV and NPV in a naturalistic sample using the PHQ-9.

MATERIALS AND METHODS

Participants

Participant data used in the current investigation were obtained from a national mental health telehealth company (ie, [Brightside Health]) and consisted of 12,908 US-based adult patients, aged 18 to 82 years (mean age, 32.81 years; SD, 8.92 years), receiving psychiatric care for depression between October 2018 and January 2022. Participants were eligible to be included in analyses if they (*a*) were diagnosed with major depressive disorder by their provider, (*b*) had moderate to severe symptom severity at intake (PHQ-9, ≥10), (*c*) were prescribed a study medication (described hereinafter), and (*d*) had complete prescription data and complete outcome data at both 12 and 14 weeks. Patients at high risk for suicide and patients with psychosis or in need of emergency psychiatric services at the initial evaluation were not eligible.

Procedure

All study procedures were approved by the WCG Institutional Review Board for the retrospective analysis of patient data obtained by [Brightside Health] as part of routine clinical care. Enrolled [Brightside Health] patients complete an initial digital intake that includes clinically validated measures of depression and anxiety, as well as questions about clinical presentation, medical history, and demographics. All [Brightside Health] patients are required to complete baseline and intake questionnaires. During a patient's first session, a licensed professional prescribed psychiatric medication(s) for each patient. Over the course of treatment, patients communicate with their provider both asynchronously via messaging and synchronously via video telehealth sessions. [Brightside Health] also uses a measurement-based approach to tracking long-term outcomes by prompting patients to complete periodic assessments during treatment. Assessments were completed at baseline/intake and periodically thereafter. Surveys were administered digitally through an email prompt. Survey completion at start, 12 weeks, and endline (14 weeks) of study were required for participation.

Measures

The Patient Health Questionnaire 9 is a 9-item self-report measure used to assess the severity of depressive symptoms present within the prior 2 weeks as outlined by *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) criteria. Respondents rate items on a 4-point Likert scale (0–3), and total scores range from 0 to 27, with >9 indicating mild-to-low symptoms and 10+ indicating moderate-to-severe symptoms.[18] The PHQ-9 shows strong reliability, demonstrating 88% sensitivity and 88% specificity for major depressive disorder.[18] There is also evidence that the PHQ-9 can be used as a measure of antidepressant response.[19] On the

PHQ-9, a score of <5 typically signifies remission.[20] The PHQ-9 is a commonly used another commonly used measure of depression symptom severity[17] and is the primary measure of depression severity used at [Brightside Health].

Other variables measured at baseline included age, sex, education, race/ethnicity, employment status, income, prior episodes of depression (none, one, or more than one), duration of the current episode, prior ATD use (yes/no), comorbid anxiety disorders (to include anxiety disorder unspecified, generalized anxiety disorder, obsessive compulsive disorder, social anxiety disorder, panic disorder, acute stress disorder, or posttraumatic stress disorder), and total number of chronic health conditions endorsed (including arrhythmia, asthma, cancer, hypercholesterolemia, diabetes, heart condition, irritable bowel syndrome or Crohn disease, lung disease, obesity, thyroid disease, seizures, and chronic pain/fibromyalgia).

Interventions

Because this is a naturalistic sample, participants were prescribed a variety of medications. There were 7 different treatment groups. The most commonly prescribed medication category of the sample (63.7%) was SSRIs, followed by norepinephrine and dopamine reuptake inhibitors (NDRIs; 19.5%), serotonin-norepinephrine reuptake inhibitor (SNRI; 5.5%), trazodone (or trazodone-SSRI) (4.2%), SSRI and NDRI combination (3.9%), mirtazapine (or mirtazapine-SSRI) (1.8%), and atypical antipsychotics (atypicals-SSRI) (1.5%). The dosage of index antidepressants remained relatively consistent throughout the study period and were prescribed in standard therapeutic ranges. Dosage adjustments were made based on participant responses to the PHQ-9 and other assessments, as well as virtual visits between participation and providers. Because this was a naturalistic study, dosages were not controlled and therefore varied to meet individual needs. All medication groups were allowed to include other non-ATD (nonantidepressant) medications (eg, noncontrolled anxiolytics). For analyses examining 14-week outcomes, medication type had to remain consistent throughout that study period.

Outcomes

The primary outcome measure, the PHQ-9 score, was collected via self-report electronically at baseline, the first week, second week, fourth week, sixth week, and then at weeks 12 and 14. Standard definitions of *remission* (PHQ-9 <5 at both weeks 12 and 14)[20] and *response* (≥50% reduction in PHQ-9 score from baseline to week 14) were used. In addition, *GTMI* was defined as >30% reduction from baseline PHQ-9 score at week 14.[21]

Data Analyses

Data analyses were performed via SPSS, version 28, and Python using the statsmodels package and evaluated with sklearn. The PPV and NPV were calculated, such that participant outcomes were classified as true positive (TP, early response and remission), false positive (FP, early response but no remission), true negative (TN, no early response and no remission), and false negative (FN, no early response but remission). Positive predictive value and NPV were calculated as PPV = TP/(TP + FP) and NPV = TN/(TN + FN). Sensitivity was calculated as the ratio of true positive outcomes to the total number of patients achieving remission (sensitivity = TP/(TP + FN)), whereas specificity was calculated as the ratio of true negative outcomes to the total number of patients not achieving remission (specificity = TN/(TN + FP)). Area under the curve (AUC) values were calculated to examine classification performance.

