School of Psychology
Ysgol Seicoleg

Do neuromodulatory treatments, such as transcranial magnetic stimulation, really reduce depressive symptoms?

Thesis submitted in partial fulfilment of the requirement for the degree of:

Doctorate of Clinical Psychology (DClinPsy)

South Wales Doctoral Programme in Clinical Psychology

Cardiff University

Lowri Mair Hadden

Supervised by: Dr Helen Penny and Dr Chris Allen

2nd October 2020

Table of contents

Acknowledgements……6

Thesis preface……..……...7

Do neuromodulatory treatments really reduce depressive symptomatology: A systematic review & meta-analysis……....9

Abstract……10

Introduction……..11

Method……16

Search strategy & eligibility……16

Selection criteria……..17

Article retrieval……....18

Data extraction……20

Data analysis……20

Results……..21

Study characteristics and demographics……..21

Inclusion criteria and measurement……21

Depression diagnosis and comorbidity……....22

Depression symptomatology measurement: pre & post stimulation……...22

Treatment resistance and measurement……...23

Medication usage and measurement……24

TMS stimulation parameters: site of stimulation, and coil parameters……...24

Primary outcome measure: pre & post stimulation…….....26

Clinical relevance……28

Risk of Bias……29

Adverse effects……32

Discussion……....32

Clinical relevance……....36

Limitations…….....38

Conclusion…….....38

References…….....40

List of Tables

Table 1. Exclusion criteria for study selection…….61

Table 2. Characteristics of included studies…….....62

Table 3. Patient demographics, baseline depression severity & assessment, medication usage, & sponsor..……..63

Table 4. TMS treatment parameters, and coil type……..…….....64

Table 5. Study design, Treatment arms, pre & post measurement on HDRS……..…65

Table 6. % Change of depressive symptom reduction…….....66

Table 7. Adverse effects…….....67

Does excitatory pre-frontal stimulation reliably increase reward responsiveness?.....68

Abstract……...….69

Introduction……..70

Aims and hypotheses of the current study……...74

Primary hypothesis…….....74

Secondary hypotheses…….....75

Replication interactions of interest.....75

Method……76

Task overview and procedure……..……76

Participants ……..76

Study design…….……78

Probabilistic leaning task ……78

Transcranial magnetic stimulation…….…….…….79

Questionnaire: Temporal Experience of Pleasure Scale……..……80

Questionnaire: Positive and negative affect schedule……..……....81

Primary dependent variable- probabilistic learning task: response bias……..…81

Bayes factors and sample size estimates……..82

Analyses (pre-registered) & prior calculations……..…..82

Primary hypothesis…….…..……83

Secondary analysis…….…….……84

Replication interactions of interest…….…….……85

Manipulation checks……..……..…….87

Results…….…87

Pre-registered: primary hypothesis……...87

Pre-registered: secondary hypothesis……..…….....88

Positive affect……...……..……..88

Exploratory analysis……..……..…….88

Negative affect…….……..…..89

Exploratory analysis…….……89

Pre-registered: Replication/s of interest…….……..89

Pre-registered: Manipulation check.……....91

Discussion…….……...92

Replications of interest…….……93

Secondary hypotheses: mood ratings……..…….....95

Limitations……...97

Conclusions…….....98

References…….....99

APPENDIX A – Submission guidelines for the Journal of Affective Disorders….……115

APPENDIX B – Systematic review & Meta-analysis searches……....122

APPENDIX C – Screening for study selection ……..……..126

APPENDIX D – Risk of Bias assessment criteria……....128

APPENDIX E - RoB: Individual ratings with author judgement…….....129

APPENDIX F – Cortex guidelines for registered reports……138

APPENDIX G – In principle acceptance confirmation, Cortex……146

APPENDIX H - Deviations from Ahn et al., (2013) and Duprat et al., (2016) ……...……147

APPENDIX I - Exclusion Criteria……...149

APPENDIX J - Counterbalancing……...149

APPENDIX K – Target location TMS…….……...…….150

APPENDIX L – TMS parameters: treatment protocol…….……...…….150

APPENDIX M - TEPS questionnaire……...151

APPENDIX N - PANAS questionnaire……152

APPENDIX O - Bayes factors & sample size estimates……..……153

APPENDIX P – Behavioural results…….……154

APPENDIX Q – Cortex cut-off Bayes Factors ....……158

Acknowledgements

Alex, thank you, for your continued support, kindness and never-ending patience throughout this journey. I promise not to do any more doctorates, honestly!! With all my love.

Thank you to my academic supervisor, Dr Chris Allen for your meticulous approach to research, your guidance with the registered report and for teaching me so much- I am a TMS convert (well in so far as the evidence supports its use!). Thank you also to my clinical supervisor, Dr Helen Penny, for your guidance and wisdom with the systematic review, I really appreciated your enthusiasm and clinical expertise. I also want to thank Dr Thomas Lancaster and Dr Alex Jones for your expertise with the PLT and programming, and all of the participants who kindly gave their time to partake in our study, without you we wouldn't have a study. Also thank you to my appraisal tutor, Dr Chris Hobson, for your support, patience and compassion throughout training.

I also want to thank my wonderful in-laws: Paula, and Alan for you compassion, care-packages, and love when I inevitably couldn't make family functions, due to writing. We will have so many afternoon teas in celebration soon. Lots of love.

Thanks also to all my friends, old and new, for your words of wisdom, laughter, and just checking in on me: I love you all- Elena, Conna, Jen, Sarah, Kevanne, Awen, Kat, Julie, Biddy, Evelyn, Anna, Amber. Michelle, with whom I appreciated going through our training journey together, it would not have been the same without you, and my other fellow trainees. I can't wait to celebrate with you all, and remember what it's like to have real fun again!

Finally, thank you to my family for teaching me grit and endurance, and to love wholeheartedly.

I dedicate this thesis to my nain- llawer o gariad am byth.

Thesis Preface (761 words)

Depression is a serious mental health condition, which can have fatal effects, with over 800,000 people dying by suicide each year. It is also the leading cause of disability across the globe (WHO, 2017). Given the high rates of mortality, and disability, it is paramount to have evidence-based and clinically relevant treatments. One emerging treatment, with a relatively new evidence-base, is transcranial magnetic stimulation (TMS). TMS treatment protocols for depression are endorsed by two regulatory bodies, NICE (2015) and the FDA (2008). However, the basic scientific validity and mechanisms of action of these protocols remain vague, particularly relating to how excitatory TMS reduces a key depressive symptom, anhedonia. For our empirical chapter, we replicate two core studies (e.g. Ahn et al., 2013; Duprat et al., 2016) that have informed the clinical evidence-base of TMS for depression. One of these core studies, Ahn et al., (2013) demonstrated an increase in reward responsiveness in controls, following excitatory TMS compared to sham TMS, as measured on a probabilistic learning task (PLT). However, in a pseudo-replication study of Ahn et al., (2013), Duprat et al., (2016) only found an increase in reward responsiveness, following excitatory TMS, when including an exploratory variable, trait hedonic capacity, as a covariate. Our empirical chapter did not replicate the core effects of neither Ahn et al., (2013) nor Duprat et al., (2016), that is we did not find an effect of excitatory TMS compared to sham for reward responsiveness, nor did we demonstrate any effect with the inclusion of trait hedonic capacity. However, our results on the PLT were commensurate with the wider PLT evidence-base, evidencing that learning had occurred across the task, as was expected. In contrast to Ahn et al., (2013) and Duprat et al., (2016), we had also included mood ratings, using the Positive and Negative Affect Schedule (PANAS), both pre and post stimulation. Intriguingly, following active TMS, participants exhibited a significant decrease of positive mood, indicative of increased anhedonia. Negative mood was unaffected by TMS. These mood findings call into question the fundamental premise of TMS, which is thought to decrease symptoms of anhedonia, rather than increase them.

For our systematic review and meta-analysis, we examined the efficacy of TMS as a treatment for depression, taking into account the clinical relevance of these findings, which previous

TMS studies have failed to do. Previous "gold" standard randomised controlled trials (RCTs) assessing the effectiveness of active compared to sham TMS have generated variable estimates of treatment efficacy, with active compared to sham TMS evidencing small, medium, and large effect sizes, or indeed, no effect at all. These divergent findings may lie in the heterogeneous patient samples with comorbid disorders, and concurrent antidepressant medication included in these RCTs. Thus, we conducted a systematic review and meta-analysis of TMS for depression, controlling for previous heterogeneity through the inclusion of double-blind RCTs that included depressed patients who were anti-depressant free. We considered the questionnaires used to assess depressive symptom change, typically the Hamilton Depression Rating Scale (HDRS), and calculated the standardised mean difference (SMD/ hedges 'g') for scores on the HDRS post stimulation (active and sham). We found a small effect (SMD) of active compared to sham stimulation (SMD = -0.29), indicating that active TMS is effective in reducing symptoms of depression. However, the clinical relevance of this effect is negligible, based on the % change of depressive symptom reduction, as outlined in Lepping et al., (2014). We found a -31% reduction in depressive symptoms for active and -14% reduction for sham stimulation, which corresponds to "no change/ minimally improved" depression, commensurate with the findings of Lepping et al., (2014). Similarly, the participants, on average failed to move out of "casedness" for their depressive symptoms.

Combined, the results of our empirical paper and systematic review & meta-analysis, fail to provide convincing support for the use of TMS for treating depression. In particular, we were not able to replicate the main findings of Ahn et al., (2013) and Duprat et al., (2016). Moreover, TMS reduced positive affect, and made no difference to negative affect, when measured on self-report questionnaires. Our meta-analysis provided evidence towards the use of active TMS for depression, prior to calculating the clinical relevance of this effect, which generated no meaningful changes, or a reduction from "casedness". Thus, our studies highlight the importance of triangulating clinical and statistical methods to truly assess the efficacy and effectiveness of TMS as a treatment for depression. Not doing so, leads to an inconclusive and vague answer about treatment effects, which could result in potentially worse treatment outcomes for patients (i.e. increased anhedonia).

Do neuromodulatory treatments really reduce depressive symptomatology: A systematic review & meta-analysis. [1]

Word count (7477)

[1] This manuscript has been completed in line with the author guidelines for the Journal of Affective Disorders, including an abstract formatted in the manner stipulated by the journal. The author guidelines are included in Appendix A.

Abstract

Background
Transcranial magnetic stimulation (TMS) has been approved as a treatment for depression by NICE (2015) and the FDA (2008). However, there are discrepant findings around its efficacy. A possible explanation for these divergent findings may lie in these "gold standard" randomised controlled trials (RCTs) including heterogeneous patient samples, who have comorbid diagnoses and concurrent antidepressant use. Moreover, these RCTs fail to evidence the clinical relevance of TMS.
Methods
We conducted a systematic review and meta-analysis of TMS for depression in double-blind RCTs, controlling for previous heterogeneity by including homogeneous depressed patients, who were antidepressant free. We calculated depressive symptom change post active and sham stimulation. We also examined clinical relevance through linking % of depressive symptom change with previous modelled effects.
Results
Active compared to sham stimulation yielded a small, but significant, treatment effect (hedges 'g'= -0.29). However, the clinical relevance of this result is questionable, with both active and sham stimulation demonstrating little clinical improvement.
Conclusions
Active TMS demonstrates a small but significant effect on depression scores, in a controlled homogeneous sample of studies. However, the efficacy of TMS does not translate into a clinically relevant response, leading to important questions about the usefulness of using TMS in treating depression, for policy and practice.
Limitations
Clinically meaningful changes, as opposed to clinical relevance, are typically measured with self-report questionnaires. We could not conduct these analyses due to strict inclusion criteria of our meta-analyses yielding 6 studies (n = 285, sham; n = 290, active), that did not all consist of relevant self-report questionnaires.

250 words

Keywords: *Transcranial magnetic stimulation, Depression, Clinical relevance, Homogeneity, Efficacy*

1. Introduction

Over the past ten years, neuromodulatory protocols have been increasingly used to treat depression following the approval from two regulatory bodies (FDA, 2008; NICE, 2015).These regulatory bodies appraise "gold standard" evidence such as randomised controlled trials (RCTs) to determine whether a treatment condition (active) is more efficacious compared to a baseline (sham) condition. The most common neuromodulatory treatments, and of interest to this review, are repetitive transcranial magnetic stimulation (rTMS) and intermittent theta burst stimulation (iTBS), which can serve to facilitate (Huang et al., 2005, 2011) or inhibit (Huang et al., 2005, 2011) neuronal activity, depending on the frequency or parameters of the treatment protocols used (Fitzgerald & Daskalakis, 2011; George et al., 2013).

The most common site of stimulation used for rTMS interventions across RCTs, is the dorsolateral prefrontal cortex (DLPFC; Herrmann & Ebmeier, 2006; Lepping et al., 2014; Slotema et al., 2010). The DLPFC is associated with the neuro-circuitry of reward responsiveness (Ballard et al., 2011), which is posited to be reduced (anhedonia) in depressive disorders (Koenigs & Grafman, 2009). As such the DLPFC has become a target of excitatory TMS protocols in an attempt to upregulate neural activity within this region (Janicak & Dokucu, 2015).

Neuromodulatory treatments such as rTMS and iTBS involve generating a magnetic field via a TMS coil and stimulator (see review by Fitzgerald & Daskalakis, 2011; Slotema et al., 2010). The rapidly changing magnetic fields induced are deemed to modify the electrical field circuits of the brain, resulting in neuronal depolarisation close to the surface of the cortex (Janicak & Dokucu, 2015; Ridding & Rothwell, 2007). Low frequency TMS (LFTMS) is associated with frequencies of 1Hz and below (Fitzgerald & Daskalakis, 2011) and usually applied to the right DLPFC

(Lefaucheur et al., 2014), whereas High frequency TMS (HFTMS) is associated with frequencies of >5HZ (Liu et al., 2014) and typically applied to the left DLPFC (Allan et al., 2011). LFTMS and HFTMS are related to inhibitory (decrease of neural activity) or facilitatory (increase in neural activity) stimulation respectively (Huerta & Volpe, 2009), whilst iTBS is facilitatory, but can produce similar effects to HFTMS over a shorter period of time (e.g. Blumberger D.M. et al., 2018). Although the precise neural mechanisms of neurotransmitter action involved in TMS stimulation remain unresolved (e.g. Janicak & Dokucu, 2015; Noda et al., 2015), it is hypothesised that the neurotransmitters glutamate (Yang et al., 2014) and gamma aminobutyric acid (GABA) are modulated (Dubin et al., 2016). These neurotransmitters are posited to impact upon circuitry such as the meso-limbic system (Baeken C., 2017) involved in depressive symptomatology such as reduced mood, and avolition (Noda et al., 2015). However, this research is still in its infancy and no firm conclusions can as yet be drawn.

Some large multi-site randomised controlled trials (RCTs; George et al., 2010; O'Reardon et al., 2007), using sham compared to active stimulation and meta-analyses (Mutz et al., 2018; Schutter, 2009; Slotema et al., 2010) provide evidence to support the use of HFTMS for depression. Conversely, other multi-site RCTs (e.g. Herwig et al., 2007) and meta-analytic studies (Couturier, 2005; J. L. R. Martin et al., 2003) have found no such effects on decreasing depressive symptomatology, when comparing active HFTMS stimulation to a sham comparator, or indeed an equivalent decrease of depressive symptomatology in both active and sham conditions (Mutz et al., 2019).

A possible explanation for these discrepant findings could be the inclusion of patients who experience co-morbid psychiatric illnesses including post-traumatic

stress disorder (Fitzgerald et al., 2012), personality disorders (e.g. Januel D. et al., 2006) or depression in the context of bipolar disorder (Avery et al., 1999; Chistyakov A.V. et al., 2015), which may obfuscate the true efficacy of neuromodulatory treatments. Indeed, an RCT (Fitzgerald et al., 2016) that included a sample of patients with bipolar depression, found no therapeutic effects for bilateral active compared to sham stimulation. This suggests that patients with bipolar depression may respond differentially to TMS compared to patients with unipolar and treatment resistant depression (TRD). It has been argued that bipolar depression is qualitatively different from depression and TRD, within the context of a major depressive episode, which could be explained by differences in structural morphometry (Fung et al., 2015) and severity of depressive episode (Moreno et al., 2012). The inclusion of patients with bipolar depression in these earlier rTMS studies may conflate true stimulation results, and explain some heterogeneity found between these studies.

Moreover comorbid diagnoses are often not examined with validated questionnaires or using a clinical interview specific to the comorbid condition (George et al., 2014;Avery et al., 1999; Blumberger D.M. et al., 2016) leaving questions unanswered about the validity of depressive symptom change and rigorous experimental control. One major multi-site study examined the efficacy of depressive symptom reduction whilst taking into account comorbid anxiety, balanced across groups, and reported an enhanced therapeutic effect of rTMS in patients with comorbid disorders (Lisanby et al., 2009), suggesting an inflated rate of remission in patients with comorbid symptoms.

Additionally, many of the “gold standard” RCTs include patients who are also taking antidepressant medications, albeit these medications are controlled within the treatment period (e.g. Fitzgerald et al., 2012; Garcia-Toro et al., 2001; Li C.-T. et al.,

2014; Loo et al., 2001). Nevertheless, studies have consistently reported increased efficacy for active compared to sham rTMS in patients taking an adjunctive antidepressant (e.g. Sehatzadeh et al., 2019; Wang et al., 2017). However, improvement in depressive symptoms have also been reported for sham stimulation, suggesting a potential inflation of stimulation efficacy for both active and control conditions (Herwig et al., 2007). Sham coil orientation has been highlighted as another potential source of bias (cf. Duecker & Sack, 2015) with the blinding of treatment integrity being called into question. However, systematic reviews and meta-analyses which examined the blinding integrity between active and sham conditions found no difference in patients' ability to correctly guess their stimulation type, indicating at least a similar level of blinding between the two conditions (e.g. Berlim et al., 2013; Broadbent et al., 2011).

Another source of potential variability is how depressive symptoms are measured. Typically, improvement in depressive symptomatology is measured using a clinical interview such as the Hamilton Depression Rating Scale (HDRS; Hamilton, 1961), with a 50% improved score from baseline defined as a response, and a score of <7 as recovery (e.g. Furukawa et al., 2002; Lam et al., 2008; Lepping et al., 2014; Leucht et al., 2013; NICE, 2015; Riedel et al., 2010). However, a proportion of patients included within TMS treatment studies do not appear to respond (50% reduction in symptoms) or exhibit remission (<7 on HDRS) (Avery et al., 2007; Blumberger D.M. et al., 2016; Fitzgerald et al., 2009) following active TMS.

Importantly, these response and remission rates have not been related to clinical importance (e.g. Moncrieff & Kirsch, 2015). Leucht et al (2013) proposed a method of examining clinical relevance linked to symptom change (equipercentile linking) on the HDRS, which can provide evidence about clinical relevance (Lepping

et al., 2014). This approach seeks to link changes on the HDRS to clinical relevance of symptom change - that is, how well the patient appears to be functioning, as determined by an experienced clinician - using an interview such as the Clinical Global Impressions Scale (CGI; e.g. see. Busner & Targum, 2007). Nevertheless, this method does not provide insight into minimal clinically meaningful gains, where patients are asked to complete self-rated measures of symptom change, which are used to "anchor" clinician rated symptom changes (Button et al., 2015; McGlothlin & Lewis, 2014). However outcome measures such as validated self-report depression questionnaires, e.g. Beck Depression Inventory I or II (BDI-I ; BDI-II; Beck, Ward, & Mendelson, 1961; Beck, Steer, & Brown, 1996) have been used to assess the reduction of symptom severity, in tandem with assessment of the level of treatment resistance, defined as at least one failed antidepressant trial, measured using the Thase and Rush (1997) staging model.

To our knowledge no study has examined who administers these questionnaires and whether this has an impact on stimulation outcomes. Mutz et al., (2019) assessed the risk of bias in 113 RCTs included in their meta-analysis using the Cochrane Risk of Bias Tool, RevMan 5.3 (RoB, Cochrane Collaboration, 2014), and found variability within allocation concealment, and blinding of participants and personnel, afforded an "unclear risk" (as seen in their supplementary information), which could have impacted on the reliability of these studies' outcome measures. Of note Mutz et al., (2019)'s study included patients with both bipolar and unipolar depressive disorders, and studies that included medication usage.

Given the mixed findings for the efficacy of TMS, it is clear that there is some variability for the therapeutic effects between active and sham stimulation, but the reasons for this variability remain unanswered. Explanations for the variability

between studies could be related to a number of factors including experimental control (Demitrack & Lisanby, 2008) different study designs (Avery et al., 2007; Fitzgerald et al., 2012; Herwig et al., 2003), participant selection (Kedzior et al., 2015), stimulation parameters (Bakker et al., 2015), adequate blinding for both participants and personnel (Duecker & Sack, 2015), inclusion of patients with comorbid disorders (e.g. Lisanby et al., 2009) and adjunctive antidepressant usage (Hunter A.M. et al., 2019; Sehatzadeh et al., 2019).

The aim of the current systematic review and meta-analysis is to measure the efficacy of active compared to sham rTMS when redressing some of this variability through limiting the inclusion criteria to those patients who have treatment resistant depression alone, and who are medication free included in double-blind RCTs. In addition, we will assess the risk of bias using the RoB tool in RevMan 5.3 (Cochrane Collaboration, 2014), report stimulation parameters, and outcome measures used at baseline and for assessing depressive symptom change. We also aim to contextualise these findings with regard to % changes of depressive symptoms for active and sham stimulation linked to proposed clinical relevance, as outlined in Lepping et al., (2014) and Leucht et al., (2013).

2. Method

2.1. Search strategy & Eligibility

We conducted searches in four databases: PsycInfo; Medline; Embase and the Cochrane Controlled Register of Trials (CENTRAL). Dates of the searches were from inception of the databases; 1806 to 27th of November 2019, when we conducted the final searches. In line with our systematic review and meta-analysis question, we selected keywords related to neuromodulatory treatments endorsed for the treatment of depression (NICE, 2015; FDA, 2008); different types of depressive disorders; and

the neural area typically targeted for the use of neuromodulation- the DLPFC. Keywords associated with neuromodulatory treatments; depressive disorders and DLPFC were combined using the Boolean operator "AND". Additionally, a wildcard operator (*) was used where relevant, to maximise the number of relevant papers identified with our search terms. Our keywords included "neuromodulat*" *or* "transcranial magnetic stimulation" *or* "brain stimulation" *or* "tms" *or* "intermittent theta burst stimulation" *or* "itbs" *or* "repetitive tms" or "rtms" or "continuous theta burst stimulation" *or* "ctbs" **AND** " major depress*" *or* "depress*" *or* "treatment resistant depression" **AND "**Prefrontal Cortex" *or* "Dorsolateral Prefrontal Cortex" *or* "DLPFC" (see Appendix B for searches).

We adhered to the PRISMA guidelines for reporting systematic reviews (e.g. Liberati et al., 2009). Searches were limited, where possible, to those that included human participants, participants over the age of 18, and articles that were published in English. Bibliographic searches were also conducted in the reference list of each included study, and prior meta-analyses and systematic reviews relating to neuromodulatory treatments and depression. Prior to conducting the final searches, preliminary scoping searches were conducted to refine the search terms and criteria to ensure we captured all relevant studies, and a COCHRANE librarian was also consulted who corroborated our searches.

2.2. Selection Criteria

The author designed a screening form similar to that proposed in Boland, Cherry, & Dickson (2017; see Appendix C) and in accordance with the PICOS criteria (Participants, Intervention, Comparator, Outcome, Study design) outlined in the PRISMA guidelines (Liberati et al., 2009; Moher et al., 2015). This form was used

to ensure consistency in the selection criteria for each article screened and read in detail (also see Meline et al. 2006).

Included studies were those that met the following criteria: (1) Participants who received a diagnosis of treatment resistant depression (TRD) or major depressive disorder (MDD); (2) Facilitatory (>1HZ) and/ or Inhibitory (<1Hz) rTMS, or iTBS targeting the DLPFC as the site of stimulation; (3) Active compared to sham stimulation, as the comparator; (4) Studies that included measurement of depressive symptom change pre and post stimulation (i.e. via questionnaire; clinical interview); (5) Double-blind randomised controlled trials.

2.3. Article Retrieval

Using the search terms in the databases detailed above, 3379 study records were identified (see Figure 1), synced and transferred into the reference manager software, Zotero. After duplicate articles (1493), conference abstracts (402), clinical trial protocols (173) without associated published full-texts, or data were removed. A further twelve clinical trial protocols which had completed data collection and had associated publications were excluded due to not including depressed patients alone (11) or not having a TMS treatment intervention (1). The remaining 1296 abstracts and titles were screened using the inclusion and exclusion criteria, detailed above and in Appendix A. Of these articles, we removed review articles, editorials, case studies, case series and articles containing other psychiatric disorders, or psychiatric co-morbidities (n = 819). The first author subsequently read and screened the remaining 477 full-text articles against the inclusion and exclusion criteria and removed 460 articles that did not meet the inclusion criteria. For example, we removed studies where patients had comorbid diagnoses or diagnoses other than depression (119), the treatment intervention was not TMS (128), there was no sham comparator condition

(138), and the site of stimulation was not the DLPFC (17) (see Table 1 for additional article exclusion). After the removal of 460 articles, our final sample was 17 articles. An independent reviewer blindly screened the 17 included articles against the inclusion and exclusion criteria and obtained 100% agreement with the lead author. However, due to identifying the inclusion of patients with personality disorder/s in one study's discussion section (Januel et al., 2006), reported comorbidity in a table of another study (Li et al., 2019) and a series of large clinical trials (n = 8) using duplicate data (see Table 2), a further nine studies were excluded from the final inclusion criteria (in line with Mutz et al., 2019), leaving a remaining eight studies for qualitative synthesis and meta-analysis (see Figure 1).

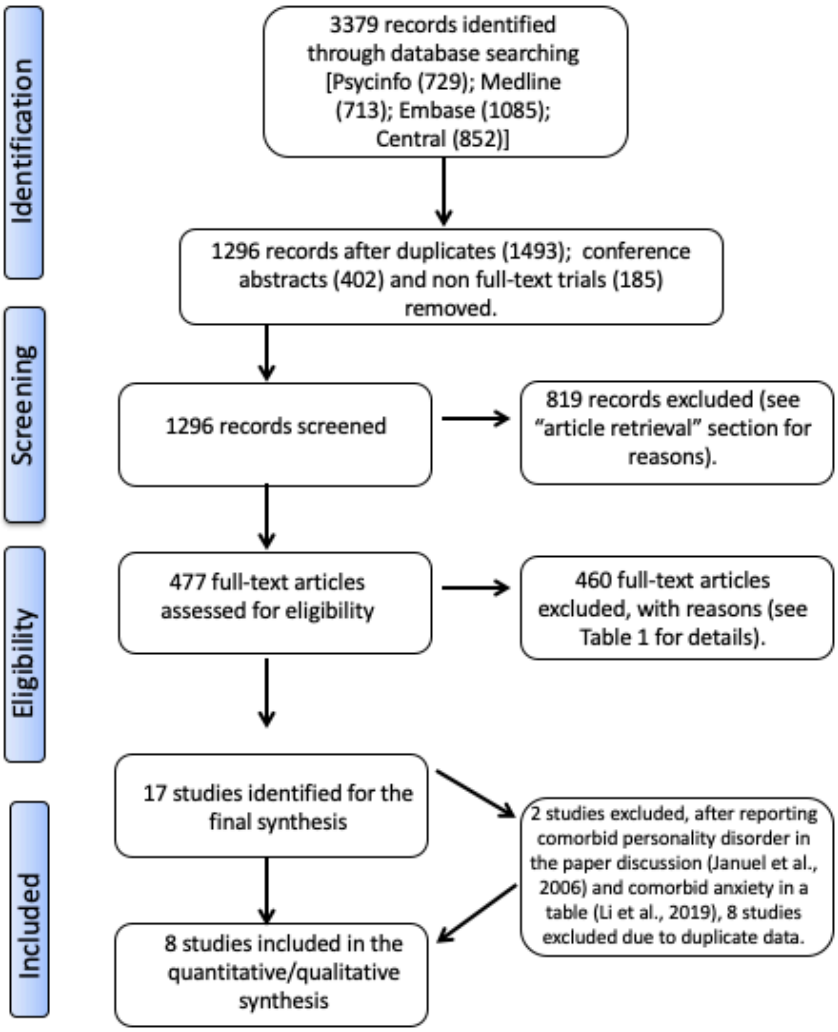


Figure 1. Study selection diagram based on PRISMA Guidelines.

---------- Insert Table 1 about here----------

2.4. Data Extraction

Data were extracted from the included papers for:

a) participants (gender, age), recruitment (inpatients/outpatients), stage of treatment resistance, depression severity, duration of current episode- and who measured these variables (e.g. independent researcher, psychiatrist);

b) intervention parameters, type of coil used, number of sessions, intensity of motor threshold, sponsor.

c) Means, *SDs*, and sample size relating to depressive symptom measurement post stimulation (active, sham) were also extracted for each included study, or where necessary from supplementary information. These data were imported into RevMan 5.3 (Cochrane Collaboration, 2014), where a subsequent meta-analysis would be conducted, and 25% of the extracted data was cross-checked by an independent reviewer.

2.5. Data analysis

Both qualitative synthesis and a meta-analysis were conducted to encapsulate themes in the data and to address the efficacy of TMS efficacy for active compared to sham stimulation in a highly controlled sample of patients with TRD (see "Primary Outcome Measure: Pre and Post Stimulation" for further details relating to the meta-analysis).

3. Results

3.1. Study Characteristics and demographics

The number of participants included across the studies was 654, with 339 included in the active stimulation (205 females) and 315 (164 females) in the sham condition, who were all aged between 18- 80 yrs (M = 46.81; SD = 10.72)[2]. Although all included studies were double-blinded RCTs, a proportion of these (n = 3) were single-centre cross-over studies (Baeken et al., 2013; Duprat et al., 2017; Vanderhasselt et al., 2009a; see Table 2), with participants counterbalanced to complete the active or sham condition first before being allocated to the other condition. The remaining studies were either single site double-blinded RCTs (Holtzheimer et al., 2004; Li et al., 2016; Stern et al., 2007) or multi-site and multi-phase double-blinded RCTs (George et al., 2010; O'Reardon et al., 2007).

---------- Insert Table 2 about here----------

3.2. Inclusion criteria & Measurement

The majority of studies included outpatients alone in their samples (n = 4; George et al., 2010; Holzheimer et al., 2004; O'Reardon et al., 2007; Stern et al., 2007), with two studies including a mixture of inpatients and outpatients (n = 2; Baeken et al., 2013; Duprat et al., 2016). There was no information regarding patients' hospital status provided for the remaining two studies (see Table 3; Li et al., 2016; Vanderhasselt et al., 2009a). Across the studies[3], the duration of the current

[2] As data was not reported for the parameters above in Vanderhasselt et al., (2009a) we were not able to include this within the data for active and sham stimulation, gender or age estimates above.
[3] Data from Stern et al., (2007) was not included in the age range, as no data relating to current depressive episode was provided in the study.

depressive episode varied from 2 months to 11 years, however, no information was provided as to how current depressive episode was calculated (i.e. through medical history, psychiatric interview).

3.3. Depression diagnosis & comorbidity

All participants received a diagnosis of depression, ranging from unipolar Major Depressive Disorder (Baeken et al., 2013; Li et al., 2016; Holzheimer et al., 2004), with the remaining studies classifying patients as having Major Depressive Disorder (see Table 3)- all being treatment resistant.

For all eight studies (see Table 3), depressive diagnoses were confirmed following a standardised clinical interview. However, the type of clinical interview differed between studies. For a proportion of studies (n = 4), the standardised neuropsychiatric interview, the Mini International Neuropsychiatric Interview (MINI; Sheehan, 1998) was used. The Structured Clinical Interview Diagnosis (SCID, 2004) was used to provide diagnosis for one study; and the remaining studies (n = 3) reported the use of the DSM-IV to make a diagnosis. Only two studies (George et al., 2010; Stern et al., 2007) provided information as to who diagnosed the patients using the clinical interviews, but none of the studies reported if the psychiatrist/ researcher who provided the diagnosis was independent to the study personnel (e.g. blinded).

The majority of studies did not explicitly measure or report whether participants experienced any other co-morbid mental health difficulties but stated within their exclusion criteria that participants with bipolar disorders (I, II), psychosis, or serious mental health disorders were excluded from the sample.

3.4. Depressive symptomatology measurement: pre and post stimulation

In the majority of studies (n = 6) baseline depression severity and post stimulation depression was measured using the 17-item HDRS (Hamilton, 1961).

However, the 21 and 24-item versions of the HDRS were also used, albeit less frequently (proportion; n = 2 studies; Stern et al., 2007; George et al., 2010, respectively). Of note, the 21 item HDRS is akin to the 17-item scale; however, concerns regarding the variability within the scores included on these measures have been raised (e.g. Williams, 2001).

All studies bar Vanderhasselt et al., (2009a) reported that either an independent blinded psychiatrist (e.g. Baeken et al., 2013; Duprat et al., 2016; Li et al., 2016; Stern et al., 2007), or a rater blind to treatment conditions conducted the assessment(e.g. George et al., 2010; Holtzheimer et al., 2004; O'Reardon et al., 2007) (see Table 3). The two multi-site RCTs (O'Reardon et al., 2007; George et al., 2010) also reported that their independent clinical raters were trained to administer the outcome measures and the standard was monitored throughout the study. All participants met the criteria for at least moderate depression, as defined by the 17-item HDRS. Two studies (Baeken et al., 2013; Stern et al., 2007) indicated baseline depression scores within the "severe" range, whilst the remaining studies included a sample of patients who met criteria for "moderate" depressive disorder. Cut-off scores for the 17- item HDRS are posited to be 0-7 for "normal range"; 8-16 are indicative of "mild depression"; scores of 17-23 "moderate depression" and scores above 24 are indicative of "severe" depression (see e.g. Zimmerman et al., 2013).

---------- Insert Table 3 about here----------

3.5. Treatment resistance and Measurement

The majority of studies (n = 4) included participants who had failed at least one antidepressant treatment (treatment resistance range 1-3; see Table 3), but apart from two studies (George et al., 2010; Stern et al., 2007) of which a psychiatrist

conducted the ratings, no other study reported who administered the questionnaires relating to treatment resistance. The most common outcome measure used to assess treatment resistance was the antidepressant treatment history form (ATHF; Sackheim, 2001; n = 3 studies), with Rush et al., (2003)'s staging model of treatment resistance being used in two studies (See Table 3), and the remaining studies (Li et al., 2016; Stern et al., 2007, Vanderhasselt et al., 2009a) not specifying how treatment resistance was measured.

3.6. Medication Usage & Measurement

Participants were free from psychotropic, and antidepressant medication prior to the beginning of brain stimulation, in all included studies. Antidepressant washout periods between studies varied between 1-2 weeks, with some studies (e.g. Vanderhasselt et al., 2009a; George et al., 2010) providing a longer washout period (e.g. 3-5 weeks) for patients who were taking fluoxetine prior to starting TMS treatment. Nevertheless, many studies (n = 6) permitted participants to take low doses (e.g. 2mg-150mg) of "rescue medications" in the form of benzodiazepines or hypnotics, if needed. Some of these studies provided details regarding the number of patients who took "rescue medications" (Baeken et al., 2013; Duprat et al., 2016; see Table 3).

