Neuroscience and Biobehavioral Reviews 125 (2021) 33–56

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Review article

Neurofeedback training in major depressive disorder: A systematic review of clinical efficacy, study quality and reporting practices

Lucas R. Trambaiolli [a,*], Simon H. Kohl [b,c], David E.J. Linden [d], David M.A. Mehler [e]

[a] Division of Basic Neuroscience, McLean Hospital, Harvard Medical School, Boston, USA
[b] JARA Institute Molecular Neuroscience and Neuroimaging (INM-11), Jülich Research Centre, Germany
[c] Department of Child and Adolescent Psychiatry, Medical Faculty, RWTH Aachen University, Germany
[d] School for Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands
[e] Department of Psychiatry, University of Münster, Germany

ARTICLE INFO

Keywords:
Neurofeedback
Biofeedback
Real-time fMRI
Electroencephalography
Functional magnetic resonance imaging
Major depressive disorder
Self-regulation
Neuroimaging

ABSTRACT

Major depressive disorder (MDD) is the leading cause of disability worldwide. Neurofeedback training has been suggested as a potential additional treatment option for MDD patients not reaching remission from standard care (i.e., psychopharmacology and psychotherapy). Here we systematically reviewed neurofeedback studies employing electroencephalography, or functional magnetic resonance-based protocols in depressive patients. Of 585 initially screened studies, 24 were included in our final sample (N = 480 patients in experimental and N = 194 in the control groups completing the primary endpoint). We evaluated the clinical efficacy across studies and attempted to group studies according to the control condition categories currently used in the field that affect clinical outcomes in group comparisons. In most studies, MDD patients showed symptom improvement superior to the control group(s). However, most articles did not comply with the most stringent study quality and reporting practices. We conclude with recommendations on best practices for experimental designs and reporting standards for neurofeedback training.

1. Introduction

Major depressive disorder (MDD) is a serious mental disorder characterized by at least one depressive episode lasting for two or more weeks (Association, 2013). This episode includes symptoms such as changes in cognition, reduced mood, interest or pleasure, and vegetative complaints (Otte et al., 2016). MDD has been recognized as a major public health challenge because of the increasing number of cases worldwide. For Western countries, it is estimated that MDD affects one in every five to six adults (Bromet et al., 2011; Patten, 2009). MDD represents a major risk factor for suicide attempts (Hoertel et al., 2015; Olfson et al., 2017). Moreover, MDD patients often suffer from comorbid psychiatric conditions (Alonso and Lépine, 2007), which increases the burden on patients and their families.

Current treatments mainly include psychotherapy or pharmacotherapy (Kupfer et al., 2012). The most widely used type of psychotherapy for depression is cognitive behavioral therapy (CBT), which aims to identify the cognitive factors leading to depressive symptoms and develop mental and behavioral strategies to cope with these (Otte et al., 2016). Another psychotherapeutic approach developed for the treatment is cognitive bias modification, which aims to readjust negative attention biases commonly observed in depressed patients (Fodor et al., 2020). The mainstay of current pharmacotherapy for depression are monoaminergic antidepressant drugs (Sharp, 2012). However, around one third of depressed patients do not respond to these conventional treatments (Fava and Davidson, 1996; Rush et al., 2006). Other therapeutic options include non-invasive brain stimulation such as transcranial magnetic stimulation (TMS) and electroconvulsive therapy (ECT), for which several stimulation protocols have been developed that show superiority compared to sham stimulation (Mutz et al., 2019). However, TMS and ECT can yield aversive effects, including local pain, headache and discomfort (Cusin and Dougherty, 2012; Rossi et al., 2009). Some ECT patients report acute but partly also persistent side effects of amnesia and cognitive disturbances following treatment (Sackeim et al., 2007). Lastly, invasive electrical deep brain stimulation (DBS) of subcortical and cortical areas is currently explored for its clinical potential (Delaloye and Holtzheimer, 2014), although most recent findings remain inconclusive and have sparked a debate in the

* Corresponding author.
E-mail address: ltrambaiolli@mclean.harvard.edu (L.R. Trambaiolli).

https://doi.org/10.1016/j.neubiorev.2021.02.015
Received 2 September 2020; Received in revised form 5 February 2021; Accepted 6 February 2021
Available online 12 February 2021

field (Bari et al., 2018). One common feature of electrical or magnetic brain stimulation treatments, shared with pharmacological treatment, is that patients remain passive recipients of the intervention.

In contrast, non-invasive neurofeedback training is a neuromodulation technique that involves patients as protagonists of their treatment. Patients learn self-regulating particular features of brain activity (Sitaram et al., 2017) by actively engaging in processes which are often adopted from techniques used in psychotherapy (Arns et al., 2017; Fovet et al., 2015). However, given the current discussion regarding the specificity and efficacy of neurofeedback protocols across psychiatric disorders (Thibault et al., 2018), a formal evaluation within specific conditions is much needed. MDD can be considered one of the most extensively studied applications of neurofeedback training, with the first case studies reported more than two decades ago (Baehr et al., 1997; Earnest, 1999; Rosenfeld et al., 1996). This systematic review pursues three main goals: First, we describe the different neurofeedback protocols that have thus far been explored with MDD patients and the main clinical and neural outcomes of these studies. Second, we summarize reported clinical changes and evaluate their efficacy. Lastly, we assess the study design and reporting quality of published research articles. We discuss limitations and open challenges, closing with a set of recommendations for future neurofeedback studies in MDD that may help advancing the field.

1.1. Description of a neurofeedback system

Neurofeedback is a non-invasive technique that provides the user with real-time feedback about their neural self-regulation performance. Feedback is commonly provided from areas that are thought of as putative neural substrates underlying specific behaviors or pathologies (Kim and Birbaumer, 2014; Sitaram et al., 2017). For instance, one well described and commonly found symptom in MDD is low mood. Several neurofeedback studies trained patients on neural correlates of emotion regulation with the aim to improve this capacity and the depressive symptom(s) related to low mood. Different imaging modalities have been used to train self-regulation in healthy participants and/or patients, including electroencephalography (EEG), magnetic encephalography (MEG), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS) (Thibault et al., 2016). Irrespective of the imaging modality, neurofeedback interventions usually consist of four main steps (Paret et al., 2019; Sitaram et al., 2017): 1) identifying the neural target (i.e., correlate of a symptom or skill) either by the means of functional data during a so-called localizer session or based on previous anatomical hypotheses using masks, 2) recording the neural activity of this neural target, 3) processing these measures while ideally controlling for potential artefacts and 4) presenting real-time feedback of this signal to the user.

At the recording stage, i.e. step one and two, the nature of brain signals needs to be considered as it differs between imaging modalities such as EEG and fMRI. For instance, EEG has been frequently used in depressed patients to search for neural correlates of mental states and later explored to develop neurofeedback protocols (Enriquez-Geppert et al., 2017; Gruzelier, 2014). EEG uses scalp electrodes, which similar to MEG, measure local field potentials (LFPs). LFPs represent the summed activity of local neural populations reflecting the electric potential in the extracellular space. Hence, EEG signals are largely determined by post-synaptic activity providing a direct measure of neural activity (Da Silva, 2009). fMRI is another neuroimaging technique that is increasingly used for neurofeedback experiments in depression (Watanabe et al., 2017; Weiskopf, 2012). This technique uses the blood oxygen level dependent (BOLD) contrast, a measure for the relative changes in local blood oxygenation that result from the metabolism of brain cells. fMRI hence provides an indirect measure of neural activity. More recently, fNIRS has gained the attention of the neurofeedback community (Kohl et al., 2020). fNIRS uses near-infrared light to measure local changes in oxygen concentrations in cortical gyri (Hoshi, 2003; Strangman et al., 2002; Villringer et al., 1993); these strongly correlate with the fMRI BOLD signal (Cui et al., 2011; Huppert et al., 2006; Strangman et al., 2002). Different neuroimaging technologies have their advantages and disadvantages, in particular, for real-time experiments (Thibault et al., 2016). For example, EEG provides higher temporal resolution and reduced cost compared to fMRI and fNIRS, and wireless-EEG systems provide new perspectives for portable therapeutic applications in the near future (De Vos et al., 2014; Ries et al., 2014). Conversely, fMRI possesses a higher spatial resolution, which allows the development of protocols that target both cortical and subcortical areas composing the circuitry of interest (Sulzer et al., 2013; Weiskopf, 2012). Multi-modal neurofeedback approaches attempt to bridge these advantages and compensate for some disadvantages by combining two or more neuroimaging techniques (Mano et al., 2017). For instance, these studies may benefit from the spatial resolution of fMRI and temporal resolution of EEG, combining these with the aim to achieve higher self-regulation performance (Perronnet et al., 2017).

The third step involves data processing methods. However, in real-time experiments, data preprocessing and data acquisition are quasi-simultaneous (depending on the delay of the respective imaging technique) such that recorded brain signals are continuously converted to an output system (Sitaram et al., 2017). For all brain imaging modalities noise-reduction methods are essential to increase the validity of the feedback; they are ideally applied to filter non-neural signal sources, such as the electrooculography (EOG) and electromyography (EMG) in EEG-based protocols (Moretti et al., 2003), or respiration and pulse waves in fMRI-based protocols (Murphy et al., 2013). Artifact corrected time-series are subsequently processed to calculate values that are subsequently used for feedback presentation. These values can be based on signal changes with respect to baseline in individual brain areas, correlations between time series of different brain areas (connectivity based feedback) (Koush et al., 2013; Ramot et al., 2017), or the output of more complex algorithms that classify different brain states based on variations in brain activity patterns (Watanabe et al., 2017). However, all these processing methods vary according to the neurofeedback paradigm and are subject to ongoing methodological research and development (Heunis et al., 2019, 2020; Hinterberger et al., 2003; Krusienski et al., 2006; Lotte et al., 2007).

The fourth and final design step concerns the presentation of real-time feedback. Although visual feedback is the most common approach, other feedback modalities can also be used in this stage and include auditory, vibrotactile, electrical or proprioceptive stimulation (Sitaram et al., 2017). The feedback setup should be carefully designed because it can cause distraction, frustration, or even induce negative emotions in users (Birbaumer et al., 2013; McFarland et al., 1998). The feedback system should constantly update the trainee about the targeted neural activity. Such real-time feedback allows the trainee to create, correct and optimize a mental or behavioral control strategy and thereby to achieve the desired level of proficiency in self-regulating neural activity (Birbaumer et al., 2013; Curran and Stokes, 2003).

1.2. Study design and non-specific effects of neurofeedback protocols

When conceptualizing this systematic review we were guided by a recently published consensus statement that discussed different mechanisms responsible for driving the outcomes of a neurofeedback experiment (Ros et al., 2020). The authors identified five potential contributors (Micoulaud-Franchi and Fovet, 2018; Ros et al., 2020): *neurofeedback-specific effects*, which are related to the actual training of a target neurophysiological variable (e.g., increased or decreased functional connectivity between trained ROIs); *non-specific neurofeedback effects*, associated to the neurofeedback context, but not to the trained neural signals (e.g., the high-tech environment); *general non-specific effects*, which are caused by psychosocial influences (e.g., believe-based expectations); *repetition related effects*, referring to the recurrence of training (e.g., test-retest improvements due to mental imagery tasks

employed in neurofeedback paradigms); and, finally, *natural effects*, associated to natural events in life (e.g., natural recovery or remission). The extent to which these factors contribute to overall clinical effects as observed in experimental (and to some degree also control) groups remains subject to ongoing and future research. Given that all potentially contributing factors as listed above likely play a role, and that they even interact with each other, some authors recently described neurofeedback as a complex intervention when studied in a clinical context (Craig et al., 2008; Sorger et al., 2019).

Similar to other interventions, developing a neurofeedback paradigm for clinical application requires several phases. Uncontrolled single-group designs are suitable for the early phase, for instance, to assess technical feasibility and acceptability of the paradigm in a healthy or patient sample. "Exploratory trials" may also serve to optimize the intervention in the targeted patient population (similarly to Phase I Clinical Trial designs) (Sorger et al., 2019). However, single-group designs cannot control for non-specific effects. Thus, further experiments with appropriate control conditions are needed during later phases to disentangle the neurofeedback-specific outcomes from those caused by other (psychosocial) mechanisms (Thibault et al., 2018; Thibault and Raz, 2016).

One main challenge that the neurofeedback field currently faces is that standards for the design of randomized controlled trials are traditionally based on the requirements that pharmacological studies need to fulfill. This challenge pertains in particular to the design of control conditions and risk of unblinding. For instance, in pharmacological studies the control group can receive a highly comparable treatment that omits the active component to drive improvement (Linden, 2014; Thibault et al., 2018). In such trials where participants receive so-called passive treatment (i.e., the intervention does not require a specific engagement in a task), the design of control conditions mainly need to account for belief-based expectations (commonly referred to as "placebo effects") rather than a range of contributing factors as listed above. Moreover, complex interventions such as neurofeedback that involve active engagement of the participant have their own requirements to ensure blinding (Linden, 2014; Sorger et al., 2019) (noteworthy, the design of appropriate placebo control conditions also remains subject for discussion in the pharmacological literature (Jensen et al., 2017; Moncrieff et al., 2004)).

Recent discussions in the field have therefore resulted in new best-practice research recommendations for different control conditions (for a detailed framework, please refer to (Sorger et al., 2019)). For this review, we grouped control conditions into three main categories:

- *Passive control*: this category includes control conditions that involve continued standard care only. Passive control conditions can reveal whether the neurofeedback has clinically significant benefit as a stand-alone, or add-on, intervention compared to standard care, for instance (Choi et al., 2011; Escolano et al., 2014; Wang et al., 2019, 2016). While this design controls for *natural effects* (e.g., regression to the mean), it does not control for any general or neurofeedback training related non-specific effects.
- *Active control outside the scanner*:this category includes control conditions where the participant is engaging in a similar mental task, but outside of the neuroimaging scanner (Jaeckle et al., 2019; Linden et al., 2012). This condition is also referred to as a mental-rehearsal control (Sorger et al., 2019). In addition to *natural effects*, it also allows to control for *repetition related* effects that occur by merely engaging in the behavioral/cognitive strategy.
- *Active control inside the scanner*: this category includes a variety of approaches in which the patient is actively performing a task inside the scanner and that may either control for *neurofeedback specific effects*, *non-specific neurofeedback effects*, or both (for a more detailed overview, please refer to (Sorger et al., 2019)). For example, patients in the control group are trained to self-regulate their brain activity in the same ROI but in the opposite direction of the experimental group. In an alternative design, they receive feedback from a different ROI or network using an alternative strategy (Mehler et al., 2018). Such designs match groups for some *general non-specific* and *non-specific neurofeedback effects* such as motivation, received reward during training, the high-tech environment, or the interaction with the experimenter and allow testing for *neurofeedback specific effects* (Thibault and Raz, 2016; Wood and Kober, 2018). Other approaches present feedback based on signals from a different brain source not associated with the brain function targeted in the experimental group (Young et al., 2017b, 2014), sham signal (e.g., randomly generated), non-neural sources (e.g., other biological features) or yoked data (e.g., data from a different participant (Hamilton et al., 2016)). However, it is essential to match groups for perceived rewards and evaluate to which extent patients remain "blind", considering that previous studies report that they could detect non-contingency and experienced adverse effects such as frustration and reduced motivation (Sorger et al., 2019).

Finally, it is also relevant to design "double-blind" or "triple blind" trial designs where the rater, participant and neurofeedback operator are blinded to the treatment condition. While only a few neurofeedback software packages are currently capable to blind the neurofeedback operator (Ros et al., 2020), double-blinding/triple-blinding can alternatively be achieved with two experimenters (these are either responsible for operating, and if necessary programming, the experiment or interacting with participants (Arnold et al., 2013)) in addition to independent and blinded research or clinical staff who assess the outcomes (Ros et al., 2020).

1.3. Rationale for the use of neurofeedback for MDD

Neurofeedback feasibility studies have yielded first promising results in different non-clinical and clinical applications ranging from athletic performance (Mirifar et al., 2017) to motor rehabilitation for neurodegenerative disorders and stroke (Krucoff et al., 2016; Linden and Turner, 2016). In neuropsychiatry, small randomized controlled studies have shown benefits for different disorders. For example, EEG-based training protocols were successfully applied to substance abuse disorders, eating disorders, attention-deficit/hyperactivity disorder, autism spectrum disorder, tinnitus, and obsessive-compulsive disorder, while fMRI-based training protocols have been successfully applied to attention-deficit/hyperactivity disorder (ADHD), post-traumatic stress disorder, schizophrenia, Alzheimer's disease, Tourette Syndrome, autism spectrum disorder, overweight/obesity, chronic pain, spider phobia, and obsessive-compulsive disorder (Arns et al., 2017; Sitaram et al., 2017; Thibault et al., 2018). These studies provide preliminary data suggesting that neurofeedback training may be effective in changing brain function and treating some neuropsychiatric symptoms, including those related to disturbances in the reward system. Noteworthy, clinical effects have been reported to last also during longitudinal follow-ups (Becerra et al., 2006; Gevensleben et al., 2010; Goldway et al., 2019; Mehler et al., 2018; Rance et al., 2018).

Most neurofeedback training paradigms are informed by neurophysiological or computational models suggested to explain the genesis of depressive symptoms. Thereby, this technique provides potentially a new way to directly test for the causal validity of reported biomarkers (Mehler and Kording, 2018; Micoulaud-Franchi et al., 2019). Similar to brain stimulation protocols (e.g. TMS or DBS) most neurofeedback protocols aim to modulate local activity. Importantly, the neurofeedback acts as an "endogenous" stimulator, reducing issues related to safety or side effects from conventional neuromodulation approaches. Additionally, such a form of non-invasive, endogenous neuromodulation puts the patient at the center of the intervention and may hence result in beneficial psychophysiological and psychosocial effects (see below) (Linden, 2014).

Besides local functional changes, several studies have also reported

remote effects of neurofeedback training at the network level. For instance, neurofeedback training has been reported to alter intrinsic functional connectivity (Hampson et al., 2011; Scheinost et al., 2013) and directed effective connectivity (Zotev et al., 2011, 2013). Moreover, these alterations (and related symptomatic improvements) were partly found to persist for months, supporting the idea of long-term changes in network organization (Megumi et al., 2015). Reports of long lasting network changes and associated clinical improvement render neurofeedback training a particularly promising approach to treat patients that present with abnormal connectivity patterns in brain networks relevant for affective and cognitive processing (Hamilton et al., 2015; Mulders et al., 2015).