To explore the effects of different drops in baseline PHQ-9 score and the assessment timeframe on PPVs and NPVs, the PPVs and NPVs for multiple percentage drops (10%, 20%, 30%, 40%, and 50%) and at various periods (weeks 1, 2, 4, and 6) were

calculated. Next, using a definition of 30% drop in PHQ-9 score at week 4, the classification accuracy statistics for predicting remission and response were calculated for general categories of ATDs, as well as specific ATDs.

Demographic and clinical characteristics were compared between those who were early responders versus not, using 30% drop at week 4 to define early responders. χ^2 Analyses compared relative proportions between groups for categorical variables (eg, education level, race/ethnicity, sex, etc), whereas *t* tests compared the groups on continuous measures (ie, age, baseline PHQ and General Anxiety Disorder scores, and number of medical comorbidities). Effect sizes were calculated using Cohen d[22] for continuous variables and Cramer V[23] for categorical variables. Only those effects that are 0.2 or greater will be presented in the Results section. Follow-up tests after significant χ^2 tests were conducted using adjusted residuals and Bonferroni corrections. Using the same group (ie, 30% drop at week 4), we calculated sensitivity, specificity, PPV, NPV, and ORs for each medication group that was sustained on that medication(s) for the full study duration.

Logistic regression was used to examine the independent effect of early improvement on predicting remission. First, univariate logistic regression models were run to determine which variables are associated with remission. Those significant predictors (that did not correlate 0.2 or greater with each other) were then added to a multivariate model to determine the unique contribution of each, controlling for all other variables.[24] Binary logistic regression analyses were used to determine if early response was predictive of remission, response, and GTMI.

Finally, for those who do not show early improvement but nonetheless went on to have remission, response, or GTMI, treatment group was investigated using binary logistic regression to see if treatment predicted later outcome. This was done using the same seven treatment groups as previous (and requiring no changes in medication regimen over the first 12 weeks of treatment).

All logistic regression model results reported are leave-one-out (LOO) cross-validation.

Leave-one-out cross-validation consists of holding out one sample for validation and training a model on the rest, repeatedly until all samples have been validated on. Model properties (ie, coefficients, *P* values, etc) are reported as means across all LOO models, and performance (ie, AUC and accuracy) are calculated on the combination of all LOO validation results.

RESULTS

Predictive Value of Varying Definitions of Early Response

There were 2814 cases of remission (21.8% of the sample), 7333 cases (56.8%) with at least a response (≥50% reduction from initial PHQ-9 score), and 9570 cases (74.1%) of at least a GTMI (≥50% reduction from initial PHQ-9 score). To determine the optimal definition of "early response" as a predictor of later remission, response, and GTMI; the effects of different drops in baseline PHQ-9 score; and assessment timeframes on PPVs and NPVs, the PPVs and NPVs for multiple percentage drops (10%, 20%, 30%, 40%, and 50%) and at various periods (weeks 1, 2, 4, and 6) were calculated. These results are presented in Figure 1. Positive predictive values for all endpoints improved with the strength of early response but did not meaningfully improve with the time allowed for that response to occur. Negative predictive values increased substantially with time—the longer patients went without an "early" response, the greater the NPV. The strength of response at each time point had a much milder effect on NPV. In all cases, lack of response by week 6 was substantially more predictive of a lack of endpoint

response or remission than lack of response by week 4, which in turn had a higher NPV than lack of response by week 2. Negative predictive values for remission were uniformly high. Based on these results (and the best trade-off between psychometric properties and maximizing time efficiency), we determined that test statistics would be optimized by defining "early response" as achieving ≥30% reduction in initial PHQ-9 score by week 4.

Comparison of Baseline Characteristics of Early Responders Versus Nonearly Responders

Using a definition of 30% drop in PHQ-9 score at week 4, 56.5% of all patients were early responders. Table 1 shows the characteristics of those who had early response and those who did not. All the significant differences between groups had negligible effect sizes (ie, <0.2).

Predictive Value of Early Response

Logistic regression models were created to ascertain the effects of age, prior depressive episodes, duration of depression, ATD naivety, baseline PHQ, and early response on likelihood of remission, response, and GTMI at week 14. Other variables collected at baseline were not entered (ie, income, race/ethnicity, education, and number of chronic medical conditions) because of being highly correlated with age ($r > 0.2$) or because of not being significant in univariate models (ie, comorbid anxiety disorder, employment, duration of depression, and sex). Results are presented in Table 2. The logistic regression model predicting remission was significant ($\chi^2 = 1227.92$, $P < 0.001$), with 78.4% accuracy and an AUC of 0.71. The early response group was about 3.2 times more likely to remit than the group with no early response, controlling for the other variables. Results were similar for response ($\chi^2 = 1382.58$; $P < 0.001$; accuracy, 66.1%; AUC, 0.68) and GTMI outcomes ($\chi^2 = 1523.72$; $P < 0.001$; accuracy, 74.2%; AUC, 0.71)—the early response group was 3.6 times more likely to show response at week 14 and were 4.9 times more likely to show GTMI than those who did not show an early response.

