A Machine Learning Approach to Identifying Placebo Responders in Late-Life Depression Trials

Sigal Zilcha-Mano, PhD[*],

Department of Psychology, University of Haifa, Mount Carmel, Haifa 31905, Israel

Steven P. Roose, MD,

Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute

Patrick J. Brown, PhD, and

Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute

Bret R. Rutherford, MD

Columbia University College of Physicians and Surgeons, New York State Psychiatric Institute

Abstract

Objectives—Despite efforts to identify characteristics associated with medication-placebo differences in antidepressant trials, few consistent findings have emerged to guide participant selection in drug development settings and differential therapeutics in clinical practice. Limitations in the methodologies used, particularly searching for a single moderator while treating all other variables as noise, may partially explain the failure to generate consistent results. The present study tested whether interactions between pre-treatment patient characteristics, rather than a single-variable solution, may better predict who is most likely to benefit from placebo vs. medication.

Methods—We analyzed data from 174 patients aged 75 years and older with unipolar depression who were randomly assigned to citalopram or placebo. We conducted model-based recursive partitioning analysis to identify the most robust significant moderators of placebo vs. citalopram response.

Results—The greatest signal detection between medication and placebo in favor of medication was among patients with fewer years of education (12) who suffered from a longer duration of depression since their first episode (> 3.47 years) (B=2.53, $t_{(32)}$=3.01, p =.004). Compared to medication, placebo had the greatest response for those who were more educated (> 12 years), to the point where placebo almost outperformed medication (B=−0.57, $t_{(96)}$=−1.90, p=.06).

---

[*]corresponding author: Telephone: 972-4-8249047, 972-52-3385944, sigalzil@gmail.com.

Conclusions—Machine learning approaches capable of evaluating the contributions of multiple predictor variables may be a promising methodology for identifying placebo vs. medication responders. Duration of depression and education should be considered in the efforts to modulate placebo magnitude in drug development settings and in clinical practice.

Keywords

placebo effect; treatment moderators; placebo responders; depression; personalized medicine

Placebo effects are substantial in the treatment of major depressive disorder (MDD), to the extent that it has become increasingly difficult and expensive to develop antidepressant medications able to outperform placebo (1–2). The increase in trial failures observed in the past decades appears to be the result of rising placebo response rather than of declining medication response (3–4). Placebo response rates have risen at an average rate of 7% per decade over the past 30 years, decreasing by half the average difference between antidepressant medication and placebo observed in published antidepressant trials between 1982 and 2008 (5). High placebo response reduces medication-placebo differences and leads investigators to make methodological modifications (i.e., use multiple study sites to increase sample size) that increase measurement error, both of which make it more difficult to demonstrate a statistically significant benefit of a putative antidepressant agent over placebo (6). Developing more efficacious medications to treat MDD is a particularly urgent need in older adults because late-life depression is often chronic, recurrent, and less responsive to antidepressant medication than MDD in younger individuals (7). A recent meta-analysis showed that only 54.5% of published RCTs report significant benefits for medication over placebo in late-life depression. At least some of the remaining 45.5% are likely false negatives owing to elevated placebo response (8).

The fact that high placebo response hinders the development and eventual approval of new antidepressants has led most pharmaceutical companies and academic researchers to undertake initiatives aimed at reducing placebo response (9–11). Thus, to improve signal detection and facilitate drug development, identifying clinical and demographic characteristics of placebo vs. medication responders has been one of the main aims of placebo research in the last decades. Brown and colleagues (12) initially identified short episode duration, few previous episodes, good response to previous antidepressant treatment, and low overall symptom severity as key determinants of increased placebo response. Other potential moderators have been identified, such as gender (13), age (14), and education (15). Recently, Weimer et al. (16) conducted a comprehensive review of 31 meta-analyses and systematic reviews of more than 500 randomized placebo-controlled trials in various areas of psychiatry, to identify consistent moderators across studies. Based on their review, only one patient characteristic was found to be consistently linked to increased placebo responses: low baseline severity of symptoms.

