Data supplement for Lam et al., Double-Blind, Randomized, Sham-Controlled Trial Testing the Efficacy of fMRI Neurofeedback on Clinical and Cognitive Measures in Children With ADHD. Am J Psychiatry (doi: 10.1176/appi.ajp.21100999)

CONTENTS

Supplementary Methods

Sample size calculation/ power analysis

The proof-of-concept study found a significant reduction in ADHD-RS total score in the rIFC fMRI-NF group from baseline to post-treatment(1) with an effect size of Cohen's *d* = 0.6. Using G*Power-3 software(2) and by entering 80% power, α = 0.05, numerator df = 1 for a 2x2 (Group × Time) analysis of covariance (ANCOVA) with three covariates (i.e., baseline score of outcome measure, medication status, and age), we estimated a minimum sample size of 45 per group. To account for an approximate drop-out rate of up to 10% based on the proof-of concept study and previous fMRI pharmacological trials in ADHD conducted in our lab(1, 3, 4), we estimated a target sample size of 50 participants per group.

Cognitive Measures

The MARS go/no-go task, adult version(5,6), measures motor response inhibition. The task requires a rapid motor response to frequent go stimuli (i.e., a spaceship) and withholding a response to infrequent no-go stimuli (i.e., an explosion; go: trials = 220; no-go trials = 80; 26.7% no-go trials). Each stimulus is displayed for 300ms, with a 1000-ms interstimulus interval (ISI). The participants performed the task once with their left index finger, and once with their right index finger. The primary outcome measure is the probability of inhibition to no-go stimuli, where higher scores indicate better motor inhibition. Secondary outcome measures are mean response time (MRT), intrasubject response time variability (RTV) to go trials, and premature responses to all trials.

The MARS Simon task, adult version(5,6), measures interference inhibition and selective attention. In this task, arrows pointing in a left or right direction appear on the same side of the monitor as the arrow direction (i.e., "congruent") or on the opposite side of their direction (i.e., "incongruent") for 400ms with a 1400ms ISI (congruent: incongruent trials = 160:60; 27.3% incongruent trials). Participants respond by pressing the arrow key that corresponds to the arrow direction. The incongruent iconic and spatial information produces cognitive interference that is typically associated with a slower response compared to the response to congruent trials, as indexed by the primary outcome measure, the Simon RT interference effect, i.e., MRT for incongruent minus MRT for congruent trials(7).

The MARS continuous performance task (CPT), adult version(5,6), measures sustained and selective attention. In this task, a string of letters is presented in a pseudo-random order with a trial time of 1 second. The letters are displayed for 300ms, followed by a blank screen for 700ms. Participants are instructed to ignore all letters except for the target letter sequence. Target letters are either an "A" followed by an "X" or an "A" followed by an "O". Participants are asked to respond to the target letters "A-X" as soon as they see this combination by pressing the left arrow key with their left index finger, and to respond to the target letters "A-O" by pressing the right arrow key with their right index finger. There is a total of 480 trials, with 60 target letter trials (12.5%) (30 trials of "A-X" trials and 30 "A-O" trials). For every three correct responses to the targets "A-X" or "A-O", a reward bar with 10 units is filled which is shown on the right side of the blue box in either red or blue. Primary measures are the percentage of omission errors, i.e., nonresponse to targets; and the percentage of commission errors, i.e., incorrect responses on nontargets. Secondary measures are MRT, intrasubject RTV to targets, and premature errors to all trials.

The computerised Mackworth clock vigilance test (MCT)(8) was used to investigate sustained attention and vigilance on the detection of signals that are difficult to detect. In this task, a clock-hand moves in short jumps like the second hand of an analogue clock, every second. At infrequent and irregular intervals (10% of trials, pseudorandomized order), the hand makes a wider jump of two seconds, which are the target trials. There are 400 trials in total: 360 regular one-second jumps, and 40 wider irregular jumps. Participantsare instructed to try their best to detect and only respond to the irregular jumps by pressing the space bar key using their right index finger. Participants receive immediate feedback of their performance. For correct trial detection, a green light flashes up. Whenever an error is made (missed jumps: omissions, and incorrect responses when there were no irregular jumps: commission) a red light flashes up. The dependent measures are the number of omission errors, and of commission errors.

The Wisconsin's card sort task, computerised version(9), requires participants to sort 64 cards according to shape, colour, or number appearing on each card. These sorting rules are not told to the participants. They instead receive a feedback "wrong" or "right" after each sorting, which allows them to deduce the sorting rule. A sorting rule stays the same for ten consecutive correct sorting, after which it changes following a predetermined order. The participants' perseveration for applying the old sorting strategy once a new rule sets in is measured as perseverative errors, which indexes cognitive flexibility, and is a key measure of this task. Non-perseverative errors are also recorded as a measure of general executive function difficulties.

The C8 sciences version of the NIH list sorting working memory task(10) measures visuospatial working memory. In this task, participants are presented with a series of stimuli (targets) which are pictures of animals or household objects, each presented visually. A 3X4 grid matrix composing of pictures of animals or objects (stimuli that appeared in the previous trial are mixed among non-target stimuli that were not shown in the previous trial) then appear on the screen. From the matrix, the participants are instructed to select the stimuli that were presented previously, and to reorder them from smallest to largest. There are two blocks of the task; the first block only presents a series of one type of item stimulus (animals or objects), while the second block show a series of stimuli that mixed both animals and objects. The second part of the task requires increased load of working memory. Each block begins with a sequence of two pictures, which increases by one picture with every correct trial. If a participant made an incorrect response twice in a row during the same sequence number, the block is then terminated. Alternatively, the block ends after the maximum sequence number of seven is reached. The primary measure is the total number of correct responses across two blocks.

Composite EF measures. Three composite measures of response prematurity, processing speed, and intrasubject response variability, were derived respectively from the variables: premature errors, MRT and intrasubject RTV, which were significantly correlated across the three similarly structured tasks (correlation ranged from r=.25; p=.018 to r=.62; p=.001) (i.e., MARS go/no-go task, Simon task, and CPT). These composite measures were constructed to reduce the data dimension and number of comparisons. They were computed as follows:

(a) *Composite response prematurity* is the average percentage of premature responses across all trials for the three tasks, where premature responses occur between 200ms pre- and 100ms post-stimulus onsets, which are considered too late a response for the previous stimulus and too early for the current response(5). The measure was derived by summing the premature responses during each task, in proportion to the number of trials per task.

$$\text{Composite response prematurity} = \frac{\sum_i n_i \text{Premature responses}_i}{\sum_i n_i}$$

(b) *Composite processing speed* is the sum of MRT to *Composite processing speed* is the sum of MRT to all trials that require a frequent motor response for each task, i.e., the go trials of the MARS go/no-go task, the congruent trials of the Simon task, and the target trials of the CPT, in proportion to the number of such trials per task., i.e., the go trials of the go/no-go task, congruent trials of the Simon task, and target trials of the CPT, in proportion to the number of such trials per task.

$$\text{Composite processing speed} = \frac{\sum_i n_i \text{MRT}_i}{\sum_i n_i}$$

c) *Composite intrasubject response variability* was derived from pooled intrasubject SDRT to the go trials of the go/no-go, the congruent trials of the Simon task and the target trials of the CPT altogether divided by the composite processing speed, in proportion to the number of trials the SDRT and MRT were derived from.

$$\text{Composite intrasubject response variability} = \frac{\sum_i n_i \text{SDRT}_i}{\sum_i n_i \text{MRT}_i}$$

Secondary statistical analyses

Mood and motivation. Mood before and during the MRI scan, and motivational state, perceived performance and liking of scan sessions were analysed using a series of 2×4 rANCOVAs with Group (active, sham fMRI-NF), Session (session 1 to 4) and Group×Session as fixed effects and age and medication status as covariates, while applying false discovery rate (FDR) multiple testing correction to control for false positives.

