Supplemental material

1 Materials and methods

1.1 Studies selection

Search terms entered in Pubmed (last check on 12/02/2018) were: «ADHD OR adhd OR attention deficit disorder with hyperactivity OR minimal brain disorders OR syndrome hyperkinetic OR hyperkinetic syndrome OR hyperactivity disorder OR hyperactive child syndrome OR childhood hyperkinetic syndrome OR attention deficit hyperactivity disorders OR attention deficit hyperactivity disorder OR adhd attention deficit hyperactivity disorder OR addh OR overactive child syndrome OR attention deficit hyperkinetic disorder OR hyperkinetic disorder OR attention deficit disorder hyperactivity OR attention deficit disorders hyperactivity OR child attention deficit disorder OR hyperkinetic syndromes OR syndromes hyperkinetic OR hyperkinetic syndrome childhood) AND (randomized control trial OR RCT OR randomized control study OR Pilot Study OR Study OR Trial OR randomized trial) AND (neurofeedback OR "EEG biofeedback" OR neurotherapy OR SCP OR "slow cortical potentials OR Theta Beta Ratio or TBR"».

1.2 Perform a meta analysis

To conduct meta-analysis, different software exist: for instance Cortese et al. [2016] used RevMan 5.3 [Cochrane Collaboration, 2011] that computes the Effect Size (ES) and its variance of each included study by applying the formula presented in Morris [2008]. However, in order to compute the variance of the ES, the pooled within-group Pearson correlation ρ (i.e the pre-post correlation) was required [James et al., 2013]. In our case, this correlation was not known and the raw data were not available so we took an approximation: Balk et al. [2012] found that a value of 0.5 yields values closer to those computed with the right value of the correlation.

In this replication of the work of Cortese et al., the same formulas were used [Borenstein et al., 2009] but instead of using RevMan, a Python code was developed in order to perform the meta-analysis. To increase replicability and transparency and promote open science, we provide the full

raw data used for this research as well as the Python code on a GitHub repository [Bussalb, 2018]; it was tested with Cortese et al. [2016] raw data to show that same results were found and could be used for replication and expansion of this work. The toolbox could also be used to run any similar meta-analysis.

To perform the meta-analysis several steps must be followed. First the choice of the model: this analysis is based on either one of the following statistical models [Borenstein et al., 2009]:

- *the fixed-effect model*: the true ES (i.e the ES that would be observed with an infinitely large sample size) is the same for all the studies in the analysis. The differences between the actually observed ESs are due to sampling errors;
- *the random-effects model*: the true ES could vary from study to study. The differences between the observed ESs are due to sampling errors but also to the various designs of the studies (for instance the number of participants or the implementation).

In the present case, although the studies included into the meta-analysis met the same criteria, they remained different from each other on various points, so the random effects model was more appropriate than the fixed-effect model.

1.3 Compute the effect size of each study

First, the scores presented in the articles were extracted and the ES of each study as defined in Morris [2008] was computed

$$\text{ES} = c_p \left[\frac{(M_{\text{post},T} - M_{\text{pre},T}) - (M_{\text{post},C} - M_{\text{pre},C})}{\sigma_{\text{pre}}} \right]. \tag{1}$$

An ES is exactly equivalent to a z-score of a standard normal distribution, it is computed as mean pre- to post-treatment score change in the Neurofeedback (NFB) group ($M_{\text{pre},T}$, $M_{\text{post},T}$) minus the mean pre- to post- treatment score change in the control group ($M_{\text{pre},C}$, $M_{\text{post},C}$), divided by the pooled pre-test standard deviation (σ_{pre})

$$\sigma_{\text{pre}} = \sqrt{\frac{(n_T - 1)\sigma^2_{\text{pre},T} + (n_C - 1)\sigma^2_{\text{pre},C}}{n_T + n_C - 2}}, \tag{2}$$

where $\sigma_{t,G}$ indicates the standard deviation for group G at time t and n_G defines the sample size of each group; c_p is a bias adjustment typically used for small sample sizes

$$c_p = 1 - \frac{3}{4(n_T + n_C - 2) - 1}. \tag{3}$$

1.4 Compute the variance of each effect size

Then, the variance of each ES was computed [Morris, 2008]

$$\sigma^2(\text{ES}) = c_p^2 \left(\frac{n_T + n_C - 2}{n_T + n_C - 4} \right) \left[\frac{2(1-\rho)(n_T + n_C)}{n_T n_C} + \text{ES}^2 \right] - \text{ES}^2. \tag{4}$$

