Trajectories of Depression Severity in Clinical Trials of Duloxetine

Insights Into Antidepressant and Placebo Responses

Ralitza Gueorguieva, PhD; Craig Mallinckrodt, PhD; John H. Krystal, MD

Context: The high percentage of failed clinical trials in depression may be due to high placebo response rates and the failure of standard statistical approaches to capture heterogeneity in treatment response.

Objective: To assess whether growth mixture modeling can provide insights into antidepressant and placebo responses in clinical trials of patients with major depression.

Design: We reanalyzed clinical trials of duloxetine to identify distinct trajectories of Hamilton Scale for Depression (HAM-D) scores during treatment. We analyzed the trajectories in the entire sample and then separately in all active arms and in all placebo arms. Effects of duloxetine hydrochloride, selective serotonin reuptake inhibitor (SSRI), and covariates on the probability of following a particular trajectory were assessed. Outcomes in different trajectories were compared using mixed-effects models.

Setting: Seven randomized double-blind clinical trials of duloxetine vs placebo and comparator SSRI.

Patients: A total of 2515 patients with major depression.

Interventions: Duloxetine and comparator SSRI.

Main Outcome Measure: Total score on the HAM-D.

Results: In the entire sample and in the antidepressant-treated subsample, we identified trajectories of responders (76.3% of the sample) and nonresponders (23.7% of the sample). However, placebo-treated patients were characterized by a single response trajectory. Duloxetine and SSRI did not differ in efficacy, and compared with placebo they significantly decreased the odds of following the nonresponder trajectory. Antidepressant responders had significantly better HAM-D scores over time than placebo-treated patients, but antidepressant nonresponders had significantly worse HAM-D scores over time than the placebo-treated patients.

Conclusions: Most patients treated with serotonergic antidepressants showed a clinical trajectory over time that is superior to that of placebo-treated patients. However, some patients receiving these medications did more poorly than patients receiving placebo. These data highlight the importance of ongoing monitoring of medication risks and benefits during serotonergic antidepressant treatment. They should further stimulate the search for biomarkers or other predictors of responder status in guiding antidepressant treatment.

Trial Registration: Clinicaltrials.gov Identifier: NCT00073411

Arch Gen Psychiatry. 2011;68(12):1227-1237
analyses of duloxetine studies 11,25,26 have addressed point analyses.10,12,13However, they assume the same mean patients, reduce bias, and improve signal detection over end-
mixture models15-17) allow identification of distinct classes over traditional analytic methods.

trajectory-based analyses improve signal detection they provide new insights into the nature of antide-
pressant clinical trials. Herein, we apply important questions that required larger sample sizes
than the typical clinical trial. Herein, we apply

cells to the active arms and to the placebo arms. The latter analy-
ses were used to evaluate whether different classes would emerge for patients in the active arms and in the placebo arms. The latter analy-

Duloxetine hydrochloride is a Food and Drug Admin-
istration–approved antidepressant medication with dem-
strated efficacy for major depressive disorder.23-25Pooled
analyses on only 2 of the protocols (HMAT and HMAY). These
dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These
dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

dose adjustment varied across trials, we could perform dose
analyses on only 2 of the protocols (HMAT and HMAY). These

T able 1. Duloxetine Hydrochloride Trials for Patients With Major Depression

<table>
<thead>
<tr>
<th>Protocola and Trial</th>
<th>Total Sample Size</th>
<th>Treatment Arms and Sample Size per Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMAQ</td>
<td></td>
<td>Duloxetine (20-60 mg Twice daily)</td>
</tr>
<tr>
<td>Part A</td>
<td>173</td>
<td>70</td>
</tr>
<tr>
<td>Part B</td>
<td>193</td>
<td>82</td>
</tr>
<tr>
<td>HMAT</td>
<td></td>
<td>Duloxetine (20 mg Twice Daily)</td>
</tr>
<tr>
<td>Part A</td>
<td>354</td>
<td>91</td>
</tr>
<tr>
<td>Part B</td>
<td>353</td>
<td>86</td>
</tr>
<tr>
<td>HMAY</td>
<td></td>
<td>Duloxetine (40 mg Twice Daily)</td>
</tr>
<tr>
<td>Part A</td>
<td>367</td>
<td>95</td>
</tr>
<tr>
<td>Part B</td>
<td>392</td>
<td>93</td>
</tr>
<tr>
<td>HMCR</td>
<td></td>
<td>Duloxetine (60 mg Twice Daily)</td>
</tr>
<tr>
<td>Part A</td>
<td>683</td>
<td>273</td>
</tr>
</tbody>
</table>

a The protocol names are an internal system that the company used to denote the study.