Parent feedback of fMRI-NF effectiveness and child feedback of fMRI-NF experience. ANCOVAs with Group (active, sham fMRI-NF) as fixed effects and age and medication status as covariates were used to test for group differences of the total score of each feedback questionnaire from parent and child.

Complier average causal effect (CACE) was used to estimate treatment efficacy while accounting for non-received treatments(11). The analysis was conducted with STATA16 (College Station, TX) with two-staged least square instrumental variable regression methods, which consisted of (a) regressing a treatment receipt variable on Group (the instrument); and (b) fitting the predicted value of treatment receipt, instead of Group, as predictor of outcomes in the substantive rANCOVA models, while bootstrapping (n=200) to compute 95% confidence intervals. Treatment receipt was defined as fulfilment of a number of “good quality” fMRI-NF runs above a predefined threshold (i.e., overall completed runs below threshold is considered non-receipt of treatment). A good quality run was defined as a fully completed run without motion artifacts and with a relative mean displacement below 0.9mm(12). We chose the threshold of (i) four runs, since increased rIFC activation has been found in adult participants with as few as four runs of fMRI-NF of the rIFC(13); and also (ii) seven runs, since a secondary data analysis of fMRI-NF learners and non-learners from our proof-of-concept trial has shown that learners of fMRI-NF of rIFC successfully completed a minimum of seven runs(14).

Sensitivity analyses. We explored the influence of medication status changes from post-treatment to follow-up (i.e., from medicated to non-medicated, and vice versa); and the UK COVID-19 lockdown-related data collection delays on the estimated treatment effectiveness at

follow-up. Analyses were conducted on two separate models for each covariate using STATA16. Group differences of outcomes were analysed at follow-up time points, while covarying for medication status changes or lockdown status and enforcing their effects to 0 at post-treatment(15), as medication status changes and lockdown primarily affected the follow-up data.

Intervention

Active Treatment. Treatment consisted of fifteen active or sham fMRI-NF runs distributed over four 1-hour fMRI scan sessions, which were completed within two-weeks with at least one rest day between sessions. There were two runs in session 1, five runs each in sessions 2 and 3, and three runs in session 4. Each run consisted of six 40-second "Self-regulation" and seven 30-second "Rest" blocks; and started and finished with a Rest block. During Self-regulation blocks, participants attempted to upregulate the rIFC activation, via a gamified visual feedback (i.e., a flying rocketeer). Participants were encouraged to discover individualised strategy but were also told that concentrating on the rocketeer might help. During Rest blocks, participants passively viewed a static image of a dolphin and were instructed to relax and keep still.

Sham Treatment. The sham group underwent identical procedures, except that they received sham NF, i.e., the rocketeer video of which motion was simulated using data from the last active participant who completed at least fMRI-NF 8 runs. Data generated from each active participant were used to create a playlist consisting of 15 runs presented to participants in the sham NF group. If the previous participant in the active group completed fewer than 15 runs, the playlist would be created by repeating completed runs while randomising their order in the playlist. A sham playlist using data from an active fMRI-NF pilot participant (18-year-old healthy control) was created, in case the first participant belonged to the sham group.

Transfer. A transfer run was completed after the last fMRI-NF run of the last session to examine consolidation of learning(16). The transfer run was nearly identical to the design of the regular fMRI-NF runs, but a static image of the rocketeer was presented instead of the animated visual feedback, although the participants were instructed to apply the same strategies used during the regular fMRI-NF runs.

fMRI-NF acquisitions and processing

Imaging data were acquired on a GE Discovery MR750 3T scanner (GE Healthcare, Chicago, USA) equipped with a 12-channel head coil for signal reception, at the Centre for Neuroimaging Sciences, King's College London, UK. The structural scan was acquired at the start of the first fMRI-NF session with a high resolution T1-weighted enhanced gradient echo 3D sequence (TR/TE = 7.312/3.016ms, FoV = 270mm, TI = 400ms, flip angle = 11°, matrix size = 256 × 256, voxel size = 1.05mm; slice thickness = 1.2mm, slice gap = 1.2mm). Functional scans were acquired using a T2*-weighted Echo Planar Imaging (EPI) sequence interleaved top to bottom (TR = 2s, TE = 30ms, FoV = 211mm, TI = 0ms, flip angle = 75°, matrix size = 64 × 64, voxel size = 3.3mm, slice thickness = 3mm, slice gap = 3.3mm). Dummy scans were run prior to fMRI-NF scans for MRI calibration purposes to allow for steady-state magnetisation for imaging.

Real-time data processing for fMRI-NF signal

Control of the rocketeer game was enabled by real-time transfer and analyses of fMRI data, facilitated by a custom fMRI interface and the AFNI software(17) that pre-processed and corrected head motion in real-time. Data were acquired from a region of interest (ROI) in rIFC opercular and triangular parts (ROI_{EXP}), co-registered to a structural localiser, the AFNI CA_N27_ML/TT_N template (14,138 voxels in the Talairach space of the template and 385

voxels when mapped to fMRI space). To control for nonspecific global signal changes activation from the white matter region (ROI_{REF}) were subtracted. The neural activity of the ROIs was extracted from the pre-processed images every 2s (i.e., new TR) to generate the fMRI-NF signal that was used to control the rocketeer feedback display, calculated as:

$$fMRINF\ signal(t)_n = \left[ROI_{Exp}(t) - ROI_{Ref}(t)\right]_n - \left[\widehat{ROI}_{Exp} - \widehat{ROI}_{Ref}\right]_{n-1}.$$

Thus, the rocketeer flies up faster with increasing rIFC activation. Furthermore, the rocketeer's movement in each Self-regulation block (n) depended on real-time activation of rIFC which was thresholded by average activation in the previous Rest block (n-1) to enable constant learning. A sliding window weighted average (weights: 0.125, 0.25, 0.625) was applied to the previous three time points to reduce brain signal noise and smooth the rocketeer's movement. The slow BOLD response and processing time delayed the real-time fMRI-NF signal by 5-7s, and all participants were made aware of it.

Offline fMRI data analyses

MRI data pre-processing and analyses were conducted using FSL v.6.0 (www.fmrib.ox.ac.uk/fsl). Standard pre-processing steps were applied to all MRI data using the fMRI Expert Analysis Tool (FEAT; 18).

