

Toward an online cognitive and emotional battery to predict treatment remission in depression

Evian Gordon¹
A John Rush²
Donna M Palmer^{3,4}
Taylor A Braund³
William Rekshan¹

¹Brain Resource, San Francisco, CA, USA; ²Duke-NUS, Singapore; ³Brain Resource, Sydney, NSW, Australia; ⁴Brain Dynamics Center, Sydney Medical School – Westmead and Westmead Millennium Institute, The University of Sydney, Sydney, NSW, Australia

Purpose: To evaluate the performance of a cognitive and emotional test battery in a representative sample of depressed outpatients to inform likelihood of remission over 8 weeks of treatment with each of three common antidepressant medications.

Patients and methods: Outpatients 18–65 years old with nonpsychotic major depressive disorder (17 sites) were randomized to escitalopram, sertraline or venlafaxine-XR (extended release). Participants scored ≥ 12 on the baseline 16-item Quick Inventory of Depressive Symptomatology – Self-Report and completed 8 weeks of treatment. The baseline test battery measured cognitive and emotional status. Exploratory multivariate logistic regression models predicting remission (16-item Quick Inventory of Depressive Symptomatology – Self-Report score ≤ 5 at 8 weeks) were developed independently for each medication in subgroups stratified by age, sex, or cognitive and emotional test performance. The model with the highest cross-validated accuracy determined the participant proportion in each arm for whom remission could be predicted with an accuracy $\geq 10\%$ above chance. The proportion for whom a prediction could be made with very high certainty (positive predictive value and negative predictive value exceeding 80%) was calculated by incrementally increasing test battery thresholds to predict remission/non-remission.

Results: The test battery, individually developed for each medication, improved identification of remitting and non-remitting participants by $\geq 10\%$ beyond chance for 243 of 467 participants. The overall remission rates were escitalopram: 40.8%, sertraline: 30.3%, and venlafaxine-XR: 31.1%. Within this subset for whom prediction exceeded chance, test battery thresholds established a negative predictive value of $\geq 80\%$, which identified 40.9% of participants not remitting on escitalopram, 77.1% of participants not remitting on sertraline, and 38.7% of participants not remitting on venlafaxine-XR (all including 20% false negatives).

Conclusion: The test battery identified about 50% of each medication group as being $\geq 10\%$ more or less likely to remit than by chance, and identified about 38% of individuals who did not remit with $\geq 80\%$ certainty. Clinicians might choose to avoid this specific medication in these particular patients.

Keywords: depression, treatment selection, cognitive tests, biomarkers, treatment prediction, antidepressant medication

Introduction

Major depressive disorder (MDD) is a disabling, life shortening, and costly disorder that affects over 15% of the population.¹ In terms of burden of disease, depression is the third most disabling medical condition worldwide, and is projected to be the first in developed nations by 2030.^{2,3}

Antidepressant medications are commonly used to treat depression. Remission is the goal of treatment,⁴ yet less than 50% of patients achieve remission with their initial treatment.^{5–7} Presently, selecting the best treatment for depressed patients rests on a trial

Correspondence: Evian Gordon
Brain Resource, 115 Sansome St, Suite
600-B, San Francisco, CA 94104, USA
Tel +1 415 399 7990
Fax +1 415 852 5198
Email eviang@brainresource.us

and error approach, with no clinically useful guidance as to which treatment is preferred for any individual patient.⁸ Overall, patients with more severe depression,⁹ more concurrent general medical conditions,¹⁰ or more concurrent psychiatric disorders,^{11,12} are more likely to not achieve remission during acute phase treatment. These features can differentiate between groups of patients who do and do not remit, hereby allowing risk stratification. However, these features are not sufficiently specific to allow the identification of individual patients who will or will not remit with enough certainty that clinical decisions can be made.^{13,14}

If one could better predict whether a particular treatment is quite likely or quite unlikely to work for a particular patient, substantial savings in time, effort, cost, and patient suffering would follow.^{15,16} While several reports have begun to suggest that pretreatment tests could inform the selection of treatment,^{17,18} none have yet entered clinical care.^{19–21} Cognitive and emotional tests are particularly well suited for use in the selection of treatment as these are scalable and guidance could be provided instantly at the point of care.

This report evaluated the performance of a well established²² and validated cognitive and emotional test battery^{23–25} in a representative group of depressed outpatients as part of the International Study to Predict Optimized Treatment in Depression (iSPOT-D) trial²⁶ to address several questions. The primary question was: can this test battery identify a meaningful proportion of these patients that exceeds chance prediction (defined as a 50–50 likelihood of remission) by at least 10% in depressed patients who are treated for 8 weeks with escitalopram, sertraline or venlafaxine-XR (extended release). To further explore the performance of the test battery, our secondary question was whether the test battery could identify a meaningful proportion of patients who, with substantial certainty (at least 80%), will either achieve or not achieve remission with 8 weeks of escitalopram, sertraline or venlafaxine-XR? This report serves to put forth preliminary hypotheses that are to be evaluated in an independent sample of depressed patients.²⁶

Methods

Overview

Details of the iSPOT-D study design and procedures have been reported elsewhere.²⁶ In short, in this multiple-phase, multi-site trial, outpatients with nonpsychotic MDD were randomly assigned to one of three antidepressants: escitalopram, sertraline or venlafaxine-XR. Participants in this report were recruited from 17 sites (in five countries) consisting of eight academic sites (four with outpatient psychiatry services,

three with outpatient psychology services, and one with combined psychiatry and primary care services) and nine private sites (four psychology/primary care, four psychology, and one psychiatry services).

This report entails a preliminary evaluation of a pre-treatment test battery in participants randomized to 8 weeks of treatment with one of these three medications to determine whether the test battery will identify individual depressed patients in each treatment group who will or will not remit in this time period.

Study participants

The study was approved by Institutional or Ethic Review Boards. Compliance with International Conference on Harmonization and Good Clinical Practice principles, the US Food and Drug Administration Code of Federal Regulations, and country-specific guidelines were implemented. Participants (N=1,008), all of whom provided written informed consent, were adults (18–65 years old) with a current diagnosis of single-episode or recurrent, nonpsychotic, MDD. Table S1 further details the inclusion and exclusion criteria. The Mini-International Neuropsychiatric Interview was used to establish *Diagnostic and Statistical Manual of Mental Disorders* (DSM)-IV diagnoses. The full iSPOT study²⁵ required a 17-item Hamilton Rating Scale for Depression (HRSD₁₇) score ≥ 16 at entry. All enrollees were either antidepressant-naïve or had undergone a washout period of at least five half-lives of a previously prescribed antidepressant. All participants were aware of the medication that they were taking.

Because this report aimed to develop an algorithm that could predict remission after 8 full weeks of antidepressant treatment for a meaningful proportion of depressed patients, and because the primary outcome was remission defined by the 16-item Quick Inventory of Depressive Symptoms – Self-Rated (QIDS-SR₁₆) at 8 weeks, this particular report focused exclusively on participants who completed all 8 weeks and who entered the study with a QIDS-SR₁₆ score ≥ 12 (equivalent to an HRSD₁₇ score ≥ 16 [www.ids-qids.org]) and who had completed at least 50% of the cognitive and emotional test battery. The focus on completers will introduce bias into the analytic sample. However, the alternative, the use of imputation, suffers from unreliability in declaring remission with certainty for each participant. Since this initial report was designed to generate hypotheses that would later be tested in a replication sample, we chose to focus only on the per protocol patients who completed 8 weeks. In this context, current results do not establish the performance of the

algorithm in actual practice. From the initial starting sample of 336 patients in each treatment arm, 157, 175, and 135 per protocol patients who completed the study were analyzed for escitalopram, sertraline, and venlafaxine-XR, respectively.

