# An updated meta-analysis on the effectiveness of Neurofeedback in children with ADHD: design choices that really matter

September 17, 2018

Aurore Bussalb[a], PhD student, Marco Congedo[b], PhD, Quentin Barthélemy[a], PhD, David Ojeda[a], PhD, Eric Acquaviva[c], MD, PhD, Richard Delorme[c], MD, PhD, Louis Mayaud[a], D.Phil.

[a]: Mensia Technolgies, Paris, France

[b]: GIPSA-Lab, CNRS, University Grenoble Alpes, National Polytechnic Institute, Grenoble, France

[c]: Hôpital Robert Debré, Paris, France

**Full postal address**

Mensia Technologies

Plateforme d'innovation Boucicaut

130 rue de Lourmel

75015 Paris - France

Aurore Bussalb's e-mail: aurore.bussalb@mensiatech.com

Louis Mayaud's e-mail: lm@mensiatech.com

A. Bussalb, Q. Barthélemy, D. Ojeda, and L. Mayaud work for Mensia Technologies. M. Congedo served as an advisor for Mensia Technologies when this work was conducted.

This research was conducted as part of a PhD thesis funded by Mensia Technologies and the Association Nationale Recherche et Technologie (ANRT) as well as the EU H2020 NEWROFEED grant 684809.

Number of words: 5369

Number of figures: 4

**Abstract**

Numerous trials and several meta-analysis have been published on the efficacy of the Neurofeedback (NFB) treatment for Attention Deficit Hyperactivity Disorder (ADHD) in children and adolescents, showing inconsistent findings. This work replicates and extents the last meta-analysis published on this subject [Cortese et al., 2016] by adding two randomized control trials (RCTs) that have been published since. We also perform a systematic analysis of biases (SAOB), which takes advantage of the technical and methodological heterogeneity of the studies included in the meta-analyses rather than suffering from it. The SAOB was performed on k = 31 trials meeting the same inclusion criteria as the earlier meta-analysis (with the exception of the requirement for a control arm). Our extended meta-analysis confirmed the results previously obtained: effect sizes in favor of NFB efficacy were significant when clinical scales of ADHD were rated by parents (non-blind, p-value = 0.0017), but not when rated by teachers (considered as probably blind, p-value = 0.14). The effect size was significant according to both raters for the subset of studies meeting the definition of "standard NFB protocols" (parents p-value = 0.0054; teachers p-value = 0.043, k = 4 studies). The SAOB identified three main factors that have an impact on NFB efficacy: first, a more intensive treatment, but not treatment duration, was associated with higher efficacy; second, high-quality EEG systems improved the effectiveness of the NFB treatment; third, teachers reported a lower improvement as compared to parents. In addition, to all raw data we used (31 studies), we release a complete Python library for performing meta-analysis, for easing the replication of this and previous studies as well as for future projects. In conclusion, in addition to extending previous findings, we introduced here a new way to look into the heterogeneity of clinical trials.

Keywords: ADHD, neurofeedback, meta-analysis, analysis of bias

# 1 Introduction

Attention Deficit/Hyperactivity Disorder (ADHD) is a common child psychiatric disorder characterized by impaired attention and/or hyperactivity/impulsivity. Symptoms may persist in adulthood with clinical significance, which makes ADHD a life-long problem for many patients [Faraone et al., 2006]. The prevalence of ADHD is about 5% in school-aged children yielding to an estimated 2.5 millions of children in Europe [The American Psychiatric Association, 2013]. ADHD impacts children's well-being with many of them suffering from low self esteem [Shaw-Zirt et al., 2005] and underachieving at school [Barry et al., 2002]. Parents are equally affected by this situation since children's behavior is commonly attributed to bad parenting [Harpin, 2005]. From a societal point

of view, ADHD has a high financial impact: a survey of 2013 in Europe estimated at between $9,860$ and $14,483$ Euros per patient/year the cost related to ADHD [Le et al., 2014].

The diagnosis of ADHD primarily relies on questionnaire-based clinical evaluation [The American Psychiatric Association, 2013], which can be supported with objective assessment metrics of executive function such as the Test of Variables of Attention (TOVA) [Forbes, 1998], the Continuous Performance Test (CPT) [Barkley, 1991], and the Sustained Attention to Response Task (SART) [Robertson et al., 1997]. Objective markers of brain function using Electroencephalogram (EEG), functional Magnetic Resonance Imaging (fMRI), or Positron Emission Tomography (PET) are not considered useful to improve diagnosis at the individual level [Neba Health, 2015], but can help differentiating groups of patients [Johnstone et al., 2005]. In particular, different phenotypes of ADHD patients present with an increase in the EEG theta waves power (4-8Hz) and/or a decrease of EEG beta waves power (12-32Hz) in frontal areas, or a decrease in the EEG Sensorimotor Rhythm (SMR) power (13-15Hz) in the central area [Monastra, 2005; Matoušek et al., 1984; Janzen et al., 1995; Loo et al., 2017].

Among all existing treatments, the most widely used is psychostimulants, which have been proven to be efficacious [Taylor, 2014; Storebo et al., 2015]. However, their long-term effectiveness and side effects are still debated and form an active area of research [DuPaul, 1998; Swanson et al., 2001; Jensen, 1999]. Moreover, ADHD children under medication commonly suffer from mild side effects such as loss of appetite and sleep disturbance, however serious adverse events are rare [Storebo et al., 2015; Cooper et al., 2011]. These drawbacks make some parents and clinicians reluctant to opt for such treatment, turning them to non-pharmaceutical alternatives such as dietary changes [Bélanger et al., 2009] and behavioral therapy, which have been proven to be less efficacious [Sonuga-Barke et al., 2013].

Neurofeedback (NFB) is a noninvasive technique aiming at the reduction of ADHD symptoms [Arns et al., 2015; Steffert and Steffert, 2010; Marzbani et al., 2016]. It is a self-paced brain neuromodulation technique that represents one's brain activity in real-time using auditory or visual modulations, on which learning paradigms, such as operant conditioning [Reynolds, 1975] or voluntary control, can be applied. To deliver this intervention, neurophysiological time series analyzed in real-time so as to be incorporated in feedback applications such as serious games leveraging learning paradigms [Wang et al., 2010]. These data represent the activity of a population of neurons involved in attentional networks, which is translated into visual or auditory cues. The sensory feedback constitutes the rewards mechanism promoting learning using operant conditioning protocol [Sherlin et al., 2011]. Operant conditioning enables neural plasticity supporting the child in the task repetition [Skinner, 1984], which is supposed to result in long-lasting neuronal reorganization [Van Doren

et al., 2017].

Several NFB protocols have been proposed and investigated to decrease the symptoms of ADHD:

- protocols based on neural oscillations, using frequency-band power training: enhance SMR while suppressing theta power, or enhance beta while suppressing theta (this is known as the Theta Beta Ratio (TBR) protocol) [Lubar and Shouse, 1976; Arns et al., 2013];
- the protocol based on the Slow Cortical Potentials (SCPs) training consisting in the regulation of cortical excitation thresholds by focusing on activity generated by external cues (similar to Event-Related Potentials (ERPs)) [Heinrich et al., 2004; Banaschewski and Brandeis, 2007];
- the protocol to enhance ERPs (P300) [Fouillen et al., 2017]: P300 amplitude can be considered as a specific neurophysiological marker of selective attention [Bouëdec, 2017].

Shortly after the discovery of the brain's electric activity by Berger [1929], Durup and Fessard [1935] proved it could be voluntarily modulated leading to a series of finding on the self-regulation of brain activity. The first indication of the therapeutic potential of brain activity operant-conditioning came forty years later when Sterman et al. [1974] serendipitously found that training the SMR activity reduces the incidence of epileptic crisis in kerozen-exposed cats. The technique, then known as NFB, quickly became investigated in various fields of neuropsychiatry including, most notably, ADHD [Lubar and Shouse, 1976; Rossiter and La Vaque, 1995; Linden et al., 1996; Maurizio et al., 2014]. Subsequently, its efficacy on the core symptoms of ADHD (inattention, hyperactivity, and impulsivity) has been the subject of several meta-analytic studies [Loo and Barkley, 2005; Lofthouse et al., 2012; Arns et al., 2009; Micoulaud-Franchi et al., 2014; Sonuga-Barke et al., 2013].

The last meta-analysis addressing the efficacy of NFB has been published by Cortese et al. [2016]. A total of 13 studies were included. The authors of this meta-analysis have made some choices that have since been debated by the community. Specifically, Micoulaud-Franchi et al. [2016] criticized the use of an uncommon behavioral scale provided by Steiner et al. [2014] for the teachers' assessments and the inclusion of a pilot study carried out by Arnold et al. [2014].