Structural MRI (sMRI) images were re-oriented and were subsequently skull-stripped using Brain Extraction Tool, with an individually-selected threshold(19). The 4-stage FSL MCFLIRT (FMRIB's Linear Image Registration Tool) was used to estimate head motion in 24 parameters (20), by registering functional images to a middle volume reference, or a nearest image substitute unaffected by head motions, using 6 degrees-of-freedom (DOF) rigid-body transformation with an 8mm search and a correlation-ratio cost function(20). The motion outlier tool identified fMRI volumes affected by large head movements beyond 75th percentile + 1.5 times the interquartile (21), to be used as nuisance regressors in the first-level analysis. Grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor was used. Co-registration of fMRI to sMRI images were undertaken in two stages by registering the EPI data to the participant's high-resolution sMRI scan, with a 6-DOF rigid-body transformation and normal search space (90°), and the sMRI scan to a standard MNI152 brain template, with a 12-DOF affine transformation and normal search space (90°), which were then combined for each subject(20, 22). Data were high-pass filtered (100s) and smoothed with a Gaussian kernel of 5mm full-width-at-half-maximum. Data were visually inspected after each processing step. A successful fMRI-NF or transfer run (i.e., good quality runs included in statistical analyses) were defined as a fully completed run without motion artifacts and with a relative mean displacement below 0.9mm(12). The motion outlier tool was used to detect timepoints in the fMRI data affected by large head movements that MCFLIRT could not correct for. The threshold used to define an outlier is the 75th percentile + 1.5 times the interquartile range(21) The outliers detected were stored as a confound matrix that was added to the GLM to regress the effects of the motion at these timepoints during the analysis.

First-level analysis.

For each participant, BOLD activation during the fMRI-NF and transfer runs was modelled as the convolution of the Gamma HRF and a box-car function to represent the Rest and Self-regulation blocks, while adding their temporal first derivatives to correct slice timing differences(23), and the 24-motion parameters as nuisance regressors. To investigate brain self-regulation, the contrast Self-regulation – Rest was computed, and the resulting COPE (contrast of parameter estimate) images were entered into a second-level analyses.

To select the optimum motion parameter nuisance regressor, we performed analyses using GLMs with 6, 12, and 24 motion parameters, with and without additional column vectors representing motion outliers. We subsequently compared the value of the models' normalised residuals [100+(value-mean)/standard deviation]. The GLM with the lowest residual values, which uses the 24-motion parameters nuisance regressors, was deemed to be the optimum model.

Second-level whole-brain analyses.

fMRI-NF sessions. The COPE images of all good quality runs for each participant were brought into a group level GLM for whole-brain analyses to investigate the repeated-measures ANCOVA of Group (active, sham) × Session (session 1, session 2, session 3, session 4) interaction. This GLM also allowed the investigation of within session effects. Medication status, age, and relative mean displacement were entered as covariates. For all fMRI analyses, a cluster threshold of $\alpha < 0.05$ with a family-wise error rate correction for multiple comparisons was applied.

Baseline and final fMRI-NF run. The second GLM estimated the effect of fMRI-NF by contrasting the final with the baseline run. Since the first fMRI-NF run is often used for training (24-26), we classed the second run as the "baseline" run. Run 11 was classified as the final run as this was the highest number of runs that was completed by at least 75% of participants. The COPE images of run 2 and run 11 were combined to their corresponding treatment group in the GLM to create the four regressors of interest: Active-Baseline, Active-Final, Sham-Baseline, and Sham-Final with age, medication status, and relative mean displacement as covariates. The comparisons between baseline and final runs were performed within (Active-Baseline *vs.* Active-Final and Sham-Baseline *vs.* Sham-Final) and between groups (Active-Baseline *vs.* Sham-Baseline and Active-Final *vs.* Sham-Final).

Transfer run. The COPE images from the transfer runs, contrasting Regulation and Rest blocks, for each participant were imported into a GLM to investigate the transfer effect within and between groups. This GLM contained two regressors of interest (Active-Transfer and Sham-Transfer) with age, medication status, and relative mean displacement added as covariates. The main effect of each regressor of interest measured the within group effect, while the contrast of these two regressor measured the between group effect.

Linear regression. To test the hypothesised linear increase of activation across all fMRI-NF runs, a linear regression analysis was conducted. To perform this analysis, an average map of each run was constructed for each group (i.e., 30 maps in total, with 15 runs for each group), which were then imported in a GLM. One regressor of interest was created for each group, with a linear weight increasing with the number of runs, with age, medication status, and relative mean displacement included as covariates. The linear progressive increase of activation was assessed within (using the main effect of each regressor of interest) and between groups (by contrasting the regressors of interest).

ROI analyses

Small volume correction. To specifically explore clusters of activation that were related to self-regulation ability within rIFC, all GLMs in the above analyses were subsequently used to analyse activation within the rIFC ROI_{EXP} mask (used for the fMRI-NF procedure) while applying a small volume correction (with a cluster threshold of $\alpha < 0.05$ for a family-wise error rate correction for multiple comparisons).

ROI data extraction. Lastly, average rIFC activations were extracted from the COPE images, using the FSL command *Featquery* and were imported into SPSS to conduct correlation analyses with the outcome measures.

Supplementary Results

Secondary analyses of clinical data

Mood and motivation. Repeated ANCOVAs indicated no significant Group or Group×Session effects on mood before and during MRI scan, performance, motivation and liking, although there was a marginal but non-significant effect of Session on mood (F[3, 257.2]=3.49, p=.055) and on motivation (F[3, 253.5]=3.27, p=.055) during scanning, FDR corrected. Mood during scan increased in session 3 *vs.* 4 (p=.012), while motivation decreased in the last relative to first session (p=.012) (Fig S2).

Parent feedback of fMRI-NF effectiveness and child feedback of fMRI-NF experience. ANCOVAs indicated no significant Group effect on the total score of parent (F[1,78]=.099, p=.75) and children feedback (F[1,83]=.077, p=.78).

CACE. Except for the group difference in parent-rated irritability that became non-significant (p=.054) at post-treatment during the CACE analyses, other group differences at post-treatment and follow-up for ADHD-RS total score, parent-rated irritability and participants' motor response inhibition on the go/no-go task remained stable with either seven or four minimum good quality runs as criteria of treatment receipt. Furthermore, the group difference of weekly evening and morning behaviour and composite intrasubject response variability became significant at post-treatment after applying both thresholds of treatment receipt. The active relative to sham fMRI-NF group had higher ratings of evening and morning difficulties (ps=.045-.046) and composite response variability (ps=.039-.041) (Table S3).

Effects of COVID-19 lockdown. Follow-up data collection was marginally but non-significantly later in 18 participants due to the UK COVID-19 lockdown (6.90 ± 1.67 months) than pre-lockdown time (6.39 ±.68 months; t = 1.96, p = .053). The impact of lockdown on treatment effects at follow-up was non-significant for ADHD-RS, parent-rated ARI and go/no-go probability of inhibition (ps = .30 - .91), which remained robustly non-significant after adjusting for the lockdown, i.e., ADHD-RS total score (B=.53; 95%CI[-3.65, 4.71]), parent-rated ARI (B=-.099; 95%CI[-.292, .093]) and Go/go-go probability of inhibition (B=4.36; 95%CI[-1.55, 10.3]).

Medication status changes at follow-up. Medication status changes between post-treatment and follow-up took place in 8 participants (5 in sham fMRI-NF) during the clinical data collection, and 13 participants (9 in sham fMRI-NF) during the neurocognitive data collection. Their impact on the follow-up treatment effects were non-significant for each measure (ps = .11 - .84), which remained robustly non-significant after adjusting for medication status changes, i.e., ADHD-RS total score (B=.47; 95%CI[-3.67, 4.62]), parent-rated ARI (B=-.097; 95%CI[-.289, .095]) and go/no-go probability of inhibition (B=4.46; 95%CI[-1.54, 10.5]).

fMRI data exploration

Previous studies have included practice run(s) before the actual NF training to allow participants to acclimatise to the scanning environment and familiarise themselves with the NF procedure(24-26). Accordingly, the first fMRI-NF run in the first session is excluded from the main analyses. Moreover, in our data, the first run of the NF training was also classified as a potential outlier after data exploration. Unusually high activation was observed in the whole

brain, and within the triangular and opercular part of rIFC, in both groups during the first run only, which may be more related to noise and be unspecific to the fMRI-NF training of the rIFC. Given that the sample were young participants with ADHD, and it was their first run of the fMRI-NF training, factors such as initial nerves, the novelty of being scanned for the first time, and/or not understanding the task properly may have been confounds.

