Response rate profiles for major depressive disorder: Characterizing early response and longitudinal nonresponse

Mary Kelley, Emory University
Boadie Dunlop, Emory University
Charles B. Nemeroff, University of Miami
Adriana Lori, Emory University
Tania Carrillo-Roa, Max-Planck Institute of Psychiatry, Germany
Elisabeth Binder, Emory University
Michael Kutner, Emory University
Vivianne Aponte Rivera, Tulane University
Wade Craighead, Emory University
Helen Mayberg, Emory University

Journal Title: Depression and Anxiety
Volume: Volume 35, Number 10
Publisher: Wiley | 2018-10-01, Pages 992-1000
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1002/da.22832
Permanent URL: https://pid.emory.edu/ark:/25593/v2pwn

Final published version: http://dx.doi.org/10.1002/da.22832

Copyright information:
© 2018 Wiley Periodical, Inc.

Accessed April 22, 2020 7:03 PM EDT
Response rate profiles for major depressive disorder: Characterizing early response and longitudinal nonresponse

Mary E. Kelley1, Boadie W. Dunlop2, Charles B. Nemeroff3, Adriana Lori2, Tania Carrillo-Roa4, Elisabeth B. Binder2,4, Michael H. Kutner1, Vivianne Aponte Rivera5, W. Edward Craighead#2,6, Helen S. Mayberg#1,7

1Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia
2Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia
3Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, Florida
4Department of Translational Research in Psychiatry, Max-Planck Institute of Psychiatry, Munich, Germany
5Department of Psychiatry and Behavioral Sciences, Tulane University, New Orleans, Louisiana
6Department of Psychology, Emory University, Atlanta, Georgia
7Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Georgia

# These authors contributed equally to this work.

Abstract

Background: Definition of response is critical when seeking to establish valid predictors of treatment success. However, response at the end of study or endpoint only provides one view of the overall clinical picture that is relevant in testing for predictors. The current study employed a classification technique designed to group subjects based on their rate of change over time, while simultaneously addressing the issue of controlling for baseline severity.

Methods: A set of latent class trajectory analyses, incorporating baseline level of symptoms, were performed on a sample of 344 depressed patients from a clinical trial evaluating the efficacy of cognitive behavior therapy and two antidepressant medications (escitalopram and duloxetine) in patients with major depressive disorder.
**Results:** Although very few demographic and illness-related features were associated with response rate profiles, the aggregated effect of candidate genetic variants previously identified in large pharmacogenetic studies and meta-analyses showed a significant association with early remission as well as nonresponse. These same genetic scores showed a less compelling relationship with endpoint response categories. In addition, consistent nonresponse throughout the study treatment period was shown to occur in different subjects than endpoint nonresponse, which was verified by follow-up augmentation treatment outcomes.

**Conclusions:** When defining groups based on the rate of change, controlling for baseline depression severity may help to identify the clinically relevant distinctions of early response on one end and consistent nonresponse on the other.

**Keywords**

antidepressants; CBT/cognitive behavior therapy; depression; genetics; treatment

---

**1 | INTRODUCTION**

The utility of treatments for major depressive disorder (MDD) is most often defined by the gold standard outcomes of nonresponse, response, and remission after an acute treatment that typically lasts 6–12 weeks. In particular, remission (Frank et al., 1991) has been identified as the goal of treatment for MDD due to its strong association with restoration of functioning and durability of wellness (Keller, 2004). However, some patients achieve remission early in a treatment course while others show a more delayed improvement, suggesting different mechanisms of recovery may be involved across remitters. Thus, classifying all remitters as a single group may confound efforts to identify predictors of recovery. Identifying an early response has been shown to be useful in predicting later response to both medication and cognitive behavior therapy (CBT) (Ilardi & Craighead, 1994; Lewis, Simons, & Kim, 2012; Lutz, Stulz, & Kock, 2009; Szegedi et al., 2009; Tadic et al., 2010; Uher et al., 2010) and may also reflect placebo responsiveness, which is highly relevant for clinical trials (Kasper, Spadone, Verpillat, & Angst, 2006; Shelton et al., 2007). Perhaps more importantly, characterizing nonresponse has been shown to be beneficial when identifying biomarkers that can aid in making recommendations for different types of treatment (Dunlop, Rajendra et al., 2017; Dunlop, Kelley, McGrath, Craighead, & Mayberg, 2015; McGrath et al., 2013; 2014). More specifically, defining and characterizing consistent nonresponse throughout a treatment course may allow us to identify those patients who may need a switch to an alternative treatment (Holtzheimer & Mayberg, 2012).

