HHS Public Access
Author manuscript
Psychol Med. Author manuscript; available in PMC 2020 May 01.

Published in final edited form as:
Psychol Med. 2019 May ; 49(7): 1118–1127. doi:10.1017/S0033291718001708.

Personalized prediction of antidepressant versus placebo response: Evidence from the EMBARC study

Christian A. Webb, Ph.D.[1], **Madhukar H. Trivedi, M.D.**[2], **Zachary D. Cohen, M.A.**[3], **Daniel G. Dillon, Ph.D.**[1], **Jay C. Fournier, Ph.D.**[4], **Franziska Goer, M.A.**[1], **Maurizio Fava, M.D.**[5], **Patrick J. McGrath, M.D.**[6], **Myrna Weissman, Ph.D.**[6], **Ramin Parsey, M.D., Ph.D.**[7], **Phil Adams, Ph.D.**[6], **Joseph M. Trombello, Ph.D.**[2], **Crystal Cooper, Ph.D.**[2], **Patricia Deldin, Ph.D.**[8], **Maria A. Oquendo, M.D., Ph.D.**[3], **Melvin G. McInnis, M.D.**[8], **Quentin Huys, M.D., Ph.D.**[9], **Gerard Bruder, Ph.D.**[6], **Benji T. Kurian, M.D.**[2], **Manish Jha, M.D.**[2], **Robert J. DeRubeis, Ph.D.**[3], and **Diego A. Pizzagalli, Ph.D.**[1]

[1]Harvard Medical School – McLean Hospital, Boston, MA [2]University of Texas, Southwestern Medical Center, Dallas, TX [3]University of Pennsylvania, Philadelphia, PA [4]University of Pittsburgh School of Medicine, Pittsburgh, PA [5]Harvard Medical School – Massachusetts General Hospital, Boston, MA [6]New York State Psychiatric Institute & Department of Psychiatry, College of

Corresponding author: Christian A. Webb, Ph.D., Center for Depression, Anxiety and Stress Research, Room 231, McLean Hospital, 115 Mill Street, Belmont, MA 02478., cwebb@mclean.harvard.edu, Phone: 617-855-4429, Fax: 617-855-4231.

Conflict of Interest

In the last three years, the authors report the following financial disclosures, for activities unrelated to the current research:

Dr. **Trivedi**: reports the following lifetime disclosures: research support from the Agency for Healthcare Research and Quality, Cyberonics Inc., National Alliance for Research in Schizophrenia and Depression, National Institute of Mental Health, National Institute on Drug Abuse, National Institute of Diabetes and Digestive and Kidney Diseases, Johnson & Johnson, and consulting and speaker fees from Abbott Laboratories Inc., Akzo (Organon Pharmaceuticals Inc.), Allergan Sales LLC, Alkermes, AstraZeneca, Axon Advisors, Brintellix, Bristol-Myers Squibb Company, Cephalon Inc., Cerecor, Eli Lilly & Company, Evotec, Fabre Kramer Pharmaceuticals Inc., Forest Pharmaceuticals, GlaxoSmithKline, Health Research Associates, Johnson & Johnson, Lundbeck, MedAvante Medscape, Medtronic, Merck, Mitsubishi Tanabe Pharma Development America Inc., MSI Methylation Sciences Inc., Nestle Health Science-PamLab Inc., Naurex, Neuronetics, One Carbon Therapeutics Ltd., Otsuka Pharmaceuticals, Pamlab, Parke-Davis Pharmaceuticals Inc., Pfizer Inc., PgxHealth, Phoenix Marketing Solutions, Rexahn Pharmaceuticals, Ridge Diagnostics, Roche Products Ltd., Sepracor, SHIRE Development, Sierra, SK Life and Science, Sunovion, Takeda, Tal Medical/Puretech Venture, Targacept, Transcept, VantagePoint, Vivus, and Wyeth-Ayerst Laboratories.

Dr. **Dillon**: funding from NIMH, consulting fees from Pfizer Inc.

Dr. **Fava**: Dr. Fava reports the following lifetime disclosures: http://mghcme.org/faculty/faculty-detail/maurizio_fava

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 Oxford University Press, Perseus Press, the American Psychiatric Association Press, and MultiHealth Systems.

Dr. **Oquendo**: funding from NIMH; royalties for the commercial use of the Columbia-Suicide Severity Rating Scale. Her family owns stock in Bristol Myers Squibb.

Dr. **McInnis**: funding from NIMH; consulting fees from Janssen and Otsuka Pharmaceuticals.

Dr. **McGrath** has received research grant support from Naurex Pharmaceuticals (now Allergan), Sunovion, and the State of New York.

Dr. **Pizzagalli**: funding from NIMH and the Dana Foundation; consulting fees from Akili Interactive Labs, BlackThorn Therapeutics, Boehringer Ingelheim, Pfizer Inc. and Posit Science.

Dr. Trombello currently owns stock in Merck and Gilead Sciences and within the past 36 months previously owned stock in Johnson & Johnson.

Drs. **Adams**, **Cohen, Bruder**, **Cooper**, **Deldin**, **DeRubeis**, **Fournier**, **Huys**, **Jha, Kurian, McGrath**, **Parsey**, **Webb**, Ms. **Goer**: report no financial conflicts.

Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Physicians and Surgeons of Columbia University, New York, NY [7]Stony Brook University, Stony Brook, NY [8]University of Michigan, Ann Arbor, MI [9]University of Zurich, Zurich, Switzerland

Abstract

Background: Major Depressive Disorder (MDD) is a highly heterogeneous condition in terms of symptom presentation and, likely, underlying pathophysiology. Accordingly, it is possible that only certain individuals with MDD are well-suited to antidepressants. A potentially fruitful approach to parsing this heterogeneity is to focus on promising endophenotypes of depression, such as neuroticism, anhedonia and cognitive control deficits.

Methods: Within an eight-week multisite trial of sertraline vs. placebo for depressed adults (n =216), we examined whether the combination of machine learning with a Personalized Advantage Index (PAI) can generate individualized treatment recommendations on the basis of endophenotype profiles coupled with clinical and demographic characteristics.

Results: Five pre-treatment variables moderated treatment response. Higher depression severity and neuroticism, older age, less impairment in cognitive control and being employed were each associated with better outcomes to sertraline than placebo. Across 1000 iterations of a 10-fold cross-validation, the PAI model predicted that 31% of the sample would exhibit a clinically meaningful advantage (post-treatment Hamilton Rating Scale for Depression [HRSD] difference ≥ 3) with sertraline relative to placebo. Although there were no overall outcome differences between treatment groups (d =.15), those identified as optimally suited to sertraline at pre-treatment had better week 8 HRSD scores if randomized to sertraline (10.7) than placebo (14.7)(d =.58).

Conclusions: A subset of MDD patients optimally suited to sertraline can be identified on the basis of pre-treatment characteristics. This model must be tested prospectively before it can be used to inform treatment selection. However, findings demonstrate the potential to improve individual outcomes through algorithm-guided treatment recommendations.

Keywords

antidepressant; placebo; prediction; depression; endophenotype; machine learning; precision medicine

Introduction

Meta-analyses reveal that average differences in depressive symptom improvement between antidepressant medications (ADMs; most commonly, selective serotonin reuptake inhibitors [SSRIs]) and placebo are often small (i.e., between-group differences in symptom change of less than 3 points on the Hamilton Depression Rating Scale (Hamilton 1960)) (Moncrieff *et al.* 2004; Kirsch *et al.* 2008; Fournier *et al.* 2010; Kirsch 2015; Jakobsen *et al.* 2017; Cipriani *et al.* 2018). A potential reason for this modest differentiation is that MDD is a highly heterogeneous condition in terms of symptom presentation and, likely, underlying pathophysiology (Wakefield & Schmitz 2013; Fried & Nesse 2015b, 2015a; Baldessarini *et al.* 2017). Accordingly, it is possible that subsets of depressed individuals are better suited to SSRIs, whereas others may derive limited benefit. For example, for certain depressed individuals the mere passage of time – possibly coupled with the expectation of improvement – may result in symptom remission (e.g., "spontaneous remitters"). Such individuals may not require SSRIs. Instead a less costly, low-intensity alternative intervention with minimal or no side effects may be sufficient for symptom remission (e.g., internet-based cognitive behavioral therapy (iCBT), which is included in the National Institute for Health and Care Excellence Guidelines (NICE 2018) as an efficacious intervention). Currently, treatment selection is largely based on trial-and-error. Approximately 55% to 75% of depressed individuals in primary care fail to achieve remission to first-line antidepressants, and 8% to 40% will switch to at least one other medication (Rush *et al.* 2006; Marcus *et al.* 2009; Schultz & Joish 2009; Vuorilehto *et al.* 2009; Milea *et al.* 2010; Saragoussi *et al.* 2012; Thomas *et al.* 2013; Ball *et al.* 2014; Mars *et al.* 2017). Identifying predictors of antidepressant response may ultimately inform the development of algorithms generating personalized predictions of optimal treatment assignment for clinicians and patients to consider in their decision-making regarding which intervention to select.