3.7. TMS Stimulation Parameters; site of stimulation and coil parameters

All included studies used HFTMS (n = 7) or iTBS (e.g. Duprat et al., 2016) as the active stimulation condition (see Table 4) applied to the left DLPFC. Only one study (Stern et al., 2007) also included a LFTMS condition to both the left and right DLPFC. Both Baeken et al., (2013) and Duprat et al., (2016) conducted accelerated treatment protocols, that is a greater number of sessions daily (5 sessions per day for a week, equating to 20 sessions) compared to the other studies which applied their

HFrTMS treatment protocols once per day but over a longer period of time. Some studies' treatment protocols were conducted for 10 sessions (e.g. Holtzheimer et al., 2004; Li et al., 2016; Stern et al., 2007), whilst others were conducted for 20 sessions (e.g. Baeken et al., 2013; Duprat et al., 2016). However, both multi-site RCTs included more than 10 sessions, with O'Reardon et al., (2007) providing 20-30 sessions (based on treatment responsiveness) in their acute phase, and George et al., (2010) providing 15 treatment sessions in the acute treatment phase. Vanderhasselt et al., (2009a) only reported one treatment session (See Table 4).

The total number of pulses applied across the different studies varied from 16,000 (Li et al., 2016) to ~31,000 for Baeken et al., (2013) and Duprat et al., (2016), with the multi-site studies providing 45,000 pulses (George et al., 2010) and 60,000 - 90,000 (O'Reardon et al., 2007) for phase 1 and acute treatment phase respectively. Although number of treatment arms varied between studies (see Table 5) all studies included active compared to sham stimulation conditions. Six studies included 2 treatment arms, whereas two studies included more than 2 arms (Li et al., 2016; Stern et al., 2007).

Studies used either neuro-navigation (n = 4) or a combination of neuro-navigation and moving the coil 5cm anterior to the motor cortex to identify the site of stimulation. Motor thresholds varied between 80% (e.g. Stern et al., 2007) to 120% for stimulation intensity, in line with TMS safety guidelines (e.g. Machii et al., 2006; NICE, 2015).

Across five studies (Baeken et al., 2013; Holtzheimer et al., 2004; Li et al., 2016; Stern et al., 2007; Vanderhasselt et al., 2009a), active and sham stimulation was conducted with coils that were identical (sound, similar visually), but the coil was rotated on the scalp for the sham condition. Two studies included coils that induced similar somatosensory effects for both active and sham- including being visually

identical (George et al., 2010; O'Reardon et al., 2007), whereas one study (Duprat et al., 2016) included a coil that looked similar, but produced different somatosensory sensations (see Table 4).

---------- Insert Table 4 about here----------

3.8. Primary Outcome Measure: Pre and Post Stimulation

Using Review Manager Software 5.3 (Cochrane Collaboration, 2014), we performed a meta-analysis for depressive symptom scores on the HDRS[4] post active compared to sham stimulation. Means, Standard deviations and number of participants for HDRS for each condition were extracted in order to calculate the weighted standardised mean difference between conditions. We were unable to obtain data for two studies (Duprat et al., 2016; Vanderhasselt et al., 2009a), thus the meta-analysis was completed for the remaining six studies (See Figure 2; Table 5). [5] For equivalence between studies we extracted data for HFTMS stimulation for the active stimulation condition in studies where there were 3 or more treatment arms (e.g. Li et al., 2016; Stern et al., 2007). Similarly, to reduce clinical heterogeneity we extracted data where at least 10 treatment sessions had taken place over the course of a week (See Tables 3 & 4 for precise stimulation parameters and TMS stimulation parameters section above for further information) [6]. To avoid carry over effects we took the first set of data for the cross-over RCTs (similar to Mutz et al., 2019).

---

[4] Four studies included the 17-item HDRS; 1 included the 21-item version, and one study included the 24-item version. Of note the SMD allows us to combine different measures, as it standardises assessments to the same scale, before assessing the intervention effect with associated variability for that study.

[5] Baseline HDRS (pre stimulation) were $M = 24.76$; $SD = 4.99$ (active) and $M = 24.33$; $SD = 5.1$ (sham), and SMD = -0.04; CI = -0.20; 0.12, p = 0.64; $I^2 = 0\%$ suggesting no differences using a frequentist approach.

[6] Data for sham and active conditions for fewer than 10 sessions was also inspected, and provided similar *M, SDs,* for active compared to sham for greater than 10 sessions.

The total number of participants included for the active stimulation and sham conditions were 285, and 290, respectively. The heterogeneity between studies was measured using Cochran's Q statistic and I^2 (Deeks, Higgins & Altman, 2008). In line with a meta-analysis by Jiang et al., (2019), values >50% for I^2 or <0.05 for Cochran's Q, indicate significant heterogeneity between studies, and signifies the use of a random rather than fixed effects model (also see Deeks, Higgins & Altman, 2008). Our data demonstrated substantial heterogeneity, $I^2 = 69\%$, and thus we used a random effects model (in line with DerSimonian & Laird ,1986), as modelled in Cochrane's RevMan 5.3 (Cochrane Collaboration, 2014).

Overall, scores on the HDRS for post active compared to sham stimulation were significantly different (SMD = -0.54, 95% CI, -0.94, -0.14, p = 0.009, see Figure 2) indicating a greater efficacy for active compared to sham stimulation, with a moderate effect. However, when removing Stern et al., (2007), which was the source of the data heterogeneity due to an effect size (SMD) of order of magnitude greater than the other studies, there was no data heterogeneity, $I^2 = 0\%$ (also known as a sensitivity analysis). We subsequently calculated the effect of active TMS compared to sham TMS using a fixed effects model. The efficacy of active TMS compared to sham had decreased to SMD = -0.29, 95% CI, -0.46, -0.12, p = 0.0007, a small effect (with n= 275 patients included for both active and sham conditions). According to Cohen (1988) SMD of 0.2, 0.5, and 0.8 represent, small, medium and large effect sizes respectively (also see Faraone, 2008).

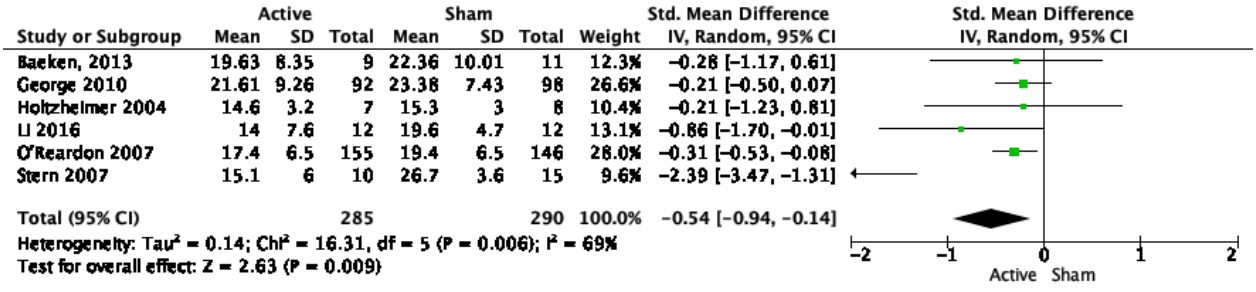


Figure 2. Forrest plot representing the efficacy of treatment effects between active and sham conditions, as measured for >10 sessions of rTMS using HDRS.

Typically, risk of publication bias is assessed using a funnel plot, where a symmetrical inverse funnel shape is indicative of low risk of bias, and an asymmetrical shape indicates a potential risk of bias. However, as stipulated in Deeks, Higgins & Altman (2008) data in funnel plots can be artificially inflated when including fewer than ten studies in a meta-analysis, and also when including continuous data, there is a risk of auto-correlation between treatment and control conditions, thus we did not explore this analysis.

---------- Insert Table 5 about here----------

3.9. Clinical Relevance

In order to understand the clinical relevance of our findings, and similar to Lepping et al., (2014), we calculated the % change (post-pre/ pre) of scores on the HDRS from pre stimulation to post stimulation, for active and sham stimulation separately (see Table 6). We subsequently linked the mean % change for reduction in the HDRS scores to the calculated clinical relevance that Leucht et al., (2013) and Lepping et al., (2014) had mapped out in previous studies. These authors computed % changes on the HDRS through linking them to a clinician rated scale, the Clinical Global Impressions scale, which provides an index of functioning and recovery following a medical treatment (Busner & Targum, 2007).

In our study, the mean % change on the HDRS was -31% for active and -14% for sham stimulation. These % changes correspond with the clinical response of "no change/ minimally improved" as calculated by Leucht et al., (2013) and Lepping et al., (2014). The active stimulation % change is almost in the "minimally improved"

bracket which is conceptualised as a change of -33 and above, whereas “no change” is thought to be between -9 to -32% reduction. Sham stimulation change is closer to the “no change” bracket. These % change scores indicate that depressive symptoms have reduced as a consequence of stimulation, with a greater reduction for active compared to sham, albeit the clinical relevance of these findings are minimal.

Similarly, depression scores, on average, reduced from pre to post stimulation, irrespective of stimulation type. However, in terms of depression severity, as measured on the HDRS, on average, no patients moved out of “casedness” for depression. Depression scores following active stimulation remained in the “mild” or moderate” depression range (See Table 6). Similarly post sham stimulation patients’ depression scores, on average, still remained in the mild, moderate or severe (e.g. Stern et al., 2007) range, indicating little clinical improvement.

---------- Insert Table 6 about here----------

3.10. Risk of Bias

Two independent raters assessed Risk of Bias (RoB) for each included paper, using a validated RoB tool, RevMan 5.3 (Cochrane Collaboration, 2014), recommended as the “gold” standard for RCTs (e.g. Higgins & Green, 2008). This tool has also been used in many current systematic reviews and meta-analyses (e.g. Jiang et al., 2019; Lage et al., 2016; D. M. Martin et al., 2017; Mutz et al., 2019) to appraise the quality of RCTs. Five sources of bias were appraised (selection, performance, detection, attrition, and performance) as demonstrating “low, unclear or high” risk of bias, based on the criteria specified in Higgins & Green (2008). For example, selection bias refers to how participants were allocated into the studies, e.g. were they randomised, and how was this randomisation conducted. A score of 0 was

provided for “low risk”, a score of 1 for “unclear risk” and a score of 2 for “high risk”, following the criteria outlined in Higgins & Green (2008) (see Figure 3, and for scores for each study independently Figure 4). Similar to Mutz et al., (2019) we calculated an overall risk of bias, based on the number of “low, unclear or high” biases for each source of bias (see Appendix D for full criteria; and Appendix E for individual study RoB ratings). For an overall “low” RoB a study must have been rated either low risk across all categories or have one “unclear” category. Overall “Unclear” RoB was provided if a study had 2 or more domains that were rated as “unclear”; and “High” RoB was provided as overall rating, if a study had rated as “High” in any one domain (see Figure 4 for individual study scores across domains).

Inter-rater reliability was assessed using Intraclass correlation, for the absolute agreement between raters one and two for each RoB category. The total level of agreement was 94%, CI: .90- .97, indicating excellent reliability between the raters in their assessment of RoB (Koo & Li, 2016; Liljequist et al., 2019).

The greatest source of bias across the studies was allocation bias, that is how patients were allocated into groups, with 100% of the studies included not providing clear enough information to be rated as a low risk of bias. Similarly, 50% of studies did not provide sufficient information about their blinding (performance bias) to confer adequate blinding of participants and personnel during treatment sessions. Finally, only 40% of studies reported sufficient information to conclude how participants were randomised to active or sham treatment conditions (See Figure 3). Only one study (Vanderhasselt et al., 2009a; see Figure 4; and Appendix E) was rated as an unclear RoB for detection bias (blinding of outcome assessment). When averaging the RoB ratings across the studies to determine the overall RoB, 75% of

studies indicated met an “unclear” RoB, which is in concordance with Mutz et al., (2019).

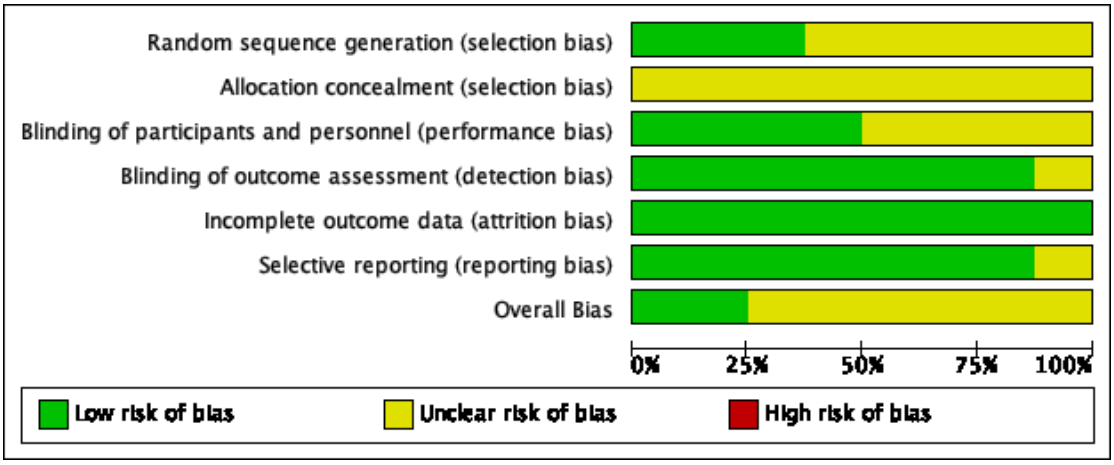


Figure 3. Risk of bias graph consisting of review author’s judgements for each risk of bias item presented as percentages across all studies included.

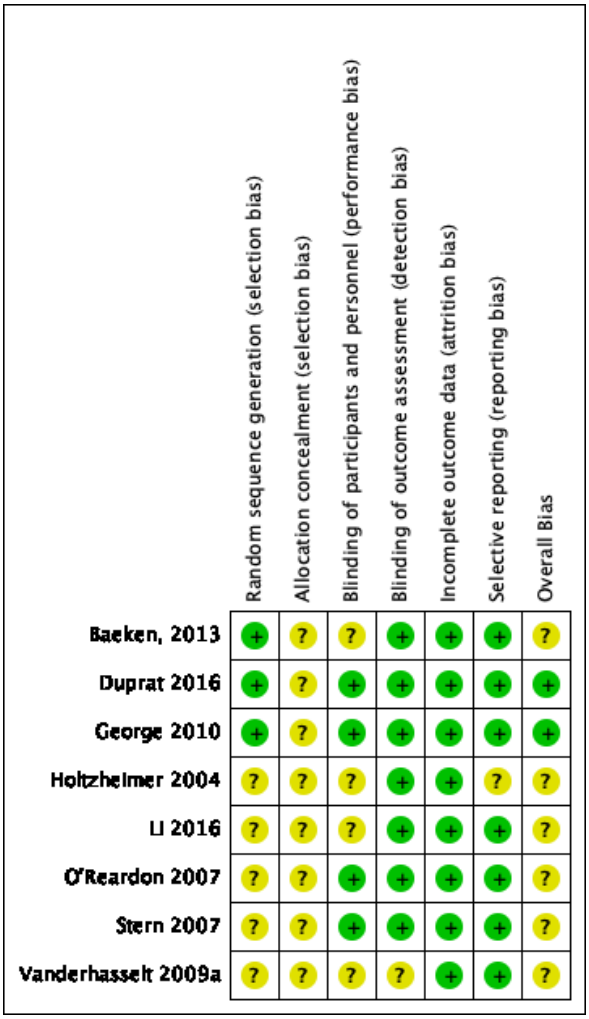


Figure 4. Individual ratings across domains for each included study in the systematic review.

3.11. Adverse effects

All studies apart from Vanderhasselt et al., (2009a) had reported that patients had experienced some form of adverse effect, including headache, fatigue, pain, scalp discomfort (see Table 7). However, the two multi-site RCTs (O'Reardon et al., 2007; George et al., 2010) reported that a proportion of patients, 15% and 5.4% respectively withdrew from the studies due to severe headache, pain, syncope and worsening of depressive symptoms and suicidality. Stern et al., (2007) also reported that 3 patients withdrew due to headache.

---------- Insert Table 7 about here----------

4. Discussion

The current systematic review and meta-analysis, to our knowledge, is the first to examine the efficacy of active compared to sham TMS in a sample of TRD patients, with no reported psychiatric comorbidity, and who were antidepressant free.

Congruent with previous meta-analyses that included patients with depression (e.g. Allan et al., 2011; Herrmann & Ebmeier, 2006; Lam et al., 2008; Mutz et al., 2018; Schutter, 2009; Slotema et al., 2010) we found that active HFTMS to the left DLPFC was more efficacious compared to sham stimulation, with a small effect; SMD/ hedge's g = -0.29. This effect size is commensurate with extant TMS meta-analytic studies such as Schutter et al., (2009) that reported a small effect size (*d*) of 0.39 for the difference in % change in the HDRS from baseline to end of treatment in active stimulation. Similarly, Slotema et al., (2010) reported a moderate effect size of 0.55 for active compared to a sham comparator in patients who had demonstrated a response of 50% improvement in depressive symptoms. However, despite having low

statistical heterogeneity, as measured using I^2, the study sample in Slotema et al., (2010) included patients with comorbid psychiatric disorders, which may have inflated the true efficacy of treatment effects for rTMS for depression.

In contrast to the extant literature, our meta-analysis and review included a homogeneous sample of depressed patients, using double-blind RCT designs, which enabled us to examine the efficacy of rTMS for depression and confer a small effect for active compared to sham rTMS. Furthermore, although our meta-analysis included six studies, due to not being able to obtain data for all nine, 285 patients were included in the active condition, and 290 in the sham condition, which is similar to participant numbers in other meta-analyses (Lepping et al., 2014; Leucht et al., 2013). Moreover, despite our meta-analysis demonstrating low methodological heterogeneity, with all included studies using HFTMS to the left DLPFC for at least 10 sessions, there were differences in the speed in which treatments were delivered, with Baeken et al., (2013) providing accelerated HFrTMS over the course of a week. However, when examining our forest plot (e.g. Figure 2), there were similarities in effect sizes between all included studies irrespective of treatment parameters.

For the majority of our included studies (n = 4/6 included in the meta-analysis) a version of the same outcome measure – the 17- item HDRS- was used across all studies, providing a commensurate measure of depressive symptom change across studies. Also, using the SMD enables us to compare depressive symptom changes on various versions of the same questionnaire, due to standardising the measure to the same scale (e.g. Higgins & Green, 2008).

It is important though to note that the HDRS was developed in the 1960's to measure depressive symptom changes following antidepressant trials (Bagby et al., 2004), in tandem with depression diagnoses provided using the Diagnostic and

Statistical Manual III. It does not currently reflect depressive diagnosis included within the DSM-IV, in particular the dimension related to anhedonia-or loss of pleasure. This could be problematic for TMS intervention studies, as the anhedonic dimension of depression is that which TMS is thought to upregulate, through targeting meso-limbic cortical structures, such as the DLPFC (Janicak & Dokucu, 2015). This factor could make the use of the HDRS questionable in inferring meaningful depressive changes following TMS, although this questionnaire is the most commonly used across the TMS literature, making findings comparable. Similarly, caution must be applied as to whether the HDRS captures the complexity of depression's multidimensionality when depressive symptoms are reported as a single unified dimension (Bagby et al., 2004; Kyle et al., 2016; Williams, 2001)- although validity and reliability appear to be better on the overall depression construct than for individual items (e.g. see Bagby et al., 2004). Only one study (O'Reardon et al., 2010) in our meta-analysis provided scores across the depression dimensions of the HDRS, which precluded meaningful interpretation across studies.

We did, however, determine that across all studies, the HDRS was administered by independent raters (albeit a psychiatrist, or researcher), who were reported to be blinded to the treatment conditions. However, an issue of contention within TMS research (e.g. Duecker & Sack, 2015) is the ability to be truly blinded to active compared to sham stimulation due to coil properties such as coil orientation and the somatosensory experience of the sham coil, potentially leading to unblinding of treatment conditions (C. K. Loo et al., 2000).

We examined the potential RoB related to this issue (blinding of personnel and patients) and across other domains (e.g. allocation bias, selection bias) and similar to Mutz et al., (2019) the majority of our included studies (75%) were rated as having

an “unclear” RoB across domains. In particular, the “unclear” RoB for random sequence generation and selection bias was 60% and 100% respectively, which relates to the randomisation of patients into active and sham groups, and how this information was concealed. Similarly, 50% of studies were judged to have an “unclear” RoB for blinding of patients and personnel. Taken together, these RoB ratings make it difficult to determine whether patients were able to tell which treatment they received for these studies, and thereby could modulate the true efficacy of rTMS. Nevertheless, Berlim et al., (2013) have reported that patients were only slightly better than chance at determining the treatment arm they received when asked to guess whether they were given active or sham TMS. However, across all studies, blinded personnel, who were required to be independent of the treatment sessions conducted the HDRS ratings both pre and post stimulation sessions, suggesting little impact on the potential bias on outcome measures. Although scores on the HDRS can vary based upon skill level and the experience of the rater providing the outcome measure (e.g. Hooijer et al., 1991), it was the same rater/s who conducted the HDRS between active and sham stimulation, which would mean that level of experience would not introduce a systematic bias between conditions. Future studies should measure the level of rater expertise empirically.

Despite including a homogeneous patient population in terms of diagnosis in our study, the hospital status of the patients varied with some being inpatients, and some being a mixture of outpatients and inpatients, which is similar to those included in the extant literature (Jin & Phillips, 2014; Kang et al., 2016; Taylor et al., 2018). Furthermore, the patients included in our study had a variable length of current depressive episode ranging from 2 months to 11 years, which could modulate the treatment effectiveness, as level of treatment resistance has been reported to be a

predictor of response to rTMS (Beuzon et al., 2017) with patients who have greater treatment resistance responding more poorly to active TMS (e.g. Kedzior et al., 2015).Similarly, although we included a sample of medication free patients- with a wash-out period of at least 1-week pre stimulation sessions, in the majority of these studies, "rescue medication" (benzodiazepines) were permitted, which could have decreased the efficacy of active TMS (cf. Hunter et al., 2019).

We also acknowledge that our effect size was moderate (SMD/ Hedge's g = -0.54) with the inclusion of Stern et al., (2007), however, due to the level of statistical heterogeneity created by the magnitude of Stern's effect (towards active stimulation), we reported the more conservative SMD of -0.29, in line with our sensitivity analysis, and decreasing statistical heterogeneity. It is possible that the true effect size of active compared to sham TMS is greater that reported (-0.29) when considering the use of "rescue medications" and level of current treatment resistance outlined above.

4.1. Clinical relevance

It is likely that TMS is an efficacious treatment for depression, but with the small effect size we found, comparable to the larger evidence-base (Slotema et al., 2010; Schutter et al., 2009; Herrmann & Ebmeier, 2006), we need to be cautious about how clinically meaningful these effects are for patients with depression. Indeed, when comparing the % change of reduction in depressive symptoms following active and sham stimulation we found "no change/minimal improvement" in relation to changes mapped out for clinical improvement (Lepping et al., 2014; Leucht et al., 2013). The "no change"/ "minimal improvement" is congruent to the level of clinical significance TMS reported by Lepping et al., (2014). This finding also coincided with patients not moving out of "casedness" following neither active nor sham stimulation. Thus, the clinical relevance of this small effect size should be called into

question in terms of the number of patients needed to treat and rates of remission and recovery, in order to make clinically meaningful gains (e.g. Button et al., 2015; Lepping et al., 2014). Moreover, a minimal clinically important difference, which is the smallest clinical difference that is important to patients (Button et al., 2015; McGlothlin & Lewis, 2014) is typically measured using self-report patient scales, which can be "anchored" to a clinician-rated questionnaire. However, we were not able to measure this variable empirically in our meta-analysis due to no studies including self-report questionnaires.

It is also important to consider other approved evidence-based treatments for depression such as cognitive behaviour therapy (CBT) and interpersonal therapy that also demonstrate slightly larger (moderate – large) effect sizes for intervention compared to treatment as usual using the HDRS as an outcome measure (e.g. see Cuijpers et al., 2011; Driessen et al., 2010, respectively). However, these psychological treatments are endorsed for mild- moderate depression (e.g. NICE; 2009) whilst TMS has been approved for treatment resistant depression (e.g. NICE, 2015). Though, mindfulness based cognitive therapy (MBCT) is a promising future avenue for treatment resistant depression, with efficacious and long lasting effects compared to treatment as usual (Kuyken et al., 2015, 2019), and should be compared against TMS. Furthermore, a proportion of patients, as indicated by our meta-analysis, are likely to experience short-term mild adverse effects such as headache, fatigue, dizziness, and "worsening of depressive symptoms" as a consequence of TMS, which are unlikely to occur for psychological therapies of a similar level of efficacy. Although, interestingly, a proportion of patients have also reported to experience psychological distress including "worsening of depressive symptoms (29%)" following CBT (e.g. Schermuly-Haupt et al., 2018).

Nevertheless, NICE (2015) have approved the use of rTMS to treat depression, based on the results of a number of meta analyses and systematic reviews, irrespective of site of stimulation (left, right or bilateral), and frequency of stimulation. Since these guidelines have a direct impact upon patient care and treatment, it is vital policy makers examine both the short term and long-term usage of TMS as a treatment strategy taking into account clinical relevance, and clinically meaningful changes, thereby improving its clinical utility and usefulness.

4.2. Limitations

Due to the heterogeneity in previous TMS efficacy literature, we aimed to conduct a meta-analysis with a homogeneous sample of depressive patients, who were medication-free, had a similar number of treatment sessions, and did not display comorbid psychiatric diagnoses. This enabled us to determine the strength of evidence for depressive symptom change, for active compared to sham TMS. However, due to our inclusion criteria, we included a small number of studies, that did not all include self-report measures, or questionnaires related to clinical relevance. This made it difficult to calculate clinically meaningful changes through the anchoring of self-report methods with clinician rated symptom changes such as the HDRS.

4.3. Conclusion

In conclusion, we find that active TMS has a small but significant effect on depression scores as measured on the HDRS, in a controlled homogeneous sample of studies. This tallies with existing findings in the literature, however, we caution that this significant result does not translate into clinical significance nor demonstrate a robust reduction of depression “casedness”. Furthermore, the key symptom of depression, anhedonia, which TMS is thought to target, is unlikely to be adequately captured in the HDRS, which was normed alongside an earlier version of the DSM.

Thus, the clinical relevance of TMS remains unresolved, which necessitates further research and raises important questions about the continued use of this treatment for depression. Future research needs to include additional measures of depressive symptom change, such as self-report questionnaires and clinician rated symptom change, CGI-S, to triangulate information and ascertain precise clinically meaningful differences to better inform policy and guidelines. These effect sizes also should be compared with other effective treatments for treatment resistant depression such as MBCT.

References[7]

Allan, C. L., Herrmann, L. L., & Ebmeier, K. P. (2011). Transcranial magnetic stimulation in the management of mood disorders. *Neuropsychobiology*, *64*(3), 163–169. https://doi.org/10.1159/000328951

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders (3rd ed.).* Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th ed., Text Revision).* Washington, DC: Author.

Avery, D. H., Claypoole, K., Robinson, L., Neumaier, J. F., Dunner, D. L., Scheele, L., Wilson, L., & Roy-Byrne, P. (1999). Repetitive transcranial magnetic stimulation in the treatment of medication-resistant depression: Preliminary data. *Journal of Nervous and Mental Disease*, *187*(2), 114–117. https://doi.org/10.1097/00005053-199902000-00009

Avery, D. H., Holtzheimer, P. E. I., Fawaz, W., Russo, J., Neumaier, J., Dunner, D. L., Haynor, D. R., Claypoole, K. H., Wajdik, C., & Roy-Byrne, P. (2007). Transcranial magnetic stimulation reduces pain in patients with major depression: A sham-controlled study. *Journal of Nervous and Mental Disease*, *195*(5), 378–381.

Baeken C. (2017). The impact of accelerated HF-rTMS on neurochemicals in major depression: Insights from 1H MR spectroscopy. *European Archives of Psychiatry and Clinical Neuroscience*, *267*(Supplement 2), S138. https://doi.org/10.1007/s00406-017-0856-0

[7] The references provided in bold are those that are included in the systematic review in APA 6 style, congruent with the submission guidelines of the JAD.

Baeken, C., Marinazzo, D., Wu, G.-R., Van Schuerbeek, P., De Mey, J., Marchetti, I., Vanderhasselt, M.-A., Remue, J., Luypaert, R., & De Raedt, R. (2014). Accelerated HF-rTMS in treatment-resistant unipolar depression: Insights from subgenual anterior cingulate functional connectivity. *The World Journal of Biological Psychiatry*, *15*(Clinical Psychopharmacology [3340]), 286–297. https://doi.org/10.3109/15622975.2013.872295

Baeken, C., Vanderhasselt, M.-A., Remue, J., Herremans, S., Vanderbruggen, N., Zeeuws, D., Santermans, L., & de Raedt, R. (2013). Intensive HF-rTMS treatment in refractory medication-resistant unipolar depressed patients. *Journal of Affective Disorders*, *151*(Health&Mental Health Treatment&Prevention [3300]), 625–631. https://doi.org/10.1016/j.jad.2013.07.008

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? *The American Journal of Psychiatry*, *161*(12), 2163–2177. https://doi.org/10.1176/appi.ajp.161.12.2163

Bakker, N., Shahab, S., Giacobbe, P., Blumberger, D. M., Daskalakis, Z. J., Kennedy, S. H., & Downar, J. (2015). rTMS of the Dorsomedial Prefrontal Cortex for Major Depression: Safety, Tolerability, Effectiveness, and Outcome Predictors for 10 Hz Versus Intermittent Theta-burst Stimulation. *Brain Stimulation*, *8*(2), 208–215. https://doi.org/10.1016/j.brs.2014.11.002

Ballard, I. C., Murty, V. P., Carter, R. M., MacInnes, J. J., Huettel, S. A., & Adcock, R. A. (2011). Dorsolateral Prefrontal Cortex Drives Mesolimbic Dopaminergic Regions to Initiate Motivated Behavior. *The Journal of*

Neuroscience : The Official Journal of the Society for Neuroscience, *31*(28), 10340–10346. https://doi.org/10.1523/JNEUROSCI.0895-11.2011

Beck, A.T., Ward, C., Mendelson, M. (1961). Beck Depression Inventory (BDI). *Archives of General Psychiatry,* 4 (6), 561-571.

Beck, A.T., Steer, R.A., Brown, G.K. (1996). *Manual for the Beck Depressive Inventory: II*. San Antonio, Texas: Psychological Corporation.

Berlim, M. T., Broadbent, H. J., & Van den Eynde, F. (2013). Blinding integrity in randomized sham-controlled trials of repetitive transcranial magnetic stimulation for major depression: A systematic review and meta-analysis. *International Journal of Neuropsychopharmacology*, *16*(5), 1173–1181. https://doi.org/10.1017/S1461145712001691

Beuzon, G., Timour, Q., & Saoud, M. (2017). Predictors of response to repetitive transcranial magnetic stimulation (rTMS) in the treatment of major depressive disorder. *L'Encephale*, *43*(1), 3–9. https://doi.org/10.1016/j.encep.2016.11.002

Blumberger D.M., Maller J.J., Thomson L., Mulsant B.H., Rajji T.K., Maher M., Brown P.E., Downar J., Vila-Rodriguez F., Fitzgerald P.B., & Daskalakis Z.J. (2016). Unilateral and bilateral MRI-targeted repetitive transcranial magnetic stimulation for treatmentresistant depression: A randomized controlled study. *Journal of Psychiatry and Neuroscience*, *41*(4), E58–E66. https://doi.org/10.1503/jpn.150265

Blumberger D.M., Vila-Rodriguez F., Thorpe K.E., Feffer K., Noda Y., Giacobbe P., Knyahnytska Y., Kennedy S.H., Lam R.W., Daskalakis Z.J., & Downar J. (2018). Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): A

randomised non-inferiority trial. *The Lancet*, *391*(10131), 1683–1692. https://doi.org/10.1016/S0140-6736%2818%2930295-2

Boland, A., Cherry, MG., & Dickson, R. (2017). *Doing a systematic review: A student's guide (2nd Ed.)*. LA: Sage.

Broadbent, H. J., van den Eynde, F., Guillaume, S., Hanif, E. L., Stahl, D., David, A. S., Campbell, I. C., & Schmidt, U. (2011). Blinding success of rTMS applied to the dorsolateral prefrontal cortex in randomised sham-controlled trials: A systematic review. *The World Journal of Biological Psychiatry : The Official Journal of the World Federation of Societies of Biological Psychiatry*, *12*(4), 240–248. https://doi.org/10.3109/15622975.2010.541281

Busner, J., & Targum, S. D. (2007). The Clinical Global Impressions Scale. *Psychiatry (Edgmont)*, *4*(7), 28–37.

Button, K. S., Kounali, D., Thomas, L., Wiles, N. J., Peters, T. J., Welton, N. J., Ades, A. E., & Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory—II according to the patient's perspective. *Psychological Medicine*, *45*(15), 3269–3279. https://doi.org/10.1017/S0033291715001270

Chistyakov A.V., Kreinin B., Marmor S., Kaplan B., Khatib A., Darawsheh N., Koren D., Zaaroor M., & Klein E. (2015). Preliminary assessment of the therapeutic efficacy of continuous theta-burst magnetic stimulation (cTBS) in major depression: A double-blind sham-controlled study. *Journal of Affective Disorders*, *170*((Chistyakov, Kaplan, Zaaroor) Department of Neurosurgery, Rambam Medical Center, B. Rappaport Faculty of Medicine, Israel), 225–259. https://doi.org/10.1016/j.jad.2014.08.035

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Ed.),*

Hillsdale, NJ: Erlbaum.

Couturier, J. L. (2005). Efficacy of rapid-rate repetitive transcranial magnetic stimulation in the treatment of depression: A systematic review and meta-analysis. *Journal of Psychiatry & Neuroscience : JPN*, *30*(2), 83–90.

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., & van Straten, A. (2011). Interpersonal psychotherapy for depression: A meta-analysis. *The American Journal of Psychiatry*, *168*(6), 581–592. https://doi.org/10.1176/appi.ajp.2010.10101411

Deeks, J.J., Higgings, J.P.T., Altman, D.G. (2008). Analysing data and undertaking meta-analyses. In J.P.T Higgins and S.Green (Eds.) *Cochrane handbook for systematic reviews* (pp. 243-296). Chichester: John Wiley & Sons.

Demitrack, M.A. & Lisanby, S.H. (2018). Methodological issues in clinical trial design for TMS In C.M. Epstein, E.M. Wassermann, & U. Ziemann (Eds.) *Oxford Handbook of Transcranial Stimulation* (pp.621-632). Oxford: OUP.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials,* 7, 177–188.

Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. M. (2010). Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *Journal of Consulting and Clinical Psychology*, *78*(5), 668–680. https://doi.org/10.1037/a0020570

Dubin, M. J., Mao, X., Banerjee, S., Goodman, Z., Lapidus, K. A. B., Kang, G., Liston, C., & Shungu, D. C. (2016). Elevated prefrontal cortex GABA in patients with major depressive disorder after TMS treatment measured with proton magnetic resonance spectroscopy. *Journal of Psychiatry &*

Neuroscience, *41*(Clinical Psychopharmacology [3340]), E37–E45. https://doi.org/10.1503/jpn.150223

Duecker, F., & Sack, A. T. (2015). Rethinking the role of sham TMS. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00210

Duprat, R., Wu, G.-R., De Raedt, R., & Baeken, C. (2017). Accelerated iTBS treatment in depressed patients differentially modulates reward system activity based on anhedonia. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, 1–12. https://doi.org/10.1080/15622975.2017.1355472

Faraone, S. V. (2008). Interpreting Estimates of Treatment Effects. *Pharmacy and Therapeutics*, *33*(12), 700–711.

FDA. (2008). *Neurostar TMS Therapy System.* Silver Spring, MD: US Food and Drug Administration.

First, M.B., & Spitzer, R.L., Gibbon, M., & Williams, J.B. (2002). *Structured Interview for DSM-IV-TR Axis I Disorders, Research version, Patient Edition (SCID-I/P).* New York: Biometrics Research, New York State Psychiatric Institute.

Fitzgerald, P. B., & Daskalakis, Z. J. (2011). The effects of repetitive transcranial magnetic stimulation in the treatment of depression. *Expert Review of Medical Devices*, *8*(1), 85–95. https://doi.org/10.1586/erd.10.57

Fitzgerald, P. B., Hoy, K., Daskalakis, Z. J., & 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*(3), 229–234. https://doi.org/10.1002/da.20454

Fitzgerald, P. B., Hoy, K. E., Elliot, D., McQueen, S., Wambeek, L. E., & Daskalakis, Z. J. (2016). A negative double-blind controlled trial of sequential bilateral rTMS in the treatment of bipolar depression. *Journal of Affective Disorders*, *198*(Angst, J. (2004). Bipolar disorder-a seriously underestimated health burden. Eur. Arch. Psychiatry Clin. Neurosci., 254, 59-60. http://dx.doi.org/10.1007/s00406-004-0502-5), 158–162. https://doi.org/10.1016/j.jad.2016.03.052

Fitzgerald, P. B., Hoy, K. E., Herring, S. E., McQueen, S., Peachey, A. V. J., Segrave, R. A., Maller, J., Hall, P., & Daskalakis, Z. J. (2012). A double blind randomized trial of unilateral left and bilateral prefrontal cortex transcranial magnetic stimulation in treatment resistant major depression. *Journal of Affective Disorders*, *139*(2), 193–198. https://doi.org/10.1016/j.jad.2012.02.017

Fung, G., Deng, Y., Zhao, Q., Li, Z., Qu, M., Li, K., Zeng, Y.-W., Jin, Z., Ma, Y.-T., Yu, X., Wang, Z.-R., Shum, D. H. K., & Chan, R. C. K. (2015). Distinguishing bipolar and major depressive disorders by brain structural morphometry: A pilot study. *BMC Psychiatry*, *15*, 298. https://doi.org/10.1186/s12888-015-0685-5

Furukawa, T. A., McGuire, H., & Barbui, C. (2002). Meta-analysis of effects and side effects of low dosage tricyclic antidepressants in depression: Systematic review. *BMJ : British Medical Journal*, *325*(7371), 991.

Garcia-Toro, M., Mayol, A., Arnillas, H., Capllonch, I., Ibarra, O., Crespi, M., Mico, J., Lafau, O., & Lafuente, L. (2001). Modest adjunctive benefit with transcranial magnetic stimulation in medication-resistant depression. *Journal of Affective Disorders*, *64*(2–3), 271–275.

George, M. S., Lisanby, S. H., Avery, D., McDonald, W. M., Durkalski, V., Pavlicova, M., Anderson, B., Nahas, Z., Bulow, P., Zarkowski, P., Holtzheimer, P. E., Schwartz, T., & Sackeim, H. A. (2010). Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: A sham-controlled randomized trial. ***Archives of General Psychiatry*****, *67*(5), 507–516. https://doi.org/10.1001/archgenpsychiatry.2010.46**

George, M. S., Raman, R., Benedek, D. M., Pelic, C. G., Grammer, G. G., Stokes, K. T., Schmidt, M., Spiegel, C., DeAlmeida, N., Beaver, K. L., Borckardt, J. J., Sun, X., Jain, S., & Stein, M. B. (2014). A two-site pilot randomized 3 day trial of high dose left prefrontal repetitive transcranial magnetic stimulation (rTMS) for suicidal inpatients. *Brain Stimulation*, *7*(Specialized Interventions [3350]), 421–431. https://doi.org/10.1016/j.brs.2014.03.006

George, M. S., Taylor, J. J., & Short, E. B. (2013). The expanding evidence base for rTMS treatment of depression. *Current Opinion in Psychiatry*, *26*(1), 13–18. https://doi.org/10.1097/YCO.0b013e32835ab46d

Hamilton, M.(1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, & Psychiatry, 23 (1),* 56-62. https://doi.org/10.1136/jnnp.23.1.56

Herrmann, L. L., & Ebmeier, K. P. (2006). Factors modifying the efficacy of transcranial magnetic stimulation in the treatment of depression: A review. *The Journal of Clinical Psychiatry*, *67*(12), 1870–1876.

Herwig, U., Fallgatter, A. J., Höppner, J., Eschweiler, G. W., Kron, M., Hajak, G., Padberg, F., Naderi-Heiden, A., Abler, B., Eichhammer, P., Grossheinrich, N., Haya, B., Kammer, T., Langguth, B., Laske, C., Plewnia, C., Richter, M. M.,

Schulz, M., Unterecker, S., … Schönfeldt-Lecuona, C. (2007). Antidepressant effects of augmentative transcranial magnetic stimulation: Randomised multicentre trial. *The British Journal of Psychiatry*, *191*(5), 441–448. https://doi.org/10.1192/bjp.bp.106.034371

Higgins, J.P.T. & Green, S .(2008). *Cochrane handbook for systematic reviews.* Chichester: John Wiley & Sons.

Holtzheimer, P. E. 3rd, Russo, J., Claypoole, K. H., Roy-Byrne, P., & Avery, D. H. (2004). Shorter duration of depressive episode may predict response to repetitive transcranial magnetic stimulation. *Depression and Anxiety*, *19*(1), 24–30.

Hooijer, C., Zitman, F. G., Griez, E., van Tilburg, W., Willemse, A., & Dinkgreve, M. A. H. M. (1991). The Hamilton Depression Rating Scale (HDRS): Changes in scores as a function of training and version used. *Journal of Affective Disorders*, *22*(1), 21–29. https://doi.org/10.1016/0165-0327(91)90079-8

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Huang, Y.-Z., Rothwell, J. C., Chen, R.-S., Lu, C.-S., & Chuang, W.-L. (2011). The theoretical model of theta burst form of repetitive transcranial magnetic stimulation. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *122*(5), 1011–1018. https://doi.org/10.1016/j.clinph.2010.08.016

Huerta, P. T., & Volpe, B. T. (2009). Transcranial magnetic stimulation, synaptic plasticity and network oscillations. *Journal of NeuroEngineering and Rehabilitation*, *6*, 7. https://doi.org/10.1186/1743-0003-6-7

Hunter A.M., Minzenberg M.J., Cook I.A., Krantz D.E., Levitt J.G., Rotstein N.M., Chawla S.A., & Leuchter A.F. (2019). Concomitant medication use and clinical outcome of repetitive Transcranial Magnetic Stimulation (rTMS) treatment of Major Depressive Disorder. *Brain and Behavior*, *9*(5), e01275. https://doi.org/10.1002/brb3.1275

Janicak, P. G., & Dokucu, M. E. (2015). Transcranial magnetic stimulation for the treatment of major depression. *Neuropsychiatric Disease and Treatment*, *11*, 1549–1560. https://doi.org/10.2147/NDT.S67477

Januel D., Dumortier G., Verdon C.-M., Stamatiadis L., Saba G., Cabaret W., Benadhira R., Rocamora J.-F., Braha S., Kalalou K., Vicaut Pr.E., & Fermanian J. (2006). A double-blind sham controlled study of right prefrontal repetitive transcranial magnetic stimulation (rTMS): Therapeutic and cognitive effect in medication free unipolar depression during 4 weeks. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *30*(1), 126–130. https://doi.org/10.1016/j.pnpbp.2005.08.016

Jiang, Y., Guo, Z., Xing, G., He, L., Peng, H., Du, F., McClure, M. A., & Mu, Q. (2019). Effects of High-Frequency Transcranial Magnetic Stimulation for Cognitive Deficit in Schizophrenia: A Meta-Analysis. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00135

Jin, Y., & Phillips, B. (2014). A pilot study of the use of EEG-based synchronized Transcranial Magnetic Stimulation (sTMS) for treatment of Major Depression. *BMC Psychiatry*, *14*, 13. https://doi.org/10.1186/1471-244X-14-13

Kang, J. I., Lee, H., Jhung, K., Kim, K. R., An, S. K., Yoon, K.-J., Kim, S. I., Namkoong, K., & Lee, E. (2016). Frontostriatal Connectivity Changes in Major Depressive Disorder After Repetitive Transcranial Magnetic

Stimulation: A Randomized Sham-Controlled Study. *The Journal of Clinical Psychiatry*, *77*(9), e1137–e1143. https://doi.org/10.4088/JCP.15m10110

Kedzior, K. K., Reitz, S. K., Azorina, V., & Loo, C. (2015). Durability of the antidepressant effect of the high-frequency repetitive transcranial magnetic stimulation (rtms) in the absence of maintenance treatment in major depression: A systematic review and meta-analysis of 16 double-blind, randomized, sham-controlled trials. *Depression and Anxiety*, *32*(3), 193–203.

Koenigs, M., & Grafman, J. (2009). The functional neuroanatomy of depression: Distinct roles for ventromedial and dorsolateral prefrontal cortex. *Behavioural Brain Research*, *201*(2), 239–243. https://doi.org/10.1016/j.bbr.2009.03.004

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuyken, W., Hayes, R., Barrett, B., Byng, R., Dalgleish, T., Kessler, D., Lewis, G., Watkins, E., Brejcha, C., Cardy, J., Causley, A., Cowderoy, S., Evans, A., Gradinger, F., Kaur, S., Lanham, P., Morant, N., Richards, J., Shah, P., … Byford, S. (2015). Effectiveness and cost-effectiveness of mindfulness-based cognitive therapy compared with maintenance antidepressant treatment in the prevention of depressive relapse or recurrence (PREVENT): A randomised controlled trial. *The Lancet*, *386*(9988), 63–73. https://doi.org/10.1016/S0140-6736(14)62222-4

Kuyken, W., Warren, F. C., Taylor, R. S., Whalley, B., Crane, C., Bondolfi, G., Hayes, R., Huijbers, M., Ma, H., Schweizer, S., Segal, Z., Speckens, A., Teasdale, J. D., Van Heeringen, K., Williams, M., Byford, S., Byng, R., & Dalgleish, T. (2019). Efficacy of Mindfulness-Based Cognitive Therapy in

Prevention of Depressive Relapse. *JAMA Psychiatry*, *73*(6), 565–574. https://doi.org/10.1001/jamapsychiatry.2016.0076

Kyle, P. R., Lemming, O. M., Timmerby, N., Søndergaard, S., Andreasson, K., & Bech, P. (2016). The Validity of the Different Versions of the Hamilton Depression Scale in Separating Remission Rates of Placebo and Antidepressants in Clinical Trials of Major Depression. *Journal of Clinical Psychopharmacology*, *36*(5), 453–456. https://doi.org/10.1097/JCP.0000000000000557

Lage, C., Wiles, K., Shergill, S. S., & Tracy, D. K. (2016). A systematic review of the effects of low-frequency repetitive transcranial magnetic stimulation on cognition. *Journal of Neural Transmission (Vienna, Austria: 1996)*, *123*(12), 1479–1490. https://doi.org/10.1007/s00702-016-1592-8

Lam, R. W., Chan, P., Wilkins-Ho, M., & Yatham, L. N. (2008). Repetitive transcranial magnetic stimulation for treatment-resistant depression: A systematic review and metaanalysis. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, *53*(9), 621–631. https://doi.org/10.1177/070674370805300909

Lefaucheur, J.-P., André-Obadia, N., Antal, A., Ayache, S. S., Baeken, C., Benninger, D. H., Cantello, R. M., Cincotta, M., de Carvalho, M., De Ridder, D., Devanne, H., Di Lazzaro, V., Filipović, S. R., Hummel, F. C., Jääskeläinen, S. K., Kimiskidis, V. K., Koch, G., Langguth, B., Nyffeler, T., … Garcia-Larrea, L. (2014). Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rTMS). *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *125*(11), 2150–2206. https://doi.org/10.1016/j.clinph.2014.05.021

Lepping, P., Schönfeldt-Lecuona, C., Sambhi, R. S., Lanka, S. V. N., Lane, S., Whittington, R., Leucht, S., & Poole, R. (2014). A systematic review of the clinical relevance of repetitive transcranial magnetic stimulation. *Acta Psychiatrica Scandinavica*, *130*(5), 326–341. https://doi.org/10.1111/acps.12276

Leucht, S., Fennema, H., Engel, R., Kaspers–Janssen, M., Lepping, P., & Szegedi, A. (2013). What does the HAMD mean? *Journal of Affective Disorders*, *148*(2), 243–248. https://doi.org/10.1016/j.jad.2012.12.001

Li C.-T., Chen M.-H., Juan C.-H., Huang H.-H., Chen L.-F., Hsieh J.-C., Tu P.-C., Bai Y.-M., Tsai S.-J., Lee Y.-C., & Su T.-P. (2014). Efficacy of prefrontal theta-burst stimulation in refractory depression: A randomized sham-controlled study. *Brain*, *137*(7), 2088–2098. https://doi.org/10.1093/brain/awu109

Li C.-T., Cheng C.-M., Chen M.-H., Juan C.-H., Tu P.-C., Bai Y.-M., Jeng J.-S., Lin W.-C., Tsai S.-J., & Su T.-P. (2019). Antidepressant Efficacy of Prolonged Intermittent Theta Burst Stimulation Monotherapy for Recurrent Depression and Comparison of Methods for Coil Positioning: A Randomized, Double-Blind, Sham-Controlled Study. *Biological Psychiatry*, *(Li, Cheng, Chen, Tu, Bai, Jeng, Lin, Tsai, Su) Department of Psychiatry, Taipei Veterans General Hospital, Taipei, Taiwan (Republic of China)*. https://doi.org/10.1016/j.biopsych.2019.07.031

Li, C.-T., Hsieh, J.-C., Huang, H.-H., Chen, M.-H., Juan, C.-H., Tu, P.-C., Lee, Y.-C., Wang, S.-J., Cheng, C.-M., & Su, T.-P. (2016). Cognition-Modulated Frontal Activity in Prediction and Augmentation of Antidepressant Efficacy: A Randomized Controlled Pilot Study. *Cerebral*

Cortex (New York, N.Y. : 1991), ***26***(1), 202–210. **https://doi.org/10.1093/cercor/bhu191**

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*. https://doi.org/10.1136/bmj.b2700

Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation – A discussion and demonstration of basic features. *PLoS ONE*, *14*(7). https://doi.org/10.1371/journal.pone.0219854

Lisanby, S. H., Husain, M. M., Rosenquist, P. B., Maixner, D., Gutierrez, R., Krystal, A., Gilmer, W., Marangell, L. B., Aaronson, S., Daskalakis, Z. J., Canterbury, R., Richelson, E., Sackeim, H. A., & George, M. S. (2009). Daily left prefrontal repetitive transcranial magnetic stimulation in the acute treatment of major depression: Clinical predictors of outcome in a multisite, randomized controlled clinical trial. *Neuropsychopharmacology*, *34*(Clinical Psychopharmacology [3340]), 522–534. https://doi.org/10.1038/npp.2008.118

Liu, B., Zhang, Y., Zhang, L., & Li, L. (2014). Repetitive transcranial magnetic stimulation as an augmentative strategy for treatment-resistant depression, a meta-analysis of randomized, double-blind and sham-controlled study. *BMC Psychiatry*, *14*, 342. https://doi.org/10.1186/s12888-014-0342-4

Loo, C. K., Taylor, J. L., Gandevia, S. C., McDarmont, B. N., Mitchell, P. B., & Sachdev, P. S. (2000). Transcranial magnetic stimulation (TMS) in controlled treatment studies: Are some 'sham' forms active? *Biological Psychiatry*, *47*(4), 325–331. https://doi.org/10.1016/s0006-3223(99)00285-1

Loo, C., Sachdev, P., Elsayed, H., McDarmont, B., Mitchell, P., Wilkinson, M., Parker, G., & Gandevia, S. (2001). Effects of a 2- to 4-week course of repetitive transcranial magnetic stimulation (rTMS) on neuropsychologic functioning, electroencephalogram, and auditory threshold in depressed patients. *Biological Psychiatry*, *49*(7), 615–623.

Machii, K., Cohen, D., Ramos-Estebanez, C., & Pascual-Leone, A. (2006). Safety of rTMS to non-motor cortical areas in healthy participants and patients. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *117*(2), 455–471.

Martin, D. M., McClintock, S. M., Forster, J. J., Lo, T. Y., & Loo, C. K. (2017). Cognitive enhancing effects of rTMS administered to the prefrontal cortex in patients with depression: A systematic review and meta-analysis of individual task effects. *Depression and Anxiety*, *34*(Medical Treatment of Physical Illness [3363]), 1029–1039. https://doi.org/10.1002/da.22658

Martin, J. L. R., Barbanoj, M. J., Schlaepfer, T. E., Thompson, E., Pérez, V., & Kulisevsky, J. (2003). Repetitive transcranial magnetic stimulation for the treatment of depression. Systematic review and meta-analysis. *The British Journal of Psychiatry: The Journal of Mental Science*, *182*, 480–491.

McGlothlin, A. E., & Lewis, R. J. (2014). Minimal clinically important difference: Defining what really matters to patients. *JAMA*, *312*(13), 1342–1343. https://doi.org/10.1001/jama.2014.13128

Meline, T.(2007). Selecting studies for systematic review: Inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders, 33*, 21-27.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1. https://doi.org/10.1186/2046-4053-4-1

Moncrieff, J., & Kirsch, I. (2015). Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemporary Clinical Trials*, *43*, 60–62. https://doi.org/10.1016/j.cct.2015.05.005

Moreno, C., Hasin, D. S., Arango, C., Oquendo, M. A., Vieta, E., Liu, S., Grant, B. F., & Blanco, C. (2012). Depression in bipolar disorder versus major depressive disorder: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Bipolar Disorders*, *14*(3), 271–282. https://doi.org/10.1111/j.1399-5618.2012.01009.x

Mutz, J., Edgcumbe, D. R., Brunoni, A. R., & Fu, C. H. Y. (2018). Efficacy and acceptability of non-invasive brain stimulation for the treatment of adult unipolar and bipolar depression: A systematic review and meta-analysis of randomised sham-controlled trials. *Neuroscience and Biobehavioral Reviews*, *92*(Atkinson, L., Sankar, A., Adams, T.M., Fu, C.H. (2014). Recent advances in neuroimaging of mood disorders: structural and functional neural correlates of depression, changes with therapy, and potential for clinical biomarkers. Curr. Treat. Options Psychiatry, 1, 278-293. http://dx.doi.org/10.1007/s40501-014-0022-5), 291–303. https://doi.org/10.1016/j.neubiorev.2018.05.015

Mutz, J., Vipulananthan, V., Carter, B., Hurlemann, R., Fu, C. H. Y., & Young, A. H. (2019). Comparative efficacy and acceptability of non-surgical brain stimulation for the acute treatment of major depressive episodes in adults:

Systematic review and network meta-analysis. *BMJ*, *364*. https://doi.org/10.1136/bmj.l1079

National Institute for Health and Clinical Excellence. (2009). *Depression in adults: The treatment and management of depression in adults*. *CG90*. London: National Institute for Health and Care Excellence.

NICE. (2015). *Repetitive transcranial magnetic stimulation for depression. IPG542.* National Institute for Health and Care Excellence.

Noda, Y., Silverstein, W. K., Barr, M. S., Vila-Rodriguez, F., Downar, J., Rajji, T. K., Fitzgerald, P. B., Mulsant, B. H., Vigod, S. N., Daskalakis, Z. J., & Blumberger, D. M. (2015). Neurobiological mechanisms of repetitive transcranial magnetic stimulation of the dorsolateral prefrontal cortex in depression: A systematic review. *Psychological Medicine*, *45*(Medical Treatment of Physical Illness [3363]), 3411–3432. https://doi.org/10.1017/S0033291715001609

O'Reardon, J. P., Solvason, H. B., Janicak, P. G., Sampson, S., Isenberg, K. E., Nahas, Z., McDonald, W. M., Avery, D., Fitzgerald, P. B., Loo, C., Demitrack, M. A., George, M. S., & Sackeim, H. A. (2007). Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: A multisite randomized controlled trial. *Biological Psychiatry*, *62*(11), 1208–1216. https://doi.org/10.1016/j.biopsych.2007.01.018

Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Ridding, M. C., & Rothwell, J. C. (2007). Is there a future for therapeutic use of transcranial magnetic stimulation? *Nature Reviews Neuroscience*, *8*(7), 559–567. https://doi.org/10.1038/nrn2169

Riedel, M., Möller, H.-J., Obermeier, M., Schennach-Wolff, R., Bauer, M., Adli, M., Kronmüller, K., Nickel, T., Brieger, P., Laux, G., Bender, W., Heuser, I., Zeiler, J., Gaebel, W., & Seemüller, F. (2010). Response and remission criteria in major depression – A validation of current practice. *Journal of Psychiatric Research*, *44*(15), 1063–1068. https://doi.org/10.1016/j.jpsychires.2010.03.006

Rush, A.J., Thase, M.E., & Dubé, S. (2003). Research issues in the study of difficult to treat depression. *Biological Psychiatry, 53,* 743-753.

Sackheim, H.A. (2001). The definition and meaning of treatment- resistant depression. *Journal of Clinical Psychiatry, 62(16),* 10-17

Schermuly-Haupt, M.-L., Linden, M., & Rush, A. (2018). Unwanted Events and Side Effects in Cognitive Behavior Therapy. *Cognitive Therapy and Research*, *42*. https://doi.org/10.1007/s10608-018-9904-y

Schutter, D. J. L. G. (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*(1), 65–75. https://doi.org/10.1017/S0033291708003462

Sehatzadeh, S., Daskalakis, Z. J., Yap, B., Tu, H.-A., Palimaka, S., Bowen, J. M., & O'Reilly, D. J. (2019). Unilateral and bilateral repetitive transcranial magnetic stimulation for treatment-resistant depression: A meta-analysis of randomized controlled trials over 2 decades. *Journal of Psychiatry & Neuroscience: JPN*, *44*(3), 151–163. https://doi.org/10.1503/jpn.180056

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J.,… Dunbar, G.C.(1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*. *59(20),*22–33.

Slotema, C. W., Blom, J. D., Hoek, H. W., & Sommer, I. E. C. (2010). Should we expand the toolbox of psychiatric treatment methods to include Repetitive Transcranial Magnetic Stimulation (rTMS)? A meta-analysis of the efficacy of rTMS in psychiatric disorders. *The Journal of Clinical Psychiatry*, *71*(7), 873–884. https://doi.org/10.4088/JCP.08m04872gre

Stern, W. M., Tormos, J. M., Press, D. Z., Pearlman, C., & Pascual-Leone, A. (2007). Antidepressant effects of high and low frequency repetitive transcranial magnetic stimulation to the dorsolateral prefrontal cortex: A double-blind, randomized, placebo-controlled trial. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *19*(2), 179–186.

Taylor, S. F., Ho, S. S., Abagis, T., Angstadt, M., Maixner, D. F., Welsh, R. C., & Hernandez-Garcia, L. (2018). Changes in brain connectivity during a sham-controlled, transcranial magnetic stimulation trial for depression. *Journal of Affective Disorders*, *232*(Argyelan, M., Lencz, T., Kaliora, S., Sarpal, D.K., Weissman, N., Kingsley, P.B., Malhotra, A.K., Petrides, G. (2016). Subgenual cingulate cortical activity predicts the efficacy of electroconvulsive therapy. Transl. Psychiatry, 6, e789http://dx.doi.org/10.1038/tp.2016.54), 143–151. https://doi.org/10.1016/j.jad.2018.02.019

Teng, S., Guo, Z., Peng, H., Xing, G., Chen, H., He, B., McClure, M. A., & Mu, Q. (2017). High-frequency repetitive transcranial magnetic stimulation over the

left DLPFC for major depression: Session-dependent efficacy: A meta-analysis. *European Psychiatry*, *41*(Avery, D.H., Holtzheimer, P.E., 3rd., Fawaz, W., Russo, J., Neumaier, J., Dunner, D.L., et al. (2006). A controlled study of repetitive transcranial magnetic stimulation in medication-resistant major depression. Biol Psychiatry, 59, 2, 187-194. http://dx.doi.org/10.1016/j.biopsych.2005.07.003), 75–84. https://doi.org/10.1016/j.eurpsy.2016.11.002

Thase, M.E., & Rush, A.J. (1997). When at first you don't succeed: sequential strategies for antidepressant nonresponders. *Journal of Clinical Psychiatry, 58(13),*23-29.

Vanderhasselt, M.-A., De Raedt, R., Baeken, C., Leyman, L., & D'Haenen, H. (2009a). A single session of rTMS over the left dorsolateral prefrontal cortex influences attentional control in depressed patients. *The World Journal of Biological Psychiatry : The Official Journal of the World Federation of Societies of Biological Psychiatry*, *10*(1), 34–42. https://doi.org/10.1080/15622970701816514

Wang, Y.-M., Li, N., Yang, L.-L., Song, M., Shi, L., Chen, W.-H., Li, S.-X., Wang, X.-Y., & Lu, L. (2017). Randomized controlled trial of repetitive transcranial magnetic stimulation combined with paroxetine for the treatment of patients with first-episode major depressive disorder. *Psychiatry Research*, *254*(qc4, 7911385), 18–23. https://doi.org/10.1016/j.psychres.2017.04.005

Williams, J. B. (2001). Standardizing the Hamilton Depression Rating Scale: Past, present, and future. *European Archives of Psychiatry and Clinical Neuroscience*, *251 Suppl 2*, II6-12. https://doi.org/10.1007/BF03035120

Yang, X.-R., Kirton, A., Wilkes, T. C., Pradhan, S., Liu, I., Jaworska, N., Damji, O., Keess, J., Langevin, L. M., Rajapakse, T., Lebel, R. M., Sembo, M., Fife, M., & MacMaster, F. P. (2014). Glutamate alterations associated with transcranial magnetic stimulation in youth depression: A case series. *The Journal of ECT*, *30*(3), 242–247. https://doi.org/10.1097/YCT.0000000000000094

Zimmerman, M., Martinez, J. H., Young, D., Chelminski, I., & Dalrymple, K. (2013). Severity classification on the Hamilton Depression Rating Scale. *Journal of Affective Disorders*, *150*(2), 384–388. https://doi.org/10.1016/j.jad.2013.04.028

Table 1. Exclusion Criteria for study selection

Number of studies excluded	Reason/s for exclusion
119	**Participant/s** having a reported comorbid mental health disorder, or having another diagnosis other than a depressive disorder (e.g. bipolar, PTSD, OCD), as specified in the inclusion criteria.
128	**Intervention** is not TMS (e.g. ECT; tdcs)
17	**Site of stimulation** is not the DLPFC, but another neural area, such as the motor area, or dorsomedial prefrontal cortex, for example.
138	**Comparator** condition does not include a sham control.
7	**Outcome measures**. Studies that did not include outcome measures relating to depressive symptoms (e.g. depressive symptoms questionnaires) for active and sham stimulation.
7	**Study design** is not a Randomised Controlled Trial(RCT).
5	**Blinding.** Studies are not double-blinded RCTs
39	**Concurrent antidepressant usage**. Studies that included a concurrent antidepressant/ patients were not medication free.
Total excluded: **460**	

Table 2. Characteristics of included studies (in bold) and those excluded due to duplicate data.

Study	Double-blind RCT	Cross-over	Single-Site	Multi-Site & Multi-Phase
Baeken et al. (2013)[1] Baeken et al. (2015)[1] **Vanderhasselt et al. (2009a)**[2] Vanderhasselt et al. (2009b)[2] **Duprat et al. (2016)**[3] Baeken et al. (2017) [3] Baeken et al. (2019) [3] Caeyenberghs et al. (2018)[3] Desmyter et al. (2016)[3] (n = 3 studies included)	**X**	**X**	**X**	
O'Reardon et al. (2007)[4] Lisanby et al. (2009)[4] Rosenquist et al. (2013)[4] **George et al. (2010)** (n = 2 studies included)	**X**			**X**
Holzheimer et al. (2004) Li et al. (2016) Stern et al. (2007) (n = 3 studies included)	**X**		**X**	

1 = These studies are part of the same larger clinical trial: FWO08/PDO/168; 2= These studies were part of the same larger clinical trial: BOF- 01J08107; 3 = These studies are part of the same larger clinical trial (http://clinicaltrials.gov/show/NCT01832805); 4 = These studies are part of the same larger clinical trial (NCT 0010461).

Table 3. Patient demographics; baseline depression severity & assessment, medication usage & study sponsor.

Study	Patients, Diagnosis: who screened + Comorbidity	Treatment Resistance Questionnaire (& who assessed)	Duration of Current Episode	Baseline Depression Severity (M; SD) (& who assessed)	Medication & Benzodiazepine Usage (patients = n; [A/S])	Study Sponsor
Baeken et al. (2013)	Outpatients + Inpatients; UDP. (MINI, ICD-10): ANR No comorbidity reported	At least Stage III Treatment Resistance (Rush et al. 2003) ANR	*M* = 7.83yrs; *SD* = 7.21	HDRS- 17 item, *M* = 26.4 (8.24). *Severe* Psychiatrist unrelated to the study	2week medication wash-out period prior to study enrolment. Habitual benzodiazepine allowed (n= 14; [6/8])	Scientific Fund W Gepts UZ Brussel (FWO08/ PDO/168)
Duprat et al. (2016)	Outpatients + Inpatients; MDD (MINI): ANR No comorbidity reported.	At least Stage I Treatment Resistance (Rush et al. 2003) - failed one antidepressant trial. ANR	*M =3.87yrs; SD = 6.08*	HDRS-17-Item; *M* = 21.34; (5.26) *Moderate* Psychiatrist blinded to the treatment of the patients.	2week medication wash-out period prior to study enrolment. Benzodiazepines provided if necessary (up to 40mg) (n = 15, [no data]).	Ghent University Multidisciplinary Research Partnership; Applied Biomedical Grant
George et al. (2010)	Outpatients MDD (DSM-IV): Psychiatrist No comorbidity reported	Moderate Treatment Resistance (ATHF)- 1.5 failed treatments Screened on the phone/ on site-psychiatrist	*M* = 74.1 wks.; *SD* = 64.9	24-item HDRS (> 20), *M* = 26.4 (4.9) *Severe* Trained masked clinical raters	Patients medication free 2 weeks pre baseline assessment and 5 weeks for fluoxetine. Sedatives and hypnotics or anxiolytics (14 daily doses) (no data, [no data])	National Institute of Health - Optimization of TMS treatment of depression
Holzheimer et al. (2004)	Outpatients MDD (DSM-IV): ANR No comorbidity reported.	Two previous failed antidepressant treatments (ATHF, Sackeim, 2001): ANR	Range: ≤4yrs to >10yrs	17-item HDRS (>18) *M* = 20.6 (4.1) *Moderate* Blind rater	2week medication wash-out period prior to baseline assessments.	University of Washington
Li et al. (2016)	Hospital status unknown Unipolar MDD (DSM-IV; MINI): ANR No comorbidity reported	Poor response to 2 antidepressant treatments: ANR	~*M* = 6.13 months ~ *SD* = 6.63	17-item HDRS (>18) *M* = 22.4 (5.33) *Moderate* Psychiatrist Blinded	1-week medication wash-out period. No patients on fluoxetine.	Taipei Veterans General Hospital; National Science Council; Grant; Ministry of Science & Technology; National Yang-Ming University; Yen Tjing Ling Medical Foundation
O'Reardon et al. (2007)	Outpatients MDD (DSM-IV): ANR No comorbidities reported: but same data as Lisanby et al. (2009) *	At least one failed antidepressant (ATHF) -no antidepressant medication for first phase of study ANR	~*M* = 13.4 months ~*SD* = 9.7	17-item HDRS (>20; CGSI = 4) *M* = 22.75 (3.4) *Moderate* Trained blinded raters not permitted to access treatment rooms.	Free of antidepressants Hypnotics, anxiolytics, lorazepam (up to 14 daily doses allowed)	NIMH; Neuronetics (TMS device producer)
Stern et al. (2007)	Outpatients; MDD (SCID; DSM-IV): Psychiatrist No comorbidity reported	Failed one antidepressant treatment: Psychiatrist	No date for current depressive episode	21-item HAMD (>20) *M* = 27.68 (3.45) Blinded psychiatrist who consulted with treating psychiatrist.	2week medication washout, from psychotropic medication prior to baseline assessment Lorazepam (up to 2mg daily doses allowed) during first half of washout period.	Spanish Ministerio de Education y Cienca, Milton Fund, Stanley Vada NAMI, NIMH, National alliance for Research in schizophrenia and depression
Vanderhasselt et al. (2009a)	Hospital status unknown Depressive Disorder (DSM-IV; MINI): ANR No comorbid diagnoses reported	At least one failed antidepressant treatment ANR	Range: 2 months- 11 years	17-item HAMD (>16) *Mild* ANR	2week washout period, of antidepressant medication prior to study start; 3 weeks for fluoxetine. No "rescue medications" permitted.	Scientific Fund W Gepts UZ Brussel

Diagnosis: UDP = unipolar depressed patients; MDD= major depressive disorder; Questionnaire: ATHF = antidepressant treatment history form;

Clinical Interview: Mini International Neuropsychiatric Interview (MINI); Diagnostic and Statistical Manual *No comorbidity reported, however, in a paper using the same data, comorbid anxiety disorders are reported.

ANR: administrator not reported; Sponsor: NIMH: National Institute of Mental Health

.

Table 4. TMS treatment parameters, and type coils used.

