

Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study

Kautzky A, Dold M, Bartova L, Spies M, Kranz GS, Souery D, Montgomery S, Mendlewicz J, Zohar J, Fabbri C, Serretti A, Lanzenberger R, Dikeos D, Rujescu D, Kasper S. Clinical factors predicting treatment resistant depression: affirmative results from the European multicenter study

Objectives: Clinical variables were investigated in the ‘treatment resistant depression (TRD)- III’ sample to replicate earlier findings by the European research consortium ‘Group for the Study of Resistant Depression’ (GSRD) and enable cross-sample prediction of treatment outcome in TRD.

Experimental procedures: TRD was defined by a Montgomery and Asberg Depression Rating Scale (MADRS) score ≥ 22 after at least two antidepressive trials. Response was defined by a decline in MADRS score by $\geq 50\%$ and below a threshold of 22. Logistic regression was applied to replicate predictors for TRD among 16 clinical variables in 916 patients. Elastic net regression was applied for prediction of treatment outcome.

Results: Symptom severity (odds ratio (OR) = 3.31), psychotic symptoms (OR = 2.52), suicidal risk (OR = 1.74), generalized anxiety disorder (OR = 1.68), inpatient status (OR = 1.65), higher number of antidepressants administered previously (OR = 1.23), and lifetime depressive episodes (OR = 1.15) as well as longer duration of the current episode (OR = 1.022) increased the risk of TRD. Prediction of TRD reached an accuracy of 0.86 in the independent validation set, TRD-I.

Conclusion: Symptom severity, suicidal risk, higher number of lifetime depressive episodes, and comorbid anxiety disorder were replicated as the most prominent risk factors for TRD. Significant predictors in TRD-III enabled robust prediction of treatment outcome in TRD-I.

A. Kautzky¹ , M. Dold¹ ,
L. Bartova¹, M. Spies¹,
G. S. Kranz^{1,2}, D. Souery³,
S. Montgomery⁴, J. Mendlewicz⁵,
J. Zohar⁶, C. Fabbri⁷ ,
A. Serretti⁷ , R. Lanzenberger¹,
D. Dikeos⁸, D. Rujescu⁹,
S. Kasper¹ 

Significant outcome

- Four clinical factors, symptom severity, suicidal risk, higher number of lifetime depressive episodes, and comorbid anxiety disorder, were successfully replicated as predictors of treatment resistance in depression.
- Symptom severity, psychotic symptoms, suicidal risk, generalized anxiety disorder, inpatient status, higher number of antidepressants administered previously, and lifetime depressive episodes as well as longer duration of the current episode increased the risk of treatment resistance.
- The clinical variables associated with resistant depression enabled accurate prediction of treatment outcome across samples of the ‘Group for the Studies of Resistant Depression’.

Limitations

- This is a cross-sectional study, and clinical data were assessed retrospectively.
- Some clinical data as well as the treatment outcome phenotypes were coded differently in the TRD-I and TRD-III datasets.
- A wide selection of AD was prescribed to the patients, and stratification by antidepressant type was not possible for this analysis.

¹Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria, ²Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, ³Université Libre de Bruxelles and Psy Pluriel Centre Européen de Psychologie Médicale, Brussels, Belgium, ⁴Imperial College, University of London, London, UK, ⁵School of Medicine, Free University of Brussels, Brussels, Belgium, ⁶Psychiatric Division, Chaim Sheba Medical Center, Ramat Gan, Israel, ⁷Department of Biomedical and NeuroMotor Sciences, University of Bologna, Bologna, Italy, ⁸Department of Psychiatry, Athens University Medical School, Athens, Greece and ⁹University Clinic for Psychiatry, Psychotherapy and Psychosomatic, Martin-Luther-University Halle-Wittenberg, Halle, Germany

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Key words: depression; antidepressives; clinical aspects

Siegfried Kasper, Department of Psychiatry and Psychotherapy, Medical University of Vienna, Währinger Gürtel 18-20, A-1090 Vienna, Austria.

Email: sci-biopsy@meduniwien.ac.at

Accepted for publication August 15, 2018

Introduction

Major depressive disorder (MDD) currently is the leading cause of disability burden worldwide (1). Nevertheless, the repertory of antidepressants (AD) available to clinical treatment is still limited. Up to 60% of patients do not show sufficient symptom relief after the first AD trial was applied and a third of these report hardly any alleviation even when multiple ADs are administered (2, 3). Based on these shortcomings, research focused on predictive signatures of treatment resistant depression (TRD) for several decades (4, 5). However, an increasing number of competing definitions for TRD have been raised since the first categorization in 1979, making translation of findings to other patient samples and clinical routine an intricate matter. Several clinical factors have consistently been associated with TRD, but clinicians are still not able to reliably identify patients at high risk of remaining significantly ill after adequate administration of ADs. Common ground of TRD staging models is an insufficient response to at least one AD trial of adequate length and dosage, whereby patients having received only one AD are mostly labeled as non-responders (6, 7). The more rigorous definition adopted by the multinational European research consortium 'Group for the Study of Resistant Depression' (GSRD) requires at least two failed trials, either consecutive or as combination or augmentation therapy, resulting in a comparably high degree of severely ill and resistant patients in our samples (8).

Since more than 15 years, the GSRD has put emphasis on evaluation of clinical and sociodemographic predictors of TRD. A first study investigated altogether 702 MDD patients and found comorbid anxiety disorders (panic disorder and social phobia), comorbid personality disorder, suicidal risk, high symptom severity, melancholic features, more than one previous

hospitalization, recurrent depressive episodes, non-response to the first administered AD as well as an early age of onset before turning 18 to be predictors of TRD (8).

Aims of the study

Based on these previous findings by the GSRD in the TRD-I sample, this study was aimed to clarify the role of clinical predictors for treatment outcome in TRD by replication in a fresh collective, named TRD-III. In addition, the usefulness of these predictors beyond the context of TRD-III was validated in a prediction model, testing performance across the two independent GSRD datasets.