All 88 participants were included in the final fMRI data analyses. However, 17% of runs from the active and 17% of runs from the sham group were excluded from the analyses due to large head motions that were detected through visual inspection and had exceeded the relative mean displacement threshold of 0.9mm.

Within session analyses

Group differences were also analysed for each session separately.

Session 1. No significant group differences were observed at the whole brain and ROI levels.

Session 2. The active group compared to the sham group had increased activation within two clusters of the right posterior middle temporal gyrus (cluster #1: *p*=.023, peak MNI coordinates: [x = 48, y = -52, z = 8], cluster size [*k*]=253 voxels, BA21; cluster #2: *p* =.025, [48, -66, 8], *k*=248, BA37), while the sham group compared to the active group had higher activation in a cluster within left middle frontal gyrus/IFC region (*p*=.008, [-40, 54, 18], *k*=327, BA45). Within the ROI mask, only the active versus sham group showed increased activation (cluster #1: *p*=.028, [54, 18, 28], *k*=33, BA44; cluster #2: *p* =.033, [48, 24, 12], *k*=28, BA45; cluster #3: *p*=.044, [56, 38, 14], *k*=19, BA 45; cluster #4: *p*=.049, [46, 10, 38], *k*=16, BA44).

Session 3. The active versus sham group showed increased activation within the ROI (cluster #1: *p*=.043, [62, 26, 12], *k*=20, BA45; cluster #2: *p* =.047, [56, 40, 14], *k*=17, BA45).The sham group revealed higher activation compared to the active group in the whole brain level in the left middle cingulate cortex/insular areas (*p*=.001, [-20, -28, 36], *k*=476, BA48), brainstem (*p*=.012, [-4, -30, -12], *k*=331), and white matter (*p*=.012, [18, -36, 38], *k*=300) and in one cluster within the ROI mask (*p* =.014, [44, 14, 10], *k*=60, BA48).

*Session 4***.** The active group relative to the sham group had higher activation in several clusters including the left middle frontal gyrus (*p*=.000, [-36, 14, 52], *k*=811, BA9), right cerebellum, fusiform and lingual gyri, and hippocampus (*p*=.001, [34, -50, 8], *k*=504, BA30), right middle temporal gyrus (*p*=.002, [66, -34, -6], *k*=422, BA21), right superior temporal gyrus (*p*=.015, [44, -40, 14], *k*=285, BA41), right medial superior frontal gyrus and anterior cingulate cortex (*p*=.023, [10, 32, 40], *k*=253, BA32), left middle temporal gyrus (*p*=.034, [-52, -30, 6], *k*=228, BA22), and left middle occipital gyrus (*p*=.034, [-50, -74, 16], *k*=228, BA39) and within the ROI mask (*p*=.023, [56, 18, 26], *k*=40, BA44). The sham group did not show any greater activation in any brain regions when compared to the active group during this session at the whole brain or ROI levels.

Covariates. As an exploratory analysis, we examined the effects of each of the covariates (i.e., medication, age, and relative mean displacement in head motion) on whole brain activation across all fMRI-NF runs across both groups combined. Out of the three covariates, motion had the largest effect on the level of activation across the whole brain. The impact of motion was also shown by the significant positive correlation between relative mean displacement and rIFC activation (functional cluster ROI: *r* =.59, *p* <.0001; triangular part: *r* =.61, *p* <.0001; opercular part: *r* =.64, *p* <.0001).

Motion and feedback. Head motion and NF scores did not corelate significantly (r=0.037, *p*=0.73.) in the active group.

Correlations between outcomes and rIFC activation

No significant correlation was found between rIFC activation changes (final *vs*. baseline run) and changes (post-treatment *vs.* baseline) of ADHD-RS total score (Pearson's *r*=.001, *p*=.99; two-tailed, n=41), parent-rated ARI (*r*=.24, *p*=.13; two-tailed, n=42), and go/no-go probability of inhibition (*r*=.24, *p*=.13; two-tailed, n=42) in the active group. The rIFC activation change in the active group correlated significantly with the average NF performance score from the last *vs.* first run (Pearson's *r*=.53, *p*<.001; two-tailed, n=42), but no equivalent correlation was found in the sham group (*r*=.08, *p*=.63; two-tailed, n=42). In the active group, the average NF performance score was negatively correlated with the number of runs (Pearson's *r*= -.12, *p*=.038; two-tailed, n=42).

