A Novel Strategy to Identify Placebo Responders: Prediction Index of Clinical and Biological Markers in the EMBARC Trial


Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA; Duke-National University of Singapore, Singapore, Singapore; Duke Medical School, Durham, NC, USA; Texas Tech University Health Sciences Center, Permian Basin, TX, USA; Department of Statistical Science, Southern Methodist University, Dallas, TX, USA; Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; Columbia University, New York, NY, USA; Center for Depression, Anxiety and Stress Research, Mclean Hospital, Belmont, MA, USA; Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; New York State Psychiatric Institute and Department of Psychiatry, College of Physicians and Surgeons, New York, NY, USA; Stony Brook University School of Medicine, Stony Brook, NY, USA; Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

Keywords
Placebo responder · Prediction index · EMBARC trial

Abstract

Background: One in three clinical trial patients with major depressive disorder report symptomatic improvement with placebo. Strategies to mitigate the effect of placebo responses have focused on modifying study design with variable success. Identifying and excluding or controlling for individuals with a high likelihood of responding to placebo may improve clinical trial efficiency and avoid unnecessary medication trials. Methods: Participants included those assigned to the placebo arm (n = 141) of the Establishing Moderators and Biosignatures for Antidepressant Response in Clinical Care (EMBARC) trial. The elastic net was used to evaluate 283 baseline clinical, behavioral, imaging, and electrophysiological variables to identify the most robust yet parsimonious features that predicted depression severity at the end of the double-blind 8-week trial. Variables retained in at least 50% of the 100 imputed data sets were used in a Bayesian multiple linear regression model to simultaneously predict the probabilities of response and remission. Results: Lower baseline depression severity, younger age, absence of melancholic features or history of physical abuse, less anxious arousal, less anhedonia, less neuroticism, and higher average theta current density in the rostral anterior cingulate predicted a higher likelihood of improvement with placebo. The Bayesian model predicted remission and response with an actionable degree of accuracy (both AUC > 0.73). An interactive calculator was developed predicting the likelihood of placebo response at the individual level. Conclusion: Easy-
to-measure clinical, behavioral, and electrophysiological assessments can be used to identify placebo responders with a high degree of accuracy. Development of this calculator based on these findings can be used to identify potential placebo responders.

© 2018 S. Karger AG, Basel

Introduction

Major depressive disorder (MDD) affects 1 in 6 adults during their lifetime and is estimated to cost the USA over USD 200 billion per year [1, 2]. Yet, only one third of MDD patients achieve remission under any given antidepressant treatment [3], with over one third of the patients not responding to 2 or more antidepressants [4, 5]. Despite substantial financial investments [6, 7], efforts to improve MDD treatment outcomes by developing nonmonoaminergic antidepressants have failed. High placebo response rates have been a common factor in the failure of several novel antidepressant medications in phase 2 and 3 clinical trials [8, 9]. With limited/variable success, attempts to mitigate the increase in placebo response rate over the last three decades [10] have focused mainly on study design-related issues such as: (1) increasing the sample size to account for smaller drug-placebo difference; (2) incorporating a placebo lead-in period; (3) controlling measurement factors by using central raters or standardized interviews; or (4) implementing innovative study designs such as the sequential parallel comparison design [11–14]. A patient-centered approach, which identifies the individual characteristics that define placebo responders, may offer an alternative way to reduce the placebo response rate in clinical trials. These characteristics may be useful, additionally, in clinical practice, where efforts to maximize the placebo response can improve treatment outcomes [15].

