Neuroscience and Biobehavioral Reviews 121 (2021) 291–306

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Review article

The efficacy of real-time functional magnetic resonance imaging neurofeedback for psychiatric illness: A meta-analysis of brain and behavioral outcomes

Emily Dudek [a], David Dodell-Feder [a,b,*]

[a] Department of Psychology, University of Rochester, United States
[b] Department of Neuroscience, University of Rochester Medical Center, United States

ARTICLE INFO

Keywords:
Neurofeedback
Neuromodulation
Real-time fMRI
Functional magnetic resonance imaging
Psychiatric illness
Psychopathology
Intervention
Meta-analysis

ABSTRACT

Real-time functional magnetic resonance imaging neurofeedback (rtfMRI-NF) has gained popularity as an experimental treatment for a variety of psychiatric illnesses. However, there has yet to be a quantitative review regarding its efficacy. Here, we present the first meta-analysis of rtfMRI-NF for psychiatric disorders, evaluating its impact on brain and behavioral outcomes. Our literature review identified 17 studies and 105 effect sizes across brain and behavioral outcomes. We find that rtfMRI-NF produces a medium-sized effect on neural activity during training ($g = .59$, 95 % CI [.44, .75], $p < .0001$), a large-sized effect after training when no neurofeedback is provided ($g = .84$, 95 % CI [.37, 1.31], $p = .005$), and small-sized effects for behavioral outcomes (symptoms $g = .37$, 95 % CI [.16, .58], $p = .002$; cognition $g = .23$, 95 % CI [-.33, .78], $p = .288$). Mixed-effects analyses revealed few moderators. Together, these data suggest a positive impact of rtfMRI-NF on brain and behavioral outcomes, although more research is needed to determine how rtfMRI-NF works, for whom, and under what circumstances.

1. Introduction

The last several decades have seen a substantial increase in our understanding and treatment of psychiatric illness (Brady et al., 2019; Casey et al., 2013; McNaught and Mink, 2011; Millan et al., 2016; Normandeau et al., 2017). Despite these scientific and clinical gains, gold-standard treatments for most psychiatric illnesses are far from panaceas, often carrying significant side-effects and high rates of discontinuation (Baldessarini et al., 1999; Bowden et al., 2005; Coldham et al., 2002; Fernandez et al., 2015; Gersh et al., 2017; Graham et al., 2011; Harrow et al., 2012; Kautzner et al., 2011; Lieberman et al., 2005; Quagliato et al., 2019; Rozental et al., 2018; Sonuga-Barke et al., 2013; Waltman et al., 2017; Wang et al., 2018; Wunderink et al., 2013). Even in cases where the gold-standard—whether it be pharmacological or psychosocial in nature—is shown to be efficacious, tolerable, and accessible, the prevalence of treatment-resistant illness remains high (Boylan et al., 2020; Patterson and Van Ameringen, 2017; Polese et al., 2019). One reason why standard treatments do not show greater efficacy is that they fail to mechanistically target pathophysiological mechanisms. Increasing research has shown that disruption to neural circuits is associated with the onset and progression of psychiatric illness (Brohawn et al., 2010; Brown and Morey, 2012; Buse et al., 2016; E. R. Duval et al., 2015; Li et al., 2009), and yet, few treatments directly target these circuits (Sitaram et al., 2017).

In recent years, several neuromodulatory techniques have emerged that show promise in their ability to mechanistically target neural circuits disrupted by psychiatric illness, without carrying the deleterious side-effects often associated with psychopharmacological intervention. One such technique is neurofeedback, which involves providing feedback to the participant based on their own neural signal as they engage in different mental processes. By providing a window into one's own neural activity as it unfolds over time, neurofeedback presents an opportunity to gain awareness around, and subsequently control important processes associated with those brain states, which may be impaired in psychiatric illness (deCharms, 2008). Neurofeedback first emerged as a form of operant conditioning in early studies that demonstrated self-regulation of electroencephalography (EEG) signals in animals and humans during the 1960s (Clemente et al., 1964; Fetz, 1969; Sterman and Wyrwicka, 1967). Following the inception of real-time functional magnetic resonance imaging (fMRI) in 1995 (Cox et al., 1995),

* Corresponding author at: Department of Psychology, University of Rochester, 453 Meliora Hall, Rochester, NY, 14627, United States.
E-mail address: d.dodell-feder@rochester.edu (D. Dodell-Feder).

https://doi.org/10.1016/j.neubiorev.2020.12.020
Received 23 June 2020; Received in revised form 1 December 2020; Accepted 18 December 2020
Available online 25 December 2020

neurofeedback was applied to fMRI, allowing for near instantaneous access to functional neuroimaging results of high spatial resolution. Contrary to standard therapeutic and pharmacological treatments, real-time fMRI neurofeedback (rtfMRI-NF) may offer specificity by directly targeting underlying neural circuits that are known to be implicated in psychopathology.

Neurofeedback has long been rooted in clinical applications, beginning with EEG-NF treatments for conditions such as epilepsy and anxiety in the 1970s (Hardt and Kamiya, 1978; Lubar and Shouse, 1976). In the first application of rtfMRI-NF for psychiatric illness, people with depression learned to upregulate individualized regions involved in positive emotion processing (Linden et al., 2012). In the years since, controlled rtfMRI-NF studies have demonstrated positive findings across a wide range of psychiatric disorders (Bauer et al., 2020; Sukhodolsky et al., 2020; Young et al., 2014; Zilverstand et al., 2015; Zotev et al., 2016). For example, the first systematic review of rtfMRI-NF (Thibault et al., 2018), which integrated results from 99 rtfMRI-NF studies with healthy and clinical samples, demonstrated that rtfMRI-NF can be used to train self-regulation of a variety of brain regions disrupted in psychiatric illness.

In addition to rtfMRI-NF's promise, many have validly highlighted its pitfalls (deCharms, 2007; Thibault et al., 2018; Weiskopf, 2012; Weiskopf et al., 2004), which are considerable. For example, rtfMRI-NF is inevitably costly and requires extensive technical setup to allow for real-time analysis (deCharms, 2007; Koush et al., 2017; Weiskopf, 2012; Weiskopf et al., 2004). The rtfMRI-NF signal itself is inherently limited due to noise and the hemodynamic response delay (deCharms, 2007; Oblak et al., 2017). Even if these issues could be appropriately dealt with, there remain open questions about optimal neurofeedback parameters (e.g., number of sessions, frequency of feedback, type of instructions; Heunis et al., 2020; Paret et al., 2019; Thibault et al., 2018). Many of these and related issues, for example, the extent to which rtfMRI-NF-induced brain changes translate to meaningful clinical change, have been incisively raised and discussed by others (e.g., Thibault et al., 2018).

Despite the increasing promise of rtfMRI-NF, and the growing need to address these open questions, most notably, the study parameters that are most likely to result in rtfMRI-NF-induced brain and behavior change, there has yet to be a quantitative synthesis of rtfMRI-NF studies for the treatment of psychiatric illness. Such an analysis would provide a formal estimate of rtfMRI-NF efficacy, and may help to uncover study parameters that maximize effects for future studies. Thus, here, we provide the first quantitative analysis of rtfMRI-NF studies of brain and behavioral outcomes in clinical populations. Our primary aim was to determine the ability of rtfMRI-NF to successfully modulate neural activity (in the expected region and direction) and behavior. To do so, we conducted a systematic review and meta-analysis of controlled rtfMRI-NF experiments with psychiatric samples. Our meta-analysis addressed four primary questions: 1) Does rtfMRI-NF lead to volitional control of neural activity as evaluated during "training" tasks when feedback is provided to participants from the targeted region(s)-of-interest? 2) Do neural effects persist after training as evaluated by "transfer" tasks in which no neurofeedback is provided? 3) Does rtfMRI-NF lead to changes in behavior, including symptom and cognition outcomes, as well as Research Domain Criteria (RDoC; Insel et al., 2010) defined outcomes? 4) Finally, because there remains significant heterogeneity across rtfMRI-NF protocols, are brain and behavioral outcomes impacted by study characteristics? Our goal with these analyses was to provide critical information for future clinical studies by determining the magnitude of rtfMRI-NF's impact on brain and behavioral outcomes, and uncovering optimal rtfMRI-NF parameters.

2. Methods

2.1. Study selection and inclusion criteria

We searched for studies in two rounds. The initial study search began on April 7, 2019 by searching PubMed for articles in English using search terms related to fMRI, neurofeedback, and psychiatric illness. Within each category of terms, we included variations of the word and associated features (e.g., functional neuroimaging, NFT, psychiatric illness). The search terms were as follows: (fMRI OR functional MRI OR functional magnetic resonance imaging OR functional neuroimaging OR rtfMRI OR rt-fMRI OR rt-functional magnetic resonance imaging OR functional rtMRI OR functional rt-MRI) AND (neurofeedback OR neurofeedback OR neural feedback OR NF OR NFB OR NFT OR feedback-training OR real time OR real-time OR realtime OR self-regulate) AND (dsm OR diagnostic statistical manual OR psychiat* OR symptom*). In order to avoid omitting newer publications, we conducted a second round of the same search on October 27, 2019 in PubMed, PsycInfo, Web of Science, and, in order to look for unpublished literature, we also searched several preprint servers including PsyArXiv, PrePubMed, and BioRxiv. Lastly, we reviewed reference sections of relevant review papers, monitored listservs, and created Google Scholar alerts using the above search terms to identify any additional papers not returned by the databases.

In order to be included in the analysis, the study needed to meet the following criteria. First, the study involved the presentation of neurofeedback via rtfMRI methods. Second, the effects of rtfMRI-NF were compared with a control condition. Third, the study sample consisted of participants with a DSM disorder who were compared to other participants with the same disorder. Fourth, the study reported at least one brain and/or behavioral outcome with the necessary statistics (provided in the manuscript or by the author) to calculate effect sizes. Fifth, the effect of neurofeedback was evaluated in the region(s) targeted for neurofeedback. For example, if the source of neurofeedback was the amygdala, the authors needed to have analyzed changes in the amygdala. This criterion meant we excluded studies that only reported whole-brain analyses instead of hypothesis-driven region-of-interest analyses.

Records were screened using Abstrackr (http://abstrackr.cebm.brown.edu), an online tool for screening articles in systematic reviews that allows for simultaneous independent screening, organizational tags, and a machine learning algorithm with screening predictions (Wallace et al., 2012). Screening was performed by the first author under supervision of the second author. For relevant papers that did not include the necessary statistics for calculating an effect size, the corresponding author was contacted for the information. When contacting authors, we also asked whether they had any unpublished data from a rtfMRI-NF study, although authors did not provide any additional data that we had not already found through our other search methods. See Fig. 2 for an overview of study screening and selection.

2.2. Data extraction and study coding

All data were extracted independently by the two authors. We coded for the following variables with generally high inter-rater reliability (*Mdn* Cohen's $\kappa = 1.00$ and ICC = 1.00; see Supplementary Material for all values):

2.2.1. Publication characteristics

This included author(s), year, and publication status (published/unpublished).

2.2.2. Participant characteristics

This included *M* age of sample, *M* age of active group, *M* age of control group, percentage of female participants in entire sample, percentage of female participants in active group, percentage of female participants in control group, DSM-5 superordinate category included in the sample (e.g., neurodevelopmental disorder), specific DSM diagnosis (e.g., attention deficit hyperactivity disorder), and whether the sample was medicated or not.

2.2.3. Control condition

Researchers compared rtfMRI-NF to a variety of control conditions, which we classified initially into four categories. Studies in which the control group received no neurofeedback were classified as *no feedback*. Studies in which the control group received neurofeedback, physiological feedback, or any other kind of true feedback signal from the participant that wasn't related to the feedback signal of interest were classified as *non-hypothesized signal* feedback. We classified two other forms of feedback including *other person feedback* in which the participant received a feedback signal derived from another participant, and *random feedback* in which the feedback signal was randomly generated. We found only two studies with these latter two conditions. Thus, for parsimony, we collapsed these two categories into a *sham feedback* category that included *non-hypothesized signal*, *other person*, and *random feedback*.

2.2.4. Instruction

Two aspects of experimenter instruction were coded including what participants were told to do during the rtfMRI-NF paradigm (if anything), and the direction in which they should regulate the neural signal. Participants were generally provided either *explicit* or *implicit* instructions for regulating the targeted neural signal. Explicit paradigms are those that involve presenting the participant with any kind of instruction—explicit, vague, or otherwise—for how the participant should attempt to regulate the targeted neural signal. Implicit paradigms are those in which participants are given no such guidance. In addition to the nature of the instruction, we coded whether participants were told to *upregulate* neural signal, *downregulate* neural signal, or to do both.

2.2.5. Feedback signal

The neurofeedback delivered to participants varied as a function of several characteristics including frequency, format, signal type, and signal origin (i.e., the ROI(s)). We coded frequency as either *continuous*, in which feedback was provided after each volume was acquired, or *intermittent*, in which feedback was provided after collecting several volumes as a summary measure (e.g., *M*, *Mdn*). We coded format in terms of how the feedback was visually presented to participants, which took the form of either *line graph*, *thermometer*, or *video*. We coded signal type as either reflecting *percent signal change* or *connectivity*. We coded several aspects of signal origin, including the names of the specific ROIs, and the number of ROIs used to calculate the feedback signal.

2.2.6. Duration

This included the number of separate testing sessions, and the total number of minutes spent performing the neurofeedback task (not including time spent by the participant performing a control task or resting).