The cumulative findings suggest that although the search for a single factor to explain variability in placebo response helped identify important potential moderators, it also produced little consistency and many mixed results across studies (16). One reason for these divergent results may be that the search for a single moderator treats all other variables as

merely noise, while it is more intuitive to hypothesize that no single factor is as important in predicting placebo response as a set of interrelated ones. Moreover, traditional approaches to subgroup analysis, which test each moderating factor as a separate hypothesis, can lead to erroneous conclusions because of multiple comparisons (inflated type I errors), model misspecification, and multicollinearity. Findings may also be affected by publication bias, because the statistically significant moderators have better chance of being reported in the literature.

Novel, systematic approaches to subgroup analysis have recently shown differential effects for different treatment conditions across patient subgroups (17–18). Of note, these differential effects were found in trials in which the primary outcome analyses (which ignored differences between subgroups of patients) failed to find any differences between conditions (17). Such methods for identifying clinical profiles of patients showing differential response across treatments demonstrate the utility of integrating information from distinct moderators vs. the use of a single moderator, in youth (19) as well as in late-life depression (20). In the present study, we apply machine learning methods to predict placebo vs. medication response in an RCT comparing medication with placebo in depressed patients aged 75 and older. This data-driven approach identifies a set of moderators (rather than a single predictor) that together can significantly determine, better than any single moderator in the data, which patients benefit from placebo. For a pool of potential moderators, we chose those previously identified in the empirical literature (16).

Method

Sample and clinical trial procedures

The procedures used in this multi-site, placebo-controlled RCT have been previously described (21). Briefly, 174 community-dwelling men and women aged 75 years or older, who met DSM-IV criteria (based on a SCID interview) for non-psychotic unipolar depression (single or recurrent), with a baseline 24-item Hamilton Rating Scale for Depression (HRSD; 22) score ≥ 20, participated in this 8-week RCT. All patients began the trial with a one-week, single-blind placebo lead-in, with the baseline visit conducted at the end of the lead-in period. At 15 centers, patients were randomized to citalopram (20 mg/d) or matched placebo at a ratio of 1:1, if they continued to meet inclusion criteria at the end of the placebo lead-in period. At the end of the fourth week, patients with an HRSD score > 10 had their medication dose increased to two pills per day, i.e., 40 mg of citalopram, or two placebo pills. Clinical assessments were conducted at baseline and at weeks 1, 2, 3, 4, 5, 6, and 8 (final week). For this analysis, weekly assessments of the HRSD were used as the dependent variable, together with the following potential baseline moderators: age, gender, education, duration of illness (current age minus age at onset, in years), symptom severity (assessed by HRSD), anxiety levels (assessed by HAMA, 23), Mini-Mental State Exam (MMSE, 17) to estimate global cognitive functioning (24), Stroop interference scores effect to assess the response inhibition component of executive functioning (25), the WAIS-III Digit Symbol Subtest as a measure of psychomotor speed (26), and the Buschke Selective Reminding Test (SRT) as a measure of verbal learning (27). Intake assessment of the instrumental activities of daily living (IADL; 28) was also added to the model as a potential

moderator based on a previous analysis demonstrating its ability to predict trajectories of symptom development in the population of elderly patients with MDD (29).

Statistical Analyses

Overview—The methods described below extend our previous successful application of novel machine learning approach to predicting dropout from placebo vs. antidepressant medication vs. psychotherapy (30). The outcome variable in the present study was HRSD score, measured on a weekly basis, from pre- to post-treatment. The data were thus hierarchically nested, with sessions within patients. To account for this non-independence of the data and to prevent inflation of the effects (31–32), we used the SAS PROC MIXED procedure (33), with level 1 as the session level and level 2 as the patient level. The model with the best fit for the outcome variable, based on the Akaike Information Criterion (AIC), was the one with a fixed effect of time, random intercept, and random slope of time. We used this model to estimate personal time trends for further analysis (the estimated individual random slope). In other words, each patient-specific slope of HRSD, across the course of treatment (assessed weekly, from pre- to post-treatment) was used as the outcome variable in the analyses.

Identifying the strongest moderators—To identify the strongest moderators of the association between treatment condition (medication vs. placebo) and outcome, we used the bootstrap aggregation of model-based recursive partitioning by the random forest algorithm, as implemented in the R package "mobForest" (version 1.2; 34). In this method, a thousand model-based trees (i.e., pathways for determining which variables best moderate the effect of treatment) were constructed based on bootstrapped samples from the primary dataset. For each tree, the model-based recursive partitioning searched for binary splits in the sample that result in model parameters on one side of the split being most different from those on the other side. We used a random sample of partitioning variables for splitting at each node (i.e., potential split-point). In each leaf (i.e., split) of the tree, we estimated the outcome for medication vs. placebo. Final model predictions were obtained by aggregation across the trees. The minimum alpha level for splits was set to 0.05, and the minimum leaf size for splitting was set to 30 patients. We used the following characteristics as potential baseline moderators: age, gender, education, duration of illness, HRSD, HAMA, IADL, MMSE, Stroop interference scores effect, WAIS-III digit symbol raw score, and Buschke short term memory scores.

