# The Dutch Measure for quantification of Treatment Resistance in Depression (DM-TRD)

Journal of Affective Disorders 205 (2016) 365–371

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

# The Dutch Measure for quantification of Treatment Resistance in Depression (DM-TRD): an extension of the Maudsley Staging Method

Frenk P.M.L. Peeters [a,*,1], Henricus G. Ruhe [b,c,1], Marieke Wichers [d], Latifa Abidi [a], Karin Kaub [b], H. Josephine van der Lande [b], Jan Spijker [e], Marcus J.H. Huibers [f], Aart H. Schene [b,g,h]

[a] Department of Psychiatry and Neuropsychology, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands
[b] Program for Mood Disorders, Department of Psychiatry, Academisch Medisch Centrum, University of Amsterdam, the Netherlands
[c] University of Groningen, University Medical Center Groningen, Mood and Anxiety Disorders, Department of Psychiatry, Groningen, the Netherlands
[d] University of Groningen, University Medical Center Groningen, Interdisciplinary Center for the Pathophysiology and Emotion regulation, Department of Psychiatry, Groningen, the Netherlands
[e] ProPersona Mental Healthcare, Nijmegen, the Netherlands
[f] Department of Clinical Psychology, Faculty of Psychology and Education VU University Amsterdam, the Netherlands
[g] Department of Psychiatry, Radboud University Medical Center, Nijmegen, the Netherlands
[h] Donders Institute, Radboud University, Nijmegen, the Netherlands

ARTICLE INFO

*Article history:*
Received 20 March 2016
Received in revised form
30 June 2016
Accepted 14 August 2016
Available online 16 August 2016

*Keywords:*
Treatment resistant depression
Major depressive disorder
Measurement
Profiling

ABSTRACT

*Background:* Treatment resistant depression (TRD) is common in daily practice. An empirical, widely accepted and applicable measure to quantify TRD is lacking. Previously, the Maudsley Staging Method (MSM) showed good validity. We aimed to improve the MSM by refining and extending its items resulting in the Dutch Measure for quantification of TRD (DM-TRD).
*Methods:* In addition to duration, severity and failed treatments in the current depressive episode, we added items for functional impairment, comorbid anxiety, personality disorders and psychosocial stressors. We extended the augmentation section and added items for failed psychotherapy and intensified treatment. We examined psychometric properties of the DM-TRD and tested prediction of future depressive symptoms and remission after 16 weeks in 274 (DSM-IV) depressed in- and outpatients entering naturalistic treatment.
*Results:* The DM-TRD showed excellent inter-/intra-rater reliability. Higher scores were associated with more symptoms and less remission during follow-up. The DM-TRD outperformed the MSM in prediction of future depressive symptomatology. Remission was predicted equally well by both measures. Longer duration of the current episode, larger functional impairment and larger baseline symptom severity were the strongest predictors of symptomatology at follow-up. Longer duration and larger functional impairment were negatively associated with remission.
*Limitations:* Longer follow-up could have increased predictive power. Addition of items for somatic comorbidity, childhood adversity and psychotic features must be investigated further.
*Conclusion:* The DM-TRD has excellent psychometric properties and better predictive validity for clinical outcome than other sophisticated measure published to date. Its use in clinical practice and research will improve treatment planning in TRD-patients.



## 1. Introduction

Major depressive disorder (MDD) has a significant impact on public health worldwide, in which treatment resistant depression (TRD) constitutes significantly (Greden, 2001). TRD is associated with high personal suffering, considerable functional impairment, and significant medical and mental health costs (Gibson et al., 2010; Petersen et al., 2004; Russell et al., 2004). On estimation, even after a variety of sequential treatments, at least 10–33% of depressed patients will remain clinically symptomatic (Hollon et al., 2014; Judd et al., 1998; Rush et al., 2006).

After the introduction of the concept of TRD in 1974, several staging models and measures for TRD have been proposed (Ruhe et al., 2012). However, none of these measures is thoroughly tested nor widely accepted. A widely accepted measure of TRD, applicable throughout the clinical spectrum of treatment settings for MDD, is

* Corresponding author.
[1] These authors share first authorship.

http://dx.doi.org/10.1016/j.jad.2016.08.019
0165-0327/

important for several reasons. First, it may provide a basis for realistic estimations about TRD-prevalences, as different definitions currently yield different prevalence rates. Second, it may help to identify risk factors involved in the development of TRD and its different stages. Third, prediction of high risk for TRD prior to or in an early stage of treatment would better guide therapy at an early stage , e.g. by early referral to tertiary care. Fourth, a broadly applicable measurement will improve description of characteristics and homogeneity in next-step treatment trials. This will support the identification of most efficacious next-step treatments (Ruhe et al., 2012).

### 1.1. Current staging measures

Until recently, measures for TRD only consisted of rating of previous unsuccessful biological treatments for the current episode (Berlim and Turecki, 2007; Ruhe et al., 2012). Some of them relied on dichotomous classifications, e.g. ≥2 adequate trials of antidepressants, at odds with the dimensional positive correlation with the number of previous unsuccessful treatments (Ruhe et al., 2006; Rush, 2006). From a clinical perspective, a measure for TRD should be able to i) dimensionally classify patients according to their level of resistance to MDD-treatment, ii) predict chances of future treatment response and iii) guide further treatment selection. Such instrument mimic staging and profiling-measures that are increasingly used in other areas of medicine like oncology, which aim to predict course and prognosis, supporting clinicians in selecting the most effective treatment for a particular patient (Hetrick et al., 2008; Kraemer et al., 2002).

Given that reliable and clinically useful biomarkers for MDD are unavailable (Kapur et al., 2012), a TRD-measure must be complemented with clinical variables known to be associated with treatment response in MDD. Fekadu et al. (2012, 2009a, 2009b) proposed such a measure, the Maudsley Staging Method (MSM), with predictive validity tested in a sample of 88 patients with high levels of TRD. The MSM consists of rating of biological treatment failures for the current episode combined with episode severity and duration.