Cognitive processes that have been suggested to contribute to the pathophysiology of depression include biased attention and processing of negative stimuli, recall of negative memories (Lewinsohn and Rosenbaum, 1987; Sato and Kawahara, 2011), and recurrent negative thoughts (rumination) (Beck, 2008; Clark and Beck, 2010; Disner et al., 2011). These cognitive processes share underlying brain structures which are commonly reported as showing abnormal activity, or connectivity, in patients with depression, such as the lateral and medial prefrontal cortex (PFC), anterior cingulate cortex (ACC), amygdala, hippocampus, and striatum (Groenewold et al., 2013; Kaiser et al., 2015). Thus, the majority of neurofeedback protocols for MDD aim to directly or indirectly rebalance these networks. For instance, in fMRI-based protocols these areas are the ones commonly targeted for self-regulation, as individual ROIs (Jaeckle et al., 2019; Young et al., 2017b, 2014) or as multi-ROI networks (Linden et al., 2012; Mehler et al., 2018). Although some studies using EEG-based neurofeedback claim to target some of these brain structures (e.g. as the dACC and the amygdala (Walker and Lawson, 2013)), we note that the relatively low spatial resolution and fidelity of EEG imposes substantial limitations and requires validation. The most common approach relies on recordings from frontal channels to measure potential asymmetries in the alpha frequency band (Choi et al., 2011; Hammond, 2005; Peeters et al., 2014; Ramirez et al., 2015; Wang et al., 2019, 2016). This approach assumes that the hyper- and hypoactivation of opposite hemispheres indicate the valence experienced during emotion regulation (Harmon-Jones et al., 2010) and that this marker may reflect symptoms of dysfunctional emotion regulation as commonly observed in depressed patients (Thibodeau et al., 2006). Of interest, neurofeedback protocols that use simultaneous EEG-fMRI recordings showed that EEG frontal asymmetry was correlated with activity in brain structures involved in emotion regulation in healthy subjects (Zotev et al., 2013) and patients suffering from depression (Zotev et al., 2019, 2016). However, the relationship between these biological markers and the cognitive mechanisms for depression is still debatable. In this context, neurofeedback protocols may provide additional validation of these mechanisms (Linden, 2014).

Apart from targeting neural correlates of MDD, neurofeedback training paradigms have also been designed to tap into the interaction between psychological and biological aspects of the disorder (Deldin and Chiu, 2005). For instance, some neurofeedback protocols incorporate aspects from cognitive therapy such as cognitive restructuring approaches and means of emotional self-regulation, including training to self-regulate the response for valenced figures, autobiographical memories, or affective imagery (MacDuffie et al., 2018; Skottnik and Linden, 2019). Thereby, mental imagery based neurofeedback training can potentially aid patients in developing coping strategies to mitigate negative thoughts and value positive experiences (Clark and Beck, 2011). Further, the task engagement itself in combination with contingent positive reinforcement during a neurofeedback session may result in behavioral activation and modulate self-efficacy (Dimidjian et al., 2011), i.e. an individual's sense of being in control of their environment and to cope with challenges (Bandura, 1982; Mehler et al., 2018). Moreover, such psychophysiological effects are particularly relevant for the treatment of depressed patients (but also other psychiatric patient populations) who often show deficits in these capacities. Of interest, it has been noted that psychotherapy approaches and neurofeedback strategies may be mutually translatable (MacDuffie et al., 2018; Skottnik and Linden, 2019). As such, neurofeedback training may provide a promising add-on tool to augment standard care treatment, supporting patients in the process of cognitive restructuring and other learning processes initiated in psychotherapeutic sessions.

Table 1
Eligibility criteria.

Inclusion criteria:
1) Studies presenting original results in human adults (> 18 years old)
2) Studies including patients with a formally diagnosed current depressive episode
Exclusion criteria:
1) Studies including patients with other psychiatric disorders (but not major depressive disorder) in the experimental sample, or targeting depressive symptoms in other disorders
2) Studies exclusively evaluating healthy participants
3) Studies applying biofeedback based only on non-neural signals
4) Studies without voluntary control of brain activity
5) Studies with animal models
6) Case reports (n<5), reviews, commentaries, or editorials

1.4. Aim of this review

Given the growing number of studies investigating neurofeedback applications as a treatment for MDD over the last decades, we aim here to (I) summarize and compare current findings, (II) evaluate the quality of these studies, and (III) provide guidelines for future research that can accelerate the field. Different from previous reviews (Linden, 2014; Sacchet and Gotlib, 2016; Young et al., 2018b), we note that the present study comprises to our knowledge the first attempt of a systematic investigation of EEG and fMRI neurofeedback training protocols in MDD patients. Also, to assess study design and reporting quality, we employed the Joanna Briggs Institute (JBI) critical appraisal tools (Tufanaru et al., 2017) and "Consensus on the Reporting and Experimental Design of Neurofeedback studies" (CRED-nf) checklist (Ros et al., 2020).

2. Methods

2.1. Systematic search

A systematic search on English peer-reviewed journal articles published until March 6th, 2020, was performed for this review. The PubMed bibliographic database, and pre-print servers including life science papers (arXiv, medRxiv, psyArXiv, and bioRxiv) were queried using the following search rule:

(biofeedback OR neurofeedback) AND (depression OR depressive)

Resulting articles were selected or rejected based on the criteria described in Table 1.

As shown in Fig. 1, a total of 577 journal articles were found in the PubMed and pre-print databases, and eight other papers were included from other sources (papers cited in articles screened) (the list of articles is available on: https://osf.io/k76g2/). Through relevance screening, 539 articles were rejected as they did not meet the inclusion criteria. After full-text examination, only 24 articles were included in this systematic review.

To collect relevant information, a data extraction sheet was created including 23 data items which were extracted and grouped into four categories: *Study Design* (diagnostic criteria, symptom scales, existing comorbidities, parallel treatments, randomization, blinding, experimental paradigm, control paradigm, feedback modality, number of sessions, and follow-up), *Clinical Outcomes* (within group differences post-NF, between groups differences post-NF, within group differences at follow-up, between groups at follow-up, exclusions and drop-outs), *Other Significant Outcomes* (within group differences post-NF, between groups differences post-NF, within group differences at follow-up,

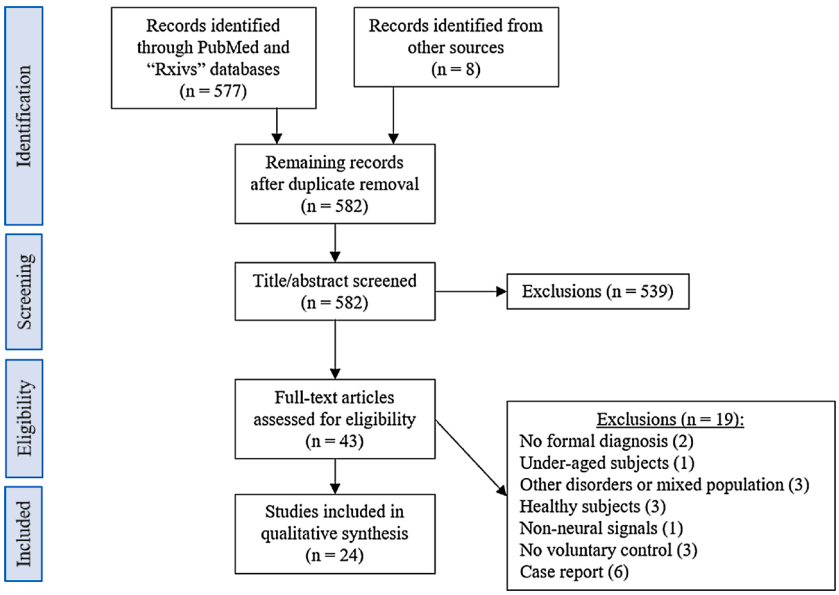

Fig. 1. Search decision flow diagram according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher et al., 2009).

between groups at follow-up). One co-author (LRT extracted data from studies and another co-author screened the extraction results (SHK). Disagreements between the reviewing authors were resolved by discussion.

2.2. Assessment of clinical efficacy

When evaluating clinical outcomes between studies, three other design aspects (besides differences in the experimental paradigm) make it difficult to compare clinical effects as reported across studies: (1) Studies vary greatly in their control conditions, ranging from no control or passive control conditions, where patients merely engage in training mental coping strategies, to active neurofeedback control conditions that are closely matched for various psychosocial factors including reward, successful self-regulation experience, regular interactions with practitioners/researchers and practicing mental coping strategies (Thibault et al., 2018). (2) Studies vary in their inclusion criteria, which may also pertain to baseline severity levels of depressive symptoms. For established treatments in depression, it is well known that baseline differences account for part of the variance of clinical improvements and superiority of treatment over non-specific psychosocial effects (Fournier et al., 2010; Kirsch et al., 2008) and hence differences in baseline severity may bias results. (3) Studies used different outcome measures (i.e., numerical scales). While calculating standardized effect sizes can account for differences in used outcome measure (see 3), they do not account for differences in baseline severity (see 2). To address the first aspect (heterogeneity in control conditions), we grouped studies into one of four categories: no control, passive control, active control outside the scanner, or active control inside the scanner. To address the second and third aspect (i.e., potential baseline differences and heterogeneity in clinical scales), we normalized clinical percentage changes reported for individual studies by the maximum score of the respective clinical scale that was being used (Fig. 2A). Symptom improvement scores were computed as the percentage change of the primary outcome measure with respect to baseline. For studies that did not declare their primary outcome (Choi et al., 2011; Hammond, 2005; Paquette et al., 2009; Peeters et al., 2014; Ramirez et al., 2015; Walker and Lawson, 2013; Wang et al., 2016; Yuan et al., 2014), we considered all clinical outcome measures (if multiple were reported) as secondary and adopted a conservative approach selecting the symptom scale with the least percentage change.

Lastly, to compare clinical effects between neurofeedback and other interventions, we computed the number needed to treat (NNT) for studies reporting remission rates. Specifically, following Altman's recommendation, we refer here to the "number needed to treat for one additional patient to benefit, or to be harmed" (i.e., worsening of depressive symptoms), i.e. NNTB and NNTH, respectively (Altman, 1998), and report these point estimates alongside their 95 % confidence intervals using the Wilson score method. In contrast to the widely used Wald method, the Wilson score method is expected to yield less biased results for studies with relatively small sample sizes or unbalanced designs (Bender, 2001; Newcombe, 1998) and thus seemed more appropriate for the current sample. Calculations were performed modifying a custom written script originally created by Bender (Bender, 2001) using the statistical software "Statistical Analysis System" (SAS, version 9.4). The SAS script that also includes the extracted data for reported remission rates is available: https://osf.io/jw7mu/.

2.3. Assessment of experimental design and reporting quality

To enhance reporting standards in the neurofeedback field, the recently published "Consensus on the Reporting and Experimental Design of Neurofeedback studies" (CRED-nf) checklist suggests "essential" and "suggested" items around design and reporting aspects, including *pre-experiment registration*, *control groups and measures*, *feedback specifications*, *outcome description* and *data storage/publishing* (Ros et al., 2020). Two of the coauthors (LRT and SHK) independently rated the studies included in this review according to the 23 criteria of the CRED-nf checklist. Disagreements between the reviewing coauthors were resolved by discussion.

Moreover, we also assessed the methodological quality of included studies based on the checklist for quasi-experimental studies of the Joanna Briggs Institute (JBI) critical appraisal tools (Tufanaru et al., 2017). The JBI checklist has been used in various experimental fields and thus allows comparing standards between neurofeedback studies but also entire fields. The JBI checklist includes items such as *clarity of cause and effect*, *similar participants*, *similar treatment in compared groups*, *existence of a control group/condition*, *multiple measurement points of the outcome*, *completion of follow-up*, *similar outcome measurements in compared groups*, *reliability of outcome measurements*, *appropriate statistical methods*. The same coauthors who rated studies according to the CRED-nf evaluated the studies included in this review according to the nine criteria of the JBI checklist. Similar to Kohl et al. (Kohl et al., 2020) we adapted some of the CRED-nf items to account for the fact that most

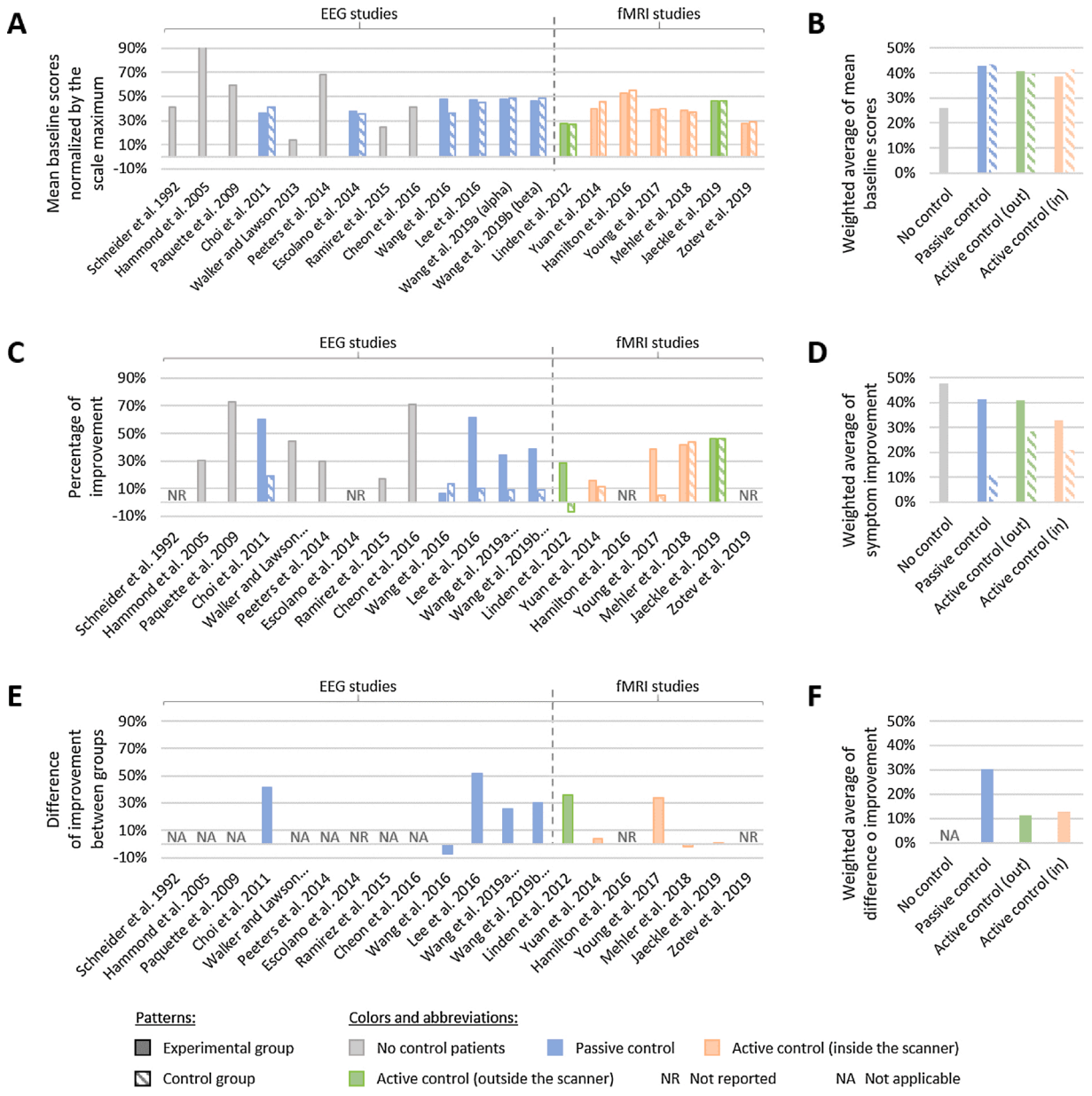

Fig. 2. (A) Depressive symptoms in percentage scores at baseline, normalized by the individual scale maxima. (B) Sample size weighted average percentages of depression severity at baseline after grouping studies according to their control condition category. (C) Percentage of within-group improvement in depressive symptoms at the primary endpoint. (D) Sample size weighted average percentages of within group improvements in depressive symptoms, after grouping studies according to by their control condition category. (E) Between group differences in improvement. (F) Sample size weighted average percentages of within group improvements in depressive symptoms after grouping studies according to their control condition category. Studies evaluating the same database were represented as a single bar. *Wang et al. (2019) is represented by two bars since they tested two different neurofeedback protocols in one single experiment (please refer to Table 2).

studies were published before these guidelines (see Supplementary Material for details).

2.4. Contacting authors

The data extraction sheets, JBI and CRED-nf score tables, and a preprint version of the manuscript were shared with all corresponding authors of the included studies to ask for corrections. Five of 16 (~31 %) corresponding authors replied to our enquiry and either approved the data extraction or suggested minor corrections.

3. Results

Thirteen of the 24 studies included in this review used EEG neurofeedback protocols: six studies including only frontal channel (targeting the structures from the frontal cortex); four studies combining frontal channels with other portions of the scalp; and three studies looking at regions other than the PFC. The remaining eleven studies applied real-time fMRI neurofeedback protocols: six studies targeting the amygdala exclusively; three studies targeting different networks (two studies including emotion processing network, one the salience network); and two studies with two or more distinct regions. Following this review's first aim, and given the heterogeneity of protocols, we first provide a detailed overview of the various experimental protocols used for EEG (Section 3.1) and fMRI (Section 3.2) neurofeedback studies, respectively. In these sections, we emphasize on study designs and training paradigms (Table 2), clinical outcomes at primary endpoint and at follow-up (if reported; Table 3) as well as other statistically significant

Table 2
Overview of experimental designs of studies using neurofeedback protocols in depressive patients (additional clinical details are provided in Table S1 in the Supplementary Material). ↑ = upregulation; ↓ = down-regulation; N = sample size; E = Experimental group; C = Control Group; NR = Not Reported; Studies with overlapping samples are highlighted in gray.