FIGURE S1. CONSORT flow diagram

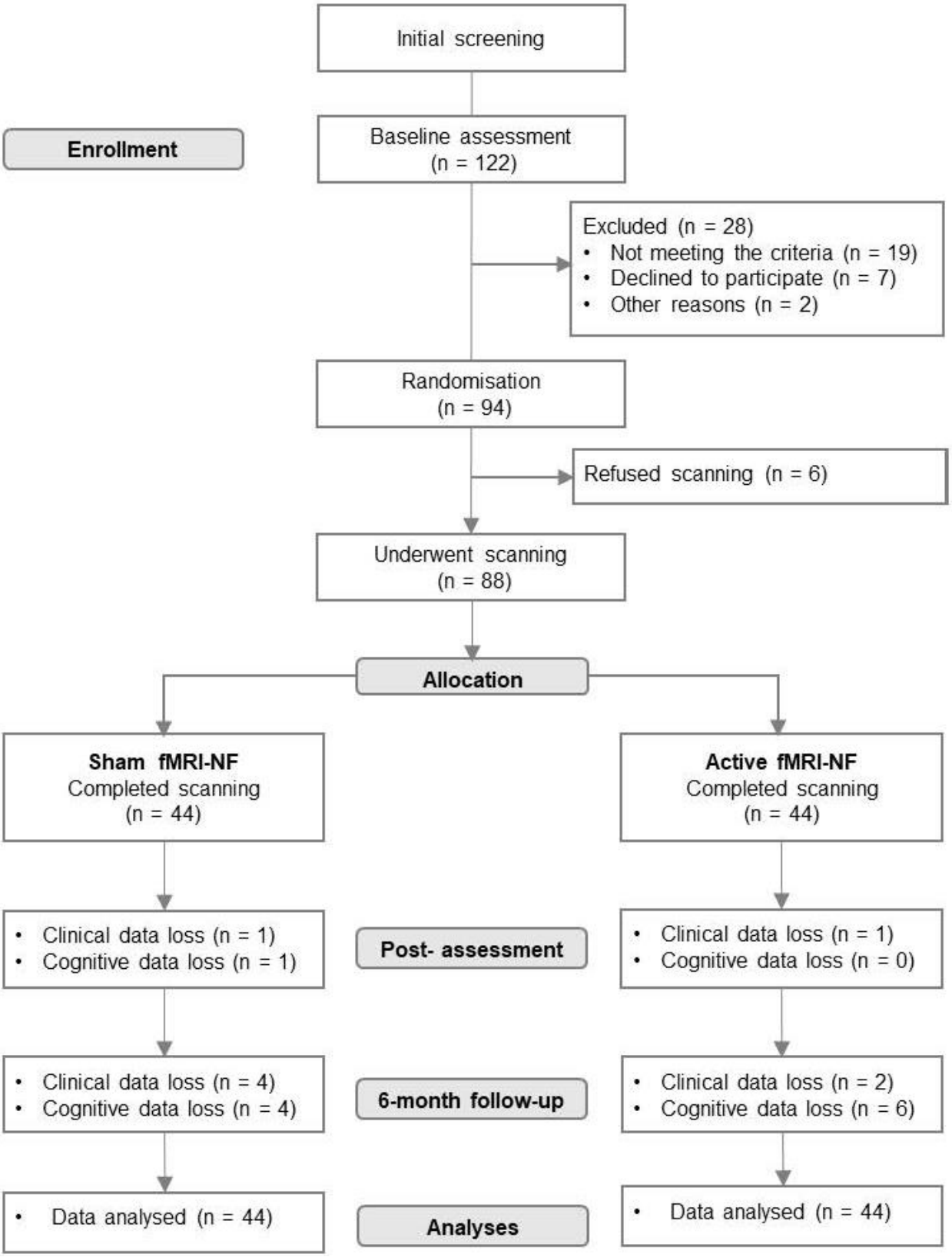
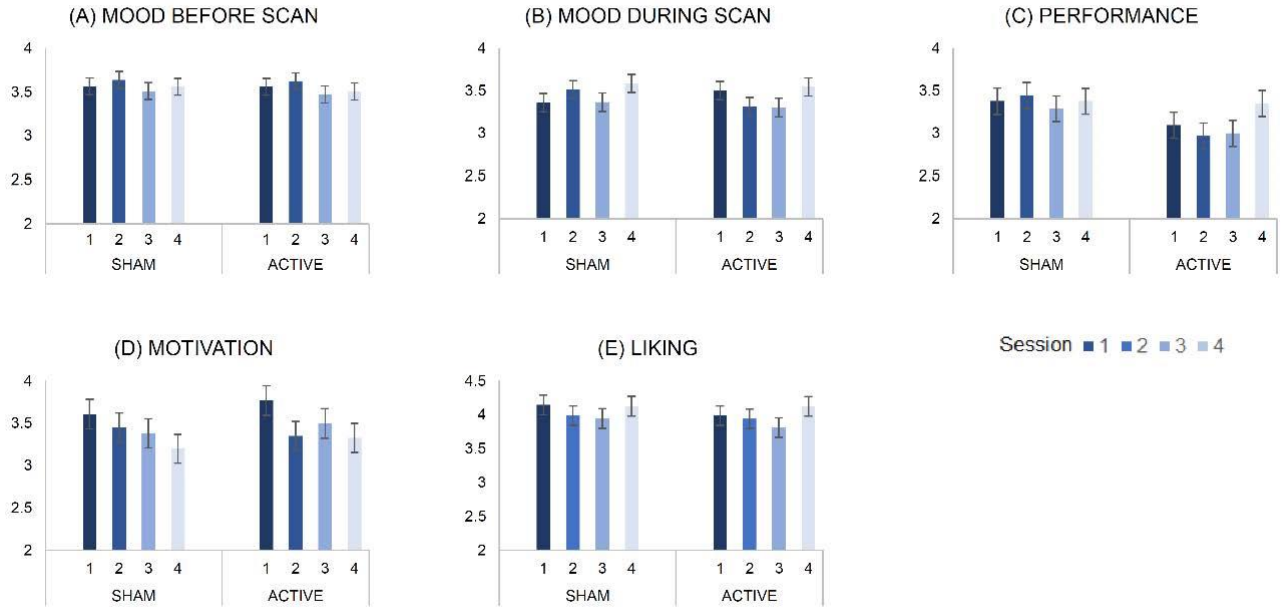
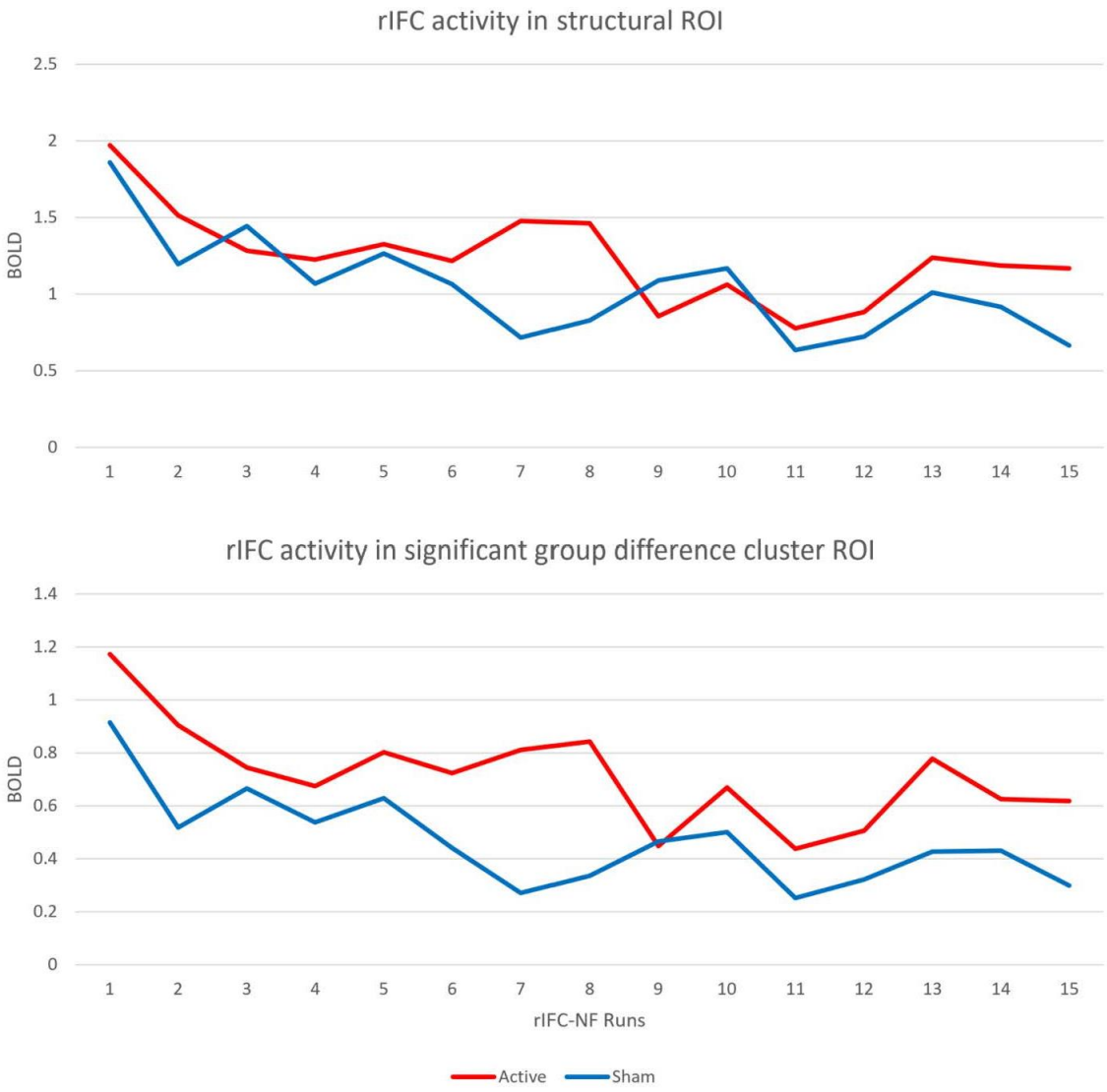


FIGURE S2. Self-rated mood state questionnaire



The figure depicts self-rated mood state before (A) and during scanning (B), self-perceived performance level (C), motivation (D) and liking of fMRI-NF training (E), stratified by treatment group and session number. Repeated ANCOVAs indicated no significant main effects of Group or interaction Group×Session. Marginal but non-significant effects of Session on Mood and on Motivation were found during scanning, FDR-corrected for multiple comparisons. Mood during scan became more positive from Session 3 to 4, while motivation was lower in the last relative to the first session.