We also performed the same cognitive and emotional test battery on 336 age-, sex- and education-matched healthy controls at baseline and 8 weeks.²⁵

Test battery

The cognitive assessment battery included eight general cognition tasks and one emotional cognition task (Table 1). These nine tests yield 18 measures of cognitive performance. For further details of the cognitive assessment tasks and construct validity see Gordon et al,²² Paul et al^{23,24} and Williams et al²⁵.

Statistical analyses

We first developed classification models to predict remission for each treatment arm to identify the proportion of depressed outpatients for whom the cognitive and emotional function test battery could provide predictions of remission with performance exceeding chance. Due to the heterogeneity of MDD,²⁷ we assumed that one test would not apply to all MDD patients. Therefore, exploratory multivariate logistic regression models to predict remission were developed in the entire sample and separately for subgroups of participants stratified by age, sex, symptom severity, and nine different metrics of cognitive and emotional performance. The median value of each measure (aside from sex) was used to stratify the participants, resulting in two groups for each measure. Age, sex, baseline depression severity (assessed by QIDS-SR₁₆), three tests from the cognitive and emotion

battery (verbal interference reaction time, number of errors on the maze task, and go/no-go reaction time), and six different summary metrics of the cognitive and emotion test battery were used to stratify the sample, resulting in 12 different stratification metrics. The cognitive and emotion tests and summary metrics were used because each was observed to have a kappa >0.5 in the iSPOT-D healthy control sample (n=336, with 8 weeks between tests). The summary metrics used to stratify the sample were not used as predictors in the multivariate analysis because the multivariate analysis seeks to, by definition, combine many different tests in an optimal way to predict a dependent variable. Overall, we evaluated 12 different methods of stratifying the participants, resulting in a total of 24 different participant subgroups.

Logistic regression methods were chosen because they are simple, more transparent than black box algorithms, and perform well in many applications of binary predictions. Logistic regression has performed well in real world comparisons of data mining methods using neuropsychological testing.²⁸ Due to the exploratory nature of this analysis, the accuracy statistics (accuracy, sensitivity, specificity, positive predictive values [PPV], and negative predictive values [NPV]) were reported for all available data and were cross-validated.

For each analysis, the following steps were followed for model development:

- Step 1: clean and normalize all predictors relative to the iSPOT-D healthy control sample (N=336).²⁶ Outliers, defined as greater or less than 3 standard deviations away from the mean, were replaced with the largest/smallest non-outlier value. Cognition, age, sex, and symptom severity measures were transformed to z-scores relative

Table 1 Cognitive and emotion tasks

Task name	Assesses	Test duration	Measures of task performance
General cognition tasks			
Switching of attention	Information processing speed	3 minutes	Completion time
Verbal interference	Cognitive flexibility	2 minutes	Reaction time – incongruent trials
Continuous performance test	Attention	6 minutes	Reaction time
Go/no-go	Response inhibition	4 minutes	Reaction time
Digit span	Working memory	6 minutes	Total number of trials correct
Motor tapping	Motor coordination	1 minute	Number of taps – dominant hand Number of taps – non-dominant hand
Maze	Executive function	10 minutes	Completion time Number of errors Number of overrun errors
Emotional cognition tasks			
Emotion identification	Emotion identification accuracy Speed of emotion identification	4 minutes	Number of correct responses Reaction time for correct responses

to the healthy control group. All cognitive performance *z*-scores of reaction time or task completion time were corrected for direction (multiplied by -1) so that higher values reflect better (ie, faster) performance.

- Step 2: to avoid co-linearity, identify and remove all highly correlated predictor measures ($r > 0.70$) by removing the predictor with the highest mean absolute correlations with all of the other predictors being evaluated.
- Step 3: to achieve a test with sufficient clinical value while minimizing patient burden, the 3, 4, or 5 predictors with the smallest *P*-values using univariate logistic regression analysis were identified.
- Step 4: perform multivariate logistic regression analysis using the predictor features selected in Step 3, including both main and all possible two-way interaction effects. The model was developed using all available participants who met the entry criteria and criteria for the specific subgroup. A random sample of 80% of the data was used to develop the regression model during cross-validation.

This was cross-validated by repeating this process 300 times, each time selecting a random 80% of the sample allocated to train the model and using the remaining 20% of the data to test the model. The cross-validated regression models were evaluated independently for each treatment arm. All cross-validated accuracy statistics are the average of the performance of the predictions of the test set for the 300 rounds of cross-validation. Of the models that achieved a cross-validated accuracy of 10 percentage points above chance, the final reported models were selected to achieve the maximum cross-validated accuracy. Because cross validation will only provide an assessment of the generalizability of a model for a specific set of data and set of statistical manipulations, a single model must be developed in order to provide a single prediction for individual patients. Therefore, once the regression models and subgroups were identified in the cross validated analyses, new regression models were developed using all available data for the subgroup and treatment arm identified in the cross-validated analyses to allow for both additional calculations regarding performance (see below) and potential application in the real world.

In order to determine the proportion of participants for whom a prediction of remission or non-remission could be made with very high certainty (PPV and NPV exceeding 80%), the predicted probability of remission was obtained from the regression models developed using all available data for the subgroup and treatment arm (not cross-validated). These values range from 0 (no probability of remission) to 1 (maximum probability of remission). The participants

were subset according to predicted probability of remission, incrementally by 0.01 probability of remission. NPV and PPV were calculated for each subset of participants and the largest number of participants for whom a prediction could be made with an NPV/PPV exceeding 80% was used to determine the proportion of participants for whom a prediction could be made with high certainty. This process was conducted separately for each treatment arm and for only participants in the specific subgroup identified in the cross-validated analyses.

All analyses were conducted using R 3.0.2.²⁹ Highly correlated predictors were identified using the caret package of R,³⁰ receiver operating characteristics curves were drawn using the Epi package of R,³¹ and pie charts were generated using the ggplot2 package of R.³²

Results Overview

Figure 1 (Consort chart) shows the formation of the analyzable sample. As expected, the sample used for this analysis presented with greater depression severity, longer duration of the depressive disorder (current age minus age at first episode), lower quality of life, and more concurrent general medical conditions compared to excluded participants. (See Tables S2 and S3 for all baseline demographic, severity, and clinical comparisons).

Overall, 157 analyzable participants completed 8 weeks on escitalopram; 175 on sertraline, and 135 on venlafaxine-XR. Remission rates at 8 weeks for these samples were 64/157 (40.8%) for escitalopram, 53/175 (30.3%) for sertraline, and 42/135 (31.1%) for venlafaxine-XR. The proportion of remitters did not significantly vary across treatment arms ($\chi^2(2, N=467)=4.78, P=0.09$).

Question 1

Can this test battery identify a meaningful proportion of these participants that exceeds chance prediction (defined as 50–50 likelihood of remission) by at least 10% for depressed participants who are treated for 8 weeks with escitalopram, sertraline or venlafaxine-XR?

We found that we could predict remission in each medication arm with a cross-validated estimate of accuracy – as well as sensitivity and specificity – that exceeded chance by 10 percentage points for nearly half of the final sample for each arm. Only when the participants in each arm were subgrouped according to performance on cognitive or emotion tests did the predictions of remission exceed 10 percentage points above chance.