We analyzed all treatment arms from 7 randomized multicenter double-blind active and placebo comparator-controlled clinical trials of duloxetine for major depressive disorder. Earlier phase II trials conducted in the dose range of 5 to 20 mg/d that were not included in the integrated summary of efficacy were excluded. Table 1 lists the trials, protocols, arms, and sample size per arm. Four different protocols were used for these studies (HMAQ, HMAT, HMAY, and HMCR). Parts A and B reflect trials run in parallel following the same protocol. Pooling of data from these trials was anticipated during study design. All trials incorporated double-blind variable-duration placebo lead-in periods to blind patients and investigators to the start of active therapy. Safety and efficacy results from these studies have been published previously as individual study findings23,24,27,28 and summarized as pooled analyses of safety29 and efficacy.30

For this analysis, the following 3 levels of the drug factor were used: duloxetine (all duloxetine doses), SSRI (fluoxetine hydrochloride, paroxetine hydrochloride, and escitalopram oxalate), and placebo (all placebo groups). Because protocols for dose adjustment varied across trials, we could perform dose analyses on only 2 of the protocols (HMAT and HMAY). These secondary analyses revealed no significant dose effects, and the results are not presented herein.

STATISTICAL ANALYSIS

The outcome variable was total score on the 17-item Hamilton Scale for Depression (HAM-D).29 We used growth mixture modeling and a commercially available computer program (MPlus; Muthen and Muthen, Los Angeles, California15-17) to identify distinct trajectories of HAM-D scores during treatment. We first fitted models to the entire sample (duloxetine, SSRI, and placebo arms combined) and then fitted separate models to the active arms and to the placebo arms. The latter analyses were used to evaluate whether different classes would emerge for patients in the active arms and in the placebo arms. We considered linear, quadratic, and cubic trends over time, with between 1 and 4 trajectory classes. We also considered piecewise models with a change point at 2 weeks, linear change before week 2, and quadratic change after week 2.
The selection of the best model was based on the Schwartz-Bayesian information criterion and on the Lo-Mendell-Rubin (LMR) likelihood ratio test. The LMR statistic tests whether a model with 1 less class than the fitted model describes the data as accurately and is used to select the number of classes. We applied the restriction that we need to have at least 5% of the patients in a class for that class to be meaningful clinically and stable numerically. Classification accuracy was assessed using the entropy value ranging between 0 and 1, with values closer to 1 corresponding to better classification accuracy.

Once the best-fitting model for the entire sample was identified, patients were classified to the most likely trajectory class. Baseline characteristics included sex, atypical flag (yes or no), melancholia flag (yes or no), age, age at onset, Hamilton Scale for Anxiety (HAM-A) total score at baseline, duration of current episode (<8 weeks, 8-18 weeks, 18 weeks to 1 year), and number of previous episodes (0, 1-2, 3-4, ≥5, or missing). We performed backward elimination at the P < .10 level. The weights were the posterior probabilities of membership in the assigned class. The association of each baseline characteristic with trajectory membership was also tested one at a time using χ² test, t test, or Wilcoxon rank sum test.

Because the separate analyses of the subsample receiving active drug and the subsample receiving placebo revealed 2 trajectory classes for patients receiving active drug (responders and nonresponders) but only 1 class for patients receiving placebo, mixed-model analyses were performed to assess whether patients receiving placebo had significantly different responses from patients receiving active drug who were classified in responder and nonresponder trajectories. To distinguish these 2 trajectory classes from the clinical definitions of responders and nonresponders, we refer to them as trajectory responders and trajectory nonresponders. The response variable in the mixed model was HAM-D total score during 8 weeks. The predictor variables were trajectory class membership (placebo, trajectory responder on duloxetine or SSRI, or trajectory nonresponder on duloxetine or SSRI), time (as a categorical variable), and their interaction.