To compute the variance of the ES, the pooled within-group Pearson correlation ρ (i.e the pre-post correlation) is required [James et al., 2013]:

$$\rho = \frac{\sum_{i=1}^{n} (\text{pre}_i - \mu_{\text{pre}})(\text{post}_i - \mu_{\text{post}})}{\sqrt{\sum_{i=1}^{n} (\text{pre}_i - \mu_{\text{pre}})^2} \sqrt{\sum_{i=1}^{n} (\text{post}_i - \mu_{\text{post}})^2}}, \tag{5}$$

where n is the number of patients included in a study, pre_i, post_i are score values for patient i at pre- and post-test respectively, and μ_{pre}, μ_{post} the mean scores over all patients. It is a measure of linear correlation between two variables. A value of 1 means that there is a positive correlation whereas a value of -1 means a negative correlation. When $\rho = 0$, there is no linear correlation. In our case, this correlation was not known and the raw data were not available so we took an approximation: Balk et al. [2012] found that a value of 0.5 yielded values close to those computed with the right value of the correlation.

Once variances were obtained with Eq 4, we could compute the standard error and the 95% confidence interval of each ES.

1.5 Compute the weight of each study

To compute the Summary Effect (SE) a weight must be assigned to each study. To obtain them several steps must be followed. At first, the fixed-effects model weight w_{fixed_k} of each study k was computed as defined in Borenstein et al. [2009]:

$$w_{\text{fixed}_k} = \frac{1}{\sigma^2(\text{ES}_k)}. \tag{6}$$

Nevertheless, we chose to use the random effects model, so the weights associated to this model are different. To compute them, the between-studies variance τ^2 is required. It was calculated in three steps described in Eq 7, Eq 8 and Eq 9 [Borenstein et al., 2009]:

$$Q = \sum_{k=1}^{K} (w_{\text{fixed}_k} \text{ES}_k^2), \tag{7}$$

$$C = \sum_{k=1}^{K} w_{\text{fixed}_k} - \frac{\sum_{k=1}^{K} (w_{\text{fixed}_k})^2}{\sum_{k=1}^{K} w_{\text{fixed}_k}}, \tag{8}$$

with K the total number of included studies, and

$$\tau^2 = \frac{Q - \mathrm{df}}{C}, \tag{9}$$

with $\mathrm{df} = K - 1$ the degrees of freedom.

The random-effects model takes into account the differences between the studies, so the weights are equal to the inverse of the addition between the within-study variance (the variance of the ES) and the between-studies variance:

$$w_k = \frac{1}{\sigma^2(\mathrm{ES}_k) + \tau^2}. \tag{10}$$

1.6 Compute the summary effect

Eventually, the weighted average of the K ES was computed to obtain the SE as described in Eq 11 [Borenstein et al., 2009]:

$$\mathrm{SE} = \frac{\sum_{k=1}^{K} w_k \mathrm{ES}_k}{\sum_{k=1}^{K} w_k}. \tag{11}$$

Once the SE is obtained, we can compute its variance, its standard error, its 95% confidence interval, its p-value, and I^2 estimating ES's between studies heterogeneity.

1.7 Scales used for the meta-analysis

All scales used for the meta-analysis are summarized here in order to facilitate the replication of this work Table 1.

1.8 Associate independent factors to effect sizes

Three different methods were used to perform the Systematic Analysis of Biases (SAOB):

- weighted multiple linear regression (Weighted Least Squares (WLS)) [Montgomery et al., 2012];
- sparsity-regularized linear regression with Least Absolute Shrinkage and Selection Operator (LASSO) [Tibshirani, 1996];
- decision tree [Quinlan, 1986].