Although this instrument is an important step forward in the development of an ideal measure, additional parameters might further optimize the quantification of TRD and increase its ability to prospectively predict treatment outcome. For example, the MSM does not incorporate functional impairment, presence of ongoing psychosocial stressors, nor psychiatric co-morbidity, all examples of factors that are associated with unfavorable treatment response. Furthermore, the MSM does not adequately reflect current widespread pharmacological augmentation strategies, nor does it include evidence-based psychotherapies or intensified interventions like clinical and/or daytime treatment.

Building on the work of Fekadu and colleagues and bearing the previous considerations in mind, we developed a modification of the MSM by refining and extending its elements. We first describe this modification; the Dutch Measure for quantification of Treatment Resistance in Depression (DM-TRD). Subsequently, we tested psychometric qualities of the DM-TRD. We finally tested its predictive validity in comparison to the MSM in a naturalistic-treated sample of depressed in- and outpatients with a follow-up of 16 weeks. We hypothesized that the DM-TRD would be a reliable list in terms of good inter- and intra-rater reliability. Second, we hypothesized that the DM-TRD would be predictive for treatment outcome at 16 weeks of treatment, in a comparable or better way than the MSM.

**Table 1**
The Dutch Measure for quantification of Treatment Resistance in Depression.

| Item and specification | Score |
|---|---|
| *Episode duration*[a] | |
| Acute | 1 |
| Sub-acute | 2 |
| Chronic | 3 |
| *Symptom severity*[1,a] | |
| Subsyndromal | 1 |
| Mild | 2 |
| Moderate | 3 |
| Severe without psychosis | 4 |
| Severe with psychosis | 5 |
| *Functional impairment*[1] | |
| No impairment (GAF 90–100) | 0 |
| Mild impairment (GAF 60–90) | 1 |
| Moderate impairment (GAF 30–60) | 2 |
| Severe impairment (GAF < 30) | 3 |
| *Comorbid anxiety symptoms* | |
| Not present | 0 |
| Present, but not fulfilling DSM-IV criteria | 0.5 |
| Fulfilling criteria ≥1 DSM-IV anxiety disorder | 1 |
| *Comorbid personality disorder* | |
| Not present | 0 |
| Present; not based on formal interview | 0.5 |
| Present; based on formal interview | 1 |
| *Psychosocial stressors*[2] | |
| No psychosocial stressor | 0 |
| ≥1 Psychosocial stressor | 1 |
| *Treatment failures*[a] | |
| *Antidepressants*[3] | |
| Level 0: not used | 0 |
| Level 1: 1–2 medications | 1 |
| Level 2: 3–4 medications | 2 |
| Level 3: 5–6 medications | 3 |
| Level 4: 7–10 medications | 4 |
| Level 5: >10 medications | 5 |
| *Augmentation/combination*[3,a] | |
| Level 0: not used | 0 |
| Level 1: 1–2 medications | 1 |
| Level 2: 3–4 medications | 2 |
| Level 3: 5–6 medications | 3 |
| *Electroconvulsive therapy*[3,4,a] | |
| Not used | 0 |
| Used | 1 |
| *Psychotherapy*[3] | |
| Not used | 0 |
| Supportive therapy | 0.5 |
| 1 empirically supported psychotherapy[5] | 1 |
| ≥2 empirically supported psychotherapies[5] | 2 |
| *Intensified treatment* | |
| Not used | 0 |
| Daypatient treatment[6] | 1 |
| Inpatient treatment[7] | 2 |
| **Total** | **(27)** |

[a] also included in MSM; identical rating, except that the MSM only considers augmentation as no/yes (0/1) condition.
[1] Maximum severity during current episode.
[2] Based on DSM-IV axis-IV.
[3] Counted only when given in the present depressive episode.
[4] Minimum of 8 sessions ECT.
[5] Minimum of 12 sessions.
[6] ≥12 weeks; ≥3 days/week.
[7] ≥4 weeks.

## 2. Methods

### 2.1. Grounds for the DM-TRD as a modification of MSM

Table 1 shows the DM-TRD. For each item a scoring is proposed, resulting in a total score between 2 and 27.

#### 2.1.1. Duration and severity

Duration and severity of an episode of MDD are strong predictors of treatment response. In the MSM, these variables

independently predicted treatment resistance (Fekadu et al., 2009b); we therefore retained these variables.

#### 2.1.2. Functional impairment

Although this may be a proxy for depression severity, we added functional impairment, because of its association with factors that influence course and treatment response that are indirectly related to depression, especially in chronic MDD (Frank et al., 2011). Patients with identical levels of depression severity may function differently in their daily lives because of differences in social support or economic resources (Verboom et al., 2012, 2011). For simplicity and acceptability, we operationalized functional impairment based on the GAF-scale of the DSM-IV (APA, 1994, 2000), which is routinely assessed in clinical practice.

#### 2.1.3. Comorbid anxiety symptoms

MDD is frequently accompanied by comorbid axis-I disorders, mostly in the anxiety spectrum. Because co-morbid anxiety symptoms influence course and treatment response (Fava et al., 2008; Penninx et al., 2011), we included this as a separate categorical variable, based on the absence/presence of anxiety symptoms, or the presence of a diagnosed DSM-IV comorbid anxiety disorder.

#### 2.1.4. Comorbid personality disorder

Comorbid personality disorder has traditionally been assumed to increase the risk of TRD. Although recent studies and meta-analyses are equivocal (De Bolle et al., 2011; Kool et al., 2005; Markowitz et al., 2007; Mulder, 2002; Newton-Howes et al., 2014; Skodol et al., 2011), we included the presence of co-morbid personality pathology to acknowledge this possible association with TRD. We distinguished three levels, acknowledging different grades of diagnostic validity: no personality disorder; clinically suspected co-morbid personality disorders, without formal structured interview; and a validly diagnosed personality disorder according to a structured interview (e.g. Structured Clinical Interview for DSM-IV (SCID)).

#### 2.1.5. Psychosocial stressors

Ongoing psychosocial stressors might influence the outcome of treatment of MDD (Brown et al., 2010; Leskela et al., 2006). To keep this criterion user-friendly and clearly operationalized, we dichotomized the absence/presence of current stressors.