Study	Intervention, Site of Stimulation + Parameters	Determination of stimulation site	Motor threshold (MT)	Type of coil	
				Active	Sham
Baeken et al. (2013)	aHFrTMS_20Hz_lDLPFC 1560 pulses; 40 trains of 1.9s duration, intertrain of 12s. 20 sessions (4 days x 5 sessions). Total: 31.200 stimuli. 20 sessions	3D-MRI neuro-navigation	EMG; rAPB MT: 110%	High speed magnetic stimulator connected to figure of eight shaped cooled coil	Same coil as active, however, coil placed at a 90° angle; coil resting on scalp.
Duprat et al. (2016)	aiTBS_lDLPFC 1620 pulses; 54 triplet bursts, train duration of 2s; intertrain interval of 8s. 20 sessions (4 days x 5 sessions). Total: 32.400 stimuli. 20 sessions	Structural MRI; Brain Sight neuro-navigation	EMG; rAPB MT: 110%	Magstim Rapid 2 Plus1 stimulator, figure of eight shaped coil	Figure of 8 shaped coil, identical to active, bar somatosensory sensations.
George et al. (2010)	HFrTMS_10Hz_lDLPFC 3000 pulses;10 pulses per second for 4s, intertrain of 26s. Sessions were 37.5 minutes with 75 trains;3 weeks fixed treatment phase (5 days per wk.). Total: 45,000. No. of sessions: 15 sessions for phase 1.	MRI with vitamin E TMS coil 5 cm anterior to MT location	EMG: weekly rMT; l & rAPB. MT: 120 & 110%	Figure-8 solid-core coil.	Similar to active coil, but metal insert to block magnetic field, scalp electrodes to match somatosensory sensation from active coil.
Holzheimer et al. (2004)	HFrTMS_10Hz_lDLPFC 1600 pulses per day; 32 trains of 5s trains, 30-60s intertrain interval. Ten sessions over 2 weeks Total: 16, 000 10 sessions + follow-up	5cm anterior to motor cortex	EMG: rFDI muscle MT: 110%	Magnetic stimulator; Dantec, Magpro; Medtronic; Shoreview; MN- figure of 8 coil- active: coil flat on the scalp, short axis of coil orientated in parasagittal plane.	Same location, same stimulation – but lateral edge of coil rotated at 45°
Li et al. (2016)	HFrTMS _10 Hz_lDLPFC 1600 pulses; 4s on; 26s off; 40 times/session; 5 sessions a week for 2 weeks Total: 16,000 10 sessions total piTBS_5Hz_lDLPFC; HFrTMS_lDLPFC_10Hz	Brain-navigation software, infra-red system for PFC using patients' MRI scans	MT: 100%	Magstim super rapid magnetic stimulator, 4 booster modules 700mm air-cooled figure eight shaped coil	Same coil as active but angled at 90 °off skull.
O'Reardon et al. (2007)	HFrTMS_10 Hz_lDLPFC_acute phase 3000 pulses per session (4 second, intertrain:26s); 5-day sequence = 30 sessions; 4-6 weeks (37.5 minutes) Total: 60, 000-90, 000 No. of sessions: 20-30 sessions for the acute phase.	5cm anterior to optimal area for stimulating the thumb	120% + first week of acute: 110% visual twitch contralateral hand muscle, beginning of each treatment week	Neuronetics Model 2100 Therapy System Investigational Device	sham- had an embedded magnetic shield- limited magnetic energy reaching the cortex to 10% to less than active coil.
Stern et al. (2007)	HFrTMS_10Hz_lDLPFC_8s train; 52s intertrain interval x 20 trains LFrTMS_1Hz_lDLPFC_1600s train_x1 train LFrTMS_1Hz_rDLPFC_1600s train x 1 train 10 sessions plus 2 weeks follow up	5cm anterior to optimal area for stimulating the thumb	EMG. rAPB MT: 80%	Dantec Magpro stimulator; and Magstim Rapid Super-rapid magnetic stimulator- 8-shaped stimulation coil	Same coil Oriented perpendicular to scalp
Vanderhasselt et al. (2009a)	HFrTMS_10Hz_lDLPFC 1560 pulses; 40 trains of 3.9 s duration, intertrain 26.1s; 20 minutes: Total: 1560 pulses No data re: number of weeks.	MRI- neuro-navigated	EMG MT: 110%	Magstim high speed magnetic stimulator, figure of 8 coil	sham- figure of 8 coil held at 90°, resting on the head

HFrTMS= high frequency repetitive transcranial magnetic stimulation; LFrTMS = low frequency repetitive transcranial magnetic stimulation; aHFrTMS = accelerated high frequency transcranial magnetic stimulation; iTBS= intermittent theta burst stimulation ; piTBS = prolonged intermittent theta burst stimulation; l_DLPFC = left Dorsolateral Prefrontal Cortex; r_DLPFC = right Dorsolateral Prefrontal Cortex; Neuro-navigation, MRI = magnetic resonance imaging; Motor Threshold; EMG = electromyography; rAPB = right abductor pollicis muscle; rFDI = right first dorsal interosseous muscle; rMT = resting motor threshold; l = left; r = right.

Table 5. Study design, number of treatment arms, pre and post stimulation measures on the HDRS, and statistical interpretation from the included studies.

Study	Research Question & Aim	Design	Arms, Stimulation Type: Gender (F/M); [Age, years, SD]	Primary Outcome: Pre-Stimulation (SD) HDRS	Primary Outcome: Post-Stimulation (SD) HDRS	Interpretation
Baeken et al. (2013)	Aim: examine whether a one-week HFrTMS protocol would increase clinical outcomes for patients with depression	Randomised Double-Blind Sham-Controlled Crossover study	2 aHFTMS (7/2) [51.77, 12.10] Sham (5/6) [47.27, 13.66]	HDRS_17item_Pre-Stim **aHFTMS** = _26.33 (8.15) **Sham** = 26.45 (8.71)	Post-Stim Wk1 (20 sessions) **aiTMS** =19.63 (8.35) [25% symptom reduction] **Sham** = 22.36 (10.01)	HDRS scores ↓ over time (pre/ post) ns differences between A/S stimulation.
Duprat et al. (2016)	Aim: examine the effect of aiTBS on depressive symptomatology (active compared to sham aiTBS)	RCT double-blind Sham-Controlled cross-over design	2 iTBS (16/6) [40.1, 11.5] Sham (17/8) [43.2, 12.2]	HDRS_17item_Pre-Stim. **iTBS** = 21.1 (no data) **Sham** = 21.6 (no data)	Post Stim_Wk1 (20 sessions) **iTBS** = 16.4 (nd) **Sham** = 18.96 (nd)	HDRS score↓ over time (pre/ post). Ns differences between A/S stimulation.
George et al. (2010)	Aim: to conduct a high quality, multi-site randomised clinical trial, addressing some of the key limitations of previous RCTs for TMS interventions	Prospective multi-site (4 sites) Randomised Double-Blind Sham-Controlled duration-adaptive study.	2 HFrTMS (58/ 34) [47.7, 10.6] Sham (50/48) [46.5, 12.3]	HDRS_24 item_Pre_Stim **HFrTMS** = 26.3 (5) **Sham** = 26.5 (4.8)	Post Stim_Wk 3 (15 sessions) **HFrTMS** = 21.61 (9.26) [18% symptom reduction] **Sham** = 23.38 (7.43)	HDRS score ↓, and a marginally sd between A/S stimulation, p = 0.06.
Holzheimer et al. (2004)	Aim: examine whether active compared to sham rTMS improve depressive symptoms.	Randomised Controlled Trial- Double-blind	2 HFrTMS (4/3) [40.4, 8.5] Sham (3/5) [45.4, 4.9]	HDRS_17item_Pre_Stim **HFrTMS** = 22.7 (5.3) **Sham** = 20.8 (6.3)	Post_Stim_Wk 2 (10 sessions) **HFrTMS** = 14.6 (3.2) [37% symptom reduction] **Sham** = 15.3 (3.0)	HDRS no different between active and sham stimulation.
Li et al. (2016)	Aim/s: Examine whether post attention task changes frontalΘ power can increase raCC activity and influence active rTMS efficacy.	Randomised double-blind Sham-Controlled trial.	3 HFrTMS + sham RECT (7/5) [39.4, 13.2] HFrTMS + active RECT (8/4) [43.4, 9] Sham +active RECT (6/6) [42.4, 12.5]	HDRS_17item_Pre_Stim **HFrTMS + Sham RECT** = 22.8 (5) **HFrTMS + active RECT** = 22.5 (6.7) **Sham + active RECT** = 21.9 (4.9)	Post_Stim_Wk 1 (10 sessions) **HFrTMS + Sham RECT** = 14.5 (6.1) [36% symptom reduction] **HFrTMS + active RECT** = 12.6 (9.5) **Sham + active RECT** = 19.3 (7.7)	HDRS scores ns between the 3 groups for the first week post stimulation + RECT task
O'Reardon et al. (2007)	Aim/s: Examining the efficacy of active compared to sham rTMS.	Multi-site (23 sites) Double-Blind Randomised, Sham-Controlled Trial.	2 HFrTMS (86/69) [47.9, ± 11.0] Sham (74/72) [48.7, ± 10.8]	HDRS_17item_Pre_Stim **HFrTMS** = 22.6 (3.3) **Sham** = 22.9 (3.5)	HDRS_17item_Post_Stim_Wk4 (20 sessions) **HFrTMS** = 17.4 (6.5) [23% symptom reduction] **Sham** = 19.4 (6.5)	HDRS score ↓, and sd between A/S, p = .006.
Stern et al. (2007)	Aim/s: Examine the efficacy of Low and HFrTMS compared to sham rTMS in reducing depressive symptomatology.	Design: Randomised Parallel group, Double-Blind, Placebo-Controlled Trial.	4 HFrTMS_l (6/4) [53.2, 12] LF_rTMS_l (6/4) [52.3, 9.4] LF_rTMS_r (7/3) [52.8, 9.5] Sham (9/6) [53.3, 9]	HDRS_21item_Pre_Stim **HFrTMS_l** = 27.8 (3.2) **LF_rTMS_l** = 27.6 (3.9) **LF_rTMS_r** = 27.9 (3.8) **Sham** = 27.4 (2.9)	HDRS_21item_Post_Stim_Wk2 (10 sessions) **HFrTMS_l** = 15.1 (6) [46% symptom reduction] **LF_rTMS_l** = 27.6 (5.9) **LF_rTMS_r** = 15.8 (4.8) **Sham** = 26.7 (3.6)	HDRS score ↓ over time for HFrTMS_l but not for sham.
Vanderhasselt et al. (2009a)	Aim/s: examine effects of active compared to sham rTMS on task switching and mood in patients with depression.	Design: Double-Blind, Placebo-Controlled, Cross-over Within Subject.	2 HFrTMS (no data) [no data] Sham (no data) [no data]	VAS_PreStimulation **HFrTMS** = no data **Sham** = no data	VAS_PreStimulation **HFrTMS** = no data **Sham** = no data	VAS ns differences between A/S for VAS ratings

Gender: F= Female; M = Male; SD = standard deviation; Stimulation: HFrTMS_l = high frequency repetitive transcranial magnetic stimulation left; HFrTMS_r = high frequency repetitive transcranial magnetic stimulation right; aHFTMS = accelerated high frequency transcranial magnetic stimulation; iTBS= intermittent theta burst stimulation ;RECT = rACC-engaging cognitive task; piTBS = prolonged intermittent theta burst stimulation; Questionniares: HDRS = hamilton depression rating scale; VAS = visual analogue scale; abbreviations: a = active; s = sham; ns = non-significant ; sd = significantly different; Time: Wk = week

Table 6. The % change of depressive symptom reduction, as measured on the HDRS for active and sham stimulation.

	Sham Stimulation			Active Stimulation			
Study	**Pre HDRS score**	**Post HDRS score**	**% change on the HDRS**	**Pre HDRS score**	**Post HDRS score**	**% change on the HDRS**	**HDRS version**
Baeken et al. (2013)	26.45 (severe)	22.36 (moderate)	-15%	26.33 (severe)	19.63 (moderate)	-25%	17-item
George et al. (2010) *	26.5	23.38	-12%	26.3	21.61	-18%	24-item*
Holzheimer et al. (2004)	20.8 (moderate)	15.3 (mild)	-26%	22.7 (moderate)	14.6 (mild)	-36%	17-item
Li et al. (2016)	21.9 (moderate)	19.3 (moderate)	-12%	22.8 (moderate)	14.5 (mild)	-36%	17-item
O'Reardon et al. (2007)	22.9 (moderate)	19.4 (moderate)	-15%	22.6 (moderate)	17.4 (moderate)	-23%	17-item
Stern et al. (2007)	27.4 (severe)	26.7 (severe)	3%	27.8 (severe)	15.1 (mild)	-46%	21-item
			Mean % Change: -14% [no change/minimally improved].			**Mean % Change: -31% [no change/minimally improved].**	

- We were not able to calculate the "cut-off" for depression severity for George et al., (2010) due to not being able to locate valid cut-off criteria for the 24-item HDRS.

Table 7. Adverse effects reported across the included studies.

Study	Adverse Effects reported A/S	Number of participants who withdrew: Active/ Sham
Baeken et al. (2013)	Headache & fatigue during the first few sessions. Paracetamol aided with this- no information regarding how many people experienced these effects, or differences between A/S.	None reported to have withdrawn due to adverse effects
Duprat et al. (2016)	Pain at site of stimulation and headache reported for the initial few sessions.	None reported to have withdrawn from the study relating adverse effects
George et al. (2010)	Headache (32%); discomfort (17%); insomnia (7%); increased anxiety/depression (6%); fatigue; muscle aches; vertigo; skin pain; facial muscle twitching.	5.4% of patients dropped-out of the study due to pain/headache, and syncope (Active stimulation). A further two had serious adverse effects: worsening of depression (Active stimulation); and paranoid ideation (Sham stimulation).
Holzheimer et al. (2004)	Mild pain in the rTMS stimulation condition	No patients reported to have dropped out for either A/S.
Li et al. (2016)	No adverse effects measured in the paper. Although subjective ratings of pain were taken using a VAS, and reported to be comparable across stimulation type/s.	No patients reported to have dropped out of the study due to adverse effects.
O'Reardon et al. (2007)	Adverse effects measured weekly. Scalp discomfort was main adverse effect reported. No further detail provided.	A = 7.7% and S = 8.2% discontinued treatment/s up to week 4, due to scalp discomfort. However, 16 serious adverse effects were also reported: suicidality, worsening of depressive symptoms (9 = A; 7 =S).
Stern et al. (2007)	Severe headache; pain.	HFrTMS = 0; S = 3 patients withdrew from the study due to severe headache
Vanderhasselt et al., (2009a)	No adverse effects reported.	NA

Stimulation: A= active; S= Sham; rTMS = repetitive transcranial magnetic stimulation; HFrTMS = high frequency repetitive transcranial magnetic stimulation; Questionnaire/s: VAS = visual analogue scale

Does excitatory pre-frontal stimulation reliably increase reward responsiveness? [1]

word count (8373)[2]

[1] Parts of this empirical chapter (introduction, methods, analyses) have reached a stage 1 in principle acceptance as a registered report in the journal, Cortex. Please see Appendix G for the confirmation of the registered report being at stage 1 of in principle acceptance, and Appendix F for Cortex's journal guidelines. Cortex requires additional analyses, such as sample size estimates and further technical details about our procedure- e.g. exclusion criteria, counterbalancing. I have included any additional analyses needed for Cortex in the appendices, which are noted throughout this manuscript.

[2] This word limit excludes footnotes, figure captions, title, abstract, in line with both Cortex's guidelines and the DClin. Psy. & Research Degree Guidelines.

Abstract

Depression is the leading cause of disability worldwide and its effects can be fatal, with over 800,000 people dying by suicide each year. Neuromodulatory treatments such as transcranial magnetic stimulation (TMS) are being used to treat depression. Despite its endorsement by two regulatory bodies: NICE (2016) and the FDA (2008), there are major questions about the treatment efficacy and biological mechanisms of TMS. Ahn et al. (2013) justified the use of TMS in a clinical context in an important study indicating that excitatory TMS increases reward responsiveness. Yet, a pseudo-replication of this study by Duprat et al. (2016) found an effect of active TMS only with the addition of an exploratory covariate to the analyses – trait reward responsiveness. These partially conflicting results call into question the basic scientific validity of using TMS as a treatment. Here we aim to replicate Ahn et al's (2013) key study, and to test the reliability of the effects depending on trait reward responsiveness as described by Duprat et al. (2016). Using excitatory and sham TMS, we tested volunteers using the probabilistic learning task to measure their reward responsiveness both before and after stimulation. We also examined affect (positive, negative) following stimulation. Participants demonstrated an increase in reward responsiveness commensurate with the reward sensitivity literature. However, we did not replicate Ahn et al. (2013) or Duprat et al. (2016) 's key findings. Moreover, following active stimulation, positive affect was reduced, suggesting an increase in anhedonia. Given our findings, we question the basic mechanisms and effects which support the use of TMS for depression, particularly considering potential deleterious effects of reduced positive affect in patients with depression.

Keywords: transcranial magnetic stimulation, reward responsiveness, anhedonia, depression, replication.

1. Introduction

Depression is the leading cause of disability across the globe (WHO, 2017), with over 300 million people estimated to suffer from the disease. It is characterised by pervasive low mood, loss of pleasure, sleep disturbance and reduced libido, amongst other symptoms (see ICD-10; WHO, 1992). These symptoms of depression impair daily functioning, which can impact on employee retention, and quality of life (Kessler, 2012). The effects of depression can be fatal, with as many as 800,000 people globally dying by suicide (WHO, 2017).

It is apparent from both the economic and personal costs of depression that effective evidence-based interventions are of paramount importance. The recommended treatment for moderate depression is psychological therapy with medication (NICE, 2009). However, the long-term effectiveness of these treatments is subject to debate (Ali et al., 2017; Cuijpers et al., 2013; Knekt et al., 2008). More recently the National Institute for Health and Care Excellence (NICE, 2015) in the UK, and the Food and Drug Administration (FDA, 2008) in the USA, approved the use of repetitive transcranial magnetic stimulation (rTMS) to treat depression. TMS has been repeatedly shown to affect cortical excitability (Huang, Rothwell, Chen, Lu, & Chuang, 2011). However, the precise neural mechanisms of action with respect to depression are unclear (Aleman, 2013; Janicak & Dokucu, 2015; Wassermann & Zimmermann, 2012). There are also important questions around its efficacy as a treatment (Chervyakov, Chernyavsky, Sinitsyn, & Piradov, 2015; Janicak & Dokucu, 2015; Martin et al., 2003; Matheson, Shemmell, De Ridder, & Reynolds, 2016; Wassermann & Zimmermann, 2012).

The dorsolateral prefrontal cortex (DLPFC) is the region most commonly targeted in the treatment of depression (Lefaucheur et al., 2014), where TMS appears to reduce symptoms of depression (Conelea et al., 2017; Fitzgerald, Hoy, Daskalakis, & Kulkarni,

2009). In particular, a reduction in negative mood is observed, as measured on the Hamilton depression scale (J. Chen et al., 2013). However, improvement of symptoms is inconsistent across studies and individuals (Fox, Liu, & Pascual-Leone, 2013; see review by Loo & Mitchell, 2005). The evidence-base also shows considerable variation for the number of treatment sessions provided (George et al., 2009) and the type of TMS protocols administered (e.g. Bakker et al., 2015; Chung, Hoy, & Fitzgerald, 2015; Duprat, De Raedt, Wu, & Baeken, 2016). Further discrepancies between treatment protocols extend to the frequencies of the rTMS applied, i.e. low-frequency TMS (i.e. <1 Hz; Liu, Zhang, Zhang, & Li, 2014) versus higher-frequency TMS (Fitzgerald et al., 2009; Liu et al., 2014). Predominantly through demonstrations of effects upon motor cortical excitability, low-frequency (LF; ~ <1Hz) TMS has been associated with reduced cortical excitability, whilst high-frequency (HF; ~ >5Hz) TMS is associated with excitation of neural activity (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005; Siebner & Rothwell, 2003). However, despite these opposing effects on physiology, a meta-analysis conducted by J. Chen et al. (2013) indicated both high and low-frequency TMS yielded similar reductions in depressive symptoms. Despite the similar efficacy of LF-TMS and HF-TMS (see RCT by Eche et al., 2012) the majority of the current protocols use HF-TMS over the left DLPFC to increase excitability in the treatment of depression (Allan, Kalu, Sexton, & Ebmeier, 2012; Wassermann & Zimmermann, 2012).

More recent rTMS protocols have begun using intermittent Theta Burst Stimulation (iTBS) to stimulate the left DLPFC (Duprat, De Raedt, et al., 2016; Duprat et al., 2017). iTBS combines low and high-frequency stimulation and has been shown to be relatively effective in reducing the TMS intensity required to produce a motor evoked response, indicating increased excitability (Huang et al., 2005). However, the precise frequency range of iTBS protocols used in depression treatment is variable (cf. Blumberger et al.,

2018; Bulteau et al., 2017; Duprat et al., 2017; Fitzgerald et al., 2018; Huang et al., 2005). The development of TMS protocols for treating depression is promising, yet the variation between competing treatment protocols has led to inconclusive results. Despite variation between protocols (iTBS, low-frequency rTMS; high-frequency rTMS) all competing treatment protocols are endorsed by NICE (2015) and the FDA (2008).

Common symptoms of depression that have been linked to the functioning of the DLPFC include reduced reward responsiveness and anhedonia (Ballard et al., 2011; Staudinger, Erk, & Walter, 2011). In particular, the DLPFC appears to be innervated via dopaminergic pathways (Der-Avakian & Markou, 2012; Fidalgo et al., 2014), that is neural pathways that are associated with reward responsiveness and pleasure. These dopaminergic pathways are linked to the DLPFC and produce neurotransmitters such as dopamine, which are thought to play a role in external reward anticipation (e.g. reward responsiveness) (Ballard et al., 2011), making this area and its associated projections particularly sensitive to reward responsiveness or lack thereof, e.g. anhedonia. It has been suggested that facilitatory rTMS applied to the left DLPFC stimulates the mesolimbic reward pathway (Janicak & Dokucu, 2015), which is hypoactive in depression (see review by Belujon & Grace, 2017; Koenigs & Grafman, 2009).The probabilistic learning task (PLT; Pizzagalli, Jahn, & O'Shea, 2005) has been robustly related to anhedonia (Pizzagalli, Iosifescu, Hallett, Ratner, & Fava, 2008; Pizzagalli, Jahn, et al., 2005) and reward processing in healthy and depressed individuals (Huys, Pizzagalli, Bogdan, & Dayan, 2013). It has also been associated with neural areas such as the left DLPFC, which have been linked to reward responsiveness using fMRI (Der-Avakian & Markou, 2012; Ott, Ullsperger, Jocham, Neumann, & Klein, 2011) and electrophysiological approaches (e.g. Pizzagalli, Sherwood, Henriques, & Davidson, 2005). More recently the probabilistic learning task

has been used to measure reward responsiveness following rTMS stimulation (Ahn, Kim, & Kim, 2013; Duprat, De Raedt, et al., 2016; Duprat et al., 2017).

One of the most important pieces of evidence that has contributed to the advancement of rTMS to treat depression is the work of Ahn et al. (2013). The authors applied facilitatory HF-TMS to the left DLPFC in controls, leading to heightened reward processing following rTMS compared to sham. Duprat et al. (2016) attempted a partial replication of Ahn et al's (2013) study, also using the probabilistic learning task, and did not find an increase in reward processing in comparison to an alternative control condition. Nevertheless, reward responsiveness appeared to be modulated by participants' trait hedonic capacity (Duprat, De Raedt, et al., 2016), which was taken as further support for the use of facilitatory TMS as a treatment. If such an effect replicates, it might provide further support for the recent suggestion of a personalised approach in the use of rTMS as a treatment (Singh et al., 2019). However, the inconsistency in the effect between Ahn et al. (2013) and Duprat et al. (2016), and the post-hoc nature of the relationship to trait capacity, challenges the reliability of this evidence. Despite the inconsistency between these findings, both of these studies have been used to justify the use of TMS on patients as treatment for depression (e.g. Blumberger et al., 2018; Duprat et al., 2017). Although the studies that we are replicating included non-depressed participants, both Ahn et al. (2013) and Duprat et al. (2016)'s work have contributed to the advancement of using TMS as a treatment for depression. Our replication will include a non-depressed participant sample for equivalence to the studies that we are replicating, and in order to test whether these effects do in fact replicate given the discrepant findings between these papers.

The above evidence indicates that the ability of TMS to alter neuronal functioning may impact key markers of depression (anhedonia and reward sensitivity), despite the mixed evidence. This promising avenue for treatment comes at a time when both the

efficacy and cost of traditional treatments of depression are a concern (Ali et al., 2017; Friborg & Johnsen, 2017; Johnsen & Friborg, 2015). For example, the long-term effectiveness of low intensity Cognitive Behavioural Therapy, once thought to be the gold standard (Layard & Clark, 2014), is now being questioned (Ali et al., 2017). In addition, the efficacy of antidepressant medication compared to placebo for depression is also mixed (cf. Cipriani et al., 2018; Kirsch, 2014). There is a need for novel treatments that show both clinical utility and long-term effectiveness, which TMS may be able to answer.

1.1. Aims and hypotheses of the current study

Here, we propose to examine the utility of TMS as a method to affect reward sensitivity, a key component of depression, aiming to reconcile the discrepant findings in the existing literature (Ahn et al., 2013; Duprat, De Raedt, et al., 2016) by conducting a replication of Ahn et al. (2013). We will include additional measures of trait hedonic capacity and apply the increasingly utilised excitatory iTBS from Duprat et al. (2016), to redress the discrepant findings (described further in the methods section). The study will use the probabilistic learning task and will be extended through the inclusion of measures of mood. As iTBS (Duprat, Desmyter, et al., 2016; Duprat et al., 2017) is starting to be more commonly used in the treatment of depression than the HF-TMS protocol used by Ahn et al (2013), this replication will apply iTBS (Blumberger, 2018; Bulteau et al., 2017; Duprat et al., 2017) (see Appendix H for any deviations from the replication and reasoning). Comparisons will be made to a sham control.

1.1.1. Primary Hypothesis

We predict an increase in response bias (RB) which measures a participant's ability to learn to favour one type of stimuli over another, typically increasing over sequential blocks (1,2,3) of the reward learning task, the PLT (both RB and the PLT are explained in greater detail in section 2.3.1 and 2.6). We predict that RB will increase as a function of

active compared to sham stimulation, the basic expression of which would be an elevation in RB in block one following active stimulation compared to the equivalent sham stimulation. This is our critical effect of interest and was demonstrated by Ahn et al. (2013) in the first block of their probabilistic learning task. Our prediction is theoretically grounded in the multiple treatment studies that have found that active stimulation decreases depressive symptomatology, such as anhedonia, compared to sham stimulation (cf. Fitzgerald et al., 2009; George et al., 2010a; Lam, Chan, Wilkins-Ho, & Yatham, 2008; O'Reardon et al., 2007).

1.1.2. Secondary Hypotheses

We expect scores from the secondary questionnaire, which measures mood (Positive and Negative Affect Scale; PANAS; Watson, Clark, & Tellegen, 1988), to show a reduction in negative mood and increase in positive mood respectively (e.g. Chaves, Lopez-Gomez, Hervas, & Vazquez, 2017; Moulier et al., 2016). In particular, we expect active compared to sham TMS will modulate mood ratings.

1.1.3. Replication interactions of interest

In addition to the above hypotheses, Ahn et al. (2013) and Duprat et al. (2016) demonstrated significant interactions and secondary effects consistent with TMS having a positive effect in reducing symptoms of depression, as measured using RB from the probabilistic learning task. However, neither Ahn et al's (2013) nor Duprat et al's (2016) significant interactions are fully in line with the theoretical prediction that active compared to sham stimulation will increase RB in the probabilistic learning task. Ahn et al. (2013) reported a significant interaction between block and stimulation in which RB increased for active stimulation in *block 1,* but interestingly also increased following *sham* stimulation compared to *active* for *block 2* of the task. Similarly, Duprat et al's (2016) main significant interaction of time (pre/post stimulation) × stimulation (active/sham) × block (1,2,3) was

only apparent when TEPS-CON – measuring hedonic capacity, was added as covariate (see Methods; Questionnaires). This indicated an increase in RB as a pre-stimulation baseline for active compared to sham stimulation when hedonic capacity was taken into account. For completeness of this replication, we will test for the exact replication of these secondary interactions.

2. Method

2.1. Task overview & Procedure

The iTBS protocol used by Duprat et al. (2016; 2017) was used to stimulate the DLPFC (for further details see 'Transcranial Magnetic Stimulation' section). Participants were asked to attend two testing sessions (see Figure 1) and were assigned to either a sham or active iTBS condition, determined through initially flipping a coin (see Duprat et al., 2016); (e.g. heads = sham; tails = active). If participants were assigned to the active condition during the first session, they received sham stimulation during the second session, and vice versa. To determine participants' reward learning, participants completed a PLT (Pizzagalli, Jahn, et al., 2005) prior to iTBS stimulation. Participants were asked to complete a further PLT following TMS stimulation.

During the first session, participants were asked to complete the Temporal Experience of Pleasure Scale questionnaire (TEPS; Gard, Gard, Kring, & John, 2006) and the PLT task, prior to TMS stimulation. In both sessions participants were asked to complete a further questionnaire, to determine whether mood ratings have changed before and after TMS stimulation (PANAS; Watson, Clark, & Tellegen, 1988; see 'Questionnaire section' for further information). The PANAS was undertaken before and after completing the PLT task.

2.2. Participants

Twenty-four participants were recruited from a TMS participant database at Cardiff University Brain Research Imaging Centre (CUBRIC) in Cardiff University. All participants had undergone safety screening for contra-indications of TMS. Nineteen participants (*M*age = 21.53; *SD* = 1.577; 15 females/ 4 males) were included in the final TMS analyses, with five excluded (equipment issues: n= 3; voluntary withdrawal: n =2). All participants were right-handed, spoke English fluently and proficiently and had no history of mental health difficulties. Participants were paid £10 per hour, and a maximum of £12 on the reward-learning tasks. The ethics committee at Cardiff University's School of Psychology has approved the study. For exclusion criteria, see Appendix I.

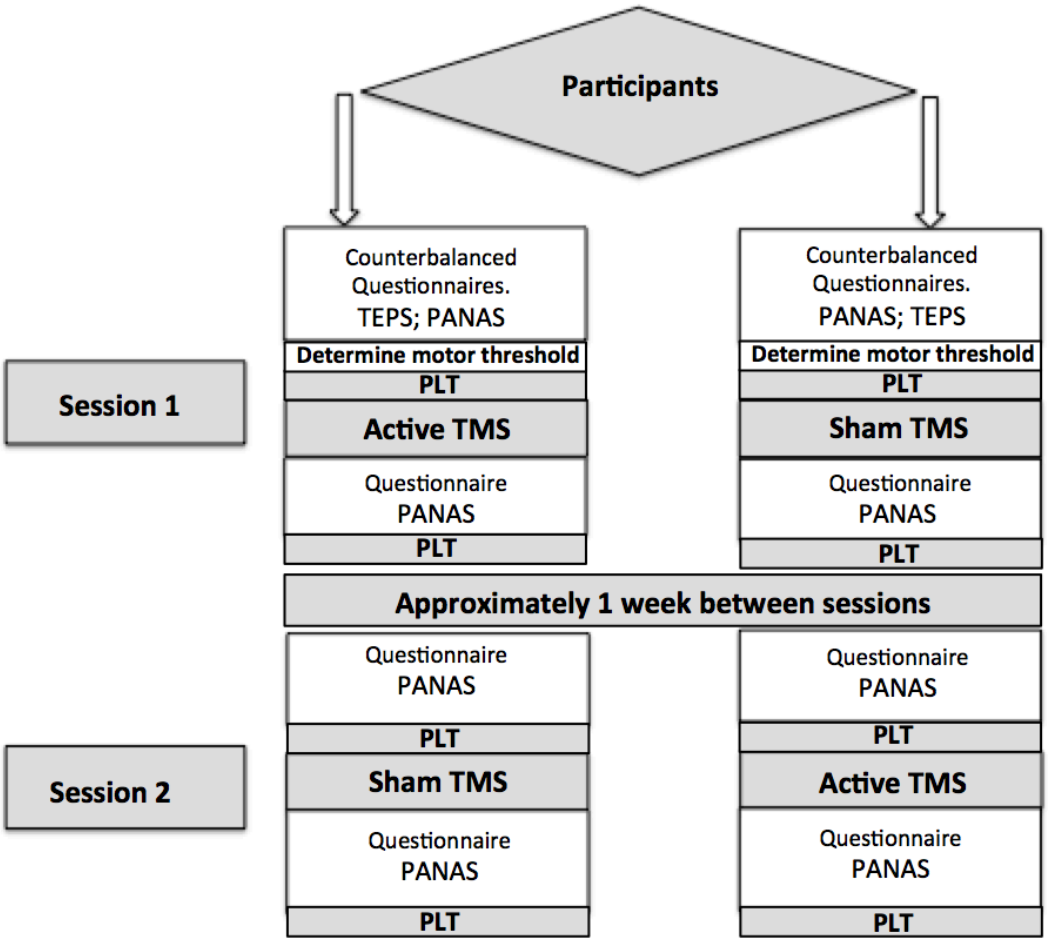


Figure 1. Task procedure for participants entering study. TEPS = Temporal Experience Pleasure Scale questionnaire; PANAS = Positive and Negative Affect Scale; PLT = Probabilistic Learning Task; TMS = Transcranial Magnetic Stimulation.

2.3. Study Design

The study is a repeated measures design, with participants undergoing all conditions; 2 (Stimulation: Active and Sham), 2 (Time: Pre and Post stimulation), 2 (Condition: Rich and Lean) and 3 (Block: One, Two, Three). In accordance with Duprat et al. (2016), we administered the hedonic capacity questionnaire, TEPS (Gard et al., 2006).

We also examined the effect of rTMS (sham, active) on mood, before and after stimulation.

2.3.1. Probabilistic Learning Task (Pizzagalli, Jahn, et al., 2005; Pizzagalli et al., 2008)

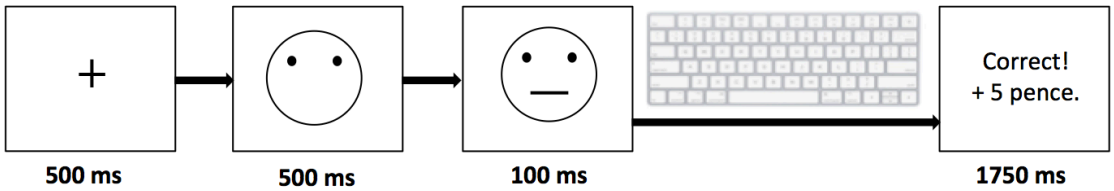


Figure 2. Trial schematic in the probabilistic learning task (based on Pizzagalli, Jahn, et al., 2005).

The PLT is based on signal detection theory (McCarthy & Davison, 1979) and measures an individual's decision to choose stimuli A over stimuli B (Pizzagalli, Iosifescu, et al., 2008; Pizzagalli, Jahn, et al., 2005) based on a prior reinforcement learning schedule. Participants received varied rewards based on an asymmetric reinforcement schedule (e.g. 'rich' or 'lean' stimuli; Pizzagalli et al., 2005). Previous literature has suggested that reward learning is biased towards the most rewarded stimuli (Pizzagalli, Goetz, Ostacher, Iosifescu, & Perlis, 2008; Pizzagalli, Iosifescu, et al., 2008; Pizzagalli, Jahn, et al., 2005). Individuals with depression tend to display a lower RB to more frequently rewarded

stimuli compared to non-depressed controls, suggesting difficulty with reinforcement leaning (Pizzagalli et al., 2008).

Each application of the task (four in total, two in session one, and two in session 2; see Figure 1) was comprised of three blocks of 100 trials (Block 1, Block 2, Block 3). For each trial, a fixation cross appeared on the screen for 500ms, followed by a cartoon face without a mouth for 500ms. Another cartoon face was subsequently presented on the screen with either a 'long' (13mm) or 'short' (11.5mm) mouth for 100 ms. The participant provided a keyboard response to assign whether the mouth was 'long' or short'. If correct, a feedback screen was presented for 1750 ms, which was either blank or announced that the participant had won five pence. For each block of 100 trials, a pseudo random sequence of 50 long and 50 short mouths was presented.