Table 2 shows the percentage of participants in each treatment arm for whom a prediction of remission, both likely

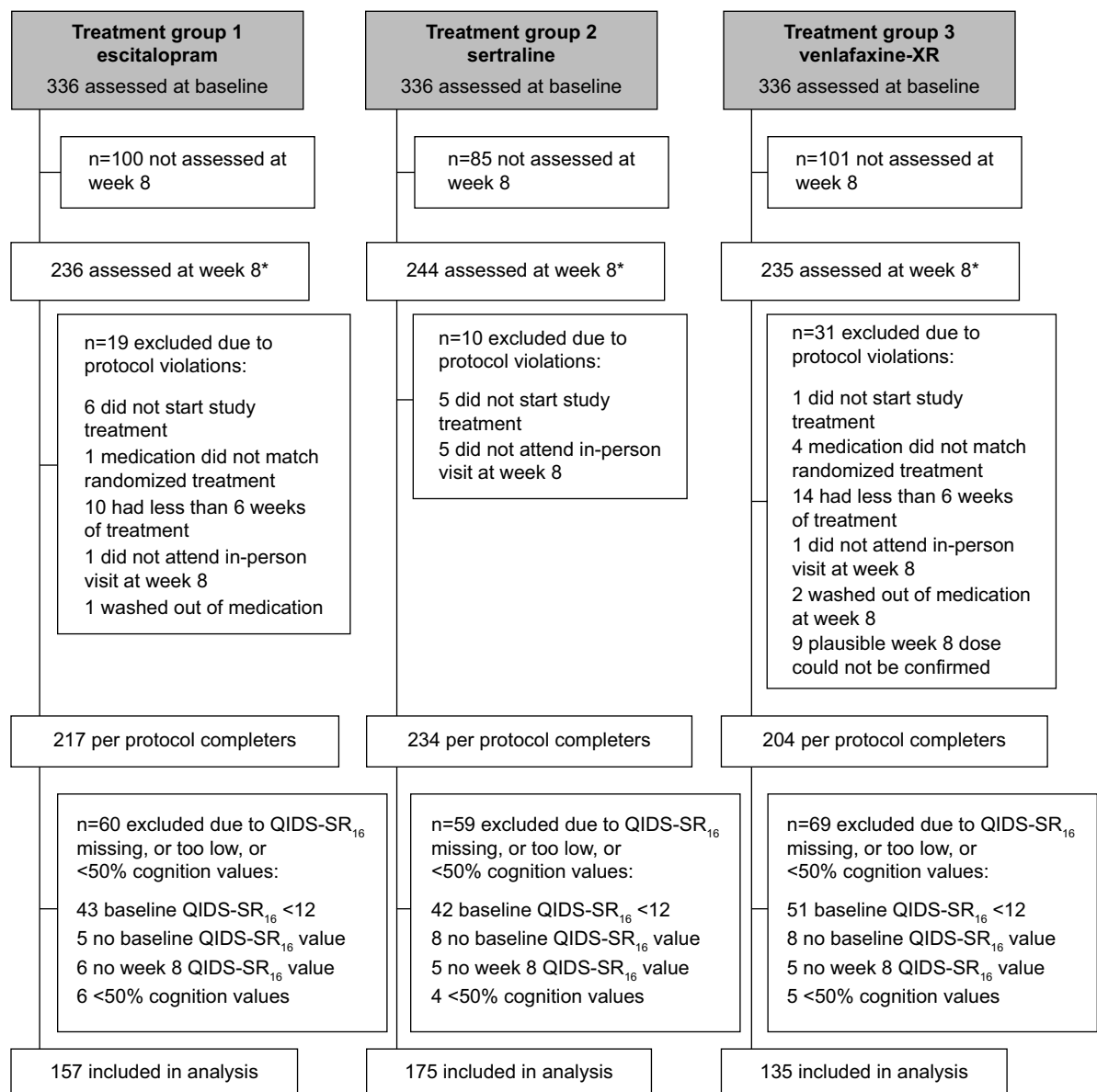


Figure 1 Consort chart.

Notes: Of the patients who met inclusion criteria for the iSPOT-D study, additional criteria were used for this specific analysis. *Minimum criteria for assessment at week 8 is either completion of the HRSD₁₇ or SOFAS assessments at the in-person visit, or completion of the week 8 QIDS-SR₁₆ in the online questionnaire at home.

Abbreviations: QIDS-SR₁₆, 16-item Quick Inventory of Depressive Symptomatology Self-Report; iSPOT-D, International Study to Predict Treatment Response in Depression; XR, extended release; HRSD₁₇, 17-item Hamilton Rating Scale for Depression.

and unlikely to remit, could be made. The test battery was able to make predictions of remission in 43.3% (68/157) of participants in the escitalopram arm, 63.4% (111/175) of those in the sertraline arm, and 47.4% (64/135) of those in venlafaxine-XR arm. If we applied all three tests in the test

battery to all participants in the three treatment arms, nearly 96% of participants would receive a prediction of remission/non-remission on at least one of the three treatments.

Table 3 reports the performance of each classification model for each treatment arm in specific subgroups of

Table 2 Proportion of patients by treatment for whom a test battery recommendation could be made

Medication	Recommendation		No recommendation
	Likely to remit	Not likely to remit	
Escitalopram	21%	22%	57%
Sertraline	24%	39%	37%
Venlafaxine-XR	21%	27%	52%

Abbreviation: XR, extended release.

participants, both with and without cross-validation. Remission could be predicted with a cross-validated sensitivity of 72% and specificity of 67% for participants taking escitalopram with below-average emotion identification performance, with 69% sensitivity and 70% specificity in participants over 30 years of age taking sertraline, and with 67% sensitivity and 62% specificity in participants taking venlafaxine-XR with above average general cognitive performance.

Figure 2 shows the specific regression model and participant subgroup in each treatment arm. For participants taking escitalopram with below average emotion identification performance, the regression model used switching of attention completion time, verbal interference incongruent correct reaction time, and sustained attention reaction time as the predictors. For participants who were at least 30 years old taking sertraline, the regression model used switching of attention completion time, digit span forward trials correct, and number of maze overrun errors as the predictors. For participants with above average general cognition performance the regression model used baseline QIDS-SR₁₆ severity as the predictor. No additional covariates were used in these regression models.

Figure 3 shows the receiver operator characteristics curves for each of these models using only the subset of participants in each medication group for whom a prediction of remission could be made. The test battery model for escitalopram and sertraline performed better than models that used age and sex or depression severity, as assessed by area under the curve. The test battery model for venlafaxine-XR performed better than a model that used age and sex, as assessed by area under the curve. While it was necessary to use baseline depression severity, defined by the QIDS-SR₁₆, as a predictor of remission for participants with above average cognition performance taking venlafaxine-XR, baseline QIDS-SR₁₆ did not

predict remission in the full treatment arm for venlafaxine-XR (cross-validated sensitivity, specificity, PPV, and NPV: 61%, 53%, 37%, and 74%) or in participants with below average general cognition performance (cross-validated sensitivity, specificity, PPV, and NPV: 38%, 45%, 21%, and 66%).

Question 2

Can the test battery identify a meaningful proportion of participants who will or will not achieve remission with a very high certainty level at least 80%, with 8 weeks of escitalopram, sertraline or venlafaxine-XR.

Table 4 reports the proportion of participants for whom predictions of remission could be made with very high certainty (PPV >80%). The predicted probability of remission was obtained for participants in the specific subgroups using the specific regression equations reported in Figure 2. While PPV never surpassed 80% for either the sertraline or venlafaxine-XR models, a threshold was found in which PPV surpassed 80% for 22/157 (14.0%) of the participants who received escitalopram.