#### 2.1.6. Treatment failure

Treatment resistance can only be adequately assessed when previous failed treatments are considered. All previous measures acknowledged that when the first treatment steps fail, remission rates of subsequent trials decrease. There is no definite evidence for a hierarchical order in which switching to other classes of antidepressants is superior to a within-class switch (Ruhe et al., 2006; Rush et al., 2006); this should therefore not be suggested in measures for TRD. For augmentation-combination strategies, evidence for superiority of any specific augmentation strategy-combination relative to each other is currently also lacking. Augmenting an ineffective antidepressant with another drug appears as effective as switching to a different antidepressant (Rush et al., 2006; Trivedi et al., 2006) and augmentation-combination possibilities are rapidly increasing (Papakostas, 2009; Zhou et al., 2015). We therefore classified both switching antidepressant medication and augmentation/combination counts into 5 and 3 levels, respectively. Like in the MSM, treatments are counted only when given in the present depressive episode and when adequate dosages have been prescribed for $\geq 4$ weeks.

The MSM's score for ECT remained unchanged. We added scores for previous psychotherapy treatment, as well for the present episode, because meta-analyses and current MDD-guidelines clearly advocate psychotherapy as efficacious as pharmacotherapy (Cuijpers et al., 2013; NICE, 2009; Parikh et al., 2009). We hierarchically distinguished supportive therapy as opposed to empirically supported psychotherapy (ESP) like cognitive behavioral therapy (CBT) or interpersonal psychotherapy (IPT). An adequate trial of ESP was defined as $\geq 12$ sessions; for patients who tried different forms of ESP, we added a next level.

We added intensified treatment like day-time/clinical treatment and hierarchically operationalized this. Intensified treatment, combining different types of treatments, is considered to be of additional benefit in more severe and complex MDD-patients (Brakemeier et al., 2015; Zobel et al., 2011). Admission in an inpatient setting ($\geq 4$ weeks to exclude mere crisis interventions) was considered more intensive than daytime treatment ($\geq 3$ days/week for $\geq 12$ weeks).

### 2.2. Procedure

We tested the measure in two independent samples of MDD-patients referred for treatment with a current episode of MDD to i) a university-affiliated outpatient treatment center (Maastricht, the Netherlands), or ii) the daytime/inpatient unit for (treatment resistant) depression of the Academic Medical Center (Amsterdam, the Netherlands). We included in-, daytime and outpatients to examine the DM-TRD in a sample with a wide range in TRD-levels. Because the study used existing, anonymized data, judgement of a Medical Ethics Committees was waved.

Our only inclusion criterion was a primary diagnosis of MDD. In the outpatients this was determined with the SCID for Axis I (First et al., 1995) and in the daytime/inpatient sample by a clinical interview using DSM-IV criteria. Exclusion criteria were primary diagnoses other than MDD (e.g., bipolar disorder; psychotic disorder or substance abuse); and insufficient fluency of Dutch. When diagnostic work-up revealed symptoms indicative of co-morbid personality pathology, the SCID-II (First et al., 1994) was administered routinely in the outpatients. In the daytime/inpatient sample formal axis-II diagnostic instruments were also administered after clinical indication. After diagnostic work-up, all participants entered a naturalistic treatment phase according to local protocolized treatment programs.

In order to determine the inter- and intra-rater reliability of the DM-TRD, for the outpatient-sample, one of the authors (LA), blinded for clinical outcome, scored the items of the DM-TRD based on retrospective information from participants' medical files. The first author (FP), also blind to clinical outcome, independently re-rated a random sample of 25 files to obtain an Intra-class Correlation Coefficient (ICC). Three months later the rater (LA) reevaluated the same 25 files. The data from the daytime/inpatient sample was abstracted likewise by two other authors (KK and HJvdL), supervised by HGR.

### 2.3. Outcome measures

Primary outcome measures were symptom severity over time and remission as assessed with the Beck Depression Inventory (BDI) (Beck et al., 1996) in Maastricht, and the Inventory of Depressive Symptomatology ($IDS\text{-}SR_{30}$) (Rush et al., 1996) in Amsterdam. In both samples, the BDI or $IDS\text{-}SR_{30}$ was administered prior to treatment (baseline) and at 8 and 16 weeks after start of treatment, obtained in regular routine outcome assessments done in these treatment settings. $IDS\text{-}SR_{30}$ scores were converted to BDI-scores according to Vittengl et al. (2005). Remission was defined as a BDI-score $< 10$ at week 16; response as $\geq 50\%$ reduction from baseline score.

### 2.4. Statistical analysis

Differences in BDI, DM-TRD and MSM scores between treatment settings were examined with *t*-tests. Prediction by the DM-TRD and MSM of outcome at week 16 during treatment were tested with linear and logistic regression for continuous BDI-scores and remission at week 16, respectively. In order to compare beta's and Odds Ratios (ORs) between the DM-TRD and MSM, these were calculated as standardized beta's and for 1S.D. of change in the scale, respectively. In order to assess whether the total DM-TRD score predicted outcome better than the model with the individual items, we compared the Akaike Information Criterion (AIC). The AIC enables comparisons of goodness of fit of a model penalized for the number of parameters describing the model (smaller is better): $AIC = 2^*k + n^*\log(RSS/n)$, with n = number of subjects, k = number of parameters, RSS = Residual sum of squares of the model. For exploratory reasons, we finally determined which items of the DM-TRD contributed most to the predictions by using a forward stepwise regression (entry at $p < 0.05$).

## 3. Results

### 3.1. Participants

We studied 274 MDD-patients (165 women) with a mean age of $44 \pm 11$ years. Most participants (63%) were diagnosed with recurrent MDD. Table 2 displays mean BDI at baseline and follow-up, mean MSM and DM-TRD scores for the total sample and the two subsamples. Differences between both samples on MDD-severity and DM-TRD and MSM scores were significant (all $p < 0.001$). Remission rates in the outpatient- and daytime/inpatient samples at 16 weeks were 27% and 15% respectively. Response rates at 16 weeks were 33% and 29%, respectively.