Experimental procedures

Sample description

In ten referral centers across Europe, 1410 patients were recruited from 2011 to 2016 as part of the GSRD project entitled TRD-III. The aim was to extend and substantiate the findings on predictors of TRD in a new sample of comparable size and clinical characteristics to TRD-I, which comprised also patients of a prospective extension study, termed and published as TRD-II (9). The participating countries include Italy (Bologna and Siena), Greece (Athens), Austria (Vienna), Switzerland (Geneva), Belgium (Brussels), Germany (Halle), France (Elancourt and Toulouse), and Israel (Tel Hashomer). All ethical committees of involved centers gave approval for this study, and informed consent was required for participation. A detailed description of the sample has been published recently (10). DSM-IV criteria were applied for diagnosis of MDD according to a modified version of the MINI-International Neuropsychiatric Interview 5.0.0 (MINI). Additionally, the 'Hamilton Rating Scale for Depression' (HAM-D) was completed for all

Table 1. Baseline characteristics of the whole sample ($n = 916$), comprising 333 responders and 583 TRD patients defined by change from baseline to current MADRS score. Distribution of patients across variable levels is provided for all 16 predictors as well as age and sex. For numerical predictors, mean values and standard deviations are provided. As all variables were affected from missing values, counts for each predictor are provided in brackets. Chronic refractory depression was not used as predictor but represents an alternative outcome to TRD assessed for comparability

Predictor	Response ($n = 333$)	TRD ($n = 583$)	Predictor	Response ($n = 333$)	TRD ($n = 583$)
Sex ($n = 914$)				Suicidal risk ($n = 916$)	
Female	216	374	None	237	265
Male	117	206	Low	55	114
			Moderate	17	121
Age ($n = 914$)	52.61 ± 15.9	52.65 ± 14.3	High	24	81
	Number of MDE ($n = 725$)			Duration ($n = 719$)	
Mean	3.0 ± 2.6	3.9 ± 3.0	Mean (weeks)	22.1 ± 26.4	35.9 ± 27.8
	Recurrent depression ($n = 725$)			Chronic refractory depression ($n = 509$)	
Single	67	83	Present		118
Recurrent	200	373	Absent		391
	Psychotic Symptoms ($n = 649$)			Symptom severity ($n = 843$)	
Present	7	38	Severe	130	374
Absent	256	346	Moderate	186	153
	In- or Outpatient ($n = 912$)			Hospitalization time ($n = 881$)	
Inpatient	101	243	Mean (weeks)	4.5 ± 8.4	5.1 ± 11.3
Outpatient	230	336			
	Panic Disorder ($n = 916$)			Melancholia ($n = 912$)	
Present	36	29	Present	212	386
Absent	297	542	Absent	119	193
	Social Phobia ($n = 912$)			Age of onset ($n = 878$)	
Present	8	17	Until 18	39	48
Absent	325	564	After 18	283	506
	GAD ($n = 843$)			Any somatic disorder ($n = 909$)	
Present	24	67	Present	186	312
Absent	309	514	Absent	143	269
	Thyroid disorder ($n = 909$)			Diabetes ($n = 909$)	
Present	53	81	Present	21	34
Absent	275	500	Absent	307	547
	Number of previous ADs ($n = 916$)				
0	132	105	3	20	46
1	92	179	4	14	41
2	53	176	>4	22	36

TRD, treatment resistant depression; MDE, major depressive episode; GAD, generalized anxiety disorder; AD, antidepressant drug.

subjects (11–14). The ‘Montgomery-Åsberg Rating Scale for Depression’ (MADRS) was applied for primary classification of treatment outcome, and patients were required to score a MADRS above a threshold of 22 at the beginning of treatment for the current episode. Patients had to be 18 years of age, show MDD as primary diagnosis and be free of any current substance abuse or addiction disorder except for nicotine. There was no upper age limit for this study; mean age was 52.6, and the oldest patients were 93 at study inclusion. Please refer also to Table 1 for baseline characteristics. Finally, diagnosis of severe personality disorder based on patients’ history or clinical judgement was treated as exclusion criteria in order not to confound MDD as primary target of this study. All diagnostics were performed according to the MINI.

For validation of the prediction model, another patient sample of the GSRD, labeled as TRD-I, was used as independent test set, comprising 314 patients with the required outcome phenotypes and complete registration of clinical data. This collective was described earlier (8).

Treatment outcome phenotypes

The two outcome variables, treatment response and TRD, were classified by change in MADRS scores over AD treatment for the current major depressive episode (MDE). Therefore, a baseline MADRS was assessed retrospectively for the time point of initiation of the first AD treatment administered. Baseline scores were compared to the current MADRS, assessed at study inclusion and therefore after failure or success of AD treatment was determined.

Treatment response was defined by two requirements, (i) a MADRS ≤ 21 at inclusion as well as (ii) a decline from baseline to current MADRS of $\geq 50\%$.

TRD was defined by failed treatment response after two or more consecutive AD or combination or augmentation therapy of adequate duration and dosage were administered. Over 60% of the patients received augmentation and/or combination therapies with an average of two AD agents prescribed simultaneously (10). At inclusion,

29.5% of patients received AD combination therapies and 25.7% received augmentation with antipsychotics. Four weeks of treatment and the minimal dosage recommended in the summary of product characteristics were required for adequacy of each AD trial, please see also the supplementary section for criteria and mean dosages for all ADs.

Patients who received only one AD trial, labelled as non-responders, were excluded from this analysis as it is unknown whether they would have responded to the second AD administered.

