A New Prediction Model for Evaluating Treatment-Resistant Depression

Alexander Kautzky, MD[a]; Pia Baldinger-Melich, MD, PhD[a]; Georg S. Kranz, MSc, PhD[a]; Thomas Vanicek, MD[a]; Daniel Souery, MD, PhD[b]; Stuart Montgomery, MD, PhD[c]; Julien Mendlewicz, MD[d]; Joseph Zohar, MD, PhD[e]; Alessandro Serretti, MD[f]; Rupert Lanzenberger, MD[a]; and Siegfried Kasper, MD[a,*]

ABSTRACT

Objective: Despite a broad arsenal of antidepressants, about a third of patients suffering from major depressive disorder (MDD) do not respond sufficiently to adequate treatment. Using the data pool of the Group for the Study of Resistant Depression and machine learning, we intended to draw new insights featuring 48 clinical, sociodemographic, and psychosocial predictors for treatment outcome.

Method: Patients were enrolled starting from January 2000 and diagnosed according to *DSM-IV*. Treatment-resistant depression (TRD) was defined by a 17-item Hamilton Depression Rating Scale (HDRS) score ≥ 17 after at least 2 antidepressant trials of adequate dosage and length. Remission was defined by an HDRS score < 8. Stepwise predictor reduction using randomForest was performed to find the optimal number for classification of treatment outcome. After importance values were generated, prediction for remission and resistance was performed in a training sample of 400 patients. For prediction, we used a set of 80 patients not featured in the training sample and computed receiver operating characteristics.

Results: The most useful predictors for treatment outcome were the timespan between first and last depressive episode, age at first antidepressant treatment, response to first antidepressant treatment, severity, suicidality, melancholia, number of lifetime depressive episodes, patients' admittance type, education, occupation, and comorbid diabetes, panic, and thyroid disorder. While single predictors could not reach a prediction accuracy much different from random guessing, by combining all predictors, we could detect resistance with an accuracy of 0.737 and remission with an accuracy of 0.850. Consequently, 65.5% of predictions for TRD and 77.7% for remission can be expected to be accurate.

Conclusions: Using machine learning algorithms, we could demonstrate success rates of 0.737 for predicting TRD and 0.850 for predicting remission, surpassing predictive capabilities of clinicians. Our results strengthen data mining and suggest the benefit of focus on interaction-based statistics. Considering that all predictors can easily be obtained in a clinical setting, we hope that our model can be tested by other research groups.

J Clin Psychiatry 2017;78(2):215–222
https://doi.org/10.4088/JCP.15m10381

[a]Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria
[b]Université Libre de Bruxelles and Psy Pluriel Centre Européen de Psychologie Medicale, Brussels, Belgium
[c]Imperial College, University of London, London, United Kingdom
[d]School of Medicine, Free University of Brussels, Brussels, Belgium
[e]Psychiatric Division, Chaim Sheba Medical Center, Ramat Gan, Israel
[f]Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy
**Corresponding author:* Siegfried Kasper, MD, Department of Psychiatry and Psychotherapy, Medical University of Vienna, Währinger Gürtel 18-20, A-1090 Vienna, Austria (sci-biolpsy@meduniwien.ac.at).

Although major depressive disorder (MDD) results in 3.0% to 3.8% of global disability-adjusted life years, clinicians still rely on a limited repertory of antidepressants.[1,2] About 30% of patients treated with these agents do not respond to the first administered antidepressant, and about 15% of patients show hardly any or no relief even after administration of multiple antidepressants.[3,4] Therefore, clinicians and scientists have put much effort into developing potent strategies for treatment-resistant depression (TRD).[5]

Since the first scientific description of TRD, a broad discussion about criteria and staging systems has arisen, however, without resulting in a general definition.[6–8] The most common definition of TRD is characterized by failure to achieve a reduction of at least 50% from baseline on a recognized rating scale for MDD after application of at least 2 antidepressant treatment trials of adequate dosage and duration.[9–13] Alternatively, a score on a recognized rating scale, such as the 17-item Hamilton Depression Rating Scale (HDRS), that indicates the presence of moderate depressive symptoms after an adequate treatment period can also be used to characterize TRD.

Over the last decade, the findings of the Group for the Study of Resistant Depression (GSRD), a multinational European research consortium, have consistently pointed toward strong effects of clinical variables on treatment outcome in TRD.[14] However, when considered separately, predictors usually showed odds ratios (ORs) around 1.5 and, thus, were not applicable for detecting patients at high risk of resistance.[15,16]

In light of an incremental augmentation of gathered data and accelerating capabilities of processing these data, interaction-based models combining predictors as suggested by recent reviews seem, for the first time, increasingly viable.[17–21] As few studies featuring a large set of clinical variables have been conducted in MDD, we performed a study on treatment outcome in TRD investigating a set of 48 clinical and sociodemographic predictors.

METHOD

Sample Description

All patients derive from the still growing sample collected by the GSRD, currently comprising 1,371

Clinical Points

- Although single clinical predictors have repeatedly been associated with treatment-resistant depression (TRD), they have not proven sufficient for predicting treatment outcome. Thus, attention shifted to interaction-based models, but only a few multivariate investigations have been performed in TRD so far. In this investigation, we focused on evaluating the influence of a variety of sociodemographic and clinical factors, adopting a machine learning algorithm for prediction of treatment outcome in major depressive disorder.
- Our results suggest that suicidality, early age at onset, and age at first treatment with an antidepressant as well as inpatient status and poor response to the first antidepressant ever administered increase the risk for TRD and lower the chance of remission. While melancholia and panic disorder increase risk for treatment-resistant depression, favorable occupation status and comorbid diabetes and thyroid disorder affect remission.
- Exploiting a machine learning algorithm, we established a multivariate model featuring 48 clinical predictors in 400 patients that we tested in a new sample of 80 patients. We scored an accuracy of 73.7% for resistance and 85.0% for remission. Reaching a probability of 65.5% for a correct prediction for TRD and 77.7% for remission exceeded the predictive capabilities of clinicians.

patients recruited in the international referral centers of our group starting from January 2000. The study was approved by the ethics committees of all participating centers.

For a detailed description, please see Souery et al.[15] All subjects had to be diagnosed with MDD according to *DSM-IV* criteria and were recruited after giving informed consent. A modified version of the MINI-International Neuropsychiatric Interview (MINI), version 5.0.0, and the HDRS were applied for diagnosis of MDD and comorbidities and assessment of symptom severity.[11,12] MDD had to be the primary diagnosis, and patients with MDD only as a secondary diagnosis to any nonaffective psychiatric disease were excluded.

Of the 1,224 patients eligible from the GSRD sample, 480 patients showed full data availability for all 48 variables and were included in our investigation. Of these, 183 patients were resistant, 157 reached an HDRS score between 8 and 16, and 140 achieved remission (HDRS ≤ 7). See Supplementary eFigure 1.

Comparison of the patients that could not be enrolled in this study with the 480 subjects included in our analysis showed no significant differences regarding ethnicity, age, and gender, while HDRS scores were slightly lower in the excluded sample. A summary of baseline characteristics of both samples can be found in Supplementary eTable 1.

Treatment Outcome Phenotypes

We focused on 2 treatment outcome phenotypes, remission and resistance.

A score of more than 16 on the 17-item HDRS after application of at least 2 adequate antidepressant trials was regarded as resistance and was compared to nonresistance, which applied to patients who had an HDRS score ≤ 16 after 1 or 2 adequate treatment trials.

Remission was defined by an HDRS score ≤ 7 and was compared to nonremission, which was characterized by an HDRS > 7.

Predictors

Predictors with more than 30% missing values as well as redundant predictors were excluded. The resulting 48 featured predictors are based on items of the MINI psychiatric interview. For a list of all predictors, see Table 1.

In more detail, the sociodemographic predictors of gender, age, and ethnicity were included. As more than 98% of the collective was white, ethnicity was regarded as a binomial predictor. Additionally, we analyzed the psychosocial predictors marital status (single, in a relationship, married, divorced, or widowed), education (no legal school, legal school, secondary inferior or superior, or university), occupation (high, medium, low, or other), and number of children. Occupation was stratified by income and social status. Higher executives, business managers, proprietors of large and medium concerns, or major professionals were coded as "high," and administrative personnel, owners of small businesses, minor professionals, clerical and sales workers, technicians, skilled manual employees, and farmers owning significant property were coded as "medium." Machine operators, semiskilled or unskilled employees, and tenant farmers were coded as "low," while students, stockholders, invalid, and unemployed workers were coded as "other."