Estimating the importance of potential moderators—To identify the strength of potential moderator splits, we constructed a variable-importance plot using the conditional permutation scheme (35), involving, within each tree, predictions for patients *left out* of the building of a given tree. To rank the moderators according to their importance in producing accurate predictions, we calculated an *importance statistic* that reflects the contribution of each variable to classifying or predicting the target variable. The importance statistic is a way of estimating the out-of-sample contribution to prediction of all tested variables. The statistic reflects the improvement in prediction using the variable in cases "held out" of a given bootstrapped sample, compared to using permutations of "fake" data to make the same prediction. We tested the strength of the predictive value of a particular variable by

examining which variables had an importance statistic above the permutated data. Although the bootstrapped scheme is exploratory, using it to select variables may result in stable predictors, often less sensitive to the unique features of a given data set.

Identifying a specific model—In the final step, we conducted decision-tree analyses with the variables found most important in the previous stage, applying the "mob" function of the R "party" package (36) to identify a model for partitioning the effect of treatment/placebo on outcome. The M-fluctuation (Mf) tests were implemented to assess the moderation effects. Based on this method, the parameter estimate is computed once for the full sample (assuming H_0 is true) together with a corresponding empirical fluctuation process that captures departures from stability. The degrees of freedom for these tests were estimated as the number of splits (37). This analysis used model-based recursive partitioning, fitting the best partitioning by M-fluctuation tests (M_f) for a given linear relationship, using Bonferroni correction for multiple tests, and providing a linear regression solution for each node of the final model. To express the results in a clinically meaningful form, we calculated the number needed to treat (NNT) for medication vs. placebo in the overall sample and subgroups, stratified by significant moderators. The NNT was calculated as the inverse of the difference between response rates in the medication and placebo conditions.

Results

Clinical and demographic characteristics of participants and clinical trial results

Eighty-four patients were randomized to citalopram and 90 to placebo. Fifty-eight percent of participants were women; mean age was 79.6 years (SD = 4.4), and the mean baseline HDRS score was 24.3 (SD = 4.1). The remission rate, defined as a final Hamilton depression scale score <10, was 35% for the citalopram group and 33% for the placebo group. Type of treatment did not significantly predict treatment outcome (18). As reported in the main outcome paper, the site where the treatment was administered was found to be a significant predictor of treatment outcome (18).

Machine learning analyses

The random forest analysis identified education, duration of depression, and baseline HRSD as important moderators of individual patients' HRSD slopes. Adding study site yielded the same results. Figure 1 shows the resulting variable-importance plot. Next, the three identified variables were entered into the "mob" decision tree analyses. Figure 2 presents the tree for the moderators of the effect of medication vs. placebo on outcome. M-fluctuation (Mf) tests were carried out to test the moderation effects. The decision tree analysis revealed a significant first split in patients' level of education (M_f = 12.59, DF=1, p = .03), and a second split in the duration of depression (M_f = 15.84, DF=1, p = .003).

For each final node, regression analysis was conducted to estimate the relationship between treatment condition and outcome. We used t-tests to assess the significance of the relationships. The dependent variable in the model was the individual HRSD slope. Thus, the coefficient refers to the slope of medication vs. placebo. In other words, the β

coefficients can be interpreted as the difference between the conditions in the HRSD slopes. The analyses revealed that the medication efficacy was greatest in patients with fewer years of education (12), who also suffered from a longer duration of depression (> 3.47 years) (B = 2.53, $t_{(32)}$ = 3.01, p = .004). Compared with medication, placebo had the greatest effect on patients who were more educated (> 12), to the point of having a marginally significant stronger effect of placebo than of medication (B = −0.57, $t_{(96)}$ = −1.90, p = .06). No significant differences were evident between medication and placebo for the subgroup of patients with fewer years of education (12), who also suffered from a shorter duration of depression (3.47 years) (B = .85, $t_{(40)}$ = 1.16, p = .25).