One mouth-type, the "rich" condition, was selected at random to be rewarded three times as often as the other mouth-type, the "lean" condition. In total 40 trials were rewarded per block, 30 of these were the "rich" condition and 10 of these were the "lean" condition. Participants were asked to try and win as much money as they can. Also, participants were told that not all trials will be rewarded, but they were not told about the rich versus lean stimulus.

The PLT was implemented in PsychoPy (Peirce, 2007), and was presented on a Microsoft PC with an Asus LCD monitor (60 Hz refresh rate). Please see Appendix J for counterbalancing.

2.4. Transcranial Magnetic Stimulation

Following the iTBS protocol of Duprat et al. (2016), we applied iTBS stimulation using a MagStim Rapid ^{2} stimulator (Magstim Company Limited, Wales, UK), which was connected to a 70mm "figure eight" shaped cooled coil (P/N 3910-00). We used the Brainsight neuronavigation system (Brainsight Rogue Research, Inc.) to accurately target

the left DLFPC. In the instances of MRI scans not being available, we used the default MNI average brain scan in Brainsight 2.3, Rogue Research, Inc to target the left DLPFC (see Appendix K for further information). As in Duprat et al's (2016) protocol, individual motor thresholds were determined using surface electromyography to produce a motor evoked potential in the right abductor pollicis brevis muscle, during the first testing session.

Participants were asked to take part in two testing sessions, where they received either active (see Appendix L for stimulation parameters) or sham iTBS with an interval of approximately one week between the sessions.

For the sham stimulation, a sham coil was used (P/N 3950-00), which has the same visual appearance and auditory artefact as the active coil but does not deliver appreciable magnetic stimulation. Therefore, sham stimulation follows the same procedure as the active condition. Order of stimulation was counterbalanced between participants and initial allocation to active and sham conditions was randomised using a coin toss.

2.5. Questionnaires

2.5.1. Temporal Experience of Pleasure Scale (TEPS; Gard et al. 2006; see Appendix M)

The TEPS consists of 18 self-report items, of which there are two subscales: anticipatory (TEPS-ANT; ten items) and consummatory (TEPS-CON; eight items) pleasure. Consummatory pleasure has been linked to immersive pleasurable experiences (Gard et al., 2006; Gard, Kring, Gard, Horan, & Green, 2007), and anticipatory pleasure has been related to reward responsiveness (Gard et al., 2006). The sum of these two scales provides a measure of hedonic capacity. The higher the total score on these scales, the greater the hedonic capacity. Conversely, the lower the total score, the lower the hedonic capacity (anhedonia, e.g. Gard et al., 2007; Strauss, Wilbur, Warren, August, & Gold,

2011). Internal consistency for each subscale has good reliability, consummatory (α = 0.71) and anticipatory (α = 0.74) (Gard et al., 2006). The TEPS scale has been used with both non-clinical (the target of this replication; Ahn et al., 2013; Duprat et al., 2016) and clinical populations (e.g. Gard et al., 2006, 2007; Sherdell, Waugh, & Gotlib, 2012).

2.5.2. Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988; see Appendix N)

To determine participants' current mood, we used the PANAS (Watson, Clark, & Tellegen, 1988). This scale consists of ten positive (e.g. 'proud') and ten negative (e.g. 'jittery') adjectives. Participants rated on a Likert scale (1-5), the extent to which they currently feel that emotion. Scores on the PANAS range between 10-50, on both positive and negative subscales separately. A relatively higher score on the positive affect scale (PA) reflects a more positive mood, whereas a higher score on the negative affect (NA) subscale reflects a more negative mood (see Watson, Clark, & Carey, 1988). Each subscale of the PANAS has good internal reliability, PA (α = 0.89) and NA (α = 0.85) (Crawford and Henry, 2004). The PANAS has been used with both clinical populations (people diagnosed with depression and anxiety) and in non-clinical populations (Watson, Clark, & Carey, 1988). Lower scores on the PA scale have been related to anhedonia and depression (e.g. Crawford & Henry, 2004; Watson, Clark, & Carey, 1988). Higher scores on the NA scale have been related to depression and anxiety disorders (e.g. Crawford & Henry, 2004). The PANAS has good discriminant and convergent validity (Crawford & Henry, 2004).

2.6. Primary Dependent Variable - Probabilistic Learning Task: Response Bias

The RB relates to the participant's preference to the most frequently rewarded stimulus ("rich") when compared to the least rewarded stimulus ("lean"). RB was calculated using the formula below. The response rate will increase if the participant selects "rich" stimuli more frequently than "lean" stimuli, regardless of accuracy. In

accordance with previous studies, i.e. Ahn et al. (2013), Duprat et al. (2016), Pizzagalli et al. (2005), RB is likely to increase between blocks 1 & 2 as a consequence of reinforcement learning. The number of trials presented in the PLT are equal to those used in previous studies applying the PLT (e.g. Ahn et al., 2013; Chevallier et al., 2016; Duprat et al., 2016; Lancaster, Heerey, Mantripragada, & Linden, 2015; Lancaster et al., 2012; Pizzagalli, Goetz, et al., 2008; Pizzagalli, Iosifescu, et al., 2008; Pizzagalli, Jahn, et al., 2005). Similar to Ahn et al. (2013), effect sizes were reported using partial eta-squared (η_p^2). In accordance with Duprat et al. (2016), we also calculated Cohen's *d* to evaluate effect sizes, and if the assumption of sphericity was violated Greenhouse Geiser correction was applied to the data.

$$RB = \left[\log b = \frac{1}{2} \log \left(\frac{Rich\ correct * Lean\ incorrect}{Rich\ incorrect * Lean\ correct} \right) \right]$$

Figure 3. Formula for calculating response bias (RB), taken from Pizzagalli, Jahn, et al. (2005).

2.7. Bayes Factors and sample size estimates

We used a Bayesian approach (Dienes, 2014) to estimate the likely sample size needed to provide support for the null or alternative hypothesis, related to our effects of interest (see Appendix O for precise calculations). For all hypotheses, we estimated a sample size of 30[3] to provide substantial support for the null or alternative (substantial support detailed further below).

2.8. Analyses (pre-registered) & prior calculations

In line with our hypotheses (primary; secondary; main replication of interest), we conducted Bayesian analyses as our primary decision statistic of interest. In contrast to frequentist statistics, Bayesian statistics allow us to estimate how much evidence our data

[3] I tested 24 participants over 2 x 2hr sessions (with 19 useable datasets), however, with the advent of COVID, testing was truncated and therefore we did not meet the inclusion threshold of 30.

provides for the null, compared to the alternative hypothesis (Dienes, 2014; Dienes & Mclatchie, 2018). As our report is geared to determine whether the effects of Ahn et al.(2013) and Duprat et al. (2016) replicate, Bayes analyses enabled us to model the distribution of prior effects (prior distribution) against observed values to determine the strength of evidence towards an effect or the null (e.g. posterior distribution)(Dienes, 2008). For our primary hypothesis and replication of interest, our priors were modelled to be equivalent to the effects reported by Ahn et al., (2013) and Duprat et al., (2016) for equivalence, and as recommended by Dienes et al., (2008) for replication studies. Namely, we have made the assumption that our effects would be comparable to those of Ahn et al., (2013) and Duprat et al., (2016) based upon drawing from a similar participant sample and assuming no additional variance in our study and measures. We also computed traditional frequentist statistics for each hypothesis. We include a pre-planned analysis section, as specified by Cortex, to provide clarity regarding our pre-registered results and prior calculations. In our results section, we provide a full exposition of the analyses related to our pre-registered analyses (see Appendix P for behavioural results) and detail any exploratory (non-registered) tests. Bayes Factors (BF) greater than 3 are thought to demonstrate substantial evidence for the alternative hypothesis (Jeffreys, 1998; Dienes, 2016) and those that are less than 1/3 have been also thought to show substantial evidence in favour of the null hypothesis (Jeffreys, 1998; Dienes, 2016), with BFs between 1/3 and 3 considered as "insensitive" (Jeffreys, 1998; Dienes, 2016) (see Appendix Q for Cortex's cut-off).We calculated our Bayesian results using Dienes (2008)'s BF calculator.

2.8.1. Main analyses: Primary Hypothesis

We designed our replication attempt to find an effect that uses the same task (the PLT) and stimulation (active, sham) as the target publications (Ahn et al., 2013; Duprat et al., 2016), and is consistent with the evidence-base that RB will increase as a function of

active TMS. Therefore, our main critical one-degree of freedom test of interest is based on that of Ahn et al's (2013) direct comparison of the effects of active compared to sham stimulation on RB for Block 1 of the PLT (active block1 – sham block 1).

As our primary analysis, we calculated a BF applied to our data modelling H1 as a *t*-distribution using the mean difference and *SE* equivalent to those observed by Ahn et al. (2013) for their corresponding one *df* test of interest ($M = 0.14$; $SE = 0.063$, $df = 17$) (Dienes & Mclatchie, 2018).

In addition, we computed a complimentary frequentist paired *t*-test for active compared to sham stimulation for RB in Block 1.

2.8.2. Secondary analyses

Our secondary hypothesis that active compared to sham TMS will increase positive mood and decrease negative mood was based on Chaves et al's (2017) clinical trial study. Amongst other variables, Chaves et al. (2017) measured mood using the PA and NA subscales of the PANAS questionnaire, before and after psychological therapies such as Cognitive Behaviour Therapy (CBT). Since Chaves et al., (2017)'s work is a recent clinical trial that tested a large participant sample (n =95) and measured change following psychological therapy, using the same measure (e.g. the PANAS) as we propose here, we set our prior to expect a similar effect. As a prior is a "best guess" of an effect size based on a set of assumption and similar samples, and study designs (e.g. Dienes, 2008), Chaves et al., (2017)'s study provides an estimate of an effect of a psychological therapy, change following a mood questionnaire and a large sample size which aids precision. It is important to note, however, that our intervention was TMS and not a psychological therapy per se, but no studies to our knowledge have been conducted using mood measures and TMS in a similar demographic sample.

For our secondary analyses, to assess the evidence that active stimulation compared to sham will (a) increase positive affect, and (b) decrease negative affect, we calculated BFs (Dienes & Mclatchie, 2018) modelling H1 as a *t*-distribution as the mean difference and *SE* observed by Chaves et al. (2017) for treatment effects on (a) changes in positive affect (M = 4.91, SE = 0.82, df = 95) and (b) changes in negative affect (M = -5.92, SE = -0.94, df = 95) (Dienes, 2008; Dienes & Mclatchie, 2018). To obtain a value that could be used in a 1 df test, we subtracted the post stimulation PANAS PA score from the pre PANAS PA stimulation ratings for both active and sham separately. We subsequently computed the difference between active and sham for these ratings. More negative scores for this index relate to lower positive mood. PANAS NA values were computed similarly. More negative scores on this index indicate mood has decreased.

We also conducted two paired sample *t*-tests as complimentary frequentist statistics comparing post active and sham stimulation, one applied to the positive affect measure and the other applied to the negative affect measure.

2.8.3. Replication interactions of interest

The previous critical effects of interest in both Ahn et al. (2013) and Duprat et al. (2016) were significant interactions. Ahn et al. (2013) reported a significant 2 × 3 interaction between Stimulation (active, sham) and Block (1,2,3) for reward learning. Duprat et al. (2016) reported a significant 2 × 2 × 3 interaction (with the presence of the TEPS-CON covariate) for Time (pre, post) × Stimulation (active, sham) × Block (1,2,3 on the reward learning task). While these are consistent with the primary outcome of active TMS increasing reward responsiveness compared to sham (main analyses), these interactions are not the basic effects one would expect of active TMS having an effect on the PLT. Therefore, the interactions reported by Ahn et al. (2013) and Duprat et al. (2016) are of secondary interest.

To test for the interactions described by Ahn et al. (2013) and Duprat et al. (2016), we computed a BF for both of their reported interactions. Following the method outlined in Dienes (2014) for reducing interaction effects to a one-degree of freedom test, for Ahn et al's (2013) primary finding of a Stimulation × Block interaction, we computed the differences between active and sham in each block, then subjected these scores to a linear contrast (e.g., B1 + -0.5*B2 + -0.5*B3) to reduce the interaction term to a one-degree of freedom test.

We subsequently used the BF calculator (Dienes, 2008) to apply the parameters calculated from Ahn et al's (2013) primary interaction: M_{diff} = 0.16, *SE* = 0.08, *SD*= 0.33, and the obtained M_{diff} (0.087) and *SE* (0.092) for the interaction from our study, reduced to a one-degree of freedom test as described above.

For Duprat et al's (2016) interaction (Time × Stimulation × Block with the added covariate of hedonic capacity; TEPS-CON), we used the same procedure as described above to reduce the significant interaction into its constituent one-degree of freedom test, prior to using these parameters in the BF calculator (Dienes & Mclatchie, 2018). We computed a linear contrast between active and sham in each block, once for each level of the 'Time' factor, and then compute the differences between these two scores, reducing the comparison to a simple difference score. To account for the presence of the covariate, we used the covariate-adjusted means (as were reported for Duprat et al., 2016's interaction term above), to calculate our one-degree of freedom values. That is, we computed a repeated-measures ANCOVA with the factors (Time × Stimulation × Block with the covariate) to obtain the resulting adjusted means to compute our mean differences (similar to Dienes, 2014), and SE for calculating the resulting BF. These were integrated with a prior based on the results obtained from reducing Duprat et al's (2016) results into a one-degree of freedom test; M_{diff} = 0.14, SE = 0.07, *SD* = 0.33, and the obtained M_{diff} (0.205)

and *SE* (0.165) from the interaction in our study following the same principle for reducing the interaction into a 1*df* using the same weighted contrast as above.

We also report all frequentist equivalents of the above statistics, as well as a full exposition of the ANOVAs and interactions for both Ahn et al. (2013) and Duprat et al. (2016).

2.8.4. Manipulation Checks

We conducted an outcome neutral test, where we compared participants' RB scores from block three of the PLT to the value of zero, using a one-sample *t*-test. At the group level, this was achieved by computing the average RB in block three for the post sham stimulation session, before submitting the scores to a one-sample *t*-test. This test will reveal whether participants have a RB significantly larger than zero, indicating if they are responding more to rich compared to lean stimuli, after undergoing the first and second blocks of the study. Taking the values from block 3 following sham stimulation will provide a baseline measure of the propensity to acquire a RB. We computed the Bayesian equivalent of this one-degree of freedom test, that is RB for block 3 of the PLT compared to the numerical value of zero, applying priors based on the mean of the effects and standard errors reported for block three of post sham stimulation for both Ahn et al. (2013) (M = 0.19, SE = 0.061, df =17, SD = 0.26) and Duprat et al. (2016) (M = 0.26, SE = 0.070, df = 20, SD = 0.32).

3. Results

3.1. Pre-registered: Primary hypothesis

We computed a BF applied to our observed data (M = -0.15, SE = 0.26) with a prior based on the mean difference and *SE* observed by Ahn et al. (2013) for their corresponding one *df* test of interest (M = 0.14; SE= 0.063, df = 17, SD = 0.26). The resultant BF was = 0.601, which indicates a value that is insensitive –i.e. no support for either the null or the

alternative hypothesis. Our complimentary frequentist one-sample t-test indicated a non-significant difference from zero for RB (active block 1 minus sham block 1) for block 1 of the PLT, t (18) = -0.592, p = 0.562.

Combined, the result from both the Bayes and frequentist statistics indicate no conclusive support for either the alternative or null hypothesis.

3.2. Pre-registered: Secondary hypothesis

3.2.1. Positive Affect

Using Dienes (2008) online Bayes calculator with the pre-specified t-distribution of Chaves et al., (2017) for PA, (M = 4.91; SD = 8.075), and our sample Mean (-3.22) and SE = 1.159, resulted in BF = 2.55. This BF indicates a weak effect towards the alternative hypothesis, suggesting that positive mood ratings were reduced, i.e. mood was less positive for active (M = -3.500; SE = 1.269) compared to sham (M = -0.278; SE = 1.226) stimulation. However, we currently have insufficient data to be certain of this weak effect. For our pre-registered complimentary frequentist analysis, we computed a paired t-test for post PA for active compared to post sham PA. There were no significant difference following post active (M = 27.56; SE = 2.073) compared to post sham (M = 28.22; SE = 7.50) stimulation, t(17) = -0.522, p = 0.588, d = -0.130, indicating no mood changes for post stimulation conditions.

3.2.2. Positive Affect - Exploratory Analysis

We conducted a non-pre-registered paired-samples t-test for post-pre active PA versus post-pre sham PA, for the Bayes analysis above. We found support for the alternative hypothesis with a significant difference following active (M = -3.500; SE = 1.269) compared to sham (M = -0.278; SE = 1.226) stimulation, t(17) = -2.780, p = 0.013, d = -0.655. This finding compliments our Bayes analysis, suggesting that positive mood is reduced following active compared to sham stimulation.

3.2.3. Negative Affect

We calculated the BF for the NA subscale of the PANAS, based upon a model of Chaves et al., (2017) (*M* = -5.92; *SE* = -0.94; *SD* = -9.246), and the Mean (0.222), and *SE* = 0.488 of our observed data, which yielded a BF of = 0.048. This result suggests strong support for the null, with a BF of 0.048, indicating neither active nor sham TMS affects negative mood as measured on the PANAS NA. Similarly, for our complimentary paired-samples t-test (post active NA versus post sham NA) following active (*M* = 10.72; *SE* = 0.253) or sham (*M* = 10.67; *SE* = 0.256) stimulation for the NA subscale of the PANAS, there were no significant differences between conditions, $t(17) = 0.236$, $p = 0.816$, $d = 0.056$.

3.2.4. Negative Affect – Exploratory Analysis

A paired-samples t-test for post-pre active NA versus post-pre sham NA, indicated that there was no significant difference for NA mood for active (*M* = -0.278; *SE* = 0.360) or sham (*M* = -0.500, *SE* = 0.452) stimulation, $t(17) = 0.455$, $p = 0.655$, $d = 0.107$.

3.3. Pre-registered: Replication/s of Interest

As pre-specified, we reduced Ahn et al., (2013)'s primary interaction (Simulation × Block), to a one-degree of freedom test, through computing the difference between active and sham for each block, then subjecting these scores to a linear contrast, (described in the analysis section). We applied the parameters calculated from Ahn et al's (2013) primary interaction: M_{diff} = 0.16, *SE* = 0.08, *SD*= 0.33, and our obtained interaction, M_{diff} (0.087) and *SE* (0.092) using the method reported above. The resultant BF was 0.403, which provides evidence towards the null hypothesis, but is not conclusive. Congruent with Ahn et al. (2013)'s design and main analysis, we also ran a 2 × 3 frequentist Repeated Measures ANOVA, with post stimulation (active, sham) and PLT Block (1,2,3) as the within subjects' factors. No data violated the assumption of sphericity, all $ps > .762$. As

predicted, there was a main effect of block for the PLT $F(1, 18) = 5.775$, $p = 0.007$, $\eta^2 = .243$, $d = 1.133$. Post-hoc comparisons using Bonferroni correction indicates that response bias for block 3 ($M = -0.273$; $SE = 0.096$) was larger than the response bias in block 1 ($M = -.138$; $SE = 0.069$), $t = -3.599$, $p = 0.006$, $\eta^2 = 0.149$, $d = -0.826$, suggesting response bias had increased irrespective of stimulation type across the blocks (see Figure 4). There were no other significant main effects or interactions either for stimulation (active versus sham) or for stimulation x block respectively, $F(1, 18) = 1.037$, $p = 0.322$, $\eta^2 = 0.054$, $d = 0.478$; $F(2, 36) = 0.884$, $p = 0.422$, $\eta^2 = 0.047$, $d = 0.444$.

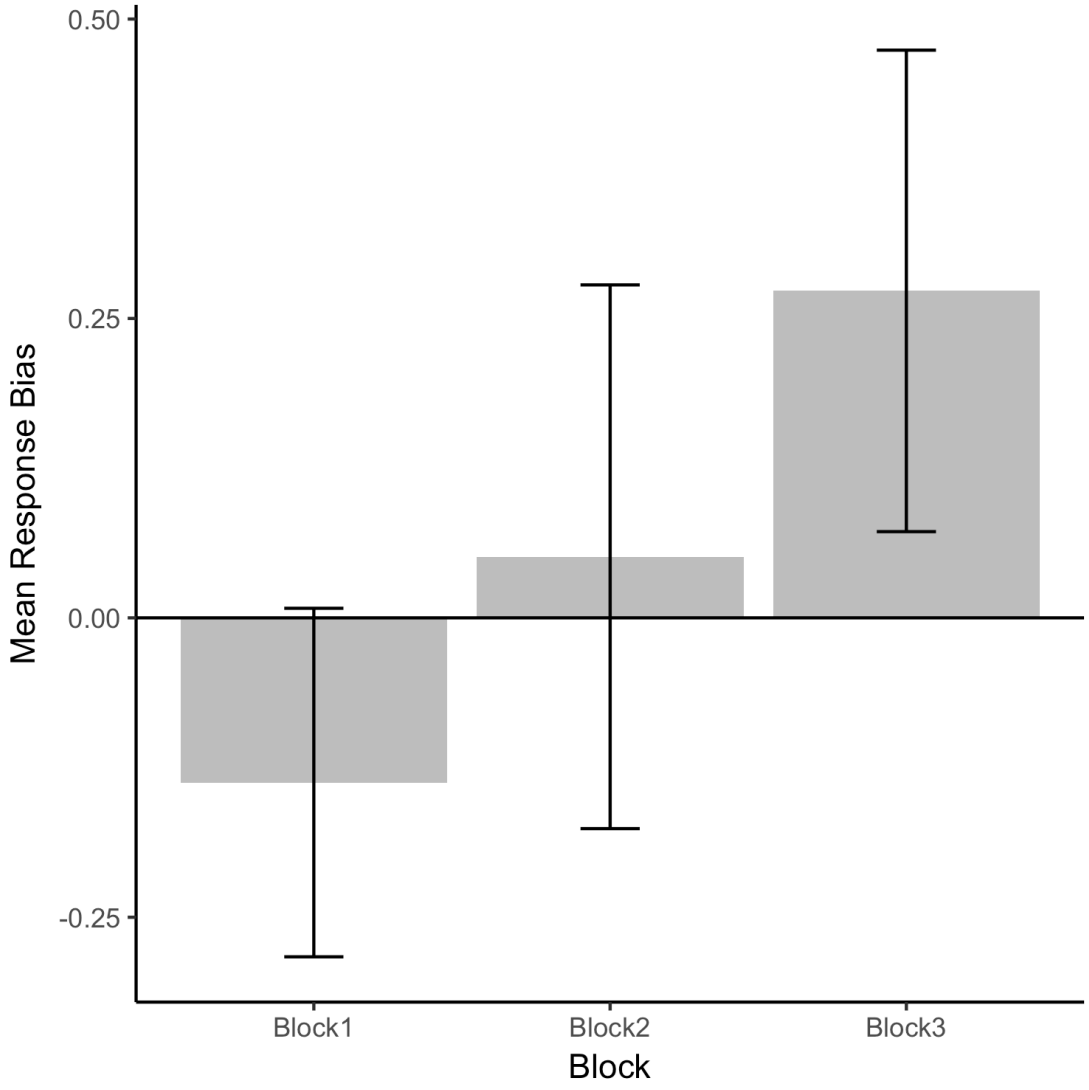


Figure 4. Graph depicting the main effect of increase in RB from blocks 1 to 3, for the combined means of active and sham conditions.

Similarly to the analyses above, we reduced Duprat et al's (2016) main interaction (Time × Stimulation × Block with the added covariate of hedonic capacity; TEPS-CON) into a one-degree of freedom test, M_{diff} = 0.14, *SE* = 0.07, *SD* = 0.33. To account for the presence of the covariate, we used the covariate-adjusted means (as were reported for Duprat et al., 2016's interaction term above), to calculate our one-degree of freedom values. That is, we computed a repeated-measures ANCOVA with the factors (Time × Stimulation × Block with the covariate) to obtain the resulting adjusted means to compute our mean differences (similar to Dienes, 2014), and SE for calculating the resulting Bayes Factor. Using the values explicated above (the M_{diff} and *SE* calculated for Duprat et al. (2016)'s primary interaction), and the obtained M_{diff} (0.205) and *SE* (0.165) from the interaction in our study ,we calculated a BF of 0.737, which does not provide any conclusive evidence for H1 or H0. For our complimentary pre-registered frequentist statistics in line with Duprat et al. (2016)'s main analyses, we conducted a 2 (time: pre, post) × 3 (block: 1,2,3) × 2 (Stimulation: active, sham) Repeated Measures ANCOVA with TEPS-Con included as the covariate. There were no significant main effects or significant interactions, all *F*s between 0.231 and 1.550, all *ps* between 0.228 and 0.775.

3.4. Pre-registered: Manipulation Checks.

We conducted a frequentist one sample t-test for the average RB in block 3 of the post sham stimulation PLT. The test indicates that RB has increased compared to zero, indicating the task has been effective, with participants displaying a propensity to acquire a response bias, that is responding more to rich compared to lean stimuli over the blocks, although this test did not reach statistical significance; $t(18) = 2.033$, $p = 0.057$, $\eta^2 = 0.052$, $d = 0.466$. However, when conducting the manipulation checks modeled on the prior parameters (i.e. based on previous existing effects) of Ahn et al., (2013), with our

observed data (M_{diff} = 0.317, SE = 0.156), we obtained a BF of 3.307. We also obtained a BF of 3.054, for H1 modeled upon Duprat et al., (2016), and our observed data. These BFs suggest evidence in favour of H1- that is participants are acquiring a RB over the blocks.

4. Discussion

We examined whether active compared to sham TMS was effective in generating an increase in reward responsiveness, in a sample of healthy control participants, akin to the studies we were replicating (e.g. Ahn et al., 2013 and Duprat et al., 2016). Overall, we found no conclusive evidence for our primary hypothesis and replication interactions of interest. For our secondary hypotheses, we noted substantial evidence towards the null for negative mood, and significantly reduced positive mood following stimulation, indicating stimulation influenced mood changes, in contrast to our predictions. We discuss these findings in depth below.

Firstly, we predicted that response bias would increase as a function of active stimulation in block one of the PLT, which was our primary hypothesis. However, our Bayes Factor was inconclusive (BF = 0.601), indicating no support for either the null or alternative hypothesis. Similarly, our frequentist *t*-test indicated a non-significant result suggesting no support for the alternative. Taken together, the frequentist and Bayesian findings provide an inconclusive result, as to whether our primary hypothesis replicates.

Why might this be? We had estimated a sample size of 30 participants to generate a Bayes Factor of 0.14, when assuming the same effect, and variability as that of Ahn et al., (2013)'s primary finding. Therefore, it is unsurprising that we obtained an inconclusive finding using a sample size of 19, of which testing was truncated by the recent global pandemic. However, it is possible that our null effect compared to Ahn et al., (2013)'s primary finding could be explained by gender differences. Namely Ahn et al., (2013) included an all-male sample (n =18) whereas our study included a primarily female

population (n = 15 females; and 4 males). Specifically, gender differences have been reported for processing styles (e.g. Boggio, Roche, da Silva, & Fregni, 2008; Byrne, & Worthy, 2015). Thus, our samples could reflect a differential proclivity to develop a RB in males and females. Moreover, our sample was more akin to the ratio and prevalence rates of females to males who suffer from depression (e.g. Albert, 2015). Therefore, one could argue our sample confers greater ecological validity in assessing the potential clinical relevance of TMS for depression.

4.1. Replications of interest

Using a Bayesian approach, we did not provide substantial support for neither Ahn et al., (2013) nor Duprat et al., (2016)'s main interactions of interest. We obtained a weak effect (BF = 0.403) favouring the null for our replication of Ahn et al., (2013)'s stimulation × block interaction. However, we did not obtain a value under the cut-off of 1/3, which would have provided "substantial" support for the null, according to broadly accepted standards (e.g. Dienes, 2014; Jeffreys,1998).

It is possible that we may not have had sufficient participants, as indicated by our sample size estimate, which suggested a sample size of 30 to obtain a BF of 0.20. However, the theoretical underpinning of this interaction appears questionable, with Ahn et al., (2013)'s post active compared to post sham stimulation providing an increase in RB for block 1 of the PLT, which the authors argue is evidence for the enhanced effect of one session of active TMS. The increased effect of RB for block one for active compared to sham stimulation in Ahn et al., (2013)'s study was surprising as the extant reward responsiveness literature evidences increased RB across sequential blocks of the task, as a function of reinforcement learning (e.g. Pizzagalli, Iosifescu,et al.,2008; Pizzagalli et al., 2005; Lancaster et al., 2012), that is participants learn to favour the most frequently rewarded stimulus. In contrast to the PLT literature conduced with healthy controls, Ahn et

al., (2013) demonstrated a decrease in learning for block 2 of the PLT for active stimulation. Conversely, our study exhibited learning across the sequential blocks of the PLT, irrespective of stimulation type, indicating reward learning had occurred in line with the PLT evidence-base (e.g. Pizzagalli, Iosifescu, et al., 2008; Pizzagalli et al., 2005; Lancaster et al., 2012; Barr et al., 2008).

We also found inconclusive evidence (BF = 0.737) for Duprat et al., (2016)'s main significant interaction (Time × Stimulation × Block with the added covariate of consummatory pleasure) and no support for the alternative hypothesis using our frequentist ANCOVA test. Duprat et al., (2016)'s main interaction of interest, found that iTBS increased reward responsiveness more for those participants with higher baseline hedonic capacity, as measured on the consummatory subscale of the TEPS (Gard et al., 2006). However, this finding is intriguing as theoretically we would have expected the anticipatory subscale of the TEPS to modulate the effects of iTBS, as this subscale measures reward responsiveness, whilst the consummatory subscale is linked to positive experiences. The significant interaction therefore calls into question how meaningful Duprat et al., (2016) 's effect was in terms of the evidence-base for TMS. Moreover, the authors themselves (Duprat et al., 2016) report that they did not correct for multiple comparisons (i.e. six ANCOVAs, which yielded one marginally significant effect), with a possible type I error. Nevertheless, we calculated a Bayesian sample size estimate of 30, necessary to find a BF of 0.20 akin to the significant interaction proposed by Duprat et al., (2016).

Given that both Ahn et al., (2013) and Duprat et al., (2016)'s main significant interactions were not supported theoretically, as explicated above, a manipulation check to see whether the task was working would have been useful. Indeed, a strength of our replication study was conducting a manipulation check to determine whether the PLT was

working as it should, which neither Ahn et al., (2013) nor Duprat et al., (2016)'s studies conducted. We demonstrated substantial evidence for the alternative hypothesis (BF = 3.307; BF = 3.054) with a t-distributed prior, scaled to the effects of Ahn et al., (2013) and Duprat et al., (2016) and a near significant difference with our frequentist test (p = 0.057), indicating that the RB had increased across the study, and that reinforcement learning had occurred.

4.2. Secondary Hypotheses: Mood Ratings

Our replication study found weak support towards the alternative hypothesis for positive mood (BF = 2.55), with PANAS PA scores decreasing, that is becoming less positive following active TMS. Similarly, our commensurate (exploratory) frequentist paired t-test provided a significant effect indicating a decrease in positive mood following active TMS. This finding was contrary to what we had predicted, that is, that positive mood would increase as a function of active TMS. Reduced positive affect on the PANAS has been linked to anhedonia in depressed patients (cf. Watson & Clark, 1984) and is indicative of reduced motivation and pleasure (e.g. Crawford & Henry, 2004). Similar to our finding, some early studies have reported decreased happiness in "healthy" controls, using self-report mood measures following TMS (cf. Pascual-Leone et al., 1996; George et al., 1996., Martin et al., 1997). Nevertheless, this finding was surprising, as we had expected active TMS to increase pleasure, as demonstrated for an increase in reward responsiveness on our behavioural PLT task, and contrary to our expectation active stimulation decreased positive mood.

It is possible that the discrepant findings between self-report ratings on positive mood, and the behavioural task could be related to demand characteristics, and social desirability effects (van der Mortel, 2008), that is participants might modulate their behaviour in line with the effects they believe the researcher is hoping to find. However, if

this was the case, we might have expected a self-reported increase in positive mood following active and sham stimulation, and indeed mood changes for negative affect. A more convincing argument could be related to length of testing sessions (~2 hrs), which could modulate a transient mood change for interest and pleasure, which should be tested empirically.

However, the discrepant result between an objective behavioural, PLT, task and self-reported mood ratings, notwithstanding whether they are measuring precisely the same construct of anhedonia, could have important implications for depressive treatment protocols, and clinical implications. Namely, if a participant's subjective view of behavioural change is incongruent with objective measures, such as the PLT, further treatments may be sought out, due to perceived distress (Bristow & Patten, 2002), which may be costly and potentially detrimental to the well-being of the individual. Given the discrepant findings between the subjective self-report PANAS PA rating and the enhanced reward responsiveness on the PLT, clinicians and policy-makers should consider using both clinical and statistical measures to ensure ecological validity of treatment effects (e.g. Button et al., 2015), as including one of these measures without the other could lead to misleading and incomplete findings.

Moreover, reduced positive mood following stimulation leads to a more fundamental question about whether TMS should be used as a treatment for depression, given the core premise of TMS treatment protocols for depression is to increase positive mood rather than decrease it. A possible future avenue for TMS protocols could be measuring individual differences in stress hormones, such as cortisol, which are thought to modulate the effectiveness of TMS in non-depressed healthy control participants (e.g. Baeken et al., 2014) and in depressed patients (Baeken et al., 2009). Using personalised TMS protocols, which take into account individual differences in neural connectivity and

variability (e.g. Singh et al., 2019; Caeyenberghs et al., 2018; Tik et al., 2017) could enhance the effectiveness of TMS treatments and indeed enhance reward sensitivity.

We also predicted that negative mood, as measured on the PANAS NA, would decrease as a function of TMS, that is participants would experience less negative mood following active rTMS. We found substantial support for the null hypothesis- no change- in our Bayesian analysis (BF = 0.048), commensurate with a non-significant result towards the alternative with our frequentist analysis. Negative mood did not change as a consequence of active or sham stimulation, which is in line with the evidence supporting no reduction in negative mood in healthy controls (e.g. Baeken et al., 2008; Moulier et al., 2016; Mousimann et al., 2000; Jenkins, Shajahan, Lappin, & Ebmeier, 2002). However, compared to other mood studies, a strength of our current study is our Bayesian approach, so we can indicate support for the null, as opposed to earlier studies that used frequentist statistics and inferred null results. In light of stimulation not decreasing negative mood, and active stimulation decreasing positive mood, we question the basic application of TMS to treat depression, as the mood change we would have expected are absent, or in the opposite direction.

4.3. Limitations

Due to the limitation on data collection, a possible limitation in our replication study may be the participant numbers that we have included are not sufficient to generate conclusive results for our primary statistic of interest. However, we did include a commensurate sample size to Ahn et al., (2013) and used the same frequentist statistics, which could call into question the basic effects that Ahn et al. (2013) reported (in addition to including an all-male sample that is not representative of depression prevalence). Therefore, our replication results must be treated with some caution, particularly the results

for our primary hypothesis and replications of interest due to the Bayes factors not meeting the required cut-off for substantial support for neither the null nor alternative hypothesis.