### 3.2. Inter- and Intra-rater reliability of the DM-TRD

In the outpatient- and daytime/inpatient samples, the ICC was excellent (0.97 ($CI_{95}$ 0.96–0.98) and 0.90 ($CI_{95}$ 0.77–0.96), respectively). The three months intra-rater reliability was 0.98 ($CI_{95}$ 0.97–0.98).

### 3.3. Predictive validity of the measures

Table 3 displays the frequency distributions of the proposed items. The total scores on the DM-TRD and MSM were significantly associated with BDI-scores and remission rates at week 16 during follow-up. Higher scores of DM-TRD (standardized ß = 0.33, $F_{1,252} = 31.21$, $p < 0.001$; $r^2 = 0.11$) and MSM (standardized ß = 0.29, $F_{1,252} = 23.81$, $p < 0.001$; $r^2 = 0.09$) were associated with higher BDI-scores during follow-up. When both measures were entered simultaneously, only the DM-TRD score ($t = 2.62$, $p = 0.009$) remained significant (MSM; $t = -0.36$, $p = 0.72$).

**Table 3**
Characteristics of participants on DM-TRD items (n = 274).

| Item | N | % |
|---|---|---|
| *Duration* | | |
| Acute | 162 | 59 |
| Sub-acute | 62 | 23 |
| Chronic | 50 | 18 |
| *Symptom severity* | | |
| Subsyndromal | – | – |
| Mild | 27 | 10 |
| Moderate | 136 | 50 |
| Severe without psychosis | 97 | 35 |
| Severe with psychosis | 14 | 5 |
| *Functional impairment* | | |
| No impairment | – | – |
| Mild impairment | 17 | 6 |
| Moderate impairment | 254 | 93 |
| Severe impairment | 3 | 1 |
| *Comorbid anxiety symptoms* | | |
| Not present | 145 | 53 |
| Present, but not fulfilling DSM-IV criteria | 49 | 18 |
| Fulfilling criteria for ≥1 DSM-IV anxiety disorder | 80 | 30 |
| *Comorbid personality disorder* | | |
| Not present | 137 | 50 |
| Present, but not based on formal interview | 112 | 41 |
| Fulfilling criteria based on formal interview | 25 | 9 |
| *Psychosocial stressors* | | |
| Not present | 11 | 4 |
| ≥1 Psychosocial stressor | 263 | 96 |
| *Treatment failures (current episode)* | | |
| *Antidepressants* | | |
| Level 0: not used | 103 | 38 |
| Level 1: 1–2 medications | 151 | 55 |
| Level 2: 3–4 medications | 16 | 6 |
| Level 3: 5–6 medications | 4 | 1 |
| Level 4: 7–10 medications | – | – |
| Level 5: >10 medications | – | – |
| *Augmentation* | | |
| Level 0: not used | 238 | 87 |
| Level 1: 1–2 medications | 35 | 13 |
| Level 2: 3–4 medications | 1 | 0.4 |
| Level 3: 5–6 medications | – | – |
| *Electroconvulsive therapy* | | |
| Not used | 272 | 99 |
| Used | 2 | 1 |
| *Psychotherapy* | | |
| Not used | 159 | 58 |
| Supportive therapy | 86 | 31 |
| Empirically supported psychotherapy | 26 | 10 |
| ≥2 Empirically supported psychotherapies | 3 | 1 |
| *Intensified treatment* | | |
| Not used | 257 | 94 |
| Day-treatment | 3 | 1 |
| Inpatient treatment | 14 | 5 |

Both measures predicted remission at week 16 equally well (DM-TRD: $OR_{SD} = 0.53$; $CI_{95}$ 0.44–0.63; $p < 0.001$; $r^2 = 0.08$ and MSM: $OR_{SD} = 0.53$; $CI_{95}$ 0.43–0.67; $p = 0.001$; $r^2 = 0.08$). When entered simultaneously, both measures were unable to predict remission ($p > 0.33$), indicative of high collinearity.

**Table 2**
Mean (S.D.) scores on BDI, MSM and DM-TRD in both samples.

| | Total sample (n = 274) | Daytime/Inpatient sample (n = 103) | Outpatient sample (n = 171) | Difference between samples |
|---|---|---|---|---|
| BDI baseline | 29.3 (9.4) | 34.9 (7.8) | 25.9 (8.6) | t = 8.7; 95% CI 7.0–11.2*** |
| BDI at 8 weeks FU | 21.8 (11.2) | 26.2 (10.3) | 19.7 (11.0) | t = 4.4; 95% CI 3.6–9.3*** |
| BDI at 16 weeks FU | 18.7 (11.6) | 23.0 (12.0) | 16.5 (10.8) | t = 4.3; 95% CI 3.5–9.4*** |
| MSM | 5.8 (1.6) | 7.0 (1.5) | 5.1 (1.2) | t = 7.6; 95% CI 1.4–2.3*** |
| DM-TRD | 9.8 (2.1) | 10.9 (2.1) | 9.1 (1.7) | t = 10.8; 95% CI 1.5–2.2*** |

BDI = Beck Depression Inventory, MSM = Maudsley Staging Method, DM-TRD = Dutch Method for quantification of TRD.
*** $p < 0.001$.

**Table 4**
Association of scores on DM-TRD items and BDI scores at 16 week follow-up.

| *Clinical variables* | Standardized β | T |
|---|---|---|
| Duration | 0.2517 | 2.54* |
| Symptom severity | 0.08 | 1.19 |
| Functional impairment | 0.13 | 2.05* |
| Comorbid anxiety | 0.06 | 0.95 |
| Comorbid personality disorder | −0.02 | −0.37 |
| Psychosocial stressors | 0.12 | 1.95^ |
| *Treatment failures (current episode)* | | |
| Antidepressants | 0.11 | 1.40 |
| Augmentation/combination | −0.01 | −0.13 |
| Electroconvulsive therapy | 0.01 | 0.13 |
| Psychotherapy | 0.01 | 0.12 |
| Intensified treatment | 0.12 | 1.87^ |

* $p < 0.05$.
^ $p < 0.07$.