2.2.7. Brain outcomes

We coded two brain outcomes towards calculating effect sizes that would address our two primary questions of interest. To address whether compared to control conditions, rtfMRI-NF training leads to the expected neural change, we extracted group *M*s and *SD*s of the brain outcome variable (i.e., percent signal change, connectivity, etc.) for the last rtfMRI-NF *training* session; that is, the last session during which participants were provided with neurofeedback. To address whether the effect of rtfMRI-NF training generalizes to a context with no rtfMRI-NF, we extracted group *M*s and *SD*s of the brain outcome variable for the last *transfer* session; that is, the last session in which participants performed the neurofeedback task without receiving neurofeedback. Accompanying group *n*s were also extracted. When *M*s and *SD*s were not reported in the text or a table, we extracted these values from relevant plots with WebPlotDigitizer (Rohatgi, 2019), which is a validated tool for extracting numerical values from figures (Arora et al., 2020; Castrellon et al., 2020; Drevon et al., 2017; Kip et al., 2020; Lim et al., 2017). We note that there was one instance in which we extracted values using WebPlotDigitizer only to later receive the values from the authors (Young et al., 2017). The values we initially extracted and the values we later received from the authors were near perfectly correlated, $r(6) = .99$ (differences were due to rounding error), confirming the accuracy of extracting data using this method. In cases where the necessary statistics were not reported in the paper nor were there plots from which we could extract these data, we directly contacted the authors for this information.

2.2.8. Behavioral outcomes

To address whether rtfMRI-NF impacted behavioral outcomes, we coded the necessary statistics for calculating effect sizes: group *M*s and *SD*s, and *n*s for all non-brain outcomes. Similar to the brain outcome values, when relevant statistics were not reported in text, we extracted these data from plots included in the paper using WebPlotDigitizer. If no relevant plots were included, authors were contacted directly for this information. We classified behavioral outcomes as assessing either *symptoms* or *cognition*. Symptom measures were those that assessed characteristics that in part defined the disorder being studied (e.g., the Beck Depression Inventory in a study of individuals with major depressive disorder; Young et al., 2017), and cognitive measures were those that assessed aspects of cognition implicated in the disorder (e.g., Go/No-Go Task in a study of individuals with ADHD; Alegria et al., 2017). Given that this distinction was not always clear, and towards grouping the behavioral outcomes in other potentially meaningful ways, we also used the RDoC Initiative (Insel et al., 2010) to classify behavioral outcomes into those assessing *negative valence systems*, *positive valence systems*, *cognitive systems*, *social processes*, *arousal and regulatory systems*, or *sensorimotor systems*. Previous work has demonstrated the benefits of this framework for conceptualizing components of psychopathology broadly including suicide risk (Glenn et al., 2017), hallucinations (Badcock and Hugdahl, 2014; Ford, 2016), and eating disorder symptoms (Wildes & Marcus, 2015). Beyond the advantages of its mechanistic specificity, the RDoC framework is particularly apt for our dataset, which spans diagnoses yet includes many shared clinical measures and outcomes.

2.3. Statistical analyses

We conducted all analyses in R (version 3.5.1; R Core Team, 2018) using the *metafor* package (Viechtbauer, 2010). The data are available on the Open Science Framework repository at the following link: https://osf.io/3qn2k/?view_only = 6b92982a56304c138bde24d337cf7422

2.3.1. Effect size calculation

We calculated the standardized mean difference as bias-corrected Hedges' *g*, which we interpret using conventional benchmarks (Cohen, 1988). Results were considered unexpected under the null hypothesis when *p* was less than .05 (hereafter referred to as "statistically significant"). There was one instance (Bauer et al., 2020) in which a cross-over design was used in which the same participants completed rtfMRI-NF training and then a control training. In this instance, we calculated the standardized mean change score using raw score standardization (Morris and DeShon, 2002). Based on a prior meta-analysis of the reliability of task-related fMRI response (Bennett and Miller, 2010), we assumed a correlation between the brain outcome variables of $r = .5$

when calculating the latter effect sizes. There was one instance whereby a behavioral outcome tested with an ANOVA was described as being not significant without providing relevant statistics (Linden et al., 2012). We conservatively estimated this effect to be 0. Additionally, there was one behavioral effect size that we were unable to calculate due to a pooled standard deviation of 0 (Alegria et al., 2017).

Data were coded in such a way that a positive effect size indicates the predicted neurofeedback effect. For example, if the rtfMRI-NF group was trained to upregulate percent signal change, a positive Hedges' *g* would mean that compared to the control group, the active group showed greater increases in percent signal change. Alternatively, if the active group was trained to downregulate percent signal change, a positive Hedges' *g* would mean that compared to the control group, the active group showed greater decreases in percent signal change.

2.3.2. Data synthesis

We had three primary questions of interest: (1) Does rtfMRI-NF training lead to predicted rtfMRI-NF-related neural changes during training? (2) Does rtfMRI-NF training lead to predicted rtfMRI-NF-related neural changes during transfer (i.e., when no feedback is provided), and (3) Does rtfMRI-NF training lead to predicted changes in behavior? We addressed the first two questions by meta-analyzing the group difference in brain value for the last training session and the last transfer session, respectively. Because the behavior outcomes assessed were broad, we conducted several meta-analyses to address the third question. First, using the symptom/cognition classification, we conducted separate meta-analyses for *symptom* and *cognition* outcomes. Second, using the RDoC Matrix, we conducted five follow-up meta-analyses, evaluating outcomes assessing the respective RDoC constructs (we did not conduct a meta-analysis for RDoC arousal and regulatory systems since all three outcomes came from the same study; Young et al., 2014).

In each meta-analysis, there was at least one instance in which more than one effect size was derived from the same sample in the same study. To deal with the statistical dependence, we used a three-level random-effects model in which we added a random effect for study (Cheung, 2014, 2019; Konstantopoulos, 2011; Van den Noortgate et al., 2013, 2015). This model allowed effects to vary at the level of sampling variance (level 1), within-study variance (level 2), and between-study variance (level 3). We compared the fit of the three-level model to that of a two-level model (using maximum likelihood estimation) by fixing level 3, and separately, level 2 variance to 0, and performing likelihood ratio tests on the full and reduced model (Assink and Wibbelink, 2016). A three-level model better fit the data for the symptom and RDoC negative valence meta-analyses; for all other meta-analyses, a two-level model better fit the data. In the results, we present the findings from the better fitting model using restricted maximum likelihood to estimate residual heterogeneity. To account for correlated sampling errors due to multiple effect sizes being derived from the sample, we generated cluster-robust tests and confidence intervals of model estimates (Hedges et al., 2010) using the *robust* function in *metafor*.

We assessed the presence and extent of heterogeneity with the Q and I^2 statistics, respectively. The Q statistic and its p value provides a test of the hypothesis that all studies share a common effect size. A statistically significant Q value ($p < .10$) indicates that the true effects vary. I^2 indicates the proportion of observed variance attributable to true variance among the effect sizes as opposed to sampling error (Borenstein et al., 2017). An I^2 value of 0% indicates that the none of the variation among the observed effects is due to variation in true effects. When interpreted alongside a forest plot, $I^2 = 0\%$ indicates that none of the variance among effects sizes would remain if sampling error was reduced to 0 (Borenstein et al., 2017). An I^2 value of 100 % indicates that all of the variation among the observed effects is due to variation in true effects. When interpreted alongside a forest plot, $I^2 = 100$ % indicates that all of the variance among effects sizes would remain if sampling error was reduced to 0. We interpret I^2 using benchmarks provided by Deeks et al. (2008). Absolute variance of the true effects is provided as τ^2.

2.3.3. Outlier and influence diagnostics

We evaluated each meta-analysis for influential outliers, which, following Viechtbauer and Cheung (2010), we defined as effect sizes with studentized residuals ±1.96 and Cook's d value larger than the 50th percentile of chi-square distribution. In analyses where we identified an influential outlier, we re-ran the analysis without that effect size.

2.3.4. Moderator analysis

To investigate whether study attributes impacted the effect of rtfMRI-NF on brain and behavioral outcomes, we conducted a series of mixed-effects moderator analyses evaluating the impact of the following variables: DSM diagnosis, number of rtfMRI-NF sessions (dichotomized into one versus more than one session), minutes of rtfMRI-NF training, control group type, feedback frequency, direction of rtfMRI-NF regulation, and type of rtfMRI-NF feedback. We decided not to include medication as a moderator given the variability in medication type and dosage, as well as a lack of information regarding the circumstances of pharmacological treatment (e.g., randomization to treatment, self-selection, etc.). In the case of a moderating effect, these factors would prevent us making meaningful conclusions regarding the impact of medication on rtfMRI-NF training. For categorical variables, we conducted these analyses only when there were at least two effects sizes in each level of a variable coming from different studies and samples. In cases where a study attribute contained more than two levels, we dropped any level not containing at least two effect sizes each from a different study before running the analysis. Moderator effects were evaluated with the F statistic and its p value, which indicate whether the relation between the moderator and rtfMRI-NF effect is stronger than would be expected by chance. We provide pseudo R^2 values denoting the percentage of heterogeneity accounted for when including the moderator in the model (Raudenbush, 2009). We note that these values may be inaccurate in analyses where the number of studies is small (López-López et al., 2014)—as is the case in several of our analyses—and so we encourage caution when interpreting these values.

2.3.5. Publication bias and sensitivity analyses

The validity of a meta-analytic finding depends on whether the meta-analysis incorporates all of the available relevant data. Selective reporting of results from a study and in the literature more broadly based on statistical significance or other conditions—i.e., publication bias—creates a situation in which the available data to analyze is not representative of the population of studies. This represents a fundamental threat to the validity of a meta-analysis. A variety of methods exist for detecting and correcting publication bias as well as other forms of bias (e.g., questionable research practices; John et al., 2012; Simmons et al., 2011), which are more or less effective depending on the meta-analytic conditions (Carter et al., 2019). However, few such methods exist that appropriately handle dependent effect sizes. Two recent simulation papers suggest that Egger's regression test (ERT) may be appropriate in the case of dependent effect sizes when using multilevel models or cluster-robust tests (Fernández Castilla, 2019; Rodgers and Pustejovsky, 2019). In ERT, effect sizes are regressed on their standard errors (Egger et al., 1997). A statistically significant slope (b_1) indicates an association between effect sizes and their precision, meaning that smaller, less precise studies, consistently produce larger effects. An important limitation of this method is that it does not speak to the underlying cause of the association, which may be due to "small-study effects" (i.e., smaller studies producing larger effects for reasons other than selection bias such as methodological differences between small and larger studies; Sterne et al., 2001; Sterne and Egger, 2006) or publication bias. It is useful to consider ERT in the context of funnel plots; a widely used visual tool for assessing small-study effects that plots effect sizes against their precision (typically standard errors). In the absence of small-study

effects (of which publication bias may be a cause), smaller, less precise studies should scatter widely at the bottom of the plot, while larger, more precise studies should cluster at the top, resembling a symmetrical funnel (Sterne and Egger, 2006). In the presence of small-study effects (of which publication bias maybe a cause), the plot will be asymmetrical. ERT is a formal evaluation of funnel plot asymmetry.

Here, we evaluate publication bias by conducting cluster-robust ERT, and, to increase interpretability of funnel plot asymmetry, provide contour-enhanced funnel plots, which depict areas of conventional statistical significance (e.g., $p < .05$; Peters et al., 2008). If funnel plot asymmetry seems to be caused by missing studies in areas of statistical non-significance, to the extent that selective reporting is based on statistical significance, publication bias may be more likely to be assumed as the source of asymmetry. This method too has its limitations, which includes the inherently subjective nature in interpreting the plots (Terrin et al., 2005). Finally, for thoroughness, we use the trim and fill (TAF) method using the R0 estimator (Duval and Tweedie, 2000), which attempts to create symmetry in the funnel plot by imputing "missing" studies, correcting the overall effect size estimate with these missing studies, and then testing the hypothesis that the number of missing studies is 0. Despite its use in multilevel meta-analysis (e.g., Weisz et al., 2017) and cluster-robust meta-analytic tests (e.g., Clark et al., 2016), we note an important limitation of this method too in that it is not designed to handle dependent effect sizes, and minimally reduces bias and Type I error rates even in the case of standard two-level meta-analyses (Carter et al., 2019). Taking into account these limitations, we treated the findings from all of the methods as a form of sensitivity analysis, and weighted the findings of ERT more heavily given its validation for dependent effect sizes.

3. Results

3.1. Participant and study characteristics

Our search returned 17 relevant studies (2 unpublished) with 105 effect sizes across brain and behavioral outcomes (brain effect sizes $n = 25$, symptom effect sizes $n = 62$, cognition effect sizes $n = 18$; Table 1, Fig. 2). In total, the studies included 410 participants, 234 of whom received rtfMRI-NF. Participants were on average 34.1 ± 9.9 years old, and 50.7 % of the sample was female. Several psychiatric disorders were studied, the most common disorder being major depressive disorder (52.9 % of studies; 60.0 % of effect sizes; Table 2). Over half of the participant samples included (52.9 %) were on psychotropic medication.

On average, participants completed 2.3 ± 1.3 sessions (min-max = 1–4) of rtfMRI-NF with an average total regulation time across sessions of 23.5 ± 18.1 min (min-max = 5–57.2). Active neurofeedback was most often compared to a sham feedback control condition (70.6 %). The overwhelming majority of studies (88.2 %) provided explicit instructions for regulating the neural signal. Most studies asked participants to upregulate the neural signal (64.7 %), which was most often task-based activation (e.g., percent-signal change; 82.4 %) as opposed to connectivity, and provided participants with continuous neurofeedback (82.4 %). A variety of brain regions were used as the source of the neurofeedback signal with the most common ROI being the amygdala (35.3 %). The majority of studies (70.6 %) included a transfer task in which no neurofeedback was provided. See Fig. 1 and Table 1 for more details regarding study characteristics.