An unstructured variance-covariance matrix was used for the errors. To control for the potential confounding of baseline covariates when comparing the randomized placebo group with the nonrandomized trajectory responder and trajectory nonresponder groups, we used a propensity scoring approach. We calculated predicted probability (propensity score) for each patient to be in the trajectory nonresponder class and used this probability as a covariate in the mixed model. The propensities were calculated based on the fitted logistic regression model for patients in the active arms using all available baseline covariates as predictors.

To assess the relationship between trajectory response and clinical response, we used χ² test and measures of agreement. Trajectory responders were patients who were classified in the responder trajectory. Clinical responders were patients with at least 50% improvement from baseline and a HAM-D total score of less than 10 using last-observation-carried-forward imputation for missing data.

Our primary analyses are valid under missing at random (MAR) assumptions. We assessed the effect of missing data on our results by performing a limited sensitivity analysis under missing not at random (MNAR) assumptions. We used the Muthén-Roy pattern mixture model that Muthén et al recommended as the most appropriate and flexible model of its class. It allows the pattern of dropout to influence the outcome of growth mixture modeling by defining 2 distinct latent class variables, one related to dropout and another related to the outcome trajectories. We identified trajectory classes based on the Muthén-Roy model and repeated all remaining analyses using the corresponding trajectory class definitions.

According to the Schwartz-Bayesian information criterion, the best-fitting set of models was the set of piece-
wise growth mixture models (Table 2). Among these, according to the LMR statistic, the model with 2 classes fit the data best. Figure 1A shows the estimated and sample means for the 2 trajectory classes over time based on all the data. The class on the bottom (class 1), with 76.3% probability of membership in this class, can be interpreted as the class of trajectory responders. The class on the top (class 2), with 23.7% probability of member-
ship in this class, can be interpreted as the class of trajectory nonresponders.

Univariate associations between classifications in the 2 trajectory classes (trajectory responder and trajectory nonresponder) and treatment, protocol, and covariates are summarized in Table 3. Based on logistic regression analysis with backward elimination, the following variables were significantly related to trajectory membership: drug, protocol, HAM-A total score at baseline, and duration of current episode. Compared with patients receiving duloxetine, patients receiving placebo had significantly lower odds of being in the responders trajectory (odds ratio [OR], 0.56; 95% CI, 0.42-0.73). The results for patients receiving SSRI were similar and statistically significant (OR, 0.69; 95% CI, 0.51-0.93). Patients receiving duloxetine and patients receiving SSRI did not have statistically different odds of being in the responders trajectory (OR, 1.24; 95% CI, 0.93-1.64). Higher HAM-A total score at baseline was associated with significantly lower odds of being in the responders trajectory (OR, 0.90; 95% CI, 0.88-0.92). Longer duration of current episode seemed to be associated with lower odds of being in the responders trajectory (P = .03), but the post hoc pairwise comparisons of longer durations with the shortest duration were not statistically significant. Compared with the other protocols, protocol HMAY was associated with significantly higher odds of being in the responders trajectory. The 2 trials using this protocol (HMAY part A and HMAY part B) were different from the other trials because they were conducted in Eastern Europe, they included 6-month extensions rather than acute phase only, and the dropout rates were 9% and 19% compared with 30% to 40% for the other trials.

Because duloxetine and SSRI were not differentiated by odds of assignment to the responders trajectory, we combined these 2 groups for the separate analyses of the active and placebo arms. Analyses of both the active and placebo subsamples showed that piecewise linear models fit the data best (Table 4). The active arms identified classes similar to the ones based on the entire sample. Figure 1B and C show the best-fitting solutions for the placebo and active data, respectively. Notably, the placebo arms failed to identify more than 1 trajectory class based on the LMR statistic, suggesting that there was no categorical difference in patients’ responses to placebo. Rather, continuous bell-shaped distributions were sufficient to describe the between-patient heterogeneity in means, rates of change, and curvature in the response over time in the placebo arms.

