Contemporary Clinical Trials Communications 6 (2017) 22–30

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

Statistical analysis plan for stage 1 EMBARC (Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care) study

Eva Petkova [a, b, *], R. Todd Ogden [c], Thaddeus Tarpey [a, d], Adam Ciarleglio [a, c], Bei Jiang [e], Zhe Su [a], Thomas Carmody [f], Philip Adams [g, h], Helena C. Kraemer [i], Bruce D. Grannemann [f], Maria A. Oquendo [g, h], Ramin Parsey [j], Myrna Weissman [g, h], Patrick J. McGrath [g, h], Maurizio Fava [k], Madhukar H. Trivedi [f]

[a] New York University, New York, NY, USA
[b] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA
[c] Columbia University, New York, NY, USA
[d] Wright State University, Dayton, OH, USA
[e] University of Alberta, Edmonton, Alberta, Canada
[f] University of Texas, Southwestern Medical Center, Dallas, TX, USA
[g] New York State Psychiatric Institute, New York, NY, USA
[h] Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York, NY, USA
[i] Stanford University, Stanford, CA, USA
[j] Stony Brook University, Stony Brook, NY, USA
[k] Massachusetts General Hospital, Boston, MA, USA

ARTICLE INFO

Article history:
Received 25 May 2016
Received in revised form
8 February 2017
Accepted 13 February 2017
Available online 24 February 2017

Keywords:
Combining biomarkers
Differential treatment response index
Moderator
Optimizing treatment decisions
Precision medicine

ABSTRACT

Antidepressant medications are commonly used to treat depression, but only about 30% of patients reach remission with any single first-step antidepressant. If the first-step treatment fails, response and remission rates at subsequent steps are even more limited. The literature on biomarkers for treatment response is largely based on secondary analyses of studies designed to answer primary questions of efficacy, rather than on a planned systematic evaluation of biomarkers for treatment decision. The lack of evidence-based knowledge to guide treatment decisions for patients with depression has lead to the recognition that specially designed studies with the primary objective being to discover biosignatures for optimizing treatment decisions are necessary. Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care (EMBARC) is one such discovery study. Stage 1 of EMBARC is a randomized placebo controlled clinical trial of 8 week duration. A wide array of patient characteristics is collected at baseline, including assessments of brain structure, function and connectivity along with electrophysiological, biological, behavioral and clinical features. This paper reports on the data analytic strategy for discovering biosignatures for treatment response based on Stage 1 of EMBARC.

1. Introduction

Major Depressive Disorder (MDD) is a highly prevalent chronic and recurrent disorder predicted to be the leading cause of disease burden in the year 2030. Despite the advent of effective pharmacological, psychotherapeutic and brain stimulation interventions, we still lack tools to predict treatment response and remission. For example, the Sequenced Treatment Alternative to Relieve Depression project (STAR*D) attempted to determine the best treatment for patients who did not remit with a standard Selective Serotonin Reuptake Inhibitor (SSRI). Disappointingly for purposes of prediction, patients were equally likely to respond to a second SSRI,

* Corresponding author. New York University , New York, NY, USA.
E-mail addresses: eva.petkova@nyumc.org (E. Petkova), to166@cumc.columbia.edu (R.T. Ogden), thaddeus.tarpey@wright.edu (T. Tarpey), adam.ciarleglio@nyumc.org (A. Ciarleglio), bei1@ualberta.ca (B. Jiang), zhe.su@nyumc.org (Z. Su), thomas.carmody@utsouthwestern.edu (T. Carmody), adamsp@nyspi.columbia.edu (P. Adams), hck@leland.stanford.edu (H.C. Kraemer), bruce.grannemann@utsouthwestern.edu (B.D. Grannemann), oquendo@nyspi.columbia.edu (M.A. Oquendo), ramin.parsey@stonybrookmedicine.edu (R. Parsey), weissman@nyspi.columbia.edu (M. Weissman), pjm5@cumc.columbia.edu (P.J. McGrath), mfava@mgh.harvard.edu (M. Fava), madhukar.trivedi@utsouthwestern.edu (M.H. Trivedi).

http://dx.doi.org/10.1016/j.conctc.2017.02.007

venlafaxine-XR or bupropion-SR, suggesting that the pharmacologic profile of prior failed treatments is insufficient to guide subsequent treatment decisions. This finding, based on the comparisons of groups of patients, raises the question of whether individual characteristics, biological or clinical, might more accurately predict the likelihood of remission with a given intervention. Prediction of outcome with commonly used interventions, namely pharmacotherapy with drugs having distinct mechanisms of action, appears a rational first step in this quest. Response to antidepressant medication in depressed patients is unpredictable, with a 30% remission rate after 12 weeks of treatment and 30–40% fail to have an adequate response even after several trials of medication or psychotherapy over a year [13,35,47]. The search for biomarkers predicting overall or specific medication response is still in its infancy [18] and, while many studies of potential biomarkers for treatment outcome have been published (e.g., Refs. [4,15,16,21,25,26,48,51]), systematic examination of the joint effects of several biomarkers together with clinical phenotypes has never been done and little practical progress has been made. The most promising biomarker strategy to date, individual pharmacogenetic profiling, has not uncovered any strongly predictive alleles, although there are now multiple single nucleotide polymorphisms (SNP) suggesting genetic variants of relatively small effect, see e.g., Refs. [4,15] among others.

While predicting treatment outcome remains an essential, though elusive research goal, the question of immediate practical importance is how to select the best treatment for each individual patient, a fundamental component of precision medicine. It has long been recognized that features that are important for predicting outcome might not be necessarily be useful for making treatment decisions (e.g. Refs. [39,50]). Interest in discovering optimal treatment decisions for individual patients is growing rapidly, both in clinical research and in statistical methodology. Optimal treatment decision for a patient was first formalized by Murphy [27] and Robins [33]. A treatment decision is a function d that maps the baseline covariates, say $\boldsymbol{X}$, to a treatment indicator {0,1}, such that a participant with covariates $\boldsymbol{X} = \boldsymbol{x}$ will receive treatment 1 if $d(\boldsymbol{x}) = 1$ and will receive treatment 0 if $d(\boldsymbol{x}) = 0$. The value of a treatment decision is the average outcome, if the decision were to be applied to the entire target population. The best treatment decision is the one that optimizes the value. Using the concept of "potential outcome" [34], let $Y^*(0)$ and $Y^*(1)$ denote the potential outcomes that would be observed if a participant was assigned treatment 0 or 1, respectively. Note that only one of the potential outcomes is observed in practice and the observed outcome under a decision $d(\boldsymbol{X})$ can be expressed as follows

$$Y(d) = Y^*(1)d(\boldsymbol{X}) + Y^*(0)[1 - d(\boldsymbol{X})].$$

That is, the observed outcome is the potential outcome under treatment 1, if the treatment decision $d(\boldsymbol{X})$ is to give treatment 1, and it is the potential outcome under treatment 0, if the treatment decision is to give treatment 0. Thus, the value of a decision d is the observed outcome averaged over the distribution of $\boldsymbol{X}$ and, from Qian and Murphy [31]; equals $E^d[Y]$, where the expectation E^d is taken with respect to the joint distribution of $(\boldsymbol{X}, A, Y)$ when d is used to assign treatments.