Table 5 reports the proportion of participants for whom predictions of non-remission could be made with very high certainty (NPV >80%). The predicted probability of remission was obtained for participants in the specific subgroups using the regression equations reported in Figure 2. Overall, 24.2% (38/157), 53.7% (94/175), and 26.6% (36/135) of all participants in each treatment arm could be predicted to not remit with an NPV >80% in escitalopram, sertraline, and venlafaxine-XR, respectively.

The test battery identified 38 of the 93 (40.9%) participants who ultimately did not remit after 8 weeks of escitalopram, 94 of the 122 (77.1%) participants who did not remit after 8 weeks of sertraline, and 36 of the 93 (38.7%)

Table 3 Sensitivity and specificity of the depression treatment test

	<u>Sensitivity</u>	<u>Specificity</u>	<u>PPV</u>	<u>NPV</u>	<u>Odds</u>	<u>NNT</u>
	<u>DT value</u>	<u>DT value</u>	<u>DT value [observed</u>	<u>DT value [observed</u>	<u>ratio</u>	
	<u>[chance rate]</u>	<u>[chance rate]</u>	<u>remission rate]</u>	<u>non-remission rate]</u>		
Full sample						
Escitalopram	75 [50]	70 [50]	64 [41]	80 [59]	7.0	2.3
Sertraline	71 [50]	73 [50]	48 [25]	88 [75]	6.9	2.7
Venlafaxine-XR	67 [50]	67 [50]	50 [33]	81 [67]	4.1	3.3
Cross-validated						
Escitalopram	72 [50]	67 [50]	61 [41]	77 [59]	4.9	2.7
Sertraline	69 [50]	70 [50]	43 [25]	87 [75]	4.4	3.5
Venlafaxine-XR	67 [50]	62 [50]	46 [33]	81 [67]	2.7	4.6

Notes: All statistics reported here are calculated only for the intended population of the prediction model (better than average emotion identification in escitalopram, >30 years of age in sertraline, and better than average general cognition performance in venlafaxine-XR). Full-sample statistics report the accuracy statistics calculated from the same sample used to develop the regression model. Cross-validated statistics are the averaged accuracy statistics of 300 cross-validation iterations. Observed remission rates are based upon the specific cohort used for the depression test (n=68, n=111, and n=64 for escitalopram, sertraline and venlafaxine-XR, respectively).

Abbreviations: DT, depression treatment test; NNT, number needed to treat; NPV, negative predictive values; PPV, positive predictive values; XR, extended release.

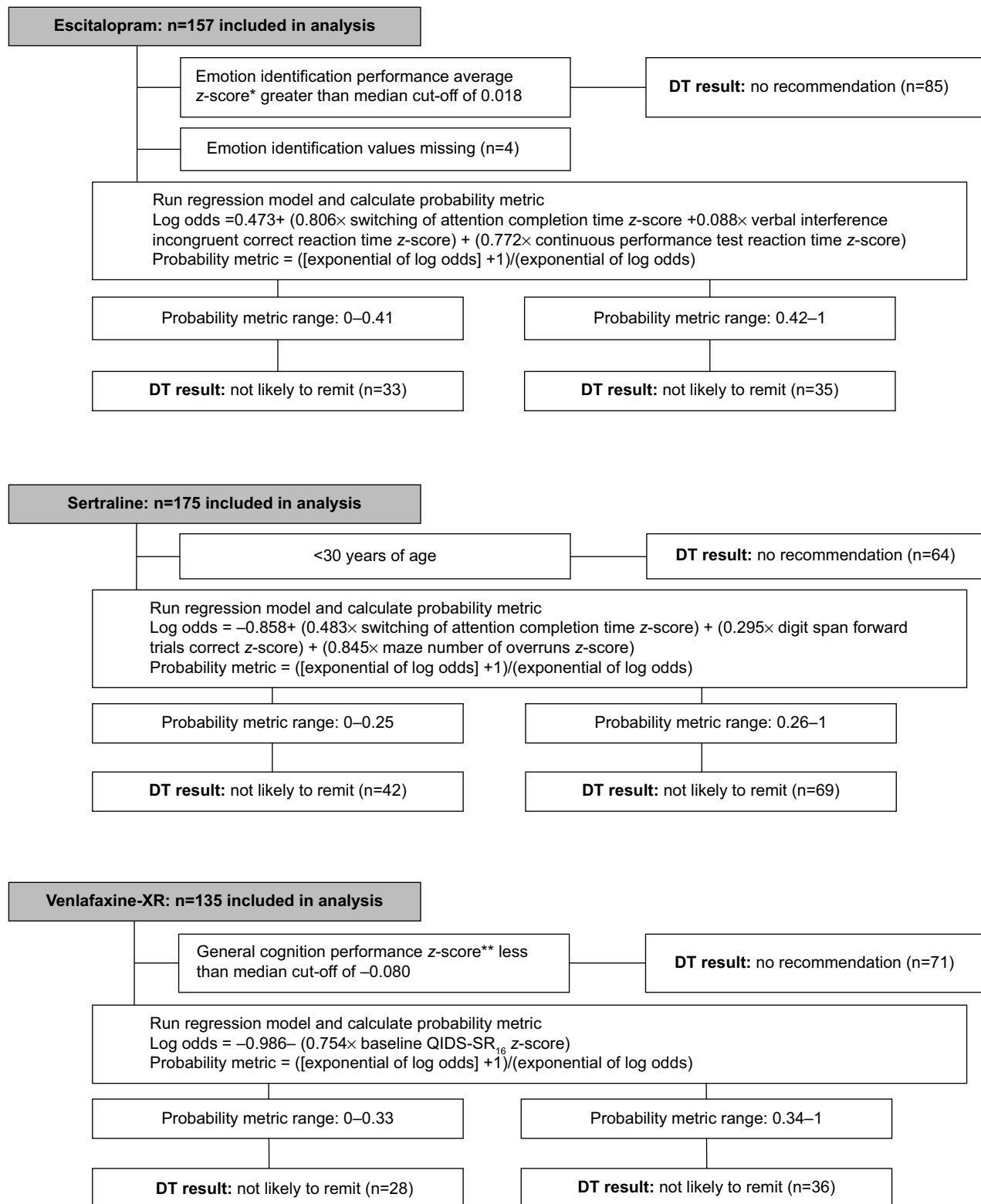


Figure 2 Final regression model to predict remission in each treatment arm and applicable subgroup.

Notes: Each treatment arm was stratified into two groups. Remission could be predicted 10 percentage points above chance. The specific subgroup and threshold is reported. The specific regression equation for each treatment arm is reported. Additionally, the number of patients in each treatment arm and each subgroup is also reported. DT result is the Depression Treatment Test predictive outcome. *Defined as average of z-scores for emotion identification number of correct responses and reaction time for correct responses. **Defined as average of z-scores for maze completion time, maze total number of errors and number of overrun errors, switching of attention completion time, verbal interference reaction time incongruent trials, continuous performance test reaction time, digit span total number of trials correct, and emotion identification number of correct response and reaction time for correct responses.

Abbreviations: DT, depression treatment test; QIDS-SR₁₆, 16-item Quick Inventory of Depressive Symptomatology Self-Report; XR, extended release.