Finally, because of the publication of new Randomized Controlled Trials (RCTs) meeting Cortese et al.'s inclusion criteria, we decided to update the meta-analysis and take the opportunity to investigate the impact of the controversial choices. We have extended the analysis with a novel method, the Systematic Analysis of Biases (SAOB), which takes advantage of the technical and methodological heterogeneity of the trials included in the meta-analysis rather than suffering from it. Indeed, the NFB domain is characterized by a clinical literature that is tremendously heterogeneous: studies differ methodologically (random assignment and presence of a blind assessment for instance),

in the NFB implementation (number of sessions, session and treatment length, and type of protocol for instance) as well as on the acquisition and processing of the EEG signal. Since methodological and technical implementations of studies are very likely to influence their outcomes [Congedo et al., 2004], we suggest here to identify which of the factors independently influence the clinical efficacy with the use of appropriate statistical tools.

# 2 Materials and Methods

## 2.1 Inclusion criteria

Search terms were directly taken from Cortese et al. [2016] to the exception of the need for a control arm, which is detailed in Supplemental Materials [Bussalb, 2018b]. The requirements included:

- studies have to assess NFB efficacy;
- subjects must have received a diagnosis of ADHD based on DSM-IV [The American Psychiatric Association, 2000], DSM-V [The American Psychiatric Association, 2013], ICD-10 [World Health Organization, 1993] criteria or by an expert psychiatrist;
- studies have to be written in English, German, Spanish, or French;
- studies have to include at least eight subjects in each group;
- patients must be younger than 25 years old;
- the publication (or subsequently their corresponding author) have to disclose sufficient details on the data to compute required metrics for the ensuing analysis.

The studies satisfying all these points were included in the SAOB. In order to replicate and update Cortese et al.'s meta-analysis, we applied the original inclusion criteria of their meta-analysis to our search (the main difference being the presence of a control group).

## 2.2 Outcome definition

In the included studies, the severity of ADHD symptoms have been assessed by parents and, whenever available, by teachers. Cortese et al. [2016] and Micoulaud-Franchi et al. [2014] defined parents as Most Proximal (MPROX) raters who were not blind to the treatment, as opposed to teachers who were considered as Probably Blind (PBLIND) raters. This distinction is meant to assess the amplitude of the placebo effect, where it is hypothesized that teachers, who are presumed more

blind to the intervention, are less influenced in their assessment. Efficacy of NFB was measured with clinical scales, such as the ADHD-RS [Pappas, 2006], on the following outcomes: inattention, hyperactivity/impulsivity, and total scores. The factor analysis was performed using the total score.

## 2.3 Meta-analysis

The goal of a meta-analysis is to aggregate results from different clinical investigations and offer a consolidated state of evidence. To do so, it is necessary to assume some level of homogeneity in the design of the studies: inclusion criteria, and the presence and type of control (active, semi-active, or non-active). Because studies occasionally use slight variations of a clinical scales and because of the clinical heterogeneity of patients and control, the scores are standardized before being pooled into a Summary Effect (SE). The between-Effect Size (ES) is one of such standardized metrics, which we have implemented in this paper (see Supplemental Materials [Bussalb, 2018b]).

The meta-analysis was performed with a Python package developed for this work. The package offers a transparent approach to the choice of parameters, in an effort to ease replicability. We have benchmarked it against RevMan version 5.1 [Cochrane Collaboration, 2011, UK, London] by replicating Cortese et al. [2016]'s work. The code is made fully available on a GitHub repository [Bussalb, 2018a], including all raw data we have used in the present study so as to allow peers to review its implementation, update it, or use it for different projects.

Before updating the Cortese et al. [2016]'s work with recently published studies [Strehl et al., 2017; Baumeister et al., 2016], we decided to run a sensitivity analysis investigating the choices that later proved controversial [Micoulaud-Franchi et al., 2016]. Altogether, the changes investigated included:

- the ES of Arnold et al.'s study was computed from the post-test clinical values taken after the completion of the 40 sessions, in contrast to Cortese et al. [2016]'s report which used the results after only 12 sessions because the end point values were not available at the time of his study;

- the ES computed from the teachers' assessment reported by Steiner et al. [2014] relied on the BOSS Classroom Observation [Shapiro, 2010]. This is an atypical scale to quantify ADHD symptoms since the Conners Rating Scale Revised [Conners et al., 1998; Christiansen et al., 2014; Bluschke et al., 2016], a well-defined [Collett et al., 2003; Epstein and Weiss, 2012], and broadly used metric, was available in this study. Thus, we decided to compute the ES based on the Conners-3, already used in this study to compute the parents' ES.

We also performed two subgroups analysis with the two choices described above: first, including only studies following standard protocol as defined by Arns et al. [2014] and then with studies whose participants took low-dose or no medication during the trial.

## 2.4 Identify factors influencing the Neurofeedback

While revisiting the existing meta-analyses, it became apparent to us that the studies pooled together were highly heterogeneous in terms of methodological and practical implementation. For instance, all NFB interventions were pooled together regardless the quality of the acquisition, the quality of EEG data, and the trained neuromarker. Likewise, the methodological implementations varied significantly, requiring the 'subgroup' analysis (gathering studies following standard protocols for instance) that are somewhat arbitrary. To circumvent these limitations, we implemented a novel approach: the SAOB. With this method, the within-ES of each intervention was considered as a dependent variable to be explained by methodological and technical factors. The results of such analysis should enable to identify known methodological biases (e.g. blind assessments negatively associated with ES) and possibly technical factors (e.g. a good control on real time data quality influences positively the treatment outcome).

### 2.4.1 Identify and pre-process factors

We classified the factors influencing the efficacy of NFB in five categories: methodological, technical, demographics, and quality of the signal acquisition. Factors were chosen based on what has been reported in the literature, presumed to influence ES, and categorized as follow:

- *the signal quality*: correction of ocular and generic (amplitude based) artifacts;
- *the population*: intake of psychostimulants during NFB treatment and the age range of children included;
- *the methodological biases*: the presence of a control group, the blindness of assessors, the randomization of subjects in controlled trials, and the approval of the study by an Institutional Review Board (IRB);
- *the NFB implementation*: the protocol used (SCP, SMR, theta up, beta up in central areas, theta down), the presence of a transfer phase during NFB training, the possibility to train at home or at school with a transfer card, the type of thresholding for discrete reward, the number of NFB sessions, the length and frequency of the sessions, and the length of the treatment;

- *the acquisition quality*: the presence of one or more active electrodes and the EEG data quality. This latter was coded as an indicator between 1 and 3, using the following criteria:

  ***the type of electrodes used***: Silver Chloride (AgCl)/Gel or Gold (Au)/Gel;

  ***the use of impedance mode***: a quality check of electrode contacts ensuring an inter-electrode impedance smaller than $40k\Omega$;

  ***the level of hardware certificate***: compliance with ISO-60601-2-26 [International Electrotechnical Commission].

  A quality score equal to 3 was assigned if all the above criteria were satisfied. If at least one was satisfied the quality score was set to 2, otherwise the score was set to 1.

We provide in a Github repository [Bussalb, 2018a] the raw data extracted from the publications. To prevent any bias in the analysis, the name of the factors was hidden during the entire analysis so that the data scientists (AB, QB, DO, and LM) were fully blind to them. The names were revealed only once the data analysis and results were accepted as valid: this included choice of variable normalization and validation of model hypothesis as detailed below.

The pre-processing of factors for the analysis included the following steps: factors for which there were too many missing observations, arbitrarily set to more than 20% of the total of observations, were removed from the analysis. Furthermore, if a factor had more than 80% similar observations it was removed as well. Categorical variables were coded as dummies meaning that the presence of the factor was represented with 1 and its absence 0. All variables were standardized by subtracting the mean and then dividing by the standard deviation (not applied before the decision tree).

### 2.4.2 Explaining effect sizes with factors

To compute the within-ES, the means of total ADHD scores given by parents and teachers were used. Besides, in case studies providing results for more than one behavioral scale, the ES scores were computed for each one as

$$\mathrm{ES} = \frac{M_{\mathrm{post,T}} - M_{\mathrm{pre,T}}}{\sqrt{\frac{\sigma^2_{\mathrm{pre,T}} + \sigma^2_{\mathrm{post,T}}}{2}}},$$

where $M_{\mathrm{t,T}}$ is the mean of clinical scale, for treatment T, taken at time t (pre-test or post-test) and $\sigma$ represents its standard deviation.