Table 1. List of All 48 Predictors Featured in the Analysis Ordered by Groups[a]

Predictor Type	Predictors
Sociodemographic predictors (no. = 7)	Age, gender, ethnicity, occupation, education, marital status, number of children
History of MDD (no. = 12)	MDD first-degree relatives, MDD second-degree relatives, BD first-degree relatives, BD second-degree relatives, number of relatives MDD, number of relatives BD, number of MDEs, age at first AD, response to first AD, timespan between first and last MDE, subsyndromal bipolarity, psychotic features lifetime
Axis II comorbidity (no. = 13)	GAD, social phobia, OCD, PTSD, panic disorder, agoraphobia, smoking, alcohol abuse, alcohol addiction, drug abuse, drug addiction, history of drug abuse, any substance use
Axis III comorbidity (no. = 2)	Diabetes, thyroid disorder
Clinical features (no. = 12)	Severity, suicidality, psychotic features currently, change of appetite, change of sleep, feelings of guilt, impaired decision-making, fatigue, social dysfunction, unrest, melancholia, autoaggressive thoughts
Other predictors (no. = 2)	Inpatient or outpatient, psychotherapy

[a]For details, see Supplementary eTable 1.
Abbreviations: AD = antidepressant, BD = bipolar disorder, GAD = generalized anxiety disorder, MDD = major depressive disorder, MDE = major depressive episode, OCD = obsessive-compulsive disorder, PTSD = posttraumatic stress disorder.

Furthermore, age at first antidepressant treatment as well as response to first antidepressant treatment, timespan between first and last depressive episode, number of depressive episodes, a family history of MDD and bipolar disorder (accounting for total number of affected relatives as well as any first-degree and second-degree relatives, respectively) were also considered. Based on items of the MINI psychiatric interview, the clinical symptoms of the current episode, namely fatigue (MINI A5d), appetite change (MINI A5a), sleep impairment (MINI A5b), unrest (MINI A5c), social dysfunctioning (MINI 6A), impaired decision-making (MINI A5f), feelings of guilt (MINI A5e), and autoaggressive thoughts (MINI A5g) were analyzed as well. Lifetime and current psychotic features, melancholia, suicidality, and severity (defined by an abundance of symptoms at the worst stage of the current episode and coded as moderate, severe, and severe with psychosis) were featured, too. Subsyndromal bipolarity was based on MINI items D1a and D2a and indicates that the patients experienced a period of feeling up or full of energy and of higher irritability during their lifetime without fulfilling criteria for the diagnosis of a bipolar disorder. Comorbidities such as panic disorder, generalized anxiety disorder, social phobia, agoraphobia, and posttraumatic stress disorder and substance use disorders such as smoking, substance use in general, and a history of substance use and alcohol or drug abuse or dependency were included. The somatic comorbidities of diabetes and thyroid disorders and psychotherapy as well as inpatient or outpatient status were also considered. See Supplementary eTable 2 for a detailed characterization of all predictors.

Statistical Analysis

We used the randomForest package for R software for our investigation (cran.r-project.org/), a machine learning algorithm designed for determining the most useful predictors for a dichotomous outcome parameter.[22] Specifically, randomForest calculates variable importance by mean decrease in Gini (MDG). MDG is computed using out-of-bag samples of rearranged values for the variables. According to the zero-hypothesis, rearrangement of nonsignificant variables should not decrease Gini values. Thus, MDG values are based on a permutation test and express the contribution of each variable to the homogeneity of the nodes and branches of the classification trees, ranging from 0 (homogeneity) to 1 (heterogeneity). The changes in Gini are summed up and normalized for all nodes split up by a specific predictor. Thus, higher MDG, meaning a higher purity of resulting nodes compared to original nodes, is an indicator for the importance of a predictor.

Because randomForest uses random starting values for growing trees, multiple runs with different randomly generated seeds were performed. Per run, the number of trees to grow was set to 1,000 in order to ensure that every input row gets predicted a few times (ntree = 1,000). The number of variables randomly sampled as candidates at each split (mtry) was set at 7 following the established rule of using the square root of the number of variables x (x = 48, mtry = 6,928).

Unfortunately, there is no established power calculation for randomForest. However, this algorithm has been shown to function with high reliability even with the number of predictors reaching the number of observations, providing sufficient patient counts and no missing data.[19,23,24]

Since randomForest requires a test sample independent from the training sample, we randomly assigned patients to one of the samples on a quota of 1 to 5. Therefore, 400 patients ended up in the training sample, and 80 patients were allocated to the test sample.

First, a 10-fold cross-validated stepwise factor reduction was applied to the training sample, repeating the prediction for remission and resistance while gradually excluding the least important variables. Subsequently, variable importance was calculated. Finally, we tried the models established in the training sample in a test sample comprising 80 untapped patients and computed prediction accuracy for treatment outcome. Results were also portrayed by receiver operating characteristics (ROC), using the ROCR package for the R-software (cran.r-project.org/).[25]

Since previous data suggest interaction effects to be more impactful on TRD than single factor effects, we intended to check for differences in interaction-based randomForest and in traditional approaches as generalized linear model (GLM) in a secondary analysis.

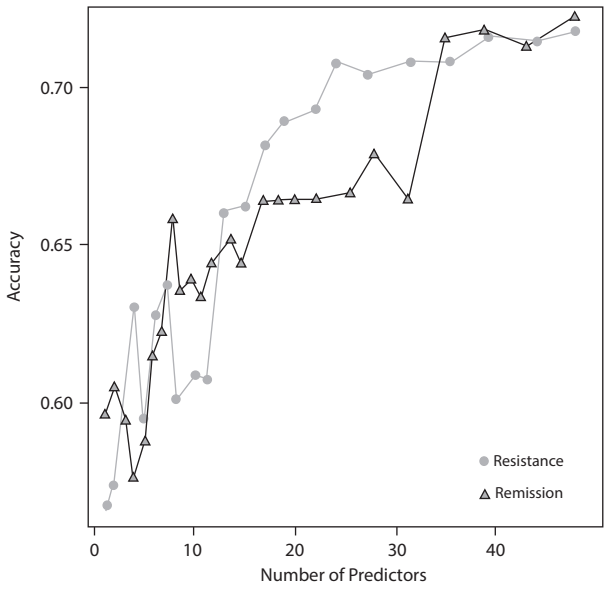
RESULTS

The 10-fold cross-validated stepwise factor reduction showed that the accuracy was inclining with the number of features for predicting resistance as well as remission, as portrayed in Figure 1. Using all 48 predictors showed the best result with an accuracy of approximately 72% for resistance and nonresistance and 77% for remission and nonremission. In contrast, when only the strongest predictor was used, the accuracy was at 56% for resistance and 60% for remission.

Average importance values for the highest scoring predictors for remission and resistance respectively are portrayed in Figure 2. Timespan between the first and last major depressive episode (MDE), age at and response to first antidepressant, suicide risk, number of lifetime depressive episodes, inpatient or outpatient status, education, and comorbid thyroid disorders were among the most influential predictors regarding both remission and resistance. While melancholia and comorbid panic disorder seemed to be more important regarding resistance, comorbid diabetes and a family history of MDD showed higher importance for remission. Importance values for all 48 predictors for remission and resistance respectively can be seen in Supplementary eFigure 2.

Investigating the 15 most important predictors for remission and resistance with a GLM resulted in significant association with treatment outcome for suicidality, thyroid disorder, and panic disorder ($P < .0001$ for resistance and remission); inpatient or outpatient status ($P = .0029$ for resistance, $P = .0002$ for remission); response to first

Figure 1. Accuracy of the Prediction Model Increases With the Number of Predictors[a]



[a]The number of predictors is portrayed on the x-axis, the accuracy on the y-axis. Prediction was performed several times with 10-fold cross validation, reducing the number of predictors at each run. Only the variables scoring the highest in the importance measurement were kept in the model. The graph indicates that the accuracy of the prediction declines with lesser numbers of predictors. When all 48 predictors were used, about 72% of predictions for resistance and 75% for remission were correct. When only 1 predictor was used, the error rate was about 44% for resistance and 40% for remission. Therefore, all 48 predictors were kept in the model.