Classification Statistics for Early Response by Medication Type

The classification accuracy statistics for predicting remission and response were calculated for general categories of ATDs, as well as specific ATDs, and are presented in Table 3. As can be seen, early improvement predicted later remission and response with high sensitivity and moderate specificity as well as low PPV but high NPV. Positive predictive values of early response for remission ranged from 23% to 38% and were 30% for all ATDs combined. Negative predictive values of failure to achieve early response for eventual failure to achieve remission ranged from 78% to 92% and was 88% for all ATDs combined. Sensitivities for individual ATDs ranged from a high of 88% for duloxetine and citalopram to a low of 69% for the combination of mirtazapine with an SSRI. Specificities for individual ATDs ranged from a high of 54% for atypical antipsychotics (+SSRI) to a low of 33% for NDRIs (bupropion). Treatment with atypical antipsychotics (+SSRI) led to the highest OR for remission (OR, 5.96; 2.10–16.90) and response (OR, 10.21; 4.20–24.84), although the number of patients prescribed that category was low (n = 131). Comparing SSRIs, citalopram had the highest OR for remission (OR, 4.02; 1.46–11.11), but sertraline had the highest OR for predicting response (OR, 3.86; 3.17–4.71). Comparing different SNRIs, duloxetine had the highest predictive value for remission (OR, 5.61; 2.49–12.60) and response (OR, 6.91; 3.83–12.48).
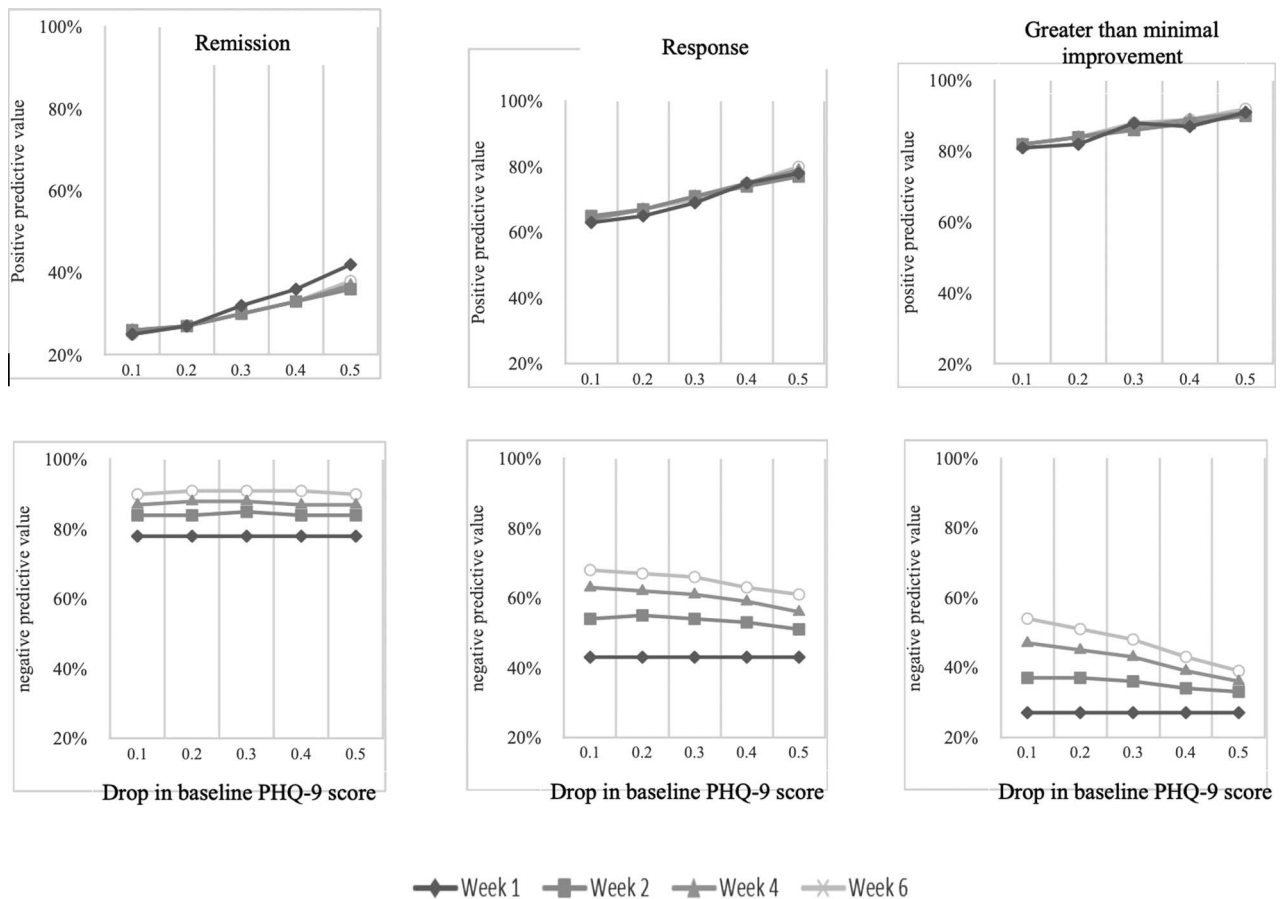
FIGURE 1. Positive predictive values and negative predictive values based on percentage drop from baseline PHQ-9 score over different time periods for remission, response, and greater than minimal improvements.