With this definition, we focused on the effect of the treatment within a group [Cohen, 1988] as commonly reported in the literature [Arns et al., 2009; Maurizio et al., 2014; Strehl et al., 2017]. This ES enables to quantify the efficacy of NFB inside the treatment group.

The ES was then considered as a dependent variable to be explained by the factors (the independent variables). The following three methods, implemented with the Scikit-Learn Python [Pedregosa et al., 2011, version 0.18.1] and the Statsmodels Python [Seabold and Perktold, 2010, version 0.8.0] libraries, were used to perform the regression:

- 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].

The aim of the linear regression is to estimate the regression coefficients linking the factors to the ES. A significant coefficient (here and hereafter meaning significantly different from zero) indicates that the associated factor has an influence on NFB efficacy and its sign the direction of the effect.

The WLS differs from a traditional linear regression estimated with Ordinary Least Squares (OLS) in that a weight is assigned to each observation so as to account for the multiplicity of reported clinical endpoints in some studies. Besides, the weight was also set as a function of the sample size to account for variations in sample sizes. Specifically, the weight of each study was taken as the ratio between the experiment group's sample size and the number of behavioral scales available. We also ran the analysis with OLS method to assess the impact of the weights on the results.

The second linear method applied was the LASSO, which naturally incorporates variable selection to the linear model thanks to L1-norm applied on the coefficients. A coefficient not set to zero means that the associated factor has an influence on NFB efficacy and its sign indicates the direction of the effect.

The last method used to determine factors influencing NFB is the decision tree [Quinlan, 1986], a hierarchical and non-linear method. It breaks down a dataset into smaller and smaller subsets using, at each iteration, a variable and a threshold chosen to optimize a simple Mean Square Error (MSE) criterion [James et al., 2013]. A tree is composed of several nodes and leafs, the importance of which is decreasing from the top node, called the root node, downward.

These methods are intrinsically different from each others, so we compared their results. For instance, the decision tree captures variable interactions and can relate factors to ES in a non-linear fashion. On the other hand, the LASSO offers an elegant mathematical framework to variable selection. Further details are given in the Supplemental Material [Bussalb, 2018b].

# 3 Results

## 3.1 Selected studies

Search terms entered in Pubmed returned 152 results during the last check on February 12, 2018, including 28 articles used in previous meta-analyses on NFB. After the selection process illustrated in Figure 1, 31 studies were included in the SAOB and 15 in the meta-analysis as summarized in Table 1. The 31 studies selected for the SAOB followed the Cortese et al.'s criteria, with the exception of the requirement for a control group. Indeed, since within-ES were considered in this analysis, a control group was not required.

## 3.2 Meta-analysis

The Python module developed and used for this work was validated (details available in Supplemental Materials) and the code was made available online [Bussalb, 2018a].

The replication and the update of Cortese et al.'s study was conducted by applying the choices described in the Materials and Methods section and the results obtained are presented in Table 2:

- when computing the ES for Arnold et al. [2014] with the values after 40 sessions of NFB, smaller ES were found than Cortese et al. [2016]. This was unexpected as the clinical efficacy might be supposed to increase with the number of NFB sessions;
- when relying on the teachers' ratings from the Conners-3 to compute the ES of Steiner et al. [2014], higher SE were found in attention but not for total and hyperactivity score. However, this different choice of scale did not affect the statistical significance of the SEs.

The meta-analysis was then extended by adding two new articles [Strehl et al., 2017; Baumeister et al., 2016] found by applying the same search criteria and finding more recent matches. Baumeister et al. [2016] provided results only for parents total outcome whereas Strehl et al. [2017] gave teachers and parents' assessments for all outcomes. Despite favorable results for NFB, particularly on parents' assessments, adding these two new studies did not change neither the magnitude nor the significance of the SE, for any outcome whatever the raters, as illustrated in Figure 2.

As initially suggested by Cortese et al., the analysis was ran on two subgroups of studies: one gathering studies following the standard protocol defined by Arns et al. [2014] and a second including only participants not taking medications during the clinical trial.

Regarding the 'standard protocol' subgroup, Cortese et al. [2016] found all the outcomes significant except for the hyperactivity symptoms rated by teachers, which only showed a statistical trend

(p-value = 0.11). Similar results were obtained when adding the most recent studies meeting this definition [Strehl et al., 2017] (p-value = 0.11). The SE for the total outcome assessed by teachers remained significant with the addition of the two new RCTs (p-value = 0.043), giving more power to this result since it is now based on four studies including 283 patients in total.

As for the no-drug subgroup, SEs were found significant for the inattention symptoms assessed by parents (p-value = 0.013). Besides, the differences in Arnold et al. [2014] values caused a loss of significance in hyperactivity outcome for parents (p-value = 0.066) compared to Cortese et al. [2016] (p-value = 0.016). The two new studies were not included in this subgroup because subjects were taking psychostimulants during the trial.

All the clinical scales used to compute the ES following our choices are summarized in the Supplemental Materials.

## 3.3 Factors influencing Neurofeedback

This analysis was performed on 31 trials assessing the efficacy of NFB as presented in Table 1. Among the 25 factors selected, six were removed because there were too many missing observations or because they were too homogeneous: beta up in frontal areas, the use of a transfer card, the type of threshold for the rewards (incremental or fixed), the EEG quality equal of 3 (meaning that the electrodes used were either AgCl/Gel or Au/Gel, a check of electrode contact quality was performed, and the hardware used was CE marked), and the presence of a control group.

All results are presented in Table 3. These results require a careful interpretation since each technique provided with slightly different results. These differences may depend on the different assumptions of the model and several other factors. In any case, we are inclined to trust those findings that are consistent across methods.

The WLS technique identified eight significant factors for an adjusted R-squared of 0.74. When applying the OLS, the same factors were significant, except the presence of a transfer phase, the protocol theta down, and the artifact correction based on amplitude, with a lower adjusted R-squared (0.42). The LASSO regression selected 12 significant factors. With these methods, a negative coefficient means that the factor was in favor of the efficacy of NFB. The decision tree is presented in Figure 3. The best predictor is at the top of the tree: in our case it was the PBLIND. Five other factors also split the subsets, however, the lower we get into the tree, the less samples are available, making the interpretation more and more doubtful.

Several factors were common to the three methods used. In particular; the treatment length, the assessment by a blind rater, and an EEG quality score equal to 2 (meaning that at least one of

the three quality criteria had to be fulfilled). The methods also agreed on the direction of the effect for these factors: a shorter treatment and recording the EEG with a good-quality system seems preferable, whereas teachers' assessment appears less favorable as compared to parents' assessment.

It is more doubtful the influence of the factors returned by only one or two methods. In particular: both WLS and LASSO found that relying on the amplitude of the signal to correct artifacts and including a transfer phase seems not to improve ADHD symptoms. Conversely, the IRB approval, a theta down protocol, and a higher number of sessions per week appears to positively influence the efficacy of the NFB treatment. The decision tree and LASSO had in common the protocol SMR: it was associated with lower ES. Five factors were returned by only one of the methods: the minimal age of the children, being on drugs during NFB treatment, randomizing the groups and the SCPs protocol.

Five factors were never selected by the three methods: the correction for the ocular artifact, the children maximum age, the number of sessions, the protocol beta up in central areas and the presence of more than one active electrode. Thus, these factors definitely appear not to have an influence on NFB efficacy.

In the next section we discuss only the factors that were selected by at least two of the three methods.

# 4 Discussion

## 4.1 Meta-analysis

In the meta-analysis performed here, we challenged some choices made by Cortese et al., which proved controversial: the computation of ES based on an unusual scale [Steiner et al., 2014] and the inclusion of a pilot study [Arnold et al., 2014] whose end point values were not available at the time Cortese et al. conducted his meta-analysis. We review here the list of changes, their justification, and their impact on the analysis.

First, relying on the Conners-3 [Conners et al., 2011] instead of the BOSS Classroom Observation [Shapiro, 2010] for teachers ratings seems preferable because this scale is more commonly used [Christiansen et al., 2014; Bluschke et al., 2016] and is the revision of the Conners Rating Scale Revised [Conners et al., 1998] whose reliability has been studied [Collett et al., 2003]. However, relying on one or the other scale did not change the significance of the ES, regardless the outcome.

Second, to compute the ES of Arnold et al. [2014] the clinical scores taken when all sessions were completed were used instead of looking at interim results as in Cortese et al.. Some studies

suggested that the number of sessions correlates positively with the changes observed in the EEG [Vernon et al., 2004] so that a lower number of sessions would lead to artificially smaller ES. Here, the ES computed with the values at post test of Arnold et al. [2014] were smaller than those obtained after 12 sessions but these differences did not lead to a change of significance of the SE.