A range of pre-treatment variables (e.g., baseline clinical, demographic and neurobiological characteristics) have been examined as predictors of SSRI response.[1] Perhaps the most well-supported clinical moderator of SSRI vs. placebo response is baseline depressive symptom severity (Khan *et al.* 2002; Kirsch *et al.* 2008; Fournier *et al.* 2010). Meta-analyses indicate that in patients with MDD, lower levels of depressive symptom severity predicts minimal to no advantage of ADM over placebo, but that as depression severity increases, so does the magnitude of the advantage of ADM over placebo (Khan *et al.* 2002; Kirsch *et al.* 2008; Fournier *et al.* 2010). Other relevant predictors of greater ADM response include younger age (Fournier *et al.* 2009), being female (Trivedi *et al.* 2006; Jakubovski & Bloch 2014), higher education (Trivedi *et al.* 2006), being employed (Fournier *et al.* 2009; Jakubovski & Bloch 2014), lower anhedonia (McMakin *et al.* 2012; Uher *et al.* 2012a), non-chronic depression (Souery *et al.* 2007) and lower anxiety (Fava *et al.* 2008). Although each of these variables has limited predictive power when considered individually, recent advances in multivariable machine learning approaches allow for the combination of large sets of variables to predict treatment response (Gillan & Whelan 2017). Critically, to be clinically useful for treatment selection, predictors of treatment response must be applicable to *individual* patients. Consistent with the goals of precision medicine, such work aims to translate treatment outcome moderation findings to actionable, algorithm-guided treatment recommendations (Cohen & DeRubeis 2018).

We sought to use machine learning coupled with a recently published Personalized Advantage Index (PAI)(DeRubeis *et al.* 2014; Huibers *et al.* 2015) to predict treatment outcome at the individual level on the basis of pre-treatment patient data. Our aim was to use the above approach to identify the subset of patients who may be optimally suited to SSRI. With regards to machine learning approaches, we used four complementary variable

[1]The term predictor is used differently in different contexts (e.g., a "prescriptive predictor" or "moderator" (i.e., defined as a treatment group × predictor variable interaction) of outcome vs. a "prognostic" (i.e., treatment nonspecific) predictor of outcome) (Kraemer 2013; Fournier *et al.* 2009). Here, we include variables that have either demonstrated moderation (e.g., baseline depression and neuroticism moderating SSRI vs. placebo differences in outcome), but also include findings from single-arm designs demonstrating that a particular variable (e.g., educational level) predicts outcome within ADM.

selection procedures in an effort to identify a reliable and stable set of predictors from the initial, larger set of baseline variables. These procedures rely on different algorithms, such as decision tree-based ensemble learning methods (e.g., Random Forests) and regression-based methods (e.g., Elastic Net Regularization). This approach encouraged the selection of a set of predictors that emerged consistently across differing variable selection algorithms (See Variable Selection below). Data were derived from the multi-site EMBARC (*Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care*) clinical trial comparing SSRI (sertraline) vs. placebo (Trivedi *et al.* 2016). Of relevance, in a recent study based on EEG and cluster analyses, we reported that the substantial heterogeneity of MDD could be parsed by considering three putative endophenotypes of depression: neuroticism, blunted reward learning, and cognitive control deficits (Webb *et al.* 2016). Endophenotypes are hypothesized to lie on the pathway between genes and downstream symptoms, and are traditionally defined as meeting the following criteria (Gottesman & Gould 2003): (1) associated with the disease, (2) heritable, (3) primarily state-independent, (4) cosegregate within families, (5) familial association and (6) measured reliably (Goldstein & Klein 2014). We posited that depressed patients with certain endophenotype profiles may be differentially responsive to certain interventions (e.g., the cluster of depressed patients defined by relatively high levels of neuroticism may be more responsive to SSRIs). Indeed, there is evidence that depressed individuals characterized by elevated neuroticism may derive relatively greater therapeutic benefit from SSRIs relative to CBT (Bagby *et al.* 2008) or placebo (Tang *et al.* 2009). Thus, we examined whether the combination of putative endophenotypes (neuroticism, reward learning, cognitive control deficits, anhedonia) with both baseline clinical (depressive symptom severity, depression chronicity, anxiety severity) and demographic (gender, age, marital status, employment status, years of education) variables previously linked with antidepressant response could be used to identify individual depressed patients optimally suited to SSRIs. Plausible neuroimaging predictor variables (McGrath *et al.* 2013; Pizzagalli *et al.* 2018) were excluded from this particular study given that they are substantially more costly and time-consuming than the above set of clinical, demographic and behavioral variables, the latter of which could be reasonably integrated into a current psychiatric clinic for the purpose of treatment selection.

Methods and Materials

After providing informed consent, participants completed several behavioral and self-report assessments prior to enrolling in an 8-week, double-blind, placebo-controlled clinical trial of sertraline vs. placebo. The clinical trial design has been described in detail in a previous publication (Trivedi *et al.* 2016).

Participants

Eligible participants (ages 18–65) met DSM-IV criteria for a current MDD episode (SCID-I/P), scored ≥ 14 on the 16-item Quick Inventory of Depression Symptomatology (QIDS-SR_{16})(Rush *et al.* 2003), and were medication-free for ≥ 3 weeks prior to completing any study measures. Exclusion criteria included: history of bipolar disorder or psychosis; substance dependence (excluding nicotine) in the past six months or substance abuse in the past two months; active suicidality; or unstable medical conditions (see Supplemental

Methods). Data from 216 MDD subjects who passed quality control criteria for both Flanker and Probabilistic Reward Task and completed at least 4 weeks of treatment (American Psychiatric Association 2010; Fournier *et al*. 2013) were included (Supplemental Methods).

Endophenotype Measures

***NEO Five-Factor Inventory-3 (NEO-FFI-3)*(McCrae & Costa 2010).—**The 12-item neuroticism subscale from the NEO-FFI was used.

Probabilistic Reward Task (PRT).—The PRT uses a differential reinforcement schedule to assess reward learning (i.e., the ability to adapt behavior as a function of rewards), and has been described in detail in previous publications (Pizzagalli *et al*. 2005, 2008a)(see Supplemental Methods).

***Snaith-Hamilton Pleasure Scale (SHAPS)*(Snaith *et al*. 1995).—**The SHAPS is a 14-item self-report scale, with items asking about hedonic experience in the "last few days" for a variety of pleasurable activities. Items consist of four response categories, with "Strongly Agree" (=1), "Agree" (=2), "Disagree" (=3), "Strongly Disagree" (=4). Higher scores indicate higher anhedonia.

***Flanker Task* (Eriksen & Eriksen 1974).—**An adapted version of the Eriksen Flanker Task that included an individually-titrated response window was used to assess cognitive control (see Supplemental Methods)(Holmes *et al*. 2010).

Clinical Measures

***Hamilton Rating Scale for Depression (HRSD)*(Hamilton 1960).—**The 17-item HRSD, a clinician-administered measure of depressive symptom severity, was administered by trained clinical evaluators.

***Mood and Anxiety Symptoms Questionnaire (MASQ)*(Watson *et al*. 1995).—**The anxious arousal subscale from a 30-item adaptation of the MASQ (MASQ-AA) assessed anxiety.