Study	*Training target in Experimental Group (N of allocated subjects / final sample)*	*Control Condition (N of allocated subjects / final sample)*	*Diagnostic criteria*	*Comorbidity (see details in Table S1)*	*Clinical and psychometric outcome measures (baseline levels)*	*Co-occurring Treatment (NF/ controls)*	*Randomization*	*Blinding*	*Number of NF sessions*	*Follow up (in weeks with reference to the primary endpoint)*	*Feedback*
EEG											
Schneider et al. (1992)	↑ or ↓ slow cortical potentials in Cz (Ninit = 8 / Nfinal = 8)	No control patients: Healthy controls ↑ or ↓ slow cortical potentials in Cz (Ninit = 8 / Nfinal = 8)	DSM-III-R	NR	HDRS-17 (NF: 21.3 ± 4.74) GAS (NF: 40.1 ± 7.95) BPRS (NF: 46.5 ± 6.72)	Psychopharm. Medication (8/0) Psychotherapy (NR)	No	No	20 (NF) 5 (controls)	No	Continuous and visual
Hammond (2005)	↑ beta and ↓ alpha and theta in Fp1 and F3 (Ninit = 9 / Nfinal = 8)	No control patients	MMPI	NR	MMPI (NF: 95.75 ± NR)	Psychopharm. Medication (Yes-NR) Psychotherapy (0)	No	No	20.75 (average)	Yes (48 - average)	Non-specified
Paquette et al. (2009)	↓ beta in AF3, AF4, T3 and T4 while ↓ negative thoughts (Ninit = 30 / Nfinal = 27)	No control patients	DSM-IV	Yes	BDI-II (NF: 37.3 ± 9.0) BAI (NF: 18.5 ± 0.3)	Psychopharm. Medication (Yes-NR) Psychotherapy (NR)	No	No	20	Yes (4)	Continuous and visual
Choi et al. (2011)	↑ alpha asymmetry in F3 and F4 (Ninit = 12 / Nfinal = 12)	Passive control: Psychoeducation (Ninit = 12 / Nfinal = 11)	DSM-IV	No	HDRS-17 (NF: 11.33 ± 7.52; C: 12.36 ± 7.67) BDI-II (NF: 22.75 ± 12.35; C: 26.18 ± 16.21) MMPI-2 (NF: 62.08 ± 12.61; C: 67.00 ± 16.07)	Psychopharm. Medication (0/NR) Psychotherapy (1/ 1)	Yes	No	10	Yes (4)	Continuous and audiovisual
Walker and Lawson (2013)	↓ alpha and ↑ beta in Fp2 (Ninit = 183 / Nfinal = 183)	No control patients	QIDS-SR_{16} DSRT	No	QIDS-SR_{16} (NF: 11.8 ± 5.0)	Psychopharm. Medication (0) Psychotherapy (NR)	No	No	6	Yes (48)	Frequency non-specified and auditory
Peeters et al. (2014)	↑ alpha asymmetry in F3 and F4 (Ninit = 9 / Nfinal = 9)	No control patients	DSM-IV	Yes	QIDS-SR_{16} (NF: 18.4 ± 7.2)	Psychopharm. Medication (Yes-NR) Psychotherapy (NR)	No	No	26.78 (average)	No	Continuous and visual
Escolano et al. (2014)	↑ upper alpha band in parieto-occipital channels (Ninit = 50 / Nfinal = 40)	Passive control: Standard care (continued pharmacological treatment) (Ninit = 24 / Nfinal = 20)	DSM-IV	Yes	BDI-II (NF: 23.70 ± 13.51; C: 22.25 ± 11.74) PHQ-9 (NF: 13.33 ± 6.84; C: 15.65 ± 5.96)	Psychopharm. Medication (37/ 18) Psychotherapy (NR)	No	No	8	No	Continuous and visual
Ramirez et al. (2015)	↑ alpha and beta rations in channels AF3, AF4, F3 and F4 during ↑ arousal and valence (Ninit = 10 / Nfinal = 6)	No control patients	Non-specified	NR	BDI (NF: 15.5 ± 9.90)	Psychopharm. Medication (NR) Psychotherapy (0)	No	No	10	No	Continuous and auditory (music)
Cheon et al. (2016)	↑ beta at F3 and ↓ alpha/theta ration in Pz (Ninit = 20 / Nfinal = 20)	No control patients	DSM-IV-TR	No	HDRS-23 (NF: 21.38 ± 5.82) HARS (NF: 19.43 ± 8.70) BDI-II (NF: 25.25 ± 7.91) BAI (NF: 19.75 ±	Psychopharm. Medication (12) Psychotherapy (NR)	No	No	16–24	No	Frequency non-specified and audiovisual

(*continued on next page*)

Table 2 (*continued*)

Study	*Training target in Experimental Group (N of allocated subjects / final sample)*	*Control Condition (N of allocated subjects / final sample)*	*Diagnostic criteria*	*Comorbidity (see details in Table S1)*	*Clinical and psychometric outcome measures (baseline levels)*	*Co-occurring Treatment (NF/ controls)*	*Randomization*	*Blinding*	*Number of NF sessions*	*Follow up (in weeks with reference to the primary endpoint)*	*Feedback*
					12.76) CGI (NF: 3.79 ± 1.30)						
Wang et al. (2016)	↑ alpha asymmetry in F3 and F4 (Ninit = 7 / Nfinal = 7)	Passive control: Standard care (continued pharmacological treatment) (Ninit = 7 / Nfinal = 7)	DSM-V	No	BDI-II (NF: 30.14 ± 10.25; C: 22.86 ± 13.03) BAI (NF: 17.86 ± 10.51; C: 16.00 ± 9.92)	Psychopharm. Medication (6/6) Psychotherapy (0/0)	Yes	No	6	No	Non-specified
Lee et al. (2019)	Self-regulate SMR or beta band in F3, T3, or T4 (according to symptoms), followed by ↓ alpha/theta ration in Pz (Ninit = 12 / Nfinal = 12)	Passive control: Standard care (continued pharmacological treatment) and placebo psychotherapy (supportive psychotherapy) (Ninit = 12 / Nfinal = 12)	DSM-IV-TR	No	HDRS-17 (NF:24.33 ± 5.77; C:23.17 ± 5.42) BDI-II (NF: 36.67 ± 14.79; C: 25.83 ± 7.99) CGI-S (NF: 4.75 ± 0.62; C: 4.17 ± 0.83)	Psychopharm. Medication (12/12) Psychoterphy (NR)	No	No	12–24	No	Frequency non-specified and audiovisual
Wang et al. (2019)	↑ alpha asymmetry in F3 and F4 (Ninit = 30 / Nfinal = 24), or ↓ beta in P3 and P4 (Ninit = 26 / Nfinal = 23)	Passive control: Standard care (continued pharmacological medication) (Ninit = 31 / Nfinal = 23)	DSM-IV	Yes	BDI-II (NFa: 30.25+-8.39; NFb: 29.17+-11.47; C: 30.44+-9.31 BAI (NFa: 21.33+-12.22; NFb: 21.52+-9.62; C: 22.04+-10.32)	Psychopharm. Medication (47/23) Psychoterphy (NR)	No	No	10	No	Continuous and audiovisual
Chen and Lin (2020) (sample partially overlapped with Wang et al. (2019))	↓ beta in P3 and P4 (Ninit = 26 / Nfinal = 23)	No control patients		Yes	See "NFb" values in Wang et al. (2019)	Psychopharm. Medication (23) Psychotherapy (NR)					
fMRI											
Linden et al. (2012)	↑ areas involved in positive emotions during mental imagery of positive emotions (Ninit = 8 / Nfinal = 8)	Active control (out): Patients performing mental imagery of positive emotions outside the scanner (Ninit = 8 / Nfinal = 8)	DSM-IV	No	HDRS-17 (NF: 14.38 ± NR; C: 13.88 ± NR) HDRS-21 (NF: 18.12 ± NR; C: 17.75 ± NR)	Psychopharm. Medication (8/8) Psychotherapy (NR)	No	No	4	No	Continuous and visual
Young et al. (2014) (samples partially overlapped with Yuan et al. (2014))	↑ of amygdala during affective memory recall (Ninit = 14 / Nfinal = 13)	Active control (in): Patients receiving feedback from non-related control region (Ninit = 7 / Nfinal = 6)	DSM-IV-TR	Yes	HDRS-21 (NF: 19.9 ± 5.15; C: 23.9 ± 5.49) MADRS (NF: 27.1 ± 6.69; C: 31.4 ± 6.71) HARS (NF: 19.1 ± 5.32; C: 23.3 ± 7.74) STAI	Psychopharm. Medication (0/0) Psychotherapy (NR)	No	Yes (double-blind)	1	No	Continuous and visual
Yuan et al. (2014)	↑ of amygdala during affective memory recall (Ninit = 14 / Nfinal = 14)	Active control (in): Patients receiving feedback from non-related control region (Ninit = 13 / Nfinal =		Yes	HDRS-21 (NF: 20.64 ± 4.63; C: 23.69 ± 4.96; H: 23.69 ± 4.96) HARS (NF: 19.93 ± 5.15; C: 22.15 ± 7.02;					Yes (0.3–4)	

(*continued on next page*)

Table 2 (*continued*)

Study	*Training target in Experimental Group (N of allocated subjects / final sample)*	*Control Condition (N of allocated subjects / final sample)*	*Diagnostic criteria*	*Comorbidity (see details in Table S1)*	*Clinical and psychometric outcome measures (baseline levels)*	*Co-occurring Treatment (NF/ controls)*	*Randomization*	*Blinding*	*Number of NF sessions*	*Follow up (in weeks with reference to the primary endpoint)*	*Feedback*
		13), and healthy subjects (Ninit = 27 / Nfinal = 27)			H: 1.31 ± 2.02) MADRS						
Zotev et al. (2016) (samples partially overlapped with Yuan et al. (2014))	↑ of left amygdala during affective memory recall (Ninit = 14 / Nfinal = 13)	Active control (in): Patients receiving feedback from non-related control region (Ninit = 13 / Nfinal = 11)		Yes	HDRS-21 (NF: 20.5 ± 4.0; C: 20.9 ± 3.3) MADRS (NF: 27.4 ± 6.8; C: 28.5 ± 3.0) HARS (NF: 17.5 ± 4.7; C: 19.3 ± 5.2) STAI	Not clear (new session in addition to the one reported by Young et al. (2014))		Not clear (new session in addition to the one reported by Young et al. (2014))	2	No	
Hamilton et al. (2016)	↓ reactivity of a node of the salience network (Ninit = 12 / Nfinal = 10)	Active control (in): Patients receiving yoked feedback from NF group (Ninit = 10 / Nfinal = 10)	DSM-IV	Yes	BDI-II (NF: 33.3 ± 2.3; C: 34.6 ± 4.0)	Psychopharm. Medication (6/4) Psychotherapy (NR)	No	Yes (double-blind)	1	No	Visual at the end of the trial
Young et al. (2017b) Young et al. (2017a) (samples partially overlapped with Young et al. (2017b)) Young et al. (2018a) (samples partially overlapped with Young et al. (2017b))	↑ of amygdala during affective memory recall (Ninit = 19 / Nfinal = 18)	Active control (in): Patients receiving feedback from non-related control region (Ninit = 17 / Nfinal = 15)	DSM-IV-TR	Yes	BDI-II (NF: 27.2 ± 10.7; C: 26.6 ± 13.4) SHAPS MADRS (NF: 23.5 ± 9.9; C: 23.8 ± 6.7) HDRS-21 (NF: 19.4 ± 7.9; C: 19.1 ± 4.4) HARS (NF: 18.8 ± 7.5; C: 18.1 ± 6.3)	Psychopharm. Medication (0/0) Psychotherapy (NR)	Yes	Yes (double-blind)	2	Yes (1)	Continuous and visual
Mehler et al. (2018)	↑ of areas involved in positive emotions during mental imagery of positive emotions (Ninit = 21 / Nfinal = 16)	Active control (in): Patients ↑ areas involved in scene processing during mental imagery of scenes (Ninit = 22 / Nfinal = 16)	DSM-IV	No	HDRS-17 (NF: 19.88 ± -3.65; C: 19.09 ± 5.09) HADS-A (NF: 12.69 ± 3.84; C: 12.63 ± 4.13) HADS-D (NF: 13.06 ± 3.43; C: 12.44 ± 4.35)	Psychopharm. Medication (16/16) Psychotherapy (0/0)	Yes	Yes (single-blind)	5	Yes (6)	Continuous and visual
Jaeckle et al. (2019)	↑ correlation between the right superior anterior temporal lobe and the posterior subgenual cortex during affective memory recall (Ninit = 22 / Nfinal = 19)	Active control (out): Cognitive reappraisal techniques outside the scanner (Ninit = 21 / Nfinal = 16)	DSM-V	Yes	BDI-II (29.14 ± 8.66)ˆ MADRS (22.84 ± 6.97)ˆ QIDS-SR_{16} (16.79 ± 6.53)ˆ ˆno details per group reported	Psychopharm. Medication (10/10) Psychotherapy (0/0)	Yes	Yes (single-blind)	3	No	Continuous and visual
Zotev et al. (2019)	↑ of fMRI (left ACC and Amygdala) and EEG (alpha and beta	Active control (in): Patients receiving feedback unrelated to	DSM-IV-TR	NR	HDRS-21 (NF: 14.4 ± 7.0; C: 15.1 ± 4.9) MADRS (NF: 19.6 ±	Psychopharm. Medication (0/0)	No	Yes (single-blind)	1	No	Continuous and visual

(*continued on next page*)

Table 2 (*continued*)

Study	*Training target in Experimental Group (N of allocated subjects / final sample)*	*Control Condition (N of allocated subjects / final sample)*	*Diagnostic criteria*	*Comorbidity (see details in Table S1)*	*Clinical and psychometric outcome measures (baseline levels)*	*Co-occurring Treatment (NF/ controls)*	*Randomization*	*Blinding*	*Number of NF sessions*	*Follow up (in weeks with reference to the primary endpoint)*	*Feedback*
	asymmetry in F3 and F4) during affective memory recall (Ninit = 16 / Nfinal = 16)	brain activity (artificially generated signals) (Ninit = 8 / Nfinal = 8)			10.7; C: 20.5 ± 5.7) HARS (NF: 13.2 ± 7.5; C: 16.1 ± 6.4) STAI (NF: 56.9 ± 9.9; C: 59.6 ± 9.6)	Psychotherapy (NR)					

Abbreviations: BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; BPRS = Brief Psychiatric Rating Scale; CGI = Clinical Global Impression; DSM = Diagnostic and Statistical Manual of Mental Disorders; DSRT = Depression Self-Rated Test; GAS = Global Assessment Scale; HDRS = Hamilton Depression Rating Scale; HARS = Hamilton Anxiety Rating Scale; MADRS = Montgomery-Asberg Depression Rating Scale; MMPI = Minnesota Multiphasic Personality Inventory; PHQ = Patient Health Questionnaire-Depression; QIDS-SR_{16} = Quick Inventory of Depressive Symptoms - Self-Report Version; STAI = State/Trait Anxiety Scale; Psychopharm. Medication = Psychopharmacological Medication; Clinical/psychometric outcome measures reported as baseline mean ± standard-deviation.

cognitive or neural effects (if evaluated; Table 4). As far as reported by the authors of the original studies, we also extracted information about co-occurring standard treatment for depression, including psychopharmacological medicine (patients are referred to as medicated) and psychotherapy.

We follow with an investigation of clinical effects grouped by control condition (Section 3.3) as well as drop-out rates and side-effects (if reported; Section 3.4). Specifically, we compared baseline scores and changes of the primary outcome measure (or if not declared, we used the secondary outcome measure with the least improvement; Fig. 2) to evaluate clinical efficacy. We further computed the number needed to treat for one additional patient to benefit (or to be harmed) [NNTB/NNTH] (Altman, 1998) for studies that declared a primary outcome measure and reported remission rates (Table 5).

Lastly, following this review's second aim, we report quality scores for study design and reporting (Fig. 3) and discuss these findings in the context of best practice recommendations (Section 3.5).

3.1. EEG neurofeedback paradigms and clinical effects

In the first controlled, non-blinded pilot EEG neurofeedback study in MDD, Schneider et al. (1992) compared the ability of medicated patients (N = 8) and healthy controls (N = 8) to regulate slow cortical potentials (SCP) at the central electrode (Cz). The patient group showed significantly higher control of the system, and the authors reported a negative correlation between on-task SCP and the onset of illness, and a correlation in the opposite direction between SCP and the number of hospitalizations pre training (Schneider et al., 1992). However, no clinical changes after the neurofeedback training were reported by the authors.

Later, neurofeedback researchers became interested in protocols that exploited spectral lateralization observed in frontal electrodes in response to mood induction (Harmon-Jones et al., 2010; Palmiero and Piccardi, 2017). Four neurofeedback studies reported frontal alpha asymmetry as the main feature. Alpha asymmetry is calculated as the difference in the alpha spectral power between left and right frontal channels F3 and F4. The first randomized, controlled and non-blinded study that used this approach in non-medicated patients compared changes in (self-rated and clinician-rated) depression scales between a group that engaged in EEG frontal alpha asymmetry neurofeedback training (N = 12) and a control group that received psychoeducation (N = 11) (Choi et al., 2011). After 10 training sessions, the neurofeedback group showed a significant improvement of the depressive symptoms (reduction of more than 60 % for the HDRS-17 and BDI-II scores), which persisted in the neurofeedback group at one-month follow-up (no follow-up reported for the control group). However, one main limitation of the study was the lack of blinding and the fact that patients were already partly remitted at enrolment (five in the experimental and two in the control group). Another research group later employed an uncontrolled, single-arm study with nine patients (medication status unclear) suffering from moderate to severe depression who underwent a similar training for a maximum of 30 sessions (three per week) (Peeters et al., 2014). Partial (defined as at least 50 % reduction) and total remission (a score of ≤ 6) were reported for four and one patient based on self-rated depression scale (QIDS-SR_{16}). Moreover, this study reported a significant correlation between symptom improvement and the normalization of the frontal alpha balance.

A similar alpha asymmetry EEG training protocol was tested later by Wang et al. (2016) in a randomized non-blinded controlled pilot study with medicated patients. However, the authors found no significant difference between patients performing neurofeedback (N = 7) and those undergoing psychoeducation (N = 7). Later, the same research group expanded this design by another control group and recruited additional patients. Specifically, they compared in a non-randomized, non-blinded two-arm follow-up study the efficacy of alpha asymmetry neurofeedback (N = 24) and beta parietal asymmetry neurofeedback training (N = 23) in medicated patients. Wang et al. (2019) found that

Table 3
Overview of main clinical outcomes from studies using neurofeedback protocols in depressive patients. ↑ = increased; ↓ = reduced; * = statistically significant effect; [1] = differences between the post-experiment measurement in each group and the pre-experiment measurement in the merged sample. Studies with overlapping samples are highlighted in gray. The primary outcome (if declared) is highlighted in bold.