To conclude on that meta-analysis, although some points from the original were controversial, the impact on the meta-analysis were minimal and did not change the statistical significance of any outcome. The addition of the two new studies [Strehl et al., 2017; Baumeister et al., 2016] further confirmed original results. Indeed, the significance did not change for any outcome: the SE remained significant for MPROX raters and non-significant for PBLIND. Adding two more studies increased the significance of the sensitivity analysis ran by Cortese et al., most notably, the SE of studies corresponding to NFB "'standard protocols"' [Arns et al., 2014]. While Cortese et al. found that this subset tended to perform better, particularly on the PBLIND outcome, adding two studies extended this result on the total clinical score as well (p-value $< 0.05$). Despite the obvious heterogeneity of the studies included in this subset (particularly in terms of protocol used), these results suggest a positive relation between the features of this *standard* design and NFB performance.

This replication and update of a meta-analysis did not meet all PRISMA recommendations [Moher et al., 2009]. In particular, the risk of bias in individual and across studies was not assessed.

## 4.2 Factors influencing Neurofeedback

Description and analysis of different types of NFB implementation was subject to several studies [Arns et al., 2014; Enriquez-Geppert et al., 2017; Vernon et al., 2004; Jeunet et al., 2018] but to our knowledge none used statistical tools to quantify their influence on clinical endpoints.

Surprisingly, the number of sessions factor was not found significant by any method, which was somewhat in contradiction with existing literature. For instance, Arns et al. [2014] stated that performing less than 20 NFB sessions led to smaller effects. Similarly, Vernon et al. [2004] observed that positive changes in the EEG and behavioral performance occurred after a minimum of 20 sessions. However, Enriquez-Geppert et al. [2017] insisted on the fact that the number of sessions should be chosen carefully in order to avoid "overtraining". The fact that the number of sessions was not identified as a positively contributing factor, might be explained by the presence of only one data point with 20 sessions or less. Possibly, the temporal threshold of efficacy was passed for all included studies making the identification of this factor unlikely on this dataset. However, regardless of its statistical significance, the coefficient found by the WLS was negative, meaning that, as expected, the more sessions are performed, the more efficient the NFB tends to be.

Interestingly, [Minder et al., 2018] suggests that the subject location of the NFB training may also be an important contributing factor to clinical effectiveness. However this has been challenged by a recent study [Minder et al., 2018] showing that performing NFB at school or at the clinic has no significant impact on treatment response.

The type of NFB protocol was not identified by all the three methods, but it seemed to influence the NFB results according to two methods. In particular, the theta down protocol appeared more efficient than the SMR protocol. This importance granted by the methods to the NFB protocols was somewhat lower to expectations given the centrality of the protocols in the neurophysiological mode of action and subsequent expected impact on therapeutic effectiveness [Vernon et al., 2004]. A possible explanation for this result is that these protocols were equally efficacious to the populations they were offered to and thereby did not constitute a significant explanatory factor. This result, however, does not preclude a combined and personalized strategy (offering personalized protocols based on phenotypes) to further improve performance, as previously suggested by Alkoby et al. [2017].

Several factors were selected by all three methods with the same direction of influence: the EEG quality, the treatment length, and the rater's probably blindness to the treatment. First, our analysis pointed out the fact that recording EEG in good conditions leads to better results. This can be explained by the fact that better signal quality enables to extract the EEG patterns linked to ADHD more correctly and henceforth leads to better learning and therapeutic efficacy [Congedo et al., 2004]. However, it remains difficult to assess the quality of EEG hardware (such as the amplifier used) because little information is provided in these studies. This calls for a greater care in the future studies, which should strive and to assess and report the quality of the data.

Next, it appears here that the longer the treatment, the less efficient it becomes. Arguably, the treatment length is a proxy for treatment intensity, suggesting that a shorter treatment is more likely to succeed because the frequency of the sessions is higher. This hypothesis was back-up by the fact that the variable *session pace* (number of sessions per week) is also associated with larger ES according to the WLS and LASSO. Impact of the intensity of treatment have been investigated by [Rogala et al., 2016] on healthy adults: it was observed that studies with at least four training sessions completed on consecutive days were all successful. Overall these results suggest to adopt a high session pace, which is not common knowledge in the domain.

As expected, the assessment of symptoms by non-blind raters leads to far more favorable results than by PBLIND raters - result widely expected and in close compliance with existing meta-analyses [Cortese et al., 2016; Micoulaud-Franchi et al., 2014]. This last point was investigated more precisely to determine if this observation could solely be explained by the placebo effect.

## 4.3 Analysis on the probably blind raters

Teachers were considered as PBLIND raters by Cortese et al. and Micoulaud-Franchi et al.. Unexpectedly, the data provided did not exactly matched the widely accepted hypothesis stating that the difference between MPROX and PBLIND can solely be explained by the placebo effect. Nonetheless, the stress put on *probably* indicated that teachers may be aware of the treatment followed. An element that corroborates this hypothesis is the fact that for all the studies included in this work the amplitude of the clinical scale at baseline suggests that teachers did not capture the full extent of the symptoms or, said differently, that they were more blind to the symptoms than to the intervention as illustrated in Figure 4. The expected differences of ratings between teachers and parents have been extensively studied [Sollie et al., 2013; Narad et al., 2015; Minder et al., 2018], noting that teachers are more likely to underrate a child's symptom severity, especially so for younger children. As a consequence, teachers might just be less likely to observe a clinical change over the course of the treatment [Sollie et al., 2013; Narad et al., 2015; Minder et al., 2018]. Besides, it is also clear that there is more variability in teachers' scores compared to parents', which could partly explain the lower ES obtained for PBLIND raters, since the variability deflates the ES. In conclusion, using PBLIND as an estimate for correcting the placebo effect does not appear an adequate choice.

Another way to highlight a possible placebo effect is to focus on the decision tree illustrated in Figure 3, whose results provide a good insight to comment on this. The top node splits: on one hand 43 observations corresponding to MPROX raters and, on the other hand, 19 observations corresponding to PBLIND. If the differences observed between PBLIND and MPROX raters were due to the placebo effect, one would expect to find in the MPROX sub-tree some factors linked to the perception of the implication in the treatment. It was, indeed the case: session and treatment length were found significant but not in the direction corroborating the hypothesis that they are as part of a placebo effect. Indeed, one would expect that the longer the session and the treatment, the higher the placebo effect and the larger the within-ES. Instead, the opposite was found, somewhat invalidating the hypothesis.

These results altogether suggest that PBLIND assessments could hardly be used to assess placebo effect as they seem to be blinder to symptoms than to intervention. In the absence of ethically [Holtmann et al., 2014] and technically [Birbaumer, 1991] feasible sham for NFB protocols [World Medical Association, 2000], it is necessary to fall back on acceptable methodological alternative for the demonstration of clinical effectiveness. Among those are the analysis of neuromarkers collected during NFB treatment demonstrating that patients do *control* the trained neuromarkers; that they *learn* (reinforce control over time), and that these possibly lead to lasting brain reorganization (e.g.,

changes in their baseline resting state activity). The specificity of these changes, in relation to, which neuromarkers were trained and to the clinical improvement will be an essential component of this demonstration.

# 5 Conclusion

This work confirms Cortese et al. [2016]'s findings in the light of recently published clinical works. In particular, studies following a standard protocol as defined by Arns et al. [2014] show significant clinical improvements on PBLIND raters ($k = 4$ studies instead of 3 Cortese et al. [2016]).

Besides a meta-analysis, a new method is suggested here to tackle the high heterogeneity of clinical data available on NFB. This method aims at identifying factors that are positively or negatively contributing to NFB efficacy. Three factors were consistently found to explain clinical within-ES. First, the quality of acquisition of the EEG was positively correlated with clinical efficacy. This supports a mode of action through specific EEG training. Likewise, treatment intensity was found to contribute, corroborating what is known from learning theory (memory consolidation) [Mowrer, 1960], that is, a more intense treatment leads to an increased clinical efficacy. Finally, results show that the therapeutic efficacy measured by teachers is reduced compared to that measured by parents. This result has long been documented and it is widely advanced that this difference is solely imputable to the amplitude of placebo effect in NFB. However, the data presented in this article, in line with the most recent work on the topic [Sollie et al., 2013; Narad et al., 2015; Minder et al., 2018] tend to refute this hypothesis and suggests instead that teachers are simply less likely to be exposed to symptoms. As a consequence, using PBLIND endpoints to address the specificity of the clinical efficacy is not recommended and we advice instead to rely on other available methodological tools. Those include sham NFB and neuromarker analysis investigating the specificity of the EEG changes with respect to trained neuromarkers as well as changes in clinical endpoints.