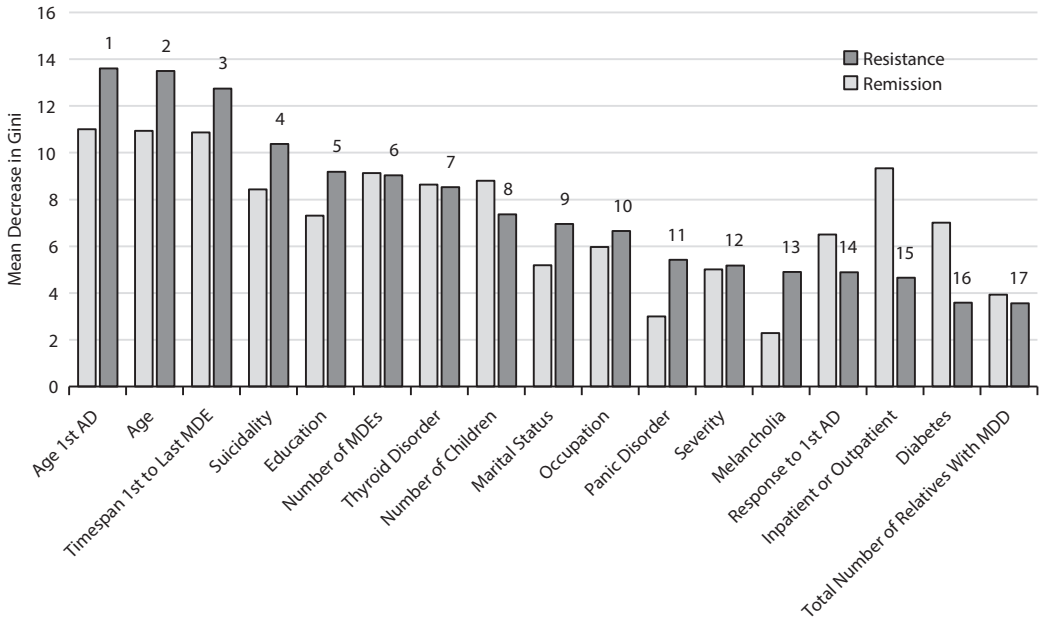
antidepressant administered lifetime ($P = .0029$ for resistance, $P = .0002$ for remission); and number of children ($P = .0339$ for resistance, $P = .0016$ for remission). Melancholia was only a significant predictor for resistance ($P < .0001$), while occupation status ($P = .0029$) and diabetes ($P = .0029$) were only relevant for remission. See Supplementary eTable 3 for a summary of the GLM results.

Most importantly, validating the prediction model in the test sample, we could reach a sensitivity of 0.633 and specificity of 0.800 for predicting resistance. The accuracy was at 0.737 using all 48 predictors. The positive predictive value (PPV) was at 0.655 and the negative predictive value (NPV) was at 0.784.

Concerning remission, 63.63% of remitters were predicted correctly, while 93.1% of nonremitters were recognized. The PPV was at 0.777, and the NPV was at 0.871. The overall accuracy was at 0.850.

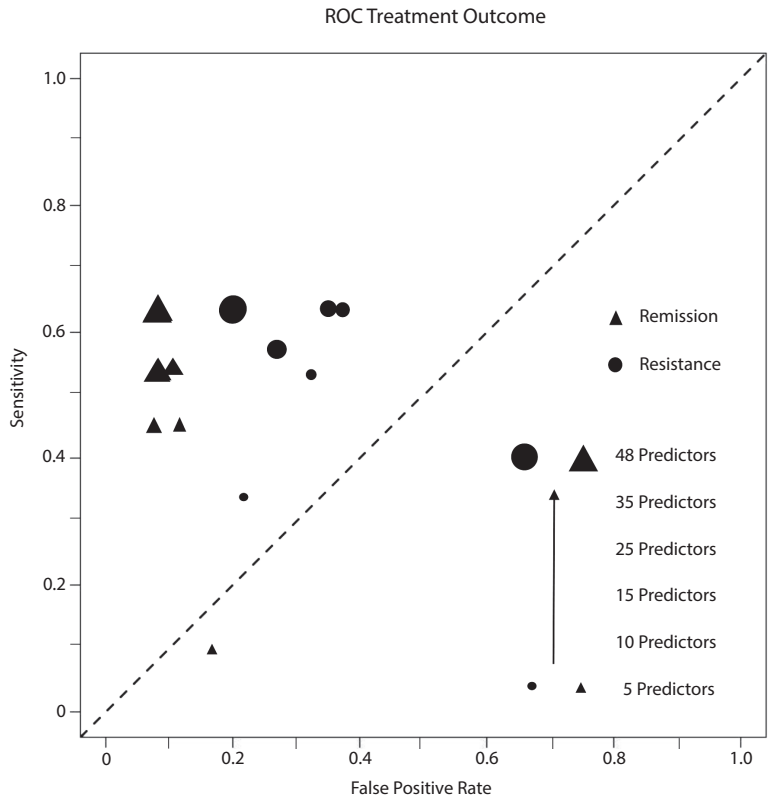
To check whether the model can be further simplified, we also ran a prediction based on the 35, 25, 15, 10, and 5 most important predictors. Notably, all parameters were subsequently declining with reduced numbers of predictors. The ROC for prediction of resistance and remission for all models are presented in Figure 3; accuracy and PPV are portrayed in Supplementary eFigure 3. All performance measures for the models can be found in Table 2. For easier interpretability, a complete list of these 6 predictor sets for remission and resistance respectively can be found in Supplementary eTable 4.

Figure 2. Importance Values for Best Predictors in the randomForest Machine Learning Algorithm for Remission and Resistance, Respectively[a]



[a]Importance refers to the usefulness of a specific predictor for predicting treatment outcome and is measured in mean decrease in Gini (MDG). MDG is shown on the y-axis. A higher value indicates more usefulness for classification decisions in the prediction model. The 15 highest ranking predictors for each, resistance and remission, are listed on the x-axis, starting with the most important predictor and ordered by importance for resistance. Age at the first antidepressant administered showed the highest MDG and was the most useful predictor to forecast whether patients will respond to antidepressant therapy or not. For importance measures of all predictors, see Supplementary eFigure 1.
Abbreviations: AD = antidepressant, BD = bipolar disorder, MDD = major depressive disorder, MDE = major depressive episode.

Figure 3. Receiver Operating Characteristic (ROC) for Predicting Remission and Resistance[a]



[a]The sensitivity is illustrated on the y-axis, the false positive rate on the x-axis. The relation of true positive (ie, sensitivity) and false positive outcome using all 48 predictors as well as only the 35, 25, 15, 10, and 5 most important predictors is shown. The diagonal dotted line indicates a random guess. Therefore, a dot above the diagonal line indicates better than random results, and the prediction results get better nearing the upper left corner. Prediction accuracy using all 48 predictors was at 0.737 for resistance and 0.850 for remission. Using fewer predictors, sensitivity was declining and the false positive rate was inclining.

Table 2. Performance Measures for the Predictions Models[a]

Number of Predictors	Sensitivity	Specificity	FPR	PPV	NPV	Accuracy
Resistance						
48	0.633	0.800	0.200	0.655	0.784	0.737
35	0.566	0.720	0.280	0.548	0.734	0.662
25	0.633	0.640	0.360	0.483	0.744	0.637
15	0.633	0.620	0.380	0.500	0.738	0.625
10	0.533	0.680	0.320	0.500	0.708	0.625
5	0.333	0.780	0.220	0.476	0.661	0.612
Remission						
48	0.636	0.931	0.069	0.777	0.871	0.850
35	0.545	0.913	0.087	0.705	0.841	0.812
25	0.545	0.896	0.104	0.666	0.838	0.800
15	0.454	0.913	0.087	0.666	0.815	0.787
10	0.454	0.862	0.138	0.555	0.806	0.750
5	0.090	0.827	0.173	0.166	0.705	0.625

[a]The most accurate prediction, meaning the highest number of correctly predicted patients, was achieved using all 48 predictors; accuracy was declining with 35, 25, 15, 10, and 5 predictors. The same holds true for the positive and negative predictive value. Using all 48 predictors, three-quarters of patients predicted to show remission and two-thirds of patients predicted to resist therapy can be expected to do so. As the complementary measure to specificity, the false positive rate was rising with reduced numbers of predictors. When all predictors were used, 20.0% of patients predicted not to resist therapy were actually resistant, and 6.9% of patients predicted not to show remission actually did show remission.

Abbreviations: FPR = false positive rate, NPV = negative predictive value, PPV = positive predictive value.

DISCUSSION

In this investigation, we focused on evaluating the influence of a variety of sociodemographic and clinical factors on treatment outcome. By adopting the machine learning algorithm randomForest, we succeeded in establishing a multivariate model based on 48 predictors in 400 patients that yielded an accuracy of 73.7% for resistance and 85.0% for remission when tested in a sample of 80 new patients.

In 2007, Souery et al[15] linked 10 clinical predictors to TRD. However, when investigated separately, these clinical features were not proven sufficient for prediction of treatment outcome, and attention shifted from single factors to combined models. Only a few multivariate investigations have been performed in TRD. Although the GSRD recently presented a combined genetic and clinical model, this model did not enable clinically significant prediction quality.[21,26–29]

Although previously not associated with TRD, timespan between first and last MDE and age at first administration of an antidepressant were the most useful predictors in this model. They are probably linked closely to already established predictors, namely number of lifetime depressive episodes, response to the first antidepressant, and severity, as well as age in general.[15,30] Concerning the number of depressive episodes, some studies indicated influence on remission while others could not demonstrate any effect.[15,31–33] Age has repeatedly been associated with TRD and older patients have shown worse treatment outcome in some studies.[33–38] On the basis of our results, we suggest that age is not a relevant predictor on its own, but is likely to moderate other factors.