Data Acquisition and Reduction

PRT.—The primary variable of interest was reward learning, which has been found to predict response to antidepressant treatment among inpatients with MDD (Vrieze *et al*. 2013). As in prior studies (Pizzagalli *et al*. 2008b; Vrieze *et al*. 2013), reward learning was defined as change in *response bias* (RB) scores throughout the task (here, from the first to the second block ($RB_{Block2} - RB_{Block1}$)).

Flanker Task.—The primary variable of interest was the *interference effect on accuracy*, defined as lower accuracy on incongruent relative to congruent trials, computed as [$Accuracy_{Compatible\ trials} - Accuracy_{Incompatible\ trials}$]. Higher scores reflect greater interference (i.e., reduced cognitive control).

Data Pre-Processing.—Missing data were imputed using a Random Forest-based imputation strategy (missForest (Stekhoven & Bühlmann 2012) package in R (R Core Team 2013)) (see Supplemental Methods)(Waljee *et al.* 2013). This approach can handle both categorical and continuous variables, and generates a single imputed dataset via averaging across multiple regression trees. Consistent with the recommendation of Kraemer and colleagues (Kraemer & Blasey 2004), continuous variables were mean-centered and categorical variables were transformed into binary variables with values of −0.5 and 0.5. Of the 216 individuals in this sample, 10.19% were missing data for the outcome variable (week 8 HRSD) and thus had their data imputed. There were no significant differences in week 8 completion rates between the SSRI (88.0%) or placebo (91.5%) conditions ($\chi 2(1)= 0.41$, $p = 0.52$). For additional analyses on dropout rates and medication/placebo adherence see Supplemental Methods.

Statistical Analyses

Variable Selection.—Prior to implementing the PAI algorithm, pre-treatment variables that interact with treatment group (SSRI or placebo) in predicting HRSD outcome (week 8 scores) must be selected. We implemented (1) Random Forests modeling (using the mobForest (Garge *et al.* 2013) package in R (R Core Team 2013)), (2) Elastic Net Regularization (glmnet package (Friedman *et al.* 2010)) and (3) Bayesian Additive Regression Trees (bartMachine package (Kapelner & Bleich 2016)). For each of these three models we entered all of our selected pre-treatment variables simultaneously: four endophenotype variables (Neuroticism [NEO-FFI-3], cognitive control [Flanker interference effect on accuracy], reward learning [PRT], and anhedonia [SHAPS]), three clinical variables (baseline severity of depressive symptoms [HRSD], baseline severity of anxiety [MASQ-AA] and chronic MDD [yes/no]) and five demographic variables (age, gender, marital status, employment status and years of education). Variables showing *Treatment Group* × *Predictor* variable interactions in two of the three models were entered into a final stepwise AIC-penalized bootstrapped variable selection (using the bootStepAIC package (Austin & Tu 2004)). For details on each of these approaches and how variables are selected from each model, see Supplemental Methods.

Generating PAIs

Briefly, to generate treatment recommendations with the PAI approach, a regression model is built and used to predict treatment outcome (week 8 HRSD) for each patient in SSRI and placebo separately. A patient's PAI is the signed difference between the two predictions (i.e., week 8 HRSD predicted in SSRI minus week 8 HRSD predicted in placebo), where a negative value reflects a predicted better outcome in SSRI, and a positive value reflects the reverse. Moreover, the magnitude of the absolute value of the PAI reflects the strength of the differential prediction, such that patients with larger PAIs, in either direction, are those who are most likely to evidence a substantially better outcome in their *PAI indicated*, relative to their *PAI non-indicated* treatment. To limit bias that could occur when evaluating model performance on individuals whose data were used to set model weights, PAIs were generated using 10-fold cross validation. This procedure ensures that each model is estimated absent any data from the patient whose outcome will be predicted (see *PAI*

Generation and *PAI Evaluation* in the Supplemental Methods for details; See also Alternative PAI Models below).

Evaluating PAIs

To assess whether PAI scores moderate treatment group differences in depression outcomes, we tested a *Treatment Group* × *PAI score* interaction with week 8 HRSD scores as the dependent variables. Next, and similar to previous PAI publications (DeRubeis *et al*. 2014; Huibers *et al*. 2015), to evaluate the utility of the PAIs we compared mean week 8 HRSD scores for SSRI-indicated individuals who were randomized to SSRI in comparison to SSRI-indicated participants who received placebo. We performed the analogous comparison for those identified as "Placebo-indicated." We then evaluated the above comparisons with only those patients for whom the absolute value of the PAI was 3 or greater (i.e., predicted to have a "clinically significant" advantage in one treatment condition)(DeRubeis *et al*. 2014). Finally, the entire ten-fold cross-validation procedure and evaluation was repeated 1000 times to generate stable estimates.

Results

Variable Selection

See Table 1 for variable selection results, including which variables were selected during each stage. The following variables survived the 4-step procedure and were included in the final model (see Figure 2 and Table 2):

$$Y = treatment * (depression\ severity\ [HRSD] + neuroticism\ [NEO-FFI-3] \\ + cognitive\ control\ [Flanker\ Interference\ (Accuracy)] + age + employment\ status)$$

Predicted Outcomes and PAIs

The average absolute value of PAI scores was 3.4 (SD = 2.6), indicating that our model predicted an average 3.4-point difference in week 8 HRSD scores between indicated and non-indicated treatment assignment. The absolute value of the PAI was 3 or greater in approximately half (48.6%) of the sample (see Figure 1 for distribution of PAI scores). Specifically, 31.5% of the sample was predicted to have a "clinically significant" advantage (DeRubeis *et al*. 2014) in the SSRI condition (PAI ≤ −3); whereas this value was 17.1% for placebo (PAI ≥ 3). In contrast, the model indicates that 51.4% of the sample was predicted to exhibit relatively minimal differences in outcome between treatment conditions.

Observed Outcomes in Indicated vs. Non-Indicated Treatment Condition

Full Sample.—First, it is important to highlight that, in the full sample, patients randomized to SSRI (M = 10.86; SD = 6.27) and placebo (M = 11.88; SD = 7.37) did not significantly differ in mean week 8 HRSD outcomes (adjusting for baseline HRSD scores) (F(1,213)= 0.92; p = .339; Cohen's d = .15; Figure 3, left panel). Critically, a significant *Treatment Group* × *PAI* interaction emerged in predicting week 8 HRSD scores, indicating that PAI scores moderated treatment group differences in outcome (F(1,212)= 6.68; p = .010). For the full sample, patients randomized to their PAI-indicated treatment condition (M = 10.39; SD = 6.97) were observed to have lower week 8 HRSD scores relative to those randomized to their contraindicated condition (M = 12.38; SD = 6.70)(d = .29, t(214)= 2.16; p = .032). For patients predicted to have better outcomes to SSRI than placebo (PAI < 0), those randomized to SSRI (M = 10.57; SD = 6.48) were observed to have lower week 8 HRSD scores than those randomized to placebo (M = 13.12; SD = 7.03)(d = .38, t(121)= 2.08; p = .040; see Figure 3, right panel). However, for patients predicted to have better outcomes to placebo (PAI > 0), those who received placebo (M = 10.18; SD = 7.54) did not differ significantly in outcome relative to those who received SSRI (M = 11.23; SD = 6.04) (d = .16; t(91)= 0.74; p= .460; see Figure 3, right panel).

***Largest PAIs* (PAI ≥ |3|).—**Among this subset, patients randomized to their indicated treatment condition (M = 9.53; SD = 6.68) were observed to have lower week 8 HRSD scores relative to those randomized to their contraindicated condition (M = 14.09; SD = 6.42) (d = .70, t(103) = 3.59; p < .001). SSRI-indicated patients randomized to SSRI (M = 10.68; SD = 7.04) were observed to have lower HRSD scores than those randomized to placebo (M = 14.66; SD = 6.83)(d = .58; t(66)= 2.34; p = .023; see Figure 3, right panel). Conversely, placebo-indicated patients randomized to placebo (M = 7.65; SD = 5.64) had better outcomes than those randomized to SSRI (M = 13.06; SD = 5.57)(d =1.01; t(35)= 3.07; p = .004; see Figure 3, right panel).

