Antidepressant Response Trajectories and Associated Clinical Prognostic Factors Among Older Adults

Stephen F. Smagula, PhD; Meryl A. Butters, PhD; Stewart J. Anderson, PhD; Eric J. Lenze, MD; Mary Amanda Dew, PhD; Benoit H. Mulsant, MD; Francis E. Lotrich, MD, PhD; Howard Aizenstein, MD, PhD; Charles F. Reynolds III, MD

IMPORTANCE More than 50% of older adults with late-life major depressive disorder fail to respond to initial treatment with first-line pharmacological therapy.

OBJECTIVES To assess typical patterns of response to an open-label trial of extended-release venlafaxine hydrochloride (venlafaxine XR) for late-life depression and to evaluate which clinical factors are associated with the identified longitudinal response patterns.

DESIGN, SETTING, AND PARTICIPANTS Group-based trajectory modeling was applied to data from a 12-week open-label pharmacological trial conducted in specialty care as part of the Incomplete Response in Late Life: Getting to Remission Study. Clinical prognostic factors, including domain-specific cognitive performance and individual depression symptoms, were examined in relation to response trajectories. Participants included 453 adults aged 60 years or older with current major depressive disorder. The study was conducted between August 2009 and August 2014.

INTERVENTION Open-label venlafaxine XR (titrated up to 300 mg/d) for 12 weeks.

MAIN OUTCOMES AND MEASURES Subgroups exhibiting similar response patterns were derived from repeated measures of overall depression severity obtained using the Montgomery-Asberg Depression Rating Scale.

RESULTS Among the 453 study participants, 3 subgroups with differing baseline depression severity clearly responded to treatment: one group with the lowest baseline severity had a rapid response (n = 69 [15.23%]), and distinct responses were also apparent among groups starting at moderate (n = 108 [23.84%]) and higher (n = 25 [5.52%]) baseline symptom levels. Three subgroups had nonresponding trajectories: 2 with high baseline symptom levels (totaling 35.98%: high, nonresponse 1, n = 110 [24.28%]; high, nonresponse 2, n = 53 [11.70%]) and 1 with moderate baseline symptom levels (n = 88 [19.43%]). Several factors were independently associated with having a nonresponsive trajectory, including greater baseline depression severity, longer episode duration, less subjective sleep loss, more guilt, and more work/activity impairment (P < .05). Higher delayed memory (list recognition) performance was independently associated with having a rapid response (adjusted odds ratio = 2.22; 95% CI, 1.18-4.20).

CONCLUSIONS AND RELEVANCE Based on the observed trajectory patterns, patients who have late-life depression with high baseline depression severity are unlikely to respond after 12 weeks of treatment with venlafaxine XR. However, high baseline depression severity alone may be neither a necessary nor sufficient predictor of treatment nonresponse.

TRIAL REGISTRATION clinicaltrials.gov Identifier: NCT00892047
More than 50% of adults treated for late-life depression (LLD) fail to respond to first-line pharmacological therapy. High pretreatment depression severity may be a powerful marker of first-line resistance; however, past research (including our own) linking baseline severity to treatment response may be limited by a simple and important methodological consideration. Specifically, use of a threshold-based definition of remission (ie, having a Montgomery-Asberg Depression Rating Scale [MADRS] score ≤10 by week 12 of treatment) results in a situation where, depending on initial symptom levels, patients recovering at the exact same rate may or may not reach the remission cutoff within the study period.

Group-based trajectory modeling provides a complementary approach to examine treatment response variability without relying on a prespecified remission threshold or assuming the entire sample follows a single trajectory. This data-driven approach identifies subgroups that share common patterns of change over time. Previous applications to clinical research focused on depression consistently demonstrate that group-based trajectory methods capture heterogeneity in the course of illness that would be neglected using traditional approaches.