Suicidality and panic disorder were important for this model and have repeatedly been associated with TRD in previous studies.[15,30] Among Axis III comorbidities, thyroid disorder and diabetes proved useful for predicting treatment outcome in our model. While Axis III comorbidity did not seem to influence treatment outcome in TRD in our previous studies, some results suggest that somatic diseases impact TRD.[15,33,39] Thyroid disorder has also been associated with MDD in a recent study that showed higher rates of autoimmune thyroiditis in unipolar and bipolar depressed patients.[40] Interestingly, thyroid disorder and diabetes both were associated with better treatment outcome in our current study. Because we did not distinguish between treated and untreated thyroid disorder and diabetes, our finding might be due patients receiving adequate treatment for their somatic comorbidity and subsequently showing better response to their psychopharmacologic medication as well. On the other hand, administration of thyroid hormone T_3 and, to a lesser

degree, T_4 has been suggested as potential augmentation therapy for TRD.[41] As patients with thyroid disorder were more likely to receive thyroid replacement therapy, this factor might have facilitated treatment response.

Furthermore, education, marital status, and occupation were important contributors. All 3 predictors have previously been associated with treatment outcome in TRD[38,42–44] but mostly showed inconclusive results in our study. Marital status was the most promising predictor and being married was suggested to protect from negative treatment outcome. On the other hand, divorce or being widowed enhanced the risk of not responding.

As hidden bipolarity has been highlighted as a potential cause for TRD, we featured subsyndromal bipolarity and the clinical predictor "unrest" in the prediction model.[45] However, both predictors showed low impact on treatment outcome. Unfortunately, agitation, which was shown to be a valid indicator for hidden bipolarity, was not recorded for this data pool.[46]

Previous studies on addiction comorbidities, family history of mood disorders, gender, and ethnicity were negative.[47–51] All of these predictors also ranked low in our prediction model. Social phobia and generalized anxiety disorder, on the other hand, have previously been associated with TRD[14,31,32] but showed low importance for our model.

In summary, our results from the randomForest and GLM analysis suggest that suicidality, early age at onset, and first treatment with antidepressant, as well as inpatient status and poor response to the first administered antidepressant increase risk for TRD and lower chances of remission. Additionally, melancholia and panic disorder are risk factors for TRD, while favorable occupation status and comorbid diabetes and thyroid disorder seem to increase chances for remission. The other predictors seem to exert their influence mainly through interaction effects that cannot be further specified by randomForest and should be clarified by future studies.

Nonetheless, our study design also displays some limitations. Even though randomForest has been demonstrated to function with a high number of predictors compared to the observations, we are aware of a risk of false positive results in this analysis. While we could successfully test our model in a new sample of 80 patients, this test sample also derived from the GSRD data pool. Whether our findings can be reproduced in other patients and whether they are independent of TRD staging methods will have to be clarified by future studies. Nevertheless, the fact that the results of the cross validation performed in the training data set were similar to the results in the test data set advocates for our findings.

Additionally, the patients from our collective received a broad range of antidepressants as well as augmentation therapy and electroconvulsive treatment, and most of them were also receiving more than 1 antidepressant.[5,34,52–56] Thus, further stratification by therapeutic agent would not be useful in this sample, as the subgroups would get too small.

While ours is one of the first studies to use new statistical approaches in TRD, another recent study[57] using a similar technique in a large STAR*D sample reached a comparable prediction accuracy of 0.71 using 15 clinical and sociodemographic predictors and logistic regression. Interestingly, predictors based on education, marital status, number of episodes, and severity showed significant influence on the prediction outcome in both studies, while other predictors, such as PTSD and ethnicity, were useful in the prediction model of Perlis[57] but showed rather small effects in our study. Most notably, the results of Perlis[57] suggested that a carefully selected set of predictors shows superior accuracy, while our results point toward increasing accuracy with the increasing number of predictors included.

In summary, we created an easily applicable model obtained from clinical and sociodemographic predictors offering an accuracy of 73.7% for TRD and 85.0% for remission. While the strongest predictors alone could not reach a prediction accuracy much different from random guessing, the combination of all 48 predictors enabled a probability of 65.5% for a correct prediction of TRD and 77.7% for remission, exceeding the predictive capabilities of clinicians. Featuring only the 15 most important predictors produced an accuracy of 62.5% for resistance and 78.7% for remission, which is less powerful, but might be of practical concern, as it is even more easily applicable. Therefore, we hope that other groups will test our model to enable implementation of machine learning algorithms in general practice. By reinforcing interaction-based approaches in TRD, we hope that this study will increase the awareness for these easily obtainable 48 predictors, which might become useful for predicting treatment outcome on an individual patient level.

Submitted: September 8, 2015; accepted March 15, 2016.

Online first: January 3, 2017.

Potential conflicts of interest: Without any relevance to this work, **Dr Kasper** has received grant/research support from Eli Lilly, Lundbeck A/S, Bristol-Myers Squibb, Servier, Sepracor, GlaxoSmithKline, Organon, and Dr Willmar Schwabe GmbH & Co KG; has served as a consultant or on advisory boards for AstraZeneca, Austrian Sick Found, Bristol-Myers Squibb, German Research Foundation, GlaxoSmithKline, Eli Lilly, Lundbeck A/S, Pfizer, Organon, Sepracor, Janssen, and Novartis; and has served on speakers' bureaus for AstraZeneca, Angelini Pharma, Eli Lilly, Krka Pharma, Lundbeck A/S, Servier, Sepracor, and Janssen. **Dr Serretti** is or has been a consultant/speaker for Abbott, AstraZeneca, Clinical Data, Boehringer, Bristol-Myers Squibb, Eli Lilly, GlaxoSmithKline, Janssen, Lundbeck, Pfizer, Sanofi, and Servier. **Dr Mendlewicz** is a member of the Board of the Lundbeck International Neuroscience Foundation and of the Advisory Board of Servier. **Dr Souery** has received grant/research support from GlaxoSmithKline and Lundbeck and has served as a consultant or on advisory boards for AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Janssen, and Lundbeck. **Dr Zohar** has received grant/research support from Lundbeck, Servier, and Pfizer; has served as a consultant or on advisory boards for Servier, Pfizer, Solvay, and Actelion; and has served on speakers' bureaus for Lundbeck, GlaxoSmithKline, Jazz, and Solvay. **Dr Montgomery** has been a consultant or served on Advisory boards for AstraZeneca, BioNevia, Bristol-Myers Squibb, Forest, GlaxoSmithKline, Grunenthal, Intellect Pharma, Johnson & Johnson, Lilly, Lundbeck, Merck, Merz, M's Science, Neurim, Otsuka, Pierre Fabre, Pfizer, PharmaNeuroBoost, Richter, Roche, Sanofi, Sepracor, Servier, Shire, Synosis, Takeda, Theracos, Targacept, Transcept, UBC, Xytis, and Wyeth. **Dr Lanzenberger** received travel grants and conference speaker honoraria from AstraZeneca, Lundbeck A/S, and Roche Austria GmbH. **Dr Baldinger-Melich** received travel grants from Roche Austria GmbH and AOP Orphan Pharmaceuticals AG and speaker honoraria from Janssen. **Dr Kranz** received travel grants from Roche and Pfizer. **Drs Kautzky** and **Vanicek** report no potential conflicts of interest.

Funding/support: None.

Role of the sponsor: None.

Acknowledgments: The authors thank all persons involved in the Group for the Study of Resistant Depression (GSRD) project and the patients who made this analysis possible.

Supplementary material: Available at PSYCHIATRIST.COM.

