Psychological Medicine

cambridge.org/psm

Can psychological features predict antidepressant response to rTMS? A Discovery–Replication approach

Noralie Krepel[1,2], A. John Rush[3,4,5], Tabitha A. Iseger[2,6], Alexander T. Sack[1,*] and Martijn Arns[2,6,7,*]

[1]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands; [2]Research Institute Brainclinics, Nijmegen, The Netherlands; [3]Duke-National University of Singapore, Singapore; [4]Duke Medical School, Durham, NC, USA; [5]Texas Tech University Health Sciences Center, Permian Basin, TX, USA; [6]Department of Experimental Psychology, Utrecht University, Utrecht, The Netherlands and [7]neuroCare Group Netherlands, Nijmegen, The Netherlands

Original Article

*Shared last author.

Cite this article: Krepel N, Rush AJ, Iseger TA, Sack AT, Arns M (2020). Can psychological features predict antidepressant response to rTMS? A Discovery–Replication approach. *Psychological Medicine* **50**, 264–272. https://doi.org/10.1017/S0033291718004191

Received: 1 June 2018
Revised: 17 December 2018
Accepted: 21 December 2018
First published online: 24 January 2019

Key words:
Major depressive disorder; psychological measures; rTMS; treatment prediction

Author for correspondence:
Martijn Arns, E-mail: martijn@brainclinics.com

Abstract

Background. Few studies focused on the relationship between psychological measures, major depressive disorder (MDD) and repetitive transcranial magnetic stimulation (rTMS) response. This study investigated several psychological measures as potential predictors for rTMS treatment response. Additionally, this study employed two approaches to evaluate the robustness of our findings by implementing immediate replication and full-sample exploration with strict *p*-thresholding.

Methods. This study is an open-label, multi-site study with a total of 196 MDD patients. The sample was subdivided in a Discovery (60% of total sample, $n = 119$) and Replication sample (40% of total sample, $n = 77$). Patients were treated with right low frequency (1 Hz) or left high frequency (10 Hz) rTMS at the dorsolateral prefrontal cortex. Clinical variables [Beck Depression Inventory (BDI), Neuroticism, Extraversion, Openness Five-Factor Inventory, and Depression, Anxiety, and Stress Scale, and BDI subscales] were obtained at baseline, post-treatment, and at follow-up. Predictors were analyzed in terms of statistical association, robustness (independent replication), as well as for their clinical relevance [positive predictive value (PPV) and negative predictive value (NPV)].

Results. Univariate analyses revealed that non-responders had higher baseline anhedonia scores. Anhedonia scores at baseline correlated negatively with total BDI percentage change over time. This finding was replicated. However, anhedonia scores showed to be marginally predictive of rTMS response, and neither PPV nor NPV reached the levels of clinical relevance.

Conclusions. This study suggests that non-responders to rTMS treatment have higher baseline anhedonia scores. However, anhedonia was only marginally predictive of rTMS response. Since all other psychological measures did not show predictive value, it is concluded that psychological measures cannot be used as clinically relevant predictors to rTMS response in MDD.

Introduction

Major depressive disorder (MDD) is a chronic mental disease with a remitting and relapsing course. Repetitive transcranial magnetic stimulation (rTMS) as a treatment method for MDD has been studied thoroughly over the past few years. High frequency (HF, 10 Hz) rTMS applied to the left dorsolateral prefrontal cortex (DLPFC) (Schutter, 2009) as well as low frequency (LF, 1 Hz) rTMS applied to the right DLPFC (Schutter, 2010) showed to have antidepressant effects. Additionally, left and right DLPFC stimulation seem to have similar clinical effects (Fitzgerald *et al.*, 2003, 2009). rTMS also showed to be effective in treatment-resistant depression (Gaynes *et al.*, 2014). However, even though rTMS is well accepted as a treatment option for MDD, response rates remain relatively low, ranging from 29.3% (Berlim *et al.*, 2014) to 58% (Carpenter *et al.*, 2012) in HF-rTMS, which is similar to a study that employed LF as well as HF rTMS [response rates 45% and 44%, respectively (Fitzgerald *et al.*, 2009)]. Finding psychological measures that are associated with treatment response may help to identify those patients who have a greater chance of achieving response. Additionally, finding predictors that can enhance treatment allocation accuracy might increase response rates by immediately indicating the optimal treatment to a given patient, thereby saving time and money.

Substantial research has been done investigating the relationship between behavior, depression, and treatment response. Frequently studied domains include the 'Big Five Personality Traits' [neuroticism, extraversion, openness, conscientiousness, and agreeableness (Goldberg, 1990)], anhedonia, depression severity, stress, and anxiety. A robust finding seems to be the relationship between the personality traits neuroticism and extraversion, wherein neuroticism

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

seems to be positively associated to MDD (Jylhä and Isometsä, 2006; Griffith *et al.*, 2010; Kotov *et al.*, 2010; Rosellini and Brown, 2011; Hayward *et al.*, 2013), and extraversion seems to be inversely associated with MDD (Jylhä and Isometsä, 2006; Kotov *et al.*, 2010; Rosellini and Brown, 2011; Hayward *et al.*, 2013). Other studies have also elaborated on the association between personality traits and MDD by including treatment response. For example, Bagby and colleagues found that MDD patients with higher scores on neuroticism are more likely to respond to pharmacotherapy, rather than to cognitive behavioral therapy (Bagby *et al.*, 2008). Similarly, Quilty and colleagues found that neuroticism was indicative of a lower probability of response, whereas conscientiousness was predictive of a higher probability of response to combined pharmacotherapy and psychotherapy (Quilty *et al.*, 2008). A review by Mulder evaluated antidepressant response in a variety of treatments, including psychotherapy, pharmacotherapy, electroconvulsive therapy (ECT), or a combination thereof, and reported that higher neuroticism generally predicts worse treatment outcome, especially in the long-term (Mulder, 2002). However, not many studies focused on rTMS as a treatment. Berlim and colleagues demonstrated that neuroticism was found to decrease during rTMS treatment, yet lacked predictive value, whereas baseline extraversion levels predicted greater treatment response (Berlim *et al.*, 2013). In a study that focused on deep TMS (dTMS), higher agreeableness and higher conscientiousness were observed in patients who achieved remission (McGirr *et al.*, 2014).

Another psychological dimension that gained interest as a predictor of treatment outcome in MDD is anhedonia. Anhedonia is a core MDD symptom. Recently, the importance of considering the role of anhedonia in MDD has been highlighted (Treadway and Zald, 2011; Pizzagalli, 2014). It has been argued that anhedonia is a difficult symptom to treat (Treadway and Zald, 2011) and multiple studies have shown that higher levels of anhedonia are predictive of poorer treatment outcome (Spijker *et al.*, 2001; McMakin *et al.*, 2012). Likewise, improvements in anhedonia levels predicted increased psychosocial functioning in patients with MDD, which is in turn an important feature of treatment response and remission (Vinckier *et al.*, 2017). In an rTMS study, Downar and colleagues found that their groups of MDD non-responders to rTMS treatment were marked by more anhedonic symptoms (Downar *et al.*, 2014). Likewise, a recent study by Rostami and colleagues found that loss of interest (an anhedonia-related symptom) predicted rTMS treatment response (Rostami *et al.*, 2017).