We first applied the WLS as described in Eq 12:

$$\mathbf{W}y = \mathbf{WX}\beta + \epsilon. \tag{12}$$

$\mathbf{X}$ is a $(n \times p)$ full rank matrix and represents n observations on each $p-1$ independent variables and an intercept term, β is a $(p \times 1)$ vector of associated regression coefficients, $\mathbf{W}$ is a $(n \times n)$ diagonal matrix with weights, y is a $(n \times 1)$ vector of dependent variables and ϵ is a $(n \times 1)$ vector of errors.

The aim of the WLS is to estimate the vector of coefficients β by minimizing the Weighted Residual Sum of Squares (WRSS)

$$\text{WRSS} = \sum_{i=1}^{n} w_i \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2. \tag{13}$$

A significant coefficient (meaning significantly different from 0) indicates that the associated factor had probably an influence on NFB efficacy and the sign of the coefficient indicates the direction of the effect.

The second method applied was the LASSO that naturally incorporates variable selection in the linear model thanks to L1-norm applied on the coefficients. The coefficients $\hat{\beta}_j$ are obtained by minimizing the term

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \,, \tag{14}$$

where λ is the regularization parameter setting more coefficients to zero as it increases. The optimal tuning parameter was determined by a leave-one-out cross-validation. This method retains 1 observation as the validation data for testing the model and the remaining n - 1 observations are used as training data. The cross-validation process is then repeated n times with each of the observation used exactly once as the testing data. For each fold, the Mean Square Error (MSE) on the test set was computed and eventually, the n results can be averaged to produce a single observation that enables to find the optimal λ: it corresponds to the abscissa of the minimum value of the MSE on the mean fold computed for a large range of λ [James et al., 2013]. A coefficient not set at zero means that the associated factor may have an influence on NFB and once again, the sign of the coefficient indicates the direction of the effect.

Eventually, the last method used to determine factors influencing NFB was the decision tree [Quinlan, 1986], a non linear method. It brakes down a dataset into smaller and smaller subsets using at each iteration a variable and a threshold chosen to optimize a simple MSE criteria

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \Big(\hat{y}_i - y_i \Big)^2, \tag{15}$$

with $\hat{y}$ the predicted values.

No weights were applied when running the LASSO and the decision tree.

2 Results

2.1 Perform a meta-analysis

First, when using the ES found by Cortese et al. [2016] thanks to RevMan [Cochrane Collaboration, 2011], and then performed the following steps of meta-analysis with the Python code, we observed no major differences between these results and those obtained with RevMan [Cochrane Collaboration, 2011] as listed in 2. The minor discrepancies, especially observed at the p-values level, were due to our choice to always use a pre-post correlation of 0.5 when computing the variance of each ES. Moreover, a sensitivity analysis was conducted to ensure the minor impact of the pre-post correlation value: when it varied between 0.2 and 0.8 the significance of the SE did not change.

Thanks to the previous step, we could conclude that the Python code yielded results close to those returned by RevMan Cochrane Collaboration [2011], so we decided to use this code to perform our meta-analysis.

2.2 Detect factors influencing the Neurofeedback

To assess the variability of each factor, box plots of their standardized values were displayed in Figure 1: the variability in treatment and sessions length as well as the number of sessions is more important from one study to another than the number of sessions per week and the age bounds.

2.3 Assumptions for applying linear regression

The first method used to detect the influencing factors was the WLS. The assumptions inherent to this method were checked:

- the moment matrix $\mathbf{X}^T\mathbf{W}^T\mathbf{W}\mathbf{X}$ was invertible;
- no apparent correlation between the continuous independent variables was found;
- the fit was significant as shown by the F-statistic (prob(F-statistic) = 2.81e-10);
- the residuals were normally distributed as demonstrated by the skew (-0.15), kurtosis (2.81) and the Omnibus test (prob(Omnibus) = 0.87).