REFERENCES

1. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*. 2013;382(9904):1575–1586.
2. Perlis RH. Pharmacogenomic testing and personalized treatment of depression. *Clin Chem*. 2014;60(1):53–59.
3. Whiteford HA, Harris MG, McKeon G, et al. Estimating remission from untreated major depression: a systematic review and meta-analysis. *Psychol Med*. 2013;43(8):1569–1585.
4. Nierenberg AA, Katz J, Fava M. A critical overview of the pharmacologic management of treatment-resistant depression. *Psychiatr Clin North Am*. 2007;30(1):13–29.
5. Carvalho AF, Berk M, Hyphantis TN, et al. The integrative management of treatment-resistant depression: a comprehensive review and perspectives. *Psychother Psychosom*. 2014;83(2):70–88.
6. Souery D, Papakostas GI, Trivedi MH. Treatment-resistant depression. *J Clin Psychiatry*. 2006;67(suppl 6):16–22.
7. Heimann H. Therapy-resistant depressions: symptoms and syndromes: contributions to symptomatology and syndromes. *Pharmakopsychiatr Neuropsychopharmakol*. 1974;7(3):139–144.
8. Berlim MT, Turecki G. Definition, assessment, and staging of treatment-resistant refractory major depression: a review of current concepts and methods. *Can J Psychiatry*. 2007;52(1):46–54.
9. Rush AJ, Thase ME. Strategies and tactics in the treatment of chronic depression. *J Clin Psychiatry*. 1997;58(suppl 13):14–22.
10. Thase ME, Rush AJ. When at first you don't succeed: sequential strategies for antidepressant nonresponders. *J Clin Psychiatry*. 1997;58(suppl 13):23–29.
11. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for *DSM-IV* and *ICD-10*. *J Clin Psychiatry*. 1998;59(suppl 20):22–33, quiz 34–57.
12. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56–62.
13. Souery D, Amsterdam J, de Montigny C, et al. Treatment resistant depression: methodological overview and operational criteria. *Eur Neuropsychopharmacol*. 1999;9(1–2):83–91.
14. Schosser A, Serretti A, Souery D, et al. European Group for the Study of Resistant Depression (GSRD)—where have we gone so far: review of clinical and genetic findings. *Eur Neuropsychopharmacol*. 2012;22(7):453–468.
15. Souery D, Oswald P, Massat I, et al; Group for the Study of Resistant Depression. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. *J Clin Psychiatry*. 2007;68(7):1062–1070.
16. Gratten J, Wray NR, Keller MC, et al. Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nat Neurosci*. 2014;17(6):782–790.
17. Scarr E, Millan MJ, Bahn S, et al. Biomarkers for psychiatry: the journey from fantasy to fact, a report of the 2013 CINP think tank. *Int J Neuropsychopharmacol*. 2015;18(10):pyv042.
18. Roetker NS, Page CD, Yonker JA, et al. Assessment of genetic and nongenetic interactions for the prediction of depressive symptomatology: an analysis of the Wisconsin Longitudinal Study using machine learning algorithms. *Am J Public Health*. 2013;103 (suppl 1):S136–S144.
19. Chen CC, Schwender H, Keith J, et al. Methods for identifying SNP interactions: a review on variations of Logic Regression, Random Forest and Bayesian logistic regression. *IEEE/ACM Trans Comput Biol Bioinformatics*. 2011;8(6):1580–1591.
20. Kennedy SH, Downar J, Evans KR, et al. The Canadian Biomarker Integration Network in Depression (CAN-BIND): advances in response prediction. *Curr Pharm Des*. 2012;18(36):5976–5989.
21. Kautzky A, Baldinger P, Souery D, et al. The combined effect of genetic polymorphisms and clinical parameters on treatment outcome in treatment-resistant depression. *Eur Neuropsychopharmacol*. 2015;25(4):441–453.
22. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18–22.
23. Botta V, Louppe G, Geurts P, et al. Exploiting SNP correlations within random forest for genome-wide association studies. *PLoS One*. 2014;9(4):e93379.
24. García-Magariños M, López-de-Ullibarri I, Cao R, et al. Evaluating the ability of tree-based methods and logistic regression for the detection of SNP-SNP interaction. *Ann Hum Genet*. 2009;73(pt 3):360–369.
25. Sing T, Sander O, Beerenwinkel N, et al. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940–3941.
26. Gudayol-Ferré E, Guàrdia-Olmos J, Peró-Cebollero M, et al. Prediction of the time-course pattern of remission in depression by using clinical, neuropsychological, and genetic variables. *J Affect Disord*. 2013;150(3):1082–1090.
27. Gudayol-Ferré E, Herrera-Guzmán I, Camarena B, et al. Prediction of remission of depression with clinical variables, neuropsychological performance, and serotonergic/dopaminergic gene polymorphisms. *Hum Psychopharmacol*. 2012;27(6):577–586.
28. Gudayol-Ferré E, Herrera-Guzmán 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.
29. Nishioka G, Yashima H, Kiuchi Y, et al. Prediction and structural equation model of sertraline treatment response in Japanese patients with major depressive disorder. *Hum Psychopharmacol*. 2013;28(6):576–585.
30. Papakostas GI, Petersen T, Pava J, et al. Hopelessness and suicidal ideation in outpatients with treatment-resistant depression: prevalence and impact on treatment outcome. *J Nerv Ment Dis*. 2003;191(7):444–449.
31. Karp JF, Whyte EM, Lenze EJ, et al. Rescue pharmacotherapy with duloxetine for selective serotonin reuptake inhibitor nonresponders in late-life depression: outcome and tolerability. *J Clin Psychiatry*. 2008;69(3):457–463.
32. Maron E, Eller T, Vasar V, et al. Effects of bupropion augmentation in escitalopram-resistant patients with major depressive disorder: an open-label, naturalistic study. *J Clin Psychiatry*. 2009;70(7):1054–1056.
33. Sagud M, Mihaljevic-Peles A, Uzun S, et al. The lack of association between components of metabolic syndrome and treatment resistance in depression. *Psychopharmacology (Berl)*. 2013;230(1):15–21.
34. Bergman J, Miodownik C, Palatnik A, et al. Efficacy of bupropion XR in treatment-resistant elderly patients: a case series study. *Clin Neuropharmacol*. 2011;34(1):17–20.
35. Petersen T, Hughes M, Papakostas GI, et al. Treatment-resistant depression and Axis II comorbidity. *Psychother Psychosom*. 2002;71(5):269–274.
36. Bares M, Novak T, Kopecek M, et al. The early improvement of depressive symptoms as a potential predictor of response to antidepressants in depressive patients who failed to respond to previous antidepressant treatments: analysis of naturalistic data. *Eur Psychiatry*. 2012;27(7):522–527.
37. Howland RH, Rush AJ, Wisniewski SR, et al. Concurrent anxiety and substance use disorders among outpatients with major depression: clinical features and effect on treatment outcome. *Drug Alcohol Depend*. 2009;99(1–3):248–260.
38. Nasso ED, Chiesa A, Serretti A, et al. Clinical and demographic predictors of improvement during duloxetine treatment in patients with major depression: an open-label study. *Clin Drug Investig*. 2011;31(6):385–405.
39. Papakostas GI, Petersen T, Denninger J, et al. Somatic symptoms in treatment-resistant depression. *Psychiatry Res*. 2003;118(1):39–45.
40. Degner D, Haust M, Meller J, et al. Association between autoimmune thyroiditis and depressive disorder in psychiatric outpatients. *Eur Arch Psychiatry Clin Neurosci*. 2015;265(1):67–72.
41. Tundo A, de Filippis R, Proietti L. Pharmacologic approaches to treatment resistant depression: evidences and personal experience. *World J Psychiatry*. 2015;5(3):330–341.
42. Papakostas GI, Fava M, Thase ME. Treatment of SSRI-resistant depression: a meta-analysis comparing within- versus across-class switches. *Biol Psychiatry*. 2008;63(7):699–704.
43. Perlis RH, Alpert J, Nierenberg AA, et al. Clinical and sociodemographic predictors of response to augmentation, or dose increase among depressed outpatients resistant to fluoxetine 20 mg/day. *Acta Psychiatr Scand*. 2003;108(6):432–438.
44. Papakostas GI, Petersen T, Mischoulon D, et al. Functioning and interpersonal relationships as predictors of response in treatment-resistant depression. *Compr Psychiatry*. 2003;44(1):44–50.
45. Rihmer Z, Dome P, Gonda X. Antidepressant response and subthreshold bipolarity in "unipolar" major depressive disorder: implications for practice and drug research. *J Clin Psychopharmacol*. 2013;33(4):449–452.
46. Iwanami T, Maeshima H, Baba H, et al. Psychomotor agitation in major depressive disorder is a predictive factor of mood-switching. *J Affect Disord*. 2015;170:185–189.
47. Bock C, Bukh JD, Vinberg M, et al. Do stressful life events predict medical treatment outcome in first episode of depression? *Soc Psychiatry Psychiatr Epidemiol*. 2009;44(9):752–760.
48. Fang Y, Yuan C, Xu Y, et al; OPERATION Study Team. Comparisons of the efficacy and tolerability of extended-release venlafaxine, mirtazapine, and paroxetine in treatment-resistant depression: a double-blind, randomized pilot study in a Chinese