916 of the 1410 patients showed either treatment response or TRD and were eligible for this analysis. 333 (216 female, mean age 52.61 ± 15.9) of these patients showed treatment response while 583 (374 female, mean age 52.65 ± 14.3) were affected by TRD according to MADRS classification. Details of distribution and levels for all predictors of the whole sample of 916 patients can be found in Table 1. Neither of these groups nor the excluded sample differed significantly in age, sex or baseline MADRS score. A detailed description of the sample of non-responders has been provided earlier (15).

Finally, the prevalence of chronic refractory depression (CRD), a phenotype established by Souery et al. in 1999, was assessed as an additional phenotype to TRD (6). CRD was defined by at least 12 months of episode duration despite treatment based on the patients' recollection of onset of treatment for the current episode or medical history whenever possible. CRD was assessed for comparability and was not included as outcome variable or predictor in any analyses (8).

Predictors

Some of the 25 variables that were analyzed in the original investigation of clinical predictors for TRD could not be implemented in the replication analysis (8). Delayed vs. abrupt onset of the depressive episode as well as delay in treatment after diagnosis of MDD were not registered in the TRD-III data base. Diagnosis of substance use disorders and axis II disorders were exclusion criteria for TRD-III. Psychiatric comorbidities obsessive compulsive disorder ($n = 9$), posttraumatic stress disorder ($n = 12$), anorexia ($n = 4$) and bulimia ($n = 16$) were excluded from the replication analysis as they were present in considerably less patients than within the TRD-I sample (4, 8, 16). Therefore, 16 predictors were included in the analysis. These were suicidal risk (based on MINI items C1 to C9 and coded accordingly, numerical from 0 = absent, 1 = low,

2 = medium and 3 = high), the number of depressive episodes (assessed over lifetime; numerical), symptom severity (defined by clinical judgement, assessing the amount of symptoms relative to required symptoms for diagnosis of MDD according to DSM-IV-TR criteria, similar to Souery et al. 2007 (8); coded as binomial; moderate and severe), absence or presence of melancholia (binomial; based on MINI items A9a to A10g), psychotic symptoms at the current episode (binomial), duration of the current episode (numerical, in weeks; calculated for the timepoint of treatment response and onset of the current episode based on the patients' recollection or medical history whenever possible), lifetime hospitalization time (numerical, in weeks), patients status (in- vs. outpatient status, daycare was regarded as outpatient; binomial), comorbid anxiety disorders (generalized anxiety disorder (GAD), panic disorder, social phobia; all binomial) and somatic comorbidities (diabetes mellitus, thyroid disorders, any diagnosed somatic disorder; all binomial), early onset (comparing MDD before turning 18 to adult onset; binomial) and the number of AD previously administered for the current and last episodes (numerical, assessed by patients memory and, if available, clinical records).

Four predictors were coded differently than in the original investigation in TRD-I: First, the number of lifetime hospitalization was not available in TRD-III, therefore absolute number of weeks was used. Second, the absolute number of lifetime depressive episodes was included here instead of the binomial variable single vs. recurrent episodes used in TRD-I. Third, suicidal risk was coded numerical as described above in TRD-III and binomial, as presence and absence of suicidality, in TRD-I. Finally, for assessment of the predictor 'number of previous AD', we used the number of not currently administered AD within the last 12 months instead of the number of AD administered in the current and last depressive episode.

For details concerning all these variables please see Table 1.

Statistical analysis

Logistic regression as implemented in the generalized linear model function 'glm' of the statistical software 'R' (<https://www.r-project.org/>) was used to define significant predictors of TRD, similar to analyses performed in the original study in TRD-I (8). More specific, the logit function for binomial family was applied. Variables associated with TRD in the TRD-III sample were post hoc also

analyzed for the whole GSRD sample (TRD-I and TRD-III). To exploit maximal patient counts, which varied for all predictors, variables were first analyzed in univariate models and odds ratio (OR) with confidence intervals (CI) were computed for each predictor.

Based on the replication results, significant predictors were subsequently implemented in an elastic net regularized logistic regression model using the 'glmnet' package of 'R' (17–19). Regularization adds a hyperparameter lambda (λ) to regression models. Instead of minimizing the residual sum of squares, regularized models flexibly give penalty to parameters insufficiently reducing residual variance. Elastic net is a method combining the penalties of the L1 and L2 norm of ridge and least absolute shrinkage and selection operator (LASSO) regression, the two most widely used regularization techniques for logistic regression. Thereby, the quadratic error term $\|\beta\|^2$ of ridge regression is added to the LASSO formula, overcoming limitations of the two methods respectively. Elastic net shows advantages for variable selection as well as handling of highly correlated variables and has been demonstrated to fit large numbers of predictors. More specifically, 'glmnet' and 'cv.glmnet' were run with alpha = 0.5 to specify elastic net instead of LASSO (alpha = 1) or ridge regression (alpha = 0), using 10-fold cross-validation. Binomial family and deviance as measure for cross-validation performance were applied. 'cv.glmnet' was used to find the optimal λ value, incurring at the minimum of the plotted function of deviance and log values of λ and indicating the best value for prediction accuracy. To get a measure for the predictive capacity, a receiver operator characteristic space (ROC) was plotted with the 'ROCR' package of 'R' (20).

Here, prediction results were computed for the 10-fold cross-validated training sample of 602 patients of TRD-III as well as for the independent test sample TRD-I as described above (8). As only HAM-D was available for the validation sample, a single evaluation of the HAM-D with a cutoff of 16 or more was used for definition of TRD for the TRD-I sample. In order to maintain similarity of data frames, for the validation analysis HAM-D was also used to define treatment outcome for the TRD-III sample. Complete data for all variables integrated in the final model were registered for 602 patients. 309 of these patients showed TRD (205 female, mean age 51.05 ± 13.81) while 293 showed treatment response (180 female, mean age 54.37 ± 15.89). Distribution for the subsample of 602 patients included in the final model can be found in Table S1.