The NIMH funded multi-site clinical trial Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care (EMBARC) was designed to systematically explore promising clinical and biological markers of antidepressant treatment outcome that would lead to personalized treatment. This paper describes the statistical analysis plan for the EMBARC study. The EMBARC study is a collaborative investigation to discover biomarker moderators and mediators of response to treatment of MDD with antidepressant medication (for full methods description see Ref. [46]. The four study sites used identical recruitment and assessment procedures and have recruited 309 participants in total with MDD. Participants had recurrent, early onset MDD (prior to age 30 years). During the 8-week first stage, patients receive either sertraline or placebo under randomized double-blind conditions in 1:1 ratio. The randomization was stratified by site, depressive symptom severity, and depression chronicity. During the second 8-week stage, non-responders to sertraline receive bupropion, non-responders to placebo receive sertraline, and responders remain on their original treatment. The study is unique in that it systematically collects a comprehensive array of carefully selected clinical, behavioral, and biological biomarkers at baseline and at one week post treatment initialization. Clinical measures include anxious depression, early trauma, gender, melancholic and atypical depression, anger attacks, Axis II disorder, hypersomnia/fatigue, and chronicity of depression. Behavioral measures result from a battery of cognitive and behavioral tasks. Biological measures include cerebral cortical thickness via structural magnetic resonance imaging (MRI), task-based functional MRI (fMRI), resting state brain connectivity, diffusion tensor imaging (DTI, collected only at baseline), arterial spin labeling (ASL), electroencephalography (EEG) and cortical evoked potentials.

One goal of the study is to quantify the effect of a selected set of candidate biomarkers as moderators of the effect of treatment (SSRI versus placebo). A major challenge in precision medicine, however, is that most baseline measures typically have small moderating effects and individually contribute little to informed treatment decisions. Thus, a key goal is to investigate possible combinations of biomarkers and clinical characteristics to generate biosignatures for making a personalized medication treatment prescription. A biosignature index can be based on patient characteristics at baseline (e.g., moderators of treatment effect). Additionally, since the study also collects biological data one week after randomization, early indicators of whether a patient will respond to the treatment (i.e., potential mediators of treatment effect) can also be identified. These early indicators could be used to refine the prediction regarding response to treatment that is solely based on pre-treatment patient characteristics, by capturing early biomarker changes in response to treatment. This paper describes the statistical analysis plan only for Stage 1 of EMBARC. Finally, in addition to being the first to institute protocols for standardizing assessment, quality control, data collection, transfer and integration of a multimodal database for depression biomarkers, the EMBARC study also aims at establishing strategies for discovery of biosignaures within such a rich and complex source of information.

2. Methods

2.1. Background

Given the sheer complexity of the brain and the fact that, despite decades of research, the causality of depression is still largely unknown, the identification of biomarkers for treatment response is a formidable challenge. Neuroimaging technologies such as structural MRI, fMRI and DTI are widely used to indirectly estimate cortical and subcortical volumes, brain activation in response to different tasks and functional and physical connectivity in the brain. Current mental health research explores the hypotheses that depression is due to the loss of cortical tissue, or due to deficient brain activation in response to stimuli, or altered connectivity in the brain, i.e., reduction in the temporal lobe volume, or aberrant connectivity within the default mode network, respectively. Correspondingly, it is hypothesized that treatments for depression work by normalizing brain structure, function, and/or

connectivity.

While insightful, this direction of research has yielded little in the way of conclusive results about causal factors for depression or the mechanisms of action of various treatments. This is likely due to several well-known challenges. First, depression is highly heterogeneous, i.e., patients tend to have widely varying combinations of symptoms and can also have vastly different underlying biology, and thus the biomarkers for treatment response among one subgroup might not apply to another. Second, the complexity of MDD makes it unlikely to find meaningful biomarkers for treatment response by focusing only at individual factors and neglecting the interrelationships between them. Finally, the intricacy of the brain and the multidimensionality of the data collected by ever evolving technologies for measuring its structure, function and connectivity, make the discovery of the treatment implications available in the collected information a daunting analytic task.

To address those challenges, the EMBARC study systematically assessed study participants over two days using several measures and characteristics that have been suggested by diverse theories and hypotheses about causes and effects of depression as well as about possible mechanisms of action of drug treatment. Suggested by studies indicating differences between patients with MDD and healthy controls with respect to a range of structural brain measures and resting state functional connectivity, the EMBARC study also collected data on structural and functional brain attributes via structural EEG MRI, DTI and resting state fMRI. Based on theoretical considerations regarding the neural circuits modulated by serotonin and dopamine, such as emotion regulation circuitry and reward circuitry, subjects were assessed with fMRI during both a specific emotion recognition task and a reward task. Since we anticipated that some of the characteristics assessed at baseline may change with the onset of treatment, they were measured again at week 1, to allow investigation as to whether early changes might contribute to improved outcomes. The use of week 1 assessments for improving the predictions and treatment decisions rules is a secondary goal and is not addressed here. For justification and details regarding all assessments, as well as alternatives that were considered, see Ref. [46].

2.2. Sample size determination

For the reasons laid out above, the EMBARC study was designed as a discovery, rather than a hypotheses-driven investigation. The sample size of the study was determined by the need to develop and validate a limited set of summary indices (i.e., composite biosignatures) for treatment response, that, if warranted by the results, would be further studied in a hypotheses-driven confirmatory investigation. Although several approaches have been proposed for developing individualized treatment decision rules (i.e. a mapping from baseline predictors to one of the treatment options, see Section 6.3, e.g., Refs. [7,14,31,53]), there are no sample size formulae for identifying composite biosignatures and constructing treatment decision rules from high-dimensional data. The necessary sample size depends on the signal to noise ratio, the complexity of the proposed models, the size of the model space to be searched, as well as the specific analytic method used. Since a careful validation is necessary for any treatment decision based on selecting and combining biomarkers, we plan to use a training set to construct a summary index and a treatment decision rule based on it. A separate test data set will then be used for validation. A 2-to-1 split of the total sample size into a training (n = 200) and testing set (n = 100) will be used. This will be a random split, stratifying for site; treatment assignment; severity and chronicity (which were used as randomization strata); and year of study entry, to control for possible secular effects. The issue of sample size and power, when testing one treatment decision rule against another, in a randomized clinical trial is discussed in Section 8.