Predictive Value of Treatment Group in Nonearly Response Patients

Of the 5465 patients (43.5% of original sample size) who did not have early response by week 4, 1694 patients (30%) were excluded from analysis because of missing treatment group. Of the 3771 included patients, there were 523 (13.9%) who nonetheless went on to have remission by week 14, 1538 (40.8%) who went on to have a response, and 2165 (57.4%) who had GTMI.

Logistic regression models were created to determine if treatment group, along with age, prior depressive episodes, duration of depression, ATD naivety, and baseline PHQ, predicted later outcome in this subset of patients. The logistic regression model predicting remission was significant ($\chi^2 = 205.01$; $P < 0.001$; accuracy, 86.1%; AUC, 0.68), although no treatment group significantly affected likelihood of remission, with the SSRI group as reference (treatment group SNRI: $\chi^2 = 3.07$, $P = 0.080$; treatment group NDRI: $\chi^2 = 0.77$, $P = 0.380$; treatment group atypical-SSRI: $\chi^2 = 1.51$, $P = 0.219$; treatment group mirtazapine: $\chi^2 = 1.15$, $P = 0.284$; treatment group trazodone: $\chi^2 = 0.05$, $P = 0.825$; treatment group SSRI-SNRI: $\chi^2 = 0.00$, $P = 0.952$).

The logistic regression model predicting response was significant but with low validation performance ($\chi^2 = 469.06$; $P < 0.001$; accuracy, 58.8%; AUC, 0.56). Treatment group significantly affected the odds of achieving response (treatment group SNRI: $\chi^2 = 0.19$, $P = 0.660$; treatment group NDRI: $\chi^2 = 5.13$, $P = 0.024$; treatment group atypical-SSRI: $\chi^2 = 6.83$, $P = 0.009$; treatment group mirtazapine: $\chi^2 = 0.27$, $P = 0.870$; treatment group trazadone: $\chi^2 = 2.62$, $P = 0.106$; treatment group SSRI-SNRI: $\chi^2 = 0.11$, $P = 0.739$). Outcomes were worse for the group prescribed atypical antipsychotics (+SSRI) and better for those prescribed NDRIs. Specifically, the group prescribed atypical antipsychotics (+SSRI) had significantly reduced odds of achieving a response (OR, 0.475 [0.3, 0.8]), whereas the NDRI group had 1.2 times greater likelihood of response (OR, 1.22 [1.0, 1.5]).

The logistic regression model predicting GTMI was significant but with low validation performance ($\chi^2 = 39.03$; $P < 0.001$; accuracy, 57.4%; AUC, 0.54), and no treatment groups significantly affected likelihood of remission (treatment group SNRI: $\chi^2 = 0.30$, $P = 0.585$; treatment group NDRI: $\chi^2 = 1.16$, $P = 0.212$; treatment group atypical-SSRI: $\chi^2 = 2.74$, $P = 0.098$; treatment group mirtazapine: $\chi^2 = 0.01$, $P = 0.946$; treatment group trazadone: $\chi^2 = 2.08$, $P = 0.149$; treatment group SNRI-SSRI: $\chi^2 = 0.00$, $P = 0.965$).

DISCUSSION

Early Responders Versus Nonearly Responders

This is the largest naturalistic study of early response to date and adds to the literature base of naturalistic studies in the treatment of depression. In the current outpatient telehealth sample, the early response group was about 3.2 times more likely to have remission at week 14. Results were similar for response and GTMI—early responders were 3.6 times more likely to show response at week 14