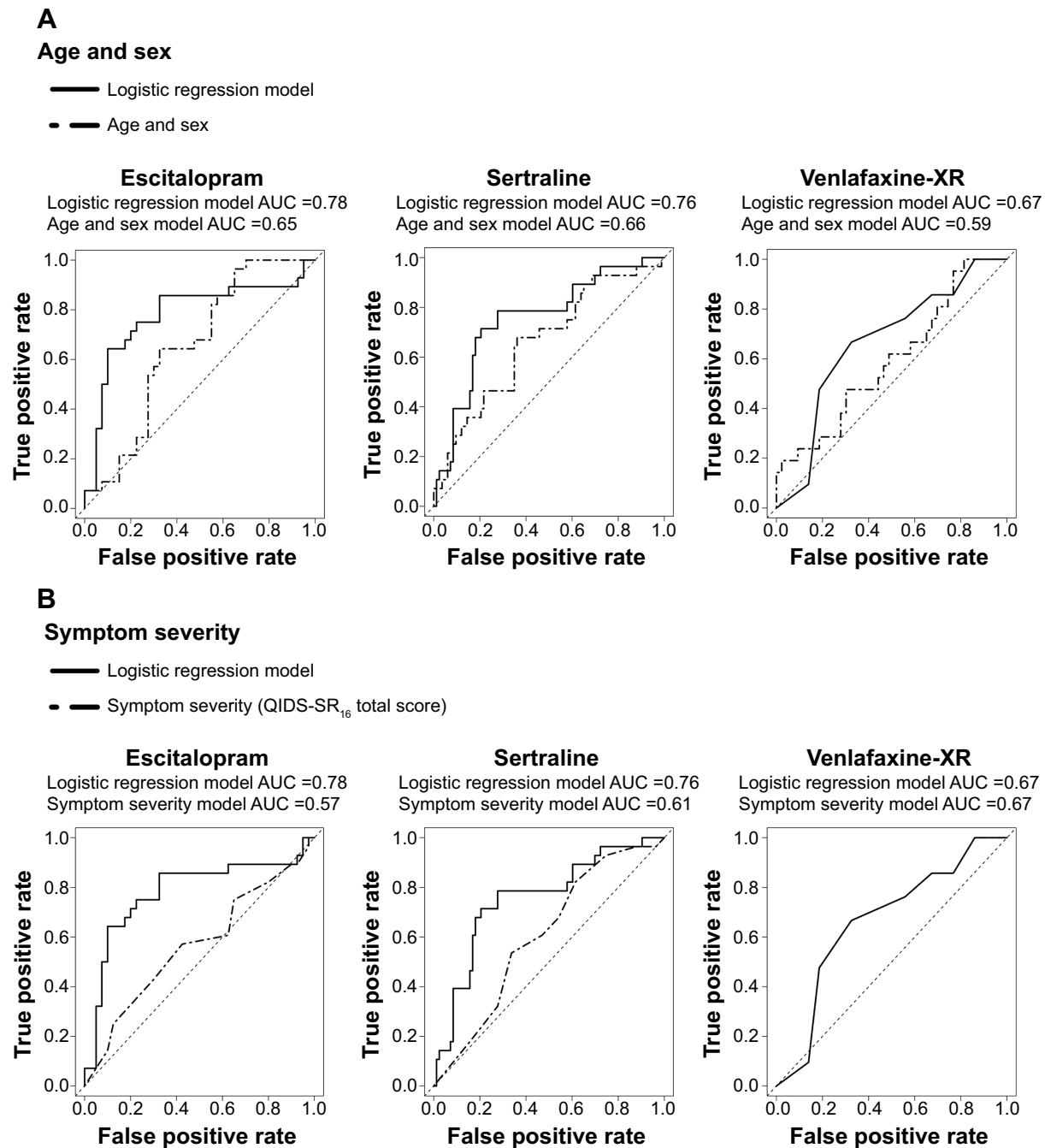


Figure 3 ROC curves comparing cognition with age and sex as well as depression severity.

Notes: This figure illustrates that regression models using cognition and emotion predictors outperform those that use only age and sex (A) or depression severity (B) (baseline QIDS-SR₁₆). The ROC curves were generated using the probability of remission obtained from each regression equation. Only the patients who met criteria for the logistic regression model were used to generate the ROC curves and calculate the AUC.

Abbreviations: AUC, area under the curve; QIDS-SR₁₆, 16-item Quick Inventory of Depressive Symptomatology Self-Report; ROC, receiver operating characteristics; XR, extended release.

participants who did not remit on venlafaxine-XR, with all counts including 20% false negatives.

Discussion

Overall, this baseline cognitive and emotional test battery, optimized for each medication separately and using only

baseline information, was able to identify a meaningful proportion (43% of participants in the escitalopram arm, 63% of those in the sertraline arm, and 47% of those in venlafaxine-XR arm) of depressed outpatients, for whom a prediction of reaching remission or non-remission after 8 weeks could be made with a certainty at least 10%

Table 4 Proportion of patients identified with greater than 80% certainty to remit

Medication	Total number	Total remitters	Number of participants >80% PPV	% of total treatment arm	% of remitters
Escitalopram	157	64	22	14.01%	34.38%
Sertraline	175	53	–	–	–
Venlafaxine-XR	135	42	–	–	–

Notes: The proportion of participants identified by the regression models to be highly certain to remit (PPV >80%). The number of participants reported represents the total number of participants identified by the regression model to be highly certain to remit and hence will reflect both true and false positives. Percentages of remitters identified include false positives.

Abbreviations: PPV, positive predictive values; XR, extended release.

greater than chance. If one were to run all of the three tests developed for each treatment on all the patients, then a prediction of remission or non-remission could be made for 96% of the patients for at least one of the three medications.

The same test battery algorithm identified a meaningful proportion of participants who ultimately did not remit after 8 weeks of treatment with a very high level of certainty (at least 80%). The test battery did not identify a meaningful proportion of participants in any treatment group who did remit with at least 80% certainty. On the other hand, the PPV were between 48% and 64% (cross-validated PPV were between 46% and 61%), indicating a 17–23 percentage point improvement from the remission rate. In other words, a participant who received a prediction of remission from any of the regression models would be at least 1.5–1.9 times as likely to remit compared to what would be expected by chance. Figure 4 illustrates the proportion of participants for whom a prediction could be made as well as the relative risks of predictions to remit and predictions to not remit. Moreover, different baseline test parameters were seen to predict remission in each of the three treatment arms. This finding is consistent with the idea that the patients who remitted on each medication may not be identical, though there may be some overlap among the three groups of remitting patients.

Present results are consistent with the general notion that cognitive measures can differentiate depressed patients who respond or remit acutely from those who do not. Dunkin

et al³³ found that non-responders to fluoxetine performed significantly worse than responders on the Wisconsin Card Sorting Test (executive function) at baseline. Alexopoulos et al³⁴ found among geriatric patients, non-responders to citalopram had poorer scores in the initiation/perseveration subscale of the Mattis Dementia Rating Scale and lower performance in the Stroop test than responders. Taylor et al⁸ reported that non-responders to fluoxetine displayed psychomotor slowing, measured with the Stroop and fluency tests. Gorlyn et al³⁵ found that responders treated with various selective serotonin reuptake inhibitors (SSRIs) outperformed non-responders across all cognitive domains, with the largest differences observed in executive function, language and working memory. Kampf-Sherf et al³⁶ found that responders to various SSRIs performed better on “simple” tasks and worse on “complex” tasks on a small cognitive battery that measured memory and executive functions. Gudayol-Ferré et al¹⁷ found that response to fluoxetine was predicted by good performance in variables of attention and low performance in planning. Conversely, Herrera-Guzmán et al³⁷ found that bupropion responders displayed low pre-treatment measures of visual memory and low levels of mental processing speed. Taken together, these studies reveal an association between baseline cognitive function and acute treatment outcomes. However, these studies assert group differences, which do not make individual patient predictions, and are, therefore, unable to guide individual treatment selection.