2.3. Analytic samples

The analyses would be conducted on two overlapping samples:

2.3.1. Adequate treatment exposure sample

This sample will include only Stage 1 participants who received 8 weeks of treatment. The sample would allow identification of moderators and mediators of treatment response among those who have had full exposure to the treatment. This sample may be more likely to reflect biological changes that are related to the exposure to antidepressant therapy and to allow a more meaningful exploration of potential moderators of treatment response. It is anticipated that biomarkers identified with this sample would be more related to physiologic response to the medication compared with biomarkers identified using participants without adequate antidepressant exposure.

2.3.2. Modified intention-to-treat sample

This sample will include all randomized Stage 1 participants who took at least one dose of study medication. Participants who were randomized, but dropped out prior to taking their first dose will be excluded, as will those who were randomized but subsequently revealed to be ineligible for the study. This definition of the study sample is in line with standards used in efficacy research, where from a public health perspective, the goal is to estimate the effect of assigning treatment.

3. Defining treatment outcomes

The primary outcome measures that we consider are based on the Hamilton Depression scale (HAMD17). Per the study protocol, assessments of HAMD17 are scheduled for the time of randomization, i.e., at baseline ($t = 0$), weekly for 4 weeks ($t = 1, \ldots, 4$) and bi-weekly for the last month of treatment, $t = 6$ and 8 weeks. Let $\boldsymbol{Y} = (Y_t, t \in \boldsymbol{t} = \{0, 1, 2, 3, 4, 6, 8\})$ denotes the vector of HAMD17 scale observations during the Stage 1 clinical trial. The time points of the observed outcomes for the ith participant will be denoted $\boldsymbol{t}_i \in \boldsymbol{t}$ (note that $\boldsymbol{t}_i$ must contain 0). To enhance clinical relevancy and interpretability, we shall consider several definitions of the outcome measure.

3.1. Course of depression symptoms

In order to obtain a scalar value to summarize an individual's outcome, the following mixed-effects model for the HAMD17 outcome, fit separately for each treatment group, will be used:

$$(1)\quad Y_{it_{ij}} = \alpha_0 + \alpha_1 t_{ij} + \boldsymbol{I}_i^{site}\boldsymbol{\alpha}_2 + t_{ij}\boldsymbol{I}_i^{site}\boldsymbol{\alpha}_3 + a_{0i} + a_{1i}t_{ij} + \varepsilon_{ij},$$

where Y is the ith participant's HAMD score at time t; $\boldsymbol{I}^{site}$ is a 4-dimensional vector of indicators for the ith participant's site (utilizing zero-sum constraint), and $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ are regression coefficient vectors for the vector of indicators; a_{0i} and a_{1i} are the random intercept and slopes, respectively, for the ith participant, and ε_{ij} are independent random errors with variance σ^2. An overall measure that summarizes the ith individual's course of depression symptoms is his/her random slope plus the fixed-effect slope coefficients, i.e., $(a_{1i} + \alpha_1)$.

Note, that if the true symptom trajectories are arbitrary smooth functions of time, rather than strictly linear, as postulated by (1), an appropriate scalar measure is the average rate of change (i.e., the average tangent slope) over the course of treatment. If the true

trajectory is quadratic, the average tangent slope is equal to the slope of the best-fitting line, i.e., the slopes from model (1). From extensive previous work with HAMD17 symptom trajectories during 6–8 weeks of treatment [28,41–43], we expect that the trajectories will be well approximated by quadratic polynomials of time. In Petkova et al. [29] we show that when missing data due to dropout is not too severe, as is in the EMBARC study, the slopes estimated with model (1) would be unbiased estimates for the average tangent slope. The reason for including the site in (1) is to eliminate any between-site differences with respect to course of symptoms over time.

This outcome will be available on all study participants with at least one post-randomization assessment of HAMD17. Smaller values of $(a_{1i} + \alpha_1)$ are desirable, as they indicate a faster rate of decline of depression symptoms severity.

3.2. Remission

A binary remission status will be available on all study participants. A participant is considered to have achieved remission if their last observed HAMD17 score is less than or equal to 7 (Tivedi et al. 2006) [47].

4. Identifying baseline patient characteristics for evaluation as moderators of treatment effect

The EMBARC study distinguishes two tiers of baseline patient characteristics that were prespecified a priori by the study team to be investigated as potential treatment effect modifiers based on published evidence supporting their relationship with treatment outcome. The First Tier and a Second Tier of variables are those that have been identified as having strong (multiple reports in the literature) and moderate (only incidental reporting) evidence for association with response to treatment. Additionally, a Third Tier of variables was identified that were not pre-specified due to lack of evidence to justify their inclusion in the First or Second Tier.

4.1. First tier prespecified variables

At the time of planning this study, a set of baseline demographic, clinical, behavioral and biological patient characteristics were identified as having evidence supporting their role as predictors of antidepressants' effect on depression. We emphasize that these variables have generally been evaluated as predictors of outcomes, not as potential moderators of the effect of several treatments. A list of 48 characteristics was generated from a review of the antidepressant treatment response literature by the study psychiatrists. These variables include: presence of melancholic depression, reaction time in the Choice-Reaction-Time task, rostral anterior cingulate cortex (ACC) theta current source density derived from EEG and thickness of the precentral cortex. The full list of First Tier variables is given in the Appendix.

Let $\boldsymbol{X} = (X_1, \ldots, X_{p_x})$ denote the set of these baseline variables. Each of these variables will be evaluated in turn as a potential effect modifier based on the following model:

$$(2)\ g(E(Y|A, X_k) = \beta_0 + \beta_1 X_k + \beta_2 A + \beta_3 A X_k,$$

where Y is one of the outcomes defined in Section 3; A is an indicator for the treatment to which a participant was randomized (A = 1 for sertraline and A = 0 for placebo); g is the logit function when the outcome is a binary remission status, and g is the identity when the outcome is the random effect slope from (1). The variables X_k will be ranked based on the magnitude of their effect size as moderators, as per Ref. [17] rather than by the p-value for significance of the interaction term β_3. This eliminates the effect of the number of participants used in the analyses, as we expect that some of the baseline characteristics might be missing for more participants than other baseline measures. Additionally, with this approach, we emphasize the importance of the magnitude of the effects, rather than their statistical significance, which is in line with the discovery nature of the study.