4.4. Conclusions

In sum, participants in our study demonstrated reward responsiveness akin to that found in the extant PLT literature. However, we failed to replicate Ahn et al. (2013) and Duprat et al. (2016)'s key findings that active stimulation would modulate reward responsiveness. Nevertheless, we demonstrated that positive affect was reduced, akin to anhedonia, following active stimulation, which is concerning for the use of TMS as a treatment for depression as TMS is thought reduce anhedonia. Our divergent findings between behavioural (PLT) and subjective self-report mood measures (decreased positive mood) highlight the need to link clinical and statistical measures. A lack of which could lead to an unclear and incomplete assessment of the efficacy of treatment effects for patients treated with TMS.

References

Ahn, H. M., Kim, S. E., & Kim, S. H. (2013). The effects of high-frequency rTMS over the left dorsolateral prefrontal cortex on reward responsiveness. *Brain Stimulation*, *6*(3), 310–314. https://doi.org/10.1016/j.brs.2012.05.013

Albert, P.R. (2015). Why is depression more prevalent in women? *Journal of Psychiatry and Neuroscience,40(4)*,219-221.https://doi.org/10.1503.jpn.150205

Aleman, A. (2013). Use of Repetitive Transcranial Magnetic Stimulation for Treatment in Psychiatry. *Clinical Psychopharmacology and Neuroscience*, *11*(2), 53–59. https://doi.org/10.9758/cpn.2013.11.2.53

Ali, S., Rhodes, L., Moreea, O., McMillan, D., Gilbody, S., Leach, C., … Delgadillo, J. (2017). How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. *Behaviour Research and Therapy*, *94*, 1–8. https://doi.org/10.1016/j.brat.2017.04.006

Allan, C., Kalu, U.-G., Sexton, C. E., & Ebmeier, K. P. (2012). Transcranial stimulation in depression. *The British Journal of Psychiatry*, *200*(1), 10–11. https://doi.org/10.1192/bjp.bp.111.101188

Allen, C., Singh, K. D., Verbruggen, F., & Chambers, C. D. (2018). Evidence for parallel activation of the pre-supplementary motor area and inferior frontal cortex during response inhibition: a combined MEG and TMS study. *Royal Society Open Science*, *5*(2), 171369. https://doi.org/10.1098/rsos.171369

Baeken, C., Leyman, L., De Raedt, R., Vanderhasselt, M.A., & D'haenen, H. (2008).Left and right high frequency repetitive transcranial magnetic stimulation of the dorsolateral prefrontal cortex does not affect mood in female volunteers. *Clinical Neurophysiology,119*, 568-575.https://doi.org.10.1016.j.clinph.2007.11.044

Baeken, C., De Raedt, R., Leyman, L., Schiettecatte, J., Kaufman, L., Poppe, K…Bossuyt, A.(2009). The impact of one HF-rTMS session on mood and salivary cortisol in treatment resistant unipolar melancholic depressed patients. *Journal of Affective Disorders,* 113, 100-108. https://doi.org.10.1016/j.jad.2008.05.008

Baeken, C., Vanderhasselt, M.A., Remue, J., Rossi, V., Schiettecatte,J…De Raedt,R. (2014). One left dorsolateral prefrontal cortical HF-rTMS session attenuates HPA-system sensitivity to critical feedback in healthy females. *Neuropsychologia, 57(1)*, 112-121.https://doi.org/10.1016/j.neuropsychologia.2014.02.019

Bakker, N., Shahab, S., Giacobbe, P., Blumberger, D. M., Daskalakis, Z. J., Kennedy, S. H., & Downar, J. (2015). rTMS of the dorsomedial prefrontal cortex for major depression: safety, tolerability, effectiveness, and outcome predictors for 10 Hz versus intermittent theta-burst stimulation. *Brain Stimulation*, *8*(2), 208–215. https://doi.org/10.1016/j.brs.2014.11.002

Ballard, I. C., Murty, V. P., Carter, R. M., MacInnes, J. J., Huettel, S. A., & Adcock, R. A. (2011). Dorsolateral Prefrontal Cortex Drives Mesolimbic Dopaminergic Regions to Initiate Motivated Behavior. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(28), 10340–10346. https://doi.org/10.1523/JNEUROSCI.0895-11.2011

Barr, R.S., Pizzagalli, D.A.,Culhane, M.A., Goff, D.C., & Evins, A.E. (2008).A single does of nicotine enhances reward responsiveness in nonsmokers: Implications for development of dependence. *Biological Psychiatry, 63(11)*, 1061-1065.https://doi.org/10.1016/j.biopsych.2007.09.015

Belujon, P., & Grace, A. A. (2017). Dopamine System Dysregulation in Major Depressive Disorders. *International Journal of Neuropsychopharmacology*, *20*(12), 1036–1046. https://doi.org/10.1093/ijnp/pyx056

Blumberger, D. M., Vila-Rodriguez, F., Thorpe, K. E., Feffer, K., Noda, Y., Giacobbe, P., … Downar, J. (2018). Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): a randomised non-inferiority trial. *Lancet (London, England)*, *391*(10131), 1683–1692. https://doi.org/10.1016/S0140-6736(18)30295-2

Boggio, P.S., Rocha, R.R., da Silva, M.T., & Fregni, F.(2008). Differential modulatory effects of transcranial direct current stimulation on a facial expression go-no-go task in males and females. *Neuroscience Letters,447*,101-105.https://doi.org.10;1016/j.neulet.2008.10.009

Bristow, K., & Patten, S. (2015). Treatment-seeking rates and associated mediating factors among individuals with depression. *Canadian Journal of Psychiatry, 47(7)*, 660-665.https://doi.org/10.1177/070674370204700708

Bulteau, S., Sébille, V., Fayet, G., Thomas-Ollivier, V., Deschamps, T., Bonnin-Rivalland, A., … Sauvaget, A. (2017). Efficacy of intermittent Theta Burst Stimulation (iTBS) and 10-Hz high-frequency repetitive transcranial magnetic stimulation (rTMS) in treatment-resistant unipolar depression: study protocol for a randomised controlled trial. *Trials*, *18*. https://doi.org/10.1186/s13063-016-1764-8

Button, K.S., Kounali, D., Thomas, L., Wiles, N.K., Peters, T.J.,…Lewis, G. (2015). Minimal clinically important difference on the Beck Depression Inventory-II according to the patient's perspective. *Psychological Medicine, 45*, 3269-3279. https://doi.org/10.1017/S0033291715001270

Byrne, K.A., & Worthy, D.A. (2015). Gender differences in reward sensitivity and information processing during decision-making. *Journal of Risk and Uncertainty*, 50, 55-71. https://doi.org/10.1007/s11166-015-9206-7

Caeyenberghs, K., Duprat, R., Leemans, A., Hosseini, H., Wilson, P.H.,…Baeken, C.(2018). Accelerated intermittent theta burst stimulation in major depression induced decreases in modularity: A connectome analysis. *Network Neuroscience, Advance Publication*. https://doi.org.10.1162/netn_a_00060

Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: realigning incentives in scientific publishing. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *66*, A1-2. https://doi.org/10.1016/j.cortex.2015.03.022

Chaves, C., Lopez-Gomez, I., Hervas, G., & Vazquez, C. (2017). A Comparative Study on the Efficacy of a Positive Psychology Intervention and a Cognitive Behavioral Therapy for Clinical Depression. *Cognitive Therapy and Research*, *41*(3), 417–433. https://doi.org/10.1007/s10608-016-9778-9

Chen, J., Zhou, C., Wu, B., Wang, Y., Li, Q., Wei, Y., … Xie, P. (2013). Left versus right repetitive transcranial magnetic stimulation in treating major depression: a meta-analysis of randomised controlled trials. *Psychiatry Research*, *210*(3), 1260–1264. https://doi.org/10.1016/j.psychres.2013.09.007

Chen, J.-J., Zhao, L.-B., Liu, Y.-Y., Fan, S.-H., & Xie, P. (2017). Comparative efficacy and acceptability of electroconvulsive therapy versus repetitive transcranial magnetic stimulation for major depression: A systematic review and multiple-treatments meta-analysis. *Behavioural Brain Research*, *320*, 30–36. https://doi.org/10.1016/j.bbr.2016.11.028

Chervyakov, A. V., Chernyavsky, A. Yu., Sinitsyn, D. O., & Piradov, M. A. (2015). Possible Mechanisms Underlying the Therapeutic Effects of Transcranial Magnetic Stimulation. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00303

Chevallier, C., Tonge, N., Safra, L., Kahn, D., Kohls, G., Miller, J., & Schultz, R. T. (2016). Measuring Social Motivation Using Signal Detection and Reward Responsiveness. *PLoS ONE*, *11*(12). https://doi.org/10.1371/journal.pone.0167024

Chung, S. W., Hoy, K. E., & Fitzgerald, P. B. (2015). Theta-burst stimulation: a new form of TMS treatment for depression? *Depression and Anxiety*, *32*(3), 182–192. https://doi.org/10.1002/da.22335

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., … Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*, *391*(10128), 1357–1366. https://doi.org/10.1016/S0140-6736(17)32802-7

Conelea, C. A., Philip, N. S., Yip, A. G., Barnes, J. L., Niedzwiecki, M. J., Greenberg, B. D., … Carpenter, L. L. (2017). Transcranial magnetic stimulation for treatment-resistant depression: Naturalistic treatment outcomes for younger versus older patients. *Journal of Affective Disorders*, *217*, 42–47. https://doi.org/10.1016/j.jad.2017.03.063

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *The British Journal of Clinical Psychology*, *43*(Pt 3), 245–265. https://doi.org/10.1348/0144665031752934

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, *58*(7), 376–385. https://doi.org/10.1177/070674371305800702

Der-Avakian, A., & Markou, A. (2012). The neurobiology of anhedonia and other reward-related deficits. *Trends in Neurosciences*, *35*(1), 68–77. https://doi.org/10.1016/j.tins.2011.11.005

Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Basingstoke: Palgrave Macmillan.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72,* 78-89.http://dx.doi.org/10.1016/j.jmp.2015.10.003

Dienes, Z., & Belfi, B. (2019). *Bayes Factor Calculator [R Source Code]*. Retrieved from http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bence%20Bayes%20factor%20calculator.html

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, *25*(1), 207–218. https://doi.org/10.3758/s13423-017-1266-z

Driessen, E., Hollon, S. D., Bockting, C. L. H., Cuijpers, P., & Turner, E. H. (2015). Does Publication Bias Inflate the Apparent Efficacy of Psychological Treatment for Major Depressive Disorder? A Systematic Review and Meta-Analysis of US National Institutes of Health-Funded Trials. *PLOS ONE*, *10*(9), e0137864. https://doi.org/10.1371/journal.pone.0137864

Duprat, R., De Raedt, R., Wu, G.-R., & Baeken, C. (2016). Intermittent Theta Burst Stimulation Increases Reward Responsiveness in Individuals with Higher Hedonic

Capacity. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00294

Duprat, R., Desmyter, S., Rudi, D. R., van Heeringen, K., Van den Abbeele, D., Tandt, H., … Baeken, C. (2016). Accelerated intermittent theta burst stimulation treatment in medication-resistant major depression: A fast road to remission? *Journal of Affective Disorders*, *200*(Specialized Interventions [3350]), 6–14. https://doi.org/10.1016/j.jad.2016.04.015

Duprat, R., Wu, G.-R., De Raedt, R., & Baeken, C. (2017). Accelerated iTBS treatment in depressed patients differentially modulates reward system activity based on anhedonia. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, 1–12. https://doi.org/10.1080/15622975.2017.1355472

Eche, J., Mondino, M., Haesebaert, F., Saoud, M., Poulet, E., & Brunelin, J. (2012). Low- vs High-Frequency Repetitive Transcranial Magnetic Stimulation as an Add-On Treatment for Refractory Depression. *Frontiers in Psychiatry*, *3*. https://doi.org/10.3389/fpsyt.2012.00013

FDA. (2008). *Neurostar TMS Therapy System.* Silver Spring, MD: US Food and Drug Administration.

Fidalgo, T. M., Morales-Quezada, J. L., Muzy, G. S. C., Chiavetta, N. M., Mendonca, M. E., Santana, M. V. B., … Fregni, F. (2014). Biological markers in noninvasive brain stimulation trials in major depressive disorder: a systematic review. *The Journal of ECT*, *30*(1), 47–61. https://doi.org/10.1097/YCT.0b013e31828b34d8

Fitzgerald, P. B., Hoy, K., Daskalakis, Z. J., & 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*(3), 229–234. https://doi.org/10.1002/da.20454

Fitzgerald, P. B., Hoy, K. E., Elliot, D., Susan McQueen, R. N., Wambeek, L. E., & Daskalakis, Z. J. (2018). Accelerated repetitive transcranial magnetic stimulation in the treatment of depression. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *43*(7), 1565–1572. https://doi.org/10.1038/s41386-018-0009-9

Fox, M. D., Liu, H., & Pascual-Leone, A. (2013). Identification of reproducible individualized targets for treatment of depression with TMS based on intrinsic connectivity. *NeuroImage*, *66*, 151–160. https://doi.org/10.1016/j.neuroimage.2012.10.082

Friborg, O., & Johnsen, T. J. (2017). The effect of cognitive-behavioral therapy as an antidepressive treatment is falling: Reply to Ljòtsson et al. (2017) and Cristea et al. (2017). *Psychological Bulletin*, *143*(3), 341–345. https://doi.org/10.1037/bul0000090

Gard, D. E., Gard, M. G., Kring, A. M., & John, O. P. (2006). Anticipatory and consummatory components of the experience of pleasure: A scale development study. *Journal of Research in Personality*, *40*(6), 1086–1102. https://doi.org/10.1016/j.jrp.2005.11.001

Gard, D. E., Kring, A. M., Gard, M. G., Horan, W. P., & Green, M. F. (2007). Anhedonia in Schizophrenia: Distinctions between Anticipatory and Consummatory Pleasure. *Schizophrenia Research*, *93*(1–3), 253–260. https://doi.org/10.1016/j.schres.2007.03.008

George, M.S., Wassermann, E.M., Williams, W.A., Steppel, J.,Pascual-Leone, A.,…Post, R.M. (1996). Changes in mood and hormone levels after rapid-rate transcranial

magnetic stimulation (rTMS) of the prefrontal cortex. *Journal of Neuropsychiatry, 8(2)*, 172-180.

George, M. S., Padberg, F., Schlaepfer, T. E., O'Reardon, J. P., Fitzgerald, P. B., Nahas, Z. H., & Marcolin, M. A. (2009). Controversy: Repetitive transcranial magnetic stimulation or transcranial direct current stimulation shows efficacy in treating psychiatric diseases (depression, mania, schizophrenia, obsessive-complusive disorder, panic, posttraumatic stress disorder). *Brain Stimulation*, *2*(1), 14–21. https://doi.org/10.1016/j.brs.2008.06.001

Grossheinrich, N., Rau, A., Pogarell, O., Hennig-Fast, K., Reinl, M., Karch, S., … Padberg, F. (2009). Theta burst stimulation of the prefrontal cortex: safety and impact on cognition, mood, and resting electroencephalogram. *Biological Psychiatry*, *65*(9), 778–784. https://doi.org/10.1016/j.biopsych.2008.10.029

Hayashi, T., Ko, J. H., Strafella, A. P., & Dagher, A. (2013). Dorsolateral prefrontal and orbitofrontal cortex interactions during self-control of cigarette craving. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(11), 4422–4427. https://doi.org/10.1073/pnas.1212185110

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*(7010), 859–862. https://doi.org/10.1038/nature02966

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206. https://doi.org/10.1016/j.neuron.2004.12.033

Huang, Y.-Z., Rothwell, J. C., Chen, R.-S., Lu, C.-S., & Chuang, W.-L. (2011). The theoretical model of theta burst form of repetitive transcranial magnetic stimulation. *Clinical Neurophysiology: Official Journal of the International Federation of*

Clinical Neurophysiology, *122*(5), 1011–1018. https://doi.org/10.1016/j.clinph.2010.08.016

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*, 12. https://doi.org/10.1186/2045-5380-3-12

Janicak, P. G., & Dokucu, M. E. (2015). Transcranial magnetic stimulation for the treatment of major depression. *Neuropsychiatric Disease and Treatment*, *11*, 1549–1560. https://doi.org/10.2147/NDT.S67477

Jeffreys, S. H. (1998). *The Theory of Probability* (Third Edition). Oxford, New York: Oxford University Press.

Jenkins, J., Shajahan, P.M., Lappin, J.M., & Ebmeier, K.P. (2002). Right and left prefrontal transcranial magnetic stimulation at 1 Hz does not affect mood in healthy volunteers. *BMC, 2:1*. https://doi.org/10.1186/1471-244X-2-1

Johnsen, T. J., & Friborg, O. (2015). The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. *Psychological Bulletin*, *141*(4), 747–768. https://doi.org/10.1037/bul0000015

Kessler, R. C. (2012). The costs of depression. *The Psychiatric Clinics of North America*, *35*(1), 1–14. https://doi.org/10.1016/j.psc.2011.11.005

Kirsch, I. (2014). Antidepressants and the Placebo Effect. *Zeitschrift Fur Psychologie*, *222*(3), 128–134. https://doi.org/10.1027/2151-2604/a000176

Knekt, P., Lindfors, O., Härkänen, T., Välikoski, M., Virtala, E., Laaksonen, M. A., … Helsinki Psychotherapy Study Group. (2008). Randomized trial on the effectiveness of long-and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. *Psychological Medicine*, *38*(5), 689–703. https://doi.org/10.1017/S003329170700164X

Koenigs, M., & Grafman, J. (2009). The functional neuroanatomy of depression: Distinct roles for ventromedial and dorsolateral prefrontal cortex. *Behavioural Brain Research*, *201*(2), 239–243. https://doi.org/10.1016/j.bbr.2009.03.004

Lancaster, T. M., Heerey, E. A., Mantripragada, K., & Linden, D. E. J. (2015). Replication study implicates COMT val158met polymorphism as a modulator of probabilistic reward learning. *Genes, Brain, and Behavior*, *14*(6), 486–492. https://doi.org/10.1111/gbb.12228

Lancaster, T. M., Linden, D. E., & Heerey, E. A. (2012). COMT val158met predicts reward responsiveness in humans. *Genes, Brain, and Behavior*, *11*(8), 986–992. https://doi.org/10.1111/j.1601-183X.2012.00838.x

Layard, R., & Clark, D., M. (2014). *Thrive. The power of psychological therapy.* London: Penguin.

Lefaucheur, J.-P., André-Obadia, N., Antal, A., Ayache, S. S., Baeken, C., Benninger, D. H., … Garcia-Larrea, L. (2014). Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rTMS). *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *125*(11), 2150–2206. https://doi.org/10.1016/j.clinph.2014.05.021

Li, C.-T., Chen, M.-H., Juan, C.-H., Huang, H.-H., Chen, L.-F., Hsieh, J.-C., … Su, T.-P. (2014). Efficacy of prefrontal theta-burst stimulation in refractory depression: a randomized sham-controlled study. *Brain: A Journal of Neurology*, *137*(Pt 7), 2088–2098. https://doi.org/10.1093/brain/awu109

Liu, B., Zhang, Y., Zhang, L., & Li, L. (2014). Repetitive transcranial magnetic stimulation as an augmentative strategy for treatment-resistant depression, a meta-analysis of randomized, double-blind and sham-controlled study. *BMC Psychiatry*, *14*, 342. https://doi.org/10.1186/s12888-014-0342-4

Loo, C. K., & Mitchell, P. B. (2005). A review of the efficacy of transcranial magnetic stimulation (TMS) treatment for depression, and current and future strategies to optimize efficacy. *Journal of Affective Disorders*, *88*(3), 255–267. https://doi.org/10.1016/j.jad.2005.08.001

Maizey, L., Allen, C. P. G., Dervinis, M., Verbruggen, F., Varnava, A., Kozlov, M., … Chambers, C. D. (2013). Comparative incidence rates of mild adverse effects to transcranial magnetic stimulation. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *124*(3), 536–544. https://doi.org/10.1016/j.clinph.2012.07.024

Martin, J.D., George, M.S., Greenberg, B.D., Wassermann, E.M., Schlaepfer, T.E.,…Post, R.M. (1997). Mood effects of prefrontal repetitive high-frequency TMS in healthy volunteers. *CNS Spectrums, 2(1)*, 53-68. https://doi.org/10.1017/S10928529000004491

Martin, J. L. R., Barbanoj, M. J., Schlaepfer, T. E., Thompson, E., Pérez, V., & Kulisevsky, J. (2003). Repetitive transcranial magnetic stimulation for the treatment of depression. Systematic review and meta-analysis. *The British Journal of Psychiatry: The Journal of Mental Science*, *182*, 480–491.

Matheson, N. A., Shemmell, J. B. H., De Ridder, D., & Reynolds, J. N. J. (2016). Understanding the Effects of Repetitive Transcranial Magnetic Stimulation on Neuronal Circuits. *Frontiers in Neural Circuits*, *10*. https://doi.org/10.3389/fncir.2016.00067

McCarthy, D., & Davison, M. (1979). Signal probability, reinforcement and signal detection. *Journal of the Experimental Analysis of Behavior*, *32*(3), 373–386. https://doi.org/10.1901/jeab.1979.32-373

Mosimann, U.P., Rihs, T.A., Engeler, J., Fisch, H-U., & Schlaepfer, T.E. (2000). Mood effects of repetitive transcranial magnetic stimulation of left prefrontal cortex in healthy volunteers. *Psychiatry Research, 94*, 251-256.

Moulier, V., Gaudeau-Bosma, C., Isaac, C., Allard, A.-C., Bouaziz, N., Sidhoumi, D., … Januel, D. (2016). Effect of repetitive transcranial magnetic stimulation on mood in healthy subjects. *Socioaffective Neuroscience & Psychology*, *6*. https://doi.org/10.3402/snp.v6.29672

NICE. (2009). *Depression in adults: the treatment and management of depression in adults. CG90.* London: National Institute for Health Care and Excellence.

NICE. (2015). *Repetitive transcranial magnetic stimulation for depression. IPG542.* National Institute for Health and Care Excellence.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science (New York, N.Y.)*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Ott, D. V. M., Ullsperger, M., Jocham, G., Neumann, J., & Klein, T. A. (2011). Continuous theta-burst stimulation (cTBS) over the lateral prefrontal cortex alters reinforcement learning bias. *NeuroImage*, *57*(2), 617–623. https://doi.org/10.1016/j.neuroimage.2011.04.038

Pascual-Leone., A., Catala, M.D., Pascual-Leone, A. (1996). Lateralized effect of rapid-rate transcranial magnetic stimulation of the prefrontal cortex on mood. *Neurology, 46,* 499-502.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pizzagalli, D. A., Goetz, E., Ostacher, M., Iosifescu, D. V., & Perlis, R. H. (2008). Euthymic patients with bipolar disorder show decreased reward learning in a probabilistic reward task. *Biological Psychiatry*, *64*(2), 162–168. https://doi.org/10.1016/j.biopsych.2007.12.001

Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008). Reduced Hedonic Capacity in Major Depressive Disorder: Evidence from a Probabilistic Reward Task. *Journal of Psychiatric Research*, *43*(1), 76–87. https://doi.org/10.1016/j.jpsychires.2008.03.001

Pizzagalli, D. A., Jahn, A. L., & O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biological Psychiatry*, *57*(4), 319–327. https://doi.org/10.1016/j.biopsych.2004.11.026

Pizzagalli, D. A., Sherwood, R. J., Henriques, J. B., & Davidson, R. J. (2005). Frontal brain asymmetry and reward responsiveness: a source-localization study. *Psychological Science*, *16*(10), 805–813. https://doi.org/10.1111/j.1467-9280.2005.01618.x

Sherdell, L., Waugh, C. E., & Gotlib, I. H. (2012). Anticipatory pleasure predicts motivation for reward in major depression. *Journal of Abnormal Psychology*, *121*(1), 51–60. https://doi.org/10.1037/a0024945

Siebner, H. R., & Rothwell, J. (2003). Transcranial magnetic stimulation: new insights into representational cortical plasticity. *Experimental Brain Research*, *148*(1), 1–16. https://doi.org/10.1007/s00221-002-1234-2

Singh, A., Erwin-Grabner, T., Sutcliffe, G., Antal, A., Paulus, W., & Goya-Maldonado, R. (2019). Personalized repetitive transcranial magnetic stimulation temporarily alters default mode network in healthy subjects. *Scientific Reports*, *9*(1), 5631. https://doi.org/10.1038/s41598-019-42067-3

Staudinger, M. R., Erk, S., & Walter, H. (2011). Dorsolateral prefrontal cortex modulates striatal reward encoding during reappraisal of reward anticipation. *Cerebral Cortex (New York, N.Y.: 1991)*, *21*(11), 2578–2588. https://doi.org/10.1093/cercor/bhr041

Strauss, G. P., Wilbur, R. C., Warren, K. R., August, S. M., & Gold, J. M. (2011). Anticipatory vs. consummatory pleasure: what is the nature of hedonic deficits in schizophrenia? *Psychiatry Research*, *187*(1–2), 36–41. https://doi.org/10.1016/j.psychres.2011.01.012

Tik, M., Hoffmann, A., Sladky, R., Tomova, L., Hummer, A.,…Windischberger, C. (2018). Towards understanding rTMS mechanism of action: Stimulation of the DLPFC causes network-specific increase in functional connectivity. *Neuroimage, 162*, 289-296. https://doi.org/10.1016/j.neuroimage.2017.09.022

van de Mortel, T.F. (2008). Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25(4)*, 40-48.

Wassermann, E. M., & Zimmermann, T. (2012). Transcranial magnetic brain stimulation: therapeutic promises and scientific gaps. *Pharmacology & Therapeutics*, *133*(1), 98–107. https://doi.org/10.1016/j.pharmthera.2011.09.003

Watson, D., & Clark, L.A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96(3)*, 465-490.https://doi.org.10.1037/0033-2909.96.3.465

Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*, *97*(3), 346–353.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

WHO. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* Geneva: World Health Organisation.

WHO. (2017). *Depression and other common mental disorders. Global health estimates.* Geneva: World Health Organisation;2017.

Appendix A- Submission Guidelines for Journal of Affective Disorders

Review Articles and Meta-analyses (up to 8000 words, excluding references and up to 10 tables/figures)

At the discretion of the accepting Editor-in-Chief, and/or based on reviewer feedback, authors may be allowed fewer or more than these guidelines.

Preparation of Manuscripts

Articles should be in English. The title page should appear as a separate sheet bearing title (without article type), author names and affiliations, and a footnote with the corresponding author's full contact information, including address, telephone and fax numbers, and e-mail address (failure to include an e-mail address can delay processing of the manuscript).

Papers should be divided into sections headed by a caption (e.g., Introduction, Methods, Results, Discussion). A structured abstract of no more than 250 words should appear on a separate page with the following headings and order: Background, Methods, Results, Limitations, Conclusions (which should contain a statement about the clinical relevance of the research). A list of three to six key words should appear under the abstract. **Authors should note that the 'limitations' section both in the discussion of the paper AND IN A STRUCTURED ABSTRACT are essential. Failure to include it may delay in processing the paper, decision making and final publication.**

Figures and Photographs

Figures and Photographs of good quality should be submitted online as a separate file. Please use a lettering that remains clearly readable even after reduction to about 66%. For every figure or photograph, a legend should be provided. All authors wishing to use illustrations already published must first obtain the permission of the author and publisher and/or copyright holders and give precise reference to the original work. This permission must include the right to publish in electronic media.

Tables

Tables should be numbered consecutively with Arabic numerals and must be cited in the text in sequence. Each table, with an appropriate brief legend, comprehensible without reference to the text, should be typed on a separate page and uploaded online. Tables should be kept as simple as possible and wherever possible a graphical representation used instead. Table titles should be complete but brief. Information other than that defining the data should be presented as footnotes.

Please refer to the generic Elsevier artwork instructions: http://authors.elsevier.com/artwork/jad.

Preparation of supplementary data

Elsevier accepts electronic supplementary material to support and enhance your scientific research. Supplementary files offer the author additional possibilities to publish supporting

applications, movies, animation sequences, high-resolution images, background datasets, sound clips and more.

Supplementary files supplied will be published online alongside the electronic version of your article in Elsevier web products, including ScienceDirect: http://www.sciencedirect.com. In order to ensure that your submitted material is directly usable, please ensure that data is provided in one of our recommended file formats. Authors should submit the material in electronic format together with the article and supply a concise and descriptive caption for each file. For more detailed instructions please visit our Author Gateway at: https://www.elsevier.com/authors.

Colour reproduction

The Journal of Affective Disorders is now also included in a new initiative from Elsevier: 'Colourful e-Products'. Through this initiative, figures that appear in black & white in print can appear in colour, online, in ScienceDirect at http://www.sciencedirect.com.

There is no extra charge for authors who participate.

For colour reproduction in print, you will receive information regarding the costs from Elsevier after receipt of your accepted article. Please indicate your preference for colour in print or on the Web only. Because of technical complications which can arise by converting colour figures to "grey scale" (for the printed version should you not opt for colour in print) please submit in addition usable black and white versions of all the colour illustrations. For further information on the preparation of electronic artwork, please see http://authors.elsevier.com/artwork/jad.

Peer review

This journal operates a single blind review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final. More information on types of peer review.

Use of word processing software

It is important that the file be saved in the native format of the word processor used. The text should be in single-column format. Keep the layout of the text as simple as possible. Most formatting codes will be removed and replaced on processing the article. In particular, do not use the word processor's options to justify text or to hyphenate words. However, do use bold face, italics, subscripts, superscripts etc. When preparing tables, if you are using a table grid, use only one grid for each individual table and not a grid for each row. If no grid is used, use tabs, not spaces, to align columns. The electronic text should be prepared in a way very similar to that of conventional manuscripts (see also the Guide to Publishing with Elsevier). Note that source files of figures, tables and text graphics will be required whether or not you embed your figures in the text. See also the section on Electronic artwork.
To avoid unnecessary errors you are strongly advised to use the 'spell-check' and 'grammar-check' functions of your word processor.

Highlights

Highlights are mandatory for this journal as they help increase the discoverability of your article via search engines. They consist of a short collection of bullet points that capture the novel results of your research as well as new methods that were used during the study (if any). Please have a look at the examples here: example Highlights.

Highlights should be submitted in a separate editable file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point).

Highlights offer your paper a considerable advantage in the online world, as they ensure that search engines pick up your article and match it to the right audience. (Nowadays, machines read your work just as often as humans do!). Highlights have been proven to widen the reach of your work and help to ensure that your article is brought to the attention of interested colleagues, both inside and outside your usual research community. Apart from a wider distribution of your research, we hope that this will also lead to new collaborations and help accelerate the pace of science.

The small print

- Not part of editorial consideration and aren't required until the final files stage
- Only required for full research articles
- Must be provided as a Word document— select "Highlights" from the drop-down list when uploading files
- Each Highlight can be no more than 85 characters, including spaces
- No jargon, acronyms, or abbreviations: aim for a general audience and use keywords
- Consider the reader - Highlights are the first thing they'll see

Examples

From: *Cancer Cell, Volume 32, Issue 2, 14 August 2017, Pages 169-184.e7*

Abstract
A concise and factual abstract is required. The abstract should state briefly the purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided, but if essential, then cite the author(s) and year(s). Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself.

Keywords
Immediately after the abstract, provide a maximum of 6 keywords, using American spelling and avoiding general and plural terms and multiple concepts (avoid, for example, 'and', 'of'). Be sparing with abbreviations: only abbreviations firmly established in the field may be eligible. These keywords will be used for indexing purposes.

Abbreviations
Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at

their first mention there, as well as in the footnote. Ensure consistency of abbreviations throughout the article.

Acknowledgements
Collate acknowledgements in a separate section at the end of the article before the references and do not, therefore, include them on the title page, as a footnote to the title or otherwise. List here those individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.).

Formatting of funding sources
List funding sources in this standard way to facilitate compliance to funder's requirements:

Funding: This work was supported by the National Institutes of Health [grant numbers xxxx, yyyy]; the Bill & Melinda Gates Foundation, Seattle, WA [grant number zzzz]; and the United States Institutes of Peace [grant number aaaa].

It is not necessary to include detailed descriptions on the program or type of grants and awards. When funding is from a block grant or other resources available to a university, college, or other research institution, submit the name of the institute or organization that provided the funding.

If no funding has been provided for the research, please include the following sentence:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Nomenclature and units
Follow internationally accepted rules and conventions: use the international system of units (SI). If other quantities are mentioned, give their equivalent in SI. You are urged to consult IUPAC: Nomenclature of Organic Chemistry for further information.

Math formulae
Please submit math equations as editable text and not as images. Present simple formulae in line with normal text where possible and use the solidus (/) instead of a horizontal line for small fractional terms, e.g., X/Y. In principle, variables are to be presented in italics. Powers of e are often more conveniently denoted by exp. Number consecutively any equations that have to be displayed separately from the text (if referred to explicitly in the text).

Footnotes
Footnotes should be used sparingly. Number them consecutively throughout the article. Many word processors can build footnotes into the text, and this feature may be used. Otherwise, please indicate the position of footnotes in the text and list the footnotes themselves separately at the end of the article. Do not include footnotes in the Reference list.

Artwork

Electronic artwork
General points
• Make sure you use uniform lettering and sizing of your original artwork.
• Embed the used fonts if the application provides that option.

• Aim to use the following fonts in your illustrations: Arial, Courier, Times New Roman, Symbol, or use fonts that look similar.
• Number the illustrations according to their sequence in the text.
• Use a logical naming convention for your artwork files.
• Provide captions to illustrations separately.
• Size the illustrations close to the desired dimensions of the published version.
• Submit each illustration as a separate file.
• Ensure that color images are accessible to all, including those with impaired color vision.

A detailed guide on electronic artwork is available.
You are urged to visit this site; some excerpts from the detailed information are given here.
Formats
If your electronic artwork is created in a Microsoft Office application (Word, PowerPoint, Excel) then please supply 'as is' in the native document format.
Regardless of the application used other than Microsoft Office, when your electronic artwork is finalized, please 'Save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):
EPS (or PDF): Vector drawings, embed all used fonts.
TIFF (or JPEG): Color or grayscale photographs (halftones), keep to a minimum of 300 dpi.
TIFF (or JPEG): Bitmapped (pure black & white pixels) line drawings, keep to a minimum of 1000 dpi.
TIFF (or JPEG): Combinations bitmapped line/half-tone (color or grayscale), keep to a minimum of 500 dpi.
Please do not:
• Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); these typically have a low number of pixels and limited set of colors;
• Supply files that are too low in resolution;
• Submit graphics that are disproportionately large for the content.

Color artwork
Please make sure that artwork files are in an acceptable format (TIFF (or JPEG), EPS (or PDF), or MS Office files) and with the correct resolution. If, together with your accepted article, you submit usable color figures then Elsevier will ensure, at no additional charge, that these figures will appear in color online (e.g., ScienceDirect and other sites) regardless of whether or not these illustrations are reproduced in color in the printed version. **For color reproduction in print, you will receive information regarding the costs from Elsevier after receipt of your accepted article**. Please indicate your preference for color: in print or online only. Further information on the preparation of electronic artwork.