Given that baseline HRSD severity was found to be a significant moderator in previous studies, we examined, in a *post hoc* analysis, whether after omitting the duration of depression variable from the model, baseline HRSD becomes a robust moderator. In this analysis, all predictors were identical to those in the previous analyses, except that the duration of depression variable was omitted. As expected, baseline HRSD and education were found to be the most robust moderators according to the resulting variable-importance plot.

To test the robustness of the study results, we repeated the analyses using clinical response (reduction in HRSD scores of 50% or more at the final assessment), rather than the individual patient HRSD slopes, as the outcome variable. The findings were similar to those reported above, with education, duration of depression, and baseline HRSD being the variables chosen by the random forest analyses; resulting in an identical tree.

To illustrate the clinical utility of employing these machine learning methods to identifying multiple, interacting moderators of medication vs. placebo differences, we calculated the NNT for the entire study sample (N=174) and compared it to the NNTs found for the most drug responsive subgroups identified in a traditional single variable moderator analysis[1] vs. the multivariable approach to moderator analysis described here, which was to focus on patients with fewer years of education (12) who suffered from a longer duration of depression since their first episode (> 3.47 years). The NNT for citalopram vs. placebo for the overall study was 315 [1/[(32/84) – (34/90)]], reflecting the fact that the parent clinical trial failed to find any significant drug vs. placebo differences. Choosing the best single moderator of drug vs. placebo differences (which was education level, as identified by the random forest algorithm) reduced the NNT to 17 [1/[(27/53) – (20/55)]]. The methods described in this manuscript permitted the identification of multiple moderators that further reduced the NNT for drug vs. placebo treatment to 4 [1/[(7/16) – (3/18)]], illustrating the great utility of this methodology in facilitating signal detection. As the ratios indicate, the placebo response rate in the identified subgroup was reduced to .16 (3/18), compared to .37 in the full sample (34/90). The drug response rate showed a small increase to .43 (7/16), compared to .38 in the full sample (32/84)[2].

---

[1]We choose the strongest single variable moderator in our data, as identified using the random forest algorithm, which is education. We followed Smagula et al. (20) to calculate the cutoff by plotting the moderator against the expected (predicted) outcome level, stratified by treatment assignment.
[2]To calculate effect sizes, we applied the method proposed by Petkova (38), which constructs a composite moderator (defined as a linear combination of pre-treatment patient characteristics) to generate a parsimonious single-index approach for making

Discussion

The present study demonstrates how machine learning methods can help identify pre-treatment characteristics of patients most likely to respond to drug vs. placebo. Our findings suggest that the subgroup benefitting most from citalopram, in which signal detection for the efficacy of medication over placebo is strongest, is that of patients who have fewer years of education and a longer duration of depression since their first depressive episode. For this subpopulation, citalopram was significantly more effective than placebo in reducing symptoms of depression, a finding that is obscured in the overall comparison between medication and placebo in this failed trial. In fact, for this subpopulation the NNT was 4, which compares to 315 in the total sample. The increased drug-placebo difference was due primarily to decreased placebo response (from .37 to .16) rather than to increased medication response in this subpopulation. The findings also suggest that the population that benefits the most from placebo, almost to the point of benefiting more from it than from medication, is that of patients who have more than 12 years of education. Importantly, these findings were replicated when we used a categorical outcome variable of responders vs. non-responders, in addition to patient-specific HRSD trajectories across treatment.

The current findings are consistent with previous studies that identified both education and duration of depression as potential moderators of placebo vs. antidepressant response (16). Shorter disease duration was found to be related to greater placebo response in schizophrenia (40) and other psychoses (41), anxiety disorders (in children and adolescents, 42), ADHD (15), and depression (12), including late life depression (43). Similarly, education was also found to be a significant indicator of placebo response in at least one population (15). An important contribution of the present findings is that the two moderators that have been identified separately in the literature were found here to interact to best predict a differential placebo vs. medication effect.

Although in the present study pre-treatment depression severity was found to be a potentially strong moderator of treatment vs. placebo differences in outcome in the random forest analysis, it was not revealed as a significant moderator in the tree analysis. This may be explained by the correlation between pre-treatment depression severity and duration of the depression, which was one of the moderators in the final model. Our *post hoc* analyses support this interpretation and demonstrate that when duration of depression was excluded from the analyses, pre-treatment depression emerged as an important moderator in the tree analysis as well.