Next to such individual symptoms, overall depression severity is also considered an important treatment response predictor, with higher pretreatment depression severity being associated with lower response rates (Croughan *et al.*, 1988; Trivedi *et al.*, 2006), however, the interaction between depression severity and treatment response is unclear. For example, in a placebo-controlled study Fournier and colleagues found that at mild to moderate levels of MDD, beneficial effects of antidepressants were minimal to none, however, at very severe levels of depression severity, there was a substantial benefit of the usage of medications (Fournier *et al.*, 2010). A similar trend of response was observed in those with high depression severity when assigned to the treatment or placebo group. That is, those with high depression severity that were assigned to the treatment group were more likely to have a greater response (i.e. the higher the severity, the greater the response), whereas those with high severity that were assigned to the placebo group were more likely to have a smaller response (i.e. the higher the severity, the smaller the response) (Khan *et al.*, 2002). For rTMS, it has been reported that younger patients with a lower baseline depression severity had a modestly better treatment outcome (Carpenter *et al.*, 2012). Likewise, Fitzgerald and colleagues found that rTMS responders had lower baseline depression severity, however, it did not sufficiently influence response rates to base treatment decisions on (Fitzgerald *et al.*, 2016).

Finally, anxiety and stress are also associated with MDD and treatment response. Higher levels of anxiety have been associated with greater depression severity (Fava *et al.*, 2004; Uher *et al.*, 2011) and lower response to pharmacological treatment (Fava *et al.*, 2008). It has even been proposed that depression with increased levels of anxiety could be a distinguished subtype of MDD (Fava *et al.*, 2004). However, the association between higher anxiety levels and decreased response to treatment is modest (Joffe *et al.*, 1993) and inconsistently replicated, including reports of patients with anxious depression responding better to ketamine treatment (Ionescu *et al.*, 2014), or reports in which anxious depression is not found to be predictive of worse pharmacological treatment response (Uher *et al.*, 2011). In rTMS studies, it has been found that rTMS non-responders had higher baseline anxiety than responders (Brakemeier *et al.*, 2007). For stress, it has been suggested that chronic stress is predictive of depression, even more so than acute stressors (McGonagle and Kessler, 1990; Hammen, 2005). Some studies support this, see for example Deng *et al.* (2018), in which greater perceived stress in remitted older (age $\geqslant$ 60) people predicted recurrence of depression.

Given the above richness of available psychological factors that might be predictive of antidepressant treatment response, yet the lack of such research for rTMS, makes the purpose of this study to test various psychological factors that can predict rTMS treatment non-response in MDD. Finding such predictors in psychological measures is the most cost-effective way to optimize treatment allocation. To explore all possible psychological factors, this study used the total Beck Depression Inventory (BDI)-II-NL score, as well as subscales thereof, the Neuroticism, Extraversion, Openness Five-Factor Inventory (NEO-FFI), and the Depression, Anxiety, and Stress Scale (DASS) to investigate potential predictors of rTMS treatment outcome. The BDI subscales were literature-based. Next to this, the clinical relevance of these predictors was explored. That is, next to sensitivity and specificity, the positive predictive value (PPV) and negative predictive value (NPV) were examined. This method has been applied elsewhere [e.g. Kuk *et al.* (2010); Li *et al.* (2012)] and attempts to identify constructs that can accurately and reliably inform the therapist on treatment response and therefore treatment (dis)continuation (Li *et al.*, 2012), while incorporating the false positives and false negatives. Additionally, the predictors were tested for their robustness by immediate replication in an independent sample. More specifically, given the recently highlighted interest for the replication of studies (Open Science Collaboration, 2012; Pashler and Wagenmakers, 2012; Patil *et al.*, 2016), and our own recent non-replication of previous work (Krepel *et al.*, 2018), it was decided to a priori split the complete dataset into a Discovery and Replication set. Hereby, it was possible to confirm or deny any findings in the Discovery set by executing the same analyses in the Replication set.

Methods and materials

Design

This study was a multi-site, open-label study. Data were collected at three sites (Brainclinics Treatment/neuroCare Nijmegen and The

https://doi.org/10.1017/S0033291718004191 Published online by Cambridge University Press

Hague, Psychologenpraktijk Timmers, Oosterhout, The Netherlands) between May 2007 and November 2016. Inclusion criteria were: (1) a primary diagnosis of major depressive or dysthymic disorder as confirmed by M.I.N.I. (M.I.N.I. Plus Dutch version 5.0.0), (2) a BDI-II of 14 or higher, and (3) a left DLPFC HF (10 Hz) rTMS or a right DLPFC LF (1 Hz) rTMS treatment combined with psychotherapy. Exclusion criteria included: (1) prior treatment with ECT, (2) epilepsy, (3) wearing a cardiac pacemaker, (4) wearing metal parts in the head, and (5) pregnancy. All participants signed an informed consent form before treatment was initiated.

The specific treatment procedures and clinical outcomes have recently been published elsewhere (Donse *et al.*, 2018). In short, baseline clinical variable measurements consisted of the BDI-II-NL, Depression, Anxiety and Stress Scale (DASS), and NEO-FFI. To track the course of the rTMS treatment, the BDI was assessed every fifth rTMS session.

Response prediction

Differences between responders (R) and non-responders (NR) were analyzed based on clinical baseline variables. The clinical variables were assessed using the following instruments:

- The BDI-II-NL was used to assess depression severity. Of the BDI, several subscales were taken. These included:
  - the Anhedonia scale (items 4, 12, and 21) and the Non-Anhedonia scale (items 1–3, 5–11, and 13–20) (Leventhal *et al.*, 2006)
  - the Cognitive-Affective scale (items 1–13) and the Somatic and Performance scale (items 14–21) (Trentini *et al.*, 2005)
  - the Cognitive scale (items 2, 3, 5–9, and 14) and the Non-Cognitive scale (items 1, 4, 10–13, and 15–21) (Kumar *et al.*, 2002)

These subscales were computed by adding the indicated items into one variable (e.g. the anhedonia scale was computed by adding item 4, item 12, and item 21). Scores for these scales were also calculated at outtake, as well as in change over time (in absolute numbers and percentages).

- The NEO-FFI was used to examine 'the Big Five' personality traits. The NEO-FFI is a 60-item, self-report instrument that measures five personality traits, being Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. These domains have shown good internal consistency (Cronbach's α range 0.87–0.92) (Costa and McCrae, 1992).
- To measure anxiety and stress, the DASS (Henry and Crawford, 2005) was used.