The efforts to identify predictors of placebo response have been limited by the focus on subjective disease severity assessment and demographic features. This has frequently resulted in conflicting findings that often did not consider a wide variety of potential predictors from behavioral and biological domains. Among individual factors, low pretreatment symptom severity has been associated with higher likelihood of placebo response [16, 17]. However, other factors such as gender and age, while significant in venlafaxine versus placebo studies [18], did not replicate in a meta-analysis by Holmes et al. [19]. The neurobiological basis of the placebo response is characterized by an increase in the metabolic activity of the frontal and striatal cortical regions [20] and increased endogenous opioid release in the subgenual anterior cingulate cortex, nucleus accumbens, midline thalamus, and amygdala [21, 22]. The placebo response has also been linked to increased baseline resting state functional connectivity of the rostral anterior cingulate cortex (rACC) within the salience network [23] and to increased pretreatment rACC activity in 2 EEG studies [24, 25]. Previous studies of neuroimaging biomarkers of placebo response have been limited by small sample sizes and a lack of comparison with other clinical and biobehavioral markers [21, 23]. Despite extensive research to characterize placebo responders, a set of clinical and objective predictors and tools to filter out this subgroup from clinical trials has yet to be agreed upon and implemented. There is also a potential real-world clinical implication. A portion of treatment-seeking depressed individuals might not need a long-term antidepressant prescription if they are placebo responders. Hence, by identifying such placebo respondents in advance, briefer low cost, low side effect interventions may be recommended for these particular patients.

The goal of this report is to identify a parsimonious set of markers among assessments across units of analyses (clinical, demographic, neuroimaging, electrophysiological, behavioral, and cognitive assessments) that most strongly predict the likelihood of placebo response and can be implemented in research settings and clinical practice. Traditional approaches to handling a statistical problem like this – such as stepwise regression or factor analysis – are not well equipped to handle a rich database with (1) a large number of predictors relative to the number of subjects or (2) missing data. To attain this goal of identifying the parsimonious set of markers, we utilized data from the Establishing Moderators and Biosignatures for Antidepressant Response in Clinical Care (EMBARC) trial to systematically explore 283 variables using an advanced variable selection method to identify the variables that most strongly predict the likelihood of improvement with placebo. Next, we used a novel Bayesian method to simultaneously predict the degree of symptom change, as well as the probability of remission and response using these variables. Such a method allows flexibility in the definition of “placebo response” and can be implemented using a simple web-based tool.

Methods

Participants

Participants for this report were recruited as part of the EMBARC trial, which is a 2-stage, multisite, double-blind randomized controlled study designed to evaluate possible modera-

Trivedi et al.
Prediction Index in the EMBARC Trial

Statistical Analysis Plan

To accomplish the goals of the analysis, three statistical problems had to be addressed: performing variable selection, accounting for missing data, and estimating the magnitude and direction of the regression coefficients for the variables with the most predictive power after dealing with the first two listed problems. Online supplementary Figure 1 (for all online suppl. material, see www.karger.com/doi/10.1159/000491093) displays the progression of the analysis. A high-level summary of the analysis plan is given below; for a more detailed description (with more technical aspects addressed), see the online supplementary file.

Variable Selection Method

With the goal to identify a parsimonious set of variables that predict response to placebo, the elastic net [33] — a penalized regression technique that is becoming increasingly popular in the statistical literature [34] with respect to variable selection — was used. This procedure introduces bias to the regression estimates in the form of shrinkage — that is, it pulls the estimates towards 0 — while simultaneously reducing the variance of the estimates in order to increase the overall predictive power. The elastic net was selected over other methods such as random forests, support vector machines, gradient boosting machines, and multivariate adaptive regression splines, to ensure the analysis would result in an easily interpretable linear model. The aforementioned alternatives — while popular as tools for prediction — often require large amounts of data [35] and can be difficult to interpret due to the inclusion of complex, nonlinear interactions [36]. The elastic net was implemented via the glmnet package [37] in R 3.3.3 [38].

Approach for Missing Data

Due to missing values of predictor baseline variables, multivariate imputation by chained equations was used [39, 40] under the assumption that the data were missing at random [41] to generate a total of 100 imputed data sets. No imputation was used for the outcome variable, (as described above in the “Outcome” subsection) the last observed HAMD17 score. As there is no universally accepted approach to apply variable selection methods in the
Table 1. Baseline sociodemographic and clinical features of EMBARC trial participants randomized to placebo who received at least 1 dose (n = 141)