4.2. Second Tier prespecified variables

The variables in this set are patients' biological characteristics that were identified by EMBARC investigators as having a potential for being important in making treatment decisions, although less evidence supporting their relationship with treatment outcome was available at the time of the EMBARC study planning, compared to First Tier variables. The variables (total 243) are denoted by $\boldsymbol{W} = (W_1, \ldots W_{p_w})$ and will be analyzed using the same approaches used for the First Tier variables, $\boldsymbol{X}$. These Second Tier variables are primarily biological brain measures of different modalities (e.g., EEG, structure, function, connectivity).

4.3. Third Tier variables

The Third Tier consists of variables that were not pre-specified, but can be computed from the collected data. For example, the Third Tier will include biological brain measures that have been identified and reported in the literature after the EMBARC study was initiated. These variables are denoted by $\boldsymbol{U} = (U_1, \ldots, U_{p_u})$ and will be analyzed using the same approaches used for variables in the other two tiers.

5. Composite indices for personalized treatment decisions

A major goal of the EMBARC study is to develop new constructs, not previously established that could be used to decide which treatment should be given to an individual depressed patient. These are called “moderators of treatment effect”, “effect modifiers” and “tailoring” variables in statistical terminology, or also “prescriptive” measures or variables in medical parlance. A Differential Treatment Response Index (DTRI) is conceived of as a combination of patient biological, behavioral and clinical characteristics, which would be used to decide which treatment would be more beneficial to a particular patient. The idea for such index is motivated by the Framingham Risk Score (see e.g., Ref. [2]), however, rather than measuring individual subject's “risk” (i.e., probability) for, say, response to a given treatment, the DTRI is required to measure the relative benefit of one treatment compared to another. In other words, the index should indicate the ranges where “treatment 1 is better than treatment 0”, to where “no difference in response to treatments 1 and 0”, to where “treatment 0 is better than treatment 1”. Such an index, constructed as a linear combination of baseline characteristics, is proposed in Cloitre et al. [10] in the context of selecting treatment for subjects with post-traumatic stress disorder. In the current setting, a DTRI can be used to determine if a patient would benefit more from the active treatment (sertraline) compared to a placebo treatment. Given the numerous adequately conducted randomized placebo controlled antidepressant clinical trials that failed to show efficacy against placebo, the question of whether or not to prescribe a medication to a specific patient with MDD is of utmost importance. One major benefit of this comprehensive approach is to ensure that all possible variables are considered together. This also allows us to account for the inter-relationships among all variables.

To develop DTRIs for making treatment decisions, the analytic sample will be split into a training set and a test set as described in Section 2.2. This will be done for both the modified intent-to- treat and the adequate treatment exposure samples, see Section 2.3. The DTRIs developed on the training data will then be applied and evaluated on the test data set. DTRIs developed through the application of different analytic approaches and using data of different modalities will be selected based on their performance in the test set. This approach (development on the training set followed by validation on the test set) will provide evidence of stability of the index within the same study and reduce the likelihood of a spurious finding.

5.1. Analytic methods for making optimal treatment decisions

It has long been recognized that baseline features that are important for predicting outcome might not necessarily be useful for making treatment decisions (e.g. Refs. [39,49,50]). Much recent research has focused on identification of baseline covariates that are specific to the treatment effect (i.e., variables that exhibit interactions with the treatment indicator in predicting treatment outcome), rather than being important in the baseline model (i.e., prognostic of outcome under either treatment, or prognostic of outcome under the standard treatment), see e.g., Refs. [6,14, 22,24,31]. Thus, we differentiate between "prescriptive" variables (that can inform clinicians in prescribing treatment to a particular patient) and "prognostic" variables that can help forecast a patient outcome but do not aid in treatment selection.

A major challenge in precision medicine is that most baseline patient measures typically have small moderating effects and thus individually contribute little to informed treatment decisions. Unconstrained regression models with p predictors that may also include the treatment variable and predictor-by-treatment interactions become unwieldy, unstable and difficult to interpret when p is large, or even moderate. Various strategies have been proposed to deal with the problem identifying prescriptive variables and estimating decision rules when several baseline measures are available. Gunter et al. [14] propose a ranking procedure to be applied to the individual baseline measures, after which a forward variable selection algorithm is employed with the restriction that a main effect of a variable be included when the interaction between a variable and treatment is selected in the model. Qian and Murphy [31] on the other hand, consider a least absolute shrinkage and selection operator (LASSO; Tibshirani [45]) penalty for choosing baseline predictors with a focus on choosing a model for the outcome that ensures good performance with respect to the value of the estimated treatment decisions, see Section 1. Lu et al. [24] propose a method for obtaining a good model for the treatment effect that is robust to misspecifying the baseline model. Ciarleglio et al. [7] extend that methodology to allow functional data objects (such as spectra estimated from EEG assessments) to be incorporated as baseline features. Recognizing that estimation based on minimizing the prediction error may not necessarily result in a decision that maximizes the clinical benefit, Zhao et al. [53] proposed an alternative method, using support vector machines [11], for developing treatment decision rules that are based on directly maximizing the clinical benefit. Petkova et al. [30] develops a methodology for combining several baseline measures for the specific purpose of finding a single powerful treatment effect modifier in the context of the classic linear model, which is called a generated effect modifier (GEM).

Based on available methodology at the time of writing this manuscript, we have identified the following approaches summarized in Table 1 that are applicable to the EMBARC study. They were selected based on the criteria that (i) the methods should be able to estimate optimal treatment decisions when the outcome variable is either continuous (e.g., rate of symptoms improvement) or binary (e.g., remission status), and (ii) there should be a variable selection algorithm embedded in the method.

Each of these methods has been shown to be useful in particular situations, but to our knowledge, there are no studies that compare them directly and make recommendations for their utility in different situations. In a preliminary simulation study, the results of which are not shown here, these methods were compared in terms of value of the treatment decision rule across (i) varying numbers of "true" and "noise" predictors; (ii) different true data generating models; and (iii) a range of magnitudes of the error variances. The results indicated that the comparative advantages of one method versus another depended on the true data generating model with no method uniformly dominating the rest. For this reason, all methods in Table 1 will be employed to determine treatment decision rules based on the training data set, and these rules afterwards will be applied to the test data set. In this way, the methods will be compared based on their performance in the test data set with respect to value of the treatment decision. Based on the comparison of the rules in the test data set, the best-performing treatment decision rules will be nominated for further validation in a future randomized clinical trial.