Discovery and Replication set

To obtain a Discovery and Replication set, the complete, original dataset ($n = 196$) was randomly divided into a 60% Discovery sample and 40% Replication sample. Differences in gender, response, and remission were tested for using χ^2 statistics. Differences in age and BDI at intake were tested for using one-way analysis of variance tests. Once none of these variables differed significantly between the two groups, the random division was frozen and the two groups were designated as Discovery set and Replication set. This resulted in a Discovery set ($n = 119$; 62 females) and Replication set ($n = 77$; 37 females). All exploratory analyses were performed in the Discovery set. Only when a significant result was obtained in the Discovery set, the same statistical test was used in the Replication set to confirm or deny the prior obtained finding.

Statistics

SPSS version 24 was used for statistical analyses. Response was defined as a ⩾50% decrease on the BDI score from intake to outtake (outtake BDI scores were taken around the last session of a patient, on average at session 21). All predictors were tested for statistical differences, robustness (independent replication), dimensional association, predictive value, and clinical relevance.

First, a generalized linear model (GLM) univariate analysis was performed to test for baseline psychological differences between R and NR, with response as a fixed factor, age as a covariate, and the BDI subscales (including total BDI, Anhedonia, Non-Anhedonia, Cognitive Affective, Somatic and Performance, Cognitive, and Non-Cognitive), NEO-FFI subscales (including Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness), and DASS scales (Depression, Anxiety, and Stress) as dependent variables. Effect sizes reported are Cohen's *d*.

Next, the sensitivity, specificity, PPV, and NPV for predicting non-response of all the above described variables were examined, using a discriminant analysis. The reason for this extra dimension of statistical evaluation was that a psychological measure may not be significantly different between R and NR, as tested by a GLM univariate analysis, yet it may prove to be highly predictive of non-response in a smaller specific subset of NR. Typically, PPV and NPV are set relatively high, i.e. ⩾0.75 or 75% (Li *et al.*, 2012). Thus, if it were the case that NPV or PPV exceeded 75% and replicated, this psychological measure was still examined in the following analyses.

Replicated psychological measures (or those showing a PPV or NPV higher than 75%) were also tested for dimensional association by partially correlating these measures with BDI percentage change (BDI% change) from intake to outtake, while controlling for age.

Next, a discriminant analysis on non-response was performed, using the predictors as indicated by GLM univariates of PPV and NPV. More specifically, the replicated baseline psychological measures that showed to be significantly different between R and NR, or that had a replicated NPV or PPV of 75% or higher, were used as independent variables in the discriminant analysis. In addition to these items, age was also used in this analysis. From these variables, a receiver operator curve (ROC) was derived to establish the predictive value of the discriminant analysis on treatment response.

It was also attempted to establish a useful construct based on the replicated predictors by examining the severity levels of these predictors. That is, a cut-off score of these predictors was established using sensitivity, specificity, and Youden's *J*. This cut-off score provided a useful construct on which a therapist can act.

A further confirmatory approach that also compares a Discovery–Replication approach to a more traditional approach applied to the full sample (albeit using strict Bonferroni corrected *p*-values) is reported in the online Supplementary material. In previous treatment prediction studies in MDD we have reported many sex-specific predictors (van Dinteren *et al.*, 2015; Arns *et al.*, 2016; Iseger *et al.*, 2017), yet statistically testing such interactions requires relatively large sample sizes (Leon and Heo, 2009). Therefore, in this full sample a post-hoc analysis on the complete dataset was conducted and statistically examined the

https://doi.org/10.1017/S0033291718004191 Published online by Cambridge University Press

Table 1. Clinical outcome measures of the Discovery and Replication sample

Clinical response	Discovery sample	Replication sample
Responders	77 (65.3%)	53 (68.8%)
Total number of sessions	22.3 (s.d. 8.0)	19.0 (s.d. 6.0)
BDI intake	31.3 (s.d. 10.3)	31.2 (s.d. 9.7)
BDI outtake BDI percentage change	13.6 (s.d. 11.6) 57.1%	14.8 (s.d. 13.0) 54.1%
Cohen's *d* intake–outtake	ES = 1.6	ES = 1.4
Trial center (Nijmegen, The Hague, Oosterhout)	109 (91.6%), 2 (1.7%), 8 (6.8%)	68 (88.3%), 2 (2.6%), 7 (9.1%)
rTMS protocol		
HF (10 Hz left)	45 (38.1%)	29 (37.7%)
LF (1 Hz right)	70 (58.8%)	45 (58.4%)
Both sequentially	4 (3.4%)	3 (3.9%)

This table depicts the number of responders, total number of sessions, total BDI at intake, total BDI at outtake, the BDI change from intake to outtake (in percentage), as well as the ES (Cohen's *d*) of total BDI change from intake to outtake, trial center, and rTMS protocol. BDI at intake, BDI at outtake, BDI percentage change, and number of responders did not differ significantly between groups ($p \geqslant 0.502$).

data for potential sex × response interactions by including sex as a between-subject factor. In the online supplementary material, it is also investigated whether anhedonia could accurately predict long-term response.

Primary and secondary hypotheses

By dividing our complete sample into a Discovery and Replication set, the sample sizes naturally decreased in both samples. To prevent a type-II error from occurring due to reduced sample sizes, it was decided that the hypotheses were defined as primary or secondary hypotheses. A primary hypothesis was defined as a hypothesis that was predicted from the literature and the *p*-value was set $p \leqslant 0.1$ in the Discovery set. In the Replication set this finding had to reach $p \leqslant 0.05$ to be considered a replication. A secondary hypothesis was defined as a hypothesis that was not predicted from the literature, i.e. an exploratory analysis, yet showed statistical significance in the dataset. This *p*-value was set at $p \leqslant 0.05$. Given the confirmatory nature of the Replication set, it was decided that the correlation would be one-tailed (under the condition that the direction of the two-tailed test was the same as in the Discovery set).

As primary hypotheses, the following were defined as predictors for non-response: high total BDI, high anhedonia, high neuroticism, low extraversion, high anxiety, and high stress.

All other psychological measures were marked as secondary hypotheses.

Results

A total of 196 MDD patients were enrolled in this study (average age: 43.62, range 18–78 years; 99 females and 97 males). Dividing the complete dataset ($n = 196$) in a 60–40 distribution (60% Discovery, 40% Replication) resulted in a Discovery set of 119 MDD patients (average age: 43.60, range 18–73; 62 females and 57 males) and a Replication set of 77 MDD patients (average age: 43.64, range 19–78 years; 37 females and 40 males). The clinical outcome measures of the Discovery and Replication sample are summarized in Table 1. One subject was excluded based on missing data. Note that, given that most the MDD patients were from Nijmegen, the following analyses were also performed in the Nijmegen-only cohort. These analyses for data collected in Nijmegen only yielded similar statistical outcomes and did not result in different conclusions.