These elements converge to the conclusion that existing methodologies, in particular the double blind design, traditionally used to assess pharmacological treatments in neuropsychiatric conditions may not be fully fitted to the evaluation of medical devices. The series of results presented here, however, suggest the presence of a genuine signal in favor of the therapeutic efficacy of NFB. A signal that should nonetheless be studied further using the aforementioned methodological tools, neuromarker analysis in the first place.

This work also offers an open-source toolbox for running meta-analysis and SAOB: the code and data used are available assuring the transparency and replicability of these analysis, as well as fostering future ones.

## Conflict of Interest Statement

A. Bussalb, Q. Barthélemy, D. Ojeda, and L. Mayaud work for Mensia Technologies. M. Congedo served as an advisor for Mensia Technologies when this work was conducted.

## Author Contributions

AB extracted all data from articles and performed the analysis, MC provided advice and expertise concerning the design of the study and the methods used, RD and EA provided clinical interpretation of the results. QB and DO supported the toolbox implementation and validation. LM supervised all this work.

## Funding

This research was conducted as part of a PhD thesis funded by Mensia Technologies, the Association Nationale Recherche et Technologie (ANRT), and the EU H2020 NEWROFEED grant 684809.

## Acknowledgments

We would like to thank Jean-Arthur Micoulaud-Franchi, MD, PhD for his helpful comments and ideas stimulating this work. We also would to thank the Association Nationale Recherche et Technologie (ANRT) that partly funded the PhD thesis of the first author as well as the EU H2020 NEWROFEED grant 684809.

## List of abbreviations

**ADHD** Attention Deficit/Hyperactivity Disorder. 2–7, 9, 12, 15

**AgCl** Silver Chloride. 8, 12

**Au** Gold. 8, 12

**CPT** Continuous Performance Test. 3

**EEG** Electroencephalogram. 3, 5, 8, 12, 13, 15, 17, 38

**EOG** Electro-Oculogram. 38

**ERP** Event-Related Potential. 4

**ES** Effect Size. 5–7, 9–17, 32, 33

**fMRI** functional Magnetic Resonance Imaging. 3

**IRB** Institutional Review Board. 8, 12, 38

**LASSO** Least Absolute Shrinkage and Selection Operator. 9, 10, 12, 13, 15, 38

**MPROX** Most Proximal. 6, 14, 16, 34

**MSE** Mean Square Error. 10, 12, 33

**NFB** Neurofeedback. 3–15, 17, 18, 31, 32, 36–38

**OLS** Ordinary Least Squares. 9, 10, 12

**PBLIND** Probably Blind. 6, 12, 14, 16, 17, 34, 38

**PET** Positron Emission Tomography. 3

**RCT** Randomized Controlled Trial. 4, 11

**SAOB** Systematic Analysis of Biases. 5, 7, 10, 18, 36

**SART** Sustained Attention to Response Task. 3

**SCP** Slow Cortical Potential. 4, 8, 13, 38

**SE** Summary Effect. 11, 13, 14, 32, 37

**SMR** Sensorimotor Rhythm. 3, 4, 8, 13, 15, 38

**TBR** Theta Beta Ratio. 4

**TOVA** Test of Variables of Attention. 3

**WLS** Weighted Least Squares. 9, 12, 15, 38

# References


O. Alkoby, A. Abu-Rmileh, O. Shriki, and D. Todder. Can we predict who will respond to neurofeedback? a review of the inefficacy problem and existing predictors for successful eeg neurofeedback learning. *Neuroscience*, 2017. doi: https://doi.org/10.1016/j.neuroscience.2016.12.050.

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.

M. Arns, S. de Ridder, U. Strehl, M. Breteler, and A. Coenen. Efficacy of neurofeedback treatment in adhd: the effects on inattention, impulsivity and hyperactivity: a meta-analysis. *Clinical EEG and neuroscience*, 40(3):180–189, 2009. doi: https://doi.org/10.1177/155005940904000311.

M. Arns, C. K. Conners, and H. C. Kraemer. A decade of eeg theta/beta ratio research in adhd: a meta-analysis. *Journal of attention disorders*, 17(5):374–383, 2013. doi: https://doi.org/10.1177/1087054712460087.

M. Arns, H. Heinrich, and U. Strehl. Evaluation of neurofeedback in adhd: the long and winding road. *Biological psychology*, 95:108–115, 2014. doi: https://doi.org/10.1016/j.biopsycho.2013.11.013.

M. Arns, H. Heinrich, T. Ros, A. Rothenberger, and U. Strehl. Neurofeedback in adhd. *Frontiers in human neuroscience*, 9:602, 2015. doi: https://doi.org/10.3389/fnhum.2015.00602.

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.

T. Banaschewski and D. Brandeis. Annotation: what electrical brain activity tells us about brain function that other techniques cannot tell us–a child psychiatric perspective. *Journal of Child Psychology and Psychiatry*, 48(5):415–435, 2007. doi: https://doi.org/10.1111/j.1469-7610.2006.01681.x.

R. A. Barkley. The ecological validity of laboratory and analogue assessment methods of adhd symptoms. *Journal of abnormal child psychology*, 19(2):149–178, 1991.

T. D. Barry, R. D. Lyman, and L. G. Klinger. Academic underachievement and attention-deficit/hyperactivity disorder: The negative impact of symptom severity on school performance. *Journal of school psychology*, 40(3):259–283, 2002. doi: https://doi.org/10.1016/S0022-4405(02)00100-0.

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.

S. A. Bélanger, M. Vanasse, S. Spahis, M.-P. Sylvestre, S. Lippé, F. l'Heureux, P. Ghadirian, C.-M. Vanasse, and E. Levy. Omega-3 fatty acid treatment of children with attention-deficit hyperactivity disorder: A randomized, double-blind, placebo-controlled study. *Paediatrics & child health*, 14 (2):89–98, 2009. doi: https://doi.org/10.1093/pch/14.2.89.

H. Berger. Über das elektrenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570, 1929.

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.

N. Birbaumer. Clinical-psychological treatment of epileptic seizures: a controlled study. *Perspectives and promises in clinical psychology*, 1991.

A. Bluschke, F. Broschwitz, S. Kohl, V. Roessner, and C. Beste. The neuronal mechanisms underlying improvement of impulsivity in adhd by theta/beta neurofeedback. *Scientific reports*, 6:31178, 2016. doi: https://doi.org/10.1038/srep31178.

R. L. Bouëdec. Mind your brain, un jeu pour entrainer le cerveau. *Le Mag*, 2017.

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

A. Bussalb. Supplemental material. https://github.com/AuroreBussalb/meta-analysis-statistical-tools/blob/master/documentation/article-latex-source/supplemental_material.pdf, 2018b.

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.

J. Cohen. Statistical power analysis for the behavioral sciences 2nd edn. Erlbaum Associates, Hillsdale, 1988.

B. R. Collett, J. L. Ohan, and K. M. Myers. Ten-year review of rating scales. v: scales assessing attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(9):1015–1037, 2003. doi: https://doi.org/10.1097/01.CHI.0000070245.24125.B6.

M. Congedo, J. F. Lubar, and D. Joffe. Low-resolution electromagnetic tomography neurofeedback. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(4):387–397, 2004.

C. K. Conners, G. Sitarenios, J. D. Parker, and J. N. Epstein. The revised conners' parent rating scale (cprs-r): factor structure, reliability, and criterion validity. *Journal of abnormal child psychology*, 26(4):257–268, 1998.

C. K. Conners, J. Pitkanen, and S. R. Rzepa. Conners 3rd edition (conners 3; conners 2008). In *Encyclopedia of clinical neuropsychology*, pages 675–678. Springer, 2011.

W. O. Cooper, L. A. Habel, C. M. Sox, K. A. Chan, P. G. Arbogast, T. C. Cheetham, K. T. Murray, V. P. Quinn, C. M. Stein, S. T. Callahan, et al. Adhd drugs and serious cardiovascular events in children and young adults. *New England Journal of Medicine*, 365(20):1896–1904, 2011. doi: https://doi.org/10.1056/NEJMoa1110212.

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.

M. Deilami, A. Jahandideh, Y. Kazemnejad, Y. Fakour, S. Alipoor, F. Rabiee, G. S. Pournesaie, R. N. Heidari, and S. A. Mosavi. The effect of neurofeedback therapy on reducing symptoms associated with attention deficit hyperactivity disorder: A case series study. *Basic and clinical neuroscience*, 7(2):167, 2016. doi: https://doi.org/10.15412/J.BCN.03070211.