3.2. The neural effect of rtfMRI-NF during the training task

3.2.1. Overall effect

We first asked whether compared to control training, rtfMRI-NF led to the predicted changes in neural signal during the training task in which participants received neurofeedback. Data were analyzed from 12 studies contributing to a total of 16 effect sizes. We found a medium-sized-effect of rtfMRI-NF during the training task, $g = .52$, 95 % CI [.34, .71], which was statistically significant, $p < .0001$. We identified one influential outlier, which demonstrated a negative effect of rtfMRI-NF (Bauer et al., 2020). After removing this outlier, the effect was slightly larger, although similar, $g = .59$, 95 % CI [.44, .75], $p < .0001$ (Table 3, Fig. 3A). Regarding heterogeneity, the Q test was not statistically significant, and I^2 was 0% indicating that none of the variance was due to variation in the true effects. Rather than there being no variation in true effects, it is likely that this finding is being driven, at least in part, by the fairly small sample sizes, which led to larger CIs with greater overlap. Regarding publication bias, the funnel plot showed the majority of studies in areas of non-statistical significance, with some studies tracking the line of statistical significance ($p < .05$), though some of this clustering was the result of effect sizes coming from the same study (see Fig. S2). ERT was not statistically significant, b_1=-.71, 95 % CI [-3.15, 1.72], $p = .524$, and the TAF estimate was similar, $g = .61$, 95 % CI [.41, .81], $p < .001$, imputing one missing study on the right side of the plot, which was not statistically significant, $p = .250$.

3.2.2. Moderator analysis

Given that all the variance could be attributed to sampling variance ($\tau^2 = 0$; $I^2 = 0\%$), as opposed to variation in true effects, we did not perform moderator analyses in order to reduce the likelihood of Type I error.

3.3. The neural effect of rtfMRI-NF during the transfer task

3.3.1. Overall effect

Next, we asked whether compared to control training, rtfMRI-NF led to the predicted changes in neural signal during the transfer task in which participants received no neurofeedback. This task serves as a test of generalizability to evaluate whether volitional control is observed in instances in which the training signal is not provided. Data were analyzed from 9 studies contributing 9 effect sizes total. We found a medium-sized-effect of rtfMRI-NF during the transfer task, $g = .68$, 95 % CI [.13, 1.23], which was statistically significant, $p = .022$. We identified one medium effect size as an influential outlier in which the control condition outperformed rtfMRI-NF (Sukhodolsky et al., 2020). After removal of this effect size, the effect of rtfMRI-NF on neural signal during the transfer task increased to a large effect, $g = .84$, 95 % CI [.37, 1.31], which was statistically significant, $p = .005$ (Table 3, Fig. 3B). The Q test for heterogeneity was not statistically significant, $Q = 10.51$, $p = .162$, and the proportion of observed variance attributable to true variance may represent moderate heterogeneity, $I^2 = 31.5$ %. The majority of studies fell in the area of non-statistical significance in the contour-enhanced funnel plot, with some clustering of studies close to the line of statistical significance, which was in part due to effect sizes coming from the same studies (Fig. S2). ERT was not statistically significant, b_1=-5.72, 95 % CI [-17.12, 5.68], $p = .253$, and the TAF estimate was the same, $g = .84$, 95 % CI [.46, 1.22], $p < .0001$, imputing zero missing studies, $p = .500$.

3.3.2. Moderator analysis

We were able to conduct moderator analyses for DSM diagnosis (restricting the analyses to studies of MDD and ADHD due to a limited number of effect sizes from other diagnoses), number of sessions, and training time (Table 4). Though the inclusion of DSM diagnosis and training time substantially reduced unexplained variance (DSM diagnosis $R^2 = 34.8$ %, training time $R^2 = 48.1$ %), this reduction in variance was not statistically significant, nor was the reduction in variance with inclusion of number of sessions ($R^2 = 0\%$). We note that including a category for other diagnoses (i.e., non-ADHD, non-MDD; $n = 2$) also did not change the findings of the DSM diagnosis moderator analysis, $F(2, 5) = 2.05$, $p = .224$, $R^2 = 37.2$ %, other $g = .11$, 95 % CI [-1.26, 1.49], $p = .843$.

Table 1
Study characteristics.

Study	Diagnosis	Control Group	Sample Size	Mean Age	Training Paradigm	Timing	Transfer	Behavioral Outcomes
Alegria et al. (2017)	ADHD	Sham Feedback	31 (EG = 18, CG = 13)	13.90	*ROI* = rIFG *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = implicit	4 sessions, 55 min total	Yes	ADHD Symptoms Motor Inhibition Sustained Attention Time Perception
Bauer et al. (2020)	SZ	Sham Feedback	11 (EG = 11, CG = 11)	43.50	*ROI* = DMN-CEN *Feedback* = connectivity *Direction* = decrease *Frequency* = continuous *Instructions* = explicit	1 session, 8 min total	Yes	Auditory Hallucinations
Hamilton et al. (2016)	MDD	Sham Feedback	20 (EG = 10, CG = 10)	32.85	*ROI* = salience network *Feedback* = PSC *Direction* = decrease *Frequency* = intermittent *Instructions* = explicit	1 session, 5.4 min total	Yes	Mood
Hartwell et al. (2016)	Nicotine Dependence	No Feedback	33 (EG = 16, CG = 17)	35.18	*ROI* = ACC/PFC *Feedback* = PSC *Direction* = decrease *Frequency* = intermittent *Instructions* = explicit	3 sessions, 16.5 min total	No	Craving
Jaeckle et al. (2019)	MDD	No Feedback	35 (EG = 19, CG = 16)	37.15	*ROI* = ATL & SCC *Feedback* = connectivity *Direction* = decrease *Frequency* = continuous *Instructions* = explicit	3 sessions, 33.6 min total	No	Depression Mood Self-Blame Self-Esteem
Linden et al. (2012)	MDD	No Feedback	16 (EG = 8, CG = 8)	48.44	*ROI* = VLPFC, insula, DLPFC, MTL, OFC *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = implicit	4 sessions, 42 min total	No	Depression Mood
Mehler et al. (2018)	MDD	Sham Feedback	32 (EG = 16, CG = 16)	47.07	*ROI* = variable *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = explicit	4 sessions, 32 min total	Yes	Anxiety Depression Motivation Self-Efficacy Thought Control
Misaki et al. (2018)	PTSD	Sham Feedback	22 (EG = 16, CG = 6)	30.27	*ROI* = left amygdala *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = explicit	3 sessions, 24 min total	Yes	Anxiety
Sukhodolsky et al. (2020)	Tourette's Disorder	Sham Feedback	21 (EG = 11, CG = 10)	16.05	*ROI* = SMA *Feedback* = PSC *Direction* = both *Frequency* = continuous *Instructions* = explicit	2 sessions, 57.2 min total	Yes	Tic Severity
Young et al. (2014)	MDD	Sham Feedback	21 (EG = 14, CG = 7)	37.33	*ROI* = left amygdala *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = explicit	1 session, 8 min total	Yes	Anxiety Mood
Young et al. (2017)	MDD	Sham Feedback	33 (EG = 18, CG = 15)	31.55	*ROI* = left amygdala *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = explicit	2 sessions, 16 min total	Yes	Anhedonia Anxiety Depression Memory
Zahn et al. (2019)	MDD	Sham Feedback	28 (EG = 14, CG = 14)	45.20	*ROI* = ATL & SCC *Feedback* = connectivity *Direction* = increase *Frequency* = continuous *Instructions* = explicit	1 session, 8 min total	No	Depression Guilt Indignation Mood Self-Esteem
Zilverstand et al. (2015)	Specific Phobia	No Feedback	18 (EG = 9, CG = 9)	21.20	*ROI* = DLPFC, insula *Feedback* = PSC *Direction* = both *Frequency* = intermittent *Instructions* = explicit	1 session, 5 min total	No	Anxiety Spider Fear
Zilverstand et al. (2017)	ADHD	No Feedback	13 (EG = 7, CG = 6)	36.68	*ROI* = dorsal ACC *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = explicit	4 sessions, 48 min total	Yes	ADHD Symptoms Cognitive Interference Sustained Attention Working Memory
Zotev et al. (2016)	MDD	Sham Feedback		37.79	*ROI* = left amygdala *Feedback* = PSC	1 session, 8 min total	Yes	Anxiety Mood

(*continued on next page*)

Table 1 (*continued*)

Study	Diagnosis	Control Group	Sample Size	Mean Age	Training Paradigm	Timing	Transfer	Behavioral Outcomes
			24 (EG = 13, CG = 11)		*Direction* = increase *Frequency* = continuous *Instructions* = explicit			
Zotev et al. (2018)	PTSD	Sham Feedback	28 (EG = 18, CG = 10)	33.09	*ROI* = left amygdala *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = explicit	3 sessions, 24 min total	Yes	Depression PTSD Symptoms
Zotev et al. (2020)	MDD	Sham Feedback	24 (EG = 16, CG = 8)	32.67	*ROI* = left amygdala *Feedback* = PSC *Direction* = increase *Frequency* = continuous *Instructions* = explicit	1 session, 8 min total	Yes	Anxiety Mood

Note. Abbreviations in alphabetical order: ACC = anterior cingulate cortex; ADHD = attention deficit hyperactivity disorder; ATL = anterior temporal lobe; CEN = central executive network; CG = control group; DLPFC = dorsolateral prefrontal cortex; DMN = default mode network; EG = experimental group; MDD = major depressive disorder; MTL = medial temporal lobe; OFC = orbitofrontal cortex; PFC = prefrontal cortex; PSC = percent signal change; PTSD = posttraumatic stress disorder; ROI = region of interest; SCC = subgenual cingulate cortex; SMA = supplementary motor area; SZ = schizophrenia; VLPFC = ventrolateral prefrontal cortex.

Table 2
Count of DSM diagnoses across studies and effect sizes.

DSM Category	Number of Studies			Number of Effect Sizes		
	Brain	*Symptom*	*Cognition*	*Brain*	*Symptom*	*Cognition*
Depressive Disorders	6	9	2	10	47	6
Neurodevelopmental Disorders	3	3	2	7	6	12
Trauma and Stressor Related Disorders	1	2	0	2	5	0
Schizophrenia and other Psychotic Disorders	1	1	0	3	1	0
Anxiety Disorders	1	1	0	2	2	0
Substance-Related and Addictive Disorders	1	1	0	1	1	0

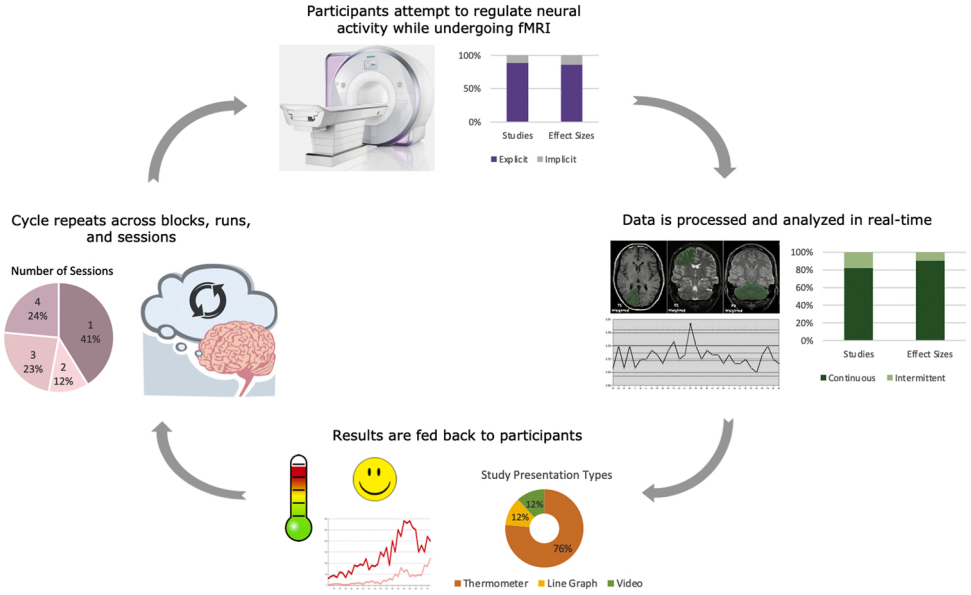

Fig. 1. Depiction of the rtfMRI-NF protocol with statistics from the studies included in our analysis. RtfMRI-NF software image depicts OpenNFT software.