Discovery set

Primary-analyses only yielded an effect of response for the Anhedonia scale ($p = 0.072$; $F = 3.298$; df = 1). For the secondary analyses only Openness differed between R and NR ($p = 0.029$; $F = 4.889$; df = 1). All other variables were not significantly different between R and NR (see Table 2). In Table 2, all the effect sizes (Cohen's *d*) for each baseline psychological measure, subdivided for R and NR, are reported as well.

Replication set

A GLM univariate test was performed on the Replication set, with the exact same parameters as in the Discovery set. When focusing on only the predicted psychological measures that were significantly related to response in the Discovery set (i.e. the Anhedonia scale and the NEO-FFI Openness scale), a significant effect was found for the Anhedonia scale ($p = 0.005$; $F = 8.516$; df = 1). No significant effect was found for Openness ($p = 0.227$; $F = 1.490$; df = 1). The results are shown in Table 2. Since all other variables showed to be non-significant in the Discovery set, significant differences between R and NR that were confined to only the Replication set were ignored. However, for purposes of completeness, all other variables and their statistical values are shown as well.

Sensitivity, specificity, NPV, and PPV

As described, baseline psychological measures may not be significantly different between R and NR, yet be of high predictive value for a (specific) (sub)group. Therefore, each baseline psychological measure was individually assessed for these features, using one psychological measure and age in a discriminant analysis. The preset value of both PPV and NPV was 75%. As can be observed

https://doi.org/10.1017/S0033291718004191 Published online by Cambridge University Press

Table 2. This table depicts the mean, standard deviation, *p*-value, and ES for each baseline psychological measure, showing values for Discovery as well as Replication sample, subdivided into R and NR

(Sub)scale	Responder [*m*, (s.d.)]		Non-responder [*m*, (s.d.)]		*p*		ES (Cohen's *d*)	
	Discovery	Replication	Discovery	Replication	Discovery	Replication	Discovery	Replication
Age	41.9 (10.8)	43.9 (14.3)	46.6 (14.5)	43.0 (12.6)	0.049	0.788	0.37	−0.07
BDI total	30.7 (9.5)	29.4 (8.4)	32.4 (11.8)	35.3 (11.1)	0.198	0.012	0.16	0.60
Anhedonia	**5.2 (2.1)**	**4.5 (1.8)**	**6.0 (2.3)**	**5.8 (1.6)**	**0.072**	**0.005**	**0.36**	**0.76**
Non-Anhedonia	25.9 (7.7)	24.5 (8.5)	26.8 (10.1)	29.0 (10.2)	0.314	0.066	0.10	0.48
Cognitive-Affective	18.4 (6.1)	16.9 (7.4)	19.1 (8.1)	20.7 (7.7)	0.315	0.064	0.10	0.50
Somatic and Performance	12.8 (4.0)	12.1 (3.1)	13.6 (4.8)	14.1 (4.2)	0.259	0.029	0.18	0.54
Cognitive	10.3 (4.5)	9.8 (5.5)	10.0 (5.7)	10.9 (5.7)	0.821	0.483	−0.06	0.20
Non-Cognitive	21.0 (6.2)	19.3 (4.9)	22.7 (7.8)	23.9 (6.4)	0.133	0.002	0.24	0.81
NEO-FFI Neuroticism	32.7 (6.3)	30.4 (7.8)	32.7 (7.2)	31.8 (6.8)	0.844	0.455	0	0.19
NEO-FFI Extraversion	19.1 (6.8)	20.1 (7.1)	18.9 (8.6)	17.0 (8.0)	0.919	0.106	−0.03	−0.41
NEO-FFI Openness	**27.0 (6.2)**	**25.7 (7.3)**	**24.6 (7.4)**	**23.4 (6.8)**	**0.029**	**0.227**	**−0.35**	**−0.33**
NEO-FFI Agreeableness	32.1 (5.3)	31.8 (6.8)	32.0 (5.9)	32.9 (4.3)	0.854	0.476	−0.02	0.19
NEO-FFI Conscientiousness	26.3 (6.5)	28.6 (7.7)	27.3 (7.4)	26.7 (5.2)	0.561	0.294	0.14	−0.29
DASS Depression	29.0 (9.2)	26.4 (11.5)	29.1 (9.5)	31.0 (9.4)	0.666	0.088	0.01	0.44
DASS Anxiety	13.9 (8.7)	12.3 (8.4)	14.7 (9.3)	14.5 (8.8)	0.801	0.343	0.09	0.26
DASS Stress	24.8 (10.0)	18.8 (11.1)	21.8 (10.4)	23.8 (8.9)	0.164	0.053	−0.29	0.50

A univariate analysis, controlled for age, showed that the Anhedonia scale is significantly different between R and NR in both the Discovery ($p = 0.072$; $d = 0.36$) as well as for the Replication set ($p = 0.005$; $d = 0.76$). Openness shows to be significantly different between R and NR in the Discovery sample ($p = 0.029$; $d = -0.35$), however, this results fails to be replicated in the Replication sample ($p = 0.227$; $d = -0.33$). These results are indicated in bold.

from Table 3, none of the baseline variables reached or exceeded and replicated the preset value of PPV and NPV.

Correlations between psychological scales and BDI percentage change

A two-tailed correlation while controlling for age showed a significant correlation between BDI% change and the Anhedonia scale at intake ($p = 0.023$; $r = -0.221$; $r^2 = 4.9\%$). Additionally, BDI% change and Openness at baseline were shown to be significantly correlated ($p = 0.025$; $r = 0.214$; $r^2 = 4.6\%$).

In the Replication set, there was a significant one-tailed partial correlation between BDI% change and the Anhedonia scale ($p = 0.025$; $r = -0.244$; $r^2 = 6.0\%$), and between BDI% change and Openness ($p = 0.047$; $r = 0.206$; $r^2 = 4.2\%$).

Discriminant analysis

Using the Discovery set, a discriminant analysis was performed using the Anhedonia scale and age. The resulting model was shown to be significant ($p = 0.034$; Wilks' $\lambda = 0.937$; $\chi^2 = 6.771$; df = 2), with a sensitivity of 62.5% and a specificity of 53.7%. The PPV and NPV were 44.6% and 70.6%, respectively. The area under the curve of the ROC, regressed on non-response, was 0.643 (Fig. 1). This analysis was repeated in the Replication set. This model also reached significance ($p = 0.018$; Wilks' $\lambda = 0.880$; $\chi^2 = 8.024$; df = 2), with a sensitivity of 73.9%, and a specificity of 74.4%. The PPV and NPV were 60.7% and 80.0%, respectively. The area under the ROC curve was 0.726 (Fig. 2).