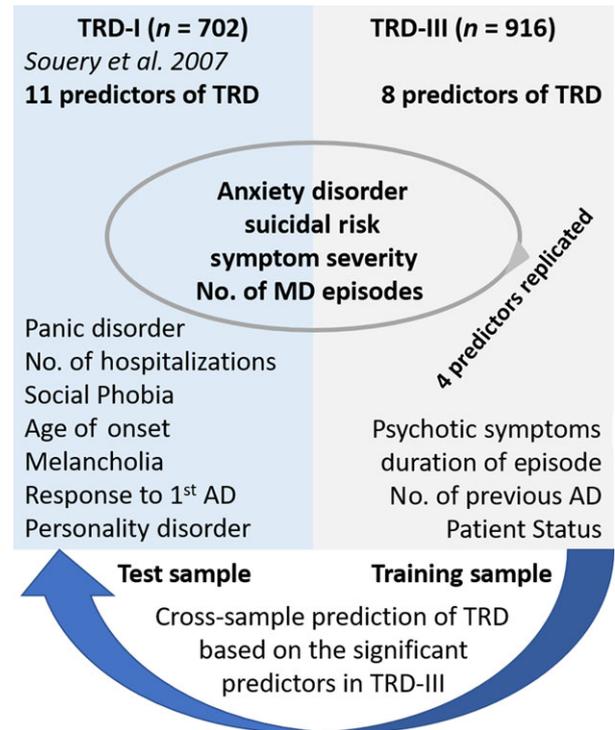


Fig. 1. Study design and summary of replication results. Predictors associated with treatment outcome for the samples TRD-I and TRD-III are listed respectively. The four predictors associated with TRD in both samples, signifying successful replication, are emphasized by the circle. Predictors associated with TRD in TRD-III were used for cross-trial prediction of treatment outcome in the independent TRD-I sample. TRD, treatment resistant depression; AD, antidepressant.

Results

Replication analysis

Logistic regression revealed severe depression ($P < 0.001$; OR = 3.31), psychotic symptoms ($P = 0.001$; OR = 2.52), low, moderate, and high suicidal risk ($P < 0.001$; OR = 1.74), GAD ($P = 0.003$; OR = 1.68), inpatient status ($P = 0.001$; OR = 1.65) increasing number of depressive episodes ($P < 0.001$; OR = 1.15), higher number of AD administered previously (OR = 1.23), and longer duration of the episode measured by weeks ($P < 0.001$; OR per SD = 1.42) as predictors of TRD after Bonferroni correction. A graphical overview of predictors associated with TRD and replication results can be found in Figure 1. In order to provide an easily interpretable OR for the episode duration, we also computed a binomial predictor, comparing patients with a duration of 3 month or longer to the rest ($P < 0.001$; OR: 2.58).

Except for psychotic symptoms and inpatient status, all of these variables were also associated with TRD in analyses in the TRD-I sample (8, 21).

Table 2. Logistic regression results for all variables available for the TRD-III and TRD-I sample that reached statistical significance in the TRD-III sample. Bold letters for predictors indicate successful replication. Original *P*-values are provided as well as odds ratio with confidence intervals; bold letters for *P*-values indicate significant results after Bonferroni correction. Differences in variable characterization between TRD samples are indicated in italic. For duration, an alternative binomial variable definition is provided

Predictor	TRD-III		TRD-I	
	Estimate	Pr. (> z)	OR (CI 5%–95%)	<i>P</i> -value & OR
Symptom severity	1.20	<0.001	3.31 (2.61–4.21)	0.001, OR 1.7
Psychotic Symptoms	1.39	0.001	2.52 (1.89–6.98)	n.s.
Suicidal Risk	0.55	<0.001	1.74 (1.54–1.98)	<0.001, OR 2.2
GAD	0.52	0.003	1.68 (1.13–2.56)	<0.001, OR 2.6
In- or Outpatient	0.50	0.001	1.65 (1.30–2.10)	n.s.
Number of MDE	0.14	<0.001	1.15 (1.09–1.21)	0.009, OR 1.5
Number of previous ADs	0.21	<0.001	1.23 (1.14–1.33)	n.s.
Duration (weeks, per SD)	0.35	<0.001	1.42 (1.23–1.66)	n.s.
Duration >3 month	0.95	<0.001	2.58 (1.94–3.44)	n.a.

GAD, generalized anxiety disorder, TRD, treatment resistant depression, MDE, major depressive episode, Pr. (>|z|), probability value according to Wald test for significance, OR, odds ratio; CI, confidence interval; SD, standard deviation; n.s., not significant; n.a., not analyzed.

Table 3. Logistic regression results for the whole GSRD sample (TRD-I and TRD-III) for the eight variables significantly associated with TRD in the TRD-III sample. Predictors are ordered by declining OR and *P*-values, and confidence intervals are provided. For better interpretability of OR, duration was coded as a binomial variable, comparing patients with an index episode longer than three months to the rest. All variables remained significant after Bonferroni correction

Predictor	Whole GSRD sample (TRD-I & TRD-III)		
	Estimate	Pr. (> z)	OR (CI 5–95%)
Psychotic symptoms	2.04	<0.0001	7.66 (4.36–13.47)
Symptom severity	0.78	<0.0001	2.18 (1.77–2.67)
Duration >3 month	0.61	<0.0001	1.85 (1.50–2.27)
GAD	0.52	0.003	1.68 (1.13–2.56)
Suicidal risk	0.41	<0.0001	1.51 (1.35–1.68)
Number of previous AD	0.34	<0.0001	1.41 (1.30–1.53)
In- or outpatient	0.29	0.004	1.34 (1.01–1.64)
Number of MDE	0.11	<0.0001	1.12 (1.07–1.16)

GAD, generalized anxiety disorder, TRD, treatment resistant depression, MDE, major depressive episode, Pr. (>|z|), probability value according to Wald test for significance, OR, odds ratio; CI, confidence interval.