population. *J Clin Psychopharmacol*. 2010;30(4):357–364.
49. Dudek D, Rybakowski JK, Siwek M, et al. Risk factors of treatment resistance in major depression: association with bipolarity. *J Affect Disord*. 2010;126(1–2):268–271.
50. Papakostas GI, Petersen TJ, Farabaugh AH, et al. Psychiatric comorbidity as a predictor of clinical response to nortriptyline in treatment-resistant major depressive disorder. *J Clin Psychiatry*. 2003;64(11):1357–1361.
51. Petersen T, Gordon JA, Kant A, et al. Treatment resistant depression and axis I co-morbidity. *Psychol Med*. 2001;31(7):1223–1229.
52. Agid O, Lerer B. Algorithm-based treatment of major depression in an outpatient clinic: clinical correlates of response to a specific serotonin reuptake inhibitor and to trilodothyronine augmentation. *Int J Neuropsychopharmacol*. 2003;6(1):41–49.
53. Mazeh D, Shahal B, Aviv A, et al. A randomized, single-blind, comparison of venlafaxine with paroxetine in elderly patients suffering from resistant depression. *Int Clin Psychopharmacol*. 2007;22(6):371–375.
54. Romera I, Pérez V, Menchón JM, et al. Early switch strategy in patients with major depressive disorder: a double-blind, randomized study. *J Clin Psychopharmacol*. 2012;32(4):479–486.
55. Papakostas GI, Petersen TJ, Green C, et al. A description of next-step switching versus augmentation practices for outpatients with treatment-resistant major depressive disorder enrolled in an academic specialty clinic. *Ann Clin Psychiatry*. 2005;17(3):161–165.
56. Papakostas GI, Petersen T, Denninger JW, et al. Treatment-related adverse events and outcome in a clinical trial of fluoxetine for major depressive disorder. *Ann Clin Psychiatry*. 2003;15(3–4):187–192.
57. Perlis RH. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry*. 2013;74(1):7–14.

Editor's Note: We encourage authors to submit papers for consideration as a part of our Early Career Psychiatrists section. Please contact Erika F. H. Saunders, MD, at esaunders@psychiatrist.com.

See supplementary material for this article at PSYCHIATRIST.COM.

THE JOURNAL OF CLINICAL PSYCHIATRY
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF CLINICAL PSYCHOPHARMACOLOGY

Supplementary Material

Article Title: A New Prediction Model for Evaluating Treatment-Resistant Depression

Author(s): Alexander Kautzky, MD; Pia Baldinger-Melich, MD, PhD; Georg S. Kranz, MSc, PhD; Thomas Vanicek, MD; Daniel Souery, MD, PhD; Stuart Montgomery, MD, PhD; Julien Mendlewicz, MD; Joseph Zohar, MD, PhD; Alessandro Serretti, MD; Rupert Lanzenbergera; and Siegfried Kasper, MD

DOI Number: 10.4088/JCP.15m10381

List of Supplementary Material for the article

Disclaimer

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Baseline Characteristics	Study Sample (n=480)		*Excluded Sample* (n=317)
	Training Sample (n=400)	Test Sample (n=80)	
Resistance/Non-Resistance (n)	153 / 247	30 / 50	114 / 203
% Resistance	38%	37.5%	36%
Remission/Non-Remission	118/282	22/58	92/225
% Remission	29.5%	27.5%	29%
Mean Age (SD)	49.74 (14.49)	49.02 (13.26)	49.64 (13.79)
Sex (% Female)	67.75%	72.5%	69.89%
Ethnicity (% Caucasian)	98%	100%	99%

Supplementary eTable 1. Baseline characteristics of the study sample and the excluded sample. For the study sample, distinctive characteristics for the training sample (n=400) and the test sample (n=80) for the machine learning classification algorithm are provided. None of the three patient samples significantly differed from each other.

Predictor	Type & Levels	**Resistance** (n=183) *vs* **Non-Resistance** (n=297)	**Remission** (n=140) *vs* **Non-Remission** (n=340))
Sociodemographic Predictors			
Age	Metric *(mean)*	49.0 / 49.9	50.2 / 49.0
Gender	Binomial *(female)*	125 / 204	92 / 237
Ethnicity	Binomial *(caucasian)*	179 / 293	138 /334
Psychosocial Predictors			
Occupation	High Medium Low Other	22 / 44 35 / 115 26 / 51 70 / 87	23 / 43 59 / 121 23 / 54 35 / 122
Education	Legal School No Legal School Secondary Inferior Secondary Superior University	49 / 64 11 / 32 47 / 64 37 / 64 39 / 73	82 / 31 16 / 27 75 / 36 24 / 77 33 / 79
Marital Status	Single Living With Married Divorced Widowed	30 / 50 9 / 16 106 / 163 26 / 40 12 / 28	23 / 57 7 / 18 84 / 185 15 / 51 11 / 29
Number of Children	Metric *(mean)*	1.6 / 1.9	2.2 / 1.6
Cinical Features			
Severity	Moderate Severe Severe with Psychsis	120 / 151 60 / 128 3 / 18	74 / 197 53 / 135 13 / 8
Suicididality	Binomial *(yes)*	41 / 150	81 / 110
Melancholia	Binomial *(yes)*	35 / 109	41 / 103
Psychotic Features currently	Binomial *(yes)*	23 / 34	16 / 41
Psychotic Features lifetime	Binomial *(yes)*	14 / 42	26 / 30
Change of Appetite	Binomial *(yes)*	129 / 227	114 / 242
Change of Sleep	Binomial *(yes)*	156 / 232	114 / 274
Unrest	Binomial *(yes)*	162 / 227	121 / 268
Fatigue	Binomial *(yes)*	175 / 281	136 / 320

Feelings of Guilt	Binomial *(yes)*	142 / 203	110 / 235
Decision Making impaired	Binomial *(yes)*	172 / 263	130 / 305
Autoaggressive Thoughts	Binomial *(yes)*	121 / 181	95 / 207
Social Disfunctioning	Binomial *(yes)*	178 / 286	136 / 328
Personal MDD History			
Number of MDE	Metric *(mean)*	4.4 / 4.6	4.2 / 4.9
Timespan 1st to last MDE	Metric *(mean)*	12.2 / 11.2	11.8 / 11.5
Age first AD	Metric *(mean)*	39.3 / 40.4	40.8 / 39.6
Response 1st AD	Good Mediocre Little	88 / 192 65 / 68 30 / 37	104 / 176 21 / 112 15 / 52
Subsyndromal Bipolarity	Binomial *(yes)*	45 / 86	35 / 96
Axis II Comorbidity			
PD	Binomial *(yes)*	53 / 39	10 / 82
Agoraphobia	Binomial *(yes)*	17 / 13	3 / 27
Social Phobia	Binomial *(yes)*	24 / 19	5 / 38
OCD	Binomial *(yes)*	6 / 7	1 / 12
PTSD	Binomial *(yes)*	9 / 6	2 / 13
GAD	Binomial *(yes)*	18 / 10	2 / 26
Addiction & Substance Use			
History of Drug Abuse	Binomial *(yes)*	12 / 18	6 / 24
Smoking	Binomial *(yes)*	71 / 107	47 / 131
Alcohol Abuse	Binomial *(yes)*	7 / 15	8 / 14
Alcohol Addiction	Binomial *(yes)*	9 / 10	2 / 17
Any Substance Use	Binomial *(yes)*	11 / 16	6 / 21
Drug Abuse	Binomial *(yes)*	6 / 5	2 / 9
Drug Addiction	Binomial *(yes)*	7 / 2	1 / 8
Axis III Comorbidities			
Diabetes	Binomial *(yes)*	25 / 62	49 / 38
Thyroid Disorder	Binomial *(yes)*	22 / 86	56 / 52
Family History			
MDD 1° Relatives	Binomial *(yes)*	78 / 116	59 / 135

MDD 2° Relatives	Binomial *(yes)*	30 / 46	24 / 52
BD 1° Relatives	Binomial *(yes)*	12 / 26	13 / 25
BD 2° Relatives	Binomial *(yes)*	7 / 9	5 / 11
Number of Relatives MDD	Metric *(mean)*	0.6 / 0.6	0.6 / 0.6
Number of Relatives BD	Metric *(mean)*	0.1 / 0.1	0.1 / 0.1
Other Predictors			
In- or Outpatient	Binomial *(inpatient)*	116 / 156	64 / 218
Psychotherapie	Binomial *(yes)*	39 / 73	28 / 84