Clinical relevance (sensitivity, specificity, NPV, PPV, Youden's J of the Anhedonia scale)

For the current calculations, the Discovery and Replication set were merged, since in this section the clinical relevance of this metric was investigated, rather than its methodological value. To specify the clinical relevance of the Anhedonia scale, a cut-off score was established by calculating Youden's index.

A complete overview of statistics can be found in the online supplementary material. In short, using an ROC the cut-off, from which predicting NR based on the Anhedonia scale was most accurate, was estimated. Youden's *J* was highest at severity level 6 ($J = 0.235$). Also, the PPV and NPV do not reach the preset value of 75%, and therefore the model is, based on this method, considered to not be clinically relevant.

Discussion

This study focused on finding psychological measures and their potential ability to predict rTMS treatment response in an MDD sample. Additionally, attempting to overcome the issue of (non-)replication, this study tried to immediately replicate obtained findings by a priori dividing the complete dataset into a Discovery and Replication set.

https://doi.org/10.1017/S0033291718004191 Published online by Cambridge University Press

Table 3. This table depicts the sensitivity, specificity, PPV, and NPV per individual baseline psychological measure

(Sub)scale	Sensitivity (%)		Specificity (%)		PPV (%)		NPV (%)	
	Discovery	Replication	Discovery	Replication	Discovery	Replication	Discovery	Replication
BDI total	63.4	54.2	57.1	64.2	44.1	40.6	74.6	75.6
Anhedonia	62.5	73.9	53.7	74.4	44.6	60.7	70.6	84.2
Non-Anhedonia	56.4	56.5	54.5	64.3	42.3	46.4	67.9	73.0
Cognitive-Affective	57.5	47.8	53.0	59.5	42.6	39.3	67.3	67.6
Somatic and Performance	59.0	56.5	59.7	62.8	46.0	44.8	71.4	73.0
Cognitive	56.4	47.8	54.5	50.0	42.3	34.4	67.9	63.6
Non-Cognitive	62.5	47.8	58.2	72.1	47.2	47.8	72.2	72.1
NEO-FFI Neuroticism	56.4	59.1	56.9	45.7	41.5	32.5	70.7	67.9
NEO-FFI Extraversion	56.4	63.6	56.9	58.7	41.5	42.4	70.7	77.1
NEO-FFI Openness	56.4	59.1	58.3	47.8	42.3	35.1	71.2	71.0
NEO-FFI Agreeableness	59.0	59.1	55.6	47.8	41.8	35.1	71.4	71.0
NEO-FFI Conscientiousness	51.3	59.1	55.6	63.0	38.5	43.3	67.8	76.3
DASS Depression	56.1	58.3	53.4	55.1	40.4	38.9	68.4	73.0
DASS Anxiety	56.1	54.2	54.8	62.5	41.1	41.9	69.0	73.2
DASS Stress	48.8	54.2	57.5	59.2	39.2	39.4	66.67	72.5

A discriminant analysis was performed using one psychological measure plus age as independent variables. Based on the absolute number of true positives, true negative, false positives, and false negatives (as predicted by the discriminant analysis), the above metrics were calculated. If PPV or NPV exceeded 75% and replicated, this psychological measure would still be examined in the following analyses, even though it showed not to be significantly different between R and NR in the previous univariate analyses. None of the baseline psychological measures showed a PPV or NPV of 75% or higher in both the Discovery and Replication sample.

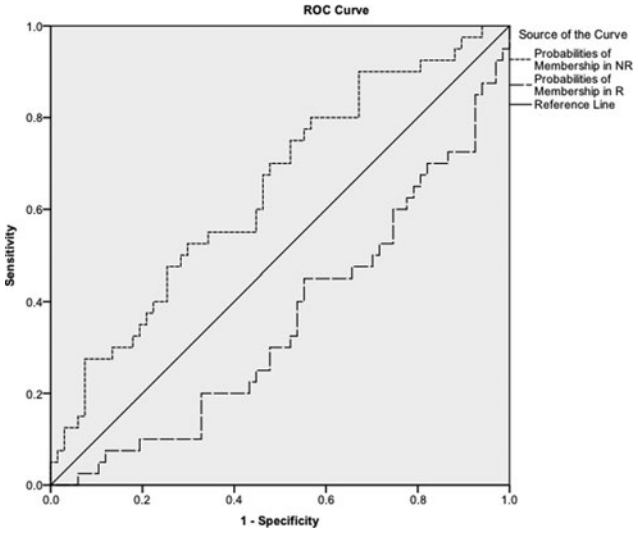


Fig. 1. ROC of the discriminant analysis on non-response using the Anhedonia scale and age as independent variables in the Discovery sample, with an area under the curve of 0.643. The ROC shows the sensitivity (62.5%) and specificity (53.7%) for non-responders (dotted line) and responders (striped line). NPV and PPV were 44.6% and 70.6%, respectively.

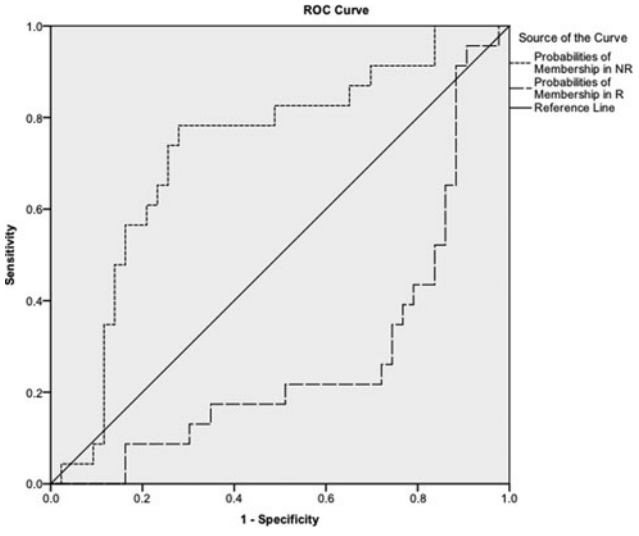


Fig. 2. ROC of the discriminant analysis on non-response using the Anhedonia scale and age as independent variables in the Replication sample, with an area under the curve of 0.726. The ROC shows the sensitivity (73.9%) and specificity (74.4%) for non-responders (dotted line) and responders (striped line). The PPV and NPV were 60.7% and 80.0%, respectively.