Four predictors, symptom severity, suicidal risk, higher number of episodes, and comorbid anxiety disorder, were replicated from the original analysis by Souery et al. in 2007 (8). In contrast to the findings in the TRD-I sample, no significant impact was detected for early age of onset, social phobia, hospitalization time, melancholic depression, or comorbid panic disorder in the TRD-III sample. For details on logistic and linear regression results and OR with CI, please also see Table 2.

Table 4. Evaluation of binary prediction outcome. The eight variables associated with TRD in the TRD-III sample were used as predictors. Prediction was performed in a 10-fold cross-validated approach in the training sample of 602 patients as well as in a validation sample of 314 patients deriving from another sample labeled TRD-I. Comparable accuracies of 0.871 and 0.869 were observed

Model	Sensitivity	Specificity	FPR	PPV	NPV	Accuracy
CV Sample	0.945	0.778	0.222	0.818	0.931	0.871
TRD-III (n = 602)						
Validation Sample	0.857	0.876	0.124	0.793	0.917	0.869
TRD-I (n = 314)						

FPR, false positive rate; PPV, positive predictive value; NPV, negative predictive value; CV, cross-validation.

All predictors significantly associated with TRD in the TRD-III sample also reached statistical significance in the whole GSRD sample (TRD-I and TRD-III). For details on logistic regression results and OR with CI, please also see Table 3.

Prediction analysis

Based on the results of the replication analysis, we performed prediction of treatment outcome exploiting the two independent datasets TRD-I and TRD-III. An elastic net model (alpha = 0.05) was computed featuring the seven predictors described above: symptom severity, suicidal risk, GAD, number of depressive episodes and AD administered previously, duration and patient status. As psychotic symptoms were only registered for a substantial smaller part of the sample and implementation in the model would have caused a drop of observations by 22%, this predictor was excluded. Optimal λ corresponding with minimal prediction error was detected with $\lambda = 0.004$ using 10-fold cross-validation. A graphical depiction of λ and change of the coefficients according to the elastic net model can be found in Figure S2. For a graphical representation of residual deviance in dependence of lambdas and depiction of optimal lambda, please see the Supplemental Information. For the cross-validated prediction in the training set, an accuracy of 0.87 was observed. Almost similar accuracy could be computed using HAM-D instead of MADRS scores for definition of TRD to maintain interoperability of the two samples. Using the TRD-I sample as independent test set, an accuracy of 0.86 was reached. For a depiction of the ROC space for both predictions, please see Figure S3. All metrics for prediction outcome are found in Table 4.

Discussion

Ten years since our original finding of 11 clinical predictors of TRD, their general importance for

TRD remained undetermined due to lack of replication and a shortage of large clinical studies in TRD. While our study from 2007 of the TRD-I sample was the largest analysis of clinical predictors specifically aimed at TRD at its time, this affirmative study benefited from two independent samples and included 916 patients of TRD-III for replication and prediction model generation and 314 patients of TRD-I for model validation.

Among the 16 predictors included, eight were significantly associated with TRD by logistic regression. Association of symptom severity, suicidal risk, GAD, and higher number of depressive episodes were replicated, while psychotic symptoms, inpatient status as well as previously administered AD were newly linked to TRD in the GSRD sample.

Severe depression compared to moderate symptoms increased the risk of TRD times 3.3, a replication of previous findings by the GSRD and others (8, 22, 23). As symptom severity expresses more pronounced and abundant depressive symptoms, significant overlap can be expected with baseline values of recognized scores as the HAM-D or MADRS, as well as with suicidality. While not included in this analysis, the baseline MADRS was demonstrated to be an effective predictor for TRD recently (24, 25). The presence of suicidal risk on the other hand increased the risk of TRD times 1.74 per rank, showing maximal risk in highly suicidal patients. Suicidality can be regarded as a definite predictor of TRD, as has been demonstrated almost univocally (8, 21, 22, 26, 27). Inpatient status was associated with TRD as well, in concordance with previous results (22). As more severely depressed patients have a higher chance of being treated in the hospital, this predictor shows high correlation with symptom severity as 81% of inpatients show severe MDD compared to 45% of outpatients. Nevertheless, inclusion of patient status in addition to symptom severity did enhance the predictive quality of the elastic net model, suggesting an independent effect of this predictor.

Different than in our previous analysis in TRD-I, a higher number of AD administered previously within the last 12 months was associated with TRD in the TRD-III sample. Similar findings have been reported previously (28).

The association of comorbid GAD could be replicated as well; however, in this analysis, social phobia and panic disorder showed negative results (8). This might be owed to the low occurrence ($n = 25$) of social phobia and higher rates of GAD ($n = 91$) in this sample compared to the TRD-I sample. Panic disorder also showed a lower

comorbidity rate in this sample ($n = 65$), and association did not withstand correction for multiple comparison. Reflecting this limitation and comparing our results to the literature, comorbid anxiety disorders seem to be predictors for TRD; however, distinctive properties of GAD, panic disorder, and social phobia need further evaluation (8, 29–31). Based on our findings, especially GAD seems to affect TRD.

Different to our results in TRD-I, psychotic symptoms were shown to increase TRD times 2.6. Altered response rates in the subgroup of patients ever showing psychotic symptoms have been demonstrated by the GSRD before, indicating better response for non-psychotic episodes and overall worse symptom severity and higher comorbidity in patients with lifetime psychotic symptoms (32). However, another study specifically targeted to surface characteristics of melancholic and psychotic depression failed to show any differences in treatment response (21). As the sample of psychotic MDD was rather small in this analysis ($n = 45$), no definite conclusion can be drawn from our results.