It is interesting to speculate whether expectancy effects, which have been shown to be important mediators of placebo effects in antidepressant trails (44), may partially explain the effects identified for education. Expectancy effects require relatively intact cognition (45), so individuals with less education and thereby diminished cognitive reserve may be less able to

---

individualized treatment decisions. Next, we calculated the effect size of the moderation effect, as introduced by Kraemer et al., (39), for each of the models. This effect size may be described as the proportion of the outcome variance (after removing the variance due to treatment) that is explained by the different relationships between outcome and moderator in the two treatment groups. The resulting effect size is a number between −1 and 1, with a null value of 0, where greater magnitudes indicate stronger moderation. Using this method, the composite moderation effect size was 0.17, the moderation for education was 0.12, and for duration of depression −0.07

benefit from them, as has been demonstrated in several meta-analyses (46–47). Longer illness duration may further diminish placebo response by making patients less likely to spontaneously remit within an acute time frame. Although patient expectancy and spontaneous remission contribute to medication response as well, their effects may be diminished by a ceiling placed on medication response by the number of enrolled patients who are categorical non-responders, misdiagnosed, or lost to follow-up. Similar phenomena of greater effect of a mechanism of change variable in placebo than in medication were also found in a previous study focusing on alliance (48). It appears that level of education can capture a unique protective factor that cannot be detected by other measures of cognitive functioning, such as those that can be assessed using the Stroop and the Buschke tests. Future studies are needed to further investigate the unique contribution reflected in the level of education among the elderly.

The machine learning approach used in the present study goes beyond previous research, taking into account interactions between moderators to better capture the richness of human complexity when seeking to identify medication vs. placebo responders. This approach has the potential to explain the inconsistencies found in previous studies, which focused on a single moderator, for example, education (16, 43). The present findings suggest that it is not enough to look at the duration of depression alone, and that the effect of this variable must be considered in the context of education. The machine learning approach used in the present study increases the likelihood of future studies being able to replicate the findings. The identified moderators were selected by random forest bootstrapping, based on their internal consistency across the sample. Predictions were made with leave-one-out cross-validation, enhancing the possibility of these relationships being replicated out-of-sample.

The clinical applications of findings emerging from this exploratory approach are contingent upon their validation in future studies (17). If prospective studies support the present findings, their conceptual and clinical implications are immense. The results of this study could inform the design of future phase III trials in drug development settings (49). When the object is to limit placebo response and find medication-placebo differences if they truly exist, one may consider modifying the selection criteria to select relatively longer duration depression, rather than requiring two-week duration of illness to make a diagnosis of MDD. Similarly, it may be useful to routinely measure the level of educational attainment in study samples, and consider stratifying the sample on the basis of education. Knowing who may benefit most from placebo can pave the way to understanding the mechanisms underlying the placebo effect, maximizing its effect in clinical practice. It is also instrumental in enabling mental health practitioners to select the treatment that is expected to offer the greatest likelihood of success for each individual patient. Additionally, focusing RCTs testing the efficacy of citalopram on patients who potentially benefit most from it can reduce the likelihood of failed trials and contribute to progress toward personalized treatment. This can assist in the development of better antidepressant medication as a result of better signal detection.

Overall, the present study is the first to use a machine learning approach to systematically examine how interactions between moderators can better capture the richness of human complexity when seeking to identify medication vs. placebo responders. It represents an

important step forward in the effort to identify patients likely to benefit most from the placebo effect, making it possible to maximize the placebo effect in the community. It also holds great promise for more sophisticated RCTs of antidepressant medications, advancing the field toward personalized treatment (50).