R. Drechsler, M. Straub, M. Doehnert, H. Heinrich, H.-C. Steinhausen, and D. Brandeis. Controlled evaluation of a neurofeedback training of slow cortical potentials in children with attention deficit/hyperactivity disorder (adhd). *Behavioral and brain functions*, 3(1):35, 2007. doi: https://doi.org/10.1186/1744-9081-3-35.

C. DuPaul, Barkley. Stimulants. *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*, pages 510–551, 1998.

N. S. Duric, J. Assmus, D. Gundersen, and I. B. Elgen. Neurofeedback for the treatment of children and adolescents with adhd: a randomized and controlled clinical trial using parental reports. *BMC psychiatry*, 12(1):107, 2012. doi: https://doi.org/10.1186/1471-244X-12-107.

G. Durup and A. Fessard. I. l'électrencéphalogramme de l'homme. observations psychophysiologiques relatives à l'action des stimuli visuels et auditifs. *L'année psychologique*, 36(1): 1–32, 1935.

S. Enriquez-Geppert, R. J. Huster, and C. S. Herrmann. Eeg-neurofeedback as a tool to modulate cognition and behavior: a review tutorial. *Frontiers in human neuroscience*, 11:51, 2017. doi: https://doi.org/10.3389/fnhum.2017.00051.

J. N. Epstein and M. D. Weiss. Assessing treatment outcomes in attention-deficit/hyperactivity disorder: a narrative review. *The primary care companion for CNS disorders*, 14(6), 2012. doi: https://doi.org/10.4088/PCC.11r01336.

C. Escolano, M. Navarro-Gil, J. Garcia-Campayo, M. Congedo, and J. Minguez. The effects of individual upper alpha neurofeedback in adhd: an open-label pilot study. *Applied psychophysiology and biofeedback*, 39(3-4):193–202, 2014. doi: https://doi.org/10.1007/s10484-014-9257-6.

S. V. Faraone, J. Biederman, and E. Mick. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological medicine*, 36(2):159–165, 2006. doi: https://doi.org/10.1017/S003329170500471X.

G. B. Forbes. Clinical utility of the test of variables of attention (tova) in the diagnosis of attention-deficit/hyperactivity disorder. *Journal of clinical psychology*, 54(4):461–476, 1998. doi: https://doi.org/10.1002/(SICI)1097-4679(199806)54:4<461::AID-JCLP8>3.0.CO;2-Q.

M. Fouillen, E. Maby, L. Le Career, V. Herbillon, and J. Mattout. ERP-based BCI for children with adhd: motivations and trial design. 2017.

T. Fuchs, N. Birbaumer, W. Lutzenberger, J. H. Gruzelier, and J. Kaiser. Neurofeedback treatment for attention-deficit/hyperactivity disorder in children: a comparison with methylphenidate. *Applied psychophysiology and biofeedback*, 28(1):1–12, 2003.

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.

V. A. Harpin. The effect of adhd on the life of an individual, their family, and community from preschool to adult life. *Archives of disease in childhood*, 90(suppl 1):i2–i7, 2005. doi: http://dx.doi.org/10.1136/adc.2004.059006.

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.

M. Holtmann, B. Pniewski, D. Wachtlin, S. Wörz, and U. Strehl. Neurofeedback in children with attention-deficit/hyperactivity disorder (adhd)–a controlled multicenter study of a non-pharmacological treatment approach. *BMC pediatrics*, 14(1):202, 2014.

International Electrotechnical Commission. 60601-2-26:2012. `https://webstore.iec.ch/publication/2637#additionalinfo`.

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

T. Janzen, K. Graap, S. Stephanson, W. Marshall, and G. Fitzsimmons. Differences in baseline eeg measures for add and normally achieving preadolescent males. *Biofeedback and self-Regulation*, 20(1):65–82, 1995.

P. S. Jensen. A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of general psychiatry*, 56(12):1073–1086, 1999.

C. Jeunet, F. Lotte, J.-M. Batail, P. Philip, and J.-A. M. Franchi. Using recent bci literature to deepen our understanding of clinical neurofeedback: A short review. *Neuroscience*, 2018. doi: https://doi.org/10.1016/j.neuroscience.2018.03.013.

J. Johnstone, J. Gunkelman, and J. Lunt. Clinical database development: characterization of eeg phenotypes. *Clinical EEG and Neuroscience*, 36(2):99–107, 2005.

J. D. Kropotov, V. A. Grin-Yatsenko, V. A. Ponomarev, L. S. Chutko, E. A. Yakovenko, and I. S. Nikishena. Erps correlates of eeg relative beta training in adhd children. *International journal of psychophysiology*, 55(1):23–34, 2005. doi: https://doi.org/10.1016/j.ijpsycho.2004.05.011.

H. H. Le, P. Hodgkins, M. J. Postma, J. Kahle, V. Sikirica, J. Setyawan, M. H. Erder, and J. A. Doshi. Economic impact of childhood/adolescent adhd in a european setting: the netherlands as a reference case. *European child & adolescent psychiatry*, 23(7):587–598, 2014. doi: https://doi.org/10.1007/s00787-013-0477-8.

E.-J. Lee and C.-H. Jung. Additive effects of neurofeedback on the treatment of adhd: A randomized controlled study. *Asian Journal of Psychiatry*, 25:16–21, 2017. doi: https://doi.org/10.1016/j.ajp.2016.09.002.

U. Leins, G. Goth, T. Hinterberger, C. Klinger, N. Rumpf, and U. Strehl. Neurofeedback for children with adhd: a comparison of scp and theta/beta protocols. *Applied psychophysiology and biofeedback*, 32(2):73–88, 2007. doi: https://doi.org/10.1007/s10484-007-9031-0.

L. Li, L. Yang, C. Zhuo, and Y.-F. Wang. A randomised controlled trial of combined eeg feedback and methylphenidate therapy for the treatment of adhd. *Swiss Med. Wkly*, 143:w13838, 2013. doi: https://doi.org/10.4414/smw.2013.13838.

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.

N. Lofthouse, L. E. Arnold, S. Hersch, E. Hurt, and R. DeBeus. A review of neurofeedback treatment for pediatric adhd. *Journal of Attention Disorders*, 2012. doi: https://doi.org/10.1177/1087054711427530.

S. K. Loo and R. A. Barkley. Clinical utility of eeg in attention deficit hyperactivity disorder. *Applied Neuropsychology*, 12:64–76, 2005. doi: https://doi.org/10.1007/s13311-012-0131-z.

S. K. Loo, J. J. McGough, J. T. McCracken, and S. L. Smalley. Parsing heterogeneity in attention-deficit hyperactivity disorder using eeg-based subgroups. *Journal of Child Psychology and Psychiatry*, 2017. doi: https://doi.org/10.1111/jcpp.12814.

J. F. Lubar and M. N. Shouse. Eeg and behavioral changes in a hyperkinetic child concurrent with training of the sensorimotor rhythm (smr). *Biofeedback and Self-regulation*, 1(3):293–306, 1976.

H. Marzbani, H. R. Marateb, and M. Mansourian. Neurofeedback: a comprehensive review on system design, methodology and clinical applications. *Basic and clinical neuroscience*, 7(2):143, 2016. doi: https://doi.org/10.15412/J.BCN.03070208.

M. Matoušek, P. Rasmussen, and C. Gillberg. Eeg frequency analysis in children with so-called minimal brain dysfunction and related disorders. In *Neurophysiological Correlates of Mental Disorders*, volume 15, pages 102–108. Karger Publishers, 1984. doi: https://doi.org/10.1159/000410508.

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.

V. Meisel, M. Servera, G. Garcia-Banda, E. Cardo, and I. Moreno. Reprint of "neurofeedback and standard pharmacological intervention in adhd: a randomized controlled trial with six-month follow-up". *Biological psychology*, 95:116–125, 2014. doi: https://doi.org/10.1016/j.biopsycho.2013.09.009.

J.-A. Micoulaud-Franchi, P. A. Geoffroy, G. Fond, R. Lopez, S. Bioulac, and P. Philip. Eeg neurofeedback treatments in children with adhd: an updated meta-analysis of randomized controlled trials. *Frontiers in human neuroscience*, 8, 2014. doi: https://doi.org/10.3389/fnhum.2014.00906.