Finally, two predictors describing the time course of MDD were associated with TRD. Each week of duration of the current episode increased the risk of TRD times 1.022. Thus, for each SD according to average episode duration in our sample, the risk is increased by roughly 40%. Interestingly, the mean duration of the depressive episode in TRD patients as well as the occurrence of CRD was lower in TRD-III compared to TRD-I (53.9 weeks mean duration with 30% of TRD patients showing CRD in TRD-I vs. 36 weeks with 23% of TRD patients showing CRD in TRD-III). Based on these observations and considering that ORs may be more meaningful for binomial predictors, we also computed ORs for patients with a duration of three months or longer compared to duration below 3 months. Patients with a duration of three month or longer were 2.6 times more likely to develop TRD.

Further, each additional depressive episode increasing the risk by 15%. Predictors based on MDD time course have been associated with TRD before and have been recently suggested as viable markers for prediction of long term outcome and symptom severity of MDD by several multivariate models (8, 21, 27, 33, 34). We also computed a binomial predictor for recurrent vs. single episode depression similar to the TRD-I ($P > 0.05$, OR = 1.51). Based on these results, we suggest that the absolute number of episodes may be the more advantageous predictor for TRD. Early age of onset and time of hospitalization did

not yield significant results in this analysis, contrary to our previous results and other reports (8, 27).

Regarding somatic comorbidities, no association was shown either in TRD-I or TRD-III. Contrary to reports from other groups as the National Institute of Mental Health-sponsored Sequenced Treatment Alternatives to Relieve Depression (STAR*D), the GSRD data therefore do not support contribution of somatic disorders to TRD (35, 36).

Nevertheless, our findings largely agree with extensive work performed within the STAR*D trial. Baseline symptom severity, longer index episodes, and comorbid anxiety disorders were all associated with lower remission rates in STAR*D (37). A multivariate prediction model computed for STAR*D highlighted the number of depressive episodes, psychotic symptoms, and baseline severity score among the most predictive for TRD (23). Other predictors associated with TRD here could not be compared to STAR*D as only outpatients were recruited, and severe suicidality was regarded as exclusion criterion.

However, some limitation must be addressed. Most importantly, we used a cross-sectional retrospective study design, and hence, a significant proportion of clinical data, including baseline symptom severity scores and previous AD trials, was assessed retrospectively. While previous data suggest that patients' reports on their AD treatment history are reliable, we cannot rule out that our results are biased by the cross-sectional data collection (38).

We used the same criteria for TRD and treatment response as in TRD-I, based on the Souery staging model for TRD established in 1999 (6). Sticking to this model allows for optimal comparability to the results in our older sample, however, involves some limitations as different AD classes or augmentation therapies did not affect TRD characterization, and minimal required dosage of AD treatment is considered sufficient for adequacy. Furthermore, similar to our investigation of clinical predictors for TRD in TRD-I, we did not look into remission. About 60% of responders reached remission according to a decline of MADRS below a score of 10, indicating that different results may have been obtained for the comparison remission vs. non-remission instead of response vs. TRD. On the other hand, although this study was designed as an affirmative replication of our previous analysis of clinical contributors to TRD, there were some differences in sample characteristics, design, and variable definition. While the TRD samples are comparable with regards to sex, age, number of

depressive episodes, melancholic and psychotic depression as well as severity ($P > 0.05$), there were significantly more patients with TRD in the TRD-III sample (51% TRD for TRD-I vs. 64% for TRD-III). Treatment outcome was defined primarily by change in MADRS here while it was based on a single HAM-D threshold in TRD-I. Furthermore, only 16 of the 25 factors originally examined could be implemented in the analyses due to lack of registration or too few counts in the TRD-III sample. Most importantly, the response to the 1st AD administered over lifetime and personality disorders, both of which were associated with TRD, were not available for the TRD-III sample. On the other hand, some predictors were defined differently. Divergent results were produced for time of hospitalization, which was coded as single vs. multiple hospital stays for TRD-I. This might be a better design than absolute time in weeks used in TRD-III as it is less prone to outliers. Also, severe personality disorders were an exclusion criterion for TRD-III but a predictor featured in TRD-I, probably resulting in some differences between the training and validation samples. While compelling evidence for the impact of personality disorders on antidepressant treatment outcome has been brought forward, we decided to exclude patients with severe axis II disorders to allow a clearer picture of TRD in unipolar depression and refrain from disorders challenging MDD as the primary diagnosis (8, 39). Considering the lack of routine screenings for personality disorders, resulting in evaluation of severity and exclusion based solely on clinical judgment, we cannot rule out bias in this approach. A similar rationale was chosen for substance use disorders, which were a predictor in TRD-I but an exclusion criterion for TRD-III. However, considering that only a small fraction of patients was affected by substance use disorders in TRD-I (3.1% with nonalcoholic substance dependence, 4.8% for alcohol dependency in the TRD group), we do not believe that this decision significantly impacts the results.

Another important limitation is the lack of a stringent treatment protocol as this was a retrospective cross-sectional study design. Only thresholds for dosage and time were applied but a wide selection of ADs was used by the patients. Consequently, a comparably large fraction of patients received two or more AD at the same time or augmentation therapy with lithium and antipsychotic drugs. Due to this polypharmacy, further stratification by AD type was not possible to implement in the prediction model (15). While these conditions might depict clinical routine more realistically than prospective studies with well-defined

treatment arms, accurate prediction of the efficacy of a specific AD is required to close in on precision medicine in depression. A recent study has demonstrated that predictors might differ considerably between AD agents, but more research is needed to address this question (25).