Studies	*Clinical Improvement*				*Drop-outs or exclusions (NF/C)*
	Post-NF (compared to baseline)		*Follow up (compared to baseline)*		
	Within groups	*Between groups*	*Within groups*	*Between groups*	
EEG					
Schneider et al. (1992)	Not reported	Not reported	Not applicable	Not applicable	Not reported
Hammond (2005)	*NF:* ↓ MMPI-2 (30%)	Time x group interaction*	Not reported	Not reported	Unmotivated (1/0)
Paquette et al. (2009)	*NF:* ↓ BDI-II (72.9%)* ↓ BAI (58.9%)*	Not applicable	Not reported	Not applicable	Tiredness (3)
Choi et al. (2011)	*NF:* ↓ HDRS-17 (64.0%)* ↓ BDI-II (60.1%)* *Controls:* ↓ HDRS-17 (10.4%) ↓ BDI-II (18.75%)	*NF > Controls:* ↓ BDI-II* ↓ HDRS-17*	Not reported	Not reported	Logistics (0/1)
Walker and Lawson (2013)	*NF:* ↓QIDS-SR_{16} (44.1%)*	Not applicable	*NF:* ↓QIDS-SR_{16} (55.1%)*	Not applicable	Not reported
Peeters et al. (2014)	*NF:* ↓ QIDS-SR_{16} (29.5%)*	Not applicable	Not applicable	Not applicable	Not reported
Escolano et al. (2014)	Not reported	Not reported	Not applicable	Not applicable	Inability to perform the cognitive assessments (4/1) Excessive noise (6/3)
Ramirez et al. (2015)	*NF:* ↓ BDI (17.2%)	Not applicable	Not applicable	Not applicable	Illness (4)
Cheon et al. (2016)	*NF:* ↓ **HDRS-17 (70.9%)*** ↓ HARS (69.1%)* ↓ BDI-II (42.0%)* ↓ BAI (41.1%)* ↓ CGI (49.1%)*	Not applicable	Not applicable	Not applicable	Adverse events of medication (1) Tiredness (1) Logistics (1) Lost to follow up (2)
Wang et al. (2016)	*NF:* ↓ BAI (21.74%) and BDI-II (18.18%) for responders ↑ BAI (124.27%) and BDI-II (24.97%)* for non-responders *Controls:* ↓ BAI (28.33%) and BDI-II (27.63%) for responders ↑ BAI (11.54%) for non-responders	No significant effects	Not applicable	Not applicable	Not reported
Lee et al. (2019)	*NF:* ↓ **HDRS-17 (61.65%)*** ↓ BDI-II (53.64%)* ↓ CGI (38.53%)* *Controls:* ↓ **HDRS-17 (10.06%)*** ↓ BDI-II (8.36%) ↓ CGI (0.00%)	*NF > Controls:* ↓ HDRS-17* ↓ BDI-II* ↓ CGI*	Not applicable	Not applicable	Not reported
Wang et al. (2019)	*NFa:* ↓ BDI-II (34.45%)* ↓ BAI (38.28%)* *NFb:* ↓ BDI-II (38.88%)* ↓ BAI (43.23%)* *Controls:* ↓ BDI-II (8.74%) ↓ BAI (-0.98%)	*NFa > Controls:* **Group x session interaction for BAI*** *NFb > Controls:* **Group x session interaction for BDI-II***	Not applicable	Not applicable	Non-specified reason for dropping out after allocation (6/3/8)
Chen and Lin (2020) (sample partially overlapped with Wang et al. (2019))	See Wang et al. (2019)	See Wang et al. (2019)	See Wang et al. (2019)	See Wang et al. (2019)	See Wang et al. (2019)
fMRI					
Linden et al. (2012)	*NF:* ↓ HDRS-17 (28.7%)* *Controls:* ↑ HDRS-17 (7.2%)	**Group x session interaction for HDRS17***	Not applicable	Not applicable	Not reported
Young et al. (2014) (samples partially overlapped with Yuan et al. (2014))	See Yuan et al. (2014)	See Yuan et al. (2014)	See Yuan et al. (2014)	See Yuan et al. (2014)	Tiredness (1/1)
Yuan et al. (2014)	NF: ↓ HDRS-21 (15.6%) ↓ HARS (18.5%)* Controls (MDD): ↓ HDRS-21 (11.7%) ↓ HARS (18.7%)*	Not reported	Not reported	Not reported	Not clear (new control participants in addition to the sample reported by Young et al. (2014))
Zotev et al. (2016) (samples partially overlapped with Yuan et al. (2014))	See Yuan et al. (2014)	See Yuan et al. (2014)	See Yuan et al. (2014)	See Yuan et al. (2014)	See Yuan et al. (2014)
Hamilton et al. (2016)	Not reported	Not reported	Not applicable	Not applicable	No response (2/0)
Young et al. (2017b)	NF: ↓ MADRS (38.7%)* ↓ BDI-II (32.4%)* ↓ HDRS-21 (34.0%)* ↓ HARS (25.0%)* *Controls:* ↓ MADRS (5.0%) ↓ BDI-II (4.9%) ↓ HDRS-21 (10.0%) ↓ HARS (7.18%)	*NF > Controls:* **Group x session interaction for MADRS***, BDI-II* and HDRS-21*	*NF:* ↓ MADRS (49.4%)* ↓ BDI-II (40.8%)* ↓ HDRS-21 (46.4%)* ↓ HARS (34.6%)* *Controls:* ↓ MADRS (8.0%) ↓ BDI-II (8.6%) ↓ HDRS-21 (9.9%) ↓ HARS (23.2%)*	NF > Controls: ↓ MADRS* ↓ BDI-II* ↓ HDRS-21* ↓ HARS	Discomfort (1/1) Excessive noise (0/1)
Young et al. (2017a) (samples partially	See Young et al. (2017b)	See Young et al. (2017b)	See Young et al. (2017b)	See Young et al. (2017b)	See Young et al. (2017b)

(*continued on next page*)

Table 3 (*continued*)

Studies	Clinical Improvement				Drop-outs or exclusions (NF/C)
	Post-NF (compared to baseline)		Follow up (compared to baseline)		
	Within groups	Between groups	Within groups	Between groups	
overlapped with Young et al. (2017b))					
Young et al. (2018a) (samples partially overlapped with Young et al. (2017b))	See Young et al. (2017b)	See Young et al. (2017b)	See Young et al. (2017b)	See Young et al. (2017b)	See Young et al. (2017b)
Mehler et al. (2018)	NF: ↓ HDRS-17 (42.0%)* ↓ HADS-A (14.0 %)↓ HADS-D (23 %) Controls: ↓ HDRS-17 (43.7%)* ↓ HADS-A (25 %)↓ HADS-D (31 %)	**No significant Group x session interaction for HDRS-17**	NF: ↓ HDRS-17 (48.5%)* ↓ HADS-A (30 %)↓ HADS-D (35 %) Controls: ↓ HDRS-17 (60.4)* ↓ HADS-A (39 %)↓ HADS-D (34 %)	No significant effects	Personal reasons (4/6) Discomfort (1/0) Lost to follow up (3/1)
Jaeckle et al. (2019)	NF: ↓ BDI-II (46.2 %)*[1] ↓ MADRS (37.1%)*[1] ↓ QIDS-SR_{16} (39.5%)*[1] Controls: ↓ BDI-II (46.0 %)*[1] ↓ MADRS (31.9%)*[1] ↓ QIDS-SR_{16} (35.2)*[1]	**No significant Group x session interaction for BDI-II**	Not applicable	Not applicable	Feeling unwell to continue (1/2) Logistics (1/2) Adverse effects (insomnia - 1/0) Worsening of symptoms (0/1)
Zotev et al. (2019)	Not reported	Not reported	Not applicable	Not applicable	Not reported

Abbreviations: BAI = Beck Anxiety Inventory; BDI-II = Beck Depression Inventory - Version 2; CGI = Clinical Global Impression; HDRS = Hamilton Depression Rating Scale; HARS = Hamilton Anxiety Rating Scale; HADS-A = Hospital Anxiety and Depression Scale (Anxiety Subscale); HADS-D = Hospital Anxiety and Depression Scale (Depression Subscale); MADRS = Montgomery-Asberg Depression Rating Scale; MMPI = Minnesota Multiphasic Personality Inventory; QIDS-SR_{16} = Quick Inventory of Depressive Symptoms - Self-Report Version.

both groups showed significant improvement in depressive (more than 10 %) and anxiety related (around 9%) symptoms. When compared to a control group receiving placebo therapy (N = 23), the alpha neurofeedback showed significant improvement of anxiety symptoms, while the beta neurofeedback showed a significant effect of depressive symptoms (Wang et al., 2019). Exploratory post-hoc analyses suggested that changes in depressive symptoms (BDI scale) were positively correlated with the beta variation in left and right parietal electrodes P3 and P4 (Chen and Lin, 2020).

Two independent studies followed a different approach combining the frontal alpha asymmetry with other EEG features: in an uncontrolled, non-blinded single-arm pilot study, Hammond (2005) combined beta up-regulation with alpha and theta down-regulation over left and right frontal electrode channels F3 and F4 in a small sample (N = 8). Seven patients showed improvement in the Minnesota Multiphasic Personality Inventory (MMPI) scale (with an average reduction of about 30 %). Of interest, these improvements in depressive symptoms were largely maintained at follow-up one year later (Hammond, 2005). Another uncontrolled, unblinded single-arm pilot study explored the up-regulation of alpha and beta ratios in frontal channels (AF3, AF4, F3 and F4) during increased arousal and valence in depressed elderly patients (N = 6) (Ramirez et al., 2015). After ten training sessions, patients presented approximately 17 % improvement of self-rated BDI scores. We note that both studies did not provide detailed information about other concomitant treatment (Table 2).

Walker and Lawson (2013) used a different approach to measure lateralization in an uncontrolled single-arm study at a non-academic institution (a commercial EEG neurofeedback clinic). They enrolled 183 MDD patients that showed no sufficient improvement after previous psychopharmacological treatment (i.e., were considered treatment resistant). This sample constitutes the largest thus far collected in the depression neurofeedback literature. However, overall, the report and documentation were very brief, no further information about patients (e.g., forms of compensation) was reported and a potential conflict of interest was not declared. Patients were trained to reduce spectral power in theta and increase spectral power in beta frequencies at the right frontopolar channel (FP2) during six training sessions. The underlying assumption of this training protocol was that modulations of these frequencies in the intended direction would mimic effects of deep brain stimulation in Brodmann area 25 and entrain inhibitory effects on the amygdala non-invasively (Walker and Lawson, 2013). However, no source localization analysis was performed. This intervention led to significant reductions of self-reported symptoms for more than 80 % of patients (group average showed a reduction of approximately 44 %) (Walker and Lawson, 2013).

Moving to other areas of the scalp, three studies combined data from frontal channels and temporal and posterior channels in their protocols. The uncontrolled single-arm study by Paquette et al. (2009) focused on the reduction of high-beta power in fronto-temporal channels (AF3, AF4, T3 and T4) while inhibiting negative thoughts. After 20 sessions, medicated patients (N = 27) presented an approximately 73 % reduction of BDI symptoms. Further, 20 patients did not meet the DSM-IV criteria for MDD anymore. One month after the end of the treatment, source localization analysis found reduced beta frequencies in emotion-related brain areas including the orbitofrontal cortex, temporal lobe, amygdala and cingulate cortex (Paquette et al., 2009). Further, Escolano et al. (2014) performed a non-randomized non-blinded trial in medicated patients, comparing an experimental group (N = 40) that underwent alpha power upregulation training over parieto-occipital channels (performing mental arithmetic) with a control group that received continued standard care (N = 20) in which patients received only their prescribed psychopharmacological medication. Behavioral outcomes suggested that the intervention group showed increased alpha EEG power and improved cognitive symptoms (working memory) (Escolano et al., 2014). However, the study did not report any changes of clinical changes despite assessing these at baseline.

In an uncontrolled single-arm study, Cheon et al. (2016) included an experimental group (N = 20) of medicated patients that trained to up-regulate beta power at F3 and down-regulate the alpha/theta ratio in the Pz electrode. After 16–24 sessions, patients presented approximately 70 % of reduction in both the Hamilton Depression Rating Scale (HDRS) and the Hamilton Anxiety Rating Scale (HARS). Later, the same research group adapted the same approach to subject-dependent protocols that were calibrated based on patients' symptoms and could include the self-regulation of sensorimotor rhythms or beta band in the electrodes F3, T3, or T4, followed by down-regulation of the alpha/theta ratio in the Pz electrode (Lee et al., 2019). The choice of the best protocol was based on a previous study evaluating the efficacy of different protocols

Table 4
Overview of other significant outcomes from studies using neurofeedback protocols in depressive patients. ↑ = increased; ↓ = reduced; + = positive; - = negative. Studies with overlapping samples are highlighted in gray.

Studies	*Other significant outcomes*			
	Post-NF (compared to baseline)		*Follow up (compared to baseline)*	
	Within groups	*Between groups*	*Within groups*	*Between groups*
EEG				
Schneider et al. (1992)	*NF:* - corr.: SCP control and onset of illness + corr.: SCP control and number of hospitalizations	*NF > Controls:* Control of SCP	Not applicable	Not applicable
Hammond (2005)	Not Reported	Not applicable	Not reported	Not applicable
Paquette et al. (2009)	*NF:* ↓ frequency of worries, frequency of negative automatic thoughts, frequency of rumination sadness, dysfunctional attitudes, behavioral inhibition ↑ frequency of positive automatic thoughts	Not applicable	NF: ↓ beta activity in orbitofrontal cortex, insula, amygdala, temporal pole and cingulate cortex + corr. (uncorrected): ↓ BDI-II and ↓ beta activity in orbitofrontal and cingulate cortices	Not applicable
Choi et al. (2011)	*NF:* ↑ accuracy in the verbal fluency task ↓ reaction time for congruent and incongruent stimuli in the Stroop task ↑ of alpha asymmetry *Controls:* No significant effects	Significant time x group interactions	*NF:* Sustained clinical, physiological, and neuropsychological improvements (values not reported)	Not reported
Walker and Lawson (2013)	Not reported	Not applicable	Not reported	Not applicable
Peeters et al. (2014)	NF: - corr.: ↓ QIDS-SR_{16} and alpha asymmetry	Not applicable	Not applicable	Not applicable
Escolano et al. (2014)	*NF:* ↓ number of errors and reaction time ↑ power in the working memory task No significant effects in alpha asymmetry *Controls:* No significant effect	Group x time interaction for the working memory task	Not applicable	Not applicable
Ramirez et al. (2015)	Not reported	Not applicable	Not applicable	Not applicable
Cheon et al. (2016)	No significant effects in alpha asymmetry	Not applicable	Not applicable	Not applicable
Wang et al. (2016)	No significant effects	No significant effects	Not applicable	Not applicable
Lee et al. (2019)	*NF:* ↑ EQ-5D-5 L and ↓ SDS *Controls:* ↓ EQ-5D-5 L and ↑ SDS	*NF > Controls:* ↑ EQ-5D-5 L *Controls > NF:* ↑ SDS	Not applicable	Not applicable
Wang et al. (2019)	*NFb:* ↓ P3 high-beta power *Controls:* ↑ P3 high-beta power	No significant effects	Not applicable	Not applicable
Chen and Lin (2020) (sample partially overlapped with Wang et al. (2019))	*NF:* ↓ beta, but not other bands, in P3 and P4 + corr. between ↓ BDI-II and ↓ beta in P3 and P4	Not applicable	Not applicable	Not applicable
fMRI				
Linden et al. (2012)	*NF:* ↓ POMS ↑ bilateral ventral striatum and left extra-striate visual cortex activity + corr.: up-regulation and HDRS *Controls:* ↓ POMS	*Controls > NF:* PANAS-NA	Not applicable	Not applicable
Young et al. (2014) (samples partially overlapped with Yuan et al. (2014))	*NF:* ↓ STAI trait and state anxiety ↓ POMS-depression and anger ↓ VAS restlessness, anxiety and irritability ↑ VAS-happiness ↑ left amygdala activity + linear trend across all runs *Controls:* ↓ VAS-sadness	*NF > Controls:* VAS-happiness Amygdala activity '*Controls > NF:* STAI state anxiety	Not applicable	Not applicable
Yuan et al. (2014)	*NF:* - corr.: amygdala-cuneus connectivity and HDRS	*Controls (healthy) > NF and Controls (MDD):* Amygdala-ACC and amygdala-cuneus connectivity before NF, but not after	*NF:* + corr.: amygdala-cuneus connectivity and the time to follow-up	Not reported
Zotev et al. (2016) (samples partially overlapped with Yuan et al. (2014))	*NF:* ↓ POMS depression, total mood disturbance ↑ VAS happiness + corr.: amygdala activity and self-reported happiness, and memory-recall, and VAS-happiness - correlation: amygdala activity and POMS-tension, and TAS-total + corr.: amygdala laterality and TAS-total + corr.: EEG asymmetry and HDRS and SHAPS-anhedonia + corr.: EEG asymmetry and amygdala laterality	Not reported	Not applicable	Not applicable
Hamilton et al. (2016)	*NF:* ↓ salience network node response ↓ emotional reactivity to negative IAPS	*NF > Controls:* Reduction in responses to IAPS negative pictures, and in negative SRET	Not applicable	Not applicable

(*continued on next page*)

Table 4 (*continued*)

Studies	*Other significant outcomes*			
	Post-NF (compared to baseline)		*Follow up (compared to baseline)*	
	Within groups	*Between groups*	*Within groups*	*Between groups*
	↓ in negative SRET *Controls:* No significant effects			
Young et al. (2017b)	NF: ↓ SHAPS ↑ recall of positive specific and overall specific memories ↓ recall of categorical positive, overall categorical, extended positive, extended negative, and overall extended memories + corr.: MADRS and amygdala activity during the final transfer run *Controls:* No significant effects	*NF > Controls:* Recall of specific extended memories, and positive specific memories *Controls > NF:* SHAPS Recall of categorical extended memories, and positive categorical and extended memories	*NF:* ↓ SHAPS* *Controls:* No significant effects	No significant effects
Young et al. (2017a) (samples partially overlapped with Young et al. (2017b))	Not reported	Not reported	*NF:* ↓ amygdala activity during n response to sad faces ↑ amygdala activity during n response to happy faces ↓ reaction time for positive faces ↓ reaction time for positive words ↑ vigilance to positive faces ↓ vigilance to negative faces *Controls:* No significant effects	*NF > Controls:* Amygdala activity during HN-NN condition Vigilance to positive faces *Controls > NF:* Amygdala activity during SN-NN condition Reaction time for positive faces Reaction time for positive words Vigilance to negative faces
Young et al. (2018a) (samples partially overlapped with Young et al. (2017b))	Not reported	*NF > Controls:* Amygdala connectivity with prefrontal cortical, striatal and subcortical regions during memory recall, and limbic regions at rest *Controls > NF:* Amygdala connectivity with right temporal lobe during positive memory recall, and bilateral temporal pole at rest	Not reported	Not reported
Mehler et al. (2018)	*NF :* ↑ ROIs across sessions *Controls:* ↑ ROIs across sessions *Both groups:* + corr.: between HDRS-17 improvement (corrected for confounds) and improvement in self-efficacy scores	No significant effects	No significant effects	No significant effects
Jaeckle et al. (2019)	*NF:* ↓ connectivity between the right superior anterior temporal lobe and the posterior subgenual cortex *Both groups:* ↓ POMS-depression dejection, Rosemberg self-esteem scale ↓ self-blame ratings during anger content ↑ self-esteem ratings - corr.: between differences in self-esteem ratings and BDI reduction	No significant effects	Not applicable	Not applicable
Zotev et al. (2019)	*NF:* ↓ POMS depression, confusion, and total mood disturbance ↑ VAS happiness ↑ alpha and beta asymmetry, and left amygdala activity during NF ↑ left amygdala-ACC connectivity + corr. between alpha asymmetry and MADRS-trait depression, and SHAPS-anhedonia - corr. between alpha asymmetry and delta POMS-state depression and POMS-total mood disturbance *Controls:* No significant effects	*NF > Controls:* left amygdala-ACC connectivity	Not applicable	Not applicable

Abbreviations: EQ-5D-5 L = 5-level version of European Quality of Life Questionnaire 5-Dimensional Classification; BDI-II = Beck Depression Inventory - Version 2; HDRS = Hamilton Depression Rating Scale; HN-NN = happy/neutral - neutral/neutral faces; IAPS = International Affective Picture System; MADRS = Montgomery-Asberg Depression Rating Scale; PANAS-NA = Positive Affect Negative Affect Schedule - Negative Affect; POMS = Profile of Mood States; QIDS-SR_{16} = Quick Inventory of Depressive Symptoms - Self-Report Version; SN-NN = sad/neutral - neutral/neutral faces; SCP = slow cortical potentials; SDS = Sheehan Disability Scale; SHAPS = Snaith–Hamilton Pleasure Scale; SRET = self-referent encoding task; STAI = State/Trait Anxiety Scale; TAS = Toronto Alexithymia Scale; VAS = Visual Analog Scale.

to reduce specific symptoms (for example, low attention, low self-esteem, high depression, or high anxiety) across different psychiatric disorders (Cheon et al., 2015). The study was non-randomized and not blinded. All patients continued pharmacological medication treatment during the course of the study. While the experimental group (N = 12) received neurofeedback training as augmentation, the control group (N = 12) received supportive psychotherapy. Results suggested that patients in the experimental group showed significantly more clinical improvements with about 60 % reduction in the primary clinical outcome after 12–24 sessions, while patients in the control group improved only by about 10 %. Lastly, we note that some studies were not included in our primary analyses due to inclusion and exclusion criteria (Table 1); they are briefly summarized in the Supplementary Material.