3.4. The effect of rtfMRI-NF on behavior

3.4.1. Overall effect

In addition to neural outcomes, we also evaluated whether compared to control training, rtfMRI-NF improved behavioral outcomes, including symptoms and cognition. All 17 of the included studies assessed symptom outcomes, contributing to a total of 62 effect sizes. We present the findings from the three-level model, which better fit the data than the two-level model. Compared to control training, rtfMRI-NF was associated with a small advantage in symptom reduction, g = .37, 95 % CI [.16, .58], which was statistically significant, p = .002 (Table 3, Fig. 4). The Q test for heterogeneity was not statistically significant, Q = 75.16, p = .105, and the proportion of variance attributable to true variance may represent moderate heterogeneity, I^2 = 36.48 %. All of the variance was attributable to between-study (i.e., level 3) differences in effect sizes. The majority of studies fell within the area of non-statistical significance, with some tracking of effects around the line of significance (Fig. S2). ERT was statistically significant, b_1 = 4.40, 95 % CI [1.13, 7.67], p = .012, and the TAF estimate (derived on the basis of a two-level symptom model) was smaller in magnitude, g = .25, 95 % CI [.13, .37], p

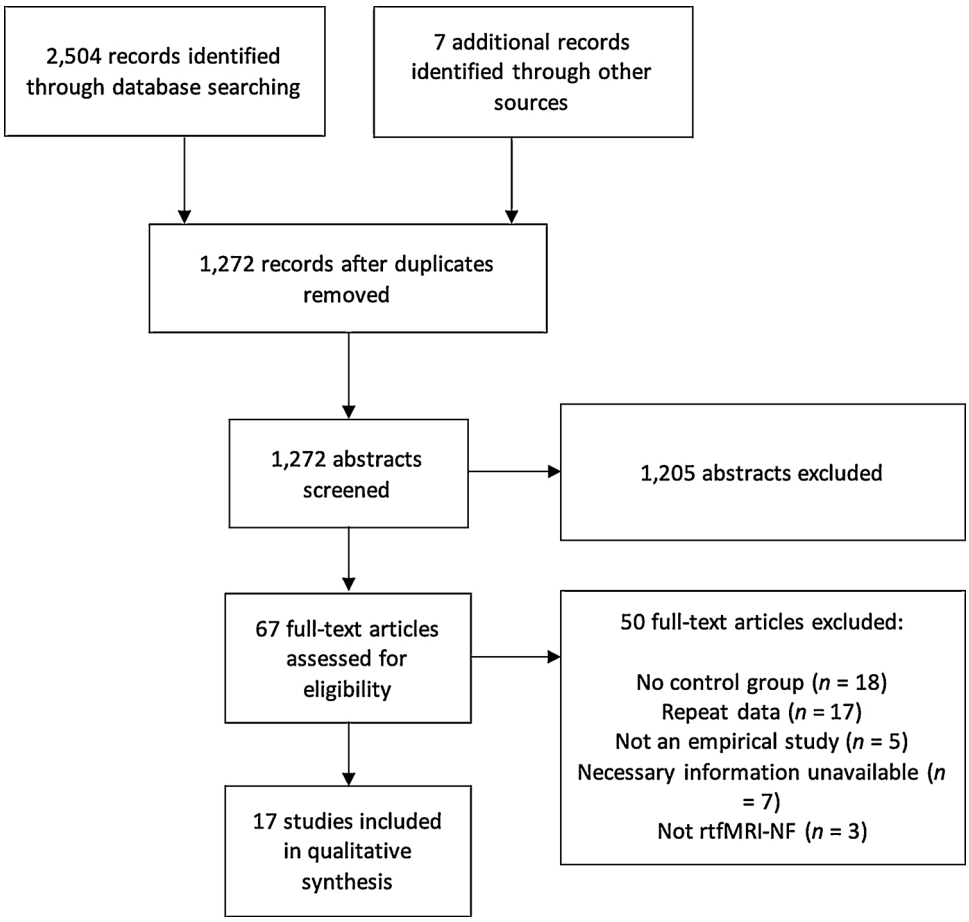

Fig. 2. Flow diagram of study selection following PRISMA criteria (Moher et al., 2009).

< .001, imputing three missing studies to the left side of the plot, which trended towards statistical significance, $p = .062$. Together, these data suggest the possibility of publication bias. As such, the true effect of rtfMRI-NF on symptoms is likely lower than $g = .37$.

On cognition, four studies assessed cognitive outcomes, contributing a total of 18 effect sizes. Compared to control conditions, rtfMRI-NF resulted in a small improvement in cognitive outcomes, $g = .23$, 95 % CI [-.33, .78], which was not statistically significant, $p = .288$ (Table 3, Fig. 5). The Q test for heterogeneity was statistically significant, $Q = 37.19$, $p = .003$, and the proportion of variance attributable to true variance may represent moderate heterogeneity, $I^2 = 54.4$ %.

3.4.2. Moderator analysis

For symptom outcomes, we were able to evaluate the effect of all moderators (Table 4). The effect of DSM diagnosis (including studies of MDD and ADHD only due to a limited number of effect sizes from other diagnoses) on symptom outcome was statistically significant, $F(1,8) = 6.06$, $p = 0.039$, $R^2 = 10.86$ %, with there being no effect of rtfMRI-NF on symptoms for ADHD, $g = .005$, 95 % CI [-.01, .02], $p = .514$, and a small effect for MDD, $g = .39$, 95 % CI [.03, .75], $p = .037$. Including a

Table 3
Meta-analytic results.

Outcome		*k*	ES	*g* [95% CI]	*p*	τ^2	*Q*	I^2
Brain								
	Neural Effect During Training Task[a]	11	15	.59 [.44, .75]	<.0001	0	8.07	0%
	Neural Effect During Transfer Task[a]	8	8	.84 [.37, 1.31]	.005	.09	10.51	31.51%
Behavior								
	Symptoms[b]	17	62	.37 [.16, .58]	.002	.09	75.16	36.48 %
	Cognition	4	18	.23 [-.33, .78]	.288	.21	37.19*	54.43 %
	RDoC Negative Valence Systems[b]	12	37	.41 [.15, .68]	.006	.09	38.60	35.51%
	RDoC Positive Valence Systems[a]	8	9	.13 [-.42, .67]	.576	.11	11.62	39.76%
	RDoC Cognitive Systems	5	20	.22 [-.28, .72]	.289	.14	32.77*	42.82 %
	RDoC Social Processes	3	8	.02 [-.44, .48]	.871	.05	10.20	30.34%
	RDoC Sensorimotor Systems	2	3	.64 [.39, .88]	.020	0	.74	0%

Note. Statistics are from models in which influential outliers were removed. *k*=number of studies, ES = number of effect sizes.
a$N = 1$ influential outlier identified and removed. Hedges' g and model statistics are reported without this influential outlier.
bData were fit with a three-level model. The sum of the variance components (across levels 2 and 3; i.e., σ^2), and the total proportion of variance attributable to heterogeneity in the true effects is provided in the τ^2 and I^2 columns, respectively.
*p<.05.
**p<.01.

category for other diagnoses (i.e., non-ADHD, non-MDD; $n = 10$) did not appreciably change the results, $F(212) = 20.56$, $p = .0001$, $R^2 = 14.3$ %, with outcomes also being superior for the other diagnosis category versus ADHD, contrast $g = .50$, 95 % CI [.31, .68], $p < .0001$, but similar to MDD, contrast $g = .12$, 95 % CI [-.25, .50], $p = .489$. Other moderators explained an additional 0% (control condition) to 29.9 % (training minutes) of variance in symptom outcomes, but none of these effects were statistically significant. Given that rtfMRI-NF did not impact cognition in a meaningful way, we did not conduct moderator analyses for cognitive outcomes.

3.4.3. RDoC analysis

As a way of providing another meaningful classification of behavioral outcomes, we conducted follow-up meta-analyses based on RDoC constructs. We were able to classify 80 behavioral outcomes as part of negative valence systems, positive valence systems, cognitive systems, social processes, sensorimotor systems, or arousal and regulatory systems. The majority of these outcomes fell within the domain of negative valence systems (46.2 %) and cognitive systems (25.0 %).

Compared to control training, a three-level meta-analysis of negative valence outcomes demonstrated that rtfMRI-NF produced a small effect, number of effect sizes = 37, $g = .41$, 95 % CI [.15, .68], $p = .006$ (Table 3, Figure S1), although the ERT indicated small-study effects, b_1

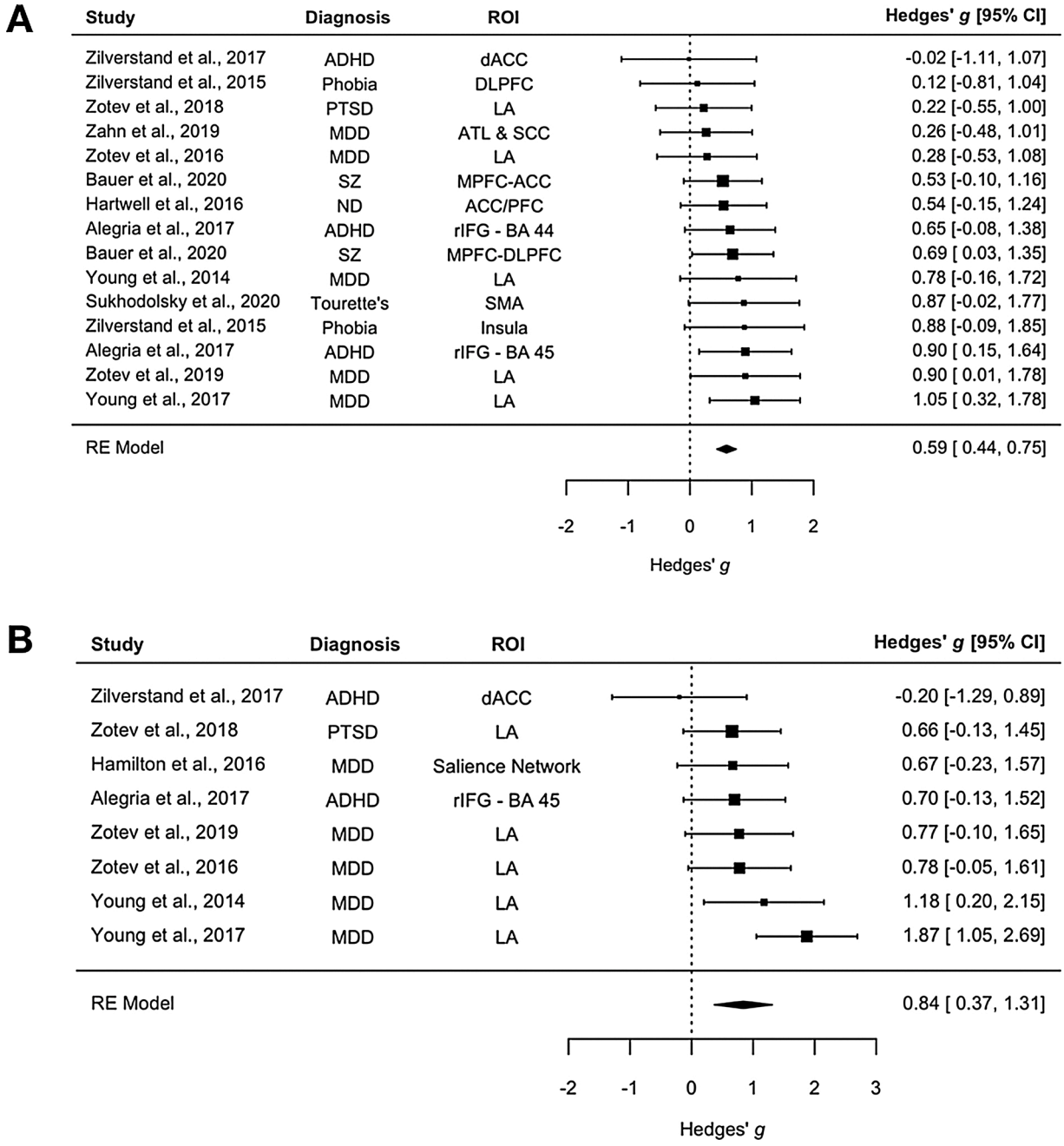

Fig. 3. Forest plots of the brain outcomes. A) Hedges' *g* effect sizes with 95 % confidence intervals comparing post-training neural activity between active and control groups. B) Hedges' *g* effect sizes with 95 % confidence intervals comparing transfer effects between active and control groups. ACC = anterior cingulate cortex; ADHD = attention deficit hyperactivity disorder; ATL = anterior temporal lobe; BA = Brodmann area; dACC = dorsal anterior cingulate cortex; DLPFC = dorsolateral prefrontal cortex; LA = left amygdala; MDD = major depressive disorder; MPFC = medial prefrontal cortex; PFC = prefrontal cortex; PTSD = posttraumatic stress disorder; rIFG = right inferior frontal gyrus; SCC = subgenual cingulate cortex; SMA = supplementary motor area; SZ = schizophrenia.

= 4.86, 95 % CI [1.14, 8.59], $p = .017$, and the TAF estimate (based on a two-level model) was smaller, $g = .29$, 95 % CI [.14, .43], $p < .001$, imputing four missing studies to the left side of the plot, $p = .031$. RtfMRI-NF also produced a medium impact on sensorimotor system outcomes, number of effect sizes=3, $g = .64$, 95 % CI [.39, .88], $p = .020$ (Table 3, Fig. S1). We did not observe a statistically significant effect for any other RDoC outcome. Of the moderators we were able to test, none impacted the effect of rtfMRI-NF on negative valence outcomes, and there were too few effect sizes to perform any moderator analyses on sensorimotor system outcomes (Table 4).

4. Discussion

In recent years, rtfMRI-NF has emerged as a promising experimental intervention for psychiatric illness. Despite several excellent qualitative reviews synthesizing these findings (Heunis et al., 2020; Paret et al., 2019; Sitaram et al., 2017; Stoeckel et al., 2014; Sulzer et al., 2013; Thibault et al., 2016, 2018; Tursic et al., 2020; Weiss et al., 2020), there has yet to be a quantitative review of these data. Here, we present data from the first quantitative analysis addressing this topic.