Table 3. Baseline Characteristics of Trajectory Responders and Trajectory Nonresponders

<table>
<thead>
<tr>
<th>Variable</th>
<th>Trajectory Responders (n = 2034)</th>
<th>Trajectory Nonresponders (n = 481)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duloxetine hydrochloride</td>
<td>978 (84.2)</td>
<td>183 (15.8)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SSRI</td>
<td>561 (79.9)</td>
<td>141 (20.1)</td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>495 (75.9)</td>
<td>157 (24.1)</td>
<td></td>
</tr>
<tr>
<td>Protocol, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMAQ</td>
<td>280 (76.5)</td>
<td>86 (23.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>HMAT</td>
<td>557 (78.8)</td>
<td>150 (21.2)</td>
<td></td>
</tr>
<tr>
<td>HMAY</td>
<td>669 (88.1)</td>
<td>90 (11.9)</td>
<td></td>
</tr>
<tr>
<td>HMCR</td>
<td>528 (77.3)</td>
<td>155 (22.7)</td>
<td></td>
</tr>
<tr>
<td>Sex, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1337 (80.5)</td>
<td>323 (19.5)</td>
<td>0.55</td>
</tr>
<tr>
<td>Male</td>
<td>697 (81.5)</td>
<td>158 (18.5)</td>
<td></td>
</tr>
<tr>
<td>Atypical flag, No. (%)</td>
<td></td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>Yes</td>
<td>72 (82.8)</td>
<td>15 (17.2)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1941 (80.9)</td>
<td>457 (19.1)</td>
<td></td>
</tr>
<tr>
<td>Melancholia flag, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1343 (79.5)</td>
<td>346 (20.5)</td>
<td>0.02</td>
</tr>
<tr>
<td>No</td>
<td>683 (83.5)</td>
<td>135 (16.5)</td>
<td></td>
</tr>
<tr>
<td>Duration of current episode, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;8 wk</td>
<td>482 (85.9)</td>
<td>79 (14.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>8-18 wk</td>
<td>571 (85.1)</td>
<td>100 (14.9)</td>
<td></td>
</tr>
<tr>
<td>18 wk to 1 y</td>
<td>489 (78.4)</td>
<td>135 (21.6)</td>
<td></td>
</tr>
<tr>
<td>&gt;1 y</td>
<td>488 (74.5)</td>
<td>167 (25.5)</td>
<td></td>
</tr>
<tr>
<td>Previous episodes, No. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>464 (83.8)</td>
<td>90 (16.2)</td>
<td>0.06</td>
</tr>
<tr>
<td>1-2</td>
<td>677 (82.1)</td>
<td>148 (17.9)</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>334 (80.5)</td>
<td>81 (19.5)</td>
<td></td>
</tr>
<tr>
<td>≥5</td>
<td>390 (77.5)</td>
<td>113 (22.5)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>169 (77.5)</td>
<td>49 (22.5)</td>
<td></td>
</tr>
<tr>
<td>Continuous variables, mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>42.47 (12.14)</td>
<td>43.38 (12.30)</td>
<td>.14</td>
</tr>
<tr>
<td>Age at onset</td>
<td>31.88 (13.87)</td>
<td>31.00 (13.80)</td>
<td>.21</td>
</tr>
<tr>
<td>HAM-A total score</td>
<td>15.39 (5.56)</td>
<td>17.71 (5.82)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Abbreviations: HAM-A, Hamilton Scale for Anxiety; SSRI, selective serotonin reuptake inhibitor.
To assess whether this result might be affected by potential confounding factors (eg, the number of treatment arms, percentage of patients randomized to placebo,31 or different sample sizes for the active and placebo groups), we performed separate analyses for the 3-arm and 4-arm trials, separate analyses of trials with different placebo response rates (20%, 25%, and 40%), and separate analyses of the duloxetine and SSRI arms. The substantive conclusions were the same. Only 1 class was identified for the placebo group ($P > .1$ for all by LMR test). Separate analyses of the duloxetine and SSRI groups favored a 2-class solution over a single-class solution ($P > .03$ for duloxetine and $P > .002$ for comparator SSRI). Adding a third class did not improve the fit of these data ($P = .07$ for duloxetine and $P = .15$ for comparator SSRI), and the third class consisted of a very small percentage of patients with unstable trajectory estimation. Therefore, 2 classes were necessary to adequately describe heterogeneity in response trajectories with duloxetine and comparator SSRI.