Next, we examined how individual items were associated with outcome defined as the BDI-score at 16 weeks. Table 4 displays the full regression model of the association between baseline scores on the DM-TRD items and BDI-score. The AIC of the model with the DM-TRD total score was 529.18 relative to 544.86 when all items were entered individually, indicating that the total score was more parsimonious. The items 'duration', and 'functional impairment' were significantly associated with outcome at 16 weeks, while 'psychosocial stressors' and 'intensified treatment' were associated at a trend level ($p > 0.07$). In exploratory stepwise regression models, the items 'duration' ($t=4.22$; $p < 0.001$), 'functional impairment'($t=2.22$; $p=0.028$), and 'symptom severity' ($t=2.03$; $p=0.043$) remained as predictors of follow-up BDI-scores with an explained variance ($r^2$) of 10.8%. For remission, the remaining DM-TRD items were 'duration' (OR=0.50, $CI_{95}$ 0.29–0.85; $p=0.01$) and 'functional impairment' (OR=0.29, $CI_{95}$ 0.10–0.89; $p < 0.03$; $r^2=10.9\%$).

## 4. Discussion

This study examined the DM-TRD, a new dimensional staging measure for TRD in MDD. The DM-TRD is able to provide a reliable, dimensional classification of treatment resistance, to detect difference in TRD levels between out- and more treatment resistant daytime/inpatient samples, and to predict both severity of future depressive symptomatology and remission. Moreover, outcomes for our measure as a whole and some of its separate items suggest that it may be an even better predictive tool for clinical outcome than the MSM (Fekadu et al., 2009a), the most sophisticated measure to date. Additional strengths of the current study are a large, thoroughly diagnosed out- and daytime/inpatient sample that displays a wide range in its level of TRD.

In line with previous work (Fekadu et al., 2009b), both measures predicted clinical outcome at odds with a recent inpatient study that was unable to predict very short (2 weeks) clinical outcome based on measures of TRD (Icick et al., 2014). Additionally, given the significant differences in scores between the two samples, both measures may be of value for treatment planning; e.g. indication for combined/intensified treatments or early referral to tertiary care. Supportive of this conclusion, Hazari et al. also differentiated tiers of treatment based on quantification of TRD (Hazari et al., 2013).

In addition to the overall DM-TRD score, some items of the DM-TRD appeared particularly associated with clinical outcome. We replicated the role of well-known predictors like duration of the index episode and symptom severity in outcome. Functional impairment, although to some extent a proxy for symptom severity, was associated independently with outcome. This finding additionally underlines the importance of assessing functional impairment in MDD-patients, apart from MDD-severity (Bentley et al., 2014; Ito et al., 2015; Zimmerman et al., 2011). Although we included other well-known factors associated with unfavorable outcome (e.g. comorbid anxiety or personality disorders), in our exploratory analyses none of these items significantly predicted outcome. This can be explained by a lack of power and/or smaller (corrected) effect-sizes for such items in our multivariate models, despite the rather large sample-size. Nevertheless, combination of items in one total DM-TRD score was more parsimonious than acknowledging items separately. Given our a-priori expectation that the total of items would predict outcome better, we consider our exploratory analyses as indicative for future refinements of the DM-TRD, to be investigated in new, larger samples.

There is a debate about the optimal duration of treatments like antidepressants, augmentation/combination strategies but also psychotherapies. For example, the NICE guideline (www.nice.org.uk/guidance/qs8) states that a minimum period of 6–8 weeks is required to assess effectiveness, however other reports advocate earlier switching (after 2 weeks) when insufficient early response is achieved (Szegedi et al., 2009; Tadic et al., 2016). Also in the Dutch multidisciplinary guideline for MDD switching is advised when after 4–8 weeks no response occurred (Spijker and Nolen, 2010). In the absence of clear evidence about the optimal duration of treatments, we decided to copy the $\geq 4$ weeks criterion for pharmacotherapy from the MSM and –potentially arbitrarily- defined the duration of an adequate evidence-based psychotherapy as $\geq 12$ sessions. Improved consensus about the optimal duration of treatments for MDD in the future, will require further exploration of the effects of adaptation of these criteria for the predictive properties of the MSM and DM-TRD.

With the present report, we examined the predictive properties of the DM-TRD and MSM-scale in a range of MDD-patients covering a broader scope of mental healthcare facilities than in the initial publication of the MSM (Fekadu et al., 2012, 2009a, 2009b). The fact that both the MSM and the DM-TRD predict nonresponse to various forms of treatment, across a wider range of MDD-patients and settings is supportive to accept these scales as a broadly applicable measurement.

We think implementation of the DM-TRD would both improve clinical treatment planning and the description of characteristics of next-step treatment trials. First, when an easily applicable instrument like the DM-TRD is used to recognize patients at baseline as being potentially treatment resistant. This information will influence treatment planning, e.g. by closer monitoring of treatment results, earlier referral to tertiary care when a first treatment step fails, intensified psychoeducation (also regarding treatment resistance and ways to avoid this) and/or closer monitoring of compliance. Whether such earlier recognition results in better outcomes needs to be assessed in prospectively studies. From a research perspective, the heterogeneity of clinical trials in next-step treatment approaches is enormous which confounds efficacy rates (Ruhe et al., 2006, 2012). When all future studies on treatment resistance would use a measure to quantify TRD, this would facilitate comparability and eventually allow and meta-regression approaches.