Toward the goal of identifying different classes of treatment response, a number of studies have used latent-class trajectory analysis to identify groups of subjects with similar rates of change (slopes) in depressive symptoms over time; these classes can be conceptualized as response rate profiles (RRP). These analyses provide a mechanism for using all longitudinal data throughout the treatment period to classify subjects into groups as opposed to using only end-of-study response criterion. These studies have included various subgroups of patients and treatment modalities (Gueorguieva, Mallinckrodt, & Krystal, 2011; Lutz et al., 2009; Muthen, Asparouhov, Hunter, & Leuchter, 2011; Smagula et al., 2015; Stulz, Thase, Klein, Manber, & Crits-Christoph, 2010; Uher et al., 2010, 2011). In general, most studies...
comprised comparisons of antidepressant medications; of these, only some were randomized trials (Lutz et al., 2009; Stulz et al., 2010), and all were of 8–12 weeks in duration. Although various measures of depression were used to define symptom change, most showed evidence of a nonresponse group (Hunter, Muthen, Cook, & Leuchter, 2010; Lam, 2012; Muthen et al., 2011; Smagula et al., 2015; Thibodeau et al., 2015; Uher et al., 2010, 2011) and a clear “early response” group (Lutz et al., 2009; Muthen et al., 2011; Thibodeau et al., 2015; Uher et al., 2010, 2011). Notably, very few of these studies have identified biomarkers related to these response classes, with only one study identifying possible genetic markers (Hunter et al., 2010; Uher et al., 2010). In contrast, a number of genetic variants have shown to be associated with endpoint response to antidepressant medication (Breitenstein, Scheuer, & Holsboer, 2014), with some verified by more recent meta-analyses (Niitsu, Fabbri, Bentini, & Serretti, 2013; Porcelli, Fabbri, & Serretti, 2012). Much less work has examined the effects of genetic variants on psychotherapy treatment outcomes (Lester & Eley, 2013).

Most importantly, the majority of these studies have not made adjustment for the baseline severity part of the classification procedure. Baseline severity, therefore, remains the most significant predictor (Hunter et al., 2010; Lutz et al., 2009; Smagula et al., 2015; Stulz et al., 2010; Thibodeau et al., 2015; Wardenaar, Monden, Conradi, & de Jonge, 2015) as well as the most prominent characteristic of the classes. In fact, most of the descriptions of the various derived classes reflect this (Thibodeau et al., 2015), and a number of studies have noted that baseline severity was the main element of the classification, hindering profile interpretation (Smagula et al., 2015; Stulz et al., 2010). Moreover, high baseline severity was consistently associated with both nonresponse and response trajectories (Smagula et al., 2015; Thibodeau et al., 2015). This indicates that severity itself is not a consistent predictor in all patients, and may be the result of regression to the mean for a subset of patients (Benedetti, Carlino, & Pollo, 2011). Given the high correlation between measures of anxiety and depression, lack of adjustment for baseline severity may also be the reason behind the widely replicated finding of baseline anxiety as a consistent negative predictor of treatment response (Fava et al., 2008; Forand & DeRubeis, 2013; Gueorguieva et al., 2011; Lewis et al., 2012; Stulz et al., 2010). In fact, a post hoc adjustment for the level of baseline depression is often employed when evaluating potential predictors of treatment outcome (Hunter et al., 2010; Porcelli et al., 2012).

Thus, to define meaningful RRPs, it would be more beneficial to adjust for baseline when defining the response groups, rather than doing it post hoc, to control for extraneous factors such as regression to the mean. In the classification setting this addresses the question: Are there groups of subjects with similar change overtime, if they were roughly equivalent at baseline? (Laird, 1983). Thus, the current study derived RRP with the baseline effect statistically removed, and then tested for predictors of rate of response.

### 2 | MATERIALS AND METHODS

#### 2.1 | Study overview and participants

The current study evaluated data collected from the Emory Predictors of Remission in Depression to Individual and Combined Treatments (PReDICIT) study, for which the rationale, methods, and design have been previously published (Dunlop et al., 2012; Dunlop,
Kelley et al., 2017). The study was conducted through the Emory Mood and Anxiety Disorders Program in English-language clinics at Emory University and in Spanish-language clinic at Grady Hospital in Atlanta. All participants (n = 344) provided a written informed consent prior to beginning of the study procedures. The study enrolled men and women, aged 18–65 years, who met the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) criteria for current MDD and who had never previously received treatment for a mood disorder. Patients were randomized 1:1:1 to one of three possible phase 1 treatments over 12 weeks: (1) escitalopram (ESC): 10–20 mg/d, (2) duloxetine (DUL): 30–60 mg/d, or (3) CBT, 16 individual 1-hr sessions. To ensure equal allocation across treatment groups, prior to opening the study for enrollment the treatment assignment was generated using randomized permuted blocks. Randomization blocks were stratified for the English- and Spanish-language clinics. Blood for genetic analysis was collected at the baseline visit. Predefined outcome definitions and exclusion criteria are detailed in (Dunlop et al., 2012). A subset of the subjects (n = 251; 73%) completed the full 12 weeks of treatment. Nonremitters at the end of phase 1 were offered an additional 12-week phase 2 treatment with augmentation therapy: a 16-session course of CBT was added to those who had initially received medication, and escitalopram was provided to the patients who had not remitted with CBT. Study approval was granted by the Emory Institutional Review Board and the Grady Hospital Research Oversight Committee.