Alternative PAI Models

See Supplement for results from two alternative PAI models. First, a PAI model was run including all 12 *a priori* baseline variables, rather than the reduced set of 5 moderators emerging from our variable selection procedure. In other words, in the former model including all a priori variables, our variable selection procedure was not performed. The fact that a similar pattern of findings emerged in this control PAI analysis suggests that our findings are likely not attributable to overfitting due to running our PAI analysis on a reduced set of variables emerging from our variable selection steps. Second, to evaluate the utility of treatment recommendations based solely on depression severity (rather than our 5 moderator variables), we re-ran the above analysis using only baseline depressive symptom (HRSD) severity to inform the PAI, which did not yield significant findings.

Discussion

This study used the variable selection approach proposed by Cohen et al. (Cohen *et al.* 2017) combining machine learning with a previously published PAI algorithm (DeRubeis *et al.* 2014; Huibers *et al.* 2015) to generate individualized treatment recommendations on the basis of (i) putative behavioral endophenotypes of depression (Goldstein & Klein 2014; Webb *et al.* 2016) and (ii) clinical and demographic characteristics previously linked with antidepressant response. Ultimately, the goal is to translate research on predictors of antidepressant response to actionable treatment recommendations for individuals. First, it is important to highlight that the baseline moderators emerging from our machine learning variable selection steps are largely consistent with prior research. In particular, depressed individuals with higher baseline severity of depressive symptoms (Khan *et al.* 2002; Kirsch *et al.* 2008; Fournier *et al.* 2010), higher neuroticism (Tang *et al.* 2009) and who were employed (Fournier *et al.* 2009; Jakubovski & Bloch 2014) had better outcomes to SSRI than placebo. In addition, relatively older patients and those with lower deficits in cognitive control (i.e., smaller Flanker accuracy interference effect) also exhibited better outcomes to SSRI. Of note, owing to their minimal cost and relatively low time burden, these baseline measurements could be more easily integrated into a treatment clinic than baseline assessments involving neuroimaging.

Perhaps the most well-supported clinical moderator of SSRI vs. placebo response is baseline depressive symptom severity (Khan *et al*. 2002; Kirsch *et al*. 2008; Fournier *et al*. 2010). It should be noted that total depression score at baseline is not the only meaningful marker of depression severity. Other relevant variables such as episode chronicity and anhedonia were included in our initial models but did not survive the variable selection steps. Chronicity is known to be linked with poor response to placebo (Khan *et al*. 1991; Dunner 2001), yet did not emerge as a moderator of SSRI vs. placebo response. Consistent with prior work, higher neuroticism was associated with greater response to SSRI relative to placebo, which may in part be due to the role of SSRIs in blunting negative affect (Quilty *et al*. 2008; Tang *et al*. 2009; Soskin *et al*. 2012). It is important to highlight that elevated neuroticism moderated SSRI vs. placebo response above and beyond the contribution of baseline depression (i.e., while the baseline HRSD × treatment group interaction was included in the model).

The interpretation of the cognitive control finding is less clear. Namely, those with more intact cognitive control exhibited better outcomes in SSRI vs. placebo; whereas those with greater impairments showed little between-group differences in outcome. Continued cognitive impairments – even following symptom remission – are among the most common residual symptoms of depression (Herrera-Guzmán *et al*. 2009; Lam *et al*. 2014). Moderation may be more likely to be observed when comparing a treatment that more successfully targets cognitive control deficits (e.g., vortioxetine, (Mahableshwarkar *et al*. 2015)) vs. one with limited pro-cognitive effects (also see Etkin *et al*. 2015).

Of the 12 a priori variables we initially included, 7 did not survive our four-step variable selection procedure. It may be that some of these variables are prognostic predictors of outcome, but were not selected as they do not moderate SSRI vs. placebo response. For example, higher anhedonia (McMakin *et al*. 2012; Uher *et al*. 2012a) and blunted reward learning (Vrieze *et al*. 2013) have each been shown to predict worse antidepressant outcome. Although anhedonia did not moderate of SSRI vs. placebo response it did emerge as a prognostic predictor of worse outcome across groups ($t = 3.51$, $p < .001$; reward learning *ns*; see Supplemental Results). With regards to the specific variable selection approaches used, both Random Forests (RF) and Bayesian Additive Regression Trees (BART) identified the same 5 variables; whereas Elastic Net Regularization (ENR) selected a larger set of 8 variables. Differences in results between these approaches are not unexpected, and may be due to the fact that both RF and BART rely on a similar decision-tree based ensemble learning algorithm, whereas ENR is a variant of classic regression. As well, unlike ENR, both RF and BART consider both unspecified non-linear relationships and higher-order interactions between variables.

Importantly, there were no overall differences in depression outcomes between outpatients randomized to SSRI and placebo in the overall sample ($d = 0.15$). These findings are in line with meta-analyses of SSRI vs. placebo indicating small overall differences in outcome (Moncrieff *et al.* 2004; Kirsch *et al.* 2008; Fournier *et al.* 2010; Kirsch 2015; Jakobsen *et al.* 2017; Cipriani *et al.* 2018). However, overall between group comparisons obscure any meaningful between-patient characteristics that may moderate SSRI vs. placebo differences in outcome. Indeed, we identified five patient characteristics that moderated group differences in depression outcome. These variables were subsequently entered into a PAI algorithm (DeRubeis *et al.* 2014; Huibers *et al.* 2015) to generate patient-specific predictions of SSRI vs. placebo outcome. Results using our PAI model indicated that approximately one-third of the sample would have a clinically significant advantage (DeRubeis *et al.* 2014) with SSRI relative to placebo (PAI ≤ −3). Intriguingly, and unexpectedly, the model also predicted that a subset (17%) of depressed individuals would exhibit a clinically significant advantage in placebo.

As the treatment recommendations for some individuals indicated almost no advantage of one treatment over the other (e.g., see distribution of PAI scores near 0 in Figure 1), one might reasonably expect that differences in outcome between patients who received their PAI-indicated vs. contraindicated treatment would be larger for those individuals predicted to have more clinically meaningful differences in outcomes (i.e., larger absolute PAI values), which our sub-analyses confirmed. Notably, when considering the subset with larger PAIs (absolute PAI values ≥ 3), the effect size for the difference in outcome for SSRI-indicated patients who were randomized to SSRI vs. placebo (d = .58) was substantially larger than the overall treatment group difference between SSRI and placebo (d = .15), as well as larger than the effect sizes reported in systematic reviews of ADM vs. placebo comparisons (d ~ .30)(Cipriani *et al.* 2018; Fournier *et al.* 2010; Kirsch *et al.* 2008; Kirsch 2015; Turner *et al.* 2008; Khin *et al.* 2011; Moncrieff & Kirsch 2015), and those observed between active treatments and controls from general medical contexts (d ~ .45)(Leucht *et al.* 2012). In sum, findings suggest that our statistical approach may identify patients who are optimally suited to SSRI treatment. Of course, this study compared SSRI vs. a placebo condition, rather than an alternative evidence-based treatment (e.g., CBT). Thus, our model identified individuals who would likely evidence greater depressive symptom improvement on an SSRI relative to an intervention providing the “non-specific” therapeutic elements associated with a pill placebo condition (i.e., the expectation of symptom improvement, the passage of time, symptom monitoring and minimal contact/support from a clinician).

Although no statistically significant advantage was observed for placebo-indicated patients who received their indicated treatment, a significant advantage of placebo over SSRI was observed for the 17% of the sample for whom placebo was more strongly indicated (PAIs ≥ 3; d = 1.01). The possibility that SSRIs are relatively ineffective or countertherapeutic for certain patients (e.g., due to side effects) requires additional research (Bet *et al.* 2013; Julien 2013; Hollon 2016). It is important to emphasize that this finding did not emerge in the full sample. Given the reduced sample size in the latter analysis, conclusions must be tempered and replications are required.

An alternative PAI model based exclusively on pre-treatment HRSD scores did not yield significant findings, suggesting that baseline depressive symptom severity alone is not as informative as our model incorporating baseline data on five variables. Second, a similar pattern of findings emerged in a control PAI analysis (in which all 12 a priori variables were included) relative to our primary analysis, suggesting that our findings are likely not attributable to overfitting due to running our PAI analysis on a reduced set of variables emerging from our variable selection steps.