Our literature search uncovered 17 controlled studies evaluating the efficacy of rtfMRI-NF in improving brain and behavioral outcomes for a variety of psychiatric disorders. Despite the range of study parameters, there was some consistency in rtfMRI-NF methods. For example, rtfMRI-NF was most often compared to sham feedback, instructions for regulating the signal were explicit, the feedback signal was derived from task-based activation (e.g., percent signal change), and the feedback signal was delivered continuously. The neural effects of rtfMRI-NF were often evaluated in two contexts: (a) when participants were regulating while receiving neurofeedback ("training sessions"), and (b) when participants were regulating in the absence of neurofeedback ("transfer sessions") towards evaluating whether regulation can be sustained in a context without receiving a feedback signal, as in one's daily life.

We first addressed whether during the training task—i.e., when receiving a neurofeedback signal—rtfMRI-NF produced an advantage over control conditions in the region targeted for training. We found that it did. Specifically, rtfMRI-NF produced a medium-sized advantage over control conditions, $g = .59$, 95 % CI [.44, .75]. Said otherwise, patients across a range of psychiatric illnesses are able to use a neurofeedback signal delivered through rtfMRI to self-regulate neural activity in the targeted region. Despite the range of neurofeedback parameters (e.g., instructions, neurofeedback format and delivery, number of sessions,

Table 4
Mixed-effects moderator analyses results.

			Moderator Test			Residual Heterogeneity		Individual Estimates		
Outcome	**Moderator**	**Level**	***F***	***p***	***R²***	***Q***	***I²***	**ES**	***g* [95% CI]**	***p***
Neural Effect During Transfer Task										
	Diagnosis		2.61	.182	34.82 %	7.04	30.60%			
		ADHD						2	.34 [-.70, 1.37]	.416
		MDD						5	1.06 [.36, 1.76]	.013
	Sessions		.43	.535	0%	17.66**	60.57 %			
		Multiple						5	.55 [-.51, 1.61]	.249
		Single						4	.84 [.63, 1.06]	<.0001
	Training Minutes		4.29	.084	48.07 %	11.49	39.97%	9	*b*=-.02 [-.04, .004]	.084
Symptoms[a]										
	Diagnosis		6.06	.039	10.86 %	67.87**	46.66 %			
		ADHD						5	.005 [-.01, .02]	.514
		MDD						47	.39 [.03, .75]	.037
	Control Condition		.00	.948	0%	74.26	38.84%			
		No Feedback						17	.38 [.03, .74]	.037
		Sham Feedback						45	.37 [.10, .64]	.011
	Sessions		1.22	.289	17.08%	69.91	35.27%			
		Multiple						34	.27 [-.01, .56]	.054
		Single						28	.50 [.16, .83]	.007
	Training Minutes		2.52	.136	29.95 %	65.89	32.20%	62	*b*=-.01 [-.02, .003]	.136
	Instructions		.01	.919	.43%	75.04*	38.96 %			
		Explicit						55	.38 [.15, .61]	.004
		Implicit						7	.35 [-.30, .99]	.267
	Regulation Direction		.39	.686	.63%	73.90*	40.68 %			
		Decrease						15	.42 [.04, .81]	.032
		Increase						44	.35 [.04, .65]	.029
		Mixed						3	.49 [.32, .65]	<.0001
	Neural Signal		.69	.421	6.87%	72.15	37.93%			
		Activation						45	.40 [.14, .66]	.006
		Connectivity						17	.26 [-.01, .52]	.054
RDoC Negative Valence Systems[a]										
	Control Condition		.11	.751	0%	38.55	40.08%			
		No Feedback						8	.51 [-.23, 1.25]	.159
		Sham Feedback						29	.40 [.09, .70]	.016
	Sessions		.27	.615	11.18%	37.57	38.01%			
		Multiple						19	.36 [-.10, .81]	.108
		Single						18	.48 [.15, .82]	.010
	Training Minutes		.09	.776	19.93%	35.84	37.22%	37	*b*=-.004 [-.04, .03]	.776
	Regulation Direction		.086	.777	2.64%	37.88	43.97%			
		Decrease						6	.36 [-.21, .93]	.181
		Increase						29	.44 [.07, .82]	.027
	Neural Signal		3.18	.112	18.08%	35.98	36.74%			
		Activation						29	.48 [.13, .83]	.013
		Connectivity						8	.19 [.04, .33]	.021

[a] Data were fit with a three-level model.
* $p<.10$.
** $p < .05$.

etc.) and patient samples tested, none of the variation in effects could be attributed to heterogeneity in the true effects ($\tau^2 = 0$; $I^2 = 0\%$). However, this might reflect, in part, the relatively small samples tested, leading to imprecise study effect size estimates, and large, overlapping CIs. Thus, we caution against interpreting these findings to mean that there truly is no variation in true effects. Nevertheless, the lack of variance in the effects precluded a meaningful analysis of potential moderators.

One of the key tests in assessing whether an individual truly learns volitional control over a brain region(s) is evaluating whether the individual shows evidence of regulation in the absence of neurofeedback. If so, this might suggest that the individual can regulate the targeted region in other contexts during which neurofeedback is not available, such as one's daily life, which is precisely the context in which one would hope that an intervention has an impact. Towards evaluating this issue, we meta-analyzed the effects of rtfMRI-NF during transfer scans from 9 studies (8 after the removal of one outlier). Compared to control trainings, rtfMRI-NF demonstrated a large advantage, $g = .84$, 95 % CI [.37, 1.31]. In other words, participants demonstrated volitional control of the targeted region(s) even in the absence of a neurofeedback signal. In fact, this effect was even larger than the effect observed during training sessions when the neurofeedback signal is provided. Given that transfer sessions are typically administered last, it is possible that these large effects reflect the benefit of learning across all training sessions, including the final session from which we measured the training effect size. Another possibility is that the feedback is distracting and/or to some extent inaccurate. For example, given the hemodynamic response lag, if provided with continuous neurofeedback, one needs to keep in mind that the feedback currently received maps onto to the mental processes engaged 4–8 seconds prior. This would be challenging at baseline, and perhaps even more so for individuals experiencing cognitive difficulties due to psychiatric illness. Further, many studies do not report denoising and quality control methods leaving open the possibility that the neurofeedback signal may be corrupted (Heunis et al., 2020). In fact, recently it was shown that rtfMRI-NF training effects on network connectivity could be attributed to physiological artifacts (Weiss et al., 2020). As suggested by others, this all serves as further evidence in support of the need for additional work determining how learning occurs, how best to facilitate the generalization of learning in the context of rtfMRI-NF (Weiskopf, 2012), and methodological

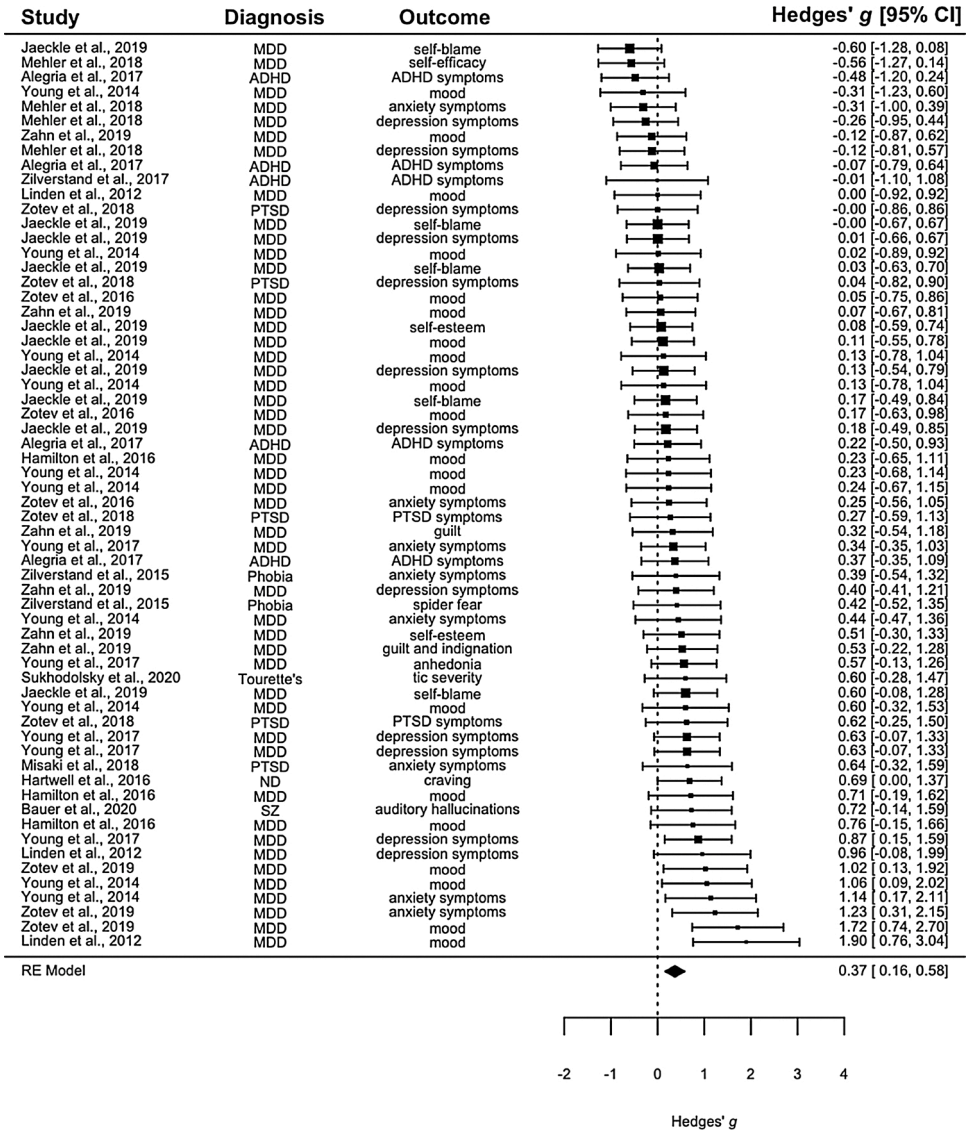

Fig. 4. Forest plot depicting Hedges' *g* effect sizes with 95 % confidence intervals comparing post-training psychiatric symptoms between active and control groups. ADHD = attention deficit hyperactivity disorder; MDD = major depressive disorder; PTSD = posttraumatic stress disorder; ND = nicotine dependence; SZ = schizophrenia.

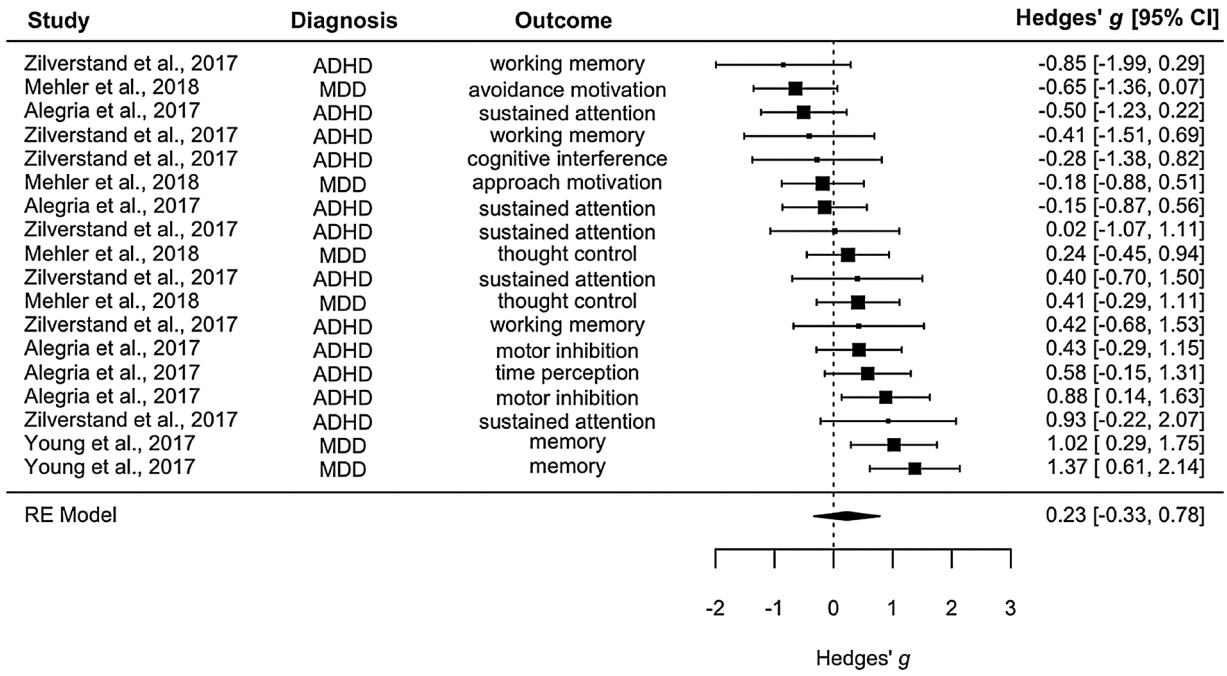

Fig. 5. Forest plot depicting Hedges' *g* effect sizes with 95 % confidence intervals comparing post-training cognition between active and control groups. ADHD = attention deficit hyperactivity disorder; MDD = major depressive disorder.

guidelines for conducting high-quality rtfMRI-NF studies that are not corrupted by noise (Fede et al., 2020; Heunis et al., 2020).