### 4.1. Limitations

Despite being the largest study of prediction of TRD on clinical variables to date, some limitations must be acknowledged. First, our follow-up period was short, which may explain the relatively low remission/response rates, especially in the inpatient sample. Longer follow-up, e.g. like in the investigations of the MSM

(Fekadu et al., 2012, 2009a, 2009b), could have resulted in an increased predictive power of both instruments, but was unavailable. Second, the DM-TRD does not include the presence of somatic co-morbidity (Rayner et al., 2010; van Milligen et al., 2012), childhood adversity (Tunnard et al., 2014; Wiersma et al., 2009) and sub threshold psychotic features (Perlis et al., 2011; Wigman et al., 2014), all candidate variables also negatively associated with treatment response. These may be added to improve prediction of TRD. Third, measuring p sychosocial stressors is complex, while we simplified it to absence/presence of current psychosocial stressors. The reliable determination whether an event is a precursor or a consequence of MDD, and quantify the personal valence/impact of an event presumably requires structured interviews (Brown and Harris, 1978). Nevertheless, this simple, clinically applicable operationalization showed a trend to predict BDI-scores. Fourth, the scoring of the different items of the DM-TRD as semi-continuous scales (with an *a-priori* determined range and inter-item distance) does not necessarily reflect the optimal solution to explain most of the variance. In order to investigate an optimal scoring range within items, alternative fitting procedures might be helpful, for which larger and replication datasets are needed. Fifth, for the assessment of the level of functioning the GAF might not be the best multidimensional scale. The Functioning Assessment Short Test (FAST; Rosa et al., 2007), Work and Social Adjustment Scale (WSAS; Mundt et al., 2002) or the World Health Organization Disability Assessment Schedule (WHODAS-II; Chwastiak and Von Korff, 2003) might be more valid measures. We pragmatically used the GAF-scale, as it is a clinically acceptable routine DSM-IV assessment. Future research should address the improvement of the DM-TRD by inclusion of alternative scales for assessment of functioning. Sixth, the study relied on retrospective information drawn from extensive medical files available in university-affiliated outpatient treatment centers. Despite the accuracy during the diagnostic work-up and good inter-rater reliability, relevant information may have been obscured, requiring prospective studies.

## 5. Conclusion

The DM-TRD is a multidimensional measure for quantification of (future) TRD, based on clinical variables in the current MDD-episode. Due to its practical applicability, reliability and validity, this instrument may be useful for clinicians and can be used in future TRD-research to characterize treated populations. To further establish its utility, the DM-TRD will have to be tested prospectively. In new, larger samples the impact of various weights given to the items might be assessed to further improve predictive properties. The DM-TRD may serve as a starting point for a staging and profiling approach based on psychopathological and biological markers that, like in other areas of medicine may individually predict course and prognosis of the disorder (Hetrick et al., 2008). This will ultimately assist clinicians and patients to select the most appropriate treatment options.

## Potential conflicts of interest

The authors report no financial or other relationship relevant to the subject of this article.

## Financial support

This research was performed without additional funding. Dr. H. G. Ruhé is supported by a NWO/ZonMW VENI-Grant #016.126.059.

## Acknowledgment

None.

## References

APA, 1994. Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association, Washington, DC.
APA, 2000. Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR®). American Psychiatric Association, Washington DC.
Beck, A.T., Steer, R.A., Ball, R., Ranieri, W., 1996. Comparison of beck depression inventories -IA and -II in psychiatric outpatients. J. Pers. Assess. 67, 588–597.
Bentley, K.H., Gallagher, M.W., Carl, J.R., Barlow, D.H., 2014. Development and validation of the overall depression severity and impairment scale. Psychol. Assess. 26, 815–830.
Berlim, M.T., Turecki, G., 2007. Definition, assessment, and staging of treatment-resistant refractory major depression: a review of current concepts and methods. Can. J. Psychiatry 52, 46–54.
Brakemeier, E.L., Radtke, M., Engel, V., Zimmermann, J., Tuschen-Caffier, B., Hautzinger, M., Schramm, E., Berger, M., Normann, C., 2015. Overcoming treatment resistance in chronic depression: a pilot study on outcome and feasibility of the cognitive behavioral analysis system of psychotherapy as an inpatient treatment program. Psychother. Psychosom. 84, 51–56.
Brown, G., Harris, T., 1978. Social Origins of Depression. Free Press, London.
Brown, G.W., Harris, T.O., Kendrick, T., Chatwin, J., Craig, T.K., Kelly, V., Mander, H., Ring, A., Wallace, V., Uher, R., 2010. Antidepressants, social adversity and outcome of depression in general practice. J. Affect. Disord. 121, 239–246.
Chwastiak, L.A., Von Korff, M., 2003. Disability in depression and back pain: evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting. J. Clin. Epidemiol. 56 , pp. 507–514.
Cuijpers, P., Sijbrandij, M., Koole, S.L., Andersson, G., Beekman, A.T., Reynolds 3rd, C. F., 2013. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 12, 137–148.
De Bolle, M., De Fruyt, F., Quilty, L.C., Rolland, J.P., Decuyper, M., Bagby, R.M., 2011. Does personality disorder co-morbidity impact treatment outcome for patients with major depression?: a multi-level analysis. J. Pers. Disord. 25, 1–15.
Fava, M., Rush, A.J., Alpert, J.E., Balasubramani, G.K., Wisniewski, S.R., Carmin, C.N., Biggs, M.M., Zisook, S., Leuchter, A., Howland, R., Warden, D., Trivedi, M.H., 2008. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. Am. J. Psychiatry.
Fekadu, A., Rane, L.J., Wooderson, S.C., Markopoulou, K., Poon, L., Cleare, A.J., 2012. Prediction of longer-term outcome of treatment-resistant depression in tertiary care. Br. J. Psychiatry 201, 369–375.
Fekadu, A., Wooderson, S., Donaldson, C., Markopoulou, K., Masterson, B., Poon, L., Cleare, A.J., 2009a. A multidimensional tool to quantify treatment resistance in depression: the Maudsley staging method. J. Clin. Psychiatry 70, 177–184.
Fekadu, A., Wooderson, S.C., Markopoulou, K., Cleare, A.J., 2009b. The Maudsley Staging Method for treatment-resistant depression: prediction of longer-term outcome and persistence of symptoms. J. Clin. Psychiatry 70, 952–957.
First, M., Spitzer, R., Gibbon, M., Williams, J., Benjamin, L., 1994. Structured Clinical Interview for DSM-IV Axis II personality disorders (SCID II). Biometric Research Department, New York.
First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 1995. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I). Biometrics Research Department, New York State Psychiatric Institute, New York.
Frank, E., Cassano, G.B., Rucci, P., Thompson, W.K., Kraemer, H.C., Fagiolini, A., Maggi, L., Kupfer, D.J., Shear, M.K., Houck, P.R., Calugi, S., Grochocinski, V.J., Scocco, P., Buttenfield, J., Forgione, R.N., 2011. Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. Psychol. Med. 41, 151–162.
Gibson, T.B., Jing, Y., Smith Carls, G., Kim, E., Bagalman, J.E., Burton, W.N., Tran, Q.V., Pikalov, A., Goetzel, R.Z., 2010. Cost burden of treatment resistance in patients with depression. Am. J. Manag. Care 16, 370–377.
Greden, J.F., 2001. The burden of disease for treatment-resistant depression. J. Clin. Psychiatry 62 (Suppl. 16), S26–S31.
Hazari, H., Christmas, D., Matthews, K., 2013. The clinical utility of different quantitative methods for measuring treatment resistance in major depression. J. Affect. Disord. 150, 231–236.
Hetrick, S.E., Parker, A.G., Hickie, I.B., Purcell, R., Yung, A.R., McGorry, P.D., 2008. Early identification and intervention in depressive disorders: towards a clinical staging model. Psychother. Psychosom. 77, 263–270.
Hollon, S.D., DeRubeis, R.J., Fawcett, J., Amsterdam, J.D., Shelton, R.C., Zajecka, J., Young, P.R., Gallop, R., 2014. Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. JAMA Psychiatry 71, 1157–1164.
Icick, R., Millet, E., Curis, E., Bellivier, F., Lepine, J.P., 2014. Predictive value of baseline resistance in early response to antidepressants. J. Affect. Disord. 161, 127–135.
Ito, M., Bentley, K.H., Oe, Y., Nakajima, S., Fujisato, H., Kato, N., Miyamae, M., Kanie, A., Horikoshi, M., Barlow, D.H., 2015. Assessing depression related severity and functional impairment: the overall depression severity and impairment scale (ODSIS). PLoS One 10, e0122969.