5.2. Extension to functional predictors

The methods described in the previous section are focused on making patient-specific treatment decisions based on a set of scalar-valued predictor variables. In the EMBARC study, many of these variables are derived from the biological brain data of various modalities collected at baseline. To supplement the analysis based only on scalar variables, a potentially more powerful approach would take advantage of the natural spatial and/or temporal structure of the imaging data using methods adapted from the general field of functional data analysis [32]. Rather than use some average of the functional brain modality data over a particular region of interest, for instance, we could instead use the entire image as a functional predictor, for example, 1-, 2- or 3-dimensional data object.

The analysis of functional data, like those described here, has been a topic of great interest in the past decade. Spurred in part by the increasing rate of generation of such data in diverse scientific fields, methods for functional data analysis are being developed at a rapid pace. We are developing and employing new methods for identification of treatment effect modifiers when the predictors are functional data objects, as well as for combining scalar and functional predictors, see, for example, Refs. [7–9].

6. Strategy for developing indices for personalized treatment decisions

6.1. Overview

The set of potential baseline moderators are gathered from six data sources/modalities: clinical, behavioral, EEG, DTI, structural MRI and fMRI. We will approach the identification of a DTRI using both scalar and functional moderators for making treatment decisions first within a data modality and then we will combine the modalities. Within a data modality, the set of predictors employed will progress from the most exclusive (First Tier only), through First and Second Tiers combined, to least restricted (First, Second and Third Tiers combined). The goal of such a progression is to be able to evaluate the values of treatment decisions based on known or anticipated patient characteristics and to quantify the improvements in value when new patient features are added, thus moving

Table 1
Methods for developing treatment decision rules.

Abbreviation	Description	Citation	Comment
Q	Q-learning	[31]	Performs variable selection using a LASSO penalty, but chooses the tuning parameters based on maximizing the value of the treatment decision resulting from the selected model. Extended to a generalized linear model (GLM)
OWL	Outcome Weighted Learning	[23,38,53]	Uses the method of Ref. [38] for variable selection and the modified estimation of the weights of Ref. [23]
QT	Estimating interactions based on the modified covariates approach	[44]	While the Tian et al. [44] performs variable selection using a LASSO penalty with tuning parameters selected to minimize the prediction error, we choose the tuning parameters to optimize the value of the treatment decision rule, as in Q-learning
ZQT	General weighted classifica- tion method	[52]	Uses QT to estimate classification weights and combines this with a classification algorithm
ZQT-SVM	ZQT with support vector machine	[11]	ZQT with SVM for classification
ZQT-CART	ZQT with classification and regression trees	[3]	ZQT with CART for classification

from least to most exploratory investigations. The analyses will be conducted in the following order:

1. Combine scalar predictors within a given data modality, e.g., EEG.
2. Identify a best treatment effect modifier based on
   - A single functional data object from a given data modality, such as, for example from EEG, the current source density over the frequency range 3–16 Hz at a given electrode using a functional linear model.
   - A combination of all functional data objects from a given data modality, such as, for ex- ample from EEG, the current source density over the frequency range 3–16 Hz, measured at all 72 electrodes.
3. Combination of scalar and functional variables from a given imaging modality and the clinical and demographic data, see Section 6.2 for justification.

To address potential issues regarding multicollinearity, we first note that the variables in the first tier were carefully vetted by experts in the respective data modalities and a single measure was nominated from possibly multiple ways of measuring the same construct. Thus, gross multicollinearity due to multiple measures of the same construct is eliminated by the expert preselection of variables and the remaining variables truly correspond to different characteristics. We also note that the clinical and demographic data are only modestly related to the imaging, EEG and behavioral data. The methods for developing treatment decision rules specified in Section 5.1 all incorporate some variable selection mechanism, which is often an effective means of dealing with multicollinearity. Since the primary objective of the analysis is in terms of prediction, methods like Q-learning [31] that use the LASSO will tend to select predictors from among a set of correlated predictors in order to optimize the value of a decision and hence will mitigate multicollinearity problems. Second, among the 2nd and 3rd tier variables, multicollinearity within a given imaging or EEG modality can happen (i) if there are several measures that represent the same construct (as noted for tier 1), such as for example, when different filters are applied prior to computing alpha band amplitudes for EEG data; and (ii) when measurements on adjacent locations in the brain are correlated. This will be dealt with by not including multiple measures of the same construct together in the models, and by treating the EEG and imaging data as functional when possible. The variables in the 3rd tier, which are most numerous and consist of everything that can be computed from the exhaustive baseline assessments and about which no hypotheses have been postulated, will be subjected to sure independence screening [12], based on model (2), prior to combining them within or between modalities.

The main analysis will only include scalar variables, unless methods for analysis of functional data are available. Different approaches will be developed and tested on the training data set. The development will involve evaluations of the statistical stability of the models defining the DTRIs using cross-validation with the training data to obtain an assessment of how well the DTRIs are likely to perform using the test set [5,19]. The approaches that perform well in terms of cross-validation with the training data, will be evaluated on the hold-out test data set. A small number (e.g., one or two) DTRIs will be nominated from each data modality to be studied further if warranted.

6.2. Dealing with missing data in the covariates

Due to potential problems with processing imaging data, it is expected that some of the brain imaging data will not be useable. In addition, some study participants might not be able to complete the entire sequence of assessments specified in the study protocol. Therefore, only a subset of all study participants is expected to have complete baseline and week-1 data. For an individual participant, the typical missing data pattern is expected to involve all measures from one or more modalities, while all measures from the remaining biomarker modalities may be complete for that participant. For example, a study participant might not have any imaging data under the Emotion Recognition task because of excessive head motion during the scan, but if s/he has a good fMRI scan under the Reward task, all measures related to that task would be observed. This typical pattern of missing data is one reason that we plan to develop biosignatures within each modality separately and initially consider only combining each biomarker modality with the clinical and demographic data, which are expected to have minimal missingness.We will also attempt to impute the missing covariate data. Multiple imputations will be employed and thorough diagnostics of the results from the imputations will be conducted [1,36]; Su et al. [40]. The diagnostic step will be particularly important given that the most common pattern of missingness is for all variables from a particular modality to be simultaneously missing. Hence all variables from a given modality (e.g., fMRI) will need to be imputed based on the other modalities (e.g., clinical and demographic, EEG, DRI and behavioral phenotyping). In these cases, data from the other modalities might not be sufficient for a quality imputation. The analyses outlined in Section 6.1 will be repeated using the imputed data sets and combined inferences will be performed, following Ref. [37]. Results from applying the DTRIs obtained using the complete training data only and the multiple imputed training data, on the test data set will inform further whether it is useful to impute baseline data in studies designed to discover biosignatures

for treatment response. If the quality of the imputations is unsatisfactory, results from only complete case analyses will be reported.