TABLE 1. Characteristics of Early Responders and Nonearly Responders

Characteristic	Early Responders	Nonearly Responders	t or χ^2	Effect Size*	P
Age	33.11	32.53	−3.55	0.06	<0.001
% Female	70.8%	66.8%	23.12	0.04	<0.001
Education			12.17	0.03	0.016
No high school	1.2%	1.7%			
High school diploma	30.0%	44.2%			
Some college	14.2%	13.9%			
College degree	36.6%	37.0%			
Graduate degree	18.0%	16.4%			
Race/ethnicity			17.67	0.04	0.007
White/Caucasian	78.4%	77.2%			
Asian	3.3%	4.6%			
Hispanic	8.0%	8.4%			
Black/African American	4.2%	4.4%			
Native American	0.4%	0.5%			
Pacific Islander	0.3%	0.3%			
Other	5.4%	4.7%			
Employed			2.19	0.01	0.206
Full time	68.7%	68.9%			
Part time	11.5%	10.7%			
Unemployed	19.8%	20.3%			
Annual income			11.94	0.03	0.008
<$30,000	28.9%	31.0%			
$30–60,000	32.3%	29.7%			
$60–100,000	20/9%	20.8%			
>$100,000	17.9	18.5%			
Prior episodes of depression			27.90	0.05	<0.001
None	32.5%	36.9%			
One	12.7%	10.8%			
More than one	54.8%	52.3%			
Comorbid anxiety disorder	46.4%	50.5%	20.90	0.04	<0.001
No. chronic medical conditions	0.52	0.56	2.42	0.04	0.015
Baseline PHQ-9	18.25	18.02	−2.87	0.05	0.004
Baseline GAD-7	14.76	14.91	1.76	0.03	0.079
How long depressed			10.01	0.03	0.040
Less than 2 wk	1.2%	1.0%			
2 wk to 2 mo	12.3%	10.8%			
2 mo to 1 y	27.5%	26.9%			
1 to 2 y	17.1%	17.5%			
More than 2 y	41.8%	43.7%			
ATD naive	49.9%	47.1%	9.51	0.03	0.002
Initial prescription			43.06	0.06	<0.001
SSRI	62.7%	64.9%			
SNRI	5.1%	6.0%			
NDRI	21.3%	17.2%			
Atypical-SSRI	1.3%	1.7%			
Mirtazapine-SSRI or SNRI	1.6%	2.0%			
Trazodone-SSRI or SNRI	4.2%	4.1%			
SSRI and SNRI	3.8%	4.1%			

*Effect sizes are Cohen *d* for continuous variables and Cramer *V* for categorical variables. Effect sizes are interpreted as small (0.2), medium (0.5), and large (0.8).[23] Mean values are presented for continuous variables, and frequency counts are presented (with %) for categorical variables. Follow-up tests revealed no significant differences between the groups on education level or how long depressed. Early responders were more likely to be female, in the income group of $30 to $60 k, in the NDRI group, have greater depression severity at baseline, and fewer number of chronic medical conditions. They are less likely to have a comorbid anxiety disorder and less likely to be Asian.

GAD-7 indicates General Anxiety Disorder 7.

TABLE 2. Logistic Regression Model Predicting Stable Remission, Response, and GTMI

Variable	β (SE)	OR	95% Confidence Interval	*P*
Remission				
Baseline PHQ-9 score	−0.120 (0.005)	0.887	0.878–0.896	<0.001
Age	0.017 (0.002)	1.017	1.012–1.022	<0.001
Prior depression (one)	0.121 (0.075)	1.129	0.975–1.308	0.106
Prior depression (several)	0.043 (0.049)	1.044	0.949–1.150	0.375
Duration (2 wk to 2 mo)	−0.288 (0.128)	0.749	0.583–0.964	0.025
Duration (2 to 12 mo)	−0.483 (0.121)	0.617	0.487–0.781	<0.001
Duration (1 to 2 y)	−0.617 (0.125)	0.534	0.422–0.690	<0.001
Duration (>2 y)	−0.708 (0.117)	0.492	0.392–0.619	<0.001
ATD naive	0.113 (0.046)	1.120	1.024–1.225	0.013
Early response	1.178 (0.050)	3.249	2.945–3.585	<0.001
Response				
Baseline PHQ-9 score	−0.009 (0.004)	0.991	0.983–0.999	0.033
Age	0.014 (0.002)	1.015	1.010–1.019	<0.001
Prior depression (one)	0.216 (0.068)	1.241	1.087–1.417	0.001
Prior depression (several)	0.116 (0.041)	1.123	1.037–1.216	0.004
Duration (2 wk to 2 mo)	−0.703 (0.119)	0.495	0.392–0.625	<0.001
Duration (2 to 12 mo)	−0.740 (0.112)	0.477	0.383–0.594	<0.001
Duration (1 to 2 y)	−0.905 (0.115)	0.404	0.323–0.506	<0.001
Duration (>2 y)	−0.977 (0.108)	0.376	0.305–0.465	<0.001
ATD naive	0.114 (0.038)	1.121	1.039–1.209	0.003
Early response	1.281 (0.038)	3.601	3.342–3.881	<0.001
GTMI				
Baseline PHQ-9 score	0.009 (0.005)	1.009	0.999–1.019	0.054
Age	0.015 (0.002)	1.015	1.010–1.020	<0.001
Prior depression (one)	0.253 (0.079)	1.288	1.103–1.505	0.001
Prior depression (several)	0.140 (0.046)	1.151	1.051–1.260	0.002
Duration (2 wk to 2 mo)	−0.447 (0.136)	0.640	0.490–0.834	0.001
Duration (2 to 12 mo)	−0.391 (0.127)	0.676	0.527–0.868	0.002
Duration (1 to 2 y)	−0.536 (0.130)	0.585	0.453–0.755	<0.001
Duration (>2 y)	−0.597 (0.123)	0.550	0.433–0.700	<0.001
ATD naive	0.149 (0.044)	1.161	1.065–1.265	<0.001
Early response	1.158 (0.045)	4.876	4.466–5.323	<0.001

For duration of depression, "2 weeks or less" was the reference group; for prior depression, "none" was the reference group.

and were 4.9 times more likely to show GTMI than those who did not show an early response.