2.2 Genetic measures for analysis

To evaluate the biological relevance of the identified RRPs, we examined relationships between the RRPs and five genetic variants, selected on the basis of their previously identified associations with treatment outcomes in meta-analytic or large clinical studies. The genetic variants examined were: (1) serotonin transporter (5HTTLPR) (Kato & Serretti, 2010; Mrazek et al., 2009; Porcelli et al., 2012; Serretti, Kato, De Ronchi, & Kinoshita, 2007; (2) rs7997012 (5HTR2A) (McMahon et al., 2006; Niitsu et al., 2013; Paddock et al., 2007); (3) rs4713916 (FKBP5) (Lekman et al., 2008; Zou et al., 2010a); (4) rs1954787 GRIK4 (Paddock et al., 2007); and (5) rs6265 (BDNF) (Colle et al., 2015; Niitsu et al., 2013; Zou et al., 2010b). All of these variants have been shown to be associated with response to multiple classes of pharmacological antidepressants.

For the purpose of analysis, we identified the allele predictive of response for each candidate gene. Each subject has zero, one, or two copies with one or two copies considered a carrier. We then calculated genetic scores across all five genetic variants using: (1) the sum of response alleles per subject (0–10), as well as the (2) response carrier counts per subject (0–5). For the BDNF (rs6265) gene, the literature has identified both G allele carriers as well as GG homozygotes as response patterns. However, in this study 97.9% of subjects were G allele carriers, which did not allow differentiation across response groups; thus, the GG homozygotes comprised the “response” category. This grouping is supported by a recent clustering approach that showed stronger evidence for the homozygote as the response category (Kautzky et al., 2015). Further details are in the Supporting Information.
### 2.3 Statistical analysis

Analyses were performed on the Hamilton Rating Scale for Depression (HRSD) measured at baseline, weekly through week 6, and at weeks 8, 10, and 12. The primary analyses examined changes in HRSD total score over the 12 weeks of treatment. The traditional criteria reported in the published study were used for endpoint classification: remission (HRSD17 \( \leq 7 \)), response (\( \geq 50\% \) improvement without reaching remission), partial response (30–49% improvement), and nonresponse (<30% improvement).

Latent class analysis was used to group subject-level trajectories; groups were based on the intercept and slope of the change for each individual (Muthen & Shedden, 1999). The current analyses were performed using the “gllamm” add-on to Stata (Rabe-Hesketh et al., 2004). Although the goal for the current study was to obtain categories for baseline-controlled change in HRSD (bc-ΔHRSD), consistent with previous approaches and for comparison of results, an analysis was conducted without the baseline control (Supporting Information). Fit criteria, including Bayesian information criterion (BIC), and the BIC log Bayes factor, were examined to choose the appropriate number of classes from the multiple solutions, similar to (Smagula et al., 2015). In addition, model fit was examined using the average posterior probability of group membership. The baseline HRSD showed considerable variation (mean 19.8 ± 3.8, range: 15–33). All available data at each follow-up were used, that is, all 344 randomized subjects were part of the trajectory analysis; however, endpoint classes can only be compared for protocol completers (n = 251). The use of all randomized subjects refers to the less-biased mixed model solution that invokes the missing at random assumption for longitudinal data (Mallinckrodt et al., 2003). To determine the effects of missing data on the latent class analysis, we also derived classes on the “per protocol” subset of the data (n = 234) defined in (Dunlop, Kelley et al., 2017) for comparison.

All patients were combined across treatments and response profiles were obtained for the entire sample, as well as within treatment group to ensure that there were no profiles that were unique to treatment group. Combined profiles are preferred (Lutz et al., 2009; Muthen et al., 2011; Stulz et al., 2010; Uher et al., 2010), as they allow for comparison of response rates in different treatment groups as well as testing for treatment-specific predictors (i.e., treatment by response interactions).

Once trajectories were derived, the RRP s were then treated as an ordinal outcome variable representing time to response, and ordinal logistic regression was used to test demographics, related clinical variables, and candidate genotypes as possible predictors of response level. Although we present group means for completeness, the test itself can be viewed as a single measure of association between time to response and the predictor that takes advantage of the full sample size. Further analysis was performed to characterize the nature of the response group differences. In these analyses we approached the RRP s as a category, and thus standard Bonferroni correction for multiple comparisons of post hoc ANOVA tests were employed. No other corrections for multiple comparisons were made; instead all nominally significant results (\( P < 0.05 \)) were also discussed in terms of effect size. Although there are no formal calculations of effect size for ordinal measures, the comparisons at the highest and lowest levels of early remission (erem), and nonresponse (nresp), though
conservative, can provide a clinically relevant view of the detected differences. For categorical predictors, we show the relative risk (RR) for group comparisons at the these levels. For mean differences, we calculated Cohen’s d for the difference between the levels.