Limitations

Several limitations should be noted. First, and importantly, prospective tests are needed in which a PAI model is built in one sample, and then tested in a separate sample. The k-fold cross-validation approach we used approximates such a test by leaving each patient's data out of the model used to generate their predicted outcomes. However, although we implemented cross-validation during the weight-setting stage, we used the full sample for variable selection which can lead to overfitting and inflated associations (Hastie *et al.* 2009; Fiedler 2011). Until such models are tested and replicated in separate samples it will be difficult to determine the extent to which overfitting contributes to findings and whether models generalize to new sets of treatment-seeking depressed individuals. Second, we focused on clinical, demographic and putative behavioral endophenotypes that could be collected at low cost and with relatively minimal clinic staff and patient burden. The extent to which neural assessments provide incremental predictive validity above and beyond such variables is an important direction for research, particularly with regards to relatively less costly and non-invasive imaging approaches (e.g., EEG). Third, it is unclear whether findings would generalize to depressed individuals who do not meet the inclusion/exclusion criteria of this trial. In addition, as others have highlighted (Uher *et al.* 2012b), measures of outcome (HRSD) and predictors include a certain amount of error, which may significantly attenuate the magnitude of observed predictor-outcome associations. Fourth, sample size was relatively small. Finally, the current PAI model relies on randomized designs (i.e., to examine outcomes for those randomly assigned to their indicated vs. non-indicated treatment). An important future direction for research is to adapt these statistical models for the investigation of optimal treatment assignment in current clinical practice settings in which patients are not randomly assigned to interventions. These limitations notwithstanding, our findings demonstrate the potential for precision medicine to improve individual outcomes through model-guided treatment recommendations rather than the current practice of trial-and-error. Findings from replicated prescriptive algorithms could ultimately be used to inform the development of web-based "treatment selection calculators" available to clinicians and patients to facilitate decision-making.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

This study was supported by NIMH grants (M.H.K., grant number U01MH092221; P.J.M, RVP, MMW, grant number U01MH092250).

Financial Support

The EMBARC study was supported by the National Institute of Mental Health of the National Institutes of Health under award numbers 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. This work was supported by the EMBARC National Coordinating Center at UT Southwestern Medical Center, Madhukar H. Trivedi, M.D., Coordinating PI, and the Data Center at Columbia and Stony Brook Universities. Christian A. Webb and Diego A. Pizzagalli were partially supported by 5K23MH108752 and 2R37MH068376, respectively Zachary D. Cohen and Robert J. DeRubeis are supported in part by a grant from MQ: Transforming mental health MQ14PM_27. The opinions and assertions contained in this article should not be construed as reflecting the views of the sponsors.

References

American Psychiatric Association (2010). Treatment of patients with major depressive disorder. 3rd ed American Psychiatric Press: Washington, DC.

Austin PC, Tu JV (2004). Bootstrap Methods for Developing Predictive Models. The American Statistician 58, 131–137.

Bagby RM, Quilty LC, Segal ZV, McBride CC, Kennedy SH, Costa PT (2008). Personality and Differential Treatment Response in Major Depression: A Randomized Controlled Trial Comparing Cognitive-Behavioural Therapy and Pharmacotherapy. The Canadian Journal of Psychiatry 53, 361–370. [PubMed: 18616856]

Baldessarini RJ, Forte A, Selle V, Sim K, Tondo L, Undurraga J, Vázquez GH (2017). Morbidity in Depressive Disorders. Psychotherapy and Psychosomatics 86, 65–72. [PubMed: 28183075]

Ball S, Classi P, Dennehy EB (2014). What happens next?: a claims database study of second-line pharmacotherapy in patients with major depressive disorder (MDD) who initiate selective serotonin reuptake inhibitor (SSRI) treatment. Annals of General Psychiatry 13, 8. [PubMed: 24645830]

Bet PM, Hugtenburg JG, Penninx BWJH, Hoogendijk WJG (2013). Side effects of antidepressants during long-term use in a naturalistic setting. European Neuropsychopharmacology 23, 1443–1451. [PubMed: 23726508]

Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, Leucht S, Ruhe HG, Turner EH, Higgins JPT, Egger M, Takeshima N, Hayasaka Y, Imai H, Shinohara K, Tajika A, Ioannidis JPA, Geddes JR (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. The Lancet 0

Cohen ZD, DeRubeis RJ (2018). Treatment Selection in Depression. Annual Review of Clinical Psychology 14, 209–236.

Cohen ZD, Kim T, Van H, Dekker J, Driessen E (2017). Recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. PsyArXiv, https://osf.io/6qxve/.

DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L (2014). The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration. PLOS ONE 9, e83875. [PubMed: 24416178]

Dunner DL (2001). Acute and maintenance treatment of chronic depression. The Journal of clinical psychiatry 62 Suppl 6, 10–16.

Eriksen BA, Eriksen CW (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception & Psychophysics 16, 143–149.

Etkin A, Patenaude B, Song YJC, Usherwood T, Rekshan W, Schatzberg AF, Rush AJ, Williams LM (2015). A Cognitive–Emotional Biomarker for Predicting Remission with Antidepressant Medications: A Report from the iSPOT-D Trial. Neuropsychopharmacology 40, 1332–1342. [PubMed: 25547711]

Fava M, Rush AJ, Alpert JE, Balasubramani GK, Wisniewski SR, Carmin CN, Biggs MM, Zisook S, Leuchter A, Howland R, Warden D, Trivedi MH (2008). Difference in Treatment Outcome in Outpatients With Anxious Versus Nonanxious Depression: A STAR*D Report. American Journal of Psychiatry 165, 342–351. [PubMed: 18172020]

Fiedler K (2011). Voodoo Correlations Are Everywhere—Not Only in Neuroscience. Perspectives on Psychological Science 6, 163–171. [PubMed: 26162135]

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, Fawcett J (2010). Antidepressant Drug Effects and Depression Severity: A Patient-Level Meta-analysis. JAMA 303, 47–53. [PubMed: 20051569]

Fournier JC, DeRubeis RJ, Hollon SD, Gallop R, Shelton RC, Amsterdam JD (2013). Differential Change in Specific Depressive Symptoms during Antidepressant Medication or Cognitive Therapy. Behaviour research and therapy 51, 392–398. [PubMed: 23644038]

Fournier JC, DeRubeis RJ, Shelton RC, Hollon SD, Amsterdam JD, Gallop R (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. Journal of Consulting and Clinical Psychology 77, 775–787. [PubMed: 19634969]

Fried EI, Nesse RM (2015a). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. Journal of affective disorders 172, 96–102. [PubMed: 25451401]

Fried EI, Nesse RM (2015b). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Medicine 13, 72. [PubMed: 25879936]

Friedman J, Hastie T, Tibshirani R (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software 33, 1–22. [PubMed: 20808728]

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

Gillan CM, Whelan R (2017). What big data can do for treatment in psychiatry. Current Opinion in Behavioral Sciences 18, 34–42.

Goldstein BL, Klein DN (2014). A review of selected candidate endophenotypes for depression. Clinical Psychology Review 34, 417–427. [PubMed: 25006008]

Gottesman II, Gould TD (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. The American Journal of Psychiatry 160, 636–645. [PubMed: 12668349]

Hamilton M (1960). A Rating Scale for Depression. Journal of Neurology, Neurosurgery & Psychiatry 23, 56–62.

Hastie T, Tibshirani R, Friedman J (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. 2nd edition Springer: New York, NY.