Our study suggests that none of the psychological measures are clinically meaningful predictors of rTMS treatment response in an MDD sample. This is in line with studies that found evidence for predictive utilities of psychological measures, but did not found it to be highly influential on response rates (Fitzgerald *et al.*, 2016). However, this study suggests that NR robustly show higher anhedonia scores at baseline, and that this score was – to some degree – related to clinical improvement. More specifically, a higher score on the Anhedonia scale (as taken from the BDI) marginally predicted non-response to rTMS. However, the ROC curves were only mildly predictive of non-response, and NPV was 70.6–80.0% and PPV was 44.6–60.7%, thereby the NPV

https://doi.org/10.1017/S0033291718004191 Published online by Cambridge University Press

just being short of the a priori preset threshold of 75%. Therefore, even though NR seem to have a higher anhedonia score at baseline, its predictive value remains relatively low. This is partly in line with Rostami and colleagues., who found that the symptom 'loss of interest' was strongly predictive in rTMS response (Rostami *et al.*, 2017). Even though in our sample it is also found that baseline anhedonia is different between R and NR, the subset of symptoms could not predict treatment response such that is was clinically relevant.

When attempting to find the optimal cut-off score on which a therapist can act, it was found that those with a baseline anhedonia score of 6 or higher are less likely to respond to rTMS treatment. However, given the overall clinical non-relevance of the model based on anhedonia, this cut-off should be taken with caution.

Replication

In this study we employed two approaches that are often advocated as tools to overcome the 'Replication crisis'. Some people have advocated to rely more strongly on replication studies (Open Science Collaboration, 2012; Pashler and Wagenmakers, 2012; Roediger, 2012; Simons, 2014) [although caution should be taken when designing and interpreting a replication study, see Maxwell *et al.* (2015)], as well as self-replication (i.e. replicate your own work before publishing) (Roediger, 2012; Simons, 2014), which is what we performed in this main manuscript. Others advocate to use stricter *p*-values, for example using a *p*-value of 0.005 (Benjamin *et al.*, 2018). In the online supplementary material this latter approach was implemented, using the full sample and a strict Bonferroni corrected *p*-value of 0.0033. Interestingly, both approaches converged on identifying anhedonia as a predictor for response. A difference was that in the latter approach, as elaborated on in the online supplementary material, another psychological scale of the BDI (the Non-Cognitive scale) was identified as being statistically different between R and NR. However, it seems that this difference is mainly driven by BDI item overlap with the Anhedonia scale (since the Anhedonia scale is also part of the Non-Cognitive scale) and thus a spurious correlation. Further elaboration on these analyses can be found in the online supplementary material.

Interestingly, Table 2 can also be inspected as an example of how many false positives one could obtain using a $p < 0.05$ approach without requiring replication (i.e. if these would be published as two separate papers). When looking at any $p < 0.05$ values in either the Discovery or Replication dataset 6 out of 17 measures (35%) are significantly different, whereas only one of those metrics (6%) actually replicates. When using stricter thresholding and a larger sample size ($p < 0.0033$; online supplementary material) only one false positive was found (the Non-Cognitive scale) and the same result for anhedonia was confirmed.

Therefore, these results further highlight that false positive findings can be easily obtained. Both approaches (replication *v.* stricter *p*-thresholding) increased the robustness of results. Based on these results we cannot draw a definite conclusion, but for future manuscripts where we will be looking at other predictors [e.g. electroencephalography (EEG)], we will use this exact same approach and hope to obtain robust and clinically relevant predictors of treatment response.

Limitations

This study did not employ a double blind placebo controlled design, hence we are unable to rule out that the obtained results can be (partly) explained by such effects. Similarly, this study was open-labeled, which poses another weakness, albeit results in ecologically valid results. Furthermore, as is described by Donse *et al.* (2018), this sample consisted of MDD patients who received simultaneous rTMS and psychotherapy, rather than rTMS only. Lastly, while the complete sample consists of 196 MDD patients, reducing it to a Discovery and Replication set and dividing it into a R/NR classification, narrows down the sample size, resulting in a smallest sample size (Replication set, NR) of 24.

Conclusion

Our study indicates that the psychological measure anhedonia, as measured by the Anhedonia scale of the BDI at baseline, is related to clinical improvement on MDD symptoms in response to rTMS treatment in an MDD sample. More specifically, lower baseline anhedonia scores were related to better clinical improvement. However, our study also suggests that anhedonia is only mildly predictive of treatment response and does not achieve predefined levels of clinical relevance. Therefore, it can be concluded that lower anhedonia scores are favorable in the treatment of MDD, however, these are unlikely to be of clinical usage and relevance due to the low PPV and NPV.

Furthermore, the current study also highlights that false positives are relatively easy obtained, when handling a 0.05 significance level, advocating for the Replication approach and the usage of stricter *p*-values, as was done in this main manuscript and online supplementary material.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291718004191.

Acknowledgements. We would like to thank Lana Donse, Vera Kruiver, Rosalinde van Ruth, Marleen Stam, Maaike Moolenaar, Sanne Bongers, Myrthe van Eerdt, Dagmar Timmers, Inge Janssen-Bouwmeester and Nicole van Merode for support and collecting the data used in this study.

Author contributions. NK managed the literature search, performed the analyses, and wrote the first draft of the manuscript. All other authors contributed, reviewed, and approved the final manuscript.

Conflict of interest. MA reports options from Brain Resource (Sydney, Australia); he is director and owner of Research Institute Brainclinics, a minority shareholder in neuroCare Group (Munich, Germany), and a co-inventor on four patent applications (A61B5/0402; US2007/0299323, A1; WO2010/139361 A1) related to EEG, neuromodulation and psychophysiology, but does not own these nor receives any proceeds related to these patents; Research Institute Brainclinics received funding from Brain Resource (Sydney, Australia) and neuroCare Group (Munich, Germany), and equipment support from Deymed, neuroConn and Magventure, however data analyses and writing of this manuscript were unconstrained. AJR reports consulting fees from the American Psychiatric Association, Brain Resource Ltd., Compass Inc., Curbstone Consultant LLC, Eli Lilly, Emmes Corp., Liva-Nova, Lundbeck A/S, National Institute of Drug Abuse, Taj Medical, Santium Inc., Sunovion, Taj Medical, Takeda USA; speaking fees from Live Nova; royalties from Guilford Publications and the University of Texas Southwestern Medical Center. The other authors report no disclosures or conflicts of interest.

References

Arns M, Bruder G, Hegerl U, Spooner C, Palmer DM, Etkin A, Fallahpour K, Gatt JM, Hirshberg L and Gordon E (2016) EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment

with different antidepressant medications in the randomized iSPOT-D study. *Clinical Neurophysiology* **127**, 509–519.

Bagby RM, Quilty LC, Segal ZV, McBride CC, Kennedy SH and Costa PT (2008) Personality and differential treatment response in major depression: a randomized controlled trial comparing cognitive-behavioural therapy and pharmacotherapy. *Canadian Journal of Psychiatry* **53**, 361–370.