Concerning advanced statistical learning algorithms, the risk of overfitting and irreproducibility of results outside of a narrow data context has been demonstrated. On the other hand, the number of predictors relative to patient counts was favorable in this dataset, and elastic net has been demonstrated to yield valid results even when the number of variables implemented surpasses observations. Additionally, 10-fold cross-validation in the 602 TRD-III patients and validation in the independent TRD-I sample resulted in comparable accuracies above 0.8. Nevertheless, our results might be dependent on variable coding and the staging methods featured by the GSRD. Prediction outcome for the cross-validation sample was hardly affected by the outcome variable applied, and TRD based on change in MADRS scores or HAM-D score. However, training with MADRS scores for TRD-III and prediction for HAM-D score in the test sample TRD-I disrupted prediction performance and decreased accuracy to 0.56. Shortcomings in the comparability of symptom severity scores have been demonstrated recently and may explain this discrepancy in accuracy (40). In fact, some patients switched treatment outcome groups when MARDS was changed to HAM-D criteria for treatment outcome phenotype determination. Therefore, our data advocate that strict abundance of data structure is essential for reproducibility in advanced statistical learning.

Our findings in the TRD-III sample advocate the importance of the eight clinical variables comorbid anxiety disorder, symptom severity, suicidal risk, psychotic features, inpatient status, long duration of the index episode as well as a high number of previously prescribed AD and of depressive episodes as predictors for TRD. We especially emphasize the relevance of comorbid anxiety disorder, symptom severity, suicidal risk, and high number of depressive episodes as these four predictors were associated with treatment outcome in both independent samples of the GSRD, TRD-I and TRD-III. The results gain additional weight as we were able to predict TRD with an accuracy of 0.87 across the independent datasets TRD-I and TRD-III using the predictors associated with TRD in the TRD-III sample. Our prediction did not only outperform judgement based on clinical expertise or other suggested stratification tools exploiting EEG and fMRI, but

also slightly surpassed our recently deployed multivariate prediction models using machine learning and ‘RandomForest’ (27). While our previous results indicated a steady increase of prediction accuracy with the number of predictors included, here, the final model consisted only of six pre-selected variables. Logistic regression on the other hand performed poorly in our previous studies for classification of treatment outcome; however, no advanced statistics with regularized models were used. This could be attributed to differences between decision tree-based techniques as ‘RandomForest’ and logistic regression as the latter might be more vulnerable to highly correlated or less informative variables. Also, the patient count was considerably higher in this analysis, allowing better prerequisites for application of advanced statistics. On the other hand, another study on multivariate prediction models for TRD also favored regularized logistic regression over machine learning techniques (23). Which statistical methods should be applied therefore remains to be resolved.

We conclude that it is clinically meaningful that eight clinical variables which can easily be obtained in routine settings within a timeframe of a few minutes may increase assessment of prospective treatment outcome decisively. While prediction outcome might be dependent on specifics of data registration and patient selection, our replication results strongly emphasize the importance of comorbid anxiety disorder, symptom severity, suicidal risk, and the number of depressive episodes. However, no prospective study allocating patients at risk to respective treatment arms, based either on the predictors highlighted here or by other advanced statistical algorithms, has been performed so far. Nevertheless, based on our findings in TRD-I and TRD-III, we advocate faster application of augmentation therapies, ECT, or ketamine treatment in inpatients with a history of MDD with several episodes, comorbid anxiety disorders, high baseline symptom severity scores, presence of any suicidality, or psychotic features and longer duration of the current episode that were already treated with a higher number of AD for the index episode.

Conflict of Interest and Funding Sources

The Group for the Study of Resistant Depression (GRSD) was supported by an unrestricted grant from Lundbeck that had no further role in the study design, data collection, analysis and interpretation, as well as in the writing and submitting of the manuscript for publication. Dr. Dold has received a travel grant from Janssen-Cilag. Dr. Kasper received grants/research support, consulting fees and/or honoraria within the last

3 years from Angelini, AOP Orphan Pharmaceuticals AG, AstraZeneca, Eli Lilly, Janssen, KRKA-Pharma, Lundbeck, Neuraxpharm, Pfizer, Pierre Fabre, Schwabe, and Servier. Dr. Souery has received grant/research support from GlaxoSmithKline and Lundbeck, and he has served as a consultant or on advisory boards for AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Janssen, and Lundbeck. Dr. Mendlewicz is a member of the board of the Lundbeck International Neuroscience Foundation and of the advisory board of Servier. Dr. Lanzenberger received travel grants and/or conference speaker honoraria from Shire, AstraZeneca, Lundbeck A/S, Dr. Willmar Schwabe GmbH, Orphan Pharmaceuticals AG, Janssen-Cilag Pharma GmbH, and Roche Austria GmbH. Dr. Serretti is or has been consultant/speaker for Abbott, Abbvie, Angelini, AstraZeneca, Clinical Data, Boehringer, Bristol-Myers Squibb, Eli Lilly, GlaxoSmithKline, Innovapharma, Italfarmaco, Janssen, Lundbeck, Naurex, Pfizer, Polifarma, Sanofi, and Servier. Dr. Zohar has received grant/research support from Lundbeck, Servier, and Pfizer; he has served as a consultant on the advisory boards for Servier, Pfizer, Solvay, and Actelion; and he has served on speakers' bureaus for Lundbeck, GSK, Jazz, and Solvay. Dr. Montgomery has been a consultant or served on advisory boards for AstraZeneca, Bio-nesia, 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. All other authors declare that they have no conflict of interests.

References

1. WHO. World Health Report: mental health – new understanding, new hope. Geneva: WHO; 2001: p 2001.
2. 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**:1575–1586.
3. PERLIS RH. Pharmacogenomic testing and personalized treatment of depression. *Clin Chem* 2014;**60**:53–59.
4. CHOSSER 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**:453–468.
5. DOLD M, KASPER S. Evidence-based pharmacotherapy of treatment-resistant unipolar depression. *Int J Psychiatry Clin Pract* 2017;**21**:1–11.
6. SOUERY D, AMSTERDAM J, de MONTIGNY C et al. Treatment resistant depression: methodological overview and operational criteria. *Eur Neuropsychopharmacol* 1999;**9**:83–91.
7. THASE ME. Management of patients with treatment-resistant depression. *J Clin Psychiatry* 2008;**69**:e8.
8. SOUERY D, OSWALD P, MASSAT I et al. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. *J Clin Psychiatry* 2007;**68**:1062–1070.
9. SOUERY D, CALATI R, PAPAGEORGIOU K et al. What to expect from a third step in treatment resistant depression: a prospective open study on escitalopram. *World J Biol Psychiatry* 2015;**16**:472–482.
10. DOLD M, BARTOVA L, MENDLEWICZ J et al. Clinical correlates of augmentation/combination treatment strategies in major depressive disorder. *Acta Psychiatr Scand* 2018;**137**:401–412.