Table 5
Numbers of remitters in experimental and control group and their percentages with respect to total number enrolled patients (dropouts were treated as non-responders), number needed to treat for one additional patient to benefit (or to be harmed) [NNTB/NNTH] and their respective 95 % confidence interval (CI). Negative NNTB/NNTH or CIs indicate that patients in the control may have shown a better outcome.

Study	Experimental group		Control group		NNTB/NNTH	Lower 95 % CI	Upper 95 % CI
	Remitters / total sample	Remission rate (%)	Remitters / total sample	Remission rate (%)			
Linden et al. (2012)	2/8	25.00	0/8	0.00	4.00	1.69	−8.31
Young et al. (2017b)	6/19	31.58	1/17	5.88	3.89	2.06	−109.44
Mehler et al. (2018)	4/21	19.05	8/22	36.36	−5.78	10.60	−2.44
Lee et al. (2019)	6/12	50.00	1/12	8.33	2.40	1.49	19.66

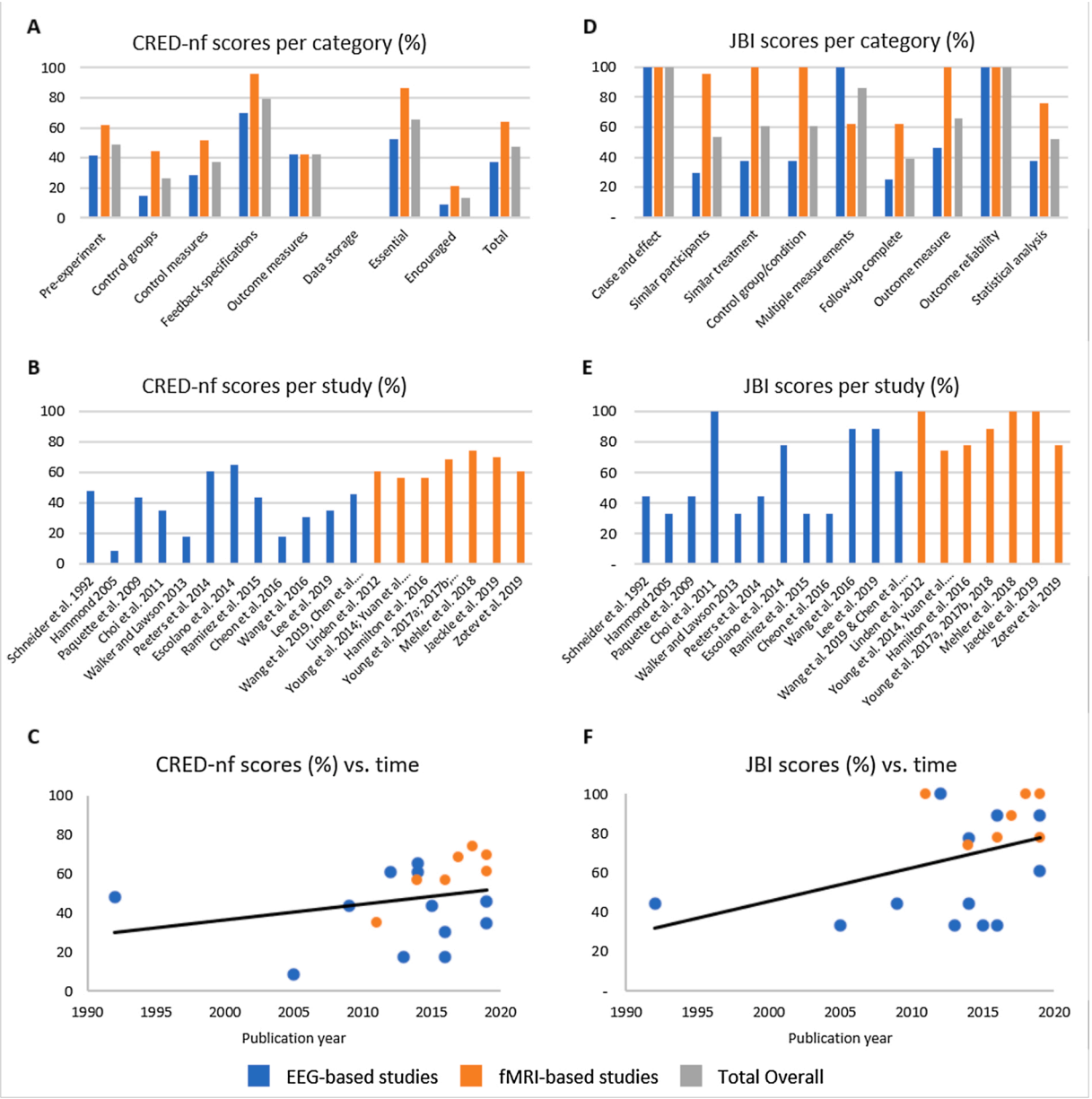

Fig. 3. Summary of CRED-nf scores. (A) the average score per CRED-nf item for category using EEG-based protocols (blue), fMRI-based protocols (orange), and the overall score across modalities (gray). (B) the average score per study using EEG- (blue) and fMRI-based (orange) protocols. Summary of JBI scores. (C) the trend of quality improvement measured with CRED-nf scores. (D) the average score per JBI item for studies using EEG-based protocols (blue), fMRI-based protocols (orange), and the overall score across modalities (gray). (E) the average score per study using EEG- (blue) and fMRI-based (orange) protocols. (C) the trend of quality improvement measured with JBI scores. Studies reporting results from the same database are reported as an averaged single bar.

3.2. fMRI neurofeedback paradigms and clinical effects

fMRI-based neurofeedback protocols commonly employ tasks of emotional self-regulation and train patients to either up- or down-regulating the BOLD signal in brain regions related to emotion processing, such as the amygdala, dorsal anterior cingulate cortex,

prefrontal cortex, insular cortex, superior temporal gyrus, precentral gyrus, and middle temporal gyrus (Johnston et al., 2010; Linhartová et al., 2019).

In a first proof-of-concept non-randomized, non-blinded study of fMRI neurofeedback in depression, Linden et al. (2012) compared two groups of eight medicated MDD patients. While the control group engaged in mere mental imagery training outside the scanner, the experimental group received four sessions of neurofeedback training, during which they used similar mental strategies to self-regulate the activity in brain areas responsive to affective visual stimulation. Specifically, to identify responsive ROIs for the algorithm, all volunteers were initially submitted to affectively charged figures with positive valence. Throughout the sessions, patients in the neurofeedback group learned to up-regulate the BOLD response of the targeted areas, including the ventrolateral and dorsolateral portions of the prefrontal cortex, insula, medial temporal lobe, and orbitofrontal cortex.

Further, the neurofeedback group, but not the control group, presented a significantly larger improvement in depressive symptoms (approximately 28 % of improvement and 7% of worsening, respectively) and two patients in the experimental but no patient in the control group were remitted at the primary endpoint (Linden et al., 2012). In a subsequent larger randomized single-blind controlled trial (N = 16 per group) by Mehler et al. (2018), medicated patients were assigned to one of two neurofeedback training interventions: whereas the experimental group trained over five sessions to activate limbic areas using positive mental imagery (NFE) similar to patients in the neurofeedback group in Linden et al. (2012), the active control group trained over five sessions to activate higher visual areas imagined relaxing scenes (NFS). Training areas in the NFS control group included regions involved in scene processing, such as the parahippocampal place area and higher visual cortices (Mehler et al., 2018). Although the NFE group was expected to show superior clinical improvements, no statistically significant group difference was found at the primary endpoint or a follow up (6 weeks later). However, patients in both groups showed substantial reductions on the HDRS-17 (about 42 % and 44 % for the NFE and NFS group, respectively), which lasted and improved slightly further at follow-up six weeks later (about 48 % and 59 % for the NFE and NFS group, respectively).

Moreover, about 38 % (12/32) of patients were remitted (based on the HDRS-17 score) at the primary endpoint (with 4/16 patients, i.e., 25 % in the NFE, and 8/16, i.e., 50 % in the NFS group, respectively). Potential reasons that may account for these findings include that both groups engaged in a potentially beneficial form of mental imagery. Further, post-experimental analyses showed that both groups presented overlapping active voxels in the anterior insula during the neurofeedback training. Of interest, a correlation between clinical improvement and a measure of self-efficacy was reported suggesting that the successful training experience may already provide clinical benefit to patients.

Another set of six studies reported clinical results and exploratory analyses from two independently conducted neurofeedback experiments (Young et al., 2017b, 2014) in which participants trained self-regulation of amygdala activity. In the first study by Young and colleagues (Young et al., 2014), unmedicated MDD patients enrolled in a non-randomized single-blinded, sham-controlled experiment, unmedicated MDD patients trained to self-regulate the amygdala using positive autobiographical memories (N = 13) in an experiment in which the control group (N = 6) received feedback from a brain area (the intraparietal sulcus) that was not associated with the mental task (Young et al., 2014). After a single session, the first group achieved effective control of the amygdala responsiveness. Psychometric testing suggested a reduction of anxiety indexes and increased happiness indexes (Young et al., 2014), but results for clinical effects were not reported albeit the HDRS-21 was assessed at baseline.

In a follow-up study the same research group included this initial data set and tested a few more patients in the patient control group (new N = 13 patients) as well as an additional control group of healthy participants (N = 27) (Yuan et al., 2014). Their results suggested slight decreases on the HDRS-21 in all groups (about 16 % of improvement for the experimental group, and 12 % for the patient control group) but no significant group difference. Of interest, post-hoc analyses indicated increased resting-state functional connectivity between the left amygdala and the left pregenual anterior cingulate cortex (pgACC), and between amygdala and the left cuneus in both groups following neurofeedback training (Yuan et al., 2014). Finally, upregulation of the left amygdala BOLD activity during a new (second) session of the same protocol was accompanied by positive average changes in frontal alpha EEG asymmetry, which significantly correlated with the MDD patients' trait depression severity (Zotev et al., 2016).

In a subsequent, larger randomized, double-blinded clinical trial, the same research group compared the clinical effects over two training sessions using a similar training protocol. The authors reported a significant group by session interaction and follow up-analyses suggested that only the experimental group (N = 18) that trained amygdala up-regulation showed improved depressive symptoms (about 39 % reduction of the Montgomery-Åsberg Depression Rating Scale - MADRS - at the primary endpoint). Of interest, about 32 % (6/19 patients) were remitted (based on the MADRS score) at the primary endpoint (Young et al., 2017b). In contrast, mean scores in the control group (N = 15) remained nearly unchanged (about 5%), and only one patient showed remission. Further, compared to the control group, the neurofeedback group presented higher hemodynamic and behavioral responses for positive visual stimuli, lower responses for negative stimuli (Young et al., 2017a). In a follow-up analysis the authors further reported functional connectivity changes between the amygdala and areas of the frontal and limbic network that correlated with the previously reported clinical improvement (Young et al., 2018a).

In a subsequent non-randomized single-blind controlled study this neurofeedback paradigm was expanded by Zotev et al. (2019) to a multimodal, single-blinded, single-session training protocol combining fMRI based self-regulation training of the left amygdala and left rostral anterior cingulate cortex (rACC), as well as EEG based training of asymmetry in the alpha and beta band. Unmedicated patients were assigned either to the experimental group (N = 16) that received veridical feedback or a control group (N = 8) that received randomly generated feedback signals that were unrelated to their brain activity. The results suggested that the experimental group showed increased activity in the left amygdala, EEG asymmetries, as well as enhanced functional connectivity between the left amygdala and the left rACC (Zotev et al., 2019). However, no information about clinical effects was reported in this study.

Hamilton et al. (2016) introduced another paradigm that employed functional connectivity based neurofeedback in which they investigated the ability of partly medicated MDD patients to down-regulate nodes from the salience network in the presence of negative stimuli. Twenty patients were presented to pictures taken from the IAPS (International Affective Picture System) that were associated with negative valence to identify nodes involved in processing negative affect. The authors then allocated patients either to an experimental group that received veridical (N = 10), or a control group (N = 10) that received a form of sham neurofeedback training where participants are provided with the replay of visual feedback from the experimental group to control (yoked feedback). When re-exposed to negative visual stimulation in a post test, only the neurofeedback group but not the control group showed reduced responses in neural nodes from the salience network. Moreover, there was a trend for lower scores for a self-reported responses to negative images (Hamilton et al., 2016). However, the study only assessed within group changes but no between group comparison, and clinical effects were not reported.

More recently, Jaeckle et al. (2019) conducted a randomized, single-blinded trial that consisted of three training sessions. Patients (majority of them under medication) were either allocated to the

experimental group (N = 19) that trained up-regulation of functional connectivity between the right superior anterior temporal lobe and the right subgenual cingulate, or to a control group (N = 16) that trained cognitive reappraisal techniques outside the fMRI scanner. Results suggested that both groups showed significant symptom improvement in the BDI scale (approximately 46 % and 37 %, respectively), but no significant difference between groups was found. Lastly, we note that some studies were not included in our primary analyses due to inclusion and exclusion criteria (Table 1); they are briefly summarized in the Supplementary Material (Section 4).

3.3. Clinical efficacy for different control condition categories

Both EEG and fMRI neurofeedback studies in MDD thus far published are heterogeneous with regards to some key design aspects of clinical studies. Albeit several studies employ similar training paradigms (e.g., alpha asymmetry EEG neurofeedback training of frontal electrodes, or self-regulation fMRI neurofeedback training of limbic areas), they vary substantially with regards to features such as randomization, blinding and control conditions. For instance, only six studies (2 applying EEG neurofeedback and 4 applying fMRI neurofeedback protocols) randomized patients to either an experimental or a control arm. Moreover, only four studies (all fMRI) used double-blinding, while the other fMRI neurofeedback studies (except for the first feasibility study) were single-blinded. Noteworthy, none of the EEG neurofeedback studies were single- or double-blinded, while all but one fMRI neurofeedback studies were at least single-blinded (Table 2). Whether neurofeedback studies allow single or double-blinded assessment depends largely on the choice of the control condition (Sorger et al., 2019): while some designs (e.g. yoked feedback) allow blinding patients (Young et al., 2017b), other active control conditions that are based on different instructions and veridical feedback do not (Mehler et al., 2018).

Normalized baseline scores of depressive symptoms (see Section 2.2) were largely comparable between studies as well as control conditions (where applicable). With the exception of five studies ((Linden et al., 2012), (Walker and Lawson, 2013), (Ramirez et al., 2015), (Wang et al., 2016), and (Zotev et al., 2019)), patients were on average moderately to severely depressed and experimental and control groups were on average largely matched for their depression severity at baseline (group differences were mostly under 10 %, with only one study (Wang et al., 2016) showing a difference of 11.55 %; see Fig. 2A–B). However, most studies did not provide sufficient clinical information regarding prior treatment experience, treatment resistance, duration of illness, number of episodes and hospitalizations of patients (see Table S1 in the Supplementary Material). We also note that the EEG-NF field has either employed only single-arm studies (58 %, in particular early studies) or passive control conditions, whereas the fMRI-NF field has exclusively employed active control conditions (71 % inside the scanner and 29 % outside the scanner).

Fig. 2C–D and E–F show the symptom improvement per group and the difference of improvement between groups, respectively. In general, all groups presented some level of symptom improvement, with the exception of one study in which the control group presented 7.20 % of mean symptom worsening (Linden et al., 2012). Regarding differences across groups, in seven studies the experimental group showed higher improvement than the control group, while in two studies the effect was in the opposite direction (Mehler et al., 2018; Wang et al., 2016). Moreover, group differences tended to be larger for studies that used a passive control groups compared to studies with active control groups, which found relatively small group differences (Fig. 2E–F). This exploratory finding is in line with the notion that non-specific psychosocial effects are additive, and they confirm previous theoretical considerations (Ros et al., 2020; Sorger et al., 2019; Thibault et al., 2016).

Average NNTB/NNTHs based on reported remission rates ranged between -5.78 and 4 and were mostly positive (Table 5), suggesting the experimental groups showed higher efficacy with respect to remission from depression. Noteworthy, only one (unblinded, non-randomized) study (Lee et al., 2019) could rule out potential superiority of the control condition, which consisted of continued standard care (psychopharmacological medication), over the experimental condition, which consisted of continued standard care augmented by EEG neurofeedback training (Table 5). In contrast, the upper bounds of 3 trials were negative and they could hence not reject the null hypothesis that patients in the control arm showed a better clinical outcome compared to patients in the main treatment arm (Altman, 1998). One main reason for this finding is likely the relatively small sample sizes of studies that could not exclude potential superiority of the control group. For instance, although Young et al. (2017b) found a remarkable difference in remission between the experimental and the control group, the upper bound of the 95 % confidence interval was -109.44; this negative values indicates that it remains possible that about 1 in 109 patients who are allocated to the experimental group will show less improvement compared to the control group.

3.4. Reported side effects and drop-outs

Side effects are rarely reported for neurofeedback interventions (Table 3, last column). This observation may be explained by the non-invasive nature of the intervention, but partly also related to reporting practices (see Section 3.5). In general, we note that one limiting factor for the wide usage of clinical neurofeedback may be physical discomfort experienced before and during each session, respectively. For example, EEG protocols require a relatively long time for the EEG cap preparation (positioning, conductive gel, calibration) (Nijholt et al., 2011). It also results in residual gel over the participant's head after the session. During fMRI protocols, on the other hand, the patient may experience claustrophobia due to the physical restriction imposed by the equipment (Sulzer et al., 2013). These aspects are particularly relevant to MDD patients because their symptoms can include diminished interest, sleeping problems, psychomotor agitation, and fatigue or loss of energy (Association, 2013). We documented reported reasons for drop-outs or exclusions (Table 2), which included lack of motivation (Hammond, 2005), tiredness (Cheon et al., 2016; Paquette et al., 2009; Young et al., 2014), discomfort (Young et al., 2017a), logistics difficulties (Cheon et al., 2016; Choi et al., 2011) and excessive noise (possibly related with the patient's agitation) (Escolano et al., 2014; Young et al., 2017b). However, we note that overall drop-out rates were relatively low and no serious side effects have been reported.