Given that the proportion of observed variance could be attributed to moderate heterogeneity in true effects (I^2 = 31.5 %), we evaluated whether some of the heterogeneity in transfer effects could be explained by study characteristics. We found the effect of rtfMRI-NF training to be higher in MDD versus ADHD, in single-session versus multiple-session designs, and the effect to be larger in studies with fewer minutes of rtfMRI-NF training. Though the pseudo R^2 values for sample diagnosis and rtfMRI-NF minutes were substantial at 34.82 % and 48.07 %, respectively, the effect of these moderators, and number of sessions, was not statistically significant. Because the number of studies included in each analysis was small, the lack of effect here does not necessarily mean that rtfMRI-NF is not impacted by these factors, but that we may have been underpowered to detect such an effect. Nonetheless, these findings intimate study characteristics that *may* be important. For example, more sessions and more training time may not be helpful; sufficient learning may occur early, and additional sessions may instead contribute to mental fatigue (Sulzer et al., 2013), which would dampen the transfer effect. In line with this idea, several prior studies have found that rtfMRI-NF can be effective after only one or a few sessions (Canterberry et al., 2013; Nicholson et al., 2018; Orlov et al., 2018; Stoeckel et al., 2014). That said, there is at least one study to suggest that while neural effects due to rtfMRI may be observable early, clinical change may require additional rtfMRI sessions (Canterberry et al., 2013). Given that rtfMRI-NF dose did not moderate any outcomes, these ideas are speculative, but would be worth evaluating in future work. In particular, it would be useful to evaluate the effect of rtfMRI-NF on the brain and behavior at the end of each training run to better understand the arc of rtfMRI-NF-induced change.

To summarize thus far, rtfMRI-NF has a moderate-sized impact on the targeted brain region(s) during training, which increases in magnitude when the neurofeedback signal is not provided. We believe this provides relatively strong evidence that volitional control over neural processes that are specifically targeted during training is possible, and that this volitional control generalizes to contexts in which no feedback is provided. Because the regions targeted for training from each study were selected based on prior research demonstrating their role in the underlying mechanisms of illness (e.g., Dunlop et al., 2017; Zahn et al., 2019), these data suggest that psychopathology-related neural disruptions may be remediable through self-regulation.

For rtfMRI-NF to have clinical utility, it should not simply restore neural function, but confer demonstrable benefits to behavior. We addressed this issue by evaluating the effect of rtfMRI-NF on symptoms and cognition. Analyzing data from all 17 studies, we found that rtfMRI-NF showed a small effect on reducing symptoms, g = .37, 95 % CI [.16, .58]. That said, sensitivity analyses suggested the possibility of publication bias. Thus, the effect of rtfMRI-NF on symptom outcomes is likely smaller than the effect size observed here. The moderator analysis showed similar trends as above whereby the effect was higher for MDD versus ADHD, in single versus multiple session protocols, and with fewer training minutes, with difference for MDD versus ADHD being statistically significant, although accounting for a small amount of variance (pseudo R^2 = 10.9 %). Also similar to the moderator effects on neural transfer outcomes, of all the moderators, training minutes accounted for the most variance (pseudo R^2 = 29.9 %). We observed the effect of rtfMRI-NF on symptoms to be equivalent for sham versus no feedback controls, explicit versus implicit instructions, up- versus down- versus mixed-regulation, and the effect to be slightly greater for activation- versus connectivity-based neurofeedback. In contrast to the impact on symptoms, the effect of rtfMRI-NF training on cognition was small, g = .23, 95 % CI [-.33, .78], and not statistically significant. Here, there were too few studies to perform a moderator analysis.

Because the distinction between symptoms and cognition is not clear cut, and the fact that the behavioral outcomes could be meaningfully classified using other schemes, we evaluated behavioral outcomes as a function of RDoC construct. We found that rtfMRI-NF was most effective at producing changes within the negative valence—a small effect, g = .41, 95 % CI [.15, .68]—and sensorimotor constructs—a medium effect, g = .64, 95 % CI [.39, .88]. We did not find moderator variables to have an impact on the negative valence effect size, and were unfortunately not able to perform a moderator analysis on the sensorimotor effect size due to our small sample size. Similar to our cognition results, we found that there was not a significant impact of training on the cognitive systems RDoC construct, and no impact of rtfMRI-NF on the other RDoC constructs.

Regarding these behavioral outcomes, it is important to note that many of them lacked diagnostic specificity (e.g., Positive and Negative Affect Schedule; Linden et al., 2012) and were administered at different time points across studies. A growing literature has suggested that the time course of clinical change may differ from the course of learned self-regulation of neural activity (Rance et al., 2018; Sukhodolsky et al., 2020) and it is possible that behavioral effects were obscured in part due to these factors. One recent study found that rather than plateauing or returning to baseline, clinical symptoms continued to improve even weeks after rtfMRI-NF in two separate samples of individuals with obsessive compulsive disorder and Tourette Syndrome (Rance et al., 2018). Thus, it is possible that our behavioral analyses underestimate the clinical benefit of rtfMRI-NF due to these effects. It would be useful for future studies and quantitative reviews to further investigate the time course of neural and clinical change. For example, a recent systematic review and meta-analysis of EEG-NF investigated the long-term clinical sustainability of training for ADHD and found lasting behavioral improvements in follow-ups of at least six to twelve months (Van Doren et al., 2019). With the steadily growing number of studies investigating rtfMRI-NF across many psychiatric illnesses and increased interest in the time course of clinical change (Rance et al., 2018; Sukhodolsky et al., 2020), we hope that this type of systematic review will soon be possible for fMRI neurofeedback as well.

Our findings should be considered in the context of several important limitations. First, the effect sizes we analyzed were derived from relatively small sample sizes. Second, we were unable to conduct moderator analyses for several outcomes, and those that we did conduct may have been underpowered. Third, though we limited our analysis to controlled studies evaluating brain outcomes in neurofeedback-targeted regions, given the wide range of rtfMRI-NF methods and applications, we analyzed a diversity of outcomes. That said, these diverse outcomes all address the broad questions we set out answer, which we believe will help identify areas for future research and assist in the planning of future studies. Fourth, given the lack of methods for addressing publication bias in multilevel and/or clustered data, we were limited in our ability to detect publication bias. Finally, though our analysis addresses, in part, *whether* rtfMRI-NF works for those with a psychiatric illness, *how* and specifically *for whom* it works remains unanswered. Addressing these issues may help to maximize the potential clinical benefits of rtfMRI-NF.

In summary, here we provide the first quantitative analysis of brain and behavioral outcomes from rtfMRI-NF studies of those with psychiatric illness. We find a medium-to-large sized effect of rtfMRI-NF for brain outcomes, and small-to-medium sized effects for behavioral outcomes. In addition to providing effect size estimates that may be used in power analyses towards conducting new rtfMRI-NF studies, our review highlights the need for more pre-registered, adequately powered, and high quality studies that follow many of the excellent guidelines suggested in other reviews (deCharms, 2007; Fede et al., 2020; Heunis et al., 2020; Paret et al., 2019; Ros et al., 2020; Sitaram et al., 2017; Stoeckel et al., 2014; Sulzer et al., 2013; Thibault et al., 2018; Tursic et al., 2020; Weiss et al., 2020). We also recommend that future studies systematically evaluate rtfMRI-NF parameters that the current analysis can only intimate as being important (e.g., training time), and address questions about the mechanism underlying rtfMRI-NF mediated change. This work will be instrumental in establishing the clinical utility of rtfMRI-NF.

Declaration of Competing Interest

None

Acknowledgments

The authors would like to thank Dr. Clemens C.C. Bauer, Dr. Michelle Hampson, Will Koller, Dr. David Linden, Dr. David Mehler, Dr. Katya Rubia, Dr. Kymberly Young, Dr. Roland Zahn, Dr. Anna Zilverstand, and Dr. Vadim Zotev for graciously providing us with data and/or information from the studies assessed for the analysis. This work was supported indirectly by the National Institute of Mental Health (grant number 1L30MH117569-01 awarded to DDF).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.neubiorev.2020.12.020.

References

Alegria, A.A., Wulff, M., Brinson, H., Barker, G.J., Norman, L.J., Brandeis, D., Stahl, D., David, A.S., Taylor, E., Giampietro, V., Rubia, K., 2017. Real-time fMRI neurofeedback in adolescents with attention deficit hyperactivity disorder. Hum. Brain Mapp. 38 (6), 3190–3209. https://doi.org/10.1002/hbm.23584.

Arora, S., Lawrence, M.A., Klein, R.M., 2020. The attention network test database: ADHD and cross-cultural applications. Front. Psychol. 11, 388. https://doi.org/10.3389/fpsyg.2020.00388.

Assink, M., Wibbelink, C.J.M., 2016. Fitting three-level meta-analytic models in R: a step-by-step tutorial. Quant. Methods Psychol. 12 (3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154.

Badcock, J.C., Hugdahl, K., 2014. A synthesis of evidence on inhibitory control and auditory hallucinations based on the Research Domain Criteria (RDoC) framework. Front. Hum. Neurosci. 8, 180. https://doi.org/10.3389/fnhum.2014.00180.

Baldessarini, R.J., Tondo, L., Viguera, A.C., 1999. Discontinuing lithium maintenance treatment in bipolar disorders: risks and implications. Bipolar Disord. 1 (1), 17–24. https://doi.org/10.1034/j.1399-5618.1999.10106.x.

Bauer, C.C.C., Okano, K., Gosh, S.S., Lee, Y.J., Melero, H., Angeles, C., de, L., Nestor, P. G., Del. Re, E.C., Northoff, G., Niznikiewicz, M.A., Whitfield-Gabrieli, S., 2020. Real-time fMRI neurofeedback reduces auditory hallucinations and modulates resting state connectivity of involved brain regions: part 2: default mode network -preliminary evidence. Psychiatry Res. 284, 112770 https://doi.org/10.1016/j.psychres.2020.112770.

Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging? Ann. N. Y. Acad. Sci. 1191, 133–155. https://doi.org/10.1111/j.1749-6632.2010.05446.x.

Borenstein, M., Higgins, J.P.T., Hedges, L.V., Rothstein, H.R., 2017. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. Res. Synth. Methods 8 (1), 5–18. https://doi.org/10.1002/jrsm.1230.

Bowden, C.L., Grunze, H., Mullen, J., Brecher, M., Paulsson, B., Jones, M., Vågerö, M., Svensson, K., 2005. A randomized, double-blind, placebo-controlled efficacy and safety study of quetiapine or lithium as monotherapy for mania in bipolar disorder. J. Clin. Psychiatry 66 (1), 111–121. https://doi.org/10.4088/jcp.v66n0116.

Boylan, K., MacQueen, G., Kirkpatrick, R., Lee, J., Santaguida, P.L., 2020. A systematic review of interventions for treatment resistant major depressive disorder in adolescents. Eur. Child Adolesc. Psychiatry 29 (4), 433–443. https://doi.org/10.1007/s00787-019-01341-5.

Brady, L.S., Potter, W.Z., Gordon, J.A., 2019. Redirecting the revolution: new developments in drug development for psychiatry. Expert Opin. Drug Discov. 14 (12), 1213–1219. https://doi.org/10.1080/17460441.2019.1666102.

Brohawn, K.H., Offringa, R., Pfaff, D.L., Hughes, K.C., Shin, L.M., 2010. The neural correlates of emotional memory in posttraumatic stress disorder. Biol. Psychiatry 68 (11), 1023–1030. https://doi.org/10.1016/j.biopsych.2010.07.018.

Brown, V.M., Morey, R.A., 2012. Neural systems for cognitive and emotional processing in posttraumatic stress disorder. Front. Psychol. 3, 449. https://doi.org/10.3389/fpsyg.2012.00449.

Buse, J., Beste, C., Herrmann, E., Roessner, V., 2016. Neural correlates of altered sensorimotor gating in boys with Tourette Syndrome: a combined EMG/fMRI study. World J. Biol. Psychiatry Off. J. World Federation Soc. Biol. Psychiatry 17 (3), 187–197. https://doi.org/10.3109/15622975.2015.1112033.

Canterberry, M., Hanlon, C.A., Hartwell, K.J., Li, X., Owens, M., LeMatty, T., Prisciandaro, J.J., Borckardt, J., Saladin, M.E., Brady, K.T., George, M.S., 2013. Sustained reduction of nicotine craving with real-time neurofeedback: exploring the role of severity of dependence. Nicotine Tob. Res. 15 (12), 2120–2124. https://doi.org/10.1093/ntr/ntt122.

Carter, E.C., Schönbrodt, F.D., Gervais, W.M., Hilgard, J., 2019. Correcting for Bias in psychology: a comparison of meta-analytic methods. Adv. Methods Pract. Psychol. Sci. https://doi.org/10.1177/2515245919847196.

Casey, B.J., Craddock, N., Cuthbert, B.N., Hyman, S.E., Lee, F.S., Ressler, K.J., 2013. DSM-5 and RDoC: progress in psychiatry research? Nature reviews. Neuroscience 14 (11), 810–814. https://doi.org/10.1038/nrn3621.

Castrellon, J.J., Meade, J., Greenwald, L., Hurst, K., Samanez-Larkin, G.R., 2020. *Dopaminergic modulation of reward discounting: a systematic review and meta-analysis* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.04.03.024364.

Cheung, M.W.-L., 2014. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. Psychol. Methods 19 (2), 211–229. https://doi.org/10.1037/a0032968.

Cheung, M.W.-L., 2019. A guide to conducting a meta-analysis with non-independent effect sizes. Neuropsychol. Rev. 29 (4), 387–396. https://doi.org/10.1007/s11065-019-09415-6.