To our knowledge, no prior study has used group-based trajectory methods to examine treatment response variability within a large sample of older adults undergoing solo pharmacological treatment for LLD. Although both sophisticated and important, previous applications of group-based longitudinal methods to study treatment response have been large but not restricted to older adults (predominantly including middle-aged adults), small plus not restricted to older adults, or small and not restricted to solo pharmacological treatment. Compared with younger adults, factors including pretreatment comorbidity and polypharmacy may contribute to an altered course of pharmacological treatment and response in LLD. Variability in neuropsychological function among older adults also predicts treatment outcomes.

We therefore applied group-based trajectory modeling to examine LLD response in the largest open-label trial of solo pharmacotherapy (extended-release venlafaxine hydrochloride [venlafaxine XR]) conducted among older adults to date. Given that this was an open-label trial delivered in a specialty care setting that included depression management support and monitoring, we did not aim to evaluate pharmacological efficacy or the causality of effect (ie, venlafaxine, monitoring, or placebo effects). Instead, we took an exploratory approach to describe the typical patterns of symptom change observed during a trial of pharmacological care for LLD. In addition, we assessed associations between previously identified prognostic factors (ie, medical comorbidity, depression symptom severity, and neuropsychological function) with data-derived response patterns. Given evidence suggesting that variability in individual sleep or core depression symptoms might predict treatment outcomes, we also explored associations between individual symptom expression and response patterns.

Methods

This article pertains to the initial phase of the Incomplete Response in Late-Life Depression (IRLGREY) Study, which was a 3-site open-label trial conducted to prospectively assess treatment response.

Participants

The open-label phase of the IRLGREY Study, conducted between August 2009 and August 2014, has been described previously. Participants were aged 60 years or older (n = 466) and had current nonpsychotic major depressive disorder as diagnosed by the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition plus a MADRS score of 15 or higher. Exclusions were lifetime diagnosis of bipolar disorder, schizophrenia, schizoaffective disorder, other psychotic disorders, or current psychotic symptoms; clinical history of dementia or cognitive impairment as indicated by a score less than 20 on the Mini-Mental State Examination; alcohol or substance abuse in the past 3 months; high suicide risk; unstable medical illness; or contraindication to venlafaxine XR or aripiprazole (which was administered in a subsequent trial). Participants were included in the analysis if they had complete outcome data from baseline and at least 1 follow-up visit (n = 453). Participants provided written informed consent, and ethical approval was obtained from site institutional review boards at the University of Pittsburgh, Washington University in St Louis, and the Center for Addiction and Mental Health, University of Toronto.

Intervention

Participants were treated with venlafaxine XR starting at 37.5 mg/d and titrated up to 300 mg/d following a study-standardized protocol for at least 12 weeks. Details on titration and use of other medications have been published previously. Depression-specific psychotherapy was not provided, although pharmacological treatment was delivered with depression management support including clinical care focused on depression symptoms, treatment, suicidal ideation, countermeasures for adverse medication effects, and adherence.

Measures

Outcome

Depression symptom burden was measured over time as the total MADRS score assessed at study baseline, week 1, week 2, and every 2 weeks thereafter until completion of the open-label study.

Baseline Prognostic Factors

Baseline assessments included the 17-item Hamilton Rating Scale for Depression. To explore potential prognostic roles of both overall symptom burden and individual symptoms, the MADRS and the 17-item Hamilton Rating Scale for Depression were examined as scored totals and individual items. Adequacy of prior antidepressant treatment was measured with the Antidepressant Treatment History Form (scores ≥3 in-
dicate a prior adequate trial). Other clinical factors were age at first episode onset, current episode duration (natural log transformed), single vs recurrent episodes, and receipt of outside psychotherapy. The Brief Symptom Inventory24 and the Scale for Suicide Ideation26 were used to assess general anxiety and suicidal ideation, respectively. Medical comorbidity was assessed using the Cumulative Illness Rating Scale for Geriatrics27 (expressed as total and count). The physical function subsection of the 36-Item Short Form Health Survey was expressed as a total score.28 Self-reported alcohol use was assessed as the number of drinks per week.

Baseline neuropsychological measures included the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS).29 Initially, we examined domain-specific RBANS summary scores. Among RBANS domains associated with trajectory group, we further examined the individual subtests that composed the relevant summary score. Subtest scores were analyzed using age-normed z scores. Two tasks from the Delis-Kaplan Executive Function System30 were also examined in this manner.