Other naturalistic studies have found that partial improvement of symptoms in the first 4 weeks of treatment resulted in 2 times greater likelihood for remission at 6 to 14 weeks (OR, 2.1) and roughly 3 times greater likelihood of response (OR, 3.3).[6] In RCTs, those with 20% to 25% reduction in depressive symptom severity after 2 weeks were 8 times more likely to show later response and 6 times more likely to have remission.[5] Collectively, there is thus convincing evidence that, regardless of setting, early improvement is a significant predictor of later outcome. Lower predictive values in naturalistic studies are likely due to a variety of factors, including uncontrolled environmental and patient factors.

Our PPV and NPVs, across a variety of definitions of "early," were remarkably similar to those found in a randomized trial of antidepressant medications in military veterans[21] in that NPVs were uniformly high for remission, whereas PPVs were lower for remission but higher for response and GTMI. They found that a ≥20% drop in depressive symptom severity at week 2 resulted in a 38% PPV and 97% NPV for remission at week 14, whereas we found PPV of 27% and NPV of 84% using the same parameters. In our

data, our highest NPVs were 90% to 91% for early response measured at week 6. In our sample, arguably, the best trade-off between psychometric properties and maximizing time efficiency occurred using a 30% drop in PHQ-9 score by week 4, with a PPV of 30% and NPV of 88%. As others have pointed out,[16,21,25] this NPV is informative to clinicians who can infer that an absence of such early improvement predicts a lack of remission in 88% of patients. The NPVs were remarkably high, even for "early response" defined at week 1, again underscoring the utility of the lack of early response. Given that ATDs are thought to exert their influence on symptom severity over the course of a few weeks[26] or more,[27] the existence of very early responders (ie, 1 week) may give credence to the theory that antidepressants induce changes in the processing of emotional stimuli very early in the course of treatment[28,29] and/or that certain people have a genetic predisposition to early response[30] such as polymorphism of serotonin 2A receptors, which may produce variability between individuals in the timing of response. Given that there was no placebo control, very early responders could also be experiencing placebo effects.

Like Fabbri et al,[31] we found that early response was associated with greater depressive symptom severity at baseline, although

TABLE 3. Predictive Value of Early Improvement on Response and Remission at Endpoint

	n	Sensitivity, %	Specificity, %	PPV, %	NPV, %	OR (95% CI)
Remission						
All ATDs	12,575	76	49	30	88	3.266 (2.959–3.605)
SSRI*	5918	77	47	33	86	3.264 (2.838–3.754)
Escitalopram	2838	82	37	31	85	3.439 (2.811–4.208)
Sertraline	1861	81	37	30	85	3.080 (2.402–3.949)
Fluoxetine	1018	81	39	27	88	3.134 (2.212–4.441)
Citalopram	115	88	36	37	88	4.617 (1.600–13.330)
Paroxetine	69	84	48	38	89	2.089 (0.508–8.588)
SNRI*	494	83	43	25	91	4.347 (2.506–7.540)
Venlafaxine	250	75	44	20	90	3.266 (2.959–3.605)
Duloxetine	238	88	41	29	93	6.032 (2.641–13.776)
NDRI	1856	83	33	31	85	2.793 (2.173–3.590)
Atypical (+SSRI)	131	81	54	31	92	6.920 (2.269–21.105)
Mirtazapine (+SSRI)	145	69	35	25	78	2.218 (0.939–5.241)
Trazodone (+SSRI)	347	75	36	23	84	2.137 (1.190–3.837)
SSRI-SNRI	344	79	41	30	86	3.014 (1.712–5.305)
Response						
All ATDs	12,575	70	61	71	61	3.643 (3.380–3.927)
SSRI*	5918	71	60	72	59	3.706 (3.318–4.139)
Escitalopram	2838	72	59	72	60	3.814 (3.247–4.479)
Sertraline	1861	72	60	72	60	3.832 (3.142–4.674)
Fluoxetine	1018	69	61	71	60	3.564 (2.736–4.642)
Citalopram	115	69	49	74	43	2.560 (1.069–6.130)
Paroxetine		66	60	74	50	2.215 (0.694–7.2074)
SNRI*	494	70	68	75	62	4.992 (3.365–7.407)
Venlafaxine	250	66	68	75	58	3.643 (3.380–3.927)
Duloxetine	238	75	67	75	68	6.599 (3.611–12.062)
NDRI	1856	74	53	73	54	3.263 (2.667–3.991)
Atypical (+SSRI)	131	72	75	70	76	12.167 (4.650–31.835)
Mirtazapine (+SSRI)	145	68	59	68	59	3.464 (1.617–7.442)
Trazodone (+SSRI)	347	75	64	73	66	5.230 (3.263–8.383)
SSRI*-SNRI*	344	68	60	67	61	3.290 (2.088–5.184)

All ATDs includes all possible combinations of ATDs without necessarily staying the same through the study period. For other groups, medication type had to remain consistent throughout that study period. Odds ratio was calculated with age, baseline PHQ-9 score, duration of depression, ATD naivety, and prior depressive episodes.

*Overarching group includes medications not examined separately because of low numbers (fluvoxamine, trintellix, viibryd, desvenlafaxine).