3 | RESULTS

3.1 | Response rate profile (RRP) definition

Linear trajectories were sufficient to summarize the classes as quadratic models provided no further increases in goodness of fit (data not shown). BIC criterion demonstrated that for the combined group analyses, the HRSD longitudinal data were best fit into five bc-ΔHRSD profiles (Table 1). Average posterior probabilities were high for all groups (mean = 0.81, range: 0.76–0.90), indicating the model fit was good. These bc-ΔHRSD profiles were labeled as: (1) early remission (3 weeks), (2) delayed remission (6 weeks) (3), gradual remission (12 weeks), (4) minimal response, and (5) no response (Figure 1). These RRP have equivalent baseline scores in the two nonresponse categories (profiles 4 and 5) as well as the three remission categories (profiles 1, 2, and 3), demonstrating the adjustment used to make the subjects comparable at baseline was successful.

Analysis within treatment group (i.e., CBT, ESC, DUL) exhibited a four class solution in each case representing the early remission, gradual remission, minimal response, and no response categories that were similar to those defined above. Thus, the delayed remission category was only evident in the combined solution, where there was enough power to detect it. More importantly, the treatment-specific analyses did not uncover any patterns that would preclude the use of the combined analyses. Similarly, the “per protocol” subset (n = 234, Supporting Information) provided the same overall pattern of results as the available case sample (n = 344). The classes from the two solutions correlated at 0.95 (Spearman’s rho), indicating no quantifiable effects of missing data on the classification.

Comparison of the RRP to endpoint categories (Table 2) revealed a significant association (Somers’ d = −0.64, P < .0005), but not duplication. Of those with endpoint nonresponse (n = 46) only 45.6% (n = 21) had consistent nonresponse throughout the period, identifying a unique group of subjects. In addition, we identified an early remission (12.5%) subgroup, distinct from other remitters. It is evident from the fitted correlations that baseline severity is much more strongly related to nonresponse rather than early response, and the adjusted baselines reflect that (Table 2).

3.2 | Clinical predictors

For testing of characteristics as possible predictors of RRP, we created an ordinal variable representing time to response (1 = early remission, 2 = delayed remission, 3 = gradual remission, 4 = minimal response, and 5 = no response). Notably, very few demographic variables were associated with RRP (Table 3). Ethnicity was a significant predictor of RRP, with Hispanics more likely to respond earlier when compared to non-Hispanic Whites (RRrem = 1.73, RResp = 0.48, Z = −2.35, P = 0.019) and non-Hispanic Blacks (RRrem = 1.45, RResp = 0.50, χ² = 5.09, df = 1, P = 0.024). Treatment group was also a significant predictor of RRP, with CBT subjects less likely to respond earlier when compared to patients
treated with escitalopram (RR_{erem} = 0.33, RR_{nresp} = 2.97, Z = 2.00, P = 0.045) or duloxetine (RR_{erem} = 0.47, RR_{nresp} = 1.88, \chi^2 = 3.92, df = 1, P = 0.048). Notably, we did not find an association with either baseline anxiety score or current anxiety diagnosis and the RRPs adjusted for baseline severity.

3.3 | Genetics predictors

The sums of response alleles (erem–nresp Cohen’s d = 0.64) and allele carriers (erem–nresp Cohen’s d = 0.80) of the five genetic variants previously associated with antidepressant treatment response were also associated with time to response (Table 4a). More specifically, there was a clear association between higher score (genetic “loading”) and earlier time to response, providing more evidence for a possible biological association. The endpoint response groups (Table 4b) did not exhibit such a consistent pattern, with the partial response group mean actually higher than the mean of the remitter group. Because all of the examined genetic variants were indicators of antidepressant medication response, we tested the combined medication group (n = 229) separate from the CBT group (n = 115). Although these data are not powered to be definitive in terms of interaction, the data do show that the pattern of association in the CBT group is consistent with that in the medication group. Post hoc comparisons of the means of the RRPs on the sum of the carrier alleles (Bonferonni corrected) indicated that only the early responders were differentiable from both the minimal response (P = 0.03) and nonresponse (P = 0.016) groups. In contrast, for endpoint response, none of the post hoc comparisons were significant. To assess the need for adjustment for population stratification, principal components (PCs) of the genetic relationship matrix were calculated (Yang, Lee, Goddard, & Visscher, 2011) and mapped to demographic characteristics. The first two PCs explained approximately 9% of the variance and clearly mapped to the three racial/ethnic categories of non-Hispanic White, Hispanic, and non-Hispanic Black, indicating the PCs can be used for racial/ethnic composition adjustment. The first component was associated with all genetic variants, but only the second component was associated with RRP. Thus, population stratification is unlikely to be a confounder of response prediction. Nevertheless, we proceeded to calculate the significance after adjusting for the first two components; the values changed only slightly (resp allele sum Z = −2.01, P = 0.045, resp carr count: Z = −2.56, P = 0.01).