Statistical Analysis

A semiparametric, group-based modeling strategy was used to classify the cohort into subgroups based on identifying heterogeneous longitudinal polynomial trajectories.31,32 We implemented this technique using PROC TRAJ32 in SAS version 9.3 statistical software (SAS Institute, Inc). For our implementation, we assumed that the error structure followed a censored normal distribution. We determined the number of groups and degree of polynomial in each of the trajectory groups using the Bayesian information criterion (BIC), which measures improvement in model fit gained by adding additional groups or shape parameters incorporating a penalty for added complexity. The BIC log Bayes factor approximation, defined as $2 \times [\Delta \text{BIC}]$ (subtracting a less complex model from a more complex model), has been shown to be a good approximation to the log Bayes factor criterion33 and was used to base selection of the number of trajectory groups that fit the data. A log Bayes factor approximation higher than 10 is considered to be strong evidence in favor of the more complex model.32 Solutions that included small trajectory groups (<5% of the sample) were rejected. The degree of the fitted polynomials was determined by examining BIC for all possible permutations of linear and quadratic trajectories. Cubic polynomials did not improve model fit as indicated by BIC. To assess model fit, we examined the average posterior probability of group membership (>70% recommended) and odds of correct classification (>5 considered adequate).

Weighted multinominal logistic regression was used to assess crude associations of prognostic factors with trajectory group membership. These regressions were weighted by the probability of group membership to account for measurement error introduced by the probabilistic nature of group assignment. The reference group was chosen to be the largest group with a distinct response. To assess whether individual depression symptoms were associated with trajectory group membership beyond a function of overall depression severity, we adjusted models including individual symptoms for total baseline depression severity (measured using the item’s scale’s total score). Models including neuropsychological variables were adjusted for age and education. Associations of group membership with prognostic factors that achieved crude statistical significance (defined as $P < .05$) were selected as predictors in a maximum model. Automated backward elimination (elimination threshold $P > .20$) was then used to select uniquely predictive prognostic factors, which were entered as baseline risk factors within the PROC TRAJ framework.

### Results

**Model Selection and Trajectory Group Description**

Regardless of whether all groups were assigned linear, quadratic, or cubic trajectories, BIC consistently indicated the same optimal number of groups. Model fit indicated by BIC improved with additional trajectory groups (Table 1); however, a 7-group solution was rejected because it included a small group (<5% of the sample). The BIC log Bayes factor approximations indicated strong evidence that 6 groups fit the data better than 5 groups. The data were best fit by a combination of linear and quadratic trajectory groups. Average posterior probabilities were high for all 6 groups (range, 0.84-0.90) and the odds of correct classification were all well above 5 (minimum odds of correct classification = 20.18). This trajectory model depicted heterogeneity in symptom course across all levels of total baseline severity (Figure). Groups were assigned names based on their relative starting position and/or response trajectory.

Three groups exhibited clear changes in symptom severity overall and were all labeled responding trajectories (totaling 44.59% of the sample). Two started at relatively lower or moderate symptom levels and followed quadratic trajectories (15.23%, rapid response; 23.84%, moderate, response; see Table 2 for baseline and 12-week symptom levels). The third responding group had higher average baseline symptom levels that followed a linear trajectory (5.52%, high, response). The number of weeks until average symptom levels within these groups reached below typical remission criteria (ie, MADRS score ≤10) appeared to differ (Figure).

**Table 1. Values of BIC by Number of Groups for Models With All Linear Exponents, and Associated BIC Log Bayes Factor Approximation**

<table>
<thead>
<tr>
<th>No. of Groups</th>
<th>BIC</th>
<th>$2 \times [\Delta \text{BIC}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−11 971.12</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>−11 243.70</td>
<td>1454.84</td>
</tr>
<tr>
<td>3</td>
<td>−11 058.70</td>
<td>370.00</td>
</tr>
<tr>
<td>4</td>
<td>−10 978.19</td>
<td>161.02</td>
</tr>
<tr>
<td>5</td>
<td>−10 931.47</td>
<td>93.44</td>
</tr>
<tr>
<td>6$^a$</td>
<td>−10 888.79</td>
<td>85.36</td>
</tr>
<tr>
<td>7$^b$</td>
<td>−10 883.20</td>
<td>11.18</td>
</tr>
</tbody>
</table>

Abbreviations: BIC, Bayesian information criterion; NA, not applicable.