Herrera-Guzmán I, Gudayol-Ferré E, Herrera-Guzmán D, Guàrdia-Olmos J, Hinojosa-Calvo E, Herrera-Abarca JE (2009). Effects of selective serotonin reuptake and dual serotonergic-noradrenergic reuptake treatments on memory and mental processing speed in patients with major depressive disorder. Journal of Psychiatric Research 43, 855–863. [PubMed: 19128810]

Hollon SD (2016). The efficacy and acceptability of psychological interventions for depression: where we are now and where we are going. Epidemiology and Psychiatric Sciences 25, 295–300. [PubMed: 26310321]

Holmes AJ, Bogdan R, Pizzagalli DA (2010). Serotonin Transporter Genotype and Action Monitoring Dysfunction: A Possible Substrate Underlying Increased Vulnerability to Depression. Neuropsychopharmacology 35, 1186–1197. [PubMed: 20090673]

Huibers MJH, Cohen ZD, Lemmens LHJM, Arntz A, Peeters FPML, Cuijpers P, DeRubeis RJ (2015). Predicting Optimal Outcomes in Cognitive Therapy or Interpersonal Psychotherapy for Depressed Individuals Using the Personalized Advantage Index Approach. PLOS ONE 10, e0140771. [PubMed: 26554707]

Jakobsen JC, Katakam KK, Schou A, Hellmuth SG, Stallknecht SE, Leth-Møller K, Iversen M, Banke MB, Petersen IJ, Klingenberg SL, Krogh J, Ebert SE, Timm A, Lindschou J, Gluud C (2017). Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. BMC Psychiatry 17, 58. [PubMed: 28178949]

Jakubovski E, Bloch MH (2014). Prognostic Subgroups for Citalopram Response in the STAR*D Trial. The Journal of clinical psychiatry 75, 738–747. [PubMed: 24912106]

Julien RM (2013). A Primer of Drug Action: A Concise Nontechnical Guide to the Actions, Uses, and Side Effects of Psychoactive Drugs, Revised and Updated. Henry Holt and Company.

Kapelner A, Bleich J (2016). bartMachine: Machine learning with Bayesian Additive Regression Trees. Journal of Statistical Software 70

Khan A, Dager SR, Cohen S, Avery DH, Scherzo B, Dunner DL (1991). Chronicity of depressive episode in relation to antidepressant-placebo response. Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology 4, 125–130. [PubMed: 2025378]

Khan A, Leventhal RM, Khan SR, Brown WA (2002). Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. Journal of Clinical Psychopharmacology 22, 40–45. [PubMed: 11799341]

Khin NA, Chen Y-F, Yang Y, Yang P, Laughren TP (2011). Exploratory Analyses of Efficacy Data From Major Depressive Disorder Trials Submitted to the US Food and Drug Administration in Support of New Drug Applications. The Journal of Clinical Psychiatry 72, 464–472. [PubMed: 21527123]

Kirsch I (2015). Clinical trial methodology and drug-placebo differences. World Psychiatry 14, 301–302. [PubMed: 26407779]

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008). Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration. PLOS Medicine 5, e45. [PubMed: 18303940]

Kraemer HC (2013). Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. Statistics in Medicine 32, 1964–1973. [PubMed: 23303653]

Kraemer HC, Blasey CM (2004). Centring in regression analyses: a strategy to prevent errors in statistical inference. International Journal of Methods in Psychiatric Research 13, 141–151. [PubMed: 15297898]

Lam RW, Kennedy SH, McIntyre RS, Khullar A (2014). Cognitive Dysfunction in Major Depressive Disorder: Effects on Psychosocial Functioning and Implications for Treatment. The Canadian Journal of Psychiatry 59, 649–654. [PubMed: 25702365]

Leucht S, Hierl S, Kissling W, Dold M, Davis JM (2012). Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. The British Journal of Psychiatry 200, 97–106. [PubMed: 22297588]

Mahableshwarkar AR, Zajecka J, Jacobson W, Chen Y, Keefe RS (2015). A Randomized, Placebo-Controlled, Active-Reference, Double-Blind, Flexible-Dose Study of the Efficacy of Vortioxetine on Cognitive Function in Major Depressive Disorder. Neuropsychopharmacology

Marcus SC, Hassan M, Olfson M (2009). Antidepressant Switching Among Adherent Patients Treated for Depression. Psychiatric Services 60, 617–623. [PubMed: 19411348]

Mars B, Heron J, Gunnell D, Martin RM, Thomas KH, Kessler D (2017). Prevalence and patterns of antidepressant switching amongst primary care patients in the UK. Journal of Psychopharmacology 31, 553–560. [PubMed: 28460603]

McCrae RR, Costa PT (2010). NEO Inventories professional manual. Lutz, FL: Psychological Assessment Resources

McGrath CL, Kelley ME, Holtzheimer PE, et al. (2013). Toward a neuroimaging treatment selection biomarker for major depressive disorder. JAMA Psychiatry 70, 821–829. [PubMed: 23760393]

McMakin DL, Olino TM, Porta G, Dietz LJ, Emslie G, Clarke G, Wagner KD, Asarnow JR, Ryan ND, Birmaher B, Shamseddeen W, Mayes T, Kennard B, Spirito A, Keller M, Lynch FL, Dickerson JF, Brent DA (2012). Anhedonia Predicts Poorer Recovery among Youth with Selective Serotonin Reuptake Inhibitor-Treatment Resistant Depression. Journal of the American Academy of Child and Adolescent Psychiatry 51, 404–411. [PubMed: 22449646]

Milea D, Guelfucci F, Bent-Ennakhil N, Toumi M, Auray J-P (2010). Antidepressant monotherapy: A claims database analysis of treatment changes and treatment duration. Clinical Therapeutics 32, 2057–2072. [PubMed: 21118742]

Moncrieff J, Kirsch I (2015). Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. Contemporary Clinical Trials 43, 60–62. [PubMed: 25979317]

Moncrieff J, Wessely S, Hardy R (2004). Active placebos versus antidepressants for depression. The Cochrane Library

NICE (2018). Depression in adults: recognition and management | Guidance and guidelines | NICE

Pizzagalli DA, Evins AE, Schetter EC, Frank MJ, Pajtas PE, Santesso DL, Culhane M (2008a). Single dose of a dopamine agonist impairs reinforcement learning in humans: Behavioral evidence from a laboratory-based measure of reward responsiveness. Psychopharmacology 196, 221–232. [PubMed: 17909750]

Pizzagalli DA, Goetz E, Ostacher M, Iosifescu DV, Perlis RH (2008b). Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task. Biological Psychiatry 64, 162–168. [PubMed: 18242583]

Pizzagalli DA, Jahn AL, O'Shea JP (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. Biological Psychiatry 57, 319–327. [PubMed: 15705346]

Pizzagalli DA, Webb CA, Dillon DG, Tenke CE, Kayser J, Goer F, Fava M, McGrath P, Weissman M, Parsey R, Adams P, Trombello J, Cooper C, Deldin P, Oquendo MA, McInnis MG, Carmody T, Bruder G, Trivedi MH (2018). Pretreatment Rostral Anterior Cingulate Cortex Theta Activity in Relation to Symptom Improvement in Depression: A Randomized Clinical Trial. JAMA Psychiatry

Quilty LC, Meusel L-AC, Bagby RM (2008). Neuroticism as a mediator of treatment response to SSRIs in major depressive disorder. Journal of Affective Disorders 111, 67–73. [PubMed: 18384882]

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biological Psychiatry 54, 573–583. [PubMed: 12946886]

Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackeim HA, Kupfer DJ, Luther J, Fava M (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. The American Journal of Psychiatry 163, 1905–1917. [PubMed: 17074942]

Saragoussi D, Chollet J, Bineau S, Chalem Y, Milea D (2012). Antidepressant switching patterns in the treatment of major depressive disorder: a General Practice Research Database (GPRD) Study. International Journal of Clinical Practice 66, 1079–1087. [PubMed: 23067031]

Schultz J, Joish V (2009). Costs Associated With Changes in Antidepressant Treatment in a Managed Care Population With Major Depressive Disorder. Psychiatric Services 60, 1604–1611. [PubMed: 19952150]

Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P (1995). A scale for the assessment of hedonic tone: The Snaith-Hamilton Pleasure Scale. British Journal of Psychiatry 167, 99–103. [PubMed: 7551619]

Soskin DP, Carl JR, Alpert J, Fava M (2012). Antidepressant Effects on Emotional Temperament: Toward a Biobehavioral Research Paradigm for Major Depressive Disorder. CNS Neuroscience & Therapeutics 18, 441–451. [PubMed: 22672296]

Souery D, Oswald P, Massat I, Bailer U, Bollen J, Demyttenaere K, Kasper S, Lecrubier Y, Montgomery S, Serretti A, Zohar J, Mendlewicz J, Group for the Study of Resistant Depression (2007). Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. The Journal of clinical psychiatry 68, 1062–70. [PubMed: 17685743]