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field AP, Forster M, George EI, Gonzalez R, Goodman S, Green E, Green DP, Greenwald A, Hadfield JD, Hedges LV, Held L, Hua Ho T, Hoijtink H, Hruschka DJ, Imai K, Imbens G, Ioannidis JPA, Jeon M, Jones JH, Kirchler M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell SE, McCarthy M, Moore D, Morgan SL, Munafó M, Nakagawa S, Nyhan B, Parker TH, Pericchi L, Perugini M, Rouder J, Rousseau J, Savalei V, Schönbrodt FD, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts DJ, Winship C, Wolpert RL, Xie Y, Young C, Zinman J and Johson VE (2018) Redefine statistical significance. *Nature Human Behaviour* **2**, 6–10.

Berlim MT, McGirr A, Beaulieu M-M, Van den Eynde F and Turecki G (2013) Are neuroticism and extraversion associated with the antidepressant effects of repetitive transcranial magnetic stimulation (rTMS)? An exploratory 4-week trial. *Neuroscience Letters* **534**, 306–310.

Berlim MT, van den Eynde F, Tovar-Perdomo S and Daskalakis ZJ (2014) Response, remission and drop-out rates following high-frequency repetitive transcranial magnetic stimulation (rTMS) for treating major depression: a systematic review and meta-analysis of randomized, double-blind and sham-controlled trials. *Psychological Medicine* **44**, 225–239.

Brakemeier E-L, Luborzewski A, Danker-Hopfe H, Kathmann N and Bajbouj M (2007) Positive predictors for antidepressive response to prefrontal repetitive transcranial magnetic stimulation (rTMS). *Journal of Psychiatric Research* **41**, 395–403.

Carpenter LL, Janicak PG, Aaronson ST, Boyadjis T, Brock DG, Cook IA, Dunner DL, Lanocha K, Solvason HB and Demitrack MA (2012) Transcranial magnetic stimulation (TMS) for major depression: a multisite, naturalistic, observational study of acute treatment outcomes in clinical practice. *Depression and Anxiety* **29**, 587–596.

Costa PT and McCrae RR (1992) Revised NEO Personality Inventory and NEO Five-Factor Inventory: Professional Manual. *Psychological Assessment Resources, Inc., Lutz, FL*.

Croughan JL, Secunda SK, Katz MM, Robins E, Mendels J, Swann A and Harris-Larkin B (1988) Sociodemographic and prior clinical course characteristics associated with treatment response in depressed patients. *Journal of Psychiatric Research* **22**, 227–237.

Deng Y, McQuoid DR, Potter GG, Steffens DC, Albert K, Riddle M, Beyer JL and Taylor WD (2018) Predictors of recurrence in remitted late-life depression. *Depression and Anxiety* **35**, 658–667.

Donse L, Padberg F, Sack AT, Rush AJ and Arns M (2018) Simultaneous rTMS and psychotherapy in major depressive disorder: clinical outcomes and predictors from a large naturalistic study. *Brain Stimulation* **11**, 337–345.

Downar J, Geraci J, Salomons TV, Dunlop K, Wheeler S, McAndrews MP, Bakker N, Blumberger DM, Daskalakis ZJ, Kennedy SH, Flint AJ and Giacobbe P (2014) Anhedonia and reward-circuit connectivity distinguish nonresponders from responders to dorsomedial prefrontal repetitive transcranial magnetic stimulation in major depression. *Biological Psychiatry* **76**, 176–185.

Fava M, Alpert JE, Carmin CN, Wisniewski SR, Trivedi MH, Biggs MM, Shores Wilson K, Morgan D, Schwartz T, Balasubramani GK and Rush AJ (2004) Clinical correlates and symptom patterns of anxious depression among patients with major depressive disorder in STAR*D. *Psychological Medicine* **34**, 1299–1308.

Fava M, Rush AJ, Alpert JE, Balasubramani GK, Wisniewski SR, Carmin CN, Biggs MM, Zisook S, Leuchter A, Howland R, Warden D and Trivedi MH (2008) Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *The American Journal of Psychiatry* **165**, 342–351.

Fitzgerald PB, Brown TL, Marston NAU, Daskalakis ZJ, de Castella A and Kulkarni J (2003) Transcranial magnetic stimulation in the treatment of depression: a double-blind, placebo-controlled trial. *Archives of General Psychiatry* **60**, 1002.

Fitzgerald PB, Hoy K, Daskalakis ZJ and Kulkarni J (2009) A randomized trial of the anti depressant effects of low- and high-frequency transcranial magnetic stimulation in treatment-resistant depression. *Depression and Anxiety* **26**, 229–234.

Fitzgerald PB, Hoy KE, Anderson RJ and Daskalakis ZJ (2016) A study of the pattern of response to rTMS treatment in depression. *Depression and Anxiety* **33**, 746–753.

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC and Fawcett J (2010) Antidepressant drug effects and depression severity: a patient-level meta analysis. *JAMA* **303**, 47–53.

Gaynes BN, Lloyd SW, Lux L, Gartlehner G, Hansen RA, Brode S, Jonas DE, Swinson Evans T, Viswanathan M and Lohr KN (2014) Repetitive transcranial magnetic stimulation for treatment-resistant depression: a systematic review and meta-analysis. *The Journal of Clinical Psychiatry* **75**, 477–489.

Goldberg LR (1990) An alternative 'description of personality': the big-five factor structure. *Journal of Personality and Social Psychology* **59**, 1216–1229.

Griffith JW, Zinbarg RE, Craske MG, Mineka S, Rose RD, Waters AM and Sutton JM (2010) Neuroticism as a common dimension in the internalizing disorders. *Psychological Medicine* **40**, 1125–1136.

Hammen C (2005) Stress and depression. *Annual Review of Clinical Psychology* **1**, 293–319.

Hayward RD, Taylor WD, Smoski MJ, Steffens DC and Payne ME (2013) Association of five-factor model personality domains and facets with presence, onset, and treatment outcomes of major depression in older adults. *The American Journal of Geriatric Psychiatry* **21**, 88–96.

Henry JD and Crawford JR (2005) The short-form version of the depression anxiety stress scales (DASS-21): construct validity and normative data in a large non-clinical sample. *British Journal Clinical Psychology* **44**, 227–239.

Ionescu DF, Luckenbaugh DA, Niciu MJ, Richards EM, Slonena EE, Vande Voort JL, Brutsche NE and Zarate CA (2014) Effect of baseline anxious depression on initial And sustained antidepressant response to ketamine. *Journal of Clinical Psychiatry* **75**, e932–e938.

Iseger TA, Korgaonkar MS, Kenemans JL, Grieve SM, Baeken C, Fitzgerald PB and Arns M (2017) EEG connectivity between the subgenual anterior cingulate and prefrontal cortices in response to antidepressant medication. *European Neuropsychopharmacology* **27**, 301–312.

Joffe RT, Bagby RM and Levitt A (1993) Anxious and nonanxious depression. *The American Journal of Psychiatry* **150**, 1257–1258.