<table>
<thead>
<tr>
<th>Categorical variables, n (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>52 (36.9)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>89 (63.1)</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>98 (69.5)</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>23 (16.3)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>20 (14.2)</td>
</tr>
<tr>
<td>Employment status</td>
<td>Employed</td>
<td>78 (55.3)</td>
</tr>
<tr>
<td></td>
<td>Unemployed</td>
<td>58 (41.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuous variables (mean ± SD)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>37.4±12.9</td>
<td></td>
</tr>
<tr>
<td>Age of onset, years</td>
<td>16.4±5.6</td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td>15.2±2.7</td>
<td></td>
</tr>
<tr>
<td>Number of MDE</td>
<td>4±2</td>
<td></td>
</tr>
<tr>
<td>Duration of current episode, months</td>
<td>41.5±25.8</td>
<td></td>
</tr>
<tr>
<td>QIDS-SR</td>
<td>17.8±2.7</td>
<td></td>
</tr>
<tr>
<td>HAMD$_{17}$</td>
<td>18.6±4.3</td>
<td></td>
</tr>
</tbody>
</table>

EMBARC, Establishing Moderators and Biosignatures for Antidepressant Response in Clinical Care; SD, standard deviation; MDE, major depressive episode; QIDS-SR, Quick Inventory of Depressive Symptomatology Self-Report version; HAMD$_{17}$, 17-item Hamilton Rating Scale for Depression. a Median and interquartile range reported due to outlying values.

Model for Prediction of Placebo Outcomes

To estimate the magnitude and direction of the regression coefficients for the variables selected by the elastic net, we employed a Bayesian linear regression model that kept the same outcome variable; this is beneficial for several reasons. First, we can mitigate some of the concerns that would arise were we to instead apply ordinary least squares regression after choosing variables via the elastic net [45, 46] by assigning prior distributions to the regression parameters that mimic the shrinkage that occurs when applying the elastic net to all variables. For more details, see the supplementary file or Makalic and Schmidt [47]. Next, we can simultaneously account for the multiply imputed data [48] by running separate analyses and then mixing posterior draws from each run [49, 50] to create a more comprehensive, singular posterior distribution. Finally, we can sample from the posterior distribution of the regression parameters to estimate the predicted HAMD$_{17}$ score at the end of stage 1 (i.e. the outcome variable) and construct a posterior distribution of these predicted scores. This distribution can then be used to estimate the probability of remission (by calculating the proportion of posterior distribution scores ≤7 in the context of multiply imputed data [42–44], the elastic net ran independently on each of the 100 data sets, and we noted the number of times each variable was selected. The variables retained in at least 50% of imputed data sets were used in a Bayesian linear regression model.

Results

Of the 141 participants randomized to the placebo arm of the EMBARC trial, the mean (standard deviation) HAMD$_{17}$ at baseline was 18.6 (4.3) and at study exit it was 12.0 (7.5). At study exit, 47/141 (33.3%) attained remission, and 55/141 (39.0%) attained response. Baseline clinical and sociodemographic features are presented in Table 1.

Prediction of Placebo Outcomes

After applying the elastic net to 100 imputed data sets, 8 out of 283 variables were retained in at least 50% of the runs; these variables, listed in Table 2, included baseline HAMD$_{17}$, age, anhedonia, and anxious arousal as measured by MASQ, neuroticism, the presence of melanchol-
ic features, history of physical abuse, and average theta current density in the rACC. These variables were then entered in a Bayesian linear regression model to predict final HAMD$_{17}$ scores at the end of stage 1 of the EMBARC trial. Table 2 gives the posterior median values for the chosen set of predictors. Higher posterior median values predicted higher HAMD$_{17}$ scores at the end of stage 1 of the EMBARC trial, thus signifying worse outcomes. For example, a 1-unit increase in baseline HAMD$_{17}$ was associated with a final HAMD$_{17}$ score that was (on average) 0.52 points higher, holding all other predictors in the model constant. Thus, while higher baseline depression severity, age, neuroticism, anxiety, and anhedonia severity, as well as the presence of melancholic features and history of physical abuse, predicted worse outcomes with placebo, larger values of pretreatment theta current density localized to the rACC predicted better outcomes with placebo.