3.4 | Augmentation success in nonremitters

To further evaluate the added value of the longitudinal classification, we examined the phase 2 treatment results for the 137 nonremitting subjects at the end of phase 1 treatment. Of the 137 endpoint nonremitters, the longitudinal RRPs identified 22 as consistent nonresponders, who might require alternative forms of treatment (i.e., a treatment intervention other than medications or psychotherapy); 73 minimal responders, who may only require augmentation of their initial treatment; and 42 longitudinal remitters who may only require a longer duration of their initial treatment. If we examine phase 2 outcomes in only the consistent and minimal nonresponse RRPs, we have data on 18 of 22 and 49 of 73 subjects. Using these data, the phase 2 odds of remission (OR = 3.36, \chi^2 = 3.87, P = 0.049) and response (OR = 4.80, \chi^2 = 6.89, P = 0.009) were significantly higher in the minimal responder group than in the consistent nonresponder group, supporting the validity of the classification for use in future studies.
The current analyses were designed to test the utility of deriving RRP s in depressed patients based on a growth mixture model that included a parameter controlling for baseline severity. Five RRP s were identified, including two trajectories of poor response and three trajectories for remission. A polygenic score representing a sum of previously identified “response” alleles was shown to be significantly related to RRP, with higher scores being associated with earlier response. The fact that the relationship is consistent across all levels of response is unique to this study, and is important as we acknowledge that depression response is not truly a binary phenomenon. However, identifying levels of early and nonresponse is particularly useful for treatment recommendations, and the data indicate that the genetic score is beneficial in identifying these groups. The score was not associated with endpoint response categories in the same way, exhibiting both an inconsistent relationship in the partial response category and no significant differences between remitters and nonresponders. These results suggest that RRP s derived in this way may provide more consistent predictors of treatment response in depression than traditional endpoint analyses.

The association between genetic score and response was similar among CBT- and medication-treated patients, indicating that the association between these single nucleotide polymorphisms (SNPs) and response in previous studies may be indicative of an overall propensity toward response rather than a treatment-specific biomarker. Notably, the only other trajectory analysis to report genetic predictors (Uher et al., 2010) also identified 5HTR2A alleles (rs9316233, rs2224721) as predictors of early response to escitalopram, however, no nonresponse groups were derived in that analysis, so it is difficult to compare results. Interestingly, although not yet replicated, the serotonin transporter (Kohen et al., 2011) and 5HTR2A (Kotte, McQuaid, & Kelsoe, 2007) polymorphisms evaluated in this study have also been associated with response to psychotherapy for depression. This provides additional support for the use of these genetic scores to identify early responders and nonresponders to conventional treatments.

The lack of demographic and clinical predictors of response is not surprising given very few have been replicated across studies (Dunlop, 2015; Weitz et al., 2015). This is similar to the results reported for endpoint categories (Dunlop, Kelley et al., 2017) in the PReDICT study. As stated in the introduction, the lack of association with baseline anxiety levels could be in part due to the combination of high correlation between baseline HRSA and HRSD scores (r = 0.65, P < 0.005) and lack of adjustment in past studies. This may also occur because the threshold for baseline HRSD in the current study was lower than in some pharmaceutical trials. However, the average baseline HRSD was approximately 20, with only 15% of the subjects having the minimum baseline score of 15. In addition, the distribution of the baseline HRSD was remarkably normal, which is a requirement of any of the longitudinal analyses that attempt to fit slope profiles over time. We feel that the spread of the severity levels of depression are comparable to what is seen in clinical practice, and thus our results are more generalizable than studies that apply a higher minimum score for inclusion.

Furthermore, there is considerable recent evidence that treatment benefits are not limited to those subjects with more severe scores (Fountoulakis, Veroniki,Siamouli,& Moller, 2013; Gibbons, Hur, Brown, Davis, & Mann, 2012; Weitz et al., 2015).
Further work is required to understand the implications of an early treatment response. There are differing opinions as to whether an early response is due to treatment (Tang & DeRubeis, 1999), other non-specific factors (Ilardi & Craighead, 1999, 1999; Thomas & Persons, 2013), or may represent placebo response in some cases. However, controlling for baseline severity, and thus for regression to the mean, does remove one source of this association, giving the trajectories a higher likelihood of being treatment related. The slower response to CBT versus antidepressant medications has been previously reported (Keller et al., 2000); nevertheless, the overall efficacy of the treatments does not differ by the end of 12 weeks of treatment.

Recent approaches for imaging biomarkers have demonstrated the utility of contrasting remitters and nonresponders at endpoint (Dunlop et al., 2015; McGrath et al., 2014). However, nonresponse at endpoint can have a number of different clinical implications, some of which would have different recommendations; for example, whether to augment with more standard treatments or switch to an intervention usually reserved for treatment-resistant depression. The fact that the polygenic score differentiated the two groups of longitudinal nonresponders from early remitters provides a potential biological indicator of the nature of nonresponse that, when used with additional markers, may provide clinical benefit through the use of a multivariate or tiered approach to response prediction. Although the polygenic score did not provide information specific to the minimal versus consistent nonresponse distinction, the phase 2 data support the use of these classifications for further biomarker studies. Although no classification system for MDD is likely to achieve 100% accuracy, there can still be great value in stratifying (Kapur, Phillips, & Insel, 2012) patients into risk groups for treatment response, and further exploration of the nonremission categories is warranted. Most likely, given the small number of these more debilitated patients, power will need to be borrowed across studies to derive true biomarkers of these subcategories of nonresponse that will inform treatment recommendations.