$^a$ Value is evidence against the null hypothesis.

$^b$ Selected solution.
Two groups did not exhibit clear changes in symptom severity and were labeled nonresponse (totaling 35.98% of the sample). These groups had high severity at baseline and followed linear trajectories without substantive change (24.28%, high, nonresponse 1; and 11.70%, high, nonresponse 2). The last group had initially moderate symptom levels that followed a linear trajectory; while some symptom change was apparent (Figure), this group retained most of their average baseline symptoms at the end of the trial and was labeled moderate, mixed/nonresponse (19.43% of the sample).

**Associations of Prognostic Factors With Trajectory Group**

Clinical characteristics are presented by trajectory group for descriptive purposes only (Table 2). We also present crude associations of neuropsychological test performance and group membership (eTable in the Supplement); in these models, better performance in the delayed memory domain was related to higher odds of being in the rapid response group (all comparisons are with the moderate, response group), and better performance in the language domain was related to lower odds of membership in the 3 nonresponding groups.

In the final multivariable-adjusted model (Table 3), longer episode duration was related to 54% higher odds of being in the high, nonresponse 1 group as well as 40% higher odds of being in the moderate, mixed/nonresponse group. Greater education was related to 17% decreased odds of being in the high, nonresponse 1 group.

Higher baseline depression severity was associated with greater odds of being in the 2 high treatment nonresponse groups as well as the high, response group (Table 3). Besides severity, no other baseline factors were associated with membership in the high, response group. Higher baseline depression severity was associated with reduced odds of being in both the rapid response group and moderate, mixed/nonresponse group.

Higher levels of the MADRS reduced sleep item were associated with lower odds of being in all 3 nonresponse groups. Greater work/activity impairment was associated with more than 3 times the odds of being in the high, nonresponse 2 group, and higher levels of guilt were associated with more than twice the odds of being in this group. The association between guilt and greater odds of membership in the high, nonresponse 1 group was nonsignificant ($P = .06$).

Each standard deviation higher in list recognition performance (a subtest from the RBANS delayed memory domain) was associated with more than twice the odds of being a rapid responder. The association between higher list recognition performance and lower odds of being in the high, nonresponse 2 group was nonsignificant ($P = .08$).

Using the fitted parameters (as odds ratios in Table 3 and as $\beta$ values in Table 4) and the values for a patient's pretreatment characteristics, the probability of any patient belonging to each of the identified groups can be calculated following the equation described in Table 4.

**Discussion**

Our study provides a novel description of the typical longitudinal patterns of depressive symptom change observed among older adults undergoing open-label pharmacological treatment for LLD. High baseline depression severity was generally related to having a nonresponsive trajectory over time. Nevertheless, a small group of patients (high, response group) did have a favorable trajectory despite starting the trial among the most severely depressed. On the other hand, the moderate, mixed/nonresponse group began treatment with relatively moderate symptoms that did not respond to treatment robustly. Taken together, these findings suggest that while patients with a high baseline symptom burden are unlikely to respond to an initial trial of venlafaxine XR, treatment outcomes are likely jointly determined by several factors in addition to depression severity, including symptom expression, episode duration, and cognitive function.

A particularly innovative aspect of our study is the exploratory examination of individual symptoms in relation to response trajectories. Higher levels of guilt and greater work/activity impairment were both strongly and independently associated with only the high, nonresponse 2 group. In contrast, greater levels of subjective sleep loss were associated with reduced odds of being in all of the nonresponding groups. These findings suggest that the severity of these symptoms may be prognostic markers of response to pharmacological treatment for LLD; however, because this was an open-label trial (that was not placebo controlled), our findings do not support concluding that these symptoms moderate venlafaxine efficacy. Also note that in the final model, 3 other sleep symptoms (early, middle, and late insomnia) from the 17-item Hamilton Rating Scale for Depression were not associated with trajectory group membership. Future placebo-controlled studies...
including objective sleep-wake assessments are needed to understand the neurobiological basis and roles of specific sleep characteristics in depression treatment response.