Illustration services
Elsevier's Author Services offers Illustration Services to authors preparing to submit a manuscript but concerned about the quality of the images accompanying their article. Elsevier's expert illustrators can produce scientific, technical and medical-style images, as well as a full range of charts, tables and graphs. Image 'polishing' is also available, where our illustrators take your image(s) and improve them to a professional standard. Please visit the website to find out more.

Tables

Please submit tables as editable text and not as images. Tables can be placed either next to the relevant text in the article, or on separate page(s) at the end. Number tables consecutively in accordance with their appearance in the text and place any table notes below the table body. Be sparing in the use of tables and ensure that the data presented in them do not duplicate results described elsewhere in the article. Please avoid using vertical rules and shading in table cells.

References

Citation in text

Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

Data references

This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

Reference management software

Most Elsevier journals have their reference template available in many of the most popular reference management software products. These include all products that support Citation Style Language styles, such as Mendeley. Using citation plug-ins from these products, authors only need to select the appropriate journal template when preparing their article, after which citations and bibliographies will be automatically formatted in the journal's style. If no template is yet available for this journal, please follow the format of the sample references and citations as shown in this Guide. If you use reference management software, please ensure that you remove all field codes before submitting the electronic manuscript. More information on how to remove field codes from different reference management software.

Users of Mendeley Desktop can easily install the reference style for this journal by clicking the following link:
http://open.mendeley.com/use-citation-style/journal-of-affective-disorders
When preparing your manuscript, you will then be able to select this style using the Mendeley plug-ins for Microsoft Word or LibreOffice.

Reference style

Text: All citations in the text should refer to:
1. *Single author:* the author's name (without initials, unless there is ambiguity) and the year of publication;
2. *Two authors:* both authors' names and the year of publication;

3. *Three or more authors:* first author's name followed by 'et al.' and the year of publication. Citations may be made directly (or parenthetically). Groups of references can be listed either first alphabetically, then chronologically, or vice versa.
Examples: 'as demonstrated (Allan, 2000a, 2000b, 1999; Allan and Jones, 1999)…. Or, as demonstrated (Jones, 1999; Allan, 2000)… Kramer et al. (2010) have recently shown …'
List: References should be arranged first alphabetically and then further sorted chronologically if necessary. More than one reference from the same author(s) in the same year must be identified by the letters 'a', 'b', 'c', etc., placed after the year of publication.
Examples:
Reference to a journal publication:
Van der Geer, J., Hanraads, J.A.J., Lupton, R.A., 2010. The art of writing a scientific article. J. Sci. Commun. 163, 51–59. https://doi.org/10.1016/j.Sc.2010.00372.
Reference to a journal publication with an article number:
Van der Geer, J., Hanraads, J.A.J., Lupton, R.A., 2018. The art of writing a scientific article. Heliyon. 19, e00205. https://doi.org/10.1016/j.heliyon.2018.e00205.
Reference to a book:
Strunk Jr., W., White, E.B., 2000. The Elements of Style, fourth ed. Longman, New York.
Reference to a chapter in an edited book:
Mettam, G.R., Adams, L.B., 2009. How to prepare an electronic version of your article, in: Jones, B.S., Smith , R.Z. (Eds.), Introduction to the Electronic Age. E-Publishing Inc., New York, pp. 281–304.
Reference to a website:
Cancer Research UK, 1975. Cancer statistics reports for the UK. http://www.cancerresearchuk.org/aboutcancer/statistics/cancerstatsreport/ (accessed 13 March 2003).
Reference to a dataset:
[dataset] Oguro, M., Imahiro, S., Saito, S., Nakashizuka, T., 2015. Mortality data for Japanese oak wilt disease and surrounding forest compositions. Mendeley Data, v1. https://doi.org/10.17632/xwj98nb39r.1.

Supplementary material
Supplementary material such as applications, images and sound clips, can be published with your article to enhance it. Submitted supplementary items are published exactly as they are received (Excel or PowerPoint files will appear as such online). Please submit your material together with the article and supply a concise, descriptive caption for each supplementary file. If you wish to make changes to supplementary material during any stage of the process, please make sure to provide an updated file. Do not annotate any corrections on a previous version. Please switch off the 'Track Changes' option in Microsoft Office files as these will appear in the published version.

Appendix B- Searches conducted for our 4 databases: PsycINFO, Medline, Embase, CENTRAL

PsycINFO search terms (1806 to 27th November 2019).

1. neuromodulat*.mp (5215) [could not map onto subject heading with truncation].

2. transcranial magnetic stimulation.mp (10514) [fewer papers with TMS when using the subject heading]

3. exp Brain Stimulation.mp (22271) (multi-purpose generated fewer papers)

4. tms.mp. (5090) [no subject heading for this term].

5. intermittent theta burst stimulation.mp (149) [no subject heading for this term].

6. itbs.mp (369) [no subject heading for this term].

7. repetitive tms.mp (329) [no subject heading for this term].

8. rtms.mp (2946) [no subject heading for this term].

9. continuous theta burst stimulation.mp (252)

10. ctbs (369)

11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 (29205)

12. major depress* .mp(130534)

13. depress* (357180)

14. exp treatment resistant depression/ or treatment resistant depression.mp (2924) [explode subject heading or .mp keyword]

15. 12 or 13 or 14 (357180)

16. 11 and 15 (5147)

17. Dorsolateral Prefrontal Cortex.mp (5861) [far fewer papers when using the subject heading]

18. DLPFC.mp (3001)

19. exp Prefrontal cortex/ or prefrontal cortex.mp [36789]

20. 17 or 18 or 19 (36980)

21. 16 and 20(1056)

22. limit 21 to (full text; peer reviewed journal; adulthood <18+ years> and English and human)

729 papers/ abstracts

mp = [multi-purpose; mp =title, abstract, heading word, table of contents, key concepts, original title, tests & measures, mesh]*= truncation; any variation of the suffix at the end of the word. Generated further results.

Medline search terms (Ovid MEDLINE ® ALL. 1946 to November 27, 2019).

1. neuromodulat* [13631]

2. transcranial magnetic stimulation.mp [13910] [when using exp Transcranial Magnetic Stimulation/ fewer papers are found; 10729].

3. brain stimulation.mp [13957]

[no subject heading- searched as keyword and .mp]

4. tms.mp. [9555]

5. intermittent theta burst stimulation.mp [161]

6. itbs.mp [277]

7. repetitive tms.mp [423]

8. rtms.mp [3471]

9. continuous theta burst stimulation.mp [272] [no subject heading for this]

10. ctbs.mp [434]

11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 [42678]

12. major depress*.mp [39662] [with major depressive disorder exp subject heading, 28749]

13. depress*.mp [496502] [with depression exp as subject heading, 112242]

14. treatment resistant depression.mp [1741] [with depressive disorder, treatment-resistant/ 1119]

15. 12 or 13 or 14 [469502]

16. 11 and 15 [4331]

17. Dorsolateral Prefrontal Cortex.mp [6089]

18. DLPFC.mp [2958]

19. Prefrontal Cortex.mp or Prefrontal Cortex/[41641]

20. 17 or 18 or 19[41775]

21. 16 and 20 [830]

22. limit 20 to (adulthood <19+ years> and English and human) [357]

713 abstracts

.mp in this case covers .ti, .ab, .ot, .nm, .hw, .fx, .kf, .ox, .px, .rx, .ui, .sy

(title, abstract, title, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, organism supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms).

Embase search terms (1947 to present (November 27, 2019).

1. neuromodulat*.mp [52528] [mapping keyword/term to subject heading but fewer papers with this].

2. transcranial magnetic stimulation.mp [25077]

3. brain stimulation.mp [25142]

4. tms.mp. [17556]

5. intermittent theta burst stimulation.mp [425]

6. itbs.mp [711]

7. exp repetitive tms/ [2406]

8. rtms.mp [6953]

9. continuous theta burst stimulation.mp [500]

10. ctbs [887]

11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 [103189]

12. major depress*.mp [83309] [when exp major depression/ generated fewer than the .mp search; 61775].

13. depress* [802677][when exp major depression/ generated fewer than the .mp search; 466055].

14. treatment resistant depression.mp [4406] [when exp treatment resistant depression/ generated fewer than the .mp search; 2826].

15. 12 or 13 or 14

16. 11 and 15 [12981]

17. Dorsolateral Prefrontal Cortex.mp [12266] [when exp Dorsolateral Prefrontal Cortex/ generated fewer than the .mp search; 5520].

18. DLPFC.mp [6164]

19. Prefrontal cortex [73153] [when exp. Prefrontal cortex: 72774]

20. 17 or 18 or 19 [73684]

21. 16 and 20 [2493]

22. limit 20 to (adulthood <18 to 64 years> and English and human)

1090 papers/ abstracts (minus 5 duplicates in Embase itself) = 1085.

.mp in this case covers .ti, .ab, .hw, .tn, .ot., .dm, .mf, .dv, .kw, .fx, .dq(title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate word term).

CENTRAL Cochrane database 27th of November, 2019

1. Neuromodulat* [MESH; 2298]
2. transcranial magnetic stimulation [4749]
3. brain stimulation [9040]
4. tms [4096]
5. intermittent theta burst stimulation [239]
6. itbs [228]
7. repetitive tms [2303]
8. rtms [2580]
9. continuous theta burst stimulation [196]
10. ctbs [171]
11. #1 or #2 or #3 or #4 or #5 or #6 or #7 or # 8 or #9 or #10 [13661]
12. major depress* [original search with "major depression" is 26292; now- 29703]
13. depress* [98915]
14. treatment resistant depression [2131]
15. #12 or #13 or #14 [98915]
16. #11 and #15 [2670]
17. Dorsolateral Prefrontal Cortex [2055]
18. DLPFC [1258]
19. Prefrontal cortex [4399]
20. #17 or #18 or #19 [4538]
21. #16 and #20 [880]

21= systematic reviews; 7 = protocols

880 abstracts/ papers (minus 21 systematic reviews and minus 7 protocols in CENTRAL itself) = 852.

3379 papers in total

Appendix C - Screening Form designed for study selection (modelled on that of Boland et al., 2017): systematic review and meta-analysis.

Review question: Do neuromodulatory treatments really reduce depressive symptomatology?

Inclusion Criteria:-

Population: Adults over the age of 18, who have received a diagnosis of Treatment Resistant Depression (TRD) or Major Depressive Disorder (MDD).

Intervention: Facilitatory (>1Hz) or Inhibitory (<1Hz) TMS protocols; Intermittent theta burst; combination of inhibitory and facilitatory treatment protocols applied to the left and/or right dorsolateral prefrontal cortex (DLPFC)

Comparator: Double-blinded sham-controlled trials, with sham TMS compared to active TMS.

Outcomes: Questionnaires measuring depressive symptom change.

Study Design: Double-blinded randomised controlled trials.

Screening and Selection Tool for TMS treatment protocols with sham comparator		
Reviewer Name		**Date:**
Author Name/ DOI		**Year:**
Title		**Journal:**
Patient Population	**Include**	**Exclude**
	Adults >18 who have received a diagnosis of Treatment Resistant Depression (TRD) or Major Depressive Disorder (MDD).	People who are <18. Adults who are 18+ who have been diagnosed with other conditions than MDD or TRD, including co-morbid conditions (e.g. anxiety + depression) or bipolar disorder (including the depressive phase).
Interventions	**Include**	**Exclude**
	Facilitatory TMS protocols Inhibitory TMS protocols Combination of inhibitory + facilitatory protocols Intermittent Theta Burst Protocols.	Other brain stimulation- i.e. transcranial direct current stimulation; deep brain stimulation; electroconvulsive therapy; vagal nerve stimulation. Adjunctive Psychotropic Treatments.

	Stimulation to the left and/or right Dorsolateral Prefrontal Cortex	
Comparators	**Include**	**Exclude**
	Sham stimulation as a comparator condition to active TMS.	Studies without a sham comparator, or which do not involve brain stimulation protocols as specified in the intervention section.
Outcome	**Include**	**Exclude**
	Questionnaires measuring depressive symptom change.	No questionnaires measuring depressive symptom change.
Study Design	**Include**	**Exclude**
	Randomised Controlled Trials (Double-Blind)	Any other study design apart from RCTs
Overall decision	**Included**	**Excluded**

Appendix D- Similar to Higgins & Green (2008) & Mutz et al., (2019)'s criteria, Risk of Bias (RoB) was assessed for our included studies against the criteria below.

Random sequence generation:

Low RoB: Randomisation was clearly specified (coin "toss", random number generator, throwing dice).
Unclear RoB: Unclear method of randomisation- not described clearly.
High RoB: Arbitrary randomisation process that could lead , e.g. to uneven groups, or other systematic bias.

Allocation concealment:

Low RoB: assignment concealed from personnel and participant through means such as e.g. concealed envelopes, central allocation, computer database.
Unclear RoB: no concealment method provided
High RoB: allocation not concealed

Blinding of participants and personnel:

Low RoB: Appropriate blinding of key personnel, and patients. If in the event of inefficient blind for key personnel or patient, outcome assessment must be blinded, outcome assessment must be blinded, and no compromise to blinding integrity.
Unclear RoB: No blinding information provided
High RoB: Blind broken, or outcome measure will be affected through insufficient blind.

Blinding of outcome assessment:

Low RoB: Raters Blind to treatment conditions, and allocation concealed.
Unclear RoB: Insufficient information to determine whether raters blinded.
High RoB: Raters not blinded.

Incomplete outcome data:

Low RoB: No missing data, but if this occurs, valid explanation for why this is the case (e.g. participant attrition, with valid reasons).
Unclear RoB: Unclear reasons for missing data/ participant attrition.
High RoB: High drop-out rate, and very unbalanced treatment groups.

Selective outcome reporting

Low RoB: The "primary" and "secondary" outcome measures that were pre-specified have been reported clearly.
Unclear RoB: Only "primary" outcome measures were reported.
High RoB: Missing "primary" outcome measure.

Overall bias:

Low RoB: Low RoB for all domains, or 1 "Unclear" RoB.
Unclear RoB: Two or more domains with "Unclear" RoB.
High RoB: If any one domain has a rating of "High" RoB.

Appendix E. RoB: Individual ratings with author judgement (concordant with excellent inter-rater reliability, 94 %, CI: .90- .97).

Baeken, 2013

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomly assigned to active or sham condition for the first week- using a coin toss.
Allocation concealment (selection bias)	Unclear risk	No details provided as to how patients were initially selected for the study, no screening information.
Blinding of participants and personnel (performance bias)	Unclear risk	Patients blindfolded and wore ear plugs. Sham coil was identical to the active coil- held at 90 degrees (experimenters not blind to treatment condition). “we cannot exclude that our procedure resulted in a lower blinding success”
Blinding of outcome assessment (detection bias)	Low risk	Independent rater- psychiatrist administered the 17-item HDRS "before and after" rTMS treatment
Incomplete outcome data (attrition bias)	Low risk	1 patient dropped out due to no-improvement after the first week. 11 patients allocated to sham; 9 to active.
Selective reporting (reporting bias)	Low risk	35% of patients defined as "clinical responders", No data for the active and sham conditions, separately, so difficult to determine whether data was different between conditions, and whether some patients may have been "responders" in the sham condition. However, no interaction

		between A/S suggests a similar performance for both type/s of stimulation.
Overall Bias	Unclear risk	at least two domains with unclear RoB

Duprat 2016

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"patients were randomized (flipping a coin) to receive in the first week either real or sham iTBS" (p. 9)
Allocation concealment (selection bias)	Unclear risk	No details provided as to how patients were initially selected for the study, no screening information
Blinding of participants and personnel (performance bias)	Low risk	"patients were blindfolded, wore earplugs and were kept unaware of the type of stimulation they received" -how? (p.9) No information as to how the experimenters were blinded to treatment conditions, in particular with swapping coils.
Blinding of outcome assessment (detection bias)	Low risk	17-item HDRS used to assess depressive symptoms each week, by a psychiatrist, blinded to the treatment conditions. What about the other clinical assessments, was the rater blind to the treatment conditions?
Incomplete outcome data (attrition bias)	Low risk	3 people excluded from the dataset with appropriate reasons (suicide attempt, administration of active stimulation x 2, and spontaneous improvement).

Selective reporting (reporting bias)	Low risk	All outcome measures reported were accounted for (HDRS, BDI, VAS). However, HDRS data separated into responders/non-responders so unable to determine differences between sham and active treatment conditions. However, no interaction for order (i.e. Active/ Sham) in the ANOVA produces some evidence that A/S have similar effects.
Overall Bias	Low risk	1 domain unclear, or all Low risk

George 2010

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Study describes that 199 patients were randomised, after screening 860 patients. Patients randomised to active or sham conditions using permuted blocks by site and treatment resistance.
Allocation concealment (selection bias)	Unclear risk	"Patients recruited using media advertisments, and physician referrals" "physicians telephone screened potential participants- using inclusion and exclusion criteria". Unclear how allocation concealment was maintained, although sounds promising. Screened 860 to randomise 199, with the final sample consisting of 190 patients- selection bias? How were these patients allocated?
Blinding of participants and personnel (performance bias)	Low risk	The sham coil mimicked the somatosensory experience of the active rTMS- as this was an active coil, which would aid with patients and experimenters

		not knowing the treatment condition assignment. Also masked rTMS administrators.
Blinding of outcome assessment (detection bias)	Low risk	The paper reported continuous assessment of evaluator reliability compared with a masked evaluator- training was also provided for the evaluators to mitigate any potential unmasking. Data managed via coordination unit, analyses conducted via independent statisticians and cross-checked by coordination unit.
Incomplete outcome data (attrition bias)	Low risk	190 patients included in the final analyses due to 9 being excluded (7 due to sham procedure being piloted, and two patients did not take part in the study). Incomplete reporting of pilot testing- and number of patients included in this, although likely not to affect main hypotheses. All other participants from the 199 accounted for (860 initially screened, but 601 excluded, discussed in selection bias section).
Selective reporting (reporting bias)	Low risk	All analyses seemed appropriate to assess the hypotheses delineated. Power analyses for an odds ratio of >2, and 80% power , with a sample size of 240. However, sample size was 190, with an odds ratio of 4.2- so powered sufficiently.
Overall Bias	Low risk	1 domain unclear, or all domains low risk

Holtzheimer 2004

Risk of bias table

Bias	Authors' judgement	Support for judgement

Random sequence generation (selection bias)	Unclear risk	The authors reported that patients were randomly allocated to receive active and sham stimulation. However, no further details about how patients were randomised into the groups.
Allocation concealment (selection bias)	Unclear risk	Referrals from doctors, ECT centres, advertisments.
Blinding of participants and personnel (performance bias)	Unclear risk	Unsure- no information provided about how participants or experimenters were blinded. Unlikely to have been kept blinded, as active and sham coil were the same (just rotated for sham stimulation-sensation would be different for the two sessions).
Blinding of outcome assessment (detection bias)	Low risk	Raters were blinded to the treatment groups, administered the HDRS, and self-reported BDI. Blinded to treatment groups.
Incomplete outcome data (attrition bias)	Low risk	Initially 15 patients included in the treatment (n =7 for active; n= 8 for sham). However, 3 patients excluded from further analyses for the neuropsychological portion of the results. The study is likely to be under-powered to find any true differences between active and sham conditions.
Selective reporting (reporting bias)	Low risk	Authors report that 3 patients are excluded from analyses (so n= 12 in final analyses). However, it is unclear what data is included in both clinical outcomes and neuropsychological tests. Discrepancy between sample size reported (n =12), and data for n =15 patients included in tables (1 and 2).
Overall Bias	Unclear risk	at least 2 domains with unclear risk of bias

Li 2016

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The study is described as randomised, with n=12 patients being allocated to 3 treatment groups. However, no further details are provided either in the paper or supplementary material, as to how patients are randomised.
Allocation concealment (selection bias)	Unclear risk	No information about allocation concealment in either the paper or supplementary material.
Blinding of participants and personnel (performance bias)	Unclear risk	No details provided.
Blinding of outcome assessment (detection bias)	Low risk	Blinded raters (psychiatrist) assessed efficacy outcome measures- blind to randomisation and study design.
Incomplete outcome data (attrition bias)	Low risk	53 participants assessed for eligibility with 17 excluded (with reasons)- 12 patients allocated to each of the three treatment arms.
Selective reporting (reporting bias)	Low risk	Data analyses appropriate for hypotheses (influence of frontalΘ on clinical outcomes: questionnaires).
Overall Bias	Unclear risk	at least two domains with unclear risk of bias

O'Reardon 2007

Risk of bias table

Bias	Authors' judgement	Support for judgement

Random sequence generation (selection bias)	Unclear risk	"patients randomized 1:1 to either receive active TMS or sham TMS" "Tapered in a blinded manner in six sessions across 3 weeks" (p.1209). How?
Allocation concealment (selection bias)	Unclear risk	No information provided as to how patients were allocated to groups, and how this information was concealed from experimenters and patients.
Blinding of participants and personnel (performance bias)	Low risk	From the larger dataset (O'Reardon et al., 2007) raters were blinded to outcome measures, and were not permitted to access treatment sessions. Quality control conducted for outcome assessors.
Blinding of outcome assessment (detection bias)	Low risk	"All efficacy outcome measures were assessed by blinded raters, who were not permitted access to the treatment sessions" (p 1207) - reliability and quality control of the raters was monitored through video monitoring. Patients asked to not disclose any details of their treatment to raters. Raters, and other personnel unaware of primary efficacy measures.
Incomplete outcome data (attrition bias)	Low risk	325 patients were screened and 24 were excluded with reasons, so final sample of 301 patients included in data analyses.
Selective reporting (reporting bias)	Low risk	Data analyses and reporting appear to be appropriate to study aim: efficacy of rTMS compared to sham. Data included medication free patients in the double-blinded RCT portion of the study.
Overall Bias	Unclear risk	at least 2 domains unclear RoB

Stern 2007

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The authors describe that patients were randomised into groups, with the active and sham coils being identical, and only receiving stimulation from one machine. No detailed provided though as to how participants were randomised into respective groups.
Allocation concealment (selection bias)	Unclear risk	No data provided as to how patients were allocated to treatment groups. "patients were interviewed by a psychiatrist, who was able to consult with their treating psychiatrist" (p.180).
Blinding of participants and personnel (performance bias)	Low risk	"patients and TMS technicians wore earplugs to prevent auditory threshold shifts" (p.181). However, sham coil was orientated perpendicular, which could alert the experimenter to the fact that they were delivering sham stimulation (experimenter not adequately blinded). But outcome measures conducted by blinded personnel.
Blinding of outcome assessment (detection bias)	Low risk	"psychiatrist blinded to group assignment conducted all assessments of patients' symptoms (p.181).
Incomplete outcome data (attrition bias)	Low risk	All patients included in the study reported. Data provided for each outcome measure.
Selective reporting (reporting bias)	Low risk	All outcome measures reported appropriately.
Overall Bias	Unclear risk	at least 2 unclear RoB

Vanderhasselt 2009a

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Unsure whether randomisation was upheld from the larger trial (where a coin was flipped to determine first stimulation type; e.g. Baeken et al., 2017).
Allocation concealment (selection bias)	Unclear risk	Unsure as to how this was conducted.
Blinding of participants and personnel (performance bias)	Unclear risk	"subjects were fully aware that one of the sessions was placebo" (p. 38). Same coil used for active and sham stimulation conditions, but coil held at 90˚ for sham, which the experimenter would be aware was sham. Also, sensation on scalp may feel different for patient/s, so unclear re: integrity of blind. Participants were blindfolded, and wore earplugs.
Blinding of outcome assessment (detection bias)	Unclear risk	The authors did not report who conducted the outcome measures, and whether integrity of the blind was upheld.
Incomplete outcome data (attrition bias)	Low risk	Out of the 16 patients initially selected for the trial, data for 14 were used for the VAS, with reasons.
Selective reporting (reporting bias)	Low risk	No interaction between time (active/ sham session first) for mood ratings, indicating similar performance. However, was power sufficient to detect any effect? All other outcomes (switching task effects reported). All other outcomes (switching task effects) reported.
Overall Bias	Unclear risk	at least two domains unclear RoB

Appendix F – Cortex Guidelines: Registered Reports

Guidelines for authors

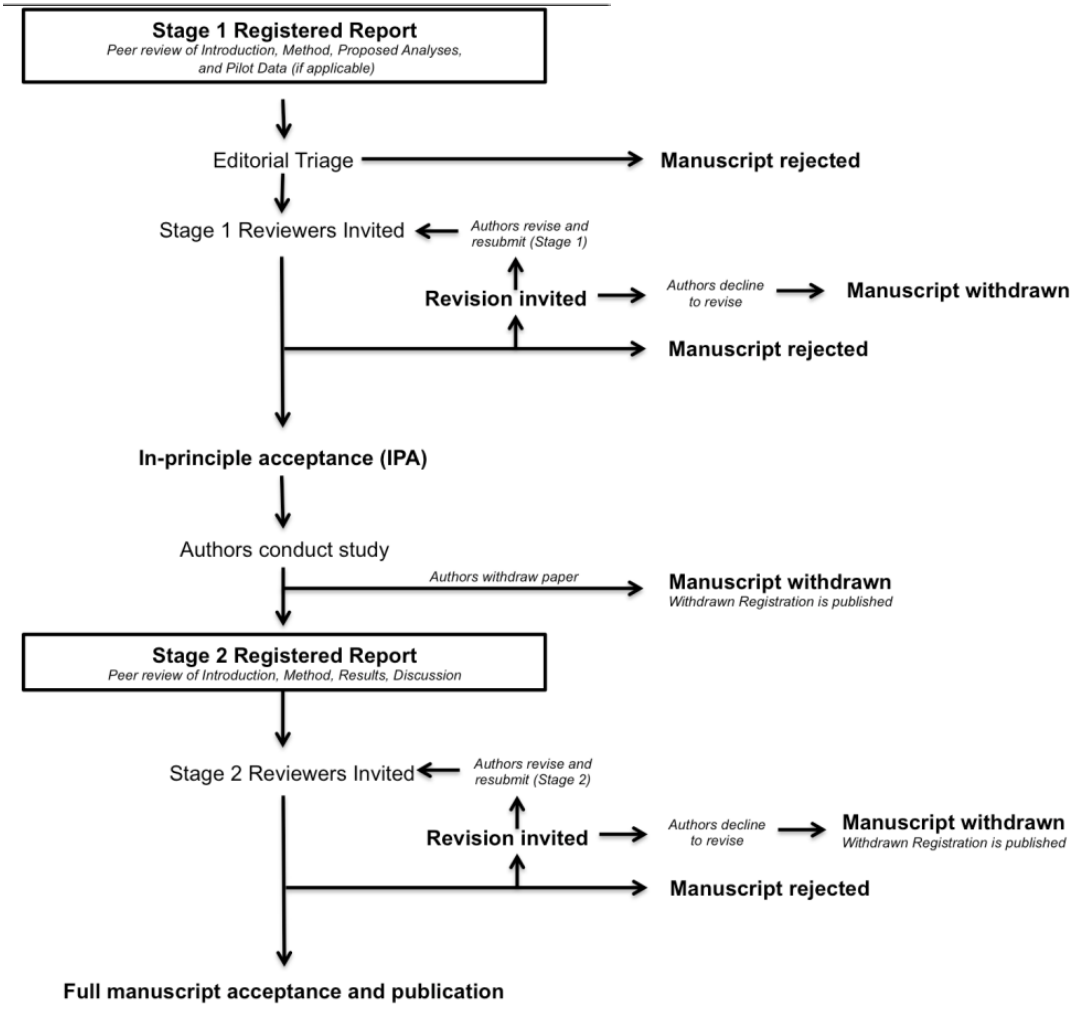
Registered Reports are a form of empirical article in which the methods and proposed analyses are pre- registered and reviewed prior to research being conducted. This format is designed to minimise bias in deductive science, while also allowing complete flexibility to conduct exploratory (unregistered) analyses and report serendipitous findings.

The cornerstone of the Registered Reports format is that a significant part of the manuscript will be assessed prior to data collection, with the highest quality submissions accepted in advance. Initial submissions will include a description of the key research question and background literature, hypotheses, experimental procedures, analysis pipeline, a statistical power analysis (or Bayesian equivalent), and pilot data (where applicable).

Initial submissions will be triaged by an editorial team for suitability. Those that pass triage will then be sent for in-depth peer review (Stage 1). Following review, the article will then be either rejected or accepted in principle for publication. Following in principle acceptance (IPA), the authors will then proceed to conduct the study, adhering exactly to the peer-reviewed procedures. When the study is complete the authors will submit their finalised manuscript for re-review (Stage 2) and will upload their raw data, digital study materials, and laboratory log to a publicly accessible file-sharing service. Pending quality checks and a sensible interpretation of the findings, the manuscript will be published regardless of the results.

Registered Reports are not subject to a word limit. Submissions should be as concise as possible, but as long as necessary to ensure that the description of methods is clear and comprehensive, and that all methods are reproducible.

The review process for *Registered Reports*



Stage 1: Initial manuscript submission and review

Stage 1 submissions should include the manuscript (details below) and a brief cover letter. Please note that the editorial board will not agree to send manuscripts for in-depth review until a complete Stage 1 submission has been considered.

The Stage 1 cover letter should include:

- A brief scientific case for consideration. Authors are encouraged to refer to the likely replication value of the research. Replication studies are welcome in addition to novel studies.
- A statement confirming that all necessary support (e.g. funding, facilities) and approvals (e.g. ethics) are in place for the proposed research. Note that manuscripts will be generally considered only for studies that are able to commence immediately; however, authors with alternative plans are encouraged to contact the journal office for advice.
- An anticipated timeline for completing the study if the initial submission is accepted.
- A statement confirming that the authors agree to share their raw data, any digital study materials,

and analysis code as appropriate.

- A statement confirming that, following Stage 1 *in principle acceptance*, the authors agree to register their approved protocol on the Open Science Framework (https://osf.io/) or other recognised repository, either publicly or under private embargo until submission of the Stage 2 manuscript.
- A statement confirming that if the authors later withdraw their paper, they agree to Cortex publishing a short summary of the pre-registered study under a section *Withdrawn Registrations*.

Manuscript preparation guidelines – Stage 1
Initial Stage 1 submissions should include the following sections:

• Abstract
o A succinct summary of the research question, hypotheses and proposed methods. The Stage 1 Abstract is provisional and subject to revision at Stage 2 to include results and conclusions.

• Introduction

o A review of the relevant literature that motivates the research question and a full description of the experimental aims and hypotheses. Please note that following IPA, the Introduction section cannot be altered apart from correction of factual errors, typographic errors and altering of tense from future to past (see below).

• Methods

o Full description of proposed sample characteristics, including criteria for data inclusion and exclusion (e.g. outlier extraction). Procedures for objectively defining exclusion criteria due to technical errors or for any other reasons must be specified, including details of how and under what conditions data would be replaced.

o A description of experimental procedures in sufficient detail to allow another researcher to repeat the methodology exactly, without requiring further information. These procedures must be adhered to exactly in the subsequent experiments or any Stage 2 manuscript can be rejected.

o Proposed analysis pipeline, including all pre-processing steps, and a precise description of all planned analyses, including appropriate correction for multiple comparisons. Any covariates or regressors must be stated. Where analysis decisions are contingent on the outcome of prior analyses, these contingencies must be specified. Only pre-planned analyses can be reported in the main Results section of Stage 2 submissions. However, unplanned exploratory analyses will be admissible in a separate section of the Results (see below).

o Authors are welcome to proposed blinded analysis techniques that control bias without requiring detailed pre-specification of analyses (e.g. as discussed here and deployed here).

o Studies involving Neyman-Pearson inference should include a statistical power analysis. 3

Estimated effect sizes should be justified with reference to the existing literature or theory. Since publication bias overinflates published estimates of effect size, power analysis must be based on the *lowest* available or meaningful estimate of the effect size. Where relevant, the *a priori* power must be 0.9 or higher for all proposed hypothesis tests, **with α set to .02** (not

.05). In the case of highly uncertain effect sizes, a variable sample size and interim data analysis is permissible but with inspection points stated in advance, appropriate Type I error correction for 'peeking' employed, and a final stopping rule for data collection outlined.

o Methods involving Bayesian hypothesis testing are encouraged. For studies involving analyses with Bayes factors, the predictions of the theory must be specified so that a Bayes factor can be calculated. Authors should indicate what distribution will be used to represent the predictions of the theory and how its parameters will be specified. For example, will you use a uniform distribution up to some specified maximum, or a normal/half-normal distribution to represent a likely effect size, or a JZS/Cauchy distribution with a specified scaling constant? For inference by Bayes factors, authors must be able to guarantee data collection until the Bayes factor is at least 6 times in favour of the experimental hypothesis over the null hypothesis (or *vice versa*). Authors with resource limitations are permitted to specify a maximum feasible sample size at which data collection must cease regardless of the Bayes factor; however to be eligible for advance acceptance this number must be sufficiently large that inconclusive results at this sample size would nevertheless be an important message for the field. For further advice on Bayes factors or Bayesian sampling methods, prospective authors are encouraged to read this key article by Schönbrodt and Wagenmakers. These two articles by *Cortex* editor Zoltan Dienes (here and here) also provide practical advice for specifying theoretically relevant effect sizes for hypothesis testing.

o For fMRI studies, various tools are available to support power calculations (e.g. www.neuropowertools.org and http://fmripower.org/) or authors are welcome to propose custom methods. In place of conventional hypothetico-deductive methods, authors are welcome to propose more innovative analytic approaches such as integration of Bayesian optimisation and hypothesis testing.

o **Please note that, due to the increasing volume of submissions, proposals that propose power analyses or Bayes factors thresholds but which fail to meet the minimum requirements (Power ≥ .9 and α ≤ .02; BF ≥6 or ≤1/6 for *all* hypothesis tests) may be desk rejected without the opportunity to resubmit. Authors who are unable to meet minimum sample size requirements (e.g. due to resource limitations) are encouraged to consider consortia-based methods of collaboration and recruitment, including the Psychological Science Accelerator and StudySwap.**

o Full descriptions must be provided of any outcome-neutral criteria that must be met for successful testing of the stated hypotheses. Such quality checks might include the absence of floor or ceiling effects in data distributions, positive controls, or other quality checks that are orthogonal to the experimental hypotheses.

o Timeline for completion of the study and proposed resubmission date if Stage 1 review is successful. Extensions to this deadline can be negotiated with the Registered Reports editor.

o Any description of prospective methods or analysis plans should be written in future tense. • Pilot Data

o Optional. Can be included to establish proof of concept, effect size estimations, or feasibility of proposed methods. Any pilot experiments will be published with the final version of the manuscript and will be clearly distinguished from data obtained for the pre-registered experiment(s).

Stage 1 submissions that are judged by the editorial board to be of sufficient quality and within journal scope will be sent for in-depth peer review. In considering papers at the registration stage, reviewers will be asked to assess:

1. The scientific validity of the research question(s).
2. The logic, rationale, and plausibility of the proposed hypotheses.
3. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis where appropriate).
4. Whether the clarity and degree of methodological detail is sufficient to exactly replicate the proposed experimental procedures and analysis pipeline.
5. Whether the authors have pre-specified sufficient outcome-neutral tests for ensuring that the obtained results can test the stated hypotheses, including positive controls and quality checks.