The mixed models comparing HAM-D scores over time of trajectory responders receiving active drug, trajectory nonresponders receiving active drug, and patients receiving placebo (Figure 2) showed a significant interaction between trajectory class membership and time ($F_{10,163} = 73.3, P < .001$). Not surprisingly, trajectory responders receiving active drug showed reduced HAM-D scores compared with patients receiving placebo at week 8 ($t_{2081} = −4.08, P < .001$). However, pairwise comparisons at week 8 also showed significantly higher HAM-D scores for trajectory nonresponders receiving active drug compared with patients receiving placebo ($t_{2081} = 7.95, P < .001$). When controlling for the propensity of a patient to be a trajectory responder, the differences between trajectory nonresponders receiving active drug vs patients receiving placebo and between trajectory responders receiving active drug vs patients receiving placebo at week 8 remained significant ($t_{2082} = 21.23$ and $t_{2083} = −3.34$, respectively; $P < .001$ for both). Therefore, it seems that the baseline characteristics we studied were insufficient to account for the differences observed between the nonresponders receiving active drug and the patients receiving placebo.

Using the definition of clinical response as at least 50% improvement from baseline and a HAM-D total score of less than 10 and a last-observation-carried-forward imputation method for missing data, the correspondence between trajectory responders and clinical responders indicated that 480 of 481 (99.8%) trajectory nonresponders were also clinical nonresponders. However, 1318 of 2034 (64.8%) trajectory responders were classified as clinical responders. The remaining trajectory responders (716 of 2034 [35.2%]) were classified as clinical nonresponders.

Overall, dropout proportions were 23.5% of patients receiving duloxetine, 23.0% receiving SSRI, and 26.7% receiving placebo. Among trajectory responders on active drug, the dropout proportion was 18.5%, and among trajectory nonresponders receiving active drug, the dropout proportion was 24.4%. In the entire sample, the Muthen-Roy pattern mixture model under MNAR assumptions identified 2 classes of trajectory response similar to those identified under MAR assumptions (63.7% trajectory responders and 36.3% trajectory nonresponders) and 2 classes of dropout patterns (22.6% with high probability of dropout and 77.4% with low probability of dropout). More patients were classified as trajectory nonresponders in the MNAR analysis (36.3%) vs 19.1%). Univariate differences in baseline characteristics between trajectory responders and trajectory nonresponders similar to those given in Table 2 emerged, with 2 exceptions. There was no longer a significant difference in class membership by melancholia depression type ($P = .38$), and the trajectory responders identified under MNAR assumptions had older mean (SD) age at onset than trajectory nonresponders identified under MNAR assumptions (32.1 [11.7] vs 31.0 [14.1] years, $P = .04$). Backward elimination with MNAR trajectory class definitions resulted in the same final model as that reported for MAR assumptions, with similar ORs.

In the placebo group, 1 trajectory class and 2 dropout classes (23.3% with high probability of dropout and 76.7% with low probability of dropout) were identified, while in the active group 2 trajectory classes (68.0% trajectory responders and 32.0% trajectory nonresponders) and 2 dropout classes (19.0% with high probability of dropout and 81.0% with low probability of dropout) were identified. According to the MNAR analysis, the mixed models comparing HAM-D scores over time of trajectory responders receiving active drug, trajec-

### Table 4. Results From Fitting of Piecewise Linear Models to the Subset of Placebo Arms and to the Subset of Active Arms

<table>
<thead>
<tr>
<th>Model</th>
<th>Likelihood</th>
<th>Schwartz-Bayesian Information Criterion</th>
<th>Entropy</th>
<th>Lo-Mendell-Rubin Likelihood Ratio Test</th>
<th>Proportion of Individuals in Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Criterion</td>
<td></td>
<td>Likelihood Ratio</td>
<td>1</td>
</tr>
<tr>
<td>Only placebo arms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1P</td>
<td>−11024.88</td>
<td>22237.68</td>
<td>...</td>
<td>1.00</td>
<td>...</td>
</tr>
<tr>
<td>4.2P</td>
<td>−10999.68</td>
<td>22226.16</td>
<td>0.94</td>
<td>0.96</td>
<td>0.04</td>
</tr>
<tr>
<td>4.3P</td>
<td>−10986.31</td>
<td>22238.30</td>
<td>0.90</td>
<td>0.93</td>
<td>0.03</td>
</tr>
<tr>
<td>4.4P</td>
<td>−10969.92</td>
<td>22244.40</td>
<td>0.72</td>
<td>0.57</td>
<td>0.38</td>
</tr>
<tr>
<td>Only active arms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1A</td>
<td>−30967.15</td>
<td>62152.67</td>
<td>...</td>
<td>1.00</td>
<td>...</td>
</tr>
<tr>
<td>4.2A</td>
<td>−30869.97</td>
<td>62093.49</td>
<td>0.67</td>
<td>0.77</td>
<td>0.23</td>
</tr>
<tr>
<td>4.3A</td>
<td>−30811.50</td>
<td>61931.74</td>
<td>0.77</td>
<td>0.74</td>
<td>0.25</td>
</tr>
<tr>
<td>4.4A</td>
<td>−30771.27</td>
<td>61896.45</td>
<td>0.80</td>
<td>0.73</td>
<td>0.24</td>
</tr>
</tbody>
</table>
tory nonresponders receiving active drug, and patients receiving placebo showed a significant interaction between trajectory class membership and time ($F_{16,1826}=82.6$), significantly lower HAM-D scores for trajectory responders receiving active drug than for patients receiving placebo at week 8 ($t_{2181}=21.72$), and significantly higher HAM-D scores for trajectory nonresponders receiving active drug than for patients receiving placebo at week 8 ($t_{2047}=-17.99$) ($P<.001$ for all). Therefore, the MNAR sensitivity analyses confirmed our results under MAR assumptions.