Consistent with prior analyses, in our final model, prolonged episode duration was associated with nonresponse. Prolonged episode duration may thus contribute to treatment resistance, or alternatively, prolonged episode duration might be a consequence of other underlying resistance factors. The current analysis cannot resolve whether episode duration is a cause or consequence of treatment resistance. Nevertheless, episode duration should be considered when initiating a treatment strategy, and the early detection and treatment of depression may be advantageous.

We found that better performance on a delayed memory word list recognition test was associated with having a rapid response. This finding suggests that, despite older age and depressive illness, relative preservation of retention ability, which is associated with hippocampal function, may facilitate a rapid response. This hypothesis is consistent with prior evidence linking hippocampal volume to LLD treatment outcomes. In crude models, we found that worse performance on a semantic fluency test was associated with being in all 3 of the nonresponsive groups. Mild cognitive impairment, although not formally adjudicated in this study, is common among older adults with LLD and may explain this finding. Future studies tracking the role of mild cognitive impairment in the development of depression can determine whether semantic fluency deficits mark a distinct phenotype that (by nature of being a downstream result of cognitive impairment) will hinder response to traditional depression treatment.

Contrary to prior evidence, this analysis did not detect associations between physical comorbidity and treatment outcomes. However, the analysis did not include a full medical history or biological assessments. We did detect crude associations between perceived physical functioning, anxiety symptoms, age at onset, and pharmacological treatment history with trajectory group membership; however, these associations were attenuated in multivariable models, suggesting shared variance between these factors, those retained in the final model (i.e., work/activity impairment, depression severity, episode duration), and response.

Several limitations should be noted, including the need to replicate the observed response patterns in other samples. The IRLGREY Study consisted of older adult outpatients and these findings cannot necessarily be generalized to middle-aged or younger patients or to older inpatients. Sufficient monitor-
Table 3. Final Multivariable Model Predicting Trajectory Groupa

<table>
<thead>
<tr>
<th>Variable</th>
<th>Responders</th>
<th>Nonresponders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rapid</td>
<td>Moderate,</td>
</tr>
<tr>
<td>Age</td>
<td>1.01 (0.95-1.09)</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Education</td>
<td>0.83 (0.68-1.01)</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Baseline MADRS score</td>
<td>0.56 (0.46-0.69)b</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Reduced sleep</td>
<td>0.83 (0.59-1.17)b</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Work/activity impairment</td>
<td>1.29 (0.63-2.65)</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Feelings of guilt</td>
<td>0.62 (0.32-1.19)</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>List recognition</td>
<td>2.22 (1.18-4.20)b</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Adequate prior pharmacological treatment</td>
<td>2.12 (0.73-6.16)</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Log of episode duration</td>
<td>1.11 (0.78-1.60)</td>
<td>1 [Reference]</td>
</tr>
</tbody>
</table>

Abbreviations: AOR, adjusted odds ratio; MADRS, Montgomery-Asberg Depression Rating Scale.

a Estimates were also adjusted for study site.

b P < .05.

Table 4. Final Model Parameters and Method to Calculate the Probability of Membership in Each of the 6 Groups Using Pretreatment Factorsa

<table>
<thead>
<tr>
<th>Parameter Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Rapid response</td>
</tr>
<tr>
<td>High, response</td>
</tr>
<tr>
<td>Moderate, mixed/nonresponse</td>
</tr>
<tr>
<td>High, nonresponse 1</td>
</tr>
<tr>
<td>High, nonresponse 2</td>
</tr>
</tbody>
</table>

Abbreviation: MADRS, Montgomery-Asberg Depression Rating Scale.