3.5. Experimental design and reporting quality

As noted above, a first overview of study designs (Table 2) suggests that while most neurofeedback studies published thus far employed control groups, only a minority conducted blinded assessment or randomized patients. We next assessed the quality of experimental designs and study reporting more systematically employing the JBI critical appraisal tools (Tufanaru et al., 2017) and CRED-nf checklist (Ros et al., 2020).

As shown in Fig. 3, EEG neurofeedback studies received on average lower scores in all CRED-nf points except for "Outcome measures" and "Data storage" (the latter was not fulfilled by any study included in this review). Similarly, EEG neurofeedback studies received on average lower scores for six of nine items of the JBI checklist. However, regarding the items "cause and effect", "outcome reliability items", both imaging methods presented full scores across studies, while for "multiple measurements", EEG neurofeedback studies tended to score higher on average. Both the CRED-nf and JBI checklist allowed identifying some major limitations in the field which we discus below. These will inform our recommendations formulated in Section 4.

Regarding the CRED-nf checklist, we first note that only five studies preregistered their experimental protocol. Complementarily, concerning the experimental design, one main limitation of EEG neurofeedback

studies in depression is the lack of adequate control groups (only present in 38 % of studies). A more general limitation, related to both imaging techniques, is the limited description of the online brain signal processing and artifact control. Although all included studies at least partly report how data is extracted and preprocessed (step 3), reporting often remained insufficient.

When reviewing studies for reported outcome measures, we found that while some studies defined "success", or "control" measures explicitly, many studies did not: according to the CRED-nf scores (see Table S3 in the Supplementary Material), only 51 % of studies reported neurofeedback success based on neural signals (33 % EEG and 81 % fMRI neurofeedback studies), while 54 % plotted within- or between sessions (38 % EEG and 81 % fMRI studies).

We also note that, only about 26 % of studies declared the primary clinical outcome measure (only 17 % EEG and 43 % fMRI neurofeedback studies). The distinction between primary and secondary outcome measures is considered a quality standard in clinical research: it is central to evaluating the clinical efficacy of an intervention (e.g., to estimate remission rates) and to control for error rates (in contrast to test results for secondary outcome measures, test results for predeclared primary outcome measures usually do not require correction for multiple testing). Further, only 19 % of studies evaluated psychosocial factors before or after the experiment. However, some EEG neurofeedback studies did not report if specific self-regulation strategies were provided/suggested to patients (58 % EEG and 100 % fMRI studies providing this information), and only very few studies reported debriefing results and thus could capture the strategies used (8% using EEG and 19 % using fMRI). Lastly, none of the studies stored the resulting (clinical or physiological) data or analysis code in publicly available domains.

4. Discussion

In this first systematic review of neurofeedback studies across imaging modalities conducted in depressed patients, we found that both EEG and fMRI studies report statistically significant and clinically meaningful within group improvements of clinical measures between 6% and 73 %. In comparison, between group comparisons showed numerically smaller changes ranging from -7% to 52 %. These findings may be explained by differences in used controlled conditions. It is assumed, however, that overall clinical effects following neurofeedback training can be partly or largely attributed to various non-specific factors: patient's positive expectancies, the rewarding experience of positive feedback, but also regression to the mean likely contribute substantially to observed within group improvements.

Neurofeedback training is a complex intervention and involves various degrees of freedom in designing control conditions. These range from passive control designs (e.g. continued standard care vs. continued standard care and neurofeedback augmentation training); these are expected to provide the least control of non-specific factors, to active control designs (e.g. continued standard care and neurofeedback augmentation training with vertical from a control region vs. continued standard care and targeted neurofeedback augmentation); these are expected to provide the most control for non-specific factors (Ros et al., 2020; Sorger et al., 2019; Thibault et al., 2016). As recently discussed in-depth by the neurofeedback community (Lubianiker et al., 2019; Sorger et al., 2019), the choice of optimal control conditions poses a challenge for neurofeedback experiments. Control conditions are important to evaluate non-specific effects and to compute more informative effect sizes such as NNTB that allow comparisons to other therapeutic approaches. To compare between-group clinical effects across neurofeedback studies, we therefore grouped these according to their control condition. Results indeed showed that active control conditions presented smaller group differences in favor of neurofeedback compared to more lenient passive control conditions (Fig. 2E). Of interest, these findings are comparable to those reported for EEG neurofeedback training in ADHD (Cortese et al., 2016; Group et al., 2020; Van Doren et al., 2019).

Reported clinical effects for neurofeedback training seem substantial; however, therapeutic effects specific to neural targets are likely relatively small, and hence future RCTs will require larger samples to study *neurofeedback-specific effects* in depression. Further, longer follow-up periods are desirable; clinical effects following neurofeedback interventions have been documented to last, and partly further improve for up to several months after the last neurofeedback session (Becerra et al., 2006; Gevensleben et al., 2010; Goldway et al., 2019; Mehler et al., 2018; Rance et al., 2018). In addition, it remains of interest to investigate to which degree observed effects occur within or between training sessions and whether there is an interaction thereof (Mehler et al., 2021).

Moving on to comparing EEG and fMRI, substantial differences in designs were found: Whereas most EEG studies lacked control conditions and were not blinded, recent fMRI studies increasingly fulfill these standards. In general, we found that fMRI studies tended to fulfill more study design and reporting quality criteria. One possible explanation for this result may be that most studies were planned and reported more recently compared to EEG-based protocols. They may also have been able to incorporate criticism raised against previous EEG neurofeedback studies, benefit from methodological advancements, and broader debates around adequate statistical aspects (Button et al., 2013; Nieuwenhuis et al., 2011). This trend is exemplified in Fig. 3C and F.

While most EEG neurofeedback studies can be considered (uncontrolled) phase IIa trials that aim to demonstrate feasibility, most fMRI neurofeedback studies represent (controlled) phase IIb trials that aim to demonstrate clinical efficacy. However, common to almost all studies are relatively small sizes, which render these statistically underpowered to detect small or medium effects. From RCTs conducted on the clinical effects of antidepressant medication, for instance, relatively small effects (Cohen's $d = 0.2$ to 0.3) are documented for treatment vs. placebo controls (Cuijpers et al., 2014; Kirsch et al., 2008) To detect an effect size within this range with 80 % probability, studies would need to feature at least about 176 patients per group for a two-arm controlled study. At least for fMRI-based neurofeedback protocols, such scales are likely only achievable in multi center studies. Some further ideas on this matter are listed in the recommendations section below. To further illustrate the limited power of existing studies, 3 of 4 studies that also reported remission rates could not rule out superiority of the control group in an NNTH analysis that we conducted (Table 5).

Evaluating reporting practices, most included studies lacked information about several aspects that are considered essential or highly desirable such as declaring the primary outcome measure, reporting a sampling plan, reporting feedback controllability or remission rates. Hence, on average studies in the field still bear considerable risk for bias, which restricts generalizations that can be drawn from reported findings. Moreover, we note that several published studies included partly overlapping samples, which made it sometimes difficult to assess their quality in a coherent way. Further, such practice indicated that authors may have employed flexible sampling stopping rules (without adequate adjustment), which risks increasing type-I error rate (see the Recommendations section below for some suggestions). Most of these aspects could be addressed by comprehensive study preregistrations, including the declaration of the primary outcome measure, main hypotheses, intended sample size and planned analyses. Originally introduced in clinical medicine (DeAngelis et al., 2005), study preregistrations can restrict degrees of freedom and avoid sources for researcher bias, including outcome switching, inadequately used flexible stopping rules and analytical degrees of freedom (Nosek et al., 2018) as well as publication bias (Allen and Mehler, 2019).

We also note that many studies, and in particular EEG neurofeedback studies, did not report neurofeedback success measures. A clear definition on success measures allows assessing the proportion of individuals who show relatively poor neurofeedback control, a phenomenon that

has also been labeled as "illiteracy" (Allison and Neuper, 2010), and which likely pertains 10–50% of neurofeedback users (Alkoby et al., 2018; Allison and Neuper, 2010; Edlinger et al., 2015). Estimating the proportion of non-learners, and ideally identifying predictors for self-regulation success, seem in particular important for neurofeedback studies with depressed patients who tend to process negative experiences (e.g. no self-regulation success) more negatively (Disner et al., 2017; Peckham et al., 2010).

Further, documenting experiment factors, such as attention from the staff, comfort in the experiment room, or motivation, measures of confidence, or frustration, and personal believes might help to understand variations in self-regulation performance (Paret et al., 2019), but also explain observed clinical effects. Constructs such as self-efficacy that are related to the psychopathology in depression (Bandura, 1982) may be modifiable through self-regulation training (Linden, 2014; Mehler et al., 2018). Ratings also showed that none of the included studies has shared their imaging and/or clinical data publicly and that only a few studies were preregistered. While such reservation may be an expression of data protection concerns, we note that data anonymization tools are widely available and it should be in the best interest of the community to make use of these and follow recent efforts of the neuroimaging community clinical medicine in tackling issues around reproducibility and replicability (Poldrack and Gorgolewski, 2014).

Finally, it is crucial to use appropriate and robust methods for data extraction and preprocessing. For instance, most EEG and fMRI studies do not use state-of-the-art artifact control methods (e.g., electrooculography and electromyography) when calculating the feedback signal. Similarly, for fMRI-neurofeedback, control for confounding factors such as online correction of head motion, breathing, and cardiovascular artefacts are often insufficiently reported, although they may have a major impact on reported findings (Weiss et al., 2020). This finding is in line with earlier findings for fMRI neurofeedback studies more broadly (Heunis et al., 2020; Thibault et al., 2018).

Overall, our findings indicate that, CRED-NF and JBI checklist ratings suggest that fMRI neurofeedback studies featured on average better reporting quality. Yet, we note that the CRED-nf guidelines were published only very recently and hence the authors of the investigated studies could not use neurofeedback specific guidelines as orientation for design and reporting practices. Comparing JBI ratings reported here with other fields, the present sample featured an average rating of 6.17, which is similar to those reported in systematic reviews (that included a similar number of studies) conducted about fMRI neurofeedback training in stroke patients (mean 6.24) and non-clinical/clinical fNIRS neurofeedback (mean 5.55) (Kohl et al., 2020; Wang et al., 2018). Also, with regard to essential, encouraged and total CRED-NF ratings, we found similar results (with 65 % vs. 63 %, 13 % vs 10 % and 47 % vs. 45 %, respectively) compared to the fNIRS-NF field (Kohl et al., 2020). Lastly, we note that one main limitation of this review was the relatively small number of studies that could be included, and which precluded employing other established meta-research techniques such as p-curve analysis (Simonsohn et al., 2014) or funnel plots to test for small study effects (e.g. due to publication bias). Further, the heterogeneity in study designs that controlled for non-specific effects to different degrees – which ranged from no control to very conservative active neurofeedback control conditions – rendered an aggregated effect size across studies rather meaningless. We therefore decided to merely provide estimates of clinical improvement in percentages averages for studies with similar control conditions.

4.1. Recommendations

Despite promising first results with patient groups, current neurofeedback protocols present methodological challenges for real-world therapeutic applications (Arns et al., 2017; Thibault et al., 2016). Heterogeneity of protocols and inconsistent reporting make replication and standardization difficult. These aspects are crucial not only for the

Table 6
Recommendations for future experiments with depressive patients (some of these points are discussed in more detail in Section 6 of the Supplementary Material).

a) More comprehensive clinical documentation and phenotyping	To ensure reliable clinical results and to allow comparison between studies, we recommend that future neurofeedback experiments in depressive patients use formal and standardized procedures to diagnose and evaluate clinical changes with clinician-rated scales (e.g., HDRS-21, MADRS) and self-rated scales (e.g., BDI-II or QIDS-SR_{16}). Besides changes in sum scores of scales, we encourage reporting changes in individual items to assess changes in specific symptoms or symptom networks and cluster different types of responses (Fried and Nesse, 2015; Fried et al., 2017; Hofmann et al., 2016). Further detailed descriptions of previous antidepressant treatment and patients' duration of illness should be provided to allow to assess the level of chronicity and treatment resistance of included patients, factors that may impact clinical outcomes (Kiebs et al., 2019). Further, we note that etiology of developing MDD is likely quite heterogeneous across patients (Winokur, 1997) and hence a more comprehensive clinical and phenotypic characterization may help identifying patients subgroups who benefit in particular from neurofeedback training.
b) *Choice of appropriate control conditions*	The use of control conditions is fundamental to determine if any positive effect is caused by the neurofeedback protocol or by other reasons. The best control design depends on the research interest, and a decision tree for control conditions for neurofeedback applications was recently described by Sorger et al. (2019). In the context of depressive patients, different control conditions should be considered.
c) Adequately powered studies	Powering studies to be able to detect meaningful effect sizes or rule these out (Algermissen and Mehler, 2018). For instance, studies may set minimal clinically important differences (MCID) reported for depressed patients as their target effect size (Lakens et al., 2018). Further, alternative sampling strategies such as sequential Bayes Factor (SBF) sampling may be worthwhile exploring for clinical neurofeedback studies (Schönbrodt and Wagenmakers, 2018). Lastly, we recommend that null findings are followed up with appropriate statistical tests that allow providing evidence for the absence of an effect (Mehler et al., 2019).
d) Online and offline quality control of signals	Although several studies report the exclusion of subjects due to excessive artifacts, only few studies intended to perform online quality control and denoising. This is not a particular problem in studies applying neurofeedback in MDD populations, but a current issue in the field (Heunis et al., 2020). Thus, we recommend that more rigorous approaches should be conducted during experiments and the reporting of results, for instance with regards to EOG and EMG noises in EEG-based protocols (Moretti et al., 2003), or respiration and pulse waves in fMRI-based experiments (Murphy et al., 2013). Further, it is fundamental to evaluate and report differences in artifacts between groups (Ros et al., 2020), since group-biased noisy data can lead to false conclusions.
e) *Standardization of protocols*	Several clinical neurofeedback studies targeting MDD patients do not focus on new methodological approaches (for example, testing signal processing and feedback presentation), but on potential clinical,

(*continued on next page*)

Table 6 (*continued*)

	cognitive, or neural benefits of targeting one, or more, brain regions (fMRI), or frequencies (EEG). In this context, the use of standardized methods to extract information from the source signal, or to present the feedback would allow direct comparison between studies. Also, potential comparisons depend on a clear definition of success/learning, as well as the detailed report of responders/literates and non-responders/illiterates. In particular for depressed patients, insufficient self-regulation success may result in frustration and potentially deteriorate clinical outcome in individuals. In line with previous consensus (Ros et al., 2020), we recommend that individual self-regulation performances should be ideally reported and potential predictors of self-regulation success explored and researchers should aim to standardize approaches (Paret et al., 2019).
f) *Basic methodological research*	As pointed out by others (Paret et al., 2019), more basic research is needed to solve the many open methodological questions and increase standardization and agreements to finally inform translational work. This work seems particularly relevant for the treatment of MDD, which affects the reward system. Lastly, it has been suggested that neurofeedback may serve as a tool to test neural models (Nielson et al., 2020) or biomarkers suggested for MDD. However, there is reason for skepticism and discussions about the reliability and validity of biomarker research remain controversial (see Section 6 of in the Supplementary Material).
g) *Exploring the potential for children and young adults*	The current review was limited to studies conducted in adults. However, given the low risk profile of non-invasive neurofeedback training and the promising clinical findings found in adults, we recommend that this approach should also be explored in younger patients. In particular modulating self-referential beliefs such as self-efficacy may provide substantial clinical benefits related to anxious (Lewis et al., 2020) but also depressive symptoms. Noteworthy, first results from an fMRI neurofeedback study in depressed adolescence showed feasibility and promising clinical potential (Quevedo et al., 2019) (see also Section 6 g in the Supplementary Material). Feasibility has also been recently demonstrated in targeting anxiety (Zich et al., 2020) and depression (Quevedo et al., 2020) in adolescents.
h) *Appropriate reporting of methods and results*	In addition to the proper experimental design, an appropriate report of methods and results is crucial to advance the neurofeedback field and propagate reliable results. For example, an extensive methodological review showed that a substantial portion of neurofeedback studies do not apply or report adequate denoising methods in fMRI-based protocols (Heunis et al., 2020) (complete data base available here: https://rtfmri-methods.herokuapp.com/). the CRED-nf checklist was created in a collaborative effort between several dozen laboratories to support this matter (Ros et al., 2020), including an easy-to-use app for quick validation (rtfin.org/CREDnf).
i) *Study preregistration and open science research practices*	To make neurofeedback findings transparent and reliable, as well as to allow further collaboration between research groups, we strongly recommend that researchers explore and implement open science research practices where possible (Allen and Mehler, 2019; Nosek et al., 2015) by preregistering their study protocol and sharing the data that support their final results. Analytical degrees of freedom remain a controversial topic in neuroimaging (Botvinik-Nezer et al., 2020; Carp, 2012); real-time experiments already predeclare a substantial part of their analysis pipeline when setting parameters for real-time data analysis and it is hence in particular suited for study preregistration (e.g., Mehler et al. (2020)) or publishable research protocols (e.g., Cox et al. (2016)). Regarding data sharing practices, researchers can benefit from recommendations for reliable analysis pipelines (Nichols et al., 2017), as well tools to standardize data accessibility and reproducibility (Gorgolewski et al., 2017) and facilitate data sharing (Gorgolewski et al., 2016; Poldrack et al., 2013).

research community to understand and progress the neurofeedback technology (Thibault et al., 2017), but also for patients, since a poor setup can cause frustration and lead to discontinued training (Müller-Putz et al., 2015). Thus, in line with the final aim of this review, we provide here an overview of recommendations that future researchers should adopt for experiments with depressive patients (Table 6). A more detailed discussion of these with a particular focus on points a) - c) can be found in the Supplementary Material (Section 6).

5. Conclusion

Neurofeedback presents a complex, non-invasive intervention which aims to target cognitive and affective processes affected in patients with depression through mental imagery-based self-regulation of functionally relevant brain areas or network. As such the approach has good face validity for MDD. Patients have shown significant clinical improvements as well as cognitive and neural changes following neurofeedback training with both EEG and fMRI-based protocols. Moreover, given the relatively low risk of side effects due to its non-invasive nature, we consider neurofeedback in particular worth exploring as an augmentation therapy for patients who have already received standard care but remain symptomatic. However, our review also found that most studies published thus far still lag current best practice standards of study design and reporting quality. Some main issues are the lack of study preregistration, the use of mostly small and/or unbalanced samples as well as the lack of control conditions, randomized treatment allocation or blinding. These issues render the evaluation of clinical effects difficult and require improvements in future studies. Following a first attempt to quantify the contribution of different non-specific effects for studies that included a control group, our results suggest that non-specific effects add up such that more passive control conditions (e.g., continued standard care) yield larger group differences compared to more conservative active control conditions (e.g., successful neurofeedback self-regulation training from an alternative brain region). We close with a set of recommendations for future studies, which include suggestions for more comprehensive clinical documentation, considerations regarding adequate control conditions, a synopsis of some statistical and study design aspects that can help achieving more adequately powered and hence more informative studies, aspects concerning signal quality and protocol standardization, and lastly pointers to open science resources.