11. SHEEHAN DV, LECRUBIER Y, SHEEHAN KH et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): 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 4-57.
12. HAMILTON M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;**23**:56–62.
13. MONTGOMERY SA, ASBERG M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–389.
14. YOUNG RC, BIGGS JT, ZIEGLER VE, MEYER DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978;**133**:429–435.
15. DOLD M, KAUTZKY A, BARTOVA L et al. Pharmacological treatment strategies in unipolar depression in European tertiary psychiatric treatment centers – a pharmacoepidemiological cross-sectional multicenter study. *Eur Neuropsychopharmacol* 2016;**26**:1960–1971.
16. SERRETTI A, CHIESA A, CALATI R et al. Family history of major depression and residual symptoms in responder and non-responder depressed patients. *Compr Psychiatry* 2014;**55**:51–55.
17. FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22.
18. PARK MY, HASTIE T. L-1-regularization path algorithm for generalized linear models. *J Roy Stat Soc B* 2007;**69**:659–677.
19. LOCKHART R, TAYLOR J, TIBSHIRANI RJ, TIBSHIRANI R. A significance test for the lasso. *Ann Stat* 2014;**42**:413–468.
20. SING T, SANDER O, BEERENWINKEL N, LENGAUER T. ROCr: visualizing classifier performance in R. *Bioinformatics* 2005;**21**:3940–3941.
21. ZANINOTTO L, SOUERY D, CALATI R et al. Treatment resistance in severe unipolar depression: no association with psychotic or melancholic features. *Ann Clin Psychiatry* 2013;**25**:97–106.
22. BALESTRI M, CALATI R, SOUERY D et al. Socio-demographic and clinical predictors of treatment resistant depression: a prospective European multicenter study. *J Affect Disord* 2016;**189**:224–232.
23. PERLIS RH. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry* 2013;**74**:7–14.
24. PEETERS FP, RUHE HG, WICHERS M et al. The Dutch measure for quantification of treatment resistance in depression (DM-TRD): an extension of the Maudsley Staging Method. *J Affect Disord* 2016;**205**:365–371.
25. CHEKROUD AM, ZOTTI RJ, SHEHZAD Z et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 2016;**3**:243–250.
26. PAPAPOSTAS 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**:444–449.
27. KAUTZKY A, BALDINGER-MELICH P, KRANZ GS et al. A new prediction model for evaluating treatment-resistant depression. *J Clin Psychiatry* 2017;**78**:215–222.
28. DE CARLO V, CALATI R, SERRETTI A. Socio-demographic and clinical predictors of non-response/non-remission in treatment resistant depressed patients: a systematic review. *Psychiatry Res* 2016;**240**:421–430.
29. PAPAPOSTAS 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**:1357–1361.

30. 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**:457–463.
31. NASSO ED, CHIESA A, SERRETTI A, DE RONCHI D, MENCACCI C. Clinical and demographic predictors of improvement during duloxetine treatment in patients with major depression: an open-label study. *Clin Drug Investig* 2011;**31**:385–405.
32. SOUERY D, ZANINOTTO L, CALATI R et al. Phenomenology of psychotic mood disorders: lifetime and major depressive episode features. *J Affect Disord* 2011;**135**:241–250.
33. van LOO HM, CAI T, GRUBER MJ et al. Major depressive disorder subtypes to predict long-term course. *Depress Anxiety* 2014;**31**:765–777.
34. KESSLER RC, van LOO HM, WARDENAAR KJ et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry* 2016;**21**:1366–1371.
35. DEGNER D, HAUST M, MELLER J, RUTHER E, REULBACH U. Association between autoimmune thyroiditis and depressive disorder in psychiatric outpatients. *Eur Arch Psychiatry Clin Neurosci* 2015;**265**:67–72.
36. PAPA KOSTAS GI, PETERSEN T, DENNINGER J et al. Somatic symptoms in treatment-resistant depression. *Psychiatry Res* 2003;**118**:39–45.
37. 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**:28–40.
38. POSTERNAK MA, ZIMMERMAN M. How accurate are patients in reporting their antidepressant treatment history? *J Affect Disord* 2003;**75**:115–124.
39. BOCK C, BUKH JD, VINBERG M, GETHER U, KESSING LV. The influence of comorbid personality disorder and neuroticism on treatment outcome in first episode depression. *Psychopathology* 2010;**43**:197–204.
40. FRIED EI, van BORKULO CD, EPSKAMP S, SCHOEVEERS RA, TUERLINCKX F, BORSBOOM D. Measuring depression over time Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychol Assess* 2016;**28**:1354–1367.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Binomial deviance in dependence of lambda for the ten-fold cross validation model and the training data.

Figure S2. Coefficients changing by L1 Norm.

Figure S3. Receiver operator characteristics plot showing sensitivity on the y-axis and false positive rate on the X-axis.

Table S1. Baseline characteristics of the training sample (n=602), comprising 205 responders and 391 TRD patients based on a single HAM-D evaluation.

Table S2. Criteria for adequate doses for antidepressants, based on the minimal dosage described effective for monotherapy in the summary of product characteristics.

Table S3. Logistic regression results for all variables available for the TRD-III and TRD-I sample.