FIGURE S3. rIFC activation across runs



The figure depicts the activation of the rIFC across the 15 NF runs in the active (red) and sham (blue) groups. (A) Activation in the structural ROI used in the NF procedure and the small volume correction. (B) Activation in the significant cluster from the whole brain group difference (i.e., active > sham group).

TABLE S1. Sample characteristics at baseline

	Sham fMRI-NF (n = 44)		Active fMRI-NF (n = 44)		Statistics	
(a) Demographics	M	SD	M	SD	*t*(86)	*p*
Age, in months	13.3	8.14	13.1	7.30	0.46	.65
Education, in years	8.30	2.32	8.05	1.93	0.55	.58
FSIQ	104.8	12.4	102.4	13.6	0.88	.38
(b) Dimensional trait measures	M	SD	M	SD	*t*(86)	
ADHD-RS						
Total score	37.8	9.3	37.3	9.5	0.25	.80
Inattention	21.3	4.0	19.8	4.4	1.76	.08
Hyperactivity/impulsivity	16.5	6.8	17.5	6.0	-0.78	.44
Conners-3P						
DSM-5 Inattention	81.1	7.7	80.0	8.3	0.67	.51
DSM-5 Hyperactivity/impulsivity	81.4	13.7	84.3	10.1	-1.11	.27
DSM-5 ODD	69.3	15.1	72.5	14.3	-1.00	.32
DSM-5 CD	57.7	16.1	59.6	14.1	-0.58	.57
ADHD Index	13.5	4.2	14.5	4.36	-1.05	.30
SCQ	6.98	5.93	7.50	4.63	0.46	.65
(c) KSADS diagnostic measures	N	(%)	N	(%)	$X^{2(a)}$/FET[(b)]	*p*
ADHD research diagnosis						
Combined presentation	25	56.8	36	81.8	**.01[(a)]***	
Inattentive presentation	19	43.2	8	18.2		
ODD	18	40.9	21	47.7	.52[(a)]	
CD	0	0	1	2.3	>.99[(b)]	
Alcohol use	0	0	1	2.3	>.99[(b)]	
Drug use	0	0	0	0	--	
(d) Cognitive indices	M	SD	M	SD	*t*(86)	*p*
Go/no-go probability of inhibition	48.3	18.4	47.9	16.3	0.12	.91
Simon RT interference effect	67.1	30.5	66.4	40.3	0.10	.92
CPT omission errors	17.4	15.6	15.5	10.5	0.67	.51
CPT commission errors	3.19	4.62	2.33	2.62	1.07	.29
MCT omission errors	41.7	17.3	40.7	16.3	0.25	.80
MCT commission errors	6.96	7.51	6.34	5.81	0.43	.67
WCST perseverative errors	9.34	4.38	9.09	3.63	0.29	.77
WCST non-perseverative errors	10.5	6.23	10.4	4.66	0.02	.99
Working memory total score	22.7	10.6	24.4	12.0	0.68	.50
Composite response prematurity	3.84	4.05	3.20	3.04	0.83	.41
Composite processing speed	389.1	45.9	387.2	44.4	0.20	.84
Composite response variability	0.31	0.08	0.29	0.07	0.94	.35
(e) Medication	N	(%)	N	(%)	FET[(b)]	*p*
Medication status						
Naïve	14	31.8	11	25.0	.88[(b)]	
Currently medicated – off	15	34.1	15	34.1		
Currently medicated – on	12	27.3	14	31.8		
Not currently medicated	3	6.8	4	9.1		
Current medication type						
No medication	14	31.8	11	25.0	.62[(b)]	
Methylphenidate	26	59.1	24	54.5		
(Lis)dexamfetamine	3	6.8	6	13.6		
Atomoxetine	1	2.3	2	4.5		
MPH and Guanfacine	0	0	1	2.3		

FSIQ = Full-scale IQ; CPT = Continuous Performance Task; MCT = Mackworth clock vigilance task; WCST = Wisconsin card sorting task, ODD = oppositional defiant disorder, CD = conduct disorder, DSM-5 = Diagnostic and Statistical Manual of Mental Disorders-5th edition, SCQ = social communication questionnaire, N = participants number, M = mean, SD = standard deviation, MPH = methylphenidate. $^{(a)}$ *p*-values for X^2 = Chi^2 statistics, $^{(b)}$ *p*-values for FET = Fisher's exact test, *p*-values were uncorrected for multiple testing.

TABLE S2. Within-subject planned comparison of outcomes at post-treatment and follow-up relative to baseline