In conclusion, the use of trajectory analysis for the grouping of subjects by RRP provides important information not found in endpoint response categories, and controlling for baseline severity score allows for more meaningful RRRs that are not defined primarily by their baseline severity. Future work can utilize these RRRs to provide important groupings for the testing of treatment-specific predictors of early treatment response and nonresponse so that treatments can be more precisely targeted, thereby enhancing their effectiveness.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**ACKNOWLEDGMENTS**

We thank Flavia Mercado for her assistance in operationalizing the study at the Grady Hospital location.

Dr. Mayberg reports grants from NIMH, non-financial support from Eli Lilly, non-financial support from Forrest Labs, during the conduct of the study; other from St Jude Medical, Inc, outside the submitted work; In addition, Dr. Mayberg has a patent Method for Treating Depression Mood Disorders and Anxiety Disorders using Neuromodulation licensed to St. Jude Medical Inc.
Dr. Dunlop reports grants from Assurex Health, Acadia, Axsome, Janssen, Otsuka, Forest, and Takeda, outside the submitted work.

Dr. Craighead reports grants from NIMH, during the conduct of the study; other from Mary & John Brock, Fuqua Family Foundation, Hugarheill eff, John Wiley & Sons, personal fees from George West Mental Health Foundation, Scientific Advisory Board, AIM for Mental Health, outside the submitted work.

Dr. Nemeroff reports grants from NIH, personal fees from Xhale, Takeda, Prismic, Taisho Pharmaceutical Inc., Janssen, Magstim, Bracket/Clinicute, Total Pain Solutions, Gerson Lehrman Group, Fortress Biotech, Sunovion, Sumitomo Dainippon, Celgene, Seattle Genetics, Abbvie, OPKO Health, and Network Life Sciences, other from AFSP, BBRF, ADAA, Skyland Trail, RiverMend, LIR, Gratitude America, Method and devices for transdermal delivery of lithium (US 6,375,990B1), Method of assessing antidepressant drug therapy via transport inhibition of monoamine neurotransmitters by ex vivo assay (US 7,148,027B2), personal fees from Navitor Pharmaceuticals, TC MSO, ITI Inc., Antares, BI Gen Holdings, grants from Stanley Medical Research Institute, personal fees from Corcept Therapeutics Pharmaceuticals Company, outside the submitted work.

Funding information

National Institute of Mental Health, Grant/Award Numbers: K23MH086690, P50 MH077083, R01 MH080880

REFERENCES


Depress Anxiety. Author manuscript; available in PMC 2019 July 29.


FIGURE 1.
Treatment trajectory classes based on change in HRSD 17 total controlled for baseline severity (raw data means)
### TABLE 1
Values of BIC and BIC log Bayes factor for solutions with varying number of groups

<table>
<thead>
<tr>
<th>Number of groups</th>
<th>BIC</th>
<th>$2 \times [\text{BIC}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17037.55</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>16180.87</td>
<td>1713.35</td>
</tr>
<tr>
<td>3</td>
<td>15978.37</td>
<td>405.01</td>
</tr>
<tr>
<td>4</td>
<td>15886.60</td>
<td>183.53</td>
</tr>
<tr>
<td>5</td>
<td>15869.78</td>
<td>a</td>
</tr>
<tr>
<td>6</td>
<td>15887.41</td>
<td>−35.25</td>
</tr>
</tbody>
</table>

$^a$Minimum BIC = solution.
### TABLE 2

Response profiles defined by trajectory analysis (n = 344)

<table>
<thead>
<tr>
<th>Baseline controlled (bc-ΔHRSD) class</th>
<th>N</th>
<th>Total (%)</th>
<th>Baseline HRSD total</th>
<th>Fitted baseline&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Fitted ΔHRSD/wk</th>
<th>Fitted corr w/ baseline&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Correspondence with endpoint class: (% of total n = 251 completers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nonresponder (&lt; 30%)</td>
</tr>
<tr>
<td>Early remission</td>
<td>43</td>
<td>12.5</td>
<td>20.6 (3.6)</td>
<td>14.2</td>
<td>−1.27</td>
<td>0.239</td>
<td>0</td>
</tr>
<tr>
<td>Delayed remission</td>
<td>24</td>
<td>7.0</td>
<td>21.6 (4.3)</td>
<td>20.6</td>
<td>−1.67</td>
<td>0.211</td>
<td>0</td>
</tr>
<tr>
<td>Gradual remission</td>
<td>111</td>
<td>32.3</td>
<td>19.2 (3.5)</td>
<td>16.1</td>
<td>−0.95</td>
<td>0.453</td>
<td>3 (1.2%)</td>
</tr>
<tr>
<td>Minimal response</td>
<td>138</td>
<td>40.1</td>
<td>19.9 (4.0)</td>
<td>19.0</td>
<td>−0.62</td>
<td>0.536</td>
<td>22 (8.8%)</td>
</tr>
<tr>
<td>No response</td>
<td>28</td>
<td>8.1</td>
<td>18.4 (3.0)</td>
<td>18.5</td>
<td>−0.03</td>
<td>0.552</td>
<td>21 (8.4%)</td>
</tr>
<tr>
<td>Total</td>
<td>464</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46 (18.3%)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Fitted baseline for bc-ΔHRSD class is computed for baseline mean average HRSD17 for that group.