Taking advantage of the posterior distribution of the predicted HAMD$_{17}$ scores in the Bayesian framework, the probability of remission (HAMD$_{17} \leq 7$) and response (reduction of HAMD$_{17}$ by $\geq 50\%$) at the end of stage 1 of the EMBARC trial were estimated simultaneously and compared with the observed values at individual participant level; the receiver operating curve for both remission and response are shown in Figure 1. The AUC values (0.758 for remission and 0.728 for response) indicate a moderate fit, supporting the validity of using a Bayesian linear regression model to derive these values.

In univariate analyses to predict remission, baseline depression severity had the highest AUC and the greatest decrement in AUC to the model, as shown in Table 2. Two other variables with significant univariate AUC (i.e., their bootstrapped confidence interval did not include 0.50) were age and history of physical abuse. Notably, the highest univariate AUC (0.635 for baseline depression severity) was substantially lower than the AUC including all

### Table 2. Posterior median values of variables selected by elastic net that were included in the Bayesian linear regression model to predict treatment outcomes with placebo

<table>
<thead>
<tr>
<th>Variable</th>
<th>Posterior median</th>
<th>95% CI</th>
<th>Univariate AUC for remission</th>
<th>Decrement in AUC for remission upon removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline HAMD$_{17}$ score</td>
<td>0.52</td>
<td>0.186, 0.877</td>
<td>0.635 (0.541 to 0.729)</td>
<td>0.092 (–0.003 to 0.187)</td>
</tr>
<tr>
<td>Age</td>
<td>0.12</td>
<td>0.016, 0.222</td>
<td>0.611 (0.517 to 0.705)</td>
<td>0.058 (–0.032 to 0.148)</td>
</tr>
<tr>
<td>Melancholic depression indicator$^a$</td>
<td>2.06</td>
<td>–0.764, 4.959</td>
<td>0.388 (0.309 to 0.468)</td>
<td>0.003 (–0.077 to 0.083)</td>
</tr>
<tr>
<td>Anhedonia$^b$</td>
<td>0.13</td>
<td>–0.123, 0.367</td>
<td>0.537 (0.454 to 0.619)</td>
<td>0.017 (–0.067 to 0.100)</td>
</tr>
<tr>
<td>Anxious arousal$^c$</td>
<td>0.22</td>
<td>–0.085, 0.537</td>
<td>0.545 (0.444 to 0.646)</td>
<td>0.014 (–0.072 to 0.100)</td>
</tr>
<tr>
<td>Neuroticism$^d$</td>
<td>0.12</td>
<td>–0.092, 0.336</td>
<td>0.505 (0.411 to 0.600)</td>
<td>0.004 (–0.077 to 0.084)</td>
</tr>
<tr>
<td>Physical abuse$^e$</td>
<td>0.18</td>
<td>–0.185, 0.543</td>
<td>0.625 (0.533 to 0.718)</td>
<td>–0.010 (–0.090 to 0.070)</td>
</tr>
<tr>
<td>rACC theta current density</td>
<td>–5.93</td>
<td>–13.417, 1.166</td>
<td>0.510 (0.412 to 0.608)</td>
<td>0.019 (–0.067 to 0.105)</td>
</tr>
</tbody>
</table>

HAMD$_{17}$, 17-item Hamilton Rating Scale for Depression; rACC, rostral anterior cingulate. $^a$ Based on specifier questions on the SCID. $^b$ Anhedonic depression scale from the Mood and Anxiety Symptom Questionnaire. $^c$ Anxiety specific scale (anxious arousal) from the Mood and Anxiety Symptom Questionnaire. $^d$ Based on 12 neuroticism items from the NEO Five-Factor Inventory. $^e$ Scale from the Childhood Trauma Questionnaire.

![Fig. 1. Receiver operating characteristic curves for Bayesian model with an a priori threshold of 50% variable retention.](color-version-available-online)
8 variables (0.758 for remission), thus highlighting the predictive utility of our multivariate model (Tables 2, 3).