J.-A. Micoulaud-Franchi, F. Salvo, S. Bioulac, and T. Fovet. Neurofeedbach in attention-deficit/hyperactivity disorder: Efficacy. *Journal of the American Academy of Child & Adolescent Psychiatry*, 2016. doi: https://doi.org/10.1016/j.jaac.2016.09.493.

F. Minder, A. Zuberer, D. Brandeis, and R. Drechsler. Informant-related effects of neurofeedback and cognitive training in children with adhd including a waiting control phase: a randomized-controlled trial. *European child & adolescent psychiatry*, pages 1–12, 2018. doi: https://doi.org/10.1007/s00787-018-1116-1.

A. Mohagheghi, S. Amiri, N. Moghaddasi Bonab, G. Chalabianloo, S. G. Noorazar, S. M. Tabatabaei, and S. Farhang. A randomized trial of comparing the efficacy of two neurofeedback protocols for treatment of clinical and cognitive symptoms of adhd: Theta suppression/beta enhancement and theta suppression/alpha enhancement. *BioMed Research International*, 2017, 2017. doi: https://doi.org/10.1155/2017/3513281.

M. R. Mohammadi, N. Malmir, A. Khaleghi, and M. Aminiorani. Comparison of sensorimotor rhythm (smr) and beta training on selective attention and symptoms in children with attention deficit/hyperactivity disorder (adhd): A trend report. *Iranian journal of psychiatry*, 10(3):165, 2015.

D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, 151(4):264–269, 2009.

V. J. Monastra. Electroencephalographic biofeedback (neurotherapy) as a treatment for attention deficit hyperactivity disorder: rationale and empirical foundation. *Child and Adolescent Psychiatric Clinics of North America*, 14(1):55–82, 2005. doi: https://doi.org/10.1016/j.chc.2004.07.004.

V. J. Monastra, D. M. Monastra, and S. George. The effects of stimulant therapy, eeg biofeedback, and parenting style on the primary symptoms of attention-deficit/hyperactivity disorder. *Applied psychophysiology and biofeedback*, 27(4):231–249, 2002.

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.

O. Mowrer. Learning theory and behavior. chapter 2, pages 21–62. John Wiley & Sons Inc, 1960. doi: http://dx.doi.org/10.1037/10802-000.

M. E. Narad, A. A. Garner, J. L. Peugh, L. Tamm, T. N. Antonini, K. M. Kingery, J. O. Simon, and J. N. Epstein. Parent–teacher agreement on adhd symptoms across development. *Psychological assessment*, 27(1):239, 2015. doi: https://doi.org/10.1037/a0037864.

Neba Health, 2015. Neba. https://shop.nebahealth.com/products/neba-fda-cleared-brainwave-adhd-assessment-aid.

G. Ogrim and K. A. Hestad. Effects of neurofeedback versus stimulant medication in attention-deficit/hyperactivity disorder: a randomized pilot study. *Journal of child and adolescent psychopharmacology*, 23(7):448–457, 2013. doi: https://doi.org/10.1089/cap.2012.0090.

D. Pappas. Adhd rating scale-iv: Checklists, norms, and clinical interpretation. *Journal of psychoeducational assessment*, 24(2):172–178, 2006. doi: https://doi.org/10.1177/0734282905285792.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

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

G. S. Reynolds. A primer of operant conditioning, rev. 1975.

I. H. Robertson, T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend. Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35(6):747–758, 1997. doi: https://doi.org/10.1016/S0028-3932(97)00015-8.

J. Rogala, K. Jurewicz, K. Paluch, E. Kublik, R. Cetnarski, and A. Wróbel. The do's and don'ts of neurofeedback training: a review of the controlled studies using healthy adults. *Frontiers in human neuroscience*, 10:301, 2016. doi: https://doi.org/10.3389/fnhum.2016.00301.

D. T. R. Rossiter and T. J. La Vaque. A comparison of eeg biofeedback and psychostimulants in treating attention deficit/hyperactivity disorders. *Journal of Neurotherapy*, 1(1):48–59, 1995. doi: https://doi.org/10.1300/J184v01n01_07.

S. Seabold and J. Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, volume 57, page 61. SciPy society Austin, 2010.

E. S. Shapiro. Academic skills problems fourth edition workbook. Guilford Press, 2010.

B. Shaw-Zirt, L. Popali-Lehane, W. Chaplin, and A. Bergman. Adjustment, social skills, and self-esteem in college students with symptoms of adhd. *Journal of Attention Disorders*, 8(3):109–120, 2005. doi: https://doi.org/10.1177/1087054705277775.

L. H. Sherlin, M. Arns, J. Lubar, H. Heinrich, C. Kerson, U. Strehl, and M. B. Sterman. Neurofeedback and basic learning theory: implications for research and practice. *Journal of Neurotherapy*, 15(4):292–304, 2011.

B. F. Skinner. The operational analysis of psychological terms. *Behavioral and brain sciences*, 7(4): 547–553, 1984.

H. Sollie, B. Larsson, and W.-T. Mørch. Comparison of mother, father, and teacher reports of adhd core symptoms in a sample of child psychiatric outpatients. *Journal of Attention Disorders*, 17 (8):699–710, 2013. doi: https://doi.org/10.1177/1087054711436010.

E. J. Sonuga-Barke, D. Brandeis, S. Cortese, D. Daley, M. Ferrin, M. Holtmann, J. Stevenson, M. Danckaerts, S. van der Oord, M. Döpfner, R. W. Dittmann, E. Simonoff, A. Zuddas, T. Banaschewski, J. Buitelaar, D. Coghill, C. Hollis, E. Konofal, M. Lecendreux, I. C. Wong, and J. Sergeant. Nonpharmacological interventions for adhd: Systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *American Journal of Psychiatry*, 2013. doi: https://doi.org/10.1176/appi.ajp.2012.12070991.

B. Steffert and T. Steffert. Neurofeedback and adhd. *ADHD in practice*, 2(1):16–19, 2010.

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.

M. Sterman, L. MacDonald, and R. K. Stone. Biofeedback training of the sensorimotor electroencephalogram rhythm in man: effects on epilepsy. *Epilepsia*, 15(3):395–416, 1974. doi: https://doi.org/10.1111/j.1528-1157.1974.tb04016.x.

O. J. Storebo, H. B. Krogh, E. Ramstad, C. R. Moreira-Maia, M. Holmskov, M. Skoog, T. D. Nilausen, F. L. Magnusson, M. Zwi, D. Gillies, S. Rosendal, C. Groth, K. B. Rasmussen, D. Gauci, R. Kirubakaran, B. Forsbol, E. Simonsen, and C. Gluud. Methylphenidate for attention-deficit/hyperactivity disorder in children and adolescents: Cochrane systematic review with meta-analyses and trial sequential analyses of randomised clinical trials. *BMJ*, 2015. doi: https://doi.org/10.1136/bmj.h5203.

U. Strehl, U. Leins, G. Goth, C. Klinger, T. Hinterberger, and N. Birbaumer. Self-regulation of slow cortical potentials: a new treatment for children with attention-deficit/hyperactivity disorder. *Pediatrics*, 118(5):e1530–e1540, 2006. doi: https://doi.org/10.1542/peds.2005-2478.

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.

J. M. Swanson, H. C. Kraemer, S. P. Hinshaw, L. E. Arnold, C. K. Conners, H. B. Abikoff, W. Clevenger, M. Davies, G. R. Elliott, L. L. Greenhill, et al. Clinical relevance of the primary findings of the mta: success rates based on severity of adhd and odd symptoms at the end of treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(2):168–179, 2001. doi: https://doi.org/10.1097/00004583-200102000-00011.

E. Taylor. Uses and misuses of treatments for adhd. the second birgit olsson lecture. *Nordic journal of psychiatry*, 68(4):236–242, 2014. doi: https://doi.org/10.3109/08039488.2013.804118.

The American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-4)*, 4 edition, 2000.

The American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*, 5 edition, 2013.

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.

J. Van Doren, H. Heinrich, M. Bezold, N. Reuter, O. Kratz, S. Horndasch, M. Berking, T. Ros, H. Gevensleben, G. H. Moll, et al. Theta/beta neurofeedback in children with adhd: Feasibility of a short-term setting and plasticity effects. *International Journal of Psychophysiology*, 112:80–88, 2017. doi: https://doi.org/10.1016/j.ijpsycho.2016.11.004.

D. Vernon, A. Frick, and J. Gruzelier. Neurofeedback as a treatment for adhd: A methodological review with implications for future research. *Journal of Neurotherapy*, 8(2):53–82, 2004. doi: https://doi.org/10.1300/J184v08n02_04.