Declaration of Competing Interest

SHK and DMAM receive payments for work as independent advisors to a neurofeedback start-up company (Mendi Innovations AB). LRT and DEJL declare no conflict of interest.

Acknowledgements

We would like to thank (in alphabetical order) Eunjin Cheon, Paul Hamilton, Bon Hoon Koo, Argyris Stringaris, Kymberly Young, and

Roland Zahn for helpful feedback on an earlier version of this manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at https://doi.org/10.1016/j.neubiorev.2021.02.015.

References

Algermissen, J., Mehler, D.M., 2018. May the power be with you: are there highly powered studies in neuroscience, and how can we get more of them? J. Neurophysiol. 119, 2114–2117.

Alkoby, O., Abu-Rmileh, A., Shriki, O., Todder, D., 2018. Can we predict who will respond to neurofeedback? A review of the inefficacy problem and existing predictors for successful EEG neurofeedback learning. Neuroscience 378, 155–164.

Allen, C., Mehler, D.M., 2019. Open science challenges, benefits and tips in early career and beyond. PLoS Biol. 17, e3000246.

Allison, B.Z., Neuper, C., 2010. Could Anyone Use a BCI?, Brain-Computer Interfaces. Springer, pp. 35–54.

Alonso, J., Lépine, J.-P., 2007. Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). J. Clin. Psychiatry.

Altman, D.G., 1998. Confidence intervals for the number needed to treat. BMJ 317, 1309–1312.

Arnold, L.E., Lofthouse, N., Hersch, S., Pan, X., Hurt, E., Bates, B., Kassouf, K., Moone, S., Grantier, C., 2013. EEG neurofeedback for ADHD: double-blind sham-controlled randomized pilot feasibility trial. J. Atten. Disord. 17, 410–419.

Arns, M., Batail, J.-M., Bioulac, S., Congedo, M., Daudet, C., Drapier, D., Fovet, T., Jardri, R., Le-Van-Quyen, M., Lotte, F., 2017. Neurofeedback: one of today's techniques in psychiatry? LEncphale 43, 135–145.

Association, A.P, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Baehr, E., Rosenfeld, J.P., Baehr, R., 1997. The clinical use of an alpha asymmetry protocol in the neurofeedback treatment of depression: two case studies. J. Neurother. 2, 10–23.

Bandura, A., 1982. Self-efficacy mechanism in human agency. Am. Psychol. 37, 122.

Bari, A.A., Mikell, C.B., Abosch, A., Ben-Haim, S., Buchanan, R.J., Burton, A.W., Carcieri, S., Cosgrove, G.R., D'Haese, P.-F., Daskalakis, Z.J., 2018. Charting the road forward in psychiatric neurosurgery: proceedings of the 2016 American Society for Stereotactic and Functional Neurosurgery workshop on neuromodulation for psychiatric disorders. J. Neurol. Neurosurg. Psychiatr. 89, 886–896.

Becerra, J., Fernandez, T., Harmony, T., Caballero, M., Garcia, F., Fernandez-Bouzas, A., Santiago-Rodríguez, E., Prado-Alcalá, R., 2006. Follow-up study of learning-disabled children treated with neurofeedback or placebo. Clin. EEG Neurosci. 37, 198–203.

Beck, A.T., 2008. The evolution of the cognitive model of depression and its neurobiological correlates. Am. J. Psychiatry 165, 969–977.

Bender, R., 2001. Calculating confidence intervals for the number needed to treat. Control. Clin. Trials 22, 102–110.

Birbaumer, N., Ruiz, S., Sitaram, R., 2013. Learned regulation of brain metabolism. Trends Cogn. Sci. 17, 295–302.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J.A., Adcock, R.A., 2020. Variability in the analysis of a single neuroimaging dataset by many teams. Nature 1–7.

Bromet, E., Andrade, L.H., Hwang, I., Sampson, N.A., Alonso, J., De Girolamo, G., De Graaf, R., Demyttenaere, K., Hu, C., Iwata, N., 2011. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 9, 90.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376.

Carp, J., 2012. On the plurality of (methodological) worlds: estimating the analytic flexibility of FMRI experiments. Front. Neurosci. 6, 149.

Chen, T.-C., Lin, I.-M., 2020. The learning effects and curves during high beta down-training neurofeedback for patients with major depressive disorder. J. Affect. Disord. 266, 235–242.

Cheon, E.-J., Koo, B.-H., Seo, W.-S., Lee, J.-Y., Choi, J.-H., Song, S.-H., 2015. Effects of neurofeedback on adult patients with psychiatric disorders in a naturalistic setting. Appl. Psychophysiol. Biofeedback 40, 17–24.

Cheon, E.-J., Koo, B.-H., Choi, J.-H., 2016. The efficacy of neurofeedback in patients with major depressive disorder: an open labeled prospective study. Appl. Psychophysiol. Biofeedback 41, 103–110.

Choi, S.W., Chi, S.E., Chung, S.Y., Kim, J.W., Ahn, C.Y., Kim, H.T., 2011. Is alpha wave neurofeedback effective with randomized clinical trials in depression? A pilot study. Neuropsychobiology 63, 43–51.

Clark, D.A., Beck, A.T., 2010. Cognitive theory and therapy of anxiety and depression: convergence with neurobiological findings. Trends Cogn. Sci. 14, 418–424.

Clark, D.A., Beck, A.T., 2011. Cognitive Therapy of Anxiety Disorders: Science and Practice. Guilford Press.

Cortese, S., Ferrin, M., Brandeis, D., Holtmann, M., Aggensteiner, P., Daley, D., Santosh, P., Simonoff, E., Stevenson, J., Stringaris, A., 2016. Neurofeedback for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. J. Am. Acad. Child Adolesc. Psychiatry 55, 444–455.

Cox, W.M., Subramanian, L., Linden, D.E., Lührs, M., McNamara, R., Playle, R., Hood, K., Watson, G., Whittaker, J.R., Sakhuja, R., 2016. Neurofeedback training for alcohol dependence versus treatment as usual: study protocol for a randomized controlled trial. Trials 17, 1–10.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M., 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 337.

Cui, X., Bray, S., Bryant, D.M., Glover, G.H., Reiss, A.L., 2011. A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. Neuroimage 54, 2808–2821.

Cuijpers, P., Turner, E.H., Koole, S.L., Van Dijke, A., Smit, F., 2014. What is the threshold for a clinically relevant effect? The case of major depressive disorders. Depress. Anxiety 31, 374–378.

Curran, E.A., Stokes, M.J., 2003. Learning to control brain activity: a review of the production and control of EEG components for driving brain–computer interface (BCI) systems. Brain Cogn. 51, 326–336.

Cusin, C., Dougherty, D.D., 2012. Somatic therapies for treatment-resistant depression: ECT, TMS, VNS, DBS. Biol. Mood Anxiety Disord. 2, 14.

Da Silva, F.L., 2009. EEG: origin and measurement. EEG-fMRI. Springer, pp. 19–38.

De Vos, M., Kroesen, M., Emkes, R., Debener, S., 2014. P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. J. Neural Eng. 11, 036008.

DeAngelis, C.D., Drazen, J.M., Frizelle, F.A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A.J.P., 2005. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Arch. Dermatol. 141, 76–77.

Delaloye, S., Holtzheimer, P.E., 2014. Deep brain stimulation in the treatment of depression. Dialogues Clin. Neurosci. 16, 83.

Deldin, P.J., Chiu, P., 2005. Cognitive restructuring and EEG in major depression. Biol. Psychol. 70, 141–151.

Dimidjian, S., Barrera Jr., M., Martell, C., Munoz, R.F., Lewinsohn, P.M., 2011. The origins and current status of behavioral activation treatments for depression. Annu. Rev. Clin. Psychol. 7, 1–38.

Disner, S.G., Beevers, C.G., Haigh, E.A., Beck, A.T., 2011. Neural mechanisms of the cognitive model of depression. Nat. Rev. Neurosci. 12, 467–477.

Disner, S.G., Shumake, J.D., Beevers, C.G., 2017. Self-referential schemas and attentional bias predict severity and naturalistic course of depression symptoms. Cogn. Emot. 31, 632–644.

Earnest, C., 1999. Single case study of EEG asymmetry biofeedback for depression: an independent replication in an adolescent. J. Neurother. 3, 28–35.

Edlinger, G., Allison, B.Z., Guger, C., 2015. How many people can use a BCI system? Clinical Systems Neuroscience. Springer, pp. 33–66.

Enriquez-Geppert, S., Huster, R.J., Herrmann, C.S., 2017. EEG-neurofeedback as a tool to modulate cognition and behavior: a review tutorial. Front. Hum. Neurosci. 11, 51.

Escolano, C., Navarro-Gil, M., Garcia-Campayo, J., Congedo, M., De Ridder, D., Minguez, J., 2014. A controlled study on the cognitive effect of alpha neurofeedback training in patients with major depressive disorder. Front. Behav. Neurosci. 8, 296.

Fava, M., Davidson, K.G., 1996. Definition and epidemiology of treatment-resistant depression. Psychiatr. Clin. North Am. 19, 179–200.

Fodor, L.A., Georgescu, R., Cuijpers, P., Szamoskozi, Ş., David, D., Furukawa, T.A., Cristea, I.A., 2020. Efficacy of cognitive bias modification interventions in anxiety and depressive disorders: a systematic review and network meta-analysis. Lancet Psychiatry 7, 506–514.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R. C., Fawcett, J., 2010. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 303, 47–53.

Fovet, T., Jardri, R., Linden, D., 2015. Current issues in the use of fMRI-based neurofeedback to relieve psychiatric symptoms. Curr. Pharm. Des. 21, 3384–3394.

Fried, E.I., Nesse, R.M., 2015. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study. J. Affect. Disord. 172, 96–102.

Fried, E.I., van Borkulo, C.D., Cramer, A.O., Boschloo, L., Schoevers, R.A., Borsboom, D., 2017. Mental disorders as networks of problems: a review of recent insights. Soc. Psychiatry Psychiatr. Epidemiol. 52, 1–10.

Gevensleben, H., Holl, B., Albrecht, B., Schlamp, D., Kratz, O., Studer, P., Rothenberger, A., Moll, G.H., Heinrich, H., 2010. Neurofeedback training in children with ADHD: 6-month follow-up of a randomised controlled trial. Eur. Child Adolesc. Psychiatry 19, 715–724.

Goldway, N., Ablin, J., Lubin, O., Zamir, Y., Keynan, J.N., Or-Borichev, A., Cavazza, M., Charles, F., Intrator, N., Brill, S., 2019. Volitional limbic neuromodulation exerts a beneficial clinical effect on Fibromyalgia. Neuroimage 186, 758–770.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwartz, Y., Sochat, V.V., Ghosh, S.S., Maumet, C., Nichols, T.E., Poline, J.-B., Yarkoni, T., 2016. NeuroVault. org: a repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. Neuroimage 124, 1242–1244.

Gorgolewski, K.J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., Churchill, N.W., Cohen, A.L., Craddock, R.C., Devenyi, G.A., 2017. BIDS apps: improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. PLoS Comput. Biol. 13, e1005209.

Groenewold, N.A., Opmeer, E.M., de Jonge, P., Aleman, A., Costafreda, S.G., 2013. Emotional valence modulates brain functional abnormalities in depression: evidence from a meta-analysis of fMRI studies. Neurosci. Biobehav. Rev. 37, 152–163.

Group, T.N.C., Arnold, L.E., Arns, M., Barterian, J., Bergman, R., Black, S., Conners, C.K., Connor, S., Dasgupta, S., deBeus, R., 2020. Double-blind placebo-controlled randomized clinical trial of neurofeedback for attention-deficit/hyperactivity disorder with 13 month follow-up. J. Am. Acad. Child Adolesc. Psychiatry.

Gruzelier, J.H., 2014. EEG-neurofeedback for optimising performance. I: a review of cognitive and affective outcome in healthy participants. Neurosci. Biobehav. Rev. 44, 124–141.

Hamilton, J.P., Farmer, M., Fogelman, P., Gotlib, I.H., 2015. Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. Biol. Psychiatry 78, 224–230.
Hamilton, J.P., Glover, G.H., Bagarinao, E., Chang, C., Mackey, S., Sacchet, M.D., Gotlib, I.H., 2016. Effects of salience-network-node neurofeedback training on affective biases in major depressive disorder. Psychiatry Res. Neuroimaging 249, 91–96.
Hammond, D.C., 2005. Neurofeedback treatment of depression and anxiety. J. Adult Dev. 12, 131–137.
Hampson, M., Scheinost, D., Qiu, M., Bhawnani, J., Lacadie, C.M., Leckman, J.F., Constable, R.T., Papademetris, X., 2011. Biofeedback of real-time functional magnetic resonance imaging data from the supplementary motor area reduces functional connectivity to subcortical regions. Brain Connect. 1, 91–98.
Harmon-Jones, E., Gable, P.A., Peterson, C.K., 2010. The role of asymmetric frontal cortical activity in emotion-related phenomena: a review and update. Biol. Psychol. 84, 451–462.
Heunis, S., Hellrung, L., Van der Meer, J., Bergert, S., Sladky, R., Pamplona, G., Skouras, S., 2019. rtQC: An Open-Source Toolbox for Real-Time fMRI Quality Control.
Heunis, S., Lamerichs, R., Zinger, S., Caballero-Gaudes, C., Jansen, J.F., Aldenkamp, B., Breeuwer, M., 2020. Quality and denoising in real-time functional magnetic resonance imaging neurofeedback: a methods review. Hum. Brain Mapp.
Hinterberger, T., Kübler, A., Kaiser, J., Neumann, N., Birbaumer, N., 2003. A brain–computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device. Clin. Neurophysiol. 114, 416–425.
Hoertel, N., Franco, S., Wall, M., Oquendo, M., Kerridge, B., Limosin, F., Blanco, C., 2015. Mental disorders and risk of suicide attempt: a national prospective study. Mol. Psychiatry 20, 718.
Hofmann, S.G., Curtiss, J., McNally, R.J., 2016. A complex network perspective on clinical science. Perspect. Psychol. Sci. 11, 597–605.
Hoshi, Y., 2003. Functional near-infrared optical imaging: utility and limitations in human brain mapping. Psychophysiology 40, 511–520.
Huppert, T.J., Hoge, R.D., Diamond, S.G., Franceschini, M.A., Boas, D.A., 2006. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. Neuroimage 29, 368–382.
Jaeckle, T., Williams, S.C., Barker, G.J., Basilio, R., Carr, E., Goldsmith, K., Colasanti, A., Giampietro, V., Cleare, A., Young, A.H., 2019. Self-blaming emotions in major depression: a randomised pilot trial comparing fMRI neurofeedback training with self-guided psychological strategies (NeuroMooD). medRxiv, 19004309.
Jensen, J.S., Bielefeldt, A.Ø., Hróbjartsson, A., 2017. Active placebo control groups of pharmacological interventions were rarely used but merited serious consideration: a methodological overview. J. Clin. Epidemiol. 87, 35–46.
Johnston, S.J., Boehm, S.G., Healy, D., Goebel, R., Linden, D.E., 2010. Neurofeedback: a promising tool for the self-regulation of emotion networks. Neuroimage 49, 1066–1072.
Kaiser, R.H., Andrews-Hanna, J.R., Wager, T.D., Pizzagalli, D.A., 2015. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. JAMA Psychiatry 72, 603–611.
Kiebs, M., Hurlemann, R., Mutz, J., 2019. Repetitive transcranial magnetic stimulation in non-treatment-resistant depression. Br. J. Psychiatry 215, 445–446.
Kim, S., Birbaumer, N., 2014. Real-time functional MRI neurofeedback: a tool for psychiatry. Curr. Opin. Psychiatry 27, 332–336.
Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., Johnson, B.T., 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med. 5, e45.
Kohl, S.H., Mehler, D.M.A., Lührs, M., Thibault, R.T., Konrad, K., Sorger, B., 2020. The potential of functional near-infrared spectroscopy-based neurofeedback – a systematic review and recommendations for best practice. Front. Neurosci. 14, 594.
Koush, Y., Rosa, M.J., Robineau, F., Heinen, K., Rieger, S.W., Weiskopf, N., Vuilleumier, P., Van De Ville, D., Scharnowski, F., 2013. Connectivity-based neurofeedback: dynamic causal modeling for real-time fMRI. Neuroimage 81, 422–430.
Krucoff, M.O., Rahimpour, S., Slutzky, M.W., Edgerton, V.R., Turner, D.A., 2016. Enhancing nervous system recovery through neurobiologics, neural interface training, and neurorehabilitation. Front. Neurosci. 10, 584.
Krusienski, D.J., Sellers, E.W., Cabestaing, F., Bayoudh, S., McFarland, D.J., Vaughan, T. M., Wolpaw, J.R., 2006. A comparison of classification techniques for the P300 Speller. J. Neural Eng. 3, 299.
Kupfer, D.J., Frank, E., Phillips, M.L., 2012. Major depressive disorder: new clinical, neurobiological, and treatment perspectives. The Lancet 379, 1045–1055.
Lakens, D., Scheel, A.M., Isager, P.M., 2018. Equivalence testing for psychological research: a tutorial. Adv. Methods Pract. Psychol. Sci. 1, 259–269.
Lee, Y.-J., Lee, G.-W., Seo, W.-S., Koo, B.-H., Kim, H.-G., Cheon, E.-J., 2019. Neurofeedback treatment on depressive symptoms and functional recovery in treatment-resistant patients with major depressive disorder: an open-label pilot study. J. Korean Med. Sci. 34.
Lewinsohn, P.M., Rosenbaum, M., 1987. Recall of parental behavior by acute depressives, remitted depressives, and nondepressives. J. Pers. Soc. Psychol. 52, 611.
Lewis, K.M., Matsumoto, C., Cardinale, E., Jones, E.L., Gold, A.L., Stringaris, A., Leibenluft, E., Pine, D.S., Brotman, M.A., 2020. Self-efficacy As a target for neuroscience research on moderators of treatment outcomes in pediatric anxiety. J. Child Adolesc. Psychopharmacol. 30, 205–214.
Linden, D.E., 2014. Neurofeedback and networks of depression. Dialogues Clin. Neurosci. 16, 103.
Linden, D.E., Turner, D.L., 2016. Real-time functional magnetic resonance imaging neurofeedback in motor neurorehabilitation. Curr. Opin. Neurol. 29, 412.
Linden, D.E., Habes, I., Johnston, S.J., Linden, S., Tatineni, R., Subramanian, L., Sorger, B., Healy, D., Goebel, R., 2012. Real-time self-regulation of emotion networks in patients with depression. PLoS One 7, e38115.
Linhartová, P., Látalová, A., Kóša, B., Kašpárek, T., Schmahl, C., Paret, C., 2019. fMRI neurofeedback in emotion regulation: a literature review. NeuroImage 193, 75–92.
Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B., 2007. A review of classification algorithms for EEG-based brain–computer interfaces. J. Neural Eng. 4, R1.
Lubianiker, N., Goldway, N., Fruchtman-Steinbok, T., Paret, C., Keynan, J.N., Singer, N., Cohen, A., Kadosh, K.C., Linden, D.E., Hendler, T., 2019. Process-based framework for precise neuromodulation. Nat. Hum. Behav. 3, 436–445.
MacDuffie, K.E., MacInnes, J., Dickerson, K.C., Eddington, K.M., Strauman, T.J., Adcock, R.A., 2018. Single session real-time fMRI neurofeedback has a lasting impact on cognitive behavioral therapy strategies. Neuroimage Clin. 19, 868–875.
Mano, M., Lécuyer, A., Bannier, E., Perronnet, L., Noorzadeh, S., Barillot, C., 2017. How to build a hybrid neurofeedback platform combining EEG and fMRI. Front. Neurosci. 11, 140.
McFarland, D.J., McCane, L.M., Wolpaw, J.R., 1998. EEG-based communication and control: short-term role of feedback. IEEE Trans. Rehabil. Eng. 6, 7–11.
Megumi, F., Yamashita, A., Kawato, M., Imamizu, H., 2015. Functional MRI neurofeedback training on connectivity between two regions induces long-lasting changes in intrinsic functional network. Front. Hum. Neurosci. 9, 160.
Mehler, D.M.A., Kording, K.P., 2018. The lure of misleading causal statements in functional connectivity research. arXiv preprint arXiv:1812.03363.
Mehler, D.M.A., Kunas, S.L., Sokunbi, M.O., Goebel, R., Linden, D.E.J., 2021. Trajectories for mood states during a multi-session neurofeedback training intervention in major depressive disorder. PsyArXiv. https://doi.org/10.31234/osf.io/2msqp.
Mehler, D.M., Sokunbi, M.O., Habes, I., Barawi, K., Subramanian, L., Range, M., Evans, J., Hood, K., Lührs, M., Keedwell, P., 2018. Targeting the affective brain—a randomized controlled trial of real-time fMRI neurofeedback in patients with depression. Neuropsychopharmacology 43, 2578.
Mehler, D.M., Edelsbrunner, P.A., Matić, K., 2019. Appreciating the significance of nonsignificant findings in psychology. J. Eur. Psychol. Stud. 10.
Mehler, D., Williams, A.N., Whittaker, J., Krause, F., Lührs, M., Kunas, S., Wise, R., Shetty, H., Turner, D.L., Linden, D.E., 2020. Graded fMRI neurofeedback training of motor imagery in middle cerebral artery stroke patients-a preregistered proof-of-concept study. Front. Hum. Neurosci. 14, 226.
Micoulaud-Franchi, J.-A., Fovet, T., 2018. A Framework for Disentangling the Hyperbolic Truth of Neurofeedback: Comment on Thibault and Raz (2017).
Micoulaud-Franchi, J.-A., Batail, J.-M., Fovet, T., Philip, P., Cermolacce, M., Jaumard-Hakoun, A., Vialatte, F., 2019. Towards a pragmatic approach to a psychophysiological unit of analysis for mental and brain disorders: an EEG-copeia for neurofeedback. Appl. Psychophysiol. Biofeedback 1–22.
Mirifar, A., Beckmann, J., Ehrlenspiel, F., 2017. Neurofeedback as supplementary training for optimizing athletes' performance: a systematic review with implications for future research. Neurosci. Biobehav. Rev. 75, 419–432.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The, P.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6, e1000097.
Moncrieff, J., Wessely, S., Hardy, R., 2004. Active placebos versus antidepressants for depression. Cochrane Database Syst. Rev.
Moretti, D.V., Babiloni, F., Carducci, F., Cincotti, F., Remondini, E., Rossini, P., Salinari, S., Babiloni, C., 2003. Computerized processing of EEG–EOG–EMG artifacts for multi-centric studies in EEG oscillations and event-related potentials. Int. J. Psychophysiol. 47, 199–216.
Mulders, P.C., van Eijndhoven, P.F., Schene, A.H., Beckmann, C.F., Tendolkar, I., 2015. Resting-state functional connectivity in major depressive disorder: a review. Neurosci. Biobehav. Rev. 56, 330–344.
Müller-Putz, G., Leeb, R., Tangermann, M., Höhne, J., Kübler, A., Cincotti, F., Mattia, D., Rupp, R., Müller, K.-R., Millán, Jd.R., 2015. Towards noninvasive hybrid brain–computer interfaces: framework, practice, clinical application, and beyond. Proc. IEEE 103, 926–943.
Murphy, K., Birn, R.M., Bandettini, P.A., 2013. Resting-state fMRI confounds and cleanup. Neuroimage 80, 349–359.
Mutz, J., Vipulananthan, V., Carter, B., Hurlemann, R., Fu, C.H., 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, l1079.
Newcombe, R.G., 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat. Med. 17, 873–890.
Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.-B., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci. 20, 299–303.
Nielson, D.M., Keren, H., O'callaghan, G., Jackson, S.M., Douka, I., Zheng, C.Y., Belil, P. V.-R., Pornpattananangkul, N., Camp, C.C., Gorham, L.S., 2020. Great expectations: a critical review of and recommendations for the study of reward processing as a cause and predictor of depression. BioRxiv.
Nieuwenhuis, S., Forstmann, B.U., Wagenmakers, E.-J., 2011. Erroneous analyses of interactions in neuroscience: a problem of significance. Nat. Neurosci. 14, 1105–1107.
Nijholt, A., Allison, B.Z., Jacob, R.J., 2011. Brain-computer interaction: can multimodality help?. In: Proceedings of the 13th International Conference on Multimodal Interfaces. ACM, pp. 35–40.
Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., 2015. Promoting an open research culture. Science 348, 1422–1425.