Measures	Sham				Active			
	Baseline – Post-treatment		Baseline – Follow-up		Baseline – Post-treatment		Baseline – Follow-up	
	ΔM [95%CI]	*p*	*ΔM* [95%CI]	*p*	*ΔM* [95%CI]	*p*	*ΔM* [95%CI]	*p*
(a) Primary clinical outcome								
ADHD RS Total (P)	**9.14 [5.95, 12.3]**	**<.001**	**5.14 [1.87, 8.42]**	**.002**	**6.83 [3.69, 9.98]**	**<.001**	**4.21 [1.06, 7.35]**	**.009**
(b) Secondary clinical outcomes								
Conners 3P ADHD Index (P)	2.03 [.43, 3.63]	**.014**	1.79 [.15, 3.44]	**.033**	2.34 [.75, 3.93]	**.005**	2.06 [.45, 3.66]	**.013**
ARI (P)	.18 [.03, .32]	**.018**	-.10 [-.25, .05]	.17	.001 [-.16, .16]	.99	.013 [-.14, .17]	.87
ARI (C)	.06 [-.06, .17]	.33	.05 [-.07, .17]	.45	.07 [-.03, .18]	.17	.04 [-.06, .15]	.43
MEWS (C)	.72 [-1.44, 2.88]	.51	1.82 [-.39, 4.03]	.10	1.96 [-.54, 4.45]	.12	3.74 [1.22, 6.25]	**.004**
CIS (P)	3.09 [.032, 5.85]	**.030**	--	--	.73 [-1.9, 3.4]	.58	--	--
WREMB-R (P)	4.44 [2.77, 6.11]	**<.001**	--	--	1.65 [-.63, 3.93]	.15	--	--
Side effects (P)	.13 [-0.05, .31]	.14	--	--	.13 [-.17, .42]	.39	--	--
(c) Secondary cognitive outcomes								
Go/no-go probability of inhibition	-6.25 [-10.6, -1.90]	**.005**	-2.25 [-6.71, 2.21]	0.32	-.28 [-5.18, 4.61]	.91	-6.78 [-11.9, -1.65]	**.010**
Simon RT interference effect	7.44 [-2.46, 17.4]	.14	15.2 [5.10, 25.3]	**.004**	4.22 [-7.96, 16.4]	.49	5.60 [-7.12, 18.3]	.38
CPT omission errors	.60 [.18, 1.01]	**.005**	.90 [.47, 1.32]	**<.001**	.67 [.20, 1.15]	**.006**	.98 [.48, 1.50]	**<.001**
CPT commission errors	.39 [.13, .66]	**.004**	.44 [.17, .71]	**.002**	.26 [.03, .49]	**.027**	.44 [.20, .68]	**.001**
MCT omission errors	10.3 [5.71,14.9]	**<.001**	9.37 [4.64, 14.1]	**<.001**	8.86 [3.66, 14.1]	**.001**	12.1 [6.68, 17.6]	**<.001**
MCT commission errors	.12 [.05, .19]	**.001**	.11 [.04, .19]	**.003**	.09 [.002, .18]	**.046**	.09 [-.007, .19]	.07
WCST perseverative errors	.14 [.05, .22]	**.002**	.24 [.15, .33]	**<.001**	.12 [.05, .20]	**.001**	.20 [.13, .28]	**<.001**
WCST non-perseverative errors	.09 [-0.005, .18]	.06	.17 [.07, .26]	**.001**	.09 [.01, .18]	**.030**	.21 [.13, .30]	**<.001**
Working memory total score	-4.67 [-8.48, -.86]	**.017**	-7.63 [-11.5, -3.73]	**<.001**	-2.41 [-6.68, 1.87]	.27	-4.20 [-8.68,.28]	.066
Composite response prematurity	.16 [-.02, .34]	.09	.11 [-.07, .30]	0.24	-.12 [-.30, .05]	.17	.13 [-.05, .32]	**.016**
Composite processing speed	.008 [-.002, .02]	.11	.03 [.02, .04]	**<.001**	.007 [-.004, .02]	.21	.015 [.003, .03]	**.016**
Composite response variability	.013 [.006, .02]	**.001**	.01 [.004, .02]	**.003**	.003 [-.004, .009]	.42	.01 [.003, .016]	**.006**

Abbreviations. ADHD-RS = ADHD Rating Scale; ARI = Affective Reactivity Index, MEWS = Mind Excessively Wandering Scale, CIS = Columbia Impairment Scale; WREMB-R = Weekly Rating of Evening and Morning Behavior-Revised; (P/C) = (Parent/Children); CPT = Continuous Performance Task; MCT = Mackworth clock vigilance task; WCST=Wisconsin card sorting task. *ΔM* [95%CI] = mean difference [95% confidence intervals]. *p*-values were uncorrected for multiple testing.

TABLE S3. Group effects at post and at follow-up and CACE

Clinical outcomes	Active – Sham				Active – Sham CACE (threshold = 4)				Active – Sham CACE (threshold = 7)			
Measure	Post-treatment		Follow-up		Post-treatment		Follow-up		Post-treatment		Follow-up	
	ΔM [95%CI]	*p*	*ΔM* [95%CI]	*p*	*ΔM* [95%CI]	*p*	*ΔM* [95%CI]	*p*	*ΔM* [95%CI]	*p*	*ΔM* [95%CI]	*p*
(a) Primary clinical outcome												
ADHD RS Total (P)	1.91 [-6.06, 2.23]	.36	-.53 [-4.74, 3.67]	.80	1.91 [-2.13, 5.95]	.35	.511 [-3.53, 4.55]	.80	1.95 [-2.18, 6.09]	.36	.520 [-3.61, 4.66]	.80
(b) Secondary clinical outcomes												
Conners ADHD Index (P)	.068 [-2.19, 2.06]	.95	-.099 [-2.27, 2.07]	.93	.074 [-2.08, 2.22]	.95	.128 [-2.02, 2.28]	.91	.076 [-2.33, 2.48]	.95	.131 [-2.27, 2.53]	.92
ARI (P)	.194 [-.384, .003]	**.046**	-.099 [-.292, 0.94]	.31	.200 [-.003, .397]	**.047**	-.100 [-.297, .096]	.32	.205 [-.003, .412]	.054	-.102 [-.310, .106]	.34
ARI (C)	-.003 [-.169, .175]	.97	-2.81e-5 [-.18, .18]	>.99	-.003 [-.186, .180]	.98	-.0006[-.186, .185]	>.99	-.003 [-.189, .183]	.98	-.0006[-.183, .182]	>.99
MEWS (C)	-1.23 [-1.58, 4.04]	.39	1.67 [-1.21, 4.55]	.25	-1.32 [-4.38, 1.74]	.30	-1.60 [-.4.66, 1.46]	.21	-1.35 [-3.91, 1.22]	.30	-1.64 [-.4.20, .93]	.21
CIS (P)	2.63 [-5.98, 733]	.12	--	--	2.67 [-.600, 5.96]	.11	--	--	2.74 [-.581, 6.06]	.11	--	--
WREMB (P)	2.57 [-5.25, .102]	.059	--	--	2.63 [.049, 5.21]	**.046**	--	--	2.69 [.061, 5.32]	**.045**	--	--
Side effects (P)	.029 [-3.06, 3.01]	.99	--	--	-.028 [-3.28, 3.34]	.99	--	--	-.029 [-3.36, 3.41]	.99	--	--
(c) Secondary cognitive outcomes												
GNG prob. of inhibition	-6.02 [.38, 11.7]	**.037**	4.38 [-10.3, 1.57]	.15	-5.57 [-10.8, -.334]	**.037**	5.11 [-.131, 10.4]	.056	-6.30[-11.6, -.991]	**.020**	4.56 [-.747, 9.86]	.092
Simon RT int. effect	2.55 [-12.8, 7.74]	.63	9.11 [-19.9, 17.3]	.099	2.55 [-9.48, 14.6]	.68	8.38 [-3.65, 20.4]	.17	2.31 [-9.26,13.9]	.70	8.55 [-3.65, 20.7]	.17
CPT omission errors	1.50 [-1.77, 4.76]	.37	-1.15 [-2.29, 4.59]	.51	-1.49 [-4.52, 1.54]	.34	-1.43 [-4.46,1.60]	.36	-1.61 [-4.52, 1.31]	.28	-1.57 [-4.49,1.35]	.29
CPT comm. errors	.226 [-.560, 1.01]	.57	-.045 [-.781, 872]	.91	-.220 [-1.04, .597]	.60	-.074 [-.891, .743]	.86	-.242 [-1.09, .608]	.58	-.11 [-.958, .740]	.80
MCT omission errors	.918 [-6.74, 4.91]	.76	-2.62 [-3.47, 8.70]	.40	.595 [-6.08, 7.27]	.86	-4.47 [-11.1, 2.20]	.19	.878 [-5.85, 7.61]	.80	-4.26 [-11.0, 2.47]	.22
MCT comm. errors	.773 [-2.64, 1.09]	.41	1.36 [-3.31, .588]	.17	.832 [-.816, 2.48]	.32	1.20 [-.448, 2.85]	.15	.825 [-.814, 2.46]	.32	1.16 [-.477, 2.80]	.17
WCST pers. errors	.472 [-1.91, .969]	.52	.148 [-1.65, 1.36]	.85	.501 [-1.02, 2.03]	.52	-.240 [-1.77, 1.28]	.76	.460 [-1.25, 2.17]	.60	-.335 [-2.05, 1.38]	.70
WCST non-pers. errors	1.01 [-3.20, 1.18]	.36	-.58 [-1.69, 2.84]	.61	.878 [-1.41, 3.16]	.45	-1.12 [-3.40,1.17]	.34	1.07 [-1.35, 3.50]	.39	-.957 [-3.38,1.47]	.44
WM total score	-1.05 [-4.10, 6.20]	.69	-2.29 [-7.68, 3.11]	.40	-1.82 [-6.13, 2.48]	.41	-2.44 [-6.74, 1.86]	.27	-1.06 [-5.72, 3.61]	.66	-1.46 [-6.13, 3.20]	.54
Comp. resp. prematurity	.78 [-1.70, .147]	.099	-.34 [-.638, 1.31]	.50	.716 [-.229, 1.66]	.14	-.509 [-1.45, .436]	.29	.785 [-.104, 1.67]	.084	-.470 [-1.36, .419]	.30
Comp. proc. speed	1.17 [-14.2, 11.9]	.86	12.5 [-26.5, 1.11]	.072	2.54 [-14.2, 19.3]	.77	12.1 [-4.61, 28.9]	.16	1.14 [-14.0, 16.3]	.88	10.6 [-4.53, 25.7]	.17
Comp. resp. variability	.015 [-.032, .002]	.09	-.005 [-.013, .023]	.60	.016 [.0006, .030]	**.041**	-.006 [-.021, .010]	.48	.015 [.0008, .030]	**.039**	-.007 [-.022, .007]	.33