6.3. Validation

The validation of the proposed biosignatures (both those resulting in a DTRI and those that do not explicitly produce such indices) will be based on the value of each treatment decision rule corresponding to each of the proposed biosignatures, as stated in Section 1. For example, suppose a nominated DTRI is Z, defined as a linear combination of baseline predictors. Furthermore, suppose Z has been obtained by maximizing its effect as a moderator in the training data set, based on the following linear model:

(3) $E[Y|A,Z] = \beta_0 + \beta_1 Z + \beta_2 A + \beta_3 AZ.$

If higher values of Y are preferred, the treatment decision formulated based on Z and (3) is:

if $\beta_2 + \beta_3 Z > 0$, or equivalently $Z > -\frac{\beta_2}{\beta_3}$, give treatment 1.

if $\beta_2 + \beta_3 Z \leq 0$, or equivalently $Z \leq -\frac{\beta_2}{\beta_3}$,give treatment 1, assuming $\beta_3 > 0$ (with the inequalities switched if $\beta_3 < 0$). The value of each of the decision rules based on biosignatures, derived using the training data, will be computed on the validation data set and will be compared against the values of the following decisions:

d^R: Random assignment of sertraline or placebo in a ratio 1:1;

d^S: All patients are assigned sertraline;

d^P: All patients are assigned placebo.

Confidence intervals for the differences in the values between the derived biosignature and each of the three decisions d^R, d^S and d^P will be obtained using a bootstrap procedure (see e.g., Refs. [20,31,38]). In a similar way we will compare the biosignatures obtained from different data modalities.

6.4. Final DRTIs

Of the set of DTRIs nominated based on analyses of the training data, we will select a handful that show the best performance in the validation using the test set. As a final step, the methods for developing the "optimal" DTRIs will be applied to the entire study sample (training and validation sets). The resulting DTRIs and the methods used in their development will be reported.

7. Patient characteristics one week post randomization

In EMBARC, study participants are assessed one week after randomization with the entire baseline battery except the clinical and structural MRI measurements. Of interest here is whether we can identify early correlates of treatment response and whether any early biological changes can help inform treatment decisions. The data objects here will be changes from baseline to one week post-randomization. There are no specifically identified measures prior to study completion. The analyses will follow the outline in Section 5 for developing indices for personalized treatment.

8. Discussion

Here we have presented the plan for analyses to address the major goals of the EMBARC study. This plan will be followed in the reporting of the major results from this study. The EMBARC study is generating an unparalleled resource for discovery of patient characteristics related to response to antidepressant treatment. While the main analysis will follow the outlined plan, we envision a long and extended use of this data resource. No uniformly best method for developing optimal treatment decisions is known to date and the performance of such methods depend on the size, complexity and signal to noise ratio of the true biological model. Therefore, as new methods for combining biomarkers, and estimating optimal treatment decisions with variable selection are being developed, they will be applied to the EMBARC data. The results from those later analyses will be assessed and validated in a similar way as described above and also according to new measures of performance when such measures are introduced. Furthermore, the EMBARC data collection will be used to address numerous other important research questions, such as for example, predicting treatment outcome (as opposed to finding covariates that predict differential treatment effect) and better understanding the placebo effect.

The present study is the first large scale study of its kind that has obtained clinical and extensive biological variables across multiple sites in a randomized placebo controlled trial specifically designed to evaluate the differential depression treatment response index for patients with early onset, recurrent major depressive disorder. The depth and breadth of clinical and biological variables collected affords a unique opportunity to evaluate potential biomarkers based on multi-modal baseline and week 1 assessments. These biomarkers serve as potential DTRIs, which are first developed on a training set and then validated on an independent test set. If successful, these findings will: 1) provide an index that could readily be used in clinical practice to match patients with treatment; and 2) provide a proof-of-concept for future studies to prospectively assess these and other indices in a hypothesis testing study. Furthermore, this will be the first evaluation of such an approach in developing and validating a DTRI for placebo response.

We emphasize, however, that as in all studies intended to determine an optimal treatment regime, any selected decision rule should be validated in a randomized clinical trial. In the case of EMBARC, the treatment decision is either sertraline or placebo. A randomized clinical trial to evaluate the selected decision rule might be a two parallel arms study, where in one of the arms the treatment will be assigned according to the selected treatment decision rule and, in the other arm, treatment would be assigned at random, e.g., either setraline or placebo. Alternatively, a similar design would be a three parallel arms design where in the first arm, treatment will be assigned according to the selected rule, subjects in the second arm will all be assigned to setraline and subjects in the third arm all will be assigned placebo. While the two arms design emphasizes a comparison of the selected treatment decision rule to the decision to treat depressed subjects with either the drug or placebo assigned at random, the three arms design underscores the interest in the comparison between using the selected decision rule versus treating everyone with the drug, which is a more realistic treatment strategy. Perhaps a more clinically relevant three arms study design would replace the placebo with an alternative active treatment, say an antidepressant of different class or a psychotherapy. Such a study would allow not only a direct comparison of the selected treatment decision rule with the alternative treatment, but also would generate data that might be used to develop rules for deciding between sertraline and an alternative active treatment. The follow up studies for confirming the utility of the treatment decision rules developed in the EMBARC study are standard efficacy trials and are subject to the sample size and power considerations appropriate for such investigations.

Acknowledgement

The authors would like to thank the Editors and reviewers for their constructive suggestions and comments for improving this paper. This work is supported by NIMH/NIH under awards

U01MH092221 (to Dr. Trivedi) and U01MH092250 (to Drs. McGrath, Parsey, and Weissman), as well as by the EMBARC National Coordinating Center at the University of Texas Southwestern Medical Center (Trivedi MH PI) and the Data Center at Columbia University (Dr. Adams). Valeant Pharmaceuticals donated the bupropion hydrochloride extended-release used in this study. The work is also supported by NIMH/NIH under R01MH099003 (to Dr. Petkova). The content in this article is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

Appendix. First Tier baseline characteristics

This appendix provides a brief description of the First Tier baseline characteristics.