Nosek, B.A., Ebersole, C.R., DeHaven, A.C., Mellor, D.T., 2018. The preregistration revolution. Proc. Natl. Acad. Sci. 115, 2600–2606.
Olfson, M., Blanco, C., Wall, M., Liu, S.-M., Saha, T.D., Pickering, R.P., Grant, B.F., 2017. National trends in suicide attempts among adults in the United States. JAMA Psychiatry 74, 1095–1103.
Otte, C., Gold, S.M., Penninx, B.W., Pariante, C.M., Etkin, A., Fava, M., Mohr, D.C., Schatzberg, A.F., 2016. Major depressive disorder. Nat. Rev. Dis. Primers 2, 16065.
Palmiero, M., Piccardi, L., 2017. Frontal EEG asymmetry of mood: a mini-review. Front. Behav. Neurosci. 11, 224.
Paquette, V., Beauregard, M., Beaulieu-Prévost, D., 2009. Effect of a psychoneurotherapy on brain electromagnetic tomography in individuals with major depressive disorder. Psychiatry Res. Neuroimaging 174, 231–239.
Paret, C., Goldway, N., Zich, C., Keynan, J.N., Hendler, T., Linden, D., Kadosh, K.C., 2019. Current progress in real-time functional magnetic resonance-based neurofeedback: methodological challenges and achievements. NeuroImage, 116107.
Patten, S.B., 2009. Accumulation of major depressive episodes over time in a prospective study indicates that retrospectively assessed lifetime prevalence estimates are too low. BMC Psychiatry 9, 19.
Peckham, A.D., McHugh, R.K., Otto, M.W., 2010. A meta-analysis of the magnitude of biased attention in depression. Depress. Anxiety 27, 1135–1142.
Peeters, F., Oehlen, M., Ronner, J., van Os, J., Lousberg, R., 2014. Neurofeedback as a treatment for major depressive disorder–a pilot study. PLoS One 9, e91837.
Perronnet, L., Lécuyer, A., Mano, M., Bannier, E., Lotte, F., Clerc, M., Barillot, C., 2017. Unimodal versus bimodal EEG-fMRI neurofeedback of a motor imagery task. Front. Hum. Neurosci. 11, 193.
Poldrack, R.A., Gorgolewski, K.J., 2014. Making big data open: data sharing in neuroimaging. Nat. Neurosci. 17, 1510.
Poldrack, R.A., Barch, D.M., Mitchell, J., Wager, T., Wagner, A.D., Devlin, J.T., Cumba, C., Koyejo, O., Milham, M., 2013. Toward open sharing of task-based fMRI data: the OpenfMRI project. Front. Neuroinform. 7, 12.
Quevedo, K., Liu, G., Teoh, J.Y., Ghosh, S., Zeffiro, T., Ahrweiler, N., Zhang, N., Wedan, R., Oh, S., Guercio, G., 2019. Neurofeedback and neuroplasticity of visual self-processing in depressed and healthy adolescents: a preliminary study. Dev. Cogn. Neurosci. 40, 100707.
Quevedo, K., Yuan Teoh, J., Engstrom, M., Wedan, R., Santana-Gonzalez, C., Zewde, B., Porter, D., Cohen Kadosh, K., 2020. Amygdala circuitry during neurofeedback training and symptoms' change in adolescents with varying depression. Front. Behav. Neurosci. 14, 110.
Ramirez, R., Palencia-Lefler, M., Giraldo, S., Vamvakousis, Z., 2015. Musical neurofeedback for treating depression in elderly people. Front. Neurosci. 9, 354.
Ramot, M., Kimmich, S., Gonzalez-Castillo, J., Roopchansingh, V., Popal, H., White, E., Gotts, S.J., Martin, A., 2017. Direct modulation of aberrant brain network connectivity through real-time NeuroFeedback. eLife 6, e28974.
Rance, M., Walsh, C., Sukhodolsky, D.G., Pittman, B., Qiu, M., Kichuk, S.A., Wasylink, S., Koller, W.N., Bloch, M., Gruner, P., 2018. Time course of clinical change following neurofeedback. Neuroimage 181, 807–813.
Ries, A.J., Touryan, J., Vettel, J., McDowell, K., Hairston, W.D., 2014. A comparison of electroencephalography signals acquired from conventional and mobile systems. J. Neurosci. Neuroeng. 3, 10–20.
Ros, T., Enriquez-Geppert, S., Zotev, V., Young, K., Wood, G., Whitfield-Gabrieli, S., Wan, F., Vialatte, F., Van De Ville, D., Todder, D., 2020. Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). Brain 143, 1674–1685.
Rosenfeld, J.P., Baehr, E., Baehr, R., Gotlib, I.H., Ranganath, C., 1996. Preliminary evidence that daily changes in frontal alpha asymmetry correlate with changes in affect in therapy sessions. Int. J. Psychophysiol. 23, 137–141.
Rossi, S., Hallett, M., Rossini, P.M., Pascual-Leone, A., Group, S.o.T.C, 2009. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clin. Neurophysiol. 120, 2008–2039.
Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Nierenberg, A.A., Stewart, J.W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D., 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report. Am. J. Psychiatry 163, 1905–1917.
Sacchet, M.D., Gotlib, I.H., 2016. Neurofeedback Training for Major Depressive Disorder: Recent Developments and Future Directions. Taylor & Francis.
Sackeim, H.A., Prudic, J., Fuller, R., Keilp, J., Lavori, P.W., Olfson, M., 2007. The cognitive effects of electroconvulsive therapy in community settings. Neuropsychopharmacology 32, 244–254.
Sato, H., Kawahara, J.-i., 2011. Selective bias in retrospective self-reports of negative mood states. Anxiety Stress Coping 24, 359–367.
Scheinost, D., Stoica, T., Saksa, J., Papademetris, X., Constable, R., Pittenger, C., Hampson, M., 2013. Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. Transl. Psychiatry 3, e250.
Schneider, F., Heimann, H., Mattes, R., Lutzenberger, W., Birbaumer, N., 1992. Self-regulation of slow cortical potentials in psychiatric patients: depression. Biofeedback Self. 17, 203–214.
Schönbrodt, F.D., Wagenmakers, E.-J., 2018. Bayes factor design analysis: planning for compelling evidence. Psychon. Bull. Rev. 25, 128–142.
Sharp, T., 2012. Molecular and cellular mechanisms of antidepressant action. Behavioral Neurobiology of Depression and Its Treatment. Springer, pp. 309–325.
Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. p-curve and effect size: correcting for publication bias using only significant results. Perspect. Psychol. Sci. 9, 666–681.
Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M.L., Rana, M., Oblak, E., 2017. Closed-loop brain training: the science of neurofeedback. Nat. Rev. Neurosci. 18, 86.
Skottnik, L., Linden, D.E.J., 2019. Mental imagery and brain regulation—New links between psychotherapy and neuroscience. Front. Psychiatry 10.
Sorger, B., Scharnowski, F., Linden, D.E., Hampson, M., Young, K.D., 2019. Control freaks: towards optimal selection of control conditions for fMRI neurofeedback studies. Neuroimage 186, 256–265.
Strangman, G., Boas, D.A., Sutton, J.P., 2002. Non-invasive neuroimaging using near-infrared light. Biol. Psychiatry 52, 679–693.
Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M.L., Bruehl, A.B., Cohen, L.G., DeCharms, R.C., Gassert, R., 2013. Real-time fMRI neurofeedback: progress and challenges. Neuroimage 76, 386–399.
Thibault, R.T., Raz, A., 2016. Neurofeedback: the power of psychosocial therapeutics. Lancet Psychiatry 3, e18.
Thibault, R.T., Lifshitz, M., Raz, A., 2016. The self-regulating brain and neurofeedback: experimental science and clinical promise. Cortex 74, 247–261.
Thibault, R.T., Lifshitz, M., Raz, A., 2017. Neurofeedback or neuroplacebo? Brain 140, 862–864.
Thibault, R.T., MacPherson, A., Lifshitz, M., Roth, R.R., Raz, A., 2018. Neurofeedback with fMRI: a critical systematic review. Neuroimage 172, 786–807.
Thibodeau, R., Jorgensen, R.S., Kim, S., 2006. Depression, anxiety, and resting frontal EEG asymmetry: a meta-analytic review. J. Abnorm. Psychol. 115, 715.
Tufanaru, C., Munn, Z., Aromataris, E., Campbell, J., Hopp, L., 2017. Chapter 3: Systematic reviews of effectiveness. Joanna Briggs Institute Reviewer's Manual. The Joanna Briggs Institute.
Van Doren, J., Arns, M., Heinrich, H., Vollebregt, M.A., Strehl, U., Loo, S.K., 2019. Sustained effects of neurofeedback in ADHD: a systematic review and meta-analysis. Eur. Child Adolesc. Psychiatry 28, 293–305.
Villringer, A., Planck, J., Hock, C., Schleinkofer, L., Dirnagl, U., 1993. Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neurosci. Lett. 154, 101–104.
Walker, J.E., Lawson, R., 2013. FP02 beta training for drug-resistant depression—A new protocol that usually reduces depression and keeps it reduced. J. Neurother. 17, 198–200.
Wang, S.-Y., Lin, I.-M., Peper, E., Chen, Y.-T., Tang, T.-C., Yeh, Y.-C., Tsai, Y.-C., Chu, C.-C., 2016. The efficacy of neurofeedback among patients with major depressive disorder: preliminary study. NeuroRegulation 3, 127–127.
Wang, T., Mantini, D., Gillebert, C.R., 2018. The potential of real-time fMRI neurofeedback for stroke rehabilitation: a systematic review. Cortex 107, 148–165.
Wang, S.-Y., Lin, I.-M., Fan, S.-Y., Tsai, Y.-C., Yen, C.-F., Yeh, Y.-C., Huang, M.-F., Lee, Y., Chiu, N.-M., Hung, C.-F., 2019. The effects of alpha asymmetry and high-beta down-training neurofeedback for patients with the major depressive disorder and anxiety symptoms. J. Affect. Disord. 257, 287–296.
Watanabe, T., Sasaki, Y., Shibata, K., Kawato, M., 2017. Advances in fMRI real-time neurofeedback. Trends Cogn. Sci. 21, 997–1010.
Weiskopf, N., 2012. Real-time fMRI and its application to neurofeedback. Neuroimage 62, 682–692.
Weiss, F., Zamoscik, V., Schmidt, S.N., Halli, P., Kirsch, P., Gerchen, M.F., 2020. Just a very expensive breathing training? Risk of respiratory artefacts in functional connectivity-based real-time fMRI neurofeedback. Neuroimage 210, 116580.
Winokur, G., 1997. All roads lead to depression: clinically homogeneous, etiologically heterogeneous. J. Affect. Disord. 45, 97–108.
Wood, G., Kober, S.E., 2018. EEG neurofeedback is under strong control of psychosocial factors. Appl. Psychophysiol. Biofeedback 43, 293–300.
Young, K.D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C., Bodurka, J., 2014. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. PLoS One 9, e88785.
Young, K.D., Misaki, M., Harmer, C.J., Victor, T., Zotev, V., Phillips, R., Siegle, G.J., Drevets, W.C., Bodurka, J., 2017a. Real-time functional magnetic resonance imaging amygdala neurofeedback changes positive information processing in major depressive disorder. Biol. Psychiatry 82, 578–586.
Young, K.D., Siegle, G.J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C., Bodurka, J., 2017b. Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. Am. J. Psychiatry 174, 748–755.
Young, K.D., Siegle, G.J., Misaki, M., Zotev, V., Phillips, R., Drevets, W.C., Bodurka, J., 2018a. Altered task-based and resting-state amygdala functional connectivity following real-time fMRI amygdala neurofeedback training in major depressive disorder. Neuroimage Clin. 17, 691–703.
Young, K.D., Zotev, V., Phillips, R., Misaki, M., Drevets, W.C., Bodurka, J., 2018b. Amygdala real-time functional magnetic resonance imaging neurofeedback for major depressive disorder: a review. Psychiatry Clin. Neurosci. 72, 466–481.
Yuan, H., Young, K.D., Phillips, R., Zotev, V., Misaki, M., Bodurka, J., 2014. Resting-state functional connectivity modulation and sustained changes after real-time functional magnetic resonance imaging neurofeedback training in depression. Brain Connect. 4, 690–701.
Zich, C., Johnstone, N., Lührs, M., Lisk, S., Haller, S.P., Lipp, A., Lau, J.Y., Kadosh, K.C., 2020. Modulatory effects of dynamic fMRI-based neurofeedback on emotion regulation networks in adolescent females. NeuroImage 220, 117053.
Zotev, V., Krueger, F., Phillips, R., Alvarez, R.P., Simmons, W.K., Bellgowan, P., Drevets, W.C., Bodurka, J., 2011. Self-regulation of amygdala activation using real-time fMRI neurofeedback. PLoS One 6, e24522.

Zotev, V., Phillips, R., Young, K.D., Drevets, W.C., Bodurka, J., 2013. Prefrontal control of the amygdala during real-time fMRI neurofeedback training of emotion regulation. PLoS One 8, e79184.
Zotev, V., Yuan, H., Misaki, M., Phillips, R., Young, K.D., Feldner, M.T., Bodurka, J., 2016. Correlation between amygdala BOLD activity and frontal EEG asymmetry during real-time fMRI neurofeedback training in patients with depression. Neuroimage Clin. 11, 224–238.
Zotev, V., Mayeli, A., Misaki, M., Bodurka, J., 2019. Emotion self-regulation training in major depressive disorder using simultaneous real-time fMRI and EEG neurofeedback. arXiv preprint arXiv:1909.05764.