Further, the probability threshold for classifying subjects as remitters or responders can be optimized based on the desire of individual clinician or researcher. The four quantities often of interest in classification problems are: sensitivity (in this case, the percentage of correctly identified remitters from the entire population of remitters), specificity (the percentage of correctly identified nonremitters from the entire population of nonremitters), positive predictive value (the percentage of predicted remitters who are truly remitters), and negative predictive value (the percentage of predicted nonremitters who are truly nonremitters). Assuming a high certainty (minimum of 70%) to accurately identify participants likely to be a remitter – that is, when you predict someone to be a remitter you will be correct 70% of the time – the optimal probability threshold as identified by the OptimalCutpoints package in R will make correct predictions 73.3% of the time and with 23.4% sensitivity [54]. Such a threshold might be desirable for a clinician, who would hope to be quite confident in the likelihood of placebo response before making a treatment decision for his/her patients. Clinical researchers, on the other hand, might be more interested in capturing a larger percentage of placebo responders (i.e., increased sensitivity) – at the cost of more false positives – in order to screen them out of clinical trials to get a purer estimate of a treatment effect.

Sensitivity Analyses
In contrast to the 8 variables retained with the a priori threshold of 50%, only 3 variables (baseline HAMD17, age, and presence of melancholic features) were retained in all of the 100 elastic net runs, and 15 variables were retained in at least 30% of the runs. The additional 7 variables retained in 30% of the runs that indicated improvement in depression severity included lower baseline self-reported depression severity, higher openness to experience, shorter duration of major depressive episode, higher resting-state functional connectivity between right ventral striatum and dorsal anterior cingulate, higher resting-state functional connectivity between left and right ventral striatum as well as left and right insula, and higher flanker accuracy effect (calculated as: \(\text{accuracy}_{\text{congruent trials}} - \text{accuracy}_{\text{incongruent trials}}\)).

The bootstrapped Bayesian linear regression models with ridge priors were conducted and the estimated root mean squared error (RMSE) for the HAMD17 at the end of stage 1 of the EMBARC trial, as well as AUC values for the probability of remission and response, were compared between the a priori threshold of 50% retention as well as 100% retention and 30% retention (Table 3). The comparison of AUC of the a priori threshold with 100% retention to 50% retention \(p = 0.076\) and 50% retention to 30% retention \(p = 0.055\) showed modest evidence of statistical significance while that of 100 and 30% was stronger \(p = 0.012\). With respect to RMSE, the 50% retention group had an 8% improvement in accuracy over the 100% retention group, with the 30% retention group showing a 13% improvement in accuracy. Considering all this information (including the costs of measuring the predictors in each group), the a priori defined threshold of retention in 50% of the models provided the best combination of practical and statistical utility.

Visualization of Study Findings
The Shiny package in R [53] was used to adapt the Bayesian linear regression model using the 8 variables identified by the a priori threshold of 50% retention in elastic net runs. This interactive calculator allows the user to adjust the predictor values (based on the range observed in the EMBARC trial) and obtain a distribution of posterior predicted HAMD17 scores at the end of the 8-week placebo administration. The posterior distribution of HAMD17 simultaneously allows estimation of the probability of remission and response. A screenshot of this calculator is presented in Figure 2 and will be made available to the broad scientific community using a web-based interface.

Discussion
In this large sample of depressed outpatients, a systematic exploration of a broad range of clinical and biological markers identified baseline depression severity, age, neuroticism, anhedonia, anxious arousal, presence of melan-
cholic features, history of physical abuse, and theta current density in the rACC as predictors of placebo response in a double-blind randomized clinical trial. These markers can be implemented within a Bayesian framework to simultaneously predict the likelihood of response as well as remission with an easy-to-use calculator. Clinicians or clinical researchers could use this calculator to make more informed decisions about whether to prescribe a treatment or enroll subjects in clinical trials with a placebo arm.

The current findings are consistent with previous reports that have found that higher baseline depression severity is associated with lower likelihood of improvement in the placebo arm [10, 17]. The most powerful single variable predicting poorer outcome in response to placebo treatment was greater depression severity. This is a very important finding, as enrichment in terms of severity of illness can be easily implemented in the context of a clinical trial. The failure to reduce placebo response by implementing a greater illness severity threshold for inclusion in studies has been attributed to the “grade inflation” when the severity of illness is determined only by clinicians at the site, where the bias towards enrollment can be significant, thereby driving up severity measures [12]. Independent verifications of subject severity at entry by themselves can reduce the placebo response by allowing a greater depression severity enrichment [55].