a Adjustment for study site in Table 3 was added for face validity and did not substantively alter findings; here, study site is not adjusted for to facilitate use of this model for patients not located at one of the 3 Incomplete Response in Late-Life Depression: Getting to Remission Study sites. All variables are defined as stated in Methods, including education (in years); episode duration (in weeks) natural log transformed; and subtest of the Repeatable Battery for the Assessment of Neuropsychological Status as age-normed z score. To calculate the estimated probability of membership in each group: (A) for each row above, raise e to the power of [constant + (age × value above)… + (list recognition × value above)]; (B) the estimated probability of membership in each group is equal to the number calculated for each row in step A divided by [1 + sum of all numbers (one for each row) calculated in step A]; and (C) the probability of being in the moderate, response group (reference) is equal to 1 minus the sum of probabilities of being in the 5 other groups (from step B). Taken from equation 3 in the article by Jones et al.32

ing, dose titration, and depression management support were provided to minimize the possibility that inadequate pharmacological care would confound our analysis; consequently, the patterns and frequency of response cannot necessarily generalize to other settings, eg, where monitoring and support are inadequate and attrition is more common. Analyses were prognostic and could not determine which factors, if any, moderate venlafaxine XR efficacy. We did not examine several other potentially important sources of variability in treatment response, including biological and social-contextual factors. Still, the comprehensive set of possible prognostic factors examined was large, which potentially increased the risk of type I error; the prognostic associations detected require confirmation in other samples. The severity of reduced sleep was associated with trajectory grouping; however, reliance on a single self-reported item and not a validated sleep questionnaire eliminates the possibility of comparison with prior studies, including prior depression treatment research that used objective sleep assessments (eg, the study by Troxel et al39). Nevertheless, our results are consistent with one recent study that found slower treatment responses among a patient subgroup with relatively fewer subjective sleep complaints and more guilt and activity impairment.40

Strengths of our study include application of group-based trajectory modeling to the largest open-label pharmacological trial for LLD conducted to date. The large cohort provided increased statistical power compared with prior studies, which may have enabled detection of distinct response patterns across the full range of pretreatment LLD severity. Another strength of our study is the multivariable analysis of a comprehensive set of prognostic factors, including both previously investigated (eg, depression severity) and novel (eg, individual depressive symptoms) response markers.
Conclusions

We found that even in a specialty care setting, patients with severe pretreatment depression regularly fail to respond to first-line pharmacotherapy. Nevertheless, because we observed treatment response in the presence of high baseline severity as well as treatment nonresponse in its absence, high depression severity alone may not be a completely accurate predictor of treatment response. Several other clinical prognostic factors likely contribute to the risk of treatment nonresponse, and variability in these prognostic factors potentially reflects important differences in major depressive disorder stage or phenotype. Intensive short-term monitoring and/or additional treatment strategies may be important for patients with severe depression, prolonged episode duration, and/or the specific LLD characteristics or comorbidities discussed earlier. Future investigations should examine the biological basis of these prognostic markers and test whether these factors moderate the efficacy of first-line pharmacological treatments. Isolating the specific psychological and neurobiological circuits whose dysfunction moderates treatment efficacy will greatly improve our understanding of the development, detection, and treatment of LLD.

Funding/Support: This work was supported by grants R01 MH083660 (University of Pittsburgh), R01 MH083648 (Washington University in St Louis), and R01 MH 083643 (Center for Addiction and Mental Health, University of Toronto) from the National Institute of Mental Health, center core grant P30 MH09333 from the National Institute of Mental Health, and the UPMC Endowment in Geriatric Psychiatry from the University of Pittsburgh Medical Center. Dr Smagula was supported by research training grants T32 AG000181 and T32 MH019986 from the National Institute of Mental Health. Dr Lenze was supported by the Taylor Family Institute for Innovative Psychiatric Research.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: Dr Smagula thanks his doctoral committee for their contribution to this work. Bobby L. Jones, PhD, University of Pittsburgh, Pittsburgh, Pennsylvania, provided expert consultation regarding the statistical methods used; he received no compensation.

REFERENCES

18. Pimontel MA, Culang-Reinlieb ME, Morimoto SS, Sneed Jr. Executive dysfunction and treatment