**COMMENT**

The overall trajectory-based analyses of the treatment of 2515 patients successfully classified them into response trajectories and confirmed that duloxetine and SSRI treatment increased the likelihood that most patients treated with these medications would be classified in the responders trajectory. The magnitude of the effects seemed as large as the magnitude of the effects from end-point or mixed-model analyses. Therefore, trajectory analyses allowed for strong signal detection, although treatment effects were not more significant than in end-point and mixed-model analyses. The added improvements of trajectory analyses were that patients were classified into response trajectories and that the trajectories were different for active drug and for placebo. As noted earlier, “responder” in this study refers to a favorable clinical trajectory rather than achievement of a priori criteria based on symptom thresholds, an approach that is commonly used in clinical trials.

Separate analyses of the active and placebo groups revealed that distinct trajectories were identified in the groups treated with duloxetine or SSRI but not in the placebo group. Most patients (about three-quarters) receiving active drug were classified as trajectory responders, and fewer (about one-quarter) were classified as trajectory nonresponders. Patients receiving placebo could not be separated into distinct trajectories and on average showed gradual improvement over time. The failure to identify more than 1 trajectory classes over time in the placebo group is remarkable because growth mixture models are more prone to overestimate rather than underestimate the number of trajectory classes. The finding of a single trajectory for patients assigned to placebo may reflect limited statistical power to resolve subtle differences in response trajectories in this group. However, this reasoning assumes that there are categorically different outcomes in each group, and this may not be the case. The present data suggest that widely divergent trajectories of individual patients treated with placebo are best explained as variations within a single class (ie, placebo response differences may be a dimensional rather than categorical characteristic).

The present findings challenge the prevailing view that placebo response is associated predominantly with rapid and transient clinical improvement. In some investigations, researchers classified patients...
showing this response pattern as placebo responders, although the validity of this assumption was never demonstrated. The problem with this conclusion is that growth mixture models provide valid results only under the assumption that placebo responders and nonresponders are groups differentiated by a categorical distinction. If categorically different classes do not exist, spurious latent classes are likely to be identified. Therefore, whether one finds categorically different trajectories depends directly on whether one is willing to assume a priori that they exist. This problem is exacerbated in studies with small sample sizes. This study did not assume a priori that placebo response was constituted by multiple trajectory classes. Despite the fact that this study analyzed a much larger sample than prior studies of its kind, its analyses do not support the notion that there is a specific placebo response profile. Rather, we observed an average gradual improvement over time among patients taking placebo, with noticeable dimensional but not categorical heterogeneity. Our results are consistent with the view that placebo response is a continuous measure that is manifested to varying degrees across patients. This view is consistent with the approach by Tarpey and Petkova, who used continuous rather than categorical latent variables to model patient responses to antidepressant medications.

The findings that patients receiving active drug were classified as trajectory responders and trajectory nonresponders, while there was only 1 trajectory for patients receiving placebo, may be partially explained by the effect of the active compound. Active compounds generally have the potential to produce adverse responses that could contribute to categorically different responses. For example, medications with inverted U-shaped dose-response relationships may be beneficial for patients with low baseline values but may be harmful to patients at the opposite end of the spectrum.