CI indicates confidence interval.

that study compared early to late responders. Although other characteristics significantly differentiated early responders (ie, female sex, older age, fewer chronic medical conditions, and NDRI treatment), in our sample, the effect sizes were negligible and therefore unlikely to be of much clinical utility. Wagner et al[5] found that patients treated with TCA or mirtazapine had a higher likelihood of early improvement than those treated with SSRI or monoamine oxidase inhibitors (MAO) inhibitor. In our sample, those treated with NDRI were more likely to show early improvement, although, again, the effect size was very small. We did not have TCA or MAO-inhibitor treatment groups, so comparisons are difficult, and unlike our sample, the studies included in the Wagner et al[5] were RCTs.

Predicting Remission

Using the definition of remission as achieving a PHQ-9 score of <5 at weeks 12 and 14, early response had a PPV of 30% and an NPV of 88% across all antidepressants. That is, being a nonearly

responder was a good indication that the patient would not achieve remission, but early response could not accurately predict whether a patient would achieve remission. Across agents, PPVs were poor, in the range of 23% to 38%. For individual agents, duloxetine, atypical antipsychotics (+SSRI), and venlafaxine all had NPVs of 90% or greater, and mirtazapine-SSRI had the lowest NPV at 78%.

Although the NPVs for lack of early response predicting lack of remission are high, it is important to consider the clinical situation at hand before using this value to guide treatment choices. For instance, it may not always be appropriate to aim for remission—for patients with a recurrent pattern of depression or who suffer from treatment resistant depression, response or even GTMI may be a more appropriate treatment goal. On the other hand, patients suffering from their first episode of depression or those who are medication naive might benefit more from switching agents if they do not respond early, given that remission might be a more realistic treatment goal in these populations. Furthermore, only 12% of patients who did not respond early went on to achieve remission; if only about

1 in 10 patients who do not respond early eventually remit, it may not be worthwhile to continue treatment with that agent.

Predicting Response

Using our definition of response as greater than 50% reduction in initial PHQ-9 score by week 12, early response had a PPV of 71% and NPV of 61% across all antidepressants. Compared with early response's ability to predict remission, early improvement was better able to predict the response outcome, but lack of early improvement was less able to predict lack of response by our endpoint. This is in both cases likely due to the lower requirements for the response endpoint compared with the remission endpoint. The range of values for response PPV had a fairly precise distribution (67%–75%) that was also tighter than that for remission PPV (20%–38%), whereas the range of values for response NPV was broader (43% for citalopram to 76% for treatment with an atypical antipsychotic-SSRI) than the NPV values for remission (78%–93%). Clinically, there may not be much value in an NPV of 61%, which is only a slight performance improvement over random chance that in most cases probably is not enough to justify changes to a treatment plan. However, there is modest value in a PPV of 71%; if there is early response, it is probably worthwhile to continue treatment rather than switch.

Nonearly Responders

Of the patients who did not have early response by week 4, 13.9% nonetheless went on to have remission by study endpoint, 40.8% went on to have a response, and 57.4% had GTMI. Comparatively, in the RCT of military veterans cited earlier, Hicks et al[21] noted that, of the patients who did not demonstrate early improvement after 2 weeks of antidepressant therapy, approximately 7.4% went on to have remission by week 12. Other studies have found variable rates of remission after a lack of early improvement, with rates between 13% and 30%.[32,33]

Among nonearly responders, those prescribed atypical antipsychotics (+SSRI) had significantly reduced odds of achieving a response at week 14. Given concerns about adverse side effects with atypicals (not assessed in the current study),[34] these results raise further questions about the use of atypicals for nonpsychotic depression. Clearly, these results need to be replicated in an independent sample.

Limitations and Strengths

There are both limitations and strengths of this study. Because it relied on retrospective, clinical data, there was no way to control the treatment groups or ensure that dosing, timing, and other factors were controlled. Because it was not a controlled study, patients could have been taking other medications or substances outside of their treatment with Brightside that could have influenced outcome (eg, controlled anxiolytics), and it is not known why any particular ATD was prescribed. Similarly, there is no way to ascertain the extent of any placebo effect without a control group. In addition, although anxiety disorders were considered, other psychiatric comorbidities were not considered because of low numbers. Finally, it is important to note that, comparing different ATDs is difficult because the efficacy of any drug depends on many factors, like dose or titration, as well as individual differences in metabolism and compliance. These factors were not measured or controlled for in this study. On the other hand, results of this study, which had modest NPV/PPV, might better generalize to clinical practice than RCTs.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are not available.

REFERENCES

1. Murray CJ, Atkinson C, Bhalla K, et al. The state of US health, 1990–2010: burden of diseases, injuries, and risk factors. *JAMA*. 2013;310:591–608.

2. Davidson JR. Major depressive disorder treatment guidelines in America and Europe. *J Clin Psychiatry*. 2010;71(suppl E1):e04.

3. Henssler J, Kurschus M, Franklin J, et al. Trajectories of acute antidepressant efficacy: how long to wait for response? A systematic review and meta-analysis of long-term, placebo-controlled acute treatment trials. *J Clin Psychiatry*. 2018;79:17r11470.