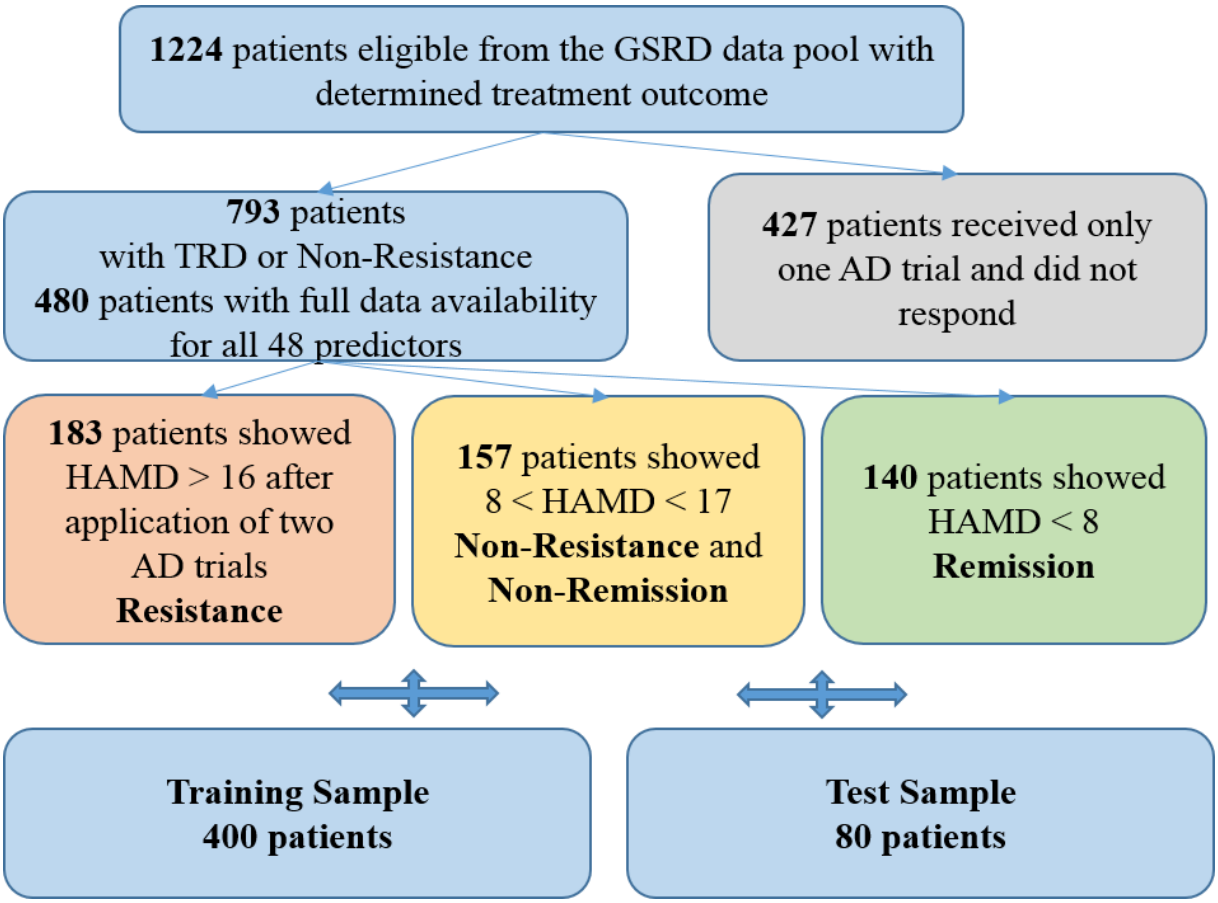
Supplementary eTable 2. List of all 48 predictors featured in the analysis ordered by groups. The second row explains the quality of the predictor and provides predictor levels. The third row shows means for the predictors among resistant and non-resistant patients. The last row shows means for the predictors among patients showing remission and non-remission. Abbreviations: MDD = major depressive disorder; MDE = major depressive episode; BD = bipolar disorder; AD = antidepressant; PTSD = posttraumatic stress disorder; GAD = generalized anxiety disorder; OCD = obstructive compulsive disorder.

Predictor		**Resistance** *(n=183)* Vs **Non-Resistance** *(n=297)*	**p =**	**Remission** *(n=140)* Vs **Non-Remission** *(n=340)*	**p =**
Thyroid Disorder	*(no)*	161 / 211	**<0.0001**	84 / 288	**>0.0001**
Panic Disorder	*(no)*	130 / 258	**<0.0001**	130 / 258	**>0.0001**
Suicidality	*(yes)*	41 / 150	**<0.0001**	81 / 110	**0.0001**
In- or Outpatient	*(inpatient)*	116 / 156	**0.0029**	64 / 218	**0.0002**
Response to 1st AD	*(good)* *(mediocre)* *(little)*	88 / 192 65 / 68 30 / 37	**0.0119**	104 / 176 21 / 112 15 / 15	**0.0006**
Number of Children	*(mean)*	1.6 / 1.9	**0.0339**	2.2 / 1.6	**0.0016**
Melancholia	*(no)*	35 / 109	**0.0001**	41 / 103	*n.s.*
Occupation	*(high)* *(medium)* *(low)* *(other)*	22 / 44 65 / 115 26 / 51 70 / 87	*n.s.*	23 / 43 53 / 97 59 / 121 35 / 122	**0.0382**
Diabetes	*(no)*	168 / 235	*n.s.*	91 / 312	**0.0027**

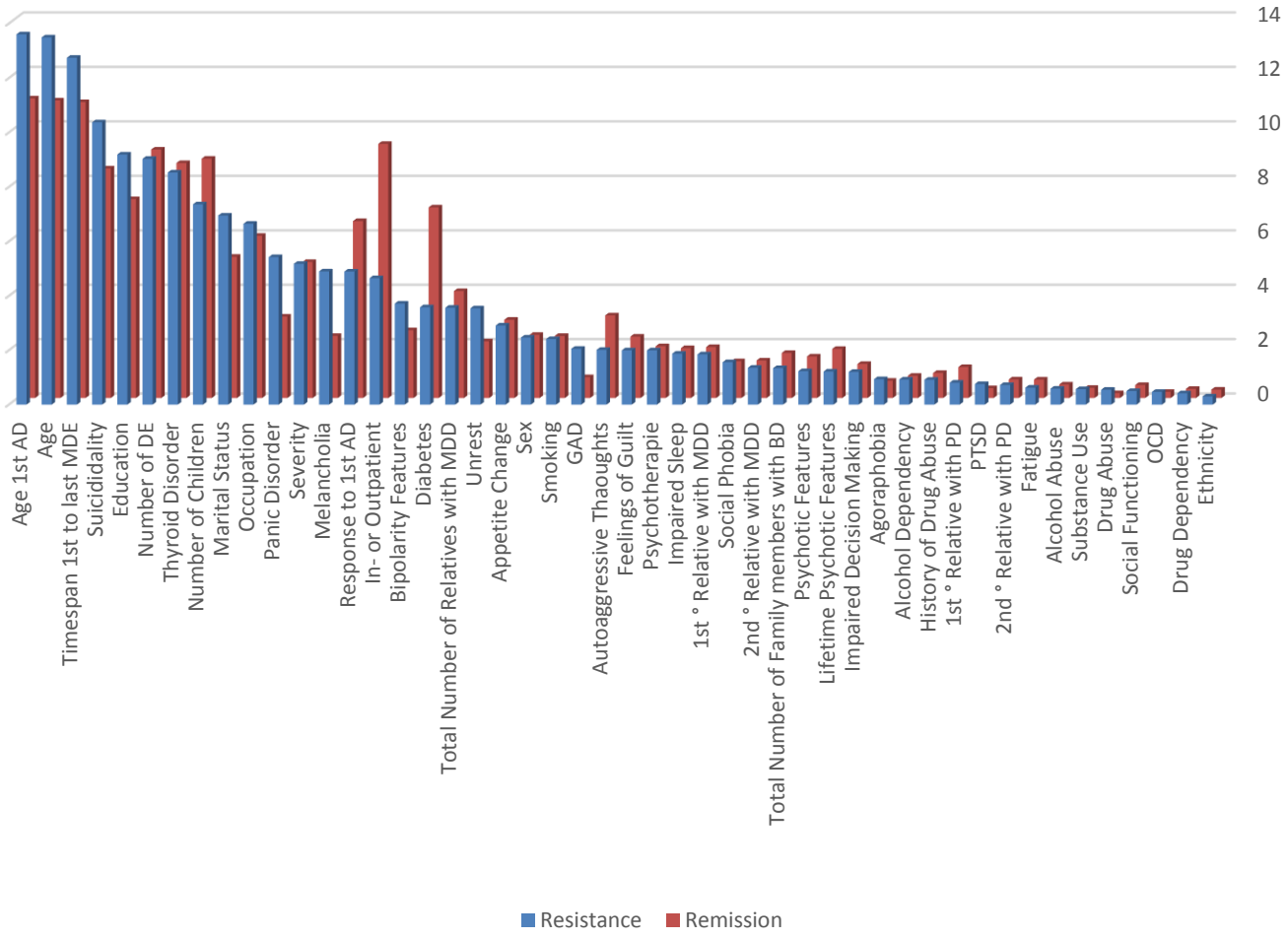
Supplementary eTable 3. Logistic regression model using the 15 most important factors for remission and resistance according to RandomForests. Predictors are ordered by p-value, Only predictors, that reached significance for either remission or resistance, are listed below. In the second and forth row, the counts for predictor levels are listed for remission and non-remission as well as resistance and non-resistance. For metric predictors, the means for are stated. Suicidality, thyroid disorder and panic disorder showed the strongest association, followed by in- or outpatient status, the response to the 1st AD administered and number of children. Melancholic depression was only significant for resistance while occupation status and diabetes were only relevant for remission. Age, education, severity, number of major depressive episodes, marital status, timespan between 1st and last depressive episode and age of 1st antidepressant treatment did not show significant associations and are not listed. Abbreviations: p = p-value for chi-square test, AD = antidepressant.