These assumptions were also satisfied for the Ordinary Least Squares (OLS).

References

L. E. Arnold, N. Lofthouse, S. Hersch, X. Pan, E. Hurt, B. Bates, K. Kassouf, S. Moone, and C. Grantier. Eeg neurofeedback for attention-deficit/hyperactivity disorder: Double-blind sham-controlled randomized pilot feasibility trial. *Journal of Atention Disorder*, 2014. doi: https://doi.org/10.1177/1087054712446173.

A. R. Bakhshayesh, S. Hänsch, A. Wyschkon, M. J. Rezai, and G. Esser. Neurofeedback in adhd: a single-blind randomized controlled trial. *European child & adolescent psychiatry*, 20(9):481, 2011. doi: https://doi.org/10.1007/s00787-011-0208-y.

E. M. Balk, A. Earley, K. Patel, T. A. Trikalinos, and I. J. Dahabreh. Empirical assessment of within-arm correlation imputation in trials of continuous outcomes. 2012.

S. Baumeister, I. Wolf, N. Holz, R. Boecker-Schlier, N. Adamo, M. Holtmann, M. Ruf, T. Banaschewski, S. Hohmann, and D. Brandeis. Neurofeedback training effects on inhibitory brain activation in adhd: A matter of learning? *Neuroscience*, 2016. doi: https://doi.org/10.1016/j.neuroscience.2016.09.025.

M. Beauregard and J. Levesque. Functional magnetic resonance imaging investigation of the effects of neurofeedback training on the neural bases of selective attention and response inhibition in children with attention-deficit/hyperactivity disorder. *Applied psychophysiology and biofeedback*, 31(1):3–20, 2006. doi: https://doi.org/10.1007/s10484-006-9001-y.

M. Bink, C. van Nieuwenhuizen, A. Popma, I. L. Bongers, and G. J. M. van Boxtel. Behavioral effects of neurofeedback in adolescents with adhd: a randomized controlled trial. *European Child and Adolescent Psychiatry*, 2014. doi: https://doi.org/10.1007/s00787-014-0655-3.

M. Borenstein, L. V. Hedges, J. Higgins, and H. R. Rothstein. *Introduction to meta-analysis*. Wiley, 2009.

A. Bussalb. Factors influencing nfb. `https://github.com/AuroreBussalb/meta-analysis-statistical-tools`, 2018.

H. Christiansen, V. Reh, M. H. Schmidt, and W. Rief. Slow cortical potential neurofeedback and self-management training in outpatient care for children with adhd: study protocol and first preliminary results of a randomized controlled trial. *Frontiers in human neuroscience*, 8, 2014. doi: https://doi.org/10.3389/fnhum.2014.00943.

Cochrane Collaboration. Revman 5.1. https://community.cochrane.org/help/tools-and-software/revman-5, 2011.

S. Cortese, M. Ferrin, D. Brandeis, M. Holtmann, P. Aggensteiner, D. Daley, P. Santosh, E. Simonoff, J. Stevenson, A. Stringaris, et al. Neurofeedback for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(6):444–455, 2016. doi: https://doi.org/10.1016/j.jaac.2016.03.007.

H. Gevensleben, B. Holl, B. Albrecht, C. Vogel, D. Schlamp, O. Kratz, P. Studer, A. Rothenberger, G. H. Moll, and H. Heinrich. Is neurofeedback an efficacious treatment for adhd? a randomised controlled clinical trial. *Journal of Child Psychology and Psychiatry*, 50(7):780–789, 2009. doi: https://doi.org/10.1111/j.1469-7610.2008.02033.x.

H. Heinrich, H. Gevensleben, F. J. Freisleder, G. H. Moll, and A. Rothenberger. Training of slow cortical potentials in attention-deficit/hyperactivity disorder: evidence for positive behavioral and neurophysiological effects. *Biological psychiatry*, 55(7):772–775, 2004. doi: https://doi.org/10.1016/j.biopsych.2003.11.013.