Clark, D.B., Tanner-Smith, E.E., Killingsworth, S.S., 2016. Digital games, design, and learning: a systematic review and meta-analysis. Rev. Educ. Res. 86 (1), 79–122. https://doi.org/10.3102/0034654315582065.
Clemente, C.D., Sterman, M.B., Wyrwicka, W., 1964. Post-reinforcement EEG synchronization during alimentary behavior. Electroencephalogr. Clin. Neurophysiol. 16 (4), 355–365. https://doi.org/10.1016/0013-4694(64)90069-0.
Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Routledge Academic., NY.
Coldham, E.L., Addington, J., Addington, D., 2002. Medication adherence of individuals with a first episode of psychosis. Acta Psychiatr. Scand. 106 (4), 286–290. https://doi.org/10.1034/j.1600-0447.2002.02437.x.
Cox, R.W., Jesmanowicz, A., Hyde, J.S., 1995. Real-time functional magnetic resonance imaging. Magn. Reson. Med. 33 (2), 230–236. https://doi.org/10.1002/mrm.1910330213.
deCharms, R.C., 2007. Reading and controlling human brain activation using real-time functional magnetic resonance imaging. Trends Cogn. Sci. (Regul. Ed.) 11 (11), 473–481. https://doi.org/10.1016/j.tics.2007.08.014.
deCharms, R.C., 2008. Applications of real-time fMRI. Nat. Rev. Neurosci. 9 (9), 720–729. https://doi.org/10.1038/nrn2414.
Deeks, J.J., Higgins, J.P., Altman, D.G., 2008. Analysing data and undertaking meta-analyses. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, Ltd., pp. 243–296. https://doi.org/10.1002/9780470712184.ch9
Drevon, D., Fursa, S.R., Malcolm, A.L., 2017. Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. Behav. Modif. 41 (2), 323–339. https://doi.org/10.1177/0145445516673998.
Dunlop, B.W., Rajendra, J.K., Craighead, W.E., Kelley, M.E., McGrath, C.L., Choi, K.S., Kinkead, B., Nemeroff, C.B., Mayberg, H.S., 2017. Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder. Am. J. Psychiatry 174 (6), 533–545. https://doi.org/10.1176/appi.ajp.2016.16050518.
Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56 (2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x.
Duval, E.R., Javanbakht, A., Liberzon, I., 2015. Neural circuits in anxiety and stress disorders: a focused review. Ther. Clin. Risk Manag. 11, 115–126. https://doi.org/10.2147/TCRM.S48528.
Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315 (7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.
Fede, S.J., Dean, S.F., Manuweera, T., Momenan, R., 2020. A guide to literature informed decisions in the design of real time fMRI neurofeedback studies: a systematic review. Front. Hum. Neurosci. 14, 60. https://doi.org/10.3389/fnhum.2020.00060.
Fernandez, E., Salem, D., Swift, J.K., Ramtahal, N., 2015. Meta-analysis of dropout from cognitive behavioral therapy: magnitude, timing, and moderators. J. Consult. Clin. Psychol. 83 (6), 1108–1122. https://doi.org/10.1037/ccp0000044.
Fernández Castilla, B., 2019. Models and Techniques for the Meta-Analysis of Dependent Effect Sizes. https://lirias.kuleuven.be/retrieve/546660.
Fetz, E.E., 1969. Operant conditioning of cortical unit activity. Science 163 (3870), 955–958. https://doi.org/10.1126/science.163.3870.955.
Ford, J.M., 2016. Studying auditory verbal hallucinations using the RDoC framework. Psychophysiology 53 (3), 298–304. https://doi.org/10.1111/psyp.12457.
Gersh, E., Hallford, D.J., Rice, S.M., Kazantzis, N., Gersh, H., Gersh, B., McCarty, C.A., 2017. Systematic review and meta-analysis of dropout rates in individual psychotherapy for generalized anxiety disorder. J. Anxiety Disord. 52, 25–33. https://doi.org/10.1016/j.janxdis.2017.10.001.
Glenn, C.R., Cha, C.B., Kleiman, E.M., Nock, M.K., 2017. Understanding suicide risk within the research domain criteria (RDoC) framework: insights, challenges, and future research considerations. Clin. Psychol. Sci. J. Assoc. Psychol. Sci. 5 (3), 568–592. https://doi.org/10.1177/2167702616686854.
Graham, J., Banaschewski, T., Buitelaar, J., Coghill, D., Danckaerts, M., Dittmann, R.W., Döpfner, M., Hamilton, R., Hollis, C., Holtmann, M., Hulpke-Wette, M., Lecendreux, M., Rosenthal, E., Rothenberger, A., Santosh, P., Sergeant, J., Simonoff, E., Sonuga-Barke, E., Wong, I.C.K., European Guidelines Group, 2011. European guidelines on managing adverse effects of medication for ADHD. Eur. Child Adolesc. Psychiatry 20 (1), 17–37. https://doi.org/10.1007/s00787-010-0140-6.
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. https://doi.org/10.1016/j.pscychresns.2016.01.016.
Hardt, J.V., Kamiya, J., 1978. Anxiety change through electroencephalographic alpha feedback seen only in high anxiety subjects. Science 201 (4350), 79–81. https://doi.org/10.1126/science.663641.
Harrow, M., Jobe, T.H., Faull, R.N., 2012. Do all schizophrenia patients need antipsychotic treatment continuously throughout their lifetime? A 20-year longitudinal study. Psychol. Med. 42 (10), 2145–2155. https://doi.org/10.1017/S0033291712000220.
Hartwell, K.J., Hanlon, C.A., Li, X., Borckardt, J.J., Canterberry, M., Prisciandaro, J.J., Moran-Santa Maria, M.M.M., LeMatty, T., George, M.S., Brady, K.T., 2016. Individualized real-time fMRI neurofeedback to attenuate craving in nicotine-dependent smokers. J. Psyc. Neurosci. JPN 41 (1), 48–55. https://doi.org/10.1503/jpn.140200.
Hedges, L.V., Tipton, E., Johnson, M.C., 2010. Robust variance estimation in meta-regression with dependent effect size estimates. Res. Synth. Methods 1 (1), 39–65. https://doi.org/10.1002/jrsm.5.
Heunis, S., Lamerichs, R., Zinger, S., Caballero-Gaudes, C., Jansen, J.F.A., Aldenkamp, B., Breeuwer, M., 2020. Quality and denoising in real-time functional magnetic resonance imaging neurofeedback: a methods review. Hum. Brain Mapp. https://doi.org/10.1002/hbm.25010.
Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., Wang, P., 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry 167 (7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379.
Jaeckle, T., Williams, S.C.R., Barker, G.J., Basilio, R., Carr, E., Goldsmith, K., Colasanti, A., Giampietro, V., Cleare, A., Young, A.H., Moll, J., Zahn, R., 2019. *Self-blaming emotions in major depression: a randomised pilot trial comparing fMRI neurofeedback training with self-guided psychological strategies (NeuroMooD)* [Preprint]. Psychiatry and Clinical Psychology. https://doi.org/10.1101/19004309.
John, L.K., Loewenstein, G., Prelec, D., 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol. Sci. 23 (5), 524–532. https://doi.org/10.1177/0956797611430953.
Kautzner, A., Vlahiotis, A., Devine, S.T., Eichholz, J., 2011. Discontinuation rates and health care costs in adult patients starting generic versus brand SSRI or SNRI antidepressants in commercial health plans. J. Manag. Care Pharm. 17 (2), 123–132. https://doi.org/10.18553/jmcp.2011.17.2.123.
Kip, A., Priebe, S., Holling, H., Morina, N., 2020. Psychological interventions for posttraumatic stress disorder and depression in refugees: a meta-analysis of randomized controlled trials. Clin. Psychol. Psychother. https://doi.org/10.1002/cpp.2446.
Konstantopoulos, S., 2011. Fixed effects and variance components estimation in three-level meta-analysis. Res. Synth. Methods 2 (1), 61–76. https://doi.org/10.1002/jrsm.35.
Koush, Y., Ashburner, J., Prilepin, E., Sladky, R., Zeidman, P., Bibikov, S., Scharnowski, F., Nikonorov, A., De Ville, D.V., 2017. OpenNFT: an open-source Python/Matlab framework for real-time fMRI neurofeedback training based on activity, connectivity and multivariate pattern analysis. NeuroImage 156, 489–503. https://doi.org/10.1016/j.neuroimage.2017.06.039.
Li, X., Branch, C.A., DeLisi, L.E., 2009. Language pathway abnormalities in schizophrenia: a review of fMRI and other imaging studies. Curr. Opin. Psychiatry 22 (2), 131–139. https://doi.org/10.1097/YCO.0b013e328324bc43.
Lieberman, J.A., Stroup, T.S., McEvoy, J.P., Swartz, M.S., Rosenheck, R.A., Perkins, D.O., Keefe, R.S.E., Davis, S.M., Davis, C.E., Lebowitz, B.D., Severe, J., Hsiao, J.K., Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators, 2005. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N. Engl. J. Med. 353 (12), 1209–1223. https://doi.org/10.1056/NEJMoa051688.
Lim, Y.H., Partridge, K., Girdler, S., Morris, S.L., 2017. Standing postural control in individuals with autism Spectrum disorder: systematic review and meta-analysis. J. Autism Dev. Disord. 47 (7), 2238–2253. https://doi.org/10.1007/s10803-017-3144-y.
Linden, D.E.J., 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 (6), e38115. https://doi.org/10.1371/journal.pone.0038115.
López-López, J.A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W., Viechtbauer, W., 2014. Estimation of the predictive power of the model in mixed-effects meta-regression: a simulation study. Br. J. Math. Stat. Psychol. 67 (1), 30–48. https://doi.org/10.1111/bmsp.12002.
Lubar, J.F., Shouse, M.N., 1976. EEG and behavioral changes in a hyperkinetic child concurrent with training of the sensorimotor rhythm (SMR): a preliminary report. Biofeedback Self. 1 (3), 293–306. https://doi.org/10.1007/BF01001170.
McNaught, K.S.P., Mink, J.W., 2011. Advances in understanding and treatment of Tourette syndrome. Nat. Rev. Neurol. 7 (12), 667–676. https://doi.org/10.1038/nrneurol.2011.167.
Mehler, D.M.A., Sokunbi, M.O., Habes, I., Barawi, K., Subramanian, L., Range, M., Evans, J., Hood, K., Lührs, M., Keedwell, P., Goebel, R., Linden, D.E.J., 2018. Targeting the affective brain-a randomized controlled trial of real-time fMRI neurofeedback in patients with depression. Neuropsychopharmacology: Off. Publ. Am. Coll. Neuropsychopharmacology 43 (13), 2578–2585. https://doi.org/10.1038/s41386-018-0126-5.
Millan, M.J., Andrieux, A., Bartzokis, G., Cadenhead, K., Dazzan, P., Fusar-Poli, P., Gallinat, J., Giedd, J., Grayson, D.R., Heinrichs, M., Kahn, R., Krebs, M.-O., Leboyer, M., Lewis, D., Marin, O., Marin, P., Meyer-Lindenberg, A., McGorry, P., McGuire, P., Weinberger, D., 2016. Altering the course of schizophrenia: progress and perspectives. Nat. Rev. Drug Discov. 15 (7), 485–515. https://doi.org/10.1038/nrd.2016.28.
Misaki, M., Phillips, R., Zotev, V., Wong, C.-K., Wurfel, B.E., Krueger, F., Feldner, M., Bodurka, J., 2018. Real-time fMRI amygdala neurofeedback positive emotional training normalized resting-state functional connectivity in combat veterans with and without PTSD: a connectome-wide investigation. Neuroimage Clin. 20, 543–555. https://doi.org/10.1016/j.nicl.2018.08.025.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339. https://doi.org/10.1136/bmj.b2535.
Morris, S.B., DeShon, R.P., 2002. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol. Methods 7 (1), 105–125. https://doi.org/10.1037/1082-989x.7.1.105.
Nicholson, A.A., Rabellino, D., Densmore, M., Frewen, P.A., Paret, C., Kluetsch, R., Schmahl, C., Théberge, J., Ros, T., Neufeld, R.W.J., McKinnon, M.C., Reiss, J.P., Jetly, R., Lanius, R.A., 2018. Intrinsic connectivity network dynamics in PTSD during amygdala downregulation using real-time fMRI neurofeedback: a preliminary