Judd, L.L., Akiskal, H.S., Maser, J.D., Zeller, P.J., Endicott, J., Coryell, W., Paulus, M.P., Kunovac, J.L., Leon, A.C., Mueller, T.I., Rice, J.A., Keller, M.B., 1998. A prospective 12-year study of subsyndromal and *syn*dromal depressive symptoms in unipolar major depressive disorders. Arch. Gen. Psychiatry 55, 694–700.

Kapur, S., Phillips, A.G., Insel, T.R., 2012. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol. Psychiatry.

Kool, S., Schoevers, R., de Maat, S., Van, R., Molenaar, P., Vink, A., Dekker, J., 2005. Efficacy of pharmacotherapy in depressed patients with and without personality disorders: a systematic review and meta-analysis. J. Affect. Disord. 88, 269–278.

Kraemer, H.C., Wilson, G.T., Fairburn, C.G., Agras, W.S., 2002. Mediators and moderators of treatment effects in randomized clinical trials. Arch. Gen. Psychiatry 59, 877–883.

Leskela, U., Rytsala, H., Komulainen, E., Melartin, T., Sokero, P., Lestela-Mielonen, P., Isometsa, E., 2006. The influence of adversity and perceived social support on the outcome of major depressive disorder in subjects with different levels of depressive symptoms. Psychol. Med. 36, 779–788.

Markowitz, J.C., Skodol, A.E., Petkova, E., Cheng, J., Sanislow, C.A., Grilo, C.M., Gunderson, J.G., McGlashan, T.H., 2007. Longitudinal effects of personality disorders on psychosocial functioning of patients with major depressive disorder. J. Clin. Psychiatry 68, 186–193.

Mulder, R.T., 2002. Personality pathology and treatment outcome in major depression: a review. Am. J. Psychiatry 159, 359–371.

Mundt, J.C., Marks, I.M., Shear, M.K., Greist, J.H., 2002. The work and social adjustment scale: a simple measure of impairment in functioning. Br. J. Psychiatry 180, 461–464.

Newton-Howes, G., Tyrer, P., Johnson, T., Mulder, R., Kool, S., Dekker, J., Schoevers, R., 2014. Influence of personality on the outcome of treatment in depression: systematic review and meta-analysis. J. Pers. Disord. 28, 577–593.

NICE, 2009. Depression: the Treatment and Management of Depression in Adults (update). National Institute for Clinical Excellence, London.

Papakostas, G.I., 2009. Managing partial response or nonresponse: switching, augmentation, and combination strategies for major depressive disorder. J. Clin. Psychiatry 70 (Suppl. 6), S16–S25.

Parikh, S.V., Segal, Z.V., Grigoriadis, S., Ravindran, A.V., Kennedy, S.H., Lam, R.W., Patten, S.B., 2009. Canadian Network for Mood and Anxiety Treatments (CANMAT) clinical guidelines for the management of major depressive disorder in adults. II. Psychotherapy alone or in combination with antidepressant medication. J. Affect. Disord. 117 (Suppl. 1), S15–S25.

Penninx, B.W., Nolen, W.A., Lamers, F., Zitman, F.G., Smit, J.H., Spinhoven, P., Cuijpers, P., de Jong, P.J., van Marwijk, H.W., der Meer, K.V., Verhaak, P., Laurant, M. G., de Graaf, R., Hoogendijk, W.J., der Wee, N.V., Ormel, J., van Dyck, R., Beekman, A.T., 2011. Two-year course of depressive and anxiety disorders: results from the Netherlands Study of Depression and Anxiety (NESDA). J. Affect Disord.

Perlis, R.H., Uher, R., Ostacher, M., Goldberg, J.F., Trivedi, M.H., Rush, A.J., Fava, M., 2011. Association between bipolar spectrum features and treatment outcomes in outpatients with major depressive disorder. Arch. Gen. Psychiatry 68, 351–360.

Petersen, T., Papakostas, G.I., Mahal, Y., Guyker, W.M., Beaumont, E.C., Alpert, J.E., Fava, M., Nierenberg, A.A., 2004. Psychosocial functioning in patients with treatment resistant depression. Eur. Psychiatry 19, 196–201.

Rayner, L., Price, A., Evans, A., Valsraj, K., Higginson, I.J., Hotopf, M., 2010. Antidepressants for depression in physically ill people. Cochrane Database Syst. Rev.