Acknowledgments

References

1. Cressey D. Psychopharmacology in crisis. Nature. 2011; 20:11–10.

2. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry. 2015; 14:294–300. [PubMed: 26407778]

3. Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA. Placebo Response in Randomized Controlled Trials of Antidepressants for Pediatric Major Depressive Disorder. Am J Psychiatry. 2009; 166:42–49. [PubMed: 19047322]

4. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo Response in Studies of Major Depression: Variable, Substantial, and Growing. JAMA. 2002; 287:1840–1847. [PubMed: 11939870]

5. Khan A, Bhat A, Kolts R, Thase ME, Brown W. Why has the Antidepressant-Placebo Difference in Antidepressant Clinical Trials Diminished over the Past Three Decades? CNS Neuroscience & Therapeutics. 2010; 16:217–226. [PubMed: 20406269]

6. Rutherford BR, Roose SP. A Model of Placebo Effects in Antidepressant Clinical Trials. Am J Psychiatry. 2013; 170:723–733. [PubMed: 23318413]

7. Rutherford RR, Taylor WD, Brown PJ, et al. Biological aging and the future of geriatric psychiatry. J Grontol A Biol Sci Med Sci. 2012; 72:343–352.

8. Rutherford BR, Tandler J, Brown PJ, et al. Clinic visits in late-life depression trials: effects on signal detection and therapeutic outcome. Am J Psychiatry. 2014; 22:1452–1461.

9. Fava M, et al. The Problem of the Placebo Response in Clinical Trials for Psychiatric Disorders: Culprits, Possible Remedies, and a Novel Study Design Approach. Psychother Psychosom. 2003; 72:115–27. [PubMed: 12707478]

10. Huang X, Tamura R. Comparison of Test Statistics for the Sequential Parallel Design. Stat Biopharm Res. 2010; 2:42–50.

11. Tamura R, Huang X. An examination of the efficiency of the sequential parallel design in psychiatric clinical trials. Clin Trials. 2007; 4:309–317. [PubMed: 17848492]

12. Brown WA, Johnson MF, Chen MG. Clinical features of depressed patients who do and do not improve with placebo. Psychiatry Res. 1992; 41:203–14. [PubMed: 1594707]

13. Mallinckrodt CH, Zhang L, Prucka WR, et al. Signal detection and placebo response in schizophrenia: parallels with depression. Psychopharmacol Bull. 2010; 43:53–72. [PubMed: 20581800]

14. Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. Eur Neuropsychopharmacol. 2009; 19:34–40. [PubMed: 18823760]

15. Buitelaar JK, Sobanski E, Stieglitz RD, et al. Predictors of placebo response in adults with attention-deficit/hyperactivity disorder: data from 2 randomized trials of osmotic-release oral system methylphenidate. J Clin Psychiatry. 2012; 73:1097–102. [PubMed: 22780962]

16. Weimer K, Colloca L, Enck P. Placebo effects in psychiatry: mediators and moderators. Lancet Psychiatry. 2015; 2:246–57. [PubMed: 25815249]

17. DeRubeis RJ, Cohen ZD, Forand NR, et al. The personalized advantage index: Translating research on prediction into individualized treatment recommendations. A demonstration. PloS one. 2014; 9:e83875. [PubMed: 24416178]

18. Wallace ML, Frank E, Kraemer HC. A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. JAMA Psychiatr. 2013; 70:1241–1247.

19. Wallace ML, McMakin DL, Tan PZ, et al. The role of day-to-day emotions, sleep, and social interactions in pediatric anxiety treatment. Behaviour Research Therapy. 2017; 90:87–95. [PubMed: 28013054]

20. Smagula SF, Wallace ML, Anderson SJ, et al. Combining moderators to identify clinical profiles of patients who will, and will not, benefit from aripiprazole augmentation for treatment resistant late-life major depressive disorder. J Psychiatr Res. 2016; 81:112–118. [PubMed: 27438687]

21. Roose SP, Sackeim HA, Krishnan KRR, et al. Antidepressant pharmacotherapy in the treatment of depression in the very old: A randomized, placebo-controlled trial. Am J Psychiatry. 2004; 161:2050–2059. [PubMed: 15514406]

22. Hamilton MA. rating scale for depression. J Neurol Neurosurg Psychiatry. 1960; 23:56–62. [PubMed: 14399272]

23. Hamilton MA. The assessment of anxiety states by rating. Brit J Med Psychol. 1959; 32:50–55. [PubMed: 13638508]

24. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975; 12:189–198. [PubMed: 1202204]

25. MacLeod CM. Half a century of research on the Stroop effect: an integrative review. Psychol Bull. 1991; 109:163–203. [PubMed: 2034749]

26. Wechsler, D. Wechsler adult intelligence scale 3rd Revision (WAIS-III). San Antonio: Psychological Corporation; 1997.