Stekhoven DJ, Bühlmann P (2012). MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 28, 112–118. [PubMed: 22039212]

Tang TZ, DeRubeis RJ, Hollon SD, Amsterdam J, Shelton R, Schalet B (2009). Personality Change During Depression Treatment: A Placebo-Controlled Trial. Archives of General Psychiatry 66, 1322–1330. [PubMed: 19996037]

Thomas L, Kessler D, Campbell J, Morrison J, Peters TJ, Williams C, Lewis G, Wiles N (2013). Prevalence of treatment-resistant depression in primary care: cross-sectional data. The British Journal of General Practice 63, e852–e858. [PubMed: 24351501]

Trivedi MH, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML, Oquendo MA, Bruder G, Pizzagalli D, Toups M, Cooper C, Adams P, Weyandt S, Morris DW, Grannemann BD, Ogden RT, Buckner R, McInnis M, Kraemer HC, Petkova E, Carmody TJ, Weissman MM (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. Journal of Psychiatric Research 78, 11–23. [PubMed: 27038550]

Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ, Shores-Wilson K, Biggs MM, Balasubramani GK, Fava M (2006). Evaluation of Outcomes With Citalopram for Depression Using Measurement-Based Care in STAR*D: Implications for Clinical Practice. American Journal of Psychiatry 163, 28–40. [PubMed: 16390886]

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008). Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. New England Journal of Medicine 358, 252–260. [PubMed: 18199864]

Uher R, Perlis RH, Henigsberg N, Zobel A, Rietschel M, Mors O, Hauser J, Dernovsek MZ, Souery D, Bajs M, Maier W, Aitchison KJ, Farmer A, McGuffin P (2012a). Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. Psychological medicine 42, 967–980. [PubMed: 21929846]

Uher R, Tansey KE, Malki K, Perlis RH (2012b). Biomarkers predicting treatment outcome in depression: what is clinically significant? Pharmacogenomics 13, 233–240. [PubMed: 22256872]

Vrieze E, Pizzagalli DA, Demyttenaere K, Hompes T, Sienaert P, de Boer P, Schmidt M, Claes S (2013). Reduced Reward Learning Predicts Outcome in Major Depressive Disorder. Biological Psychiatry 73, 639–645. [PubMed: 23228328]

Vuorilehto MS, Melartin TK, Isometsä ET (2009). Course and outcome of depressive disorders in primary care: a prospective 18-month study. Psychological Medicine 39, 1697–1707. [PubMed: 19250580]

Wakefield JC, Schmitz MF (2013). When does depression become a disorder? Using recurrence rates to evaluate the validity of proposed changes in major depression diagnostic thresholds. World Psychiatry 12, 44–52. [PubMed: 23471801]

Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, Marrero J, Zhu J, Higgins PD (2013). Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 3, e002847.

Watson D, Clark LA, Weber K, Assenheimer JS, Strauss ME, McCormick RA (1995). Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. Journal of Abnormal Psychology 104, 15–25. [PubMed: 7897037]

Webb CA, Dillon DG, Pechtel P, Goer FK, Murray L, Huys QJ, Fava M, McGrath PJ, Weissman M, Parsey R, Kurian BT, Adams P, Weyandt S, Trombello JM, Grannemann B, Cooper CM, Deldin P, Tenke C, Trivedi M, Bruder G, Pizzagalli DA (2016). Neural Correlates of Three Promising Endophenotypes of Depression: Evidence from the EMBARC Study. Neuropsychopharmacology 41, 454–463. [PubMed: 26068725]

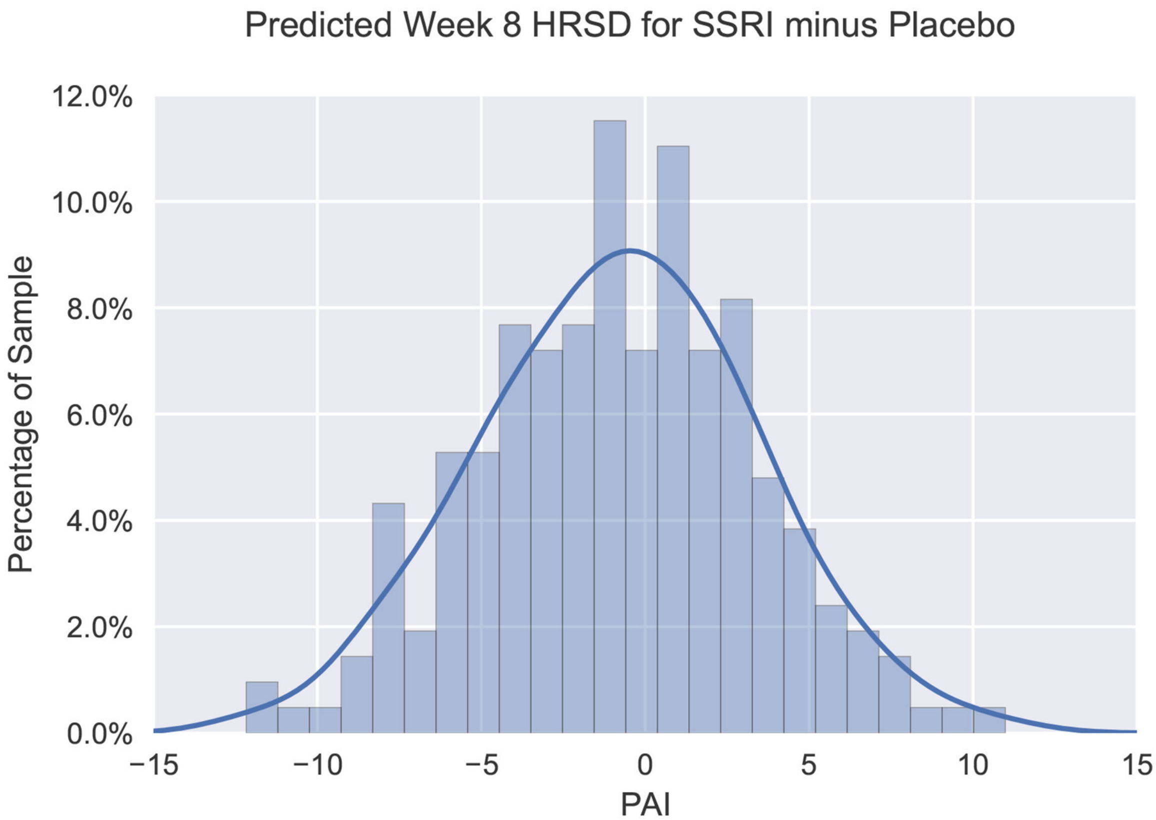

Figure 1.

Frequency histogram displaying distribution of Personalized Advantage Index (PAI) scores, computed as the predicted difference in week 8 HRSD scores for SSRI minus placebo. Accordingly, a PAI score less than 0 signifies that SSRI was indicated, whereas a PAI score greater than 0 indicates that placebo was expected to yield a better outcome. The kernel density estimate illustrates the expected distribution of PAI scores in the population.