<sup>b</sup>Represents fitted correlation between baseline HRSD and ΔHRSD.
TABLE 3
Descriptives of be-ΔHRSD response profiles: mean (SD) or count (column%)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Early Remission</th>
<th>Delayed remission</th>
<th>Gradual remission</th>
<th>Minimal response</th>
<th>No response</th>
<th>All subjects</th>
<th>Z&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRSD baseline</td>
<td>20.6 (3.6)</td>
<td>21.6 (4.3)</td>
<td>19.2 (3.5)</td>
<td>19.9 (4.0)</td>
<td>18.4 (3.0)</td>
<td>19.8 (3.8)</td>
<td>−1.82</td>
<td>0.068</td>
</tr>
<tr>
<td>HRSA&lt;sub&gt;b&lt;/sub&gt; baseline</td>
<td>16.1 (5.3)</td>
<td>16.7 (4.6)</td>
<td>15.8 (5.4)</td>
<td>16.2 (5.3)</td>
<td>15.6 (4.6)</td>
<td>16.1 (5.2)</td>
<td>−0.11</td>
<td>0.909</td>
</tr>
<tr>
<td>Age of onset MDD</td>
<td>30.0 (13.3)</td>
<td>28.9 (13.8)</td>
<td>30.1 (13.0)</td>
<td>31.2 (15.0)</td>
<td>33.3 (16.3)</td>
<td>30.7 (14.2)</td>
<td>1.2</td>
<td>0.229</td>
</tr>
<tr>
<td>CTQ&lt;sub&gt;c&lt;/sub&gt; total</td>
<td>47.4 (16.2)</td>
<td>44.3 (14.4)</td>
<td>44.8 (16.4)</td>
<td>46.3 (15.2)</td>
<td>44.0 (13.5)</td>
<td>45.6 (15.5)</td>
<td>−0.14</td>
<td>0.888</td>
</tr>
<tr>
<td>Male</td>
<td>20 (46.5%)</td>
<td>12 (50.0%)</td>
<td>47 (42.3%)</td>
<td>58 (42.0%)</td>
<td>11 (39.3%)</td>
<td>148 (43.0%)</td>
<td>−0.76</td>
<td>0.45</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>15 (38.5%)</td>
<td>11 (50.0%)</td>
<td>51 (47.7%)</td>
<td>64 (51.2%)</td>
<td>15 (57.7%)</td>
<td>156 (48.9%)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>7 (17.9%)</td>
<td>3 (13.6%)</td>
<td>17 (15.9%)</td>
<td>28 (22.4%)</td>
<td>6 (23.1%)</td>
<td>61 (19.1%)</td>
<td>0.46</td>
<td>0.649</td>
</tr>
<tr>
<td>Hispanic</td>
<td>17 (43.6%)</td>
<td>8 (36.4%)</td>
<td>39 (36.4%)</td>
<td>33 (26.4%)</td>
<td>15 (19.2%)</td>
<td>102 (32.0%)</td>
<td>−2.35</td>
<td>0.019</td>
</tr>
<tr>
<td>Previous episodes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21 (50.5%)</td>
<td>13 (54.2%)</td>
<td>65 (59.1%)</td>
<td>60 (44.4%)</td>
<td>18 (64.3%)</td>
<td>177 (52.2%)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2</td>
<td>9 (21.4%)</td>
<td>5 (20.8%)</td>
<td>16 (14.5%)</td>
<td>32 (23.7%)</td>
<td>1 (3.6%)</td>
<td>63 (18.6%)</td>
<td>0.03</td>
<td>0.974</td>
</tr>
<tr>
<td>3+</td>
<td>12 (28.6%)</td>
<td>6 (25.0%)</td>
<td>29 (26.4%)</td>
<td>43 (31.9%)</td>
<td>9 (32.1%)</td>
<td>99 (29.2%)</td>
<td>0.87</td>
<td>0.384</td>
</tr>
<tr>
<td>Current anxiety disorder</td>
<td>12 (27.9%)</td>
<td>8 (33.3%)</td>
<td>50 (45.0%)</td>
<td>59 (42.8%)</td>
<td>10 (35.7%)</td>
<td>139 (40.4%)</td>
<td>0.91</td>
<td>0.361</td>
</tr>
<tr>
<td>Melancholic subtype</td>
<td>23 (53.5%)</td>
<td>15 (62.5%)</td>
<td>62 (55.9%)</td>
<td>77 (55.8%)</td>
<td>16 (57.1%)</td>
<td>193 (56.1%)</td>
<td>0.02</td>
<td>0.984</td>
</tr>
<tr>
<td>Chronic (current episode ≥ 2 years)</td>
<td>15 (35.7%)</td>
<td>5 (21.7%)</td>
<td>30 (27.3%)</td>
<td>46 (34.3%)</td>
<td>10 (35.7%)</td>
<td>106 (31.5%)</td>
<td>0.88</td>
<td>0.381</td>
</tr>
<tr>
<td>Full time employment</td>
<td>24 (57.1%)</td>
<td>14 (58.3%)</td>
<td>40 (36.0%)</td>
<td>68 (49.6%)</td>
<td>11 (39.3%)</td>
<td>157 (45.9%)</td>
<td>−0.49</td>
<td>0.621</td>
</tr>
<tr>
<td>Family hx mood disorder</td>
<td>17 (39.5%)</td>
<td>10 (41.7%)</td>
<td>45 (40.5%)</td>
<td>60 (43.5%)</td>
<td>7 (25.0%)</td>
<td>139 (40.4%)</td>
<td>−0.41</td>
<td>0.684</td>
</tr>
<tr>
<td>Treatment group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escitalopram</td>
<td>21 (48.8%)</td>
<td>5 (20.8%)</td>
<td>34 (30.6%)</td>
<td>40 (35.5%)</td>
<td>5 (17.9%)</td>
<td>114 (33.1%)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>CBT</td>
<td>7 (16.3%)</td>
<td>8 (33.3%)</td>
<td>39 (35.1%)</td>
<td>46 (33.3%)</td>
<td>15 (53.6%)</td>
<td>115 (33.4%)</td>
<td>2.00</td>
<td>0.045</td>
</tr>
<tr>
<td>Duloxetine</td>
<td>15 (34.9%)</td>
<td>11 (45.8%)</td>
<td>38 (34.2%)</td>
<td>43 (31.2%)</td>
<td>8 (28.6%)</td>
<td>115 (33.4%)</td>
<td>0.03</td>
<td>0.972</td>
</tr>
</tbody>
</table>