27. Buschke H, Fuld P. Evaluating storage, retention, and retrieval in disordered memory and learning. Neurolog. 1974; 24:1019–1025.

28. Lawton MP. The functional assessment of elderly people. J Am Geriatr Soc. 1971; 19:465–481. [PubMed: 5094650]

29. Zilcha-Mano S, Roose SP, Brown PJ, Rutherford BR. Early symptom trajectories as predictors of treatment outcome for citalopram versus placebo. Am J Geriatr Psychiatry. 2017; 25:654–661.

30. Zilcha-Mano S, Keefe JR, Chui H, et al. Reducing dropout in treatment for depression: translating dropout predictors in to individualized treatment recommendations. J Clin Psychiatry. 2016; 77:e1584–e1590. [PubMed: 28086005]

31. Krull JL, MacKinnon DP. Multilevel modeling of individual and group level mediated effects. Multivariate Behav Res. 2001; 36:249–277. [PubMed: 26822111]

32. Laurenceau, JP., Bolger, N. Analyzing diary and intensive longitudinal data from dyads. In: Conner, TS., editor. Handbook of research methods for studying daily life. New York, NY: Guilford Press; 2012. p. 407-422.

33. SAS. SA Guide SU Version 9.1. SAS Institute Inc; Cary, NC: 2003.

34. Garge NR, Bobashev G, Eggleston B. Random forest methodology for model-based recursive partitioning: the mobForest package for R. BMC Bioinformatics. 2013; 14:125–132. [PubMed: 23577585]

35. Strobl C, Boulesteix AL, Kneib T, et al. Conditional variable importance for random forests. BMC Bioinformatics. 2008; 9:307–317. [PubMed: 18620558]

36. Zeileis A, Hothorn T, Hornik K. Model-based recursive partitioning. J Comp Graph Stat. 2008; 17:492–514.

37. Zeileis A. A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. Econom Rev. 2005; 24:445–466.

38. Petkova ET, Tarpey Su Z, Ogden RT. Generated effect modifiers (GEMs) in randomized clinical trials. Biostatistics. 2016:105–118. [PubMed: 27465235]

39. Kraemer HC. Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. Stat Med. 2013; 32:1964–1973. [PubMed: 23303653]

40. Kemp AS, Schooler NR, Kalali AH, et al. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? Schizophr Bull. 2010; 36:504–509. [PubMed: 18723840]

41. Agid O, Siu CO, Potkin SG, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970–2010. Am J Psychiatry. 2013; 170:1335–44. [PubMed: 23896810]

42. Cohen D, Consoli A, Bodeau N, et al. Predictors of placebo response in randomized controlled trials of psychotropic drugs for children and adolescents with internalizing disorders. J Child Adolesc Psychopharmacol. 2010; 20:39–47. [PubMed: 20166795]

43. Nelson JC, Delucchi K, Schneider LS. Efficacy of second generation antidepressants in late-life depression: a meta-analysis of the evidence. Am J Geriatr Psychiatry. 2008; 16:558–567. [PubMed: 18591576]

44. Rutherford BR, Wall MM, Brown PJ, et al. Patient expectancy as a mediator of placebo effects in antidepressant clinical trials. Am J Psychiatry. 2016; 174:135–142. [PubMed: 27609242]

45. Rutherford BR, Roose SP. A Model of Placebo Effects in Antidepressant Clinical Trials. Am J Psychiatry. 2013; 170:723–733. [PubMed: 23318413]

46. Rutherford BR, Pott E, Tandler JM, Wall MM, Roose SP, Lieberman JA. Placebo Response in Antipsychotic Clinical Trials: A Meta-Analysis. JAMA Psychiatry. 2014; 71:1409–1421. [PubMed: 25321611]

47. Rutherford BR, Sneed JR, Tandler J, Peterson BS, Roose SP. Deconstructing Pediatric Depression Trials: An Analysis of the Effects of Expectancy and Therapeutic Contact. J Am Acad Child Adolesc Psychiatry. 2011; 50:782–795. [PubMed: 21784298]

48. Zilcha-Mano S, Roose SP, Barber JP, et al. Therapeutic alliance in antidepressant treatment: cause or effect of symptomatic levels? Psychother Psychosom. 2015; 84:177–182. [PubMed: 25832111]