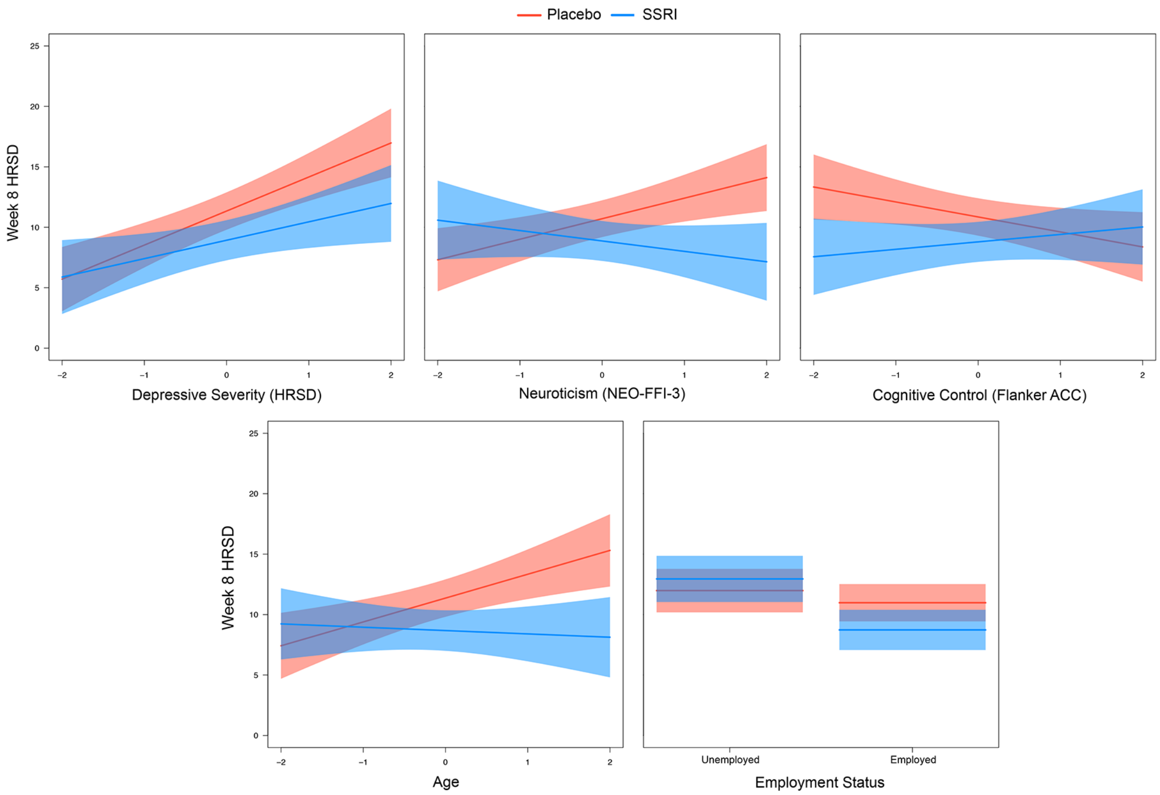

Figure 2.

Plots of baseline predictor by treatment group interactions from the final model.

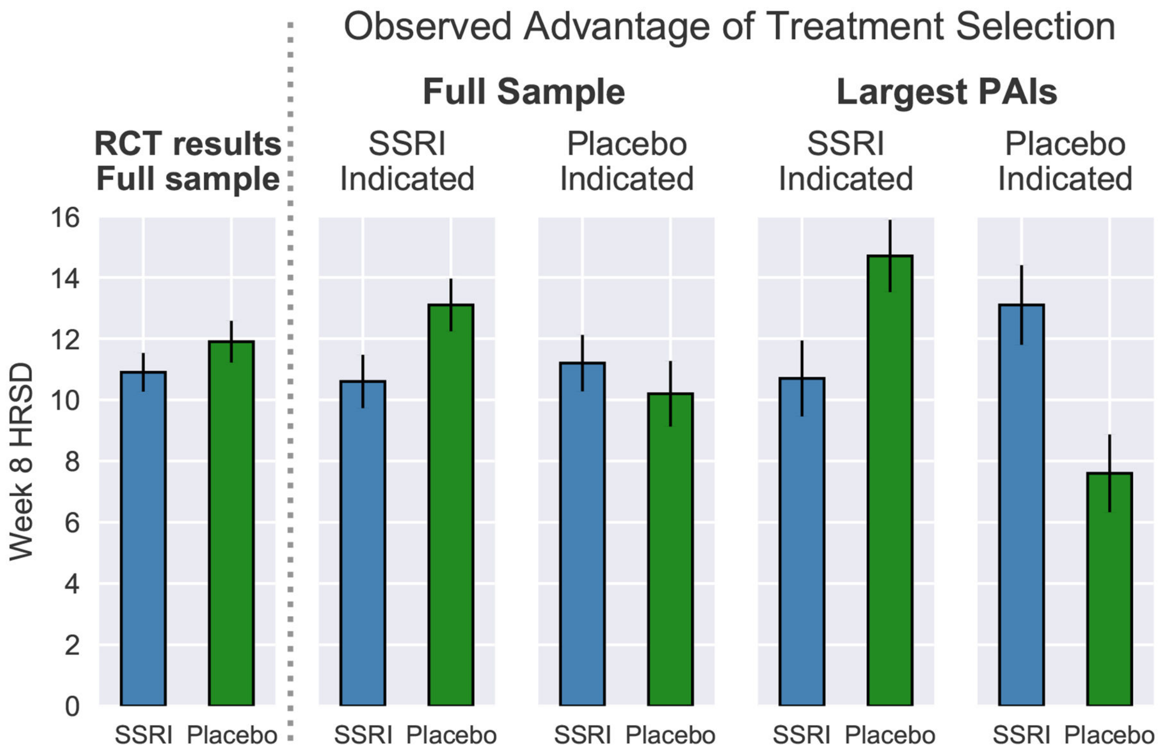

Figure 3.

Comparison of mean week 8 HRSD for patients randomized to SSRI or placebo (left panel) (n=216). Comparison of mean week 8 HRSD scores for patients randomly assigned to their PAI-indicated treatment vs. those assigned to their PAI-contraindicated treatment for the full sample (n = 216) vs. including only patients for whom the algorithm predicted a clinically significant advantage in one treatment condition (PAI ≥ |3|); n = 105) **(**right panel). Error bars represent standard error.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Variable Selection Results

Pre-Treatment Variable	Random Forest	Elastic Net	BART	Included in Bootstep AIC?
Depression Severity (HDRS)[a]	Yes	Yes	Yes	Yes
Anxiety Severity (MASQ-AA)	No	Yes	No	No
Chronic MDD (yes/no)	No	Yes	No	No
Neuroticism (NEO-FFI-3)[a]	Yes	Yes	Yes	Yes
Anhedonia (SHAPS)	No	No	No	No
Reward Learning (PRT)	No	No	No	No
Cognitive Control (Flanker ACC)[a]	Yes	Yes	Yes	Yes
Gender	No	No	No	No
Age[a]	Yes	Yes	Yes	Yes
Marital Status	No	No	No	No
Employment Status[a]	Yes	Yes	Yes	Yes
Years of Education	No	Yes	No	No

Note. HDRS: Hamilton Depression Rating Scale (17-item)(Hamilton 1960); MASQ-AA: Mood and Anxiety Symptoms Questionnaire, Anxious Arousal subscore (Watson *et al.* 1995), MDD: Major Depressive Disorder; NEO-FFI-3: NEO Five-Factor Inventory – 3 (McCrae & Costa 2010); SHAPS: Snaith-Hamilton Pleasure Scale (Snaith *et al.* 1995); PRT: Probabilistic Reward Task (Pizzagalli *et al.* 2005); Flanker ACC: Flanker Interference Accuracy score (= $\text{Accuracy}_{\text{Compatible trials}} - \text{Accuracy}_{\text{Incompatible trials}}$); Higher scores indicate more interference (i.e., reduced cognitive control).

[a]Variables selected by BootStepAIC to be included in the final model.

Table 2.

Final Model

Variable	*B*	*SE*	*p* value
(Intercept)	11.51	0.43	0.00 **
Treatment	−0.65	0.85	0.44
Depression Severity (HDRS)	2.17	0.44	0.00 **
Neuroticism (NEO-FFI-3)	0.42	0.45	0.35
Cognitive Control (Flanker ACC)	−0.31	0.45	0.49
Age	0.85	0.45	0.06
Employment Status	−2.61	0.87	0.00 **
Treatment × Depression Severity (HDRS)	−1.29	0.88	0.14
Treatment × Neuroticism (NEO-FFI-3)	−2.56	0.90	0.01 **
Treatment × Cognitive Control (Flanker ACC)	1.86	0.89	0.04 *
Treatment × Age	−2.25	0.91	0.01 *
Treatment × Employment Status	−3.21	1.74	0.07

Note. HDRS: Hamilton Depression Rating Scale (17-item)(Hamilton 1960); NEO-FFI-3: NEO Five-Factor Inventory – 3 (McCrae & Costa 2010); Flanker ACC: Flanker Interference Accuracy score (= $\text{Accuracy}_{\text{Compatible trials}} - \text{Accuracy}_{\text{Incompatible trials}}$).

$^{+}p < 0.10$.

$^{*}p < 0.05$.

$^{**}p < 0.01$.