This report, however, goes beyond most of the above literature in that it aims to address the challenge of making

Table 5 Proportion of patients identified with greater than 80% certainty to not remit

Medication	Total number	Total non-remitters	Number of participants >80% NPV	% of total treatment arm	% of non-remitters
Escitalopram	157	93	38	24.20%	40.86%
Sertraline	175	122	94	53.71%	77.05%
Venlafaxine-XR	135	93	36	26.67%	38.71%

Notes: The proportion of participants identified by the regression models to be highly certain to not remit (NPV >80%). The number of participants reported represents the total number of participants identified by the regression model to be highly certain to remit and hence will reflect both true and false negatives. Percentages of non-remitters identified include false positives.

Abbreviations: NPV, negative predictive values; XR, extended release.

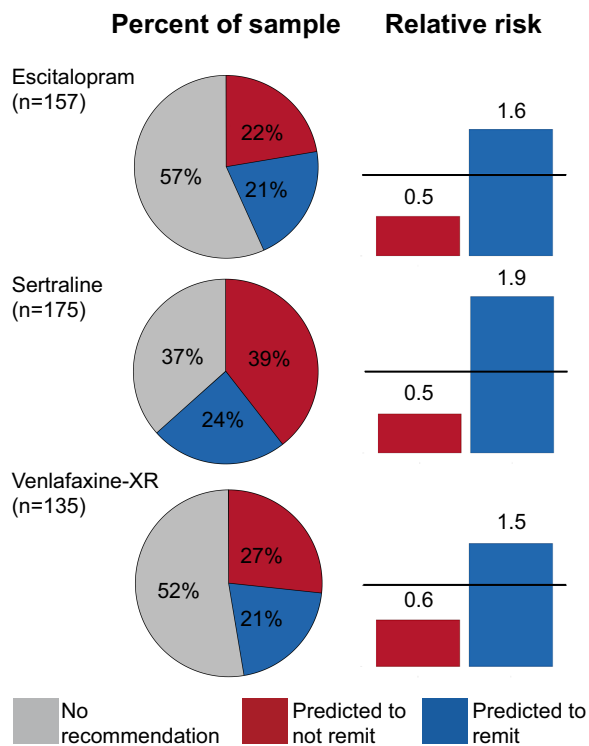


Figure 4 Percent of the sample receiving a recommendation to remit or to not remit, and relative risks for each type of prediction.

Notes: Relative risks were calculated using the remission rate for the specific patients for whom a prediction was made. A horizontal bar is drawn at a relative risk of 1 to allow for chance comparisons. Patients predicted to not remit, with a relative risk of 0.5–0.6, have one half the remission rate for the subgroup used to develop the specific regression model. Patients predicted to remit, with a relative risk of 1.5–1.9, have one and a half to two times the remission rate for the subgroup used to develop the specific regression model.

Abbreviation: XR, extended release.

individual patient predictions. In this context, the test battery was able to identify individual participants for whom recommendations could be made, outperforming chance for almost half of the participants. These findings are consistent with Etkin et al³⁸ who used a novel pattern classification analysis to show that predictions of remission can be attained for a single subgroup of one quarter of patients who show widespread cognitive deficits. This current report, focused on whether predictions of remission using measures of cognition can be obtained for a larger proportion of the patient population by using simpler statistical techniques that are more clinically actionable by pragmatically limiting the number of informative parameters needed to make predictions. With this method, the test battery was able to identify 32%–62% of individual participants who will not remit after 8 weeks of treatment with an actionable level of certainty (80%).

These findings are also consistent with a conceptual point raised by Kapur et al³⁹ and Prata et al.⁴⁰ They note that “diagnostic biomarkers themselves could be used to identify meaningful clinical sub-phenotypes” (Prata et al).⁴⁰

In fact, the test battery itself was initially formulated based on parameters that were presumed to differentiate depressed individuals from controls. Thus potentially “diagnostic” test elements may also identify treatment outcomes in subgroups of individuals. In addition, this report shows that different elements in the test battery identify individuals who remit or do not remit, suggesting that different patients are remitting with different treatments, despite the likely contribution of nonspecific treatment effects that are likely present in each group of participants who remitted.

This report has several strengths and limitations. The sample sizes are substantial. Data come from multiple sites suggesting generalizability. Analyses focused on the per protocol completer sample rather than on the full intent-to-treat sample to be certain that each individual participant whose outcome we wished to predict completed the treatment. This approach limits the generalizability of the findings and it may exaggerate or underestimate the value of the cognitive and emotional test battery as those who complete the full 8 weeks tend to be those who are benefitting. Additionally, the study exclusion criteria limit the generalizability. Further study will be required to determine how the test battery would actually perform in practice. However, the cross-validated analyses provide estimates of model performance on data not used to develop the model thereby providing an assessment of the generalizability of these models to new patients. Second, the doses of the medications used in the study were modest (mean final dosages: 12.3 mg for escitalopram, 63.6 mg for sertraline, and 84.9 mg for venlafaxine-XR), which may be representative of typical but not optimal patient care in diverse treatment settings. Higher medication doses might have resulted in more remitted participants which could affect the performance of the test battery. There is a clear need to replicate and extend these findings in new samples to evaluate the reliability, generalizability, and validity of these results, as noted in Williams et al.²⁶

In conclusion, this report suggests that, if replicated, a cognitive and emotional test battery can be clinically useful in identifying individual depressed patients who will remit or not remit beyond chance expectations. For all three medications, a meaningful number of individuals who did not remit was identified at baseline, with sufficient certainty that the clinician can decide to consider alternative treatments.

Acknowledgments

The iSPOT-D study is sponsored by Brain Resource Company Operations Pty Ltd. We received financial and material support from Brain Resource Ltd., Sydney, NSW, Australia.

We received material support from Brain Resource Inc., San Francisco, CA, USA. We gratefully acknowledge the iSPOT-D Investigators Group, and the contributions of principal investigators at each site, the monitoring support of Claire Day and the central iSPOT-D management team and PhaseForward, the data processing and quality control support of the central iSPOT-D management team, the secretarial support of Liane C Kivela, the editorial support of Jon Kilner, MS, MA (Pittsburgh, PA), and the support of Leanne William PhD in the design and execution of the iSPOT-D study.

Trial registry

Registry Name: ClinicalTrials.gov. Registration Number: NCT00693849. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00693849?term=ispot-D&rank=1>.

Disclosure

Drs Gordon and Palmer, Mr Braund and Mr Rekshan are employees of Brain Resource Ltd. Dr Palmer, Mr Braund, and Mr Rekshan hold stock options in Brain Resource Ltd. In the last 3 years, Dr Rush has received consulting fees from Brain Resource Ltd, H. Eli Lilly, Lundbeck A/S, Medavante, Inc; National Institute of Drug Abuse, Santium Inc., Takeda USA; speaking fees from the University of California at San Diego, Hershey Penn State Medical Center, the American Society for Clinical Psychopharmacology, and New York State Psychiatric Inst; royalties from Guilford Publications and from the University of Texas Southwestern Medical Center; a travel grant from CINP and research support from Duke-National University of Singapore. The authors have no other conflicts of interest to declare.

References

1. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003;289(23):3095–3105.
2. World Health Organization. *The global burden of disease: 2004 update*. Geneva: WHO; 2008.
3. World Health Organization [homepage on the Internet]. *Mental Health Atlas 2011*. Geneva, Switzerland: WHO; c2011. Available from: http://www.who.int/mental_health/publications/mental_health_atlas_2011/en/index.html. Accessed 8 September 2014.
4. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163(11):1905–1917.
5. Rush AJ, Kraemer HC, Sackeim HA, et al. Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology*. 2006;31(9):1842–1853.
6. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
7. Thase ME, Rush AJ. Treatment-resistant depression. In: Bloom FE, Kupfer DJ, editors. *Psychopharmacology: fourth generation of progress*. New York: Raven Press; 1995:1081–1097.