M. Holtmann, D. Grasmann, E. Cionek-Szpak, V. Hager, N. Panzner, A. Beyer, F. Poustka, and C. Stadler. Spezifische wirksamkeit von neurofeedback auf die impulsivität bei adhs. *Kindheit und Entwicklung*, 18(2):95–104, 2009. doi: https://doi.org/10.1026/0942-5403.18.2.95.

G. James, D. Witten, T. Hastie, and R. Tibshirani. An introduction to statistical learning. volume 112, chapter 8, pages 303–332. Springer, 2013.

M. Linden, T. Habib, and V. Radojevic. A controlled study of the effects of eeg biofeedback on cognition and behavior of children with attention deficit disorder and learning disabilities. *Applied Psychophysiology and Biofeedback*, 21(1):35–49, 1996.

S. Maurizio, M. D. Liechti, H. Heinrich, L. Jäncke, H.-C. Steinhausen, S. Walitza, D. Brandeis, and R. Drechsler. Comparing tomographic eeg neurofeedback and emg biofeedback in children with attention-deficit/hyperactivity disorder. *Biological psychology*, 95:31–44, 2014. doi: https://doi.org/10.1016/j.biopsycho.2013.10.008.

D. C. Montgomery, E. A. Peck, and G. G. Vining. Introduction to linear regression analysis. volume 821, chapter 5, pages 190–191. John Wiley & Sons, 2012.

S. B. Morris. Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2):364–386, 2008. doi: https://doi.org/10.1177/1094428106291059.

J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

N. J. Steiner, R. C. Sheldrick, D. Gotthelf, and E. C. Perrin. Computer-based attention training in the schools for children with attention deficit/hyperactivity disorder: a preliminary trial. *Clinical pediatrics*, 50(7):615–622, 2011. doi: https://doi.org/10.1177/0009922810397887.

N. J. Steiner, E. C. Frenette, K. M. Rene, R. T. Brennan, and E. C. Perrin. Neurofeedback and cognitive attention training for children with attention-deficit hyperactivity disorder in schools. *Journal of Developmental & Behavioral Pediatrics*, 35(1):18–27, 2014. doi: https://doi.org/10.1097/DBP.0000000000000009.

U. Strehl, P. Aggensteiner, D. Wachtlin, D. Brandeis, B. Albrecht, M. Arana, C. Bach, T. Banaschewski, T. Bogen, A. Flaig-Rohr, et al. Neurofeedback of slow cortical potentials in children with attention-deficit/hyperactivity disorder: A multicenter randomized trial controlling for unspecific effects. *Frontiers in human neuroscience*, 11, 2017. doi: https://doi.org/10.3389/fnhum.2017.00135.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

M. van Dongen-Boomsma, M. A. Vollebregt, D. Slaats-Willemse, and J. K. Buitelaar. A randomized placebo-controlled trial of electroencephalographic (eeg) neurofeedback in children with attention-deficit/hyperactivity disorder. *J Clin Psychiatry*, 74(8):821–827, 2013. doi: https://doi.org/10.4088/JCP.12m08321.

Figure captions

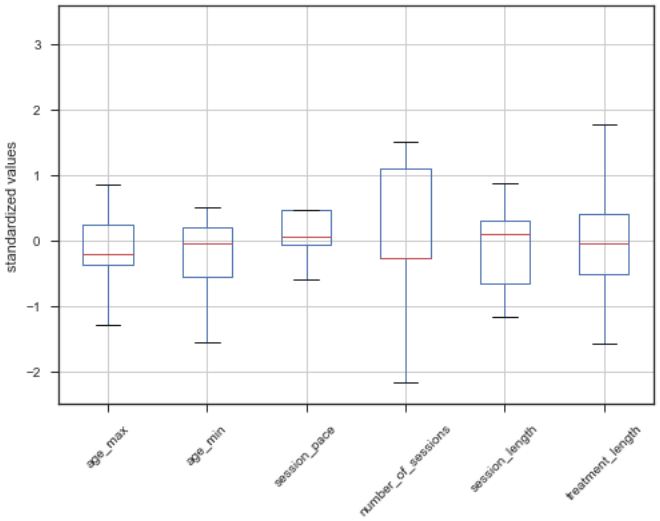

Figure 1: Boxplots of the standardized values of each continuous factor.