Rosa, A.R., Sanchez-Moreno, J., Martinez-Aran, A., Salamero, M., Torrent, C., Reinares, M., Comes, M., Colom, F., Van Riel, W., Ayuso-Mateos, J.L., Kapczinski, F., Vieta, E., 2007. Validity and reliability of the Functioning Assessment Short Test (FAST) in bipolar disorder. Clin. Pr. Epidemiol. Ment. Health 3, 5.

Ruhe, H.G., Huyser, J., Swinkels, J.A., Schene, A.H., 2006a. Switching antidepressants after a first selective serotonin reuptake inhibitor in major depressive disorder: a systematic review. J. Clin. Psychiatry 67, 1836–1855.

Ruhe, H.G., van Rooijen, G., Spijker, J., Peeters, F.P., Schene, A.H., 2012. Staging methods for treatment resistant depression. A systematic review. J. Affect. Disord. 137, 35–45.

Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol. Med. 26, 477–486.

Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Nierenberg, A.A., Stewart, J.W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D., McGrath, P.J., Rosenbaum, J.F., Sackeim, H.A., Kupfer, D.J., Luther, J., Fava, M., 2006b. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am. J. Psychiatry 163, 1905–1917.

Russell, J.M., Hawkins, K., Ozminkowski, R.J., Orsini, L., Crown, W.H., Kennedy, S., Finkelstein, S., Berndt, E., Rush, A.J., 2004. The cost consequences of treatment-resistant depression. J. Clin. Psychiatry 65, 341–347.

Skodol, A.E., Grilo, C.M., Keyes, K.M., Geier, T., Grant, B.F., Hasin, D.S., 2011. Relationship of personality disorders to the course of major depressive disorder in a nationally representative sample. Am. J. Psychiatry 168, 257–264.

Spijker, J., Nolen, W.A., 2010. An algorithm for the pharmacological treatment of depression. Acta Psychiatr. Scand. 121, 180–189.

Szegedi, A., Jansen, W.T., van Willigenburg, A.P., van der Meulen, E., Stassen, H.H., Thase, M.E., 2009. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6562 patients. J. Clin. Psychiatry 70, 344–353.

Tadic, A., Wachtlin, D., Berger, M., Braus, D.F., van Calker, D., Dahmen, N., Dreimuller, N., Engel, A., Gorbulev, S., Helmreich, I., Kaiser, A.K., Kronfeld, K., Schlicht, K.F., Tuscher, O., Wagner, S., Hiemke, C., Lieb, K., 2016. Randomized controlled study of early medication change for non-improvers to antidepressant therapy in major depression - The EMC trial. Eur. Neuropsychopharmacol. 26, 705–716.

Trivedi, M.H., Fava, M., Wisniewski, S.R., Thase, M.E., Quitkin, F., Warden, D., Ritz, L., Nierenberg, A.A., Lebowitz, B.D., Biggs, M.M., Luther, J.F., Shores-Wilson, K., Rush, A.J., 2006. Medication augmentation after the failure of SSRIs for depression. N. Engl. J. Med. 354, 1243–1252.

Tunnard, C., Rane, L.J., Wooderson, S.C., Markopoulou, K., Poon, L., Fekadu, A., Juruena, M., Cleare, A.J., 2014. The impact of childhood adversity on suicidality and clinical course in treatment-resistant depression. J. Affect. Disord. 152–154, 122–130.

van Milligen, B.A., Vogelzangs, N., Smit, J.H., Penninx, B.W., 2012. Physical function as predictor for the persistence of depressive and anxiety disorders. J. Affect. Disord. 136, 828–832.

Verboom, C.E., Ormel, J., Nolen, W.A., Penninx, B.W., Sijtsema, J.J., 2012. Moderators of the synchrony of change between decreasing depression severity and disability. Acta Psychiatr. Scand. 126, 175–185.

Verboom, C.E., Sentse, M., Sijtsema, J.J., Nolen, W.A., Ormel, J., Penninx, B.W., 2011. Explaining heterogeneity in disability with major depressive disorder: effects of personal and environmental characteristics. J. Affect. Disord. 132, 71–81.

Vittengl, J.R., Clark, L.A., Kraft, D., Jarrett, R.B., 2005. Multiple measures, methods, and moments: a factor-analytic investigation of change in depressive symptoms during acute-phase cognitive therapy for depression. Psychol. Med. 35, 693–704.

Wiersma, J.E., Hovens, J.G., van Oppen, P., Giltay, E.J., van Schaik, D.J., Beekman, A.T., Penninx, B.W., 2009. The importance of childhood trauma and childhood life events for chronicity of depression in adults. J. Clin. Psychiatry 70, 983–989.

Wigman, J.T., van Os, J., Abidi, L., Huibers, M.J., Roelofs, J., Arntz, A., Kelleher, I., Peeters, F.P., 2014. Subclinical psychotic experiences and bipolar spectrum features in depression: association with outcome of psychotherapy. Psychol. Med. 44, 325–336.

Zhou, X., Ravindran, A.V., Qin, B., Del Giovane, C., Li, Q., Bauer, M., Liu, Y., Fang, Y., da Silva, T., Zhang, Y., Fang, L., Wang, X., Xie, P., 2015. Comparative efficacy, acceptability, and tolerability of augmentation agents in treatment-resistant depression: systematic review and network meta-analysis. J. Clin. Psychiatry 76, e487–e498.

Zimmerman, M., Galione, J.N., Attiullah, N., Friedman, M., Toba, C., Boerescu, D.A., Ragheb, M., 2011. Depressed patients' perspectives of 2 measures of outcome: the Quick Inventory of Depressive Symptomatology (QIDS) and the Remission from Depression Questionnaire (RDQ). Ann. Clin. Psychiatry 23, 208–212.

Zobel, I., Kech, S., van Calker, D., Dykierek, P., Berger, M., Schneibel, R., Schramm, E., 2011. Long-term effect of combined interpersonal psychotherapy and pharmacotherapy in a randomized trial of depressed patients. Acta Psychiatr. Scand. 123, 276–282.