analysis. Hum. Brain Mapp. 39 (11), 4258–4275. https://doi.org/10.1002/hbm.24244.
Normandeau, C.P., Naumova, D., Thompson, S.L., Ebrahimzadeh, M., Liu, Y.Q., Reynolds, L., Ren, H.-Y., Hawken, E.R., Dumont, É.C., 2017. Advances in understanding and treating mental illness: proceedings of the 40th Canadian College of Neuropsychopharmacology Annual Meeting Symposia. J. Psychiatry Neurosci.: JPN 42 (5), 353–358 https://doi.org/10.1503.jpn/170120.
Oblak, E.F., Lewis-Peacock, J.A., Sulzer, J.S., 2017. Self-regulation strategy, feedback timing and hemodynamic properties modulate learning in a simulated fMRI neurofeedback environment. PLoS Comput. Biol. 13 (7) https://doi.org/10.1371/journal.pcbi.1005681 e1005681.
Orlov, N.D., Giampietro, V., O'Daly, O., Lam, S.-L., Barker, G.J., Rubia, K., McGuire, P., Shergill, S.S., Allen, P., 2018. Real-time fMRI neurofeedback to down-regulate superior temporal gyrus activity in patients with schizophrenia and auditory hallucinations: a proof-of-concept study. Transl. Psychiatry 8 (1), 46. https://doi.org/10.1038/s41398-017-0067-5.
Paret, C., Goldway, N., Zich, C., Keynan, J.N., Hendler, T., Linden, D., Cohen Kadosh, K., 2019. Current progress in real-time functional magnetic resonance-based neurofeedback: methodological challenges and achievements. NeuroImage 202, 116107. https://doi.org/10.1016/j.neuroimage.2019.116107.
Patterson, B., Van Ameringen, M., 2017. Augmentation strategies for treatment-resistant anxiety disorders: a systematic review and meta-analysis. Focus Am. Psychiatr. Publ. (Am Psychiatr Publ) 15 (2), 219–226. https://doi.org/10.1176/appi.focus.15205.
Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., Rushton, L., 2008. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J. Clin. Epidemiol. 61 (10), 991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010.
Polese, D., Fornaro, M., Palermo, M., De Luca, V., de Bartolomeis, A., 2019. Treatment-resistant to antipsychotics: a resistance to everything? Psychotherapy in treatment-resistant schizophrenia and nonaffective psychosis: a 25-Year systematic review and exploratory meta-analysis. Front. Psychiatry 10, 210. https://doi.org/10.3389/fpsyt.2019.00210.
Quagliato, L.A., Cosci, F., Shader, R.I., Silberman, E.K., Starcevic, V., Balon, R., Dubovsky, S.L., Salzman, C., Krystal, J.H., Weintraub, S.J., Freire, R.C., Nardi, A.E., International Task Force on Benzodiazepines, 2019. Selective serotonin reuptake inhibitors and benzodiazepines in panic disorder: a meta-analysis of common side effects in acute treatment. J. Psychopharmacol. (Oxford) 33 (11), 1340–1351. https://doi.org/10.1177/0269881119859372.
R Core Team, 2018. R: A Language and Environment for Statistical Computing (3.5.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org.
Rance, M., Walsh, C., Sukhodolsky, D.G., Pittman, B., Qiu, M., Kichuk, S.A., Wasylink, S., Koller, W.N., Bloch, M., Gruner, P., Scheinost, D., Pittenger, C., Hampson, M., 2018. Time course of clinical change following neurofeedback. NeuroImage 181, 807–813. https://doi.org/10.1016/j.neuroimage.2018.05.001.
Raudenbush, S.W., 2009. Analyzing effect sizes: random-effects models. The Handbook of Research Synthesis and Meta-analysis, 2nd ed. Russell Sage Foundation, pp. 295–315.
Rodgers, M.A., Pustejovsky, J.E., 2019. Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes [Preprint]. MetaArXiv. https://doi.org/10.31222/osf.io/vqp8u.
Rohatgi, A., 2019. WebPlotDigitizer (4.2) [Computer software]. https://automeris.io/WebPlotDigitizer.
Ros, T., Enriquez-Geppert, S., Zotev, V., Young, K.D., Wood, G., Whitfield-Gabrieli, S., Wan, F., Vuilleumier, P., Vialatte, F., Van De Ville, D., Todder, D., Surmeli, T., Sulzer, J.S., Strehl, U., Sterman, M.B., Steiner, N.J., Sorger, B., Soekadar, S.R., Sitaram, R., Thibault, R.T., 2020. Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). Brain: A Journal of Neurology 143 (6), 1674–1685. https://doi.org/10.1093/brain/awaa009.
Rozental, A., Castonguay, L., Dimidjian, S., Lambert, M., Shafran, R., Andersson, G., Carlbring, P., 2018. Negative effects in psychotherapy: commentary and recommendations for future research and clinical practice. BJPsych Open 4 (4), 307–312. https://doi.org/10.1192/bjo.2018.42.
Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22 (11), 1359–1366. https://doi.org/10.1177/0956797611417632.
Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M.L., Rana, M., Oblak, E., Birbaumer, N., Sulzer, J., 2017. Closed-loop brain training: the science of neurofeedback. Nat. Rev. Neurosci. 18 (2), 86–100. https://doi.org/10.1038/nrn.2016.164.
Sonuga-Barke, E.J.S., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., Stevenson, J., Danckaerts, M., van der Oord, S., Döpfner, M., Dittmann, R.W., Simonoff, E., Zuddas, A., Banaschewski, T., Buitelaar, J., Coghill, D., Hollis, C., Konofal, E., Lecendreux, M., European ADHD Guidelines Group, 2013. Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. Am. J. Psychiatry 170 (3), 275–289. https://doi.org/10.1176/appi.ajp.2012.12070991.
Sterman, M.B., Wyrwicka, W., 1967. EEG correlates of sleep: evidence for separate forebrain substrates. Brain Res. 6 (1), 143–163. https://doi.org/10.1016/0006-8993(67)90186-2.
Sterne, J.A.C., Egger, M., 2006. Regression methods to detect publication and other bias in meta-analysis. Publication Bias in Meta-Analysis. John Wiley & Sons, Ltd., pp. 99–110. https://doi.org/10.1002/0470870168.ch6
Sterne, J.A., Egger, M., Smith, G.D., 2001. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. BMJ 323 (7304), 101–105. https://doi.org/10.1136/bmj.323.7304.101.
Stoeckel, L.E., Garrison, K.A., Ghosh, S., Wighton, P., Hanlon, C.A., Gilman, J.M., Greer, S., Turk-Browne, N.B., deBettencourt, M.T., Scheinost, D., Craddock, C., Thompson, T., Calderon, V., Bauer, C.C., George, M., Breiter, H.C., Whitfield-Gabrieli, S., Gabrieli, J.D., LaConte, S.M., Evins, A.E., 2014. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. Neuroimage Clin. 5, 245–255. https://doi.org/10.1016/j.nicl.2014.07.002.
Sukhodolsky, D.G., Walsh, C., Koller, W.N., Eilbott, J., Rance, M., Fulbright, R.K., Zhao, Z., Bloch, M.H., King, R., Leckman, J.F., Scheinost, D., Pittman, B., Hampson, M., 2020. Randomized, sham-controlled trial of real-time functional magnetic resonance imaging neurofeedback for tics in adolescents with tourette syndrome. Biol. Psychiatry 87 (12), 1063–1070. https://doi.org/10.1016/j.biopsych.2019.07.035.
Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M.L., Bruehl, A.B., Cohen, L.G., DeCharms, R.C., Gassert, R., Goebel, R., Herwig, U., LaConte, S., Linden, D., Luft, A., Seifritz, E., Sitaram, R., 2013. Real-time fMRI neurofeedback: progress and challenges. NeuroImage 76, 386–399. https://doi.org/10.1016/j.neuroimage.2013.03.033.
Terrin, N., Schmid, C.H., Lau, J., 2005. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. J. Clin. Epidemiol. 58 (9), 894–901. https://doi.org/10.1016/j.jclinepi.2005.01.006.
Thibault, R.T., Lifshitz, M., Raz, A., 2016. The self-regulating brain and neurofeedback: experimental science and clinical promise. Cortex 74, 247–261. https://doi.org/10.1016/j.cortex.2015.10.024.
Thibault, R.T., MacPherson, A., Lifshitz, M., Roth, R.R., Raz, A., 2018. Neurofeedback with fMRI: a critical systematic review. NeuroImage 172, 786–807. https://doi.org/10.1016/j.neuroimage.2017.12.071.
Tursic, A., Eck, J., Lührs, M., E. J, Linden, D., Goebel, R., 2020. A systematic review of fMRI neurofeedback reporting and effects in clinical populations. Neuroimage Clin. 102496. https://doi.org/10.1016/j.nicl.2020.102496.
Van den Noortgate, W., López-López, J.A., Marín-Martínez, F., Sánchez-Meca, J., 2013. Three-level meta-analysis of dependent effect sizes. Behav. Res. Methods 45 (2), 576–594. https://doi.org/10.3758/s13428-012-0261-6.
Van den Noortgate, W., López-López, J.A., Marín-Martínez, F., Sánchez-Meca, J., 2015. Meta-analysis of multiple outcomes: a multilevel approach. Behav. Res. Methods 47 (4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2.
Van Doren, J., Arns, M., Heinrich, H., Vollebregt, M.A., Strehl, U., K Loo, S., 2019. Sustained effects of neurofeedback in ADHD: a systematic review and meta-analysis. Eur. Child Adolesc. Psychiatry 28 (3), 293–305. https://doi.org/10.1007/s00787-018-1121-4.
Viechtbauer, W., 2010. Conducting meta-analyses in r with the metafor package. J. Stat. Softw. 36 (3) https://doi.org/10.18637/jss.v036.i03.
Viechtbauer, W., Cheung, M.W.-L., 2010. Outlier and influence diagnostics for meta-analysis. Res. Synth. Methods 1 (2), 112–125. https://doi.org/10.1002/jrsm.11.
Wallace, B.C., Small, K., Brodley, C.E., Lau, J., Trikalinos, T.A., 2012. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. Proceedings of the 2nd Acm Sighit Symposium on International Health Informatics - Ihi 12. https://doi.org/10.1145/2110363.2110464, 819.
Waltman, S., Hall, B.C., McFarr, L.M., Beck, A.T., Creed, T.A., 2017. In-session stuck points and pitfalls of community clinicians learning CBT: qualitative investigation. Cogn. Behav. Pract. 24 (2), 256–267. https://doi.org/10.1016/j.cbpra.2016.04.002.
Wang, S.-M., Han, C., Bahk, W.-M., Lee, S.-J., Patkar, A.A., Masand, P.S., Pae, C.-U., 2018. Addressing the side effects of contemporary antidepressant drugs: a comprehensive review. Chonnam Med. J. 54 (2), 101–112. https://doi.org/10.4068/cmj.2018.54.2.101.
Weiskopf, N., 2012. Real-time fMRI and its application to neurofeedback. NeuroImage 62 (2), 682–692. https://doi.org/10.1016/j.neuroimage.2011.10.009.
Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., Birbaumer, N., Mathiak, K., 2004. Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). J. Physiol. Paris 98 (4–6), 357–373. https://doi.org/10.1016/j.jphysparis.2005.09.019.
Weiss, F., Zamoscik, V., Schmidt, S.N.L., 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. https://doi.org/10.1016/j.neuroimage.2020.116580.
Weisz, J.R., Kuppens, S., Ng, M.Y., Eckshtain, D., Ugueto, A.M., Vaughn-Coaxum, R., Jensen-Doss, A., Hawley, K.M., Krumholz Marchette, L.S., Chu, B.C., Weersing, V.R., Fordwood, S.R., 2017. What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. Am. Psychol. 72 (2), 79–117. https://doi.org/10.1037/a0040360.
Wildes, J.E., Marcus, M.D., 2015. Application of the Research Domain Criteria (RDoC) framework to eating disorders: emerging concepts and research. Curr. Psychiatry Rep. 17 (5) https://doi.org/10.1007/s11920-015-0572-2, 30.
Wunderink, L., Nieboer, R.M., Wiersma, D., Sytema, S., Nienhuis, F.J., 2013. Recovery in remitted first-episode psychosis at 7 years of follow-up of an early dose reduction/discontinuation or maintenance treatment strategy: long-term follow-up of a 2-year randomized clinical trial. JAMA Psychiatry 70 (9), 913–920. https://doi.org/10.1001/jamapsychiatry.2013.19.
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 (2), e88785. https://doi.org/10.1371/journal.pone.0088785.
Young, K.D., Siegle, G.J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C., Bodurka, J., 2017. Randomized clinical trial of real-time fMRI amygdala

neurofeedback for major depressive disorder: effects on symptoms and autobiographical memory recall. Am. J. Psychiatry 174 (8), 748–755. https://doi.org/10.1176/appi.ajp.2017.16060637.

Zahn, R., Weingartner, J.H., Basilio, R., Bado, P., Mattos, P., Sato, J.R., de Oliveira-Souza, R., Fontenelle, L.F., Young, A.H., Moll, J., 2019. Blame-rebalance fMRI neurofeedback in major depressive disorder: a randomised proof-of-concept trial. Neuroimage Clin. 24, 101992 https://doi.org/10.1016/j.nicl.2019.101992.

Zilverstand, A., Sorger, B., Sarkheil, P., Goebel, R., 2015. FMRI neurofeedback facilitates anxiety regulation in females with spider phobia. Front. Behav. Neurosci. 9, 148. https://doi.org/10.3389/fnbeh.2015.00148.

Zilverstand, A., Sorger, B., Slaats-Willemse, D., Kan, C.C., Goebel, R., Buitelaar, J.K., 2017. fMRI neurofeedback training for increasing anterior cingulate cortex activation in adult attention deficit hyperactivity disorder. An exploratory randomized, single-blinded study. PLoS One 12 (1). https://doi.org/10.1371/journal.pone.0170795 e0170795.

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. https://doi.org/10.1016/j.nicl.2016.02.003.

Zotev, V., Phillips, R., Misaki, M., Wong, C.K., Wurfel, B.E., Krueger, F., Feldner, M., Bodurka, J., 2018. Real-time fMRI neurofeedback training of the amygdala activity with simultaneous EEG in veterans with combat-related PTSD. Neuroimage Clin. 19, 106–121. https://doi.org/10.1016/j.nicl.2018.04.010.

Zotev, V., Mayeli, A., Misaki, M., Bodurka, J., 2020. Emotion self-regulation training in major depressive disorder using simultaneous real-time fMRI and EEG neurofeedback. Neuroimage Clin. 27, 102331 https://doi.org/10.1016/j.nicl.2020.102331.