- Clinical: includes anxious depression, melancholic depression, anger attacks, probable Axis II diagnosis, hypersomnia/fatigue, childhood trauma, family history of MDD or Bipolar disorder and chronicity
- Behavioral Phenotyping: includes number of letters reported in the Controlled Oral Word Association Test, reaction time in the Choice RT task, number of correct responses in the “A not B” Working memory task, response time in the “A not B” Working memory task, accuracy and reaction time in post-correct vs. post-incorrect trials in the Flanker task, accuracy and reaction time in incongruent vs. congruent trials in the Flanker task, reward learning (Response Bias in Block 3 - Response Bias in Block 1), response Bias in Block 3 in the Probabilistic Reward Task, total words on the Word Fluency Test and reaction time on correct responses in “A not B” Working memory task
- EEG: includes resting EEG Alpha current source density measures of condition-dependent EEG alpha, Auditory Evoked Potentials (N1 Amplitude), loudness dependency of auditory evoked potential from tone loudness of 60, 70, 80, 90, 100 dB and rostral anterior cingulate cortex theta current density
- DTI: includes fractional anisotropy of Superior Temporal Cortex
- Structural MRI: includes cortical thickness of left lateralorbitalfrontal and left precentral regions
- fMRI: includes from the Emotion recognition task, difference between activation under congruent and incongruent conditions in the pregenual anterior cingulate, dorsal anterior cingulate and pregenual cingulate/right amygdala psychophysiological interaction; from the Reward task, Beckmann Region (BR) 3 in anticipation, left ventrolateral prefrontal cortex in anticipation and right ventral striatum in outcome; from resting state fMRI functional connectivity between left amygdala and BR 2 at Time 1, left amygdala and BR 2 at Time 2, left ventral striatum to BR 3 at Time 1 and left ventral striatum to BR 3 at Time 2.

Second Tier baseline characteristics will include: Cerebral blood flow derived variables from imaging data under emotion recognition task, Reward task and resting state connectivity.

References

[1] K. Abayomi, A. Gelman, M. Levy, Diagnostics for multiple imputations, Appl. Stat. 57 (2008) 273–291.
[2] K. Anderson, P. Odell, P. Wilson, W. Kannel, Cardiovascular disease risk profiles, Am. Heart J. 121 (1991) 293–298.
[3] L. Breiman, J.H. Friedman, R.A. Alshen, C.J. Stone, CART: Classification and Regression Trees, Wadsworth, Belmont, CA, 1983.
[4] A. Cattaneo, M. Gennerelli, R. Uher, G. Breen, A. Farmer, K. Aitchison, I. Craig, C. Anacker, P. Zunsztain, P. McGuffin, C. Pariante, Candidate genes expression profile associated with antidepressants response in the GENDEP study: differentiating between baseline ‘predictors’ and longitudinal ‘targets’, Neuropsychopharmacology 38 (2013) 377–385.
[5] B. Chakraborty, E. Laber, Y. Zhao, Inference for optimal treatment regimes using adaptive m-Out-of-n bootstrap scheme, Biometrics 69 (2013) 714–723.
[6] W. Chen, D. Ghosh, T. Raghunatan, M. Norkin, D. Sargent, G. Bepler, On Bayesian methods of exploring qualitative interactions for targeted treatment, Stat. Med. 31 (2012) 3693–3707.
[7] A. Ciarleglio, E. Petkova, T. Tarpey, R.T. Ogden, Treatment decisions based on scalar and functional baseline covariates, Biometrics 71 (2015) 884–894.
[8] A. Ciarleglio, E. Petkova, T. Tarpey, R.T. Ogden, Flexible functional regression methods for estimating individualized treatment regimes, STAT 5 (2016) 185–199.
[9] A. Ciarleglio, E. Petkova, T. Tarpey, R.T. Ogden, Variable selection for treatment decision rules with scalar and functional predictors, Stat. Med. (In revision) (2017).
[10] M. Cloitre, E. Petkova, Z. Su, B. Weiss, Patient characteristics as a moderator of PTSD treatment outcome: combining symptom burden and strengths, Br. J. Psychiat. Open 2 (2016) 101–106.
[11] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[12] J. Fan, J. Lv, Sure independence screening for ultra-high dimensional feature space, J. of the R. Stat. Soc. Ser. B 70 (2008) 849–911.
[13] M. Fava, A.J. Rush, M. Trivedi, A.A. Nierenberg, M.E. Thase, H.A. Sackheim, F.M. Quitkin, S.R. Wisniewski, P.W. Lavori, J.F. Rosenbaum, D.J. Kupfer, Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study, Psychiat. Clinics of N. Am. 26 (2003) 457–494.
[14] L. Gunter, J. Zhu, S. Murphy, Variable selection for qualitative interactions in pres- onalized medicine while controlling the family-wise error rate, J. Biopharm. Stat. 21 (2011) 1063–1078.
[15] J. Hennings, M. Uhr, T. Klengel, P. Webber, B. Pütz, D. Czamara, M. Ising, F. Holsboer, S. Lucae, RNA expression profiling in depressed patients suggests retinoid-related orphan receptor alpha as a biomarker for antidepressant response, Translat. Psychiat. 5 (2015) e538.
[16] M. Korgaonkar, L. Williams, Y. Song, T. Usherwood, S. Grieve, Diffusion tensor imaging predictors of treatment outcomes in major depressive disorder, Br. J. Psychiat. 205 (2014) 321–328.
[17] H.C. Kraemer, Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach, Stat. Med. 32 (2013) 1964–1973.
[18] V. Krishnan, E. Nestler, The molecular neurobiology of depression, Nature 455 (2008) 894–902.
[19] E. Laber, D. Lizotte, M. Qian, W. Pelham, S. Murphy, Dynamic treatment regimes: technical challenges and applications, Electronic J. Stat. 8 (2014) 1225–1272.
[20] E. Laber, Y.-Q. Zhao, Tree-based methods for individualized treatment regimes, Biometrika 102 (2015) 501–514.
[21] A. Leuchter, I. Cook, L. Marangell, W. Gilmer, K. Burgoyne, R. Howland, M. Trivedi, S. Zisook, R. Jain, T. McCracken, M. Fava, D. Iosifescu, S. Greenwald, Comparative effectiveness of biomarkers and clinical indicators for predicting outcomes of SSRI treatment in major depressive disorder: results of the BRITE-MD study, Psychiat. Res. 169 (2009) 124–131.
[22] J. Li, I. Chan, Detecting qualitative interactions in clinical trials: an extension of range test, J. Biopharm. Stat. 16 (2006) 831–841.
[23] Y. Liu, Y. Wang, M. Kosorok, Y.-Q. Zhao, D. Zeng, Robust hybrid learning for estimating personalized dynamic treatment regimens, J. Am. Stat. Assoc. (Under Review) (2016).
[24] W. Lu, H. Zhang, D. Zeng, Variable selection for optimal treatment decision, Stat. Meth. Med. Res. 22 (2011) 493–504.
[25] C. McGrath, M. Kelley, B. Dunlop, P. Holtzheimer, W. Craighead, H. Mayberg, Pretreatment brain states identifying likely nonresponse to standard treatments for depression, Biol. Psychiat. 76 (2014) 527–535.
[26] J. Mundt, A. Vogel, D. Feltner, W. Lenderking, Vocal acoustic biomarkers of depres- sion severity and treatment response, Biol. Psychiat. 72 (2012) 580–587.
[27] S. Murphy, Optimal dynamic treatment regimes (with discussion), J. R. Stat. Soc. Ser. B 58 (2003) 331–366.
[28] E. Petkova, T. Tarpey, Partitioning of functional data for understanding heterogeneity in psychiatric conditions, Stat. Interface 2 (2009) 413–424.
[29] E. Petkova, T. Tarpey, A. Ciarleglio, R.T. Ogden, Deriving a scalar measure from a longitudinal trajectory with applications to placebo response, Stat. Med. (2016) (Submitted for publication).
[30] E. Petkova, T. Tarpey, R. Ogden, Z. Su, Generated effect modifiers (GEMs) in randomized clinical trials, Biostatistics 18 (1) (2017) 105–118.
[31] M. Qian, S. Murphy, Performance guarantees for individualized treatment rules, Ann. Stat. 39 (2011) 1180–1210.
[32] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[33] J. Robins, Optimal structured nested models for optimal sequential decisions, in: D. Lin, P.J. Heagerty (Eds.), Proceedings of the Second Seattle Symposium on Biostatistics, Springer, New York, 2004, pp. 189–326.
[34] D. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, J. Edu. Psychol. 66 (1974) 688–701.
[35] A.J. Rush, M.H. Trivedi, S.R. Wisniewski, A.A. Nierenberg, J.W. Stewart, D. Warden, G. Niederehe, M.E. Thase, P.W. Lavory, B.D. Lebowitz, P.J. McGrath, J.F. Rosenbaum, H.A. Sackeim, D.J. Kupfer, J. Luther, M. Fava, Acute and longer-term outcomes in depressed outpatients requiring one or several treatment