Abbreviations. CACE = complier average causal effect; ADHD-RS = ADHD Rating Scale; ARI = Affective Reactivity Index, MEWS = Mind Excessively Wandering Scale, CIS = Columbia Impairment Scale; WREMB-R = Weekly Rating of Evening and Morning Behavior-Revised; (P/C) = (Parent/Children); CPT = Continuous Performance Task; MCT = Mackworth clock vigilance task; WCST=Wisconsin card sorting task. *ΔM* [95%CI] = mean difference and 95% confidence intervals. *p*-values were uncorrected for multiple testing and *denotes significance *p* <.05

References

1. Alegria AA, Wulff M, Brinson H, Barker GJ, Norman LJ, Brandeis D, Stahl D, David AS, Taylor E, Giampietro V, Rubia K. Real-time fMRI neurofeedback in adolescents with attention deficit hyperactivity disorder. Hum Brain Mapp. 2017;38:3190-3209.

2. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39:175-191.

3. Cubillo A, Smith AB, Barrett N, Giampietro V, Brammer MJ, Simmons A, Rubia K. Shared and drug-specific effects of atomoxetine and methylphenidate on inhibitory brain dysfunction in medication-naive ADHD boys. Cereb Cortex. 2014;24:174-185.

4. Rubia K, Halari R, Cubillo A, Smith AB, Mohammad A-M, Brammer M, Taylor E. Methylphenidate normalizes fronto-striatal underactivation during interference inhibition in medication-naive boys with attention-deficit hyperactivity disorder. Neuropsychopharmacology. 2011; 36:1575-1586.

5. Rubia K, Smith AB, Taylor E. Performance of children with attention deficit hyperactivity disorder (ADHD) on a test battery of impulsiveness. Child Neuropsychol. 2007; 13:276-304.

6. Penadés R, Catalán R, Rubia K, Andrés S, Salamero M, & Gastó C. Impaired response inhibition in obsessive compulsive disorder. Eur Psychiatry, 2007; *22*:404–410.

7. Simon JR, Berbaum K. Effect of irrelevant information on retrieval time for relevant information. Acta Psychologica. 1988; 67:33-57.

8. Lichstein KL, Riedel BW, Richman SL. The Mackworth clock test: a computerized version. J Psychol. 2000; 134:153-161.

9. Heaton RK, Staff PAR: Wisconsin card sorting test-64: computer version 2-research edition (WCST-64:CV2). Lutz, FL, Psychological Assessment Resources; 2003.

10. Tulsky DS, Carlozzi N, Chiaravalloti ND, Beaumont JL, Kisala PA, Mungas D, Conway K, Gershon R. NIH toolbox cognition battery (NIHTB-CB): list sorting test to measure working memory. J Int Neuropsychol Soc. 2014; 20:599-610.

11. Dunn G, Maracy M, Dowrick C, Ayuso-Mateos JL, Dalgard OS, Page H, Lehtinen V, Casey P, Wilkinson C, Vazquez-Barquero JL, Wilkinson G, group O. Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. Br J Psychiatry. 2003; 183:323-331.

12. Owens MM, Allgaier N, Hahn S, Yuan D, Albaugh M, Adise S, Chaarani B, Ortigara J, Juliano A, Potter A, Garavan H. Multimethod investigation of the neurobiological basis of ADHD symptomatology in children aged 9-10: baseline data from the ABCD study. Transl Psychiat. 2021;11:64.

13. Rota G, Sitaram R, Veit R, Erb M, Weiskopf N, Dogil G, Birbaumer N. Self-regulation of regional cortical activity using real-time fMRI: the right inferior frontal gyrus and linguistic processing. Hum Brain Mapp. 2009; 30:1605-1614.

14. Lam SL, Criaud M, Alegria A, Barker GJ, Giampietro V, Rubia K. Neurofunctional and behavioural measures associated with fMRI-neurofeedback learning in adolescents with attention-deficit/hyperactivity disorder. Neuroimage Clin. 2020; 27:102291.

15. MacPherson H, Tilbrook H, Agbedjro D, Buckley H, Hewitt C, Frost C. Acupuncture for irritable bowel syndrome: 2-year follow-up of a randomised controlled trial. Acupunct Med. 2017; 35:17-23.

16. Sulzer J, Haller S, Scharnowski F, Weiskopf N, Birbaumer N, Blefari ML, Bruehl AB, Cohen LG, DeCharms RC, Gassert R, Goebel R, Herwig U, LaConte S, Linden D, Luft A, Seifritz E, Sitaram R. Real-time fMRI neurofeedback: progress and challenges. Neuroimage. 2013; 76:386-399.

17. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 1996; 29:162-173.

18. Woolrich MW, Ripley BD, Brady M, Smith SM. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage. 2001; 14:1370-1386.

19. Smith SM. Fast robust automated brain extraction. Human Brain Mapp. 2002; 17:143-155.

20. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 2002; 17:825-841.

21. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage. 2012; 59:2142-2154.

22. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal. 2001; 5:143-156.

23. Sladky R, Friston KJ, Tröstl J, Cunnington R, Moser E, Windischberger C. Slice-timing effects and their correction in functional MRI. Neuroimage. 2011; 58:588-594.

24. Hellrung L, Dietrich A, Hollmann M, Pleger B, Kalberlah C, Roggenhofer E, Villringer A, Horstmann A. Intermittent compared to continuous real-time fMRI neurofeedback boosts control over amygdala activation. Neuroimage. 2018; 166:198-208.

25. Young KD, Zotev V, Phillips R, Misaki M, Yuan H, Drevets WC, Bodurka J. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. PloS One. 2014;9:e88785.

26. Zotev V, Phillips R, Misaki M, Wong CK, Wurfel BE, Krueger F, Feldner M, Bodurka J. Real-time fMRI neurofeedback training of the amygdala activity with simultaneous EEG in veterans with combat-related PTSD. Neuroimage Clin. 2018; 19:106-121.