49. Cohen ZD, DeRubeis RJ. Treatment Selection in Depression. Annu Rev Clin Psychol. 2018:14.

50. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. Am J Psychiatry. 2013; 170:723–733. [PubMed: 23318413]

Highlights

- Little is known about placebo responders and their characteristics

- The present work uses a machine learning approach to search for a set of moderators that interact to identify in advance placebo responders, contributing to the field of placebo moderators

- The findings suggest that, compared to medication, placebo had the greatest response among those who had more years of education

- The greatest signal detection in favor of medication was in patients with fewer years of education, who suffered from a long duration of depression since their first episode
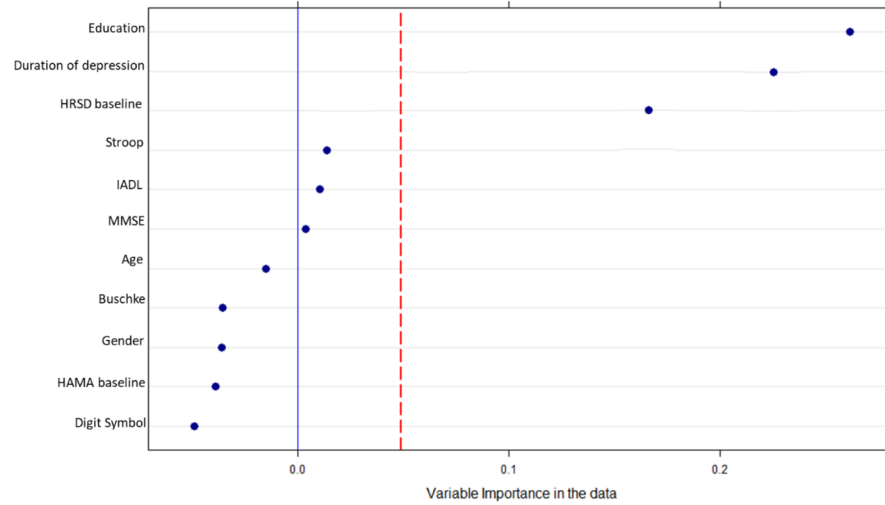
Figure 1.
Variable-importance plot for the model-based recursive partitioning trees. The horizontal axis in Figure 1 represents the average increase in classification accuracy gained by using the specific variable in the "real" data compared to use of the specific variable in permuted (i.e., "mixed up" or fake) data. Positive values indicate that a variable not only predicts patient-specific slope of HRSD outside of a given sample, but that it performs better than random noise. The red line represents the random noise of all potential moderator variables, and is constructed using the absolute value of the worst predictor. Variables to the right of the red line are selected for later modeling.
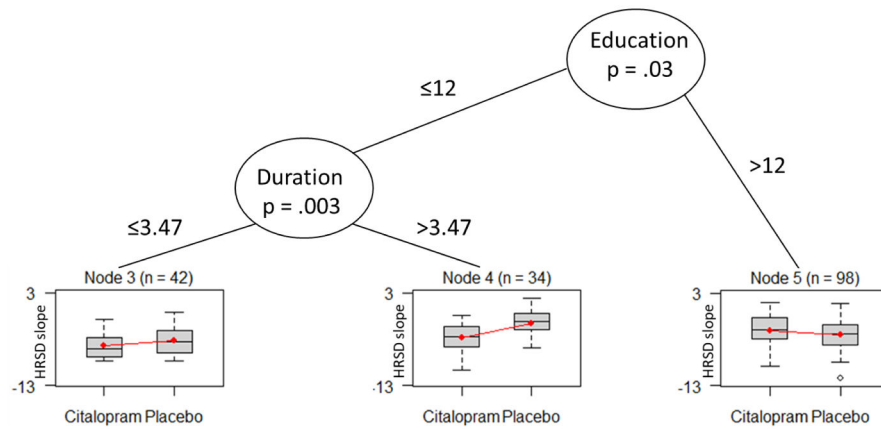
Figure 2.
Moderators of the effect of medication vs. placebo on outcome, as gleaned from decision-tree learning using model-based partitioning for condition assignment. Y = HRSD slope of change from pre-treatment to post-treatment. Duration = duration of depression (number of years since the patient's first episode). The decision tree analysis revealed a significant first split in patients' level of education (M_f = 12.59, DF=1, p = .03), and a second split in the duration of depression (M_f = 15.84, DF=1, p = .003).