Another possible explanation for the single trajectory class among patients receiving placebo is that all clinical trials in this analysis had a placebo treatment phase before treatment with randomized medication. This placebo lead-in may have acted as a filter that reduced response heterogeneity during subsequent placebo treatment because it eliminated patients with immediate placebo response from the sample. Placebo lead-in is a practice that is intended to improve signal detection by selecting out patients who are likely to respond to placebo and by selecting in patients whose poor placebo response portends superior medication responses. However, the published data overwhelmingly indicate that the rate of clinical response during placebo lead-in is very low and that exclusion of placebo responders does not improve signal detection in clinical trials. Therefore, it is unlikely that the presence of placebo lead-in periods in the contributing studies influenced the present findings or compromised their generalizability.

Our analyses do not support the view that antidepressant treatment response was slower and less transient than placebo response. Rather, consistent with a recent meta-analysis confirming early signs of SSRI efficacy, the change in HAM-D scores as a result of antidepressant treatment among responders seemed to be faster and more sustained than the change in HAM-D scores among patients on placebo. Our analyses also support the notion that early improvement during the first 2 weeks of treatment is predictive of treatment outcome and that different antidepressant medications have similar treatment response profiles.

In the present analysis, where there are 2 trajectories for patients treated with antidepressants and 1 trajectory for patients treated with placebo, some patients would seem to be more effectively treated with placebo than with a serotonergic antidepressant. The trajectory of nonresponse to antidepressants was not consistent with the timing of a transient increase in suicidality among antidepressant-treated patients with depression. Instead, the trajectories of responders and nonresponders diverged increasingly over time, suggesting that patients affected adversely by serotonin reuptake inhibitors might be better off if these medications were discontinued. At a minimum, they highlight the potential clinical importance of careful ongoing monitoring of the effect of prescribed antidepressants.

The clear separation of patients treated with antidepressants into responders and nonresponders and the observed homogeneity of treatment response on placebo, together with the significantly worse mean response of trajectory nonresponders taking active drug compared with the mean response of patients taking placebo, may help explain the failure of many clinical trials of antidepressant medication to demonstrate treatment efficacy. The main hypotheses in almost all clinical trials focus on demonstrating average treatment effects. However, when there is sizable heterogeneity in treatment response, with most patients benefiting from a treatment but with some patients demonstrating significantly worse outcomes than the average improvement while receiving placebo, it is likely that average treatment effects will be diminished and may lose statistical significance. Therefore, unless strategies are introduced to reduce this heterogeneity or study sample sizes are increased significantly to augment statistical power, it would seem that the status quo regarding failed trials of serotonin reuptake inhibitors is unlikely to change.

The fact that almost all trajectory nonresponders were clinical nonresponders but only about two-thirds of trajectory responders were clinical responders suggests that the clinical response definition is stricter than the trajectory response definition. It is possible that patients who do not meet the clinical response definition but are classified as trajectory responders may meet the clinical response definition with a longer follow-up period. This can be evaluated in follow-up analyses. Validation of our results in other studies is also necessary to assess how stable the trajectory class definitions are.

The inferences drawn from the present analyses are limited by several factors. First, we were unable to differentiate whether the poor responses to antidepressants arose from a negative effect of taking antidepressant medication or as a consequence of discontinuation of these medications. Although our limited sensitivity analysis under MNAR assumptions suggests that our results are not sensitive to the effects of missing data, further research is necessary to estimate the extent of the
influence of missing data due to different model assumptions for missing data.

Second, although the reported studies included only patients with unipolar major depression determined using rigorous diagnostic methods, it is impossible to rule out the presence of latent bipolar disorder in the study population. Antidepressants are reported to have reduced efficacy in treating depression among patients with bipolar disorder, and they may increase mood cycling. Similarly, adverse effects of other forms of psychopathologic conditions, particularly personality disorders, or the social use of alcohol cannot be determined using the existing data set.

Third, the implications are limited by the brief duration (8 weeks) of the reported studies. In the present study, it was impossible to predict whether nonresponders may become responders with extended treatment or whether the negative consequences of antidepressant discontinuation in some nonresponders outweigh the apparent benefits.

Fourth, the present analysis is predicated on the assumption that heterogeneous classes of antidepressant medications exist. An alternative view is that the probability of nonresponse may vary continuously across the population, and if this is indeed the case, alternative models with latent trait rather than latent class variables might be more appropriate.