8. Taylor BP, Bruder GE, Stewart JW, et al. Psychomotor slowing as a predictor of fluoxetine nonresponse in depressed outpatients. *Am J Psychiatry*. 2006;163(1):73–78.
9. O'Leary D, Costello F, Gormley N, Webb M. Remission onset and relapse in depression. An 18-month prospective study of course for 100 first admission patients. *J Affect Disord*. 2000;57(1–3):159–171.
10. Rush AJ. STAR*D: what have we learned? *Am J Psychiatry*. 2007;164(2):201–204.
11. Bagby RM, Ryder AG, Cristi C. Psychosocial and clinical predictors of response to pharmacotherapy for depression. *J Psychiatry Neurosci*. 2002;27(4):250–257.
12. Gaynes BN, Warden D, Trivedi MH, Wisniewski SR, Fava M, Rush AJ. What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. *Psychiatr Serv*. 2009;60(11):1439–1445.
13. Kuk AY, Li J, Rush AJ. Recursive subsetting to identify patients in the STAR*D: a method to enhance the accuracy of early prediction of treatment outcome and to inform personalized care. *J Clin Psychiatry*. 2010;71(11):1502–1508.
14. Li J, Kuk AY, Rush AJ. A practical approach to the early identification of antidepressant medication non-responders. *Psychol Med*. 2012;42(2):309–316.
15. Lutz W, Lowry J, Kopta SM, Einstein DA, Howard KI. Prediction of dose-response relations based on patient characteristics. *J Clin Psychol*. 2001;57(7):889–900.
16. Simon GE, Khandker RK, Ichikawa L, Operskalski BH. Recovery from depression predicts lower health services costs. *J Clin Psychiatry*. 2006;67(8):1226–1231.
17. Gudayol-Ferre E, Herrera-Guzman I, Camarena B, et al. The role of clinical variables, neuropsychological performance and SLC6A4 and COMT gene polymorphisms on the prediction of early response to fluoxetine in major depressive disorder. *J Affect Disord*. 2010;127(1–3):343–351.
18. Korgaonkar MS, Williams LM, Song YJ, Usherwood T, Grieve SM. Diffusion tensor imaging predictors of treatment outcomes in major depressive disorder. *Br J Psychiatry*. 2014;205(4):321–328.
19. No authors listed. Practice guideline for the treatment of patients with major depressive disorder (revision). American Psychiatric Association. *Am J Psychiatry*. 2000;157(4 Suppl):1–45.
20. NCCMH. *Depression: The Treatment and Management of Depression in Adults* (Updated Edition). Leicester and London: The British Psychological Society and the Royal College of Psychiatrists; 2010.
21. Rush AJ, Wisniewski SR, Warden D, et al. Selecting among second-step antidepressant medication monotherapies: predictive value of clinical, demographic, or first-step treatment features. *Arch Gen Psychiatry*. 2008;65(8):870–880.
22. Gordon E, Cooper N, Rennie C, Hermens D, Williams LM. Integrative neuroscience: the role of a standardized database. *Clin EEG Neurosci*. 2005;36(2):64–75.
23. Paul RH, Lawrence J, Williams LM, Richard CC, Cooper N, Gordon E. Preliminary validity of “integneuro”: a new computerized battery of neurocognitive tests. *Int J Neurosci*. 2005;115(11):1549–1567.
24. Paul RH, Gunstad J, Cooper N, et al. Cross-cultural assessment of neuropsychological performance and electrical brain function measures: additional validation of an international brain database. *Int J Neurosci*. 2007;117(4):549–568.
25. Williams LM, Mathersul D, Palmer DM, Gur RC, Gur RE, Gordon E. Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. *J Clin Exp Neuropsychol*. 2009;31(3):257–277.
26. Williams LM, Rush AJ, Koslow SH, et al. International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: rationale and protocol. *Trials*. 2011;12:4.
27. Rush AJ, Nierenberg A. Mood disorders: Treatment of depression. In: Sadock BJ, Sadock VA, Ruiz P, editors. *Kaplan & Sadock's Comprehensive Textbook of Psychiatry, Ninth Edition*. Philadelphia, Pennsylvania: Lippincott Williams & Wilkins; 2009.

28. Maroco J, Silva D, Rodrigues A, Guerreiro M, Santana I, Mendonca A. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminate analysis, logistic regression, neural networks, support vector machines, calcification tress and random forests. *BMC Research Notes*. 2011;4:299.
29. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computer; 2013. Available from: <http://www.R-project.org/>. Accessed September 25, 2013.
30. Kuhn M. caret: Classification and regression training. R package version 5.15–7. 2013. Available from: <http://CRAN.R-project.org/package=caret>. Accessed September 14, 2013.
31. Carstensen B, Plummer M, Laara E, Hills M. Epi: A package for statistical analysis in epidemiology. R package version 1.1.49. 2013. Available from: <http://CRAN.R-project.org/package=Epi>. Accessed September 14, 2013.
32. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer; 2009.
33. Dunkin JJ, Leuchter AF, Cook IA, Kasl-Godley JE, Abrams M, Rosenberg-Thompson S. Executive dysfunction predicts nonresponse to fluoxetine in major depression. *J Affect Disord*. 2000;60(1):13–23.
34. Alexopoulos GS, Kiosses DN, Heo M, Murphy CF, Shanmugham B, Gunning-Dixon F. Executive dysfunction and the course of geriatric depression. *Biol Psychiatry*. 2005;58(3):204–210.
35. Gorlyn M, Keilp JG, Grunebaum MF, et al. Neuropsychological characteristics as predictors of SSRI treatment response in depressed subjects. *J Neural Transm*. 2008;115(8):1213–1219.
36. Kampf-Sherf O, Zlotogorski Z, Gilboa A, et al. Neuropsychological functioning in major depression and responsiveness to selective serotonin reuptake inhibitors antidepressants. *J Affect Disord*. 2004;82(3):453–459.
37. Herrera-Guzman I, Gudayol-Ferre E, Lira-Mandujano J, et al. Cognitive predictors of treatment response to bupropion and cognitive effects of bupropion in patients with major depressive disorder. *Psychiatry Res*. 2008;160(1):72–82.
38. Etkin A, Patenaude B, Song Y, et al. A Cognitive-Emotional Biomarker for Predicting Remission with Antidepressant Medications: A Report from the iSPOT-D Trial. *Neuropharmacology*. In press.
39. Kapur S, Phillips AC, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012;17(12):1174–1179.
40. Prata D, Mechelli A, Kapur S. Clinically meaningful biomarkers for psychosis: A systematic and quantitative review. *Neurosci Biobehav Rev*. 2014;45:134–141.