4. Uher R, Mors O, Rietschel M, et al. Early and delayed onset of response to antidepressants in individual trajectories of change during treatment of major depression: a secondary analysis of data from the Genome-Based Therapeutic Drugs for Depression (GENDEP) study. *J Clin Psychiatry*. 2011;72:1478–1484.

5. Wagner S, Engel A, Engelmann J, et al. Early improvement as a resilience signal predicting later remission to antidepressant treatment in patients with major depressive disorder: systematic review and meta-analysis. *J Psychiatr Res*. 2017;94:96–106.

6. Olgiati P, Serretti A, Souery D, et al. Early improvement and response to antidepressant medications in adults with major depressive disorder. Meta-analysis and study of a sample with treatment-resistant depression. *J Affect Disord*. 2018;227:777–786.

7. Kudlow PA, McIntyre RS, Lam RW. Early switching strategies in antidepressant non-responders: current evidence and future research directions. *CNS Drugs*. 2014;28:601–609.

8. Papakostas GI, Fava M. A meta-analysis of clinical trials comparing milnacipran, a serotonin—norepinephrine reuptake inhibitor, with a selective serotonin reuptake inhibitor for the treatment of major depressive disorder. *Eur Neuropsychopharmacol*. 2007;17:32–36.

9. Taylor MJ, Freemantle N, Geddes JR, et al. Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Arch Gen Psychiatry*. 2006;63:1217–1223.

10. 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;70:344–353.

11. Kinrys G, Gold AK, Pisano VD, et al. Tachyphylaxis in major depressive disorder: a review of the current state of research. *J Affect Disord*. 2019; 245:488–497.

12. Insel TR. Beyond efficacy: the STAR*D trial. *Am J Psychiatry*. 2006;163:5–7.

13. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163:28–40.

14. Hierholzer R. Remission rates for depression in STAR*D study. *Am J Psychiatry*. 2006;163:1293; author reply 1293-1294.

15. Taylor DM, Barnes TRE, Young AH. *The Maudsley Prescribing Guidelines in Psychiatry*. 13th ed. New York, NY: Wiley-Blackwell; 2018.

16. Fowler JC, Patriquin M, Madan A, et al. Early identification of treatment non-response utilizing the Patient Health Questionnaire (PHQ-9). *J Psychiatr Res*. 2015;68:114–119.

17. Siu AL, , US Preventive Services Task Force (USPSTF), Bibbins-Domingo K, et al. Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*. 2016;315:380–387.

18. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606–613.

19. Lowe B, Schenkel I, Carney-Doebbeling C, et al. Responsiveness of the PHQ-9 to psychopharmacological depression treatment. *Psychosomatics*. 2006;47:62–67.

20. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Ann*. 2002;32:509–521.

21. Hicks PB, Sevilimedu V, Johnson GR, et al. Predictability of nonremitting depression after first 2 weeks of antidepressant treatment: a VAST-D trial report. *Psychiatr Res Clin Pract*. 2019;1:58–67.

22. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.

23. Rea LM, Parker RA. *Designing and Conducting Survey Research*. San Francisco, CA: Jossey-Bass; 1992.

24. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: Wiley; 2000.

25. Kudlow PA, Cha DS, McIntyre RS. Predicting treatment response in major depressive disorder: the impact of early symptomatic improvement. *Can J Psychiatry*. 2012;57:782–788.

26. Posternak MA, Zimmerman M. Is there a delay in the antidepressant effect? A meta-analysis. *J Clin Psychiatry*. 2005;66:148–158.

27. Quitkin FM, Rabkin JD, Markowitz JM, et al. Use of pattern analysis to identify true drug response. A replication. *Arch Gen Psychiatry*. 1987;44: 259–264.

28. Browning M, Kingslake J, Dourish CT, et al. Predicting treatment response to antidepressant medication using early changes in emotional processing. *Eur Neuropsychopharmacol*. 2019;29:66–75.

29. Spies M, Kraus C, Geissberger N, et al. Default mode network deactivation during emotion processing predicts early antidepressant response. *Transl Psychiatry*. 2017;7:e1008.

30. Sun Y, Tao S, Tian S, et al. Serotonin 2A receptor polymorphism rs3803189 mediated by dynamics of default mode network: a potential biomarker for antidepressant early response. *J Affect Disord*. 2021;283:130–138.

31. Fabbri C, Marsano A, Balestri M, et al. Clinical features and drug induced side effects in early versus late antidepressant responders. *J Psychiatr Res*. 2013;47:1309–1318.

32. Soares CN, Fayyad RS, Guico-Pabia CJ. Early improvement in depressive symptoms with desvenlafaxine 50 mg/d as a predictor of treatment success in patients with major depressive disorder. *J Clin Psychopharmacol*. 2014; 34:57–65.

33. Henkel V, Seemuller F, Obermeier M, et al. Does early improvement triggered by antidepressants predict response/remission? Analysis of data from a naturalistic study on a large sample of inpatients with major depression. *J Affect Disord*. 2009;115:439–449.

34. Rothschild AJ. Should antipsychotic medications be prescribed to patients with nonpsychotic depression? *J Clin Psychopharmacol*. 2022;42:231–233.