The finding that higher theta current density in the rACC predicts greater improvement with placebo differs from previous work by Korb et al. [56] which did not find any difference in rACC theta density between placebo responders and nonresponders. This difference may be related to the smaller sample size of the study of Korb et al. Moreover, in a later study, the same group reported that increased rACC theta current density predicted a greater placebo response. Notably, in the International Study to Predict Optimized Treatment in Depression (iSPOT-D)
with other variable selection methods such as random forest. Lastly, due to the unique combination of variables collected for the EMBARC trial, we cannot validate the model on an external data set. The cross-validation and bootstrap procedures employed add strength to the analysis, but there is no substitute for external validation.

To conclude, a set of 8 clinical and biological markers can predict treatment outcomes with placebo with a fair degree of accuracy. It is also possible to further increase the accuracy of prediction if all 15 variables are included. By integrating these markers in an easy-to-use interactive calculator, the findings of this report can be implemented in research and clinical care.

Acknowledgments

Valient Pharmaceuticals donated the Wellbutrin XL used in the study. This work was supported by the EMBARC National Coordinating Center at UT Southwestern Medical Center, Madhukar H. Trivedi, MD, Coordinating PI, and the Data Center at Columbia and Stony Brook Universities. We acknowledge Jennifer Furnman, PhD, and Jeremy Kee, MA, for their administrative support.

Disclosure Statement

Dr. Pizzagalli: funding from NIMH and the Dana Foundation; consulting fees from BlackThorn Therapeutics, Boehringer Ingelheim, Pfizer Inc., and Posit Science. Dr. Weissman: funding from NIMH, the National Alliance for Research on Schizophrenia and Depression (NARSAD), the Sackler Foundation, and the Templeton Foundation; royalties from the American Psychiatric Association Press, the American Psychiatric Association Press, and Multi-Health Systems. Dr. Trivedi has received consulting fees from or has served on the advisory boards of Alkeremes Inc., Akili Interactive, Navitor, and Otsuka America Pharmaceutical Inc.; he holds author agreements with Janssen Asia Pacific and Oxford University Press; honoraria from the American Psychiatric Association and grants from the Cancer Prevention and Research Institute of Texas (CPRIT), National Institute of Mental Health (NIMH), National Institute of Drug Abuse (NIDA), National Center for Advancing Translational Sciences (NCATS), Johnson & Johnson, and PCORI. Dr. Trombello currently owns stock in Merck and Gilead Sciences and within the past 36 months previously owned stock in Johnson & Johnson. Dr. Fava has received research support from Abbott Laboratories, Alkermes Inc., American Cyanamid, Aspect Medical Systems, AstraZeneca, Avanir Pharmaceuticals, BioResearch, BrainCells Inc., Bristol-Myers Squibb, CereNeRx BioPharma, Cephalon, Cerecor, Clintara LLC, Covance, Coviden, Eli Lilly and Company, EnVivo Pharmaceuticals Inc., Euthymics Bioscience Inc., Forest Pharmaceuticals Inc., Geneden Biotech Inc., GlaxoSmithKline, Harvard Clinical Research Institute, Hoffman-LaRoche, Icon Clinical Research, i3 Innovus/Ingenix, Janssen R&D LLC, Jed Foundation, Johnson & Johnson Pharmaceutical Research & Development, Lichtwer Pharma GmbH, Lorex Pharmaceuticals, Lundbeck Inc., MedAvante, Methylation Sciences Inc.,
Prediction Index in the EMBARC Trial

Prediction Index in the EMBARC Trial


Funding Sources

The EMBARC study was supported by the National Institute of Mental Health of the National Institutes of Health under award No. U01MH092221 (Trivedi, M.H.) and U01MH092250 (McGrath, P.J., Parsey, R.V., Weissman, M.M.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
References