Fifth, we considered only 1 type of linear parametric model. Alternative nonlinear models have been considered by other authors to predict treatment outcome from early treatment response. It is difficult to distinguish the fit of such models from the fit of simpler polynomial models with a limited number of time points. Such models have not been used in the context of growth mixture modeling and are unavailable in standard software, but they might provide an important tool for flexible modeling of treatment response for more frequently collected treatment outcome data.

Sixth, our analyses do not assess causal treatment effects because we consider treatment a predictor of class membership rather than a predictor of growth factors within a class, as suggested by Muthén and Brown. In the causal framework, latent classes are considered characteristics of the patients (eg, never responders, drug-only responders, placebo-only responders, and always responders), and a key assumption is that the “numbers of classes and the true, population proportions of patients in each class be the same across intervention conditions.” In our analysis, classes are considered empirically derived distinct trajectories of longitudinal response, and we have direct evidence that the number of classes and the proportion of classes vary by treatment. Although our propensity scoring approach allowed us to control for observed predictors of trajectory membership, it is possible that unmeasured confounders (eg, genotypes and environmental factors) can affect our results.

Identifying predictors of clinical trajectories might advance the personalized treatment of major depressive disorder (ie, to assist in the better matching of patients and treatments). In this regard, the present study replicated the finding from the large multicenter antidepressant trial Sequenced Treatment Alternatives to Relieve Depression (STAR*D) that high levels of anxiety predicted poorer antidepressant response. Predictors of membership in both favorable and poor response trajectories might be used to inform the selection of medications for particular patients. Despite a long history of the study of clinical, neurochemical, and genetic predictors of subtypes of depression and treatment response, there are still no objective bases for the personalized treatment of depression that are sufficiently explanatory and specific to guide the treatment of individual patients. Future research will be needed to determine whether trajectory-based analyses will be useful in advancing this objective.

Submitted for Publication: May 27, 2011; accepted July 8, 2011.

Correspondence: Ralitza Gueorguieva, PhD, Division of Biostatistics, School of Public Health, Yale University School of Medicine, 60 College St, Room 215B, New Haven, CT 06520-8034 (ralitza.gueorguieva@yale.edu).

Author Contributions: Dr Gueorguieva had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Financial Disclosure: Dr Krystal has been a consultant to the following companies: Aisling Capital LLC, Astazeneca Pharmaceuticals, Biocortech, Brintnall & Nicolini Inc, Easton Associates, Gilead Sciences Inc, GlaxoSmithKline, Janssen Pharmaceuticals, Lundbeck Research USA, Medivation Inc, Merz Pharmaceuticals, MK Medical Communications, Pfizer Pharmaceuticals, F. Hoffmann-La Roche Ltd, SK Holdings Co Ltd, Sunovion Pharmaceuticals Inc, Takeda Industries, Teva Pharmaceutical Industries Ltd, and Transcept Pharmaceuticals. Dr Krystal is on the scientific advisory board of Abbott Laboratories, Bristol-Myers Squibb, Eisai Inc, Eli Lilly and Co, Forest Laboratories, Lohocla Research Corporation, Mene-mosyne Pharmaceuticals Inc, Naurex Inc, Pfizer Pharmaceuticals, and Shire Pharmaceuticals. Dr Krystal has the following patents and inventions: (1) patent 5 447 948 (September 5, 1995) with J. P. Seibyl and D. S. Charney on dopamine and noradrenergic reuptake inhibitors in the treatment of schizophrenia and (2) patent application PCTWO06108055A1 with G. Sanacora related to targeting the glutamatergic system for the treatment of neuropsychiatric disorders.

Funding/Support: This trial was supported by grants K05 AA 14906 and 2P50 AA 012870 from the National Institute on Alcohol Abuse and Alcoholism and the Veterans Affairs National Center for Posttraumatic Stress Disorder and by Clinical and Translational Science Award UL1 RR024139 from the National Center for Research Resources (a component of the National Institutes of Health) and the National Institutes of Health Roadmap for Medical Research.

Disclaimer: The contents herein are solely the responsibility of the authors and do not necessarily represent the official view of National Center for Research Resources or the National Institutes of Health.

Additional Contributions: Brian Pittman, MS, provided helpful comments on the manuscript.
REFERENCES


©2011 American Medical Association. All rights reserved.