Jylhä P and Isometsä E (2006) The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. *Depression and Anxiety* **23**, 281–289.

Khan A, Leventhal RM, Khan SR and Brown WA (2002) Severity of depression and response to antidepressants and placebo: an analysis of the food and drug administration database. *Journal of Clinical Psychopharmacology* **22**, 40–45.

Kotov R, Gamez W, Schmidt F and Watson D (2010) Linking 'big' personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychological Bulletin* **136**, 768–821.

Krepel N, Sack AT, Kenemans JL, Fitzgerald PB, Drinkenburg WH and Arns M (2018) Non-replication of neurophysiological predictors of non-response to rTMS in depression and neurophysiological data-sharing proposal. *Brain Stimulation* **11**, 639–641.

Kuk AYC, Li J and Rush AJ (2010) Recursive subsetting to identify patients in the STAR*D: a method to enhance the accuracy of early prediction of treatment outcome and to inform personalized care. *The Journal of Clinical Psychiatry* **71**, 1502–1508.

Kumar G, Steer RA, Teitelman KB and Villacis L (2002) Effectiveness of Beck Depression Inventory-II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment* **9**, 164–170.

Leon AC and Heo M (2009) Sample sizes required to detect interactions between two binary fixed-effects in a mixed-effects linear regression model. *Computational Statistics & Data Analysis* **53**, 603–608.

Leventhal AM, Chasson GS, Tapia E, Miller EK and Pettit JW (2006) Measuring hedonic capacity in depression: a psychometric analysis of three anhedonia scales. *Journal of Clinical Psychology* **62**, 1545–1558.

https://doi.org/10.1017/S0033291718004191 Published online by Cambridge University Press

Li J, Kuk AYC and Rush AJ (2012) A practical approach to the early identification of antidepressant medication non-responders. *Psychological Medicine* **42**, 309–316.

Maxwell SE, Lau MY and Howard GS (2015) Is psychology suffering from a replication crisis? What does 'failure to replicate' really mean? *The American Psychologist* **70**, 487–498.

McGirr A, Van den Eynde F, Chachamovich E, Fleck MPA and Berlim MT (2014) Personality dimensions and deep repetitive transcranial magnetic stimulation (DTMS) for treatment-resistant depression: a pilot trial on five-factor prediction of antidepressant response. *Neuroscience Letters* **563**, 144–148.

McGonagle KA and Kessler RC (1990) Chronic stress, acute stress, and depressive symptoms. *American Journal of Community Psychology* **18**, 681–706.

McMakin DL, Olino TM, Porta G, Dietz LJ, Emslie G, Clarke G, Wagner KD, Asarnow JR, Ryan ND, Birmaher B, Shamseddeen W, Mayes T, Kennard B, Spirito A, Keller M, Lynch FL, Dickerson JF and Brent DA (2012) Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment resistant depression. *Journal of the American Academy of Child and Adolescent Psychiatry* **51**, 404–411.

Mulder RT (2002) Personality pathology and treatment outcome in major depression: a review. *The American Journal of Psychiatry* **159**, 359–371.

Open Science Collaboration (2012) An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science* **7**, 657–660.

Pashler H and Wagenmakers E-J (2012) Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspectives on Psychological Science* **7**, 528–530.

Patil P, Peng RD and Leek JT (2016) What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspectives on Psychological Science* **11**, 539–544.

Pizzagalli DA (2014) Depression, stress, and anhedonia: toward a synthesis and integrated model. *Annual Review of Clinical Psychology* **10**, 393–423.

Roediger III HL (2012) Psychology's Woes and a Partial Cure: The Value of Replication. Association for Psychological Science. Available at https://www.psychologicalscience.org/observer/psychologys-woes-and-a-partial-cure-the-value-of-replication.

Quilty LC, De Fruyt F, Rolland J-P, Kennedy SH, Rouillon PF and Bagby RM (2008) Dimensional personality traits and treatment outcome in patients with major depressive disorder. *Journal of Affective Disorders* **108**, 241–250.

Rosellini AJ and Brown TA (2011) The NEO five-factor inventory: latent structure and relationships with dimensions of anxiety and depressive disorders in a large clinical sample. *Assessment* **18**, 27–38.

Rostami R, Kazemi R, Nitsche MA, Gholipour F and Salehinejad MA (2017) Clinical and demographic predictors of response to rTMS treatment in unipolar and bipolar depressive disorders. *Clinical Neurophysiology* **128**, 1961–1970.

Schutter DJLG (2009) Antidepressant efficacy of high-frequency transcranial magnetic stimulation over the left dorsolateral prefrontal cortex in double-blind sham-controlled designs: a meta-analysis. *Psychological Medicine* **39**, 65–75.

Schutter DJLG (2010) Quantitative review of the efficacy of slow-frequency magnetic brain stimulation in major depressive disorder. *Psychological Medicine* **40**, 1789–1795.

Simons DJ (2014) The value of direct replication. *Perspectives on Psychological Science* **9**, 76–80.

Spijker J, Bijl RV, de Graaf R and Nolen WA (2001) Determinants of poor 1-year outcome of DSM-III-R major depression in the general population: results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Acta Psychiatrica Scandinavica* **103**, 122–130.

Treadway MT and Zald DH (2011) Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neuroscience and Biobehavioral Reviews* **35**, 537–555.

Trentini CM, de Xavier FMF, Chachamovich E, Rocha NSD, Hirakata VN and de Fleck MPA (2005) The influence of somatic symptoms on the performance of elders in the Beck Depression Inventory (BDI). *Revista Brasileira de Psiquiatria* (*Sao Paulo, Brazil: 1999*) **27**, 119–123.

Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ, Shores-Wilson K, Biggs MM, Balasubramani GK, Fava M and STAR*D Study Team (2006) Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *The American Journal of Psychiatry* **163**, 28–40.

Uher R, Dernovsek MZ, Mors O, Hauser J, Souery D, Zobel A, Maier W, Henigsberg N, Kalember P, Rietschel M, Placentino A, Mendlewicz J, Aitchison KJ, McGuffin P and Farmer A (2011) Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. *Journal of Affective Disorders* **132**, 112–120.

van Dinteren R, Arns M, Kenemans L, Jongsma ML, Kessels RP, Fitzgerald P, Fallahpour K, Debattista C, Gordon E and Williams LM (2015) Utility of event-related potentials in predicting antidepressant treatment response: an iSPOT-D report. *European Neuropsychopharmacology* **25**, 1981–1990.

Vinckier F, Gourion D and Mouchabac S (2017) Anhedonia predicts poor psychosocial functioning: results from a large cohort of patients treated for major depressive disorder by general practitioners. *European Psychiatry* **44**, 1–8.

https://doi.org/10.1017/S0033291718004191 Published online by Cambridge University Press