Supplementary materials

Table S1 Inclusion/exclusion criteria

Inclusion criteria

- Age 18–65
- Fluent and literate in English or Dutch
- Provide written informed consent
- Total HRSD₁₇ \geq 16
- Meets DSM-IV criteria for single or recurrent nonpsychotic MDD established by MINI Plus

Exclusion criteria

- Suicidal ideation and/or tendencies, defined by a score \geq 8 on section C of the MINI Plus
- History of bipolar disorder (I, II, not otherwise specified) (lifetime)
- History of schizophrenia, schizoaffective disorder, or psychosis not otherwise specified (lifetime)
- Current primary diagnosis of anorexia or bulimia, obsessive-compulsive disorder, or primary post-traumatic stress disorder
- Known contraindication for escitalopram, sertraline and/or venlafaxine-XR, or previous treatment failure at the highest recommended dose
- Taking any medication that is contraindicated with escitalopram, sertraline, or venlafaxine-XR
- Taking escitalopram, sertraline, or venlafaxine-XR in the current episode of MDD
- Use of any non-protocol antidepressant drug or CNS drug (antipsychotic, anticonvulsant anxiolytic, clonidine) that cannot be washed out prior to participation
- Has general medical condition that contraindicates protocol antidepressant treatments or interferes with protocol measurements (such as epileptic condition for EEG recording)
- Substance dependence (including alcohol intake equaling 29 standard alcoholic drinks per week for males; > 15 for females) in the past 6 months
- History of brain injury or blow to the head that resulted in loss of consciousness for greater than 5 minutes
- Severe impediment to vision, hearing and/or hand movement that is likely to interfere with completion of assessments, or with comprehension of instructions or study requirements
- Participation in an investigational study within 4 months prior to baseline that could affect symptoms of MDD
- Is pregnant or breast-feeding

Abbreviations: HRSD₁₇, 17-item Hamilton Rating Scale for Depression; MDD, major depressive disorder; XR, extended release; EEG, electroencephalogram; DSM, Diagnostic and Statistical Manual of Mental Disorders; MINI, Mini-International Neuropsychiatric Interview; CNS, central nervous system.

Table S2 Comparison of included and excluded samples (categorical measures)

	Excluded from sample (N=541)			Included in sample (N=467)			χ^2	P-value
	Count	%	Total N	Count	%	Total N		
Sex (female)	240	44.36%	541	197	42.18%	467	3.498	0.174
Family history of MDD	114	21.07%	541	119	25.48%	467	0.313	0.855
MDD diagnosis prior to 18	239	45.01%	531	239	51.96%	460	1.992	0.369
Race							5.146	0.161
Black	93	17.19%	541	74	15.85%	467		
Other	125	23.11%	541	87	18.63%	467		
Unknown	1	0.18%	541	3	0.64%	467		
White	322	59.52%	541	303	64.88%	467		
Employment							6.848	0.232
Employed	267	49.35%	541	239	51.18%	467		
Other	37	6.84%	541	36	7.71%	467		
Retired	28	5.18%	541	15	3.21%	467		
Student	100	18.48%	541	89	19.06%	467		
Unemployed	30	5.55%	541	36	7.71%	467		
Unknown	79	14.60%	541	52	11.13%	467		
Marital status							6.391	0.172
Divorced/separated	68	12.57%	541	74	15.85%	467		
Married/cohabiting	98	18.11%	541	95	20.34%	467		
Single	341	63.03%	541	273	58.46%	467		
Unknown	27	4.99%	541	15	3.21%	467		
Widowed	7	1.29%	541	10	2.14%	467		
Recruitment source							1.261	0.738
Ad	424	78.37%	541	371	79.44%	467		
GP	12	2.22%	541	14	3.00%	467		
Other	48	8.87%	541	40	8.57%	467		
Psych	57	10.54%	541	42	8.99%	467		
Number of episodes							6.298	0.278
0	2	0.38%	528	0	0.00%	457		
1	57	10.80%	528	46	10.07%	457		
2	53	10.04%	528	32	7.00%	457		
3	60	11.36%	528	45	9.85%	457		
4	64	12.12%	528	56	12.25%	457		
5 or more	292	55.30%	528	278	60.83%	457		
Dysthymia	101	18.67%	541	118	25.27%	467	2.660	0.264
Panic disorder	43	7.95%	541	42	8.99%	467	0.488	0.783
Agoraphobia	37	6.85%	540	37	7.92%	467	1.237	0.539
Social phobia	37	6.84%	541	56	11.99%	467	4.549	0.103
Specific phobia	24	4.44%	541	31	6.65%	466	0.607	0.738
Generalized anxiety disorder	33	6.10%	541	36	7.71%	467	3.360	0.186
Previous suicide attempt	64	11.83%	541	53	11.35%	467	0.406	0.816
Number of comorbid medication conditions							15.371	0.009
0	333	61.55%	541	236	50.54%	467		
1	111	20.52%	541	122	26.12%	467		
2	51	9.43%	541	45	9.64%	467		
3	22	4.07%	541	27	5.78%	467		
4	12	2.22%	541	19	4.07%	467		
5 or more	12	2.22%	541	18	3.85%	467		
MDD recurrence							0.393	0.822
Not recurrent	59	10.91%	541	46	9.85%	467		
Recurrent	469	86.69%	541	411	88.01%	467		
Unknown	13	2.40%	541	10	2.14%	467		

Abbreviations: MDD, major depressive disorder; GP, general practitioner; Psych, psychiatrist/psychologist.

Table S3 Comparison of included and excluded samples (continuous measures)

	Excluded from sample (N=541)			Included in sample (N=467)			t	P-value
	Mean	SD	Total N	Mean	SD	Total N		
Age (years)	37.30	12.56	541	38.47	12.57	467	-1.468	0.142
Education (years)	14.56	2.73	541	14.52	2.88	467	0.189	0.850
Age at first episode (years)	23.21	12.20	531	22.54	11.84	460	0.873	0.383
Duration of MDD (years)	14.02	11.64	531	15.83	12.72	460	-2.323	0.020
HRSD ₁₇ score	21.58	4.11	541	22.24	4.12	467	-2.521	0.012
QIDS-SR ₁₆ score	13.15	4.22	496	15.86	2.72	467	-11.920	0.000
DASS (depression)	19.73	9.45	469	24.75	8.97	451	-8.255	0.000
DASS (anxiety)	8.08	6.71	469	9.49	6.63	451	-3.199	0.001
DASS (stress)	17.00	8.15	469	19.47	8.45	451	-4.498	0.000
SOFAS	56.34	9.36	540	55.40	8.79	466	1.633	0.103
SWLS	12.10	5.51	506	11.01	5.17	461	3.184	0.002
WHOQoL-physical	54.44	14.04	485	49.02	14.35	453	5.850	0.000
WHOQoL-psychological	37.19	14.29	484	31.93	12.79	453	5.948	0.000
WHOQoL-social	40.96	20.21	485	36.09	19.22	454	3.784	0.000
WHOQoL-environmental	52.17	15.64	483	51.18	15.93	451	0.958	0.338
WHOQoL-overall	2.84	0.92	486	2.65	0.93	457	3.299	0.001
ERQ (reappraisal)	4.39	1.24	506	4.26	1.22	465	1.662	0.097
ERQ (suppression)	4.09	1.34	506	4.22	1.35	465	-1.557	0.120

Abbreviations: DASS, depression anxiety stress scale; ERQ, emotion regulation questionnaire; HRSD₁₇, 17-item Hamilton Rating Scale for Depression; MDD, major depressive disorder; QIDS-SR₁₆, 16-item Quick Inventory of Depressive Symptomatology Self-Report; SOFAS, social and occupational functioning assessment scale; SWLS, satisfaction with life scale; WHOQoL, World Health Organization quality of life; SD, standard deviation.

Neuropsychiatric Disease and Treatment

Dovepress

Publish your work in this journal

Neuropsychiatric Disease and Treatment is an international, peer-reviewed journal of clinical therapeutics and pharmacology focusing on concise rapid reporting of clinical or pre-clinical studies on a range of neuropsychiatric and neurological disorders. This journal is indexed on PubMed Central, the 'PsycINFO' database and CAS,

and is the official journal of The International Neuropsychiatric Association (INA). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/neuropsychiatric-disease-and-treatment-journal>