Table captions

Table 1: Clinical scales used to update Cortese et al. [2016] with our choices and the two new articles.

Study	Outcome	Score Names - Parents ratings	Score Names - Teachers ratings
Arnold et al.	Total	SNAP IV	SNAP IV
	Inattention	SNAP IV	SNAP IV
	Hyperactivity	SNAP IV	SNAP IV
Bakhshayesh et al.	Total	German ADHD-RS	German ADHD-RS
	Inattention	German ADHD-RS	German ADHD-RS
	Hyperactivity	German ADHD-RS	German ADHD-RS
Baumeister et al.	Total	DISYPS	-
Beauregard and Levesque	Total	CPRS	-
	Inattention	CPRS	-
	Hyperactivity	CPRS	-
Bink et al.	Total	ADHD-RS self report	-
	Inattention	ADHD-RS self report	-
	Hyperactivity	ADHD-RS self report	-
Christiansen et al.	Total	Conners-3 Parents	Conners-3 Teachers
Gevensleben et al.	Total	German ADHD-RS	German ADHD-RS
	Inattention	German ADHD-RS	German ADHD-RS
	Hyperactivity	German ADHD-RS	German ADHD-RS
Heinrich et al.	Total	German ADHD-RS	-
Holtmann et al.	Total	German ADHD-RS	-
	Inattention	German ADHD-RS	-
	Hyperactivity	German ADHD-RS	-
Linden et al.	Total	IOWA Conners	-
	Inattention	IOWA Conners	-
Maurizio et al.	Total	CPRS	CTRS
	Inattention	CPRS	CTRS
	Hyperactivity	CPRS	CTRS
Steiner et al.	Total	Conners Rating Scales Revised	Conners Rating Scales Revised
	Inattention	Conners Rating Scales Revised	Conners Rating Scales Revised
	Hyperactivity	Conners Rating Scales Revised	Conners Rating Scales Revised
Steiner et al.	Total	Conners-3 Parents	Conners-3 Teachers
	Inattention	Conners-3 Parents	Conners-3 Teachers
	Hyperactivity	Conners-3 Parents	Conners-3 Teachers
Strehl et al.	Total	German ADHD-RS	German ADHD-RS
	Inattention	German ADHD-RS	German ADHD-RS
	Hyperactivity	German ADHD-RS	German ADHD-RS
van Dongen-Boomsma et al.	Total	ADHD RS	ADHD RS
	Inattention	ADHD RS	ADHD RS
	Hyperactivity	ADHD RS	ADHD RS

SNAP: Wanson, Nolan and Pelham Questionnaire, ADHD-RS: ADHD Rating Scale, CPRS: Conners Parent Rating Scale, CTRS: Conners Teacher Rating Scale, BOSS Classroom Observation: Behavioral Observation of Students in Schools, DISYPS: Diagnostic System of Mental Disorders in Children and

Table 2: Comparison between Cortese et al. [2016] results obtained with RevMan [Cochrane Collaboration, 2011] and those obtained with the Python code. Summary effects and their corresponding p-value in parenthesis are presented. With the Python program, a negative summary effect is in favor of NFB.

Input data		Results from Cortese et al. [2016]	Effect sizes from Cortese et al. [2016]
Implementation		RevMan Cochrane Collaboration [2011]	Python program
Parents	Total	$0.35\ (0.004)$	$-0.34\ (0.004)$
	Inattention	$0.36\ (0.009)$	$-0.35\ (0.011)$
	Hyperactivity	$0.26\ (0.004)$	$-0.24\ (0.02)$
Teachers	Total	$0.15\ (0.20)$	$-0.13\ (0.25)$
	Inattention	$0.06\ (0.70)$	$-0.09\ (0.50)$
	Hyperactivity	$0.17\ (0.13)$	$-0.15\ (0.21)$