<sup>a</sup> Tests of significance are for prediction of response profile as an ordinal outcome measure (1–5); comparison of “all subjects” column to response category indicates the direction of the differences.

<sup>b</sup> Hamilton Rating Scale for Anxiety.

<sup>c</sup> Childhood Trauma Questionnaire.
TABLE 4

DNA SNP allele counts of (a) bc-ΔHRSD response profiles compared to (b) endpoint response: mean (SD)

(a)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Subjects</th>
<th>1 Early Remission (n = 41)</th>
<th>2 Delayed Remission (n = 24)</th>
<th>3 Gradual Remission (n = 107)</th>
<th>4 Minimal Response (n = 133)</th>
<th>5 No Response (n = 25)</th>
<th>Z&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resp allele count</td>
<td>All (n = 330)</td>
<td>4.46 (1.23)</td>
<td>4.08 (1.18)</td>
<td>3.92 (1.26)</td>
<td>3.92 (1.43)</td>
<td>3.60 (1.35)</td>
<td>−2.20</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>CBT (n = 109)</td>
<td>4.85 (0.90)</td>
<td>4.38 (1.06)</td>
<td>4.03 (1.16)</td>
<td>4.11 (3.83)</td>
<td>3.77 (1.24)</td>
<td>−1.21</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>med (n = 221)</td>
<td>4.38 (1.28)</td>
<td>3.93 (1.24)</td>
<td>3.86 (1.31)</td>
<td>3.83 (1.37)</td>
<td>3.42 (1.50)</td>
<td>−2.04</td>
<td>0.041</td>
</tr>
<tr>
<td>Resp allele carr sum</td>
<td>All (n = 330)</td>
<td>3.12 (0.81)</td>
<td>2.71 (0.86)</td>
<td>2.71 (0.95)</td>
<td>2.62 (0.97)</td>
<td>2.36 (1.04)</td>
<td>−2.99</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>CBT (n = 109)</td>
<td>3.14 (0.69)</td>
<td>3.00 (0.93)</td>
<td>2.78 (0.83)</td>
<td>2.76 (1.03)</td>
<td>2.38 (1.12)</td>
<td>−1.62</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>med (n = 221)</td>
<td>3.12 (0.84)</td>
<td>2.56 (0.81)</td>
<td>2.68 (1.01)</td>
<td>2.55 (0.93)</td>
<td>2.33 (0.98)</td>
<td>−2.69</td>
<td>0.007</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1 Remission (n = 112)</th>
<th>2 Response (n = 51)</th>
<th>3 Partial Response (n = 35)</th>
<th>4 No Response (n = 42)</th>
<th>Z&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resp allele count</td>
<td>4.18 (1.23)</td>
<td>3.96 (1.34)</td>
<td>4.31 (1.30)</td>
<td>3.64 (1.34)</td>
<td>1.63</td>
<td>0.102</td>
</tr>
<tr>
<td>Resp allele carr sum</td>
<td>2.84 (0.90)</td>
<td>2.71 (1.04)</td>
<td>2.91 (0.85)</td>
<td>2.45 (0.99)</td>
<td>1.63</td>
<td>0.102</td>
</tr>
</tbody>
</table>

<sup>a</sup>Tests of significance are for prediction of response profile as an ordinal outcome measure (1–5).