steps: a STAR*D report, Am. J. Psychiat. 163 (2006) 1905–1917.
[36] J.L. Schafer, Analysis of Incomplete Multivariate Data, Chapmen and Hall, New York, 1997.
[37] J.L. Schafer, Multiple imputation: a primer, Stat. Meth. Med. Res. 8 (1999) 3–15.
[38] R. Song, M. Kosorok, D. Zeng, Y. Zhao, E.B. Laber, M. Yuan, On sparse representation for optimal individualized treatment selection with penalized outcome weighted learning, STAT 4 (2015) 59–68.
[39] X. Song, M. Pepe, Evaluating markers for selecting a patient's treatment, Biometrics 60 (2004) 874–883.
[40] Y.-S. Su, A. Gelman, J. Hill, M. Yajima, Multiple imputation with diagnostics (mi) in R: opening windows into the black box, J. Stat. Softw. 45 (1) (2011) 1–31.
[41] T. Tarpey, E. Petkova, Y. Lu, U. Govindarajulu, Optimal partitioning for linear mixed effects models: applications to identifying placebo responders, J. American Statistical Association 105 (2010) 968–977.
[42] T. Tarpey, E. Petkova, R.T. Ogden, Profiling placebo responders by self-consistent partitioning of functional data, J. Am. Stat. Assoc. 98 (2003) 850–858.
[43] T. Tarpey, E. Petkova, L. Zhu, A new approach to stratified psychiatry via convexity-based clustering with applications towards moderator analysis, Stat Interface 9 (2016) 255–266.
[44] L. Tian, A. Alizadeh, A. Gentles, R. Tibshirani, A simple method for estimating interactions between a treatment and a large number of covariates, J. Am. Stat. Assoc. 109 (2014) 1517–1532.
[45] R. Tibshirani, Regression shrinkage and selection via the LASSO, J. R. Stat. Soc. Ser. B 58 (1996) 267–288.
[46] M.H. Trivedi, P. McGrath, M. Fava, R.V. Parsey, B. Kurian, M.L. Phillips, M.A. Oquendo, G. Bruder, D. Pizzagalli, M. Toups, C. Cooper, P. Adams, S. Weyandt, D. Morris, B. Grannemann, R.T. Ogden, R. Buckner, M. McInnis, H.C. Kraemer, E. Petkova, T. Carmody, M. Weissman, Establishing Moderators and Biosignatures of Antidepressant Response in Clinical care (EMBARC): rationale and design, J. Psychiat. Res. 78 (2016) 11–23.
[47] M.H. Trivedi, A.J. Rush, S.R. Wisniewski, A.A. Nierenberg, D. Warden, L. Ritz, G. Norquist, R.H. Howland, B. Lebowitz, P.J. McGrath, K. Shores-Wilson, M.M. Biggs, G.K. Balasub- ramani, M. Fava, Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice, Am. J. Psychiat. 163 (2006) 28–40.
[48] R. Uher, K. Tansey, T. Dew, W. Maier, O. Mors, J. Hauser, M. Dernovsek, N. Heingsburg, D. Souery, A.F, P. McGuffin, An inflammatory biomarker as a differential predictor of outcome of depression treatment with escitalopram and nortriptyline, Am. J. Psychiat. 171 (2014) 1278–1286.
[49] R. Wang, J. Ware, Detecting moderator effects using subgroup analyses, Prev. Sci. 14 (2013) 111–120.
[50] S. Wellek, Testing for absence of qualitative interactions between risk factors and treatment effect, Biometric. J. 39 (1997) 809–821.
[51] O.M. Wolkowitz, S.H. Mellon, E.S. Epel, J. Lin, V.I. Reus, F.S. Dhabhar, E.H. Blackburn, Resting leukocyte telomerase activity is elevated in major depression and predicts treatment response, Mol. Psychiat. 17 (2012) 164–172.
[52] B. Zhang, A.A. Tsiatis, M. Davidian, M. Zhang, E. Laber, Estimating optimal treatment regimes from classification perspective, STAT 1 (2012) 103–114.
[53] Y. Zhao, D. Zeng, A.J. Rush, M.P. Kosorok, Estimating individualized treatment rules using outcome weighted learning, J. Am. Stat. Assoc. 107 (2012) 1106–1118.