Following Stage 1 peer review, manuscripts will be rejected outright, offered the opportunity to revise, or accepted. Proposals that exceed the highest standards of scientific rigour will be issued an *in principle acceptance* (IPA), indicating that the article will be published pending completion of the approved methods and analytic procedures, passing of all pre-specified quality checks, and a defensible interpretation of the results. Stage 1 protocols are not published following IPA. Instead they are held in reserve by the journal and integrated into a single completed article following approval of the final Stage 2 manuscript.

Authors are reminded that any deviation from the stated experimental procedures, regardless of how minor it may seem to the authors, could lead to rejection of the manuscript at Stage 2. In cases where the pre-registered protocol is altered after IPA due to unforeseen circumstances (e.g. change of equipment or unanticipated technical error), the authors must consult the editorial board immediately for advice, and prior to the completion of data collection. Minor changes to the protocol may be permitted per editorial discretion. In such cases, IPA would be preserved and the deviation reported in the Stage 2 submission. If the authors wish to alter the experimental procedures more substantially following IPA but still wish to publish their article as a Registered Report then the manuscript must be withdrawn and resubmitted as a new Stage 1 submission. Note that registered analyses must be undertaken, but additional unregistered analyses can also be included in a final manuscript (see below).

Stage 2: Full manuscript review

Once the study is complete, authors prepare and resubmit their manuscript for full review, with the following additions:

• Cover letter. The Stage 2 cover letter must confirm:
o That the manuscript includes a link to the public archive containing anonymized study

data, digital materials/code and the laboratory log.
o That the manuscript contains a link to the approved and published Stage 1 protocol on the

Open Science Framework or other recognised repository.
o That, for primary Registered Reports, no data for any pre-registered study (other than pilot

data included at Stage 1) was collected prior to the date of IPA. For secondary Registered Reports, authors should confirm that no data (other than pilot data included at Stage 1) was subjected to the pre-registered analyses prior to IPA.

o That the Stage 2 manuscript tracks all changes, however minor, to any part of the submission that was approved at Stage 1. Authors should submit a tracked-changes version of the Stage 2 manuscript in addition to a clean version.

Please note that authors are welcome to update the title of the Stage 2 submission (e.g. to reflect the conclusions) and can also add or remove authors without requesting permission from the journal office. The Abstract should also be updated to state the results and conclusions.

• Submission of anonymised raw data, digital study materials, and laboratory log
o Anonymised raw data and digital study materials must be made freely available in a public repository/archive with a link provided within the Stage 2 manuscript. Authors are free to use any repository that renders data and materials freely and publicly accessible and provides a digital object identifier (DOI) to ensure that the data remain persistent, unique and citable. Potential repositories include (but are not limited to), Figshare, Harvard Dataverse, and Dryad. For a comprehensive list of available data repositories, see

http://www.re3data.org/

o Data files should be appropriately time stamped to show that data was collected *after* IPA and not before. Other than pre-registered and approved pilot data, no data acquired *prior* to the date of IPA is admissible in the Stage 2 submission. Raw data must be accompanied by guidance notes, where required, to assist other scientists in replicating the analysis pipeline. Authors are required to upload any relevant analysis scripts and other digital experimental materials that would assist in replication.

o Any supplementary figures, tables, or other text (such as supplementary methods) can either be included as standard supplementary information that accompanies the paper, or they can be archived together with the data. Please note that the raw data itself should be archived (see above) rather than submitted to the journal as supplementary material.

o A basic laboratory log must also be provided outlining the range of dates during which data collection took place. This log should be uploaded to the same public archive as the data and materials. **Authors are reminded that the laboratory log must contain no information that could identify any individual participant, including their name, initials, date of birth, or any identifying notes. Where the specific date or time of testing could identify a participant then authors should replace the log with a signed declaration that all data collection, other than any preliminary data reported in the Stage 1 manuscript, took place after the date of in principle acceptance.**

o The Stage 2 manuscript must also contain a link to the registered protocol (deposited following IPA) on the Open Science Framework or other recognised repository. If the protocol was deposited under an embargo, the embargo must be lifted at the point of Stage 2 submission to permit full public access.

• Background, Rationale and Methods
o Apart from minor revisions, **the Introduction cannot be altered from the approved**

Stage 1 submission, and the stated hypotheses cannot be amended or appended. At Stage 2, any description of the rationale or proposed methodology that was written in future tense within the Stage 1 manuscript should be changed to past tense. Any textual changes to the Introduction or Methods (e.g. correction of typographic or factual errors) must be clearly marked in the Stage 2 submission. Any relevant literature that appeared following the date of IPA should be covered in the Discussion.

• Results & Discussion
o The outcome of all registered analyses must be reported in the manuscript, except in rare instances where a registered and approved analysis is subsequently shown to be logically flawed or unfounded. In such cases, the authors, reviewers, and editor must agree that a collective error of judgment was made and that the analysis is inappropriate. In such cases the analysis would still be mentioned in the Methods but omitted with justification from the Results.

o It is reasonable that authors may wish to include additional analyses that were not included in the registered submission. For instance, a new analytic approach might become available between IPA and Stage 2 review, or a particularly interesting and unexpected finding may emerge. Such analyses

are admissible but must be clearly justified in the text, appropriately caveated, and reported in a separate section of the Results titled "*Exploratory analyses*". Authors should be careful not to base their conclusions entirely on the outcome of statistically significant *post hoc* analyses.

o Authors reporting null hypothesis significance tests are required to report exact *p* values and effect sizes for all inferential analyses.

The resubmission will most likely be considered by the same reviewers as in Stage 1, but could also be assessed by new reviewers. In considering papers at Stage 2, reviewers will be asked to decide:

1. Whether the data are able to test the authors' proposed hypotheses by satisfying the approved outcome-neutral conditions (such as quality checks, positive controls)
2. Whether the Introduction, rationale and stated hypotheses are the same as the approved Stage 1 submission (required)
3. Whether the authors adhered precisely to the registered study procedures
4. Whether any unregistered *post hoc* analyses added by the authors are justified, methodologically sound, and informative
5. Whether the authors' conclusions are justified given the data

Reviewers are informed that editorial decisions will not be based on the perceived importance, novelty or conclusiveness of the results. Thus, while reviewers are free to enter such comments on the record, they will not influence editorial decisions. Reviewers at Stage 2 may suggest that authors report additional *post hoc* tests on their data; however, authors are not obliged to do so unless such tests are necessary to satisfy one or more of the Stage 2 review criteria.

Manuscript withdrawal and *Withdrawn Registrations*
It is possible that authors with IPA may wish to withdraw their manuscript following or during data collection. Possible reasons could include major technical error, an inability to complete the study due to other unforeseen circumstances, or the desire to submit the results to a different journal. In all such cases, manuscripts can of course be withdrawn at the authors' discretion. However, the journal will publicly record each case in a section called *Withdrawn Registrations*. This section will include the authors, proposed title, the abstract from the approved Stage 1 submission, and brief reason(s) for the failure to complete the study. Partial withdrawals are not possible; i.e. authors cannot publish part of a registered study by selectively withdrawing one of the planned experiments. Such cases must lead to withdrawal of the entire paper. Studies that are not completed by the agreed Stage 2 submission deadline (which can be extended in negotiation with the editorial office) will be considered withdrawn and will be subject to a Withdrawn Registration.

Incremental Registrations
Authors may add experiments to approved submissions. In such cases the approved Stage 2 manuscript will be accepted for publication, and authors can propose additional experiments for Stage 1 consideration. Where these experiments extend the approved submission (as opposed to being part of new submissions), the editorial team will seek to fast-track the review process. This option may be particularly appropriate where an initial experiment reveals a major serendipitous finding that warrants follow-up within the same paper. In cases where an incremented submission is rejected (at either Stage 1 or 2), authors will retain the option of publishing the most recently approved version of the manuscript. For further advice on specific scenarios for incremental registration, authors are invited to contact the editorial office: cortex@ed.ac.uk

Tips for Avoiding Desk Rejection at Stage 1

Many Registered Report submissions are desk rejected at Stage 1 prior to in-depth review for failing to sufficiently meet the Stage 1 editorial criteria. For submissions that are otherwise highly promising, the journal will sometimes desk reject with the option to resubmit. However, due to the increasing

volume of submissions at Cortex, desk-based reject-and-resubmit decisions are being deployed more selectively and most manuscripts with these problems will now be desk rejected outright. To help minimize the chances of authors' submissions being desk rejected, we list below the top ten reasons why Stage 1 submissions are rejected prior to review. Authors are advised to consult this list carefully to increase their chances of proceeding immediately to Stage 1 in-depth review.

1. The cover letter doesn't make necessary statements concerning ethics, data archiving, protocol registration and so forth (see above).

2. The protocol contains insufficient methodological detail to enable replication and prevent researcher degrees of freedom. One commonly neglected area is the criteria for excluding data, both at the level of participants and at the level of data within participants. In the interests of clarity, we recommend listing these criteria systematically rather than presenting them in prose.
3. Lack of correspondence between the scientific hypotheses and the pre-registered statistical tests. This is a very common problem. To maximize the clarity of correspondence between predictions and analyses, authors are advised to number their hypotheses in the Introduction and then number the proposed analyses in the Methods to make clear *which analysis tests which prediction*. Ensure also that power analysis, where applicable, is based on the actual test procedures that will be employed to test those hypotheses; e.g. don't propose a power analysis based on an ANOVA but then suggest a linear mixed effects model to test the hypothesis. Where multiple hypotheses are being tested, each must be associated with a power analysis, Bayesian sampling plan or appropriate alternative, and all must achieve the minimum requirements (see 4). To help ensure that these requirements are met, we strongly recommend that authors include a study design table in their manuscript, as available in Section 9 of this template.
4. Power analysis, where applicable, fails to reach the minimum level (.90), or authors fail to calibrate their power analyses to the required $\alpha = .02$ (rather than the traditional $\alpha = .05$).
5. Power analysis is over-optimistic (e.g. based on previous literature but not taking into account publication bias) or insufficiently justified (e.g. based on a single point estimate from a pilot experiment or previous study). Proposals should be powered to detect the smallest effect that is plausible and of theoretical value. Pilot data can help inform this estimate but are unlikely to form an acceptable basis, alone, for choosing the target effect size due to risk of bias.
6. Intention to infer support for the null hypothesis from statistically non-significant results, without proposing use of Bayes factors or frequentist equivalence testing.
7. Inclusion of exploratory analyses in the analysis plan. Inclusion of exploratory "plans" at Stage 1 blurs the line between confirmatory and exploratory outcomes at Stage 2. Instead, such analyses can be included at Stage 2 and need not be pre-registered. Under some circumstances, exploratory analyses could be discussed at Stage 1 where they are necessary to justify study variables or procedures that are included in the design exclusively for exploratory analysis.
8. Failure to clearly distinguish work that has already been done from work that is planned. Where a Stage 1 proposal contains a mixture of preliminary/pilot work that has already been undertaken and a proposal for work not yet undertaken, authors should use the past tense for pilot work but the future tense for the proposed work. At Stage 2, all descriptions shift to past tense.
9. Lack of pre-specified positive controls or other quality checks, or an appropriate justification for their absence (See Stage 1 criterion 5). We recognise that positive controls are not possible with all study designs, in which case authors should discuss why they are not included.
10. Where applicable, lack of power analysis within proposed positive controls or manipulation checks that depend on hypothesis testing.

Appendix G: Stage 1 In Principle Acceptance Confirmation for our Registered Report , following editorial and peer review.

Dear Dr. Lowri Hadden,

Thank you for submitting your work to Cortex for consideration as a Registered Report (Stage 1). We are pleased to award in-principle acceptance (IPA) of your paper.

Appendix H - Deviations from Ahn et al. (2013) and Duprat et al. (2016)

Table 1 below highlights the principle deviations from Ahn et al's (2013) protocol. We describe each major deviation from Ahn et al. (2013) and justify alternative methods e.g. the use of an iTBS protocol, as opposed to the HF-TMS protocol used by Ahn et al. (2013). Note that iTBS is thought to be more efficient than HF-TMS, involving fewer pulses for comparable effects and both the HF-TMS and iTBS protocols, applied respectively in Ahn et al. (2013) and Duprat et al. (2016), are acknowledged as excitatory TMS protocols (Huang et al., 2011). As the more recently developed iTBS is more efficient, its use in treatment appears to be growing at a faster rate than HF-rTMS and was, therefore, the focus of this replication (Bakker et al., 2015; Grossheinrich et al., 2009). We also used a MagStim Rapid ^{2} stimulator as opposed to a MagStim Rapid ^{2} Plus 1 (see Table 1) used by Duprat et al. (2016) based on stimulator availability, which resulted in a frequency drop-off. The frequency drop-off was due to the capacitance of this system not being powerful enough to sustain the power output necessary for the higher motor thresholds (e.g. motor thresholds above 55), that is our system generated slightly lower frequency outputs (44.83 Hz as opposed to 50 Hz, see Table 2) compared to Duprat et al., (2016). Frequencies above ~5Hz are posited to induce excitatory effects (e.g. Siebner & Rothwell, 2003; Huang et al. 2005), thus our protocol remains excitatory (with the average frequency output in our study being 44.83 Hz), and still above the high frequency (10 Hz) TMS protocol used by Ahn et al., (2013).

Table 1. Major Deviations from Ahn et al., (2013) protocol, and justification of alternative methods.

Study Element	Ahn et al., (2013)	Duprat et al., (2016)	Current study	Justification
TMS protocol	HFTMS	iTBS	iTBS in accordance with Duprat et al., (2016)	In more recent treatment protocols for depression, iTBS is being used as a treatment (e.g. see Bulteau et al., 2017).
TMS Stimulator	Magstim 200 Magnetic Stimulator	MagStim Rapid ^{2} Plus1	MagStim Rapid ^{2}	Although a different stimulator will be used, with the potential of reduced frequency due to capacitance of the system, the frequencies are well within the range (i.e. ~10Hz) to generate excitatory neural changes (e.g. Huang et al., 2005; Siebner & Rothwell, 2003).
Number of Probabilistic Learning Tasks	2 (post active and post sham stimulation).	4 (pre and post active and pre and post sham stimulation).	4 probabilistic learning tasks in accordance with Duprat et al., (2016).	A baseline measure of reward bias allows for the exclusion of the possibility that day-to-day variance might explain observations.
Hedonic Capacity Questionnaire	No measures used to assess hedonic capacity.	Temporal Experience of Pleasure Scale (Gard et al., 2006).	Temporal Experience of Pleasure Scale (Gard et al., 2006).	Based on Duprat et al., (2016)'s finding that hedonic capacity interacts with reward responsiveness.
Mood questionnaire	No measures used to assess mood.	No measures used to assess mood.	Positive and Negative Affect Schedule (Watson et al., 1988).	To ensure effects of active compared to sham TMS is related to depressive symptoms, e.g. low mood. See secondary hypotheses section for precise hypotheses.

Table 2. Frequency drop-off relating to motor threshold and power output of stimulator (MagStim Rapid ^{2}). PID= participant Identification; MT= Motor Threshold.

PID	Motor Threshold (MT)	Power Output (1.10 x MT)	Frequency	Frequency Drop-off
1	50	55	48Hz	2Hz
2	53	58	46Hz	4Hz
3	44	48	50Hz	0
4	44	48	50Hz	0
5	48	53	49Hz	1Hz
6	55	61	45Hz	5Hz
7	53	58	46Hz	4Hz
8	66	73	39Hz	11Hz
9	46	51	50Hz	5Hz
10	65	72	39Hz	11Hz
11	65	72	39Hz	11Hz
12	54	59	45Hz	5Hz
13	57	63	44Hz	6Hz
14	61	67	42Hz	8Hz
15	55	61	45Hz	5Hz
16	60	66	42Hz	8Hz
17	65	72	39Hz	11Hz
18	48	53	49Hz	1Hz
19	-	-	-	-
Average	**54.944**	**60.555**	**44.83 Hz**	**5.16 Hz**

Appendix I - Exclusion Criteria

If participants did not pass initial safety measures pertaining to TMS approved by Cardiff University, they did not participate in the experiment (Allen, Singh, Verbruggen, & Chambers, 2018; Maizey et al., 2013). In line with Duprat et al. (2016), if individuals did not complete the task appropriately (e.g. only pressing one key throughout the experiment) they were not included in any analyses. In addition, if participants' reaction times were too quick (i.e. under 200ms) or too slow (i.e. over 2000ms) those responses were not included in any analyses (Ahn et al., 2013); if this occurred on greater than 10% of trials the participant's data was excluded. Participants were free to withdraw for any reason. Unanticipated technical failings also resulted in participant data being excluded (n =3, the coil overheating). If a participant only completed one testing session, their data was not included in the analyses.

Appendix J- Counterbalancing

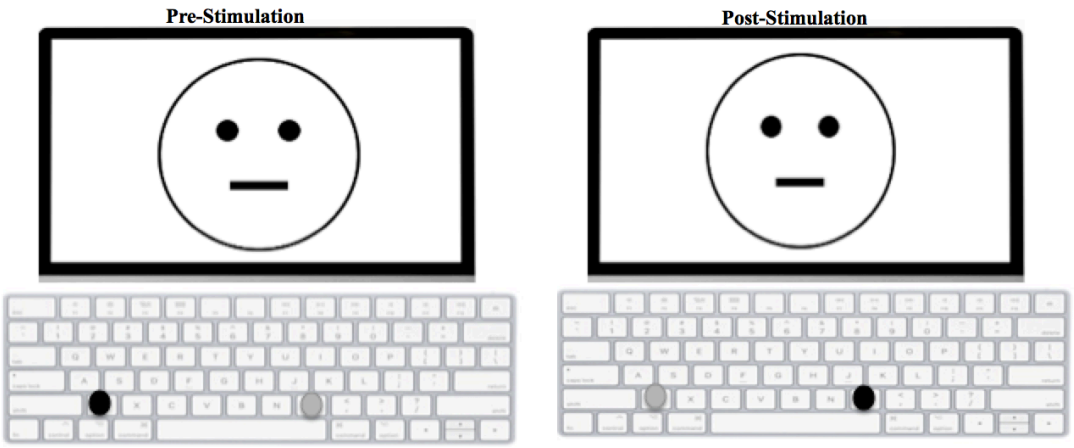


Figure A1. Counterbalancing procedure for sessions: pre and post stimulation. Keyboard presses "Z" and "M" will be counterbalanced within a testing session and between participants. The main experimental manipulation of "rich" and "lean" stimuli will be counterbalanced within stimulation sessions. The pairing will be determined randomly at the start of the stimulation session and counterbalanced across each stimulation session.

For the PLT, several parameters were counterbalanced (see Figure A1). To reduce order effects and to control for a laterality bias, the response keys ("Z" or "M") that participants press were counterbalanced between stimulation sessions. This was conducted

through the assignment of odd and even numbered participants receiving opposite keyboard presses. Each of the response keys were paired with a face with a "long" or "short" mouth, which relates to the experimental manipulation of "rich" versus "lean" reinforcement schedules. For the odd numbered participants, the "Z" key was paired with a "long" mouth and the "M" key with a "short" mouth pre-stimulation and the reverse pairing post stimulation ("Z" with "short" and "M" with "long"). The even participants received the opposite pairings. Within each stimulation session, the face that was rewarded richly or leanly – short or long – was chosen at random (using Python's *shuffle* function) at each pre-stimulation session. Then, at post-stimulation sessions, the ordering was reversed. Participants were also asked to press the "Z" key with their left index finger and the "M" key with their right index finger.

Appendix K- TMS Target location.

The study that is the target of our replication, Ahn et al (2013), used a standard location for the DLPFC for each participant, but they do not give precise information about how they located this area. We used participants' MRI scans, where available, to take into account individual structural differences between participants. The target for the DLPFC was identified in Montreal Neurological Institute (MNI) coordinates as x= –27, y=30, z=38, which is the mean of DPLFC co-ordinates reported as being active in a perceptual decision-making task (Heekeren, Marrett, Bandettini, & Ungerleider, 2004), and targeted in a TMS based cognitive control intervention (Hayashi, Ko, Strafella, & Dagher, 2013).

Appendix L- TMS parameters: treatment protocol.

The active iTBS consisted of 1620 pulses in 54 cycles of 10 bursts of 3 pulses with a train duration of 2s and inter-train interval of 8s with a power output of 110% of the resting motor threshold.

Appendix M – Questionnaire removed due to copyright

Appendix N – Questionnaire removed due to copyright

Appendix O- Bayes Factors and sample size estimates

Primary Hypothesis

We used a Bayesian approach (Dienes, 2014) to estimate the likely sample size needed to provide support for the null or alternative hypothesis, related to our critical effect of interest. Assuming a difference of zero and the same *SE* and sample size reported in Ahn et al. (2013), and applying the R Code (Dienes & Belfi, 2019) we were able to estimate the Bayes Factor for our proposed sample size of 30 under the null as 0.14. This was computed through scaling the *SE* (0.063) in Ahn et al. (2013) by the square root of their sample size divided by our proposed sample size (SE_{scaled} = 0.049) for use in the Bayes Factor function. As we are closely replicating Ahn et al. (2013) study, we have made the assumption that our testing procedures will be similar and should, therefore, yield no additional variance.

Secondary Analysis

For our secondary analysis that active compared to sham stimulation will increase positive affect and decrease negative affect, using the R Code (Dienes & Belfi, 2019), and assuming no effect and the *SE*, and sample size reported in Chaves et al. (2017), we were able to estimate the Bayes Factor for our proposed sample size of 30 under the null as 0.01 for negative affect, measured on the PANAS-NA; and the Bayes Factor of 0.02 for positive affect, as measured on the PANAS-PA.

Previous Critical Interactions of Interest

We computed the Bayes Factor for the significant interactions reported in Ahn et al. (2013) and Duprat et al. (2016) following the method outlined in Dienes (2014) and detailed further in our main analyses section above. Using the R Code (Dienes & Belfi, 2019), sample size, and SEs calculated for Ahn et al. (2013) interaction (Stimulation × Block) we estimated a Bayes factor of 0.20 under the null for our sample size of 30. The Bayes Factor was

computed through scaling the *SE* (0.08) for Ahn et al's (2013) interaction by the square root of their sample size divided by our sample size (SE_{scaled} = 0.06). We followed the same procedure for Duprat et al's (2016) significant interaction (described further in our analysis section). For Duprat et al. (2016), the Bayes Factor we obtained under the null was 0.20, using their *SE* (0.07; with a *df* = 20), scaling to our sample size of 30 (SE_{scaled} = 0.06).

Manipulation Checks

Using the method described above in the corresponding Analysis section, and R code produced by Dienes and Belfi (2019) we estimated a Bayes Factor of 0.04 under the null for our sample size of 30, using the sample size, *SE* and *df* for the RB reported in block 3 for Ahn et al. (2013) post sham stimulation (*M* = 0.19, *SD* = 0.26, *SE* = 0.061). Similarly for Duprat et al. (2016), we used the reported mean for RB in block 3 post sham stimulation (*M* = 0.26, *SD* = 0.32, *SE* = 0.070) and obtained a Bayes Factor of 0.02 under the null after scaling the SEs of the sample to our maximum sample size of 30.

Appendix P – Behavioural analyses

Exploratory analyses: Behavioural

Response Accuracy

In accordance with Ahn et al. (2013) and Duprat et al. (2016) and due to low variance in the data, we used arcsine transformation on the accuracy data. Similar to Ahn et al. (2013) we conducted a 2 × 3 × 2 Repeated Measures ANOVA with condition (lean, rich) × block (1,2,3) and post stimulation (active, sham). There was no main effect of condition, F (1,18) =1.188, p = .290, η_p^2 = 0.062, d = 0.514; block: F(2, 36) = 1.003, p 0.377, η_p^2 = 0. 053, d = 0.473 or stimulation: F(1,18) = 0.122, p = 0.731, η_p^2 = 0.007, d = 0.167. However, there was a significant interaction between condition × block, F(2,36) = 5.297, p = 0.010, η_p^2 = 0. 227,

d = 1.083. Bonferroni post-hoc comparisons revealed that response accuracy increased for the richly rewarded condition (M = 1.319, SE = .027) whilst decreasing for the "lean" rewarded (M = 1.241; SE = 0.21) condition in block 3 of the probabilistic learning task (see Figure A2), which is indicative of reinforcement learning, irrespective of rTMS application. There were no further significant interactions, all ps > 0.364. The interaction between condition and block is visualised in figure A2.

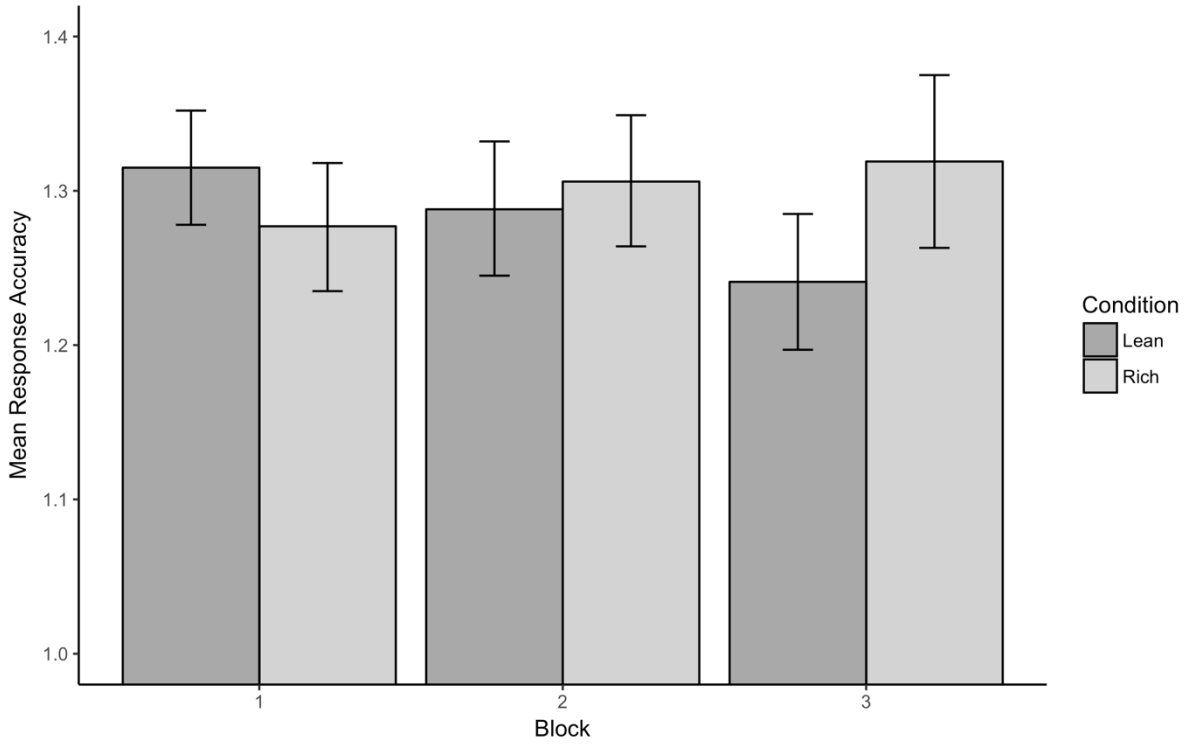


Figure A2. Response accuracy (arcsine transformed) across blocks (1,2,3) for rich and lean stimuli (averaged across active and sham stimulation), illustrating accuracy increasing for rich, but decreasing for lean stimuli.

In line with the additional factor, time (pre-post) in Duprat et al., (2016)'s paper, we conducted a further Repeated Measures ANOVA identical to the one above, but including the factor time (pre, post). There were no significant main effects, all p values ranging from 0.077 (for time) to 0.84 (for stimulation). The only significant interaction was between condition × block (1,2,3) as explicated above, F (2, 36) = 4.122, p = 0.024, η_p^2 = 0.186, d = 0.956. Pairwise comparisons using Bonferroni correction indicated that participants were

more accurate for the rich and less accurate for the lean across the blocks, (M_{diff} = -0.045), p = 0.039.

Reaction Time

In line with data analysis conducted by Ahn et al., (2013) we conducted a 2 × 3 × 2 Repeated Measures ANOVA, for post stimulation (2: active, sham) × block (1,2,3) × condition (rich, lean) as the factors. In accordance with Ahn et al., (2013) we did not log transform the RTs. However, we also conducted a further Repeated Measures ANOVA identical to that of Ahn et al., (2013), but with the addition of time (pre-post) as a factor and this data was log transformed consistent with the analyses of Duprat et al., (2016).

For the first ANOVA replicating Ahn et al., (2013)'s analyses, there was no significant main effect of stimulation, F (1,18) = 4.228, p =0.055, η_p^2= 0.190, (active, M = 0.447, SE = 0.016; Sham, M = 0.428, SE = 0.013). However, there was a main effect of condition, F (1,18) = 7.012, p = 0.016, η_p^2 = 0.280, with rich responses (M = 0.426, SE = 0.012) being faster than lean (M = 0.449, SE = 0.017). There was also a significant interaction between block (1,2,3) × condition (rich, lean), F (2, 36) = 3.643, p = 0.36, η_p^2= 0.168. Post-hoc tests using Bonferroni correction indicates that reaction time for rich stimuli are getting faster across the blocks, (M = 0.407, SE 0.015) whilst RTs for lean stimuli is decreasing across blocks (M = 0.407, SE =0.010), p = 0.001. There were no other main effects or interactions (all ps > 0.112).

We also replicated the analysis of Duprat et al. (2016) for the reaction time data, which was log transformed. This repeated measures ANOVA was identical to the one above, but included the additional factor of time (pre-post), alongside stimulation (active, sham), block (1,2,3) and condition (rich, lean). Here, we found a main effect of block, F (1.508, 27.136) = 0.012, p = 0.012, η_p^2= 0.250, d = 1. 154 (with RTs log ms, increasing across the

blocks) and a main effect of condition F (1, 18) = 12.667, p = 0.002, η_p^2= .413, d= 1.677 with faster RTs for rich (M = 5.988, SE = 0.029) compared to lean stimuli (M = 6.020, SE = 0.31). There was also a significant interaction between time × block × condition, F (2, 36) = 7.063, p = 0.003, η_p^2= 0. 282, d = 1.253. Follow-up comparisons using Bonferroni corrections revealed that log RTs for the rich (M = 5.946; SE = 0.023) condition were faster than the lean (M = 6.028; SE = 0.031) condition post (M = 5.973, SE = 0.037) at block 3 for post stimulation, p = .001, irrespective of stimulation type (see Figure A2). There were no other comparisons, all ps >.105, or other significant main effects or interactions, *ps* between 0.073 and 0.895.

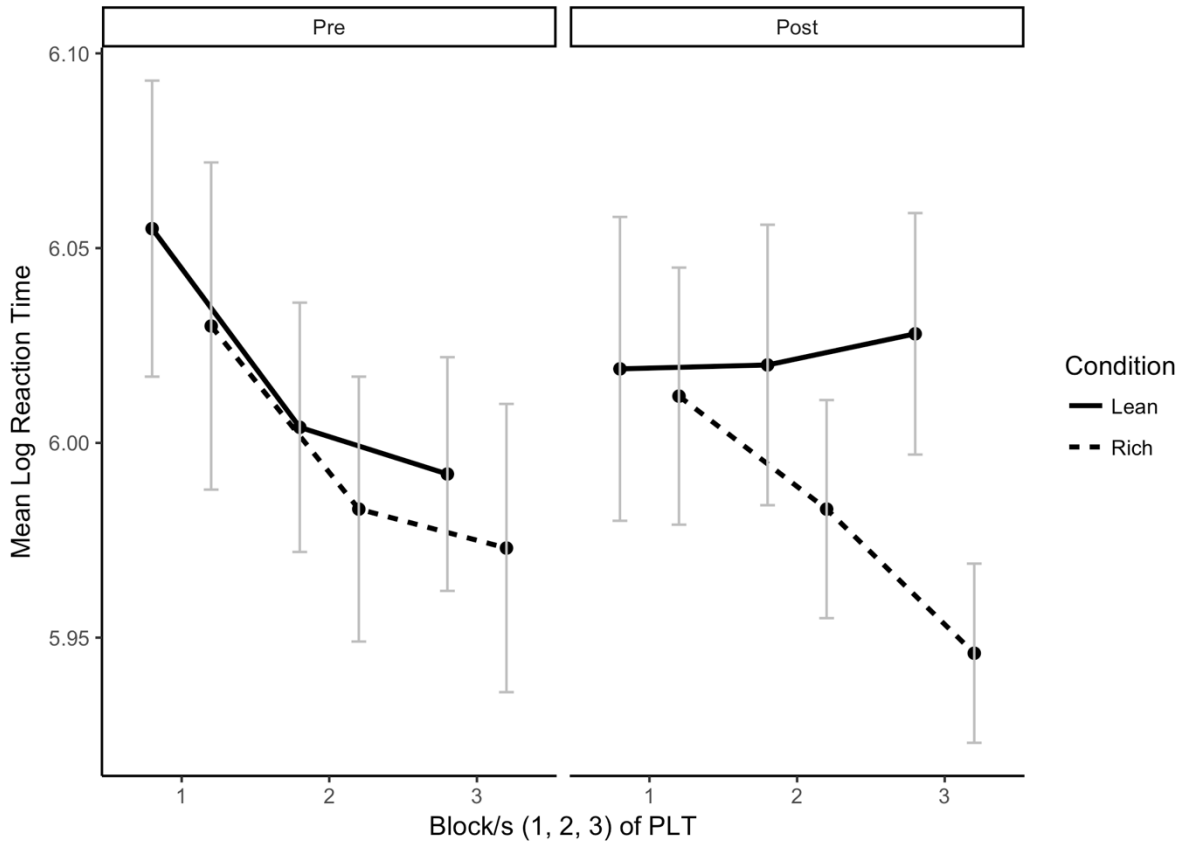


Figure A3. This graph depicts the significant interaction between condition x block x time illustrating faster reaction times for rich compared to lean for post stimulation, irrespective of stimulation type, for block 3 of the PLT.

Our behavioural analyses were commensurate with those of the extant PLT literature (Pizzagalli et al., 2005; Pizzagalli et al., 2008) indicating that response accuracy increased for the richly rewarded stimuli across blocks, whilst accuracy decreased for the lean rewarded stimuli, suggesting that participants acquired a response bias. Irrespective of stimulation type, reaction time was also quicker for rich compared to lean stimulation post stimulation. This reaction time result could be indicative of familiarity with the PLT, although we included the same number of PLTs as Duprat et al., (2016) who found faster reaction times for the rich stimuli following active iTBS alone.

Appendix Q- Bayes Factor cut-offs Cortex.

Given the Bayesian approach adopted we intended to continue collecting data until the resultant Bayes factors for primary and secondary analyses were all greater than 6 or less than 1/6 or we have collected data from 30 participants, due to feasibility constraints and in line with the above calculations. Bayes factors greater than 6 are to be interpreted as substantial evidence for the hypothesise and Bayes factors less than 1/6 are to be interpreted as substantial evidence for the null (Dienes, 2011; Jeffreys, 1998). This cut off aligns with *Cortex's* guidelines. However, due to complications arising from COVID, and not being able to continue testing, we were not able to meet this stringent threshold of obtaining Bayes Factors above 6 or below 1/6.