Set	Treatment Resistance	Remission
Top 5	Age 1st AD, Timespan, Age, Suicidality, Education	Age 1st AD, Timespan, Age, In- or Outpatient, Number of MDE
Top 10	+ Number of MDE, Thyroid Disorder, Number of Children, Marital Status, Occupation	+ Thyroid Disorder, Number of Children, Suicidality, Diabetes, Education
Top 15	+ Severity, Melancholia, Response to 1st AD, Panic Disorder, In- or Outpatient	+ Response to 1st AD, Occupation, Marital Status, Severity, Total Number of Relatives with MDD
Top25	+ Subsyndromal Bipolarity, Total Number of Relatives with MDD, Unrest, Diabetes, Appetite Change, Smoking, GAD, Autoaggressive Thoughts, Feelings of Guilt, Psychotherapy	+ Panic Disorder, Melancholia, Total Number of Relatives with PD, Impaired Sleep, Appetite Change, Smoking, Lifetime Psychotic Features, Autoaggressive Thoughts, Feelings of Guilt, Psychotherapy
Top 35	+ Sex, Impaired Sleep, ° Relatives with MDD, Social Phobia, 2° Relatives with MDD, Lifetime Psychotic Features, Current Psychotic Features, Impaired Decision Making, Total Number of Relatives with PD, Agoraphobia	+ Sex, ° Relatives with MDD, Social Phobia, 2° Relatives with MDD, Current Psychotic Features, Impaired Decision Making, Agoraphobia, GAD, Subsyndromal Bipolarity, Unrest
All 48	+ Alcohol Dependency, History of Drug Abuse, 1° Relatives with PD, PTSD, 2° Relatives with PD, Fatigue, Alcohol Abuse, Substance Use, Drug Abuse, Social Functioning, OCD, Drug Dependency, Ethnicity	+ Alcohol Dependency, History of Drug Abuse, 1° Relatives with PD, PTSD, 2° Relatives with PD, Fatigue, Alcohol Abuse, Substance Use, Drug Abuse, Social Functioning, OCD, Drug Dependency, Ethnicity

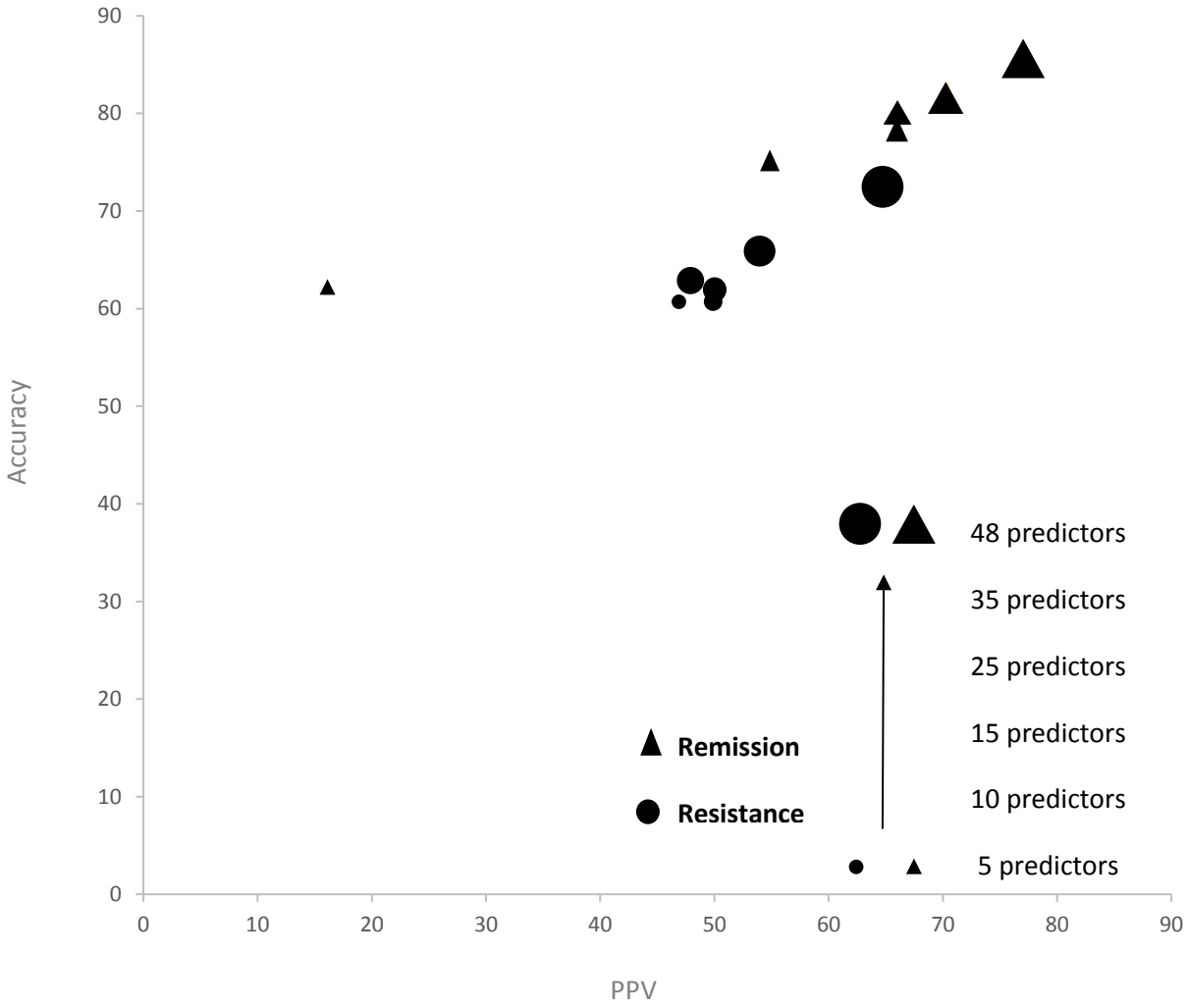
Supplementary eTable 4. Sets of predictors used for predicting resistance and remission respectively. Predictors were ordered by their importance measures for the prediction model.



Supplementary eFigure 1. Patient allocation diagram. 1224 patients of the GSRD data pool with determined treatment outcome were available. 427 of these were excluded for being non-responders indicating that they had received only one antidepressant trial to which they did not respond. Of the remaining 793 patients 480 showed full data availability, 183 of those were resistant to two antidepressant trials and 140 showed remission while 157 patients were not resistant but did not show remission. Subsequently these 480 patients were randomly assigned to the training and the test sample for the RandomForest algorithm on a quota of 5:1. Abbreviations: TRD = treatment resistant depression, HAMD = 17-item Hamilton Depression Scale; GSRD = European Group for the Study of Resistant Depression; AD = antidepressant.



Supplementary eFigure 2. Importance values measured by mean decrease in Gini for all predictors, listed on the x-axis starting with the least important. Results are shown for remission and for resistance. Mean decrease in Gini is shown on the y-axis. A higher value indicates higher usefulness for classification decisions in the prediction model. Abbreviations: MDD = major depressive disorder; MDE = depressive episode; BD = bipolar disorder; AD = antidepressant; PTSD = postraumatic stress disorder; GAD = generalized anxiety disorder; OCD = obstructive compulsive disorder.



Supplementary eFigure 3. Accuracy and negative predictive value (NPV) for the different sets of predictors. The accuracy is illustrated on the y-axis and the PPV on the x-axis. Using all 48 predictors enabled an accuracy of 0.737 for predicting resistance and 0.825 for remission. Thereby two thirds of patients predicted to stay resistant and three quarters of patients predicted to show remission can be expected to actually do so. Reducing the number of predictors subsequently weakens the performance of the prediction model and using only the 5 most important predictors leads to an accuracy of 0.612 for resistance and 0.7 for remission and about half of the predictions for resistance and most (0.84) of the predictions for remissions are wrong.