Q. Wang, O. Sourina, and M. K. Nguyen. Eeg-based" serious" games design for medical applications. In *Cyberworlds (cw), 2010 international conference on*, pages 270–276. IEEE, 2010.

World Health Organization. *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*, volume 2. World Health Organization, 1993.

World Medical Association. *The Declaration of Helsinki 52nd WMA General Assembly*. 2000.

# Figure captions

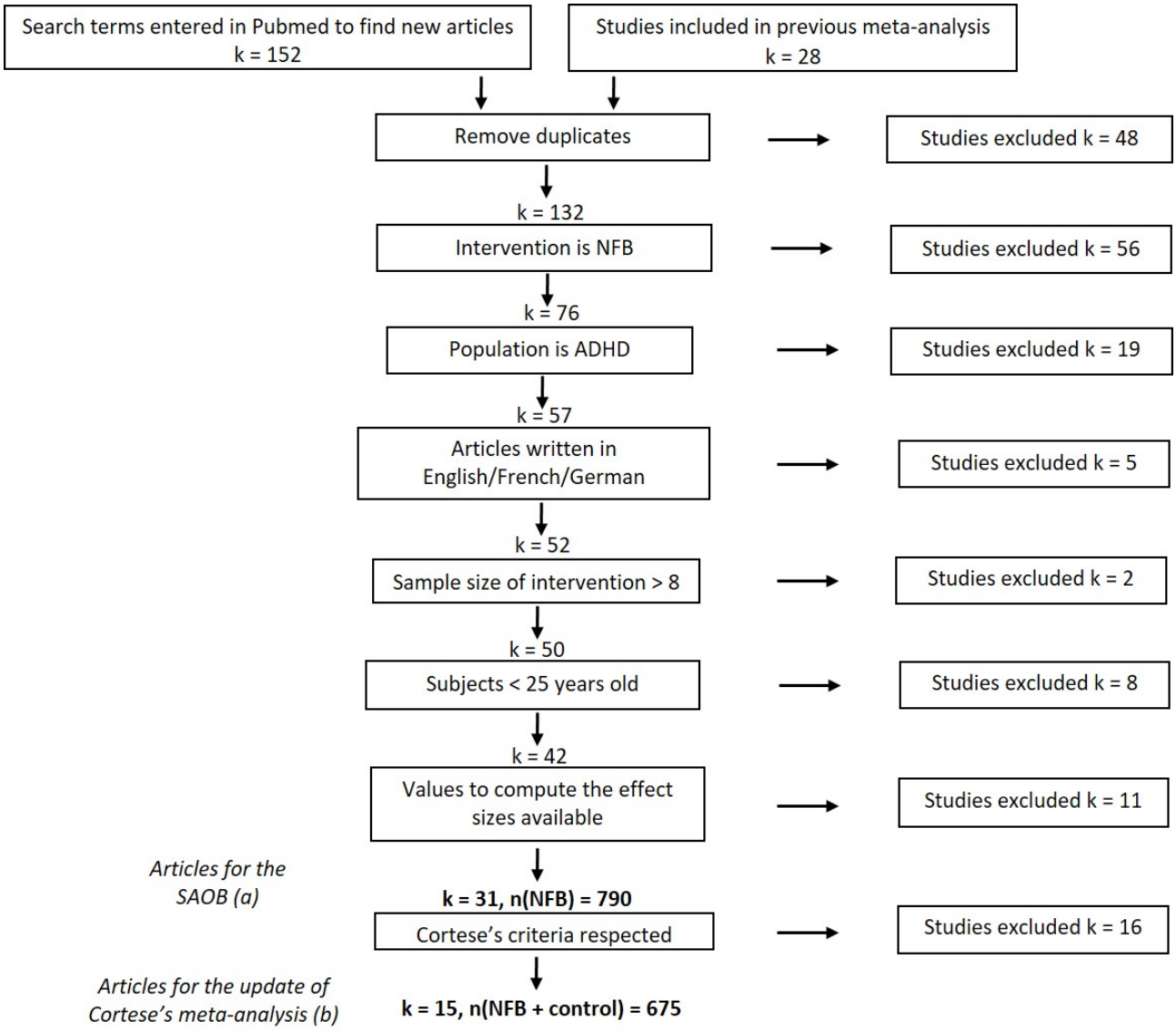


Figure 1: Flow diagram of selection of studies (last search on February 12, 2018). The subset (a) corresponds to the Cortese et al.'s inclusion criteria without the requirement for the presence of a control group. The subset (b) exactly corresponds to the studies included in Cortese et al. [2016] and more recent work meeting the same criteria.

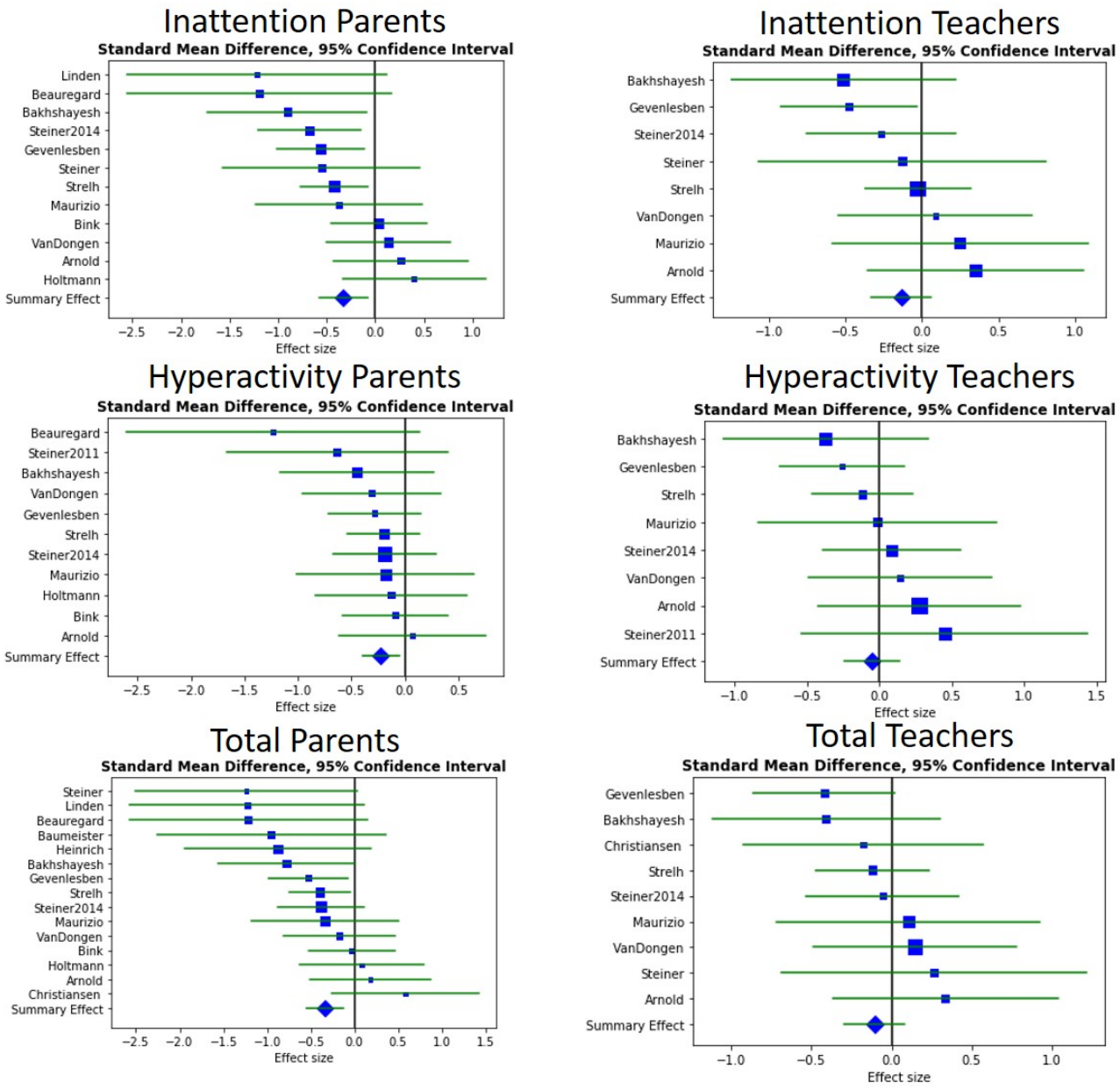


Figure 2: Forest plots showing the between-$ES$. A negative $ES$ is in favor of NFB. The blue squares correspond of the $ES$, the blue diamond to the $SE$ and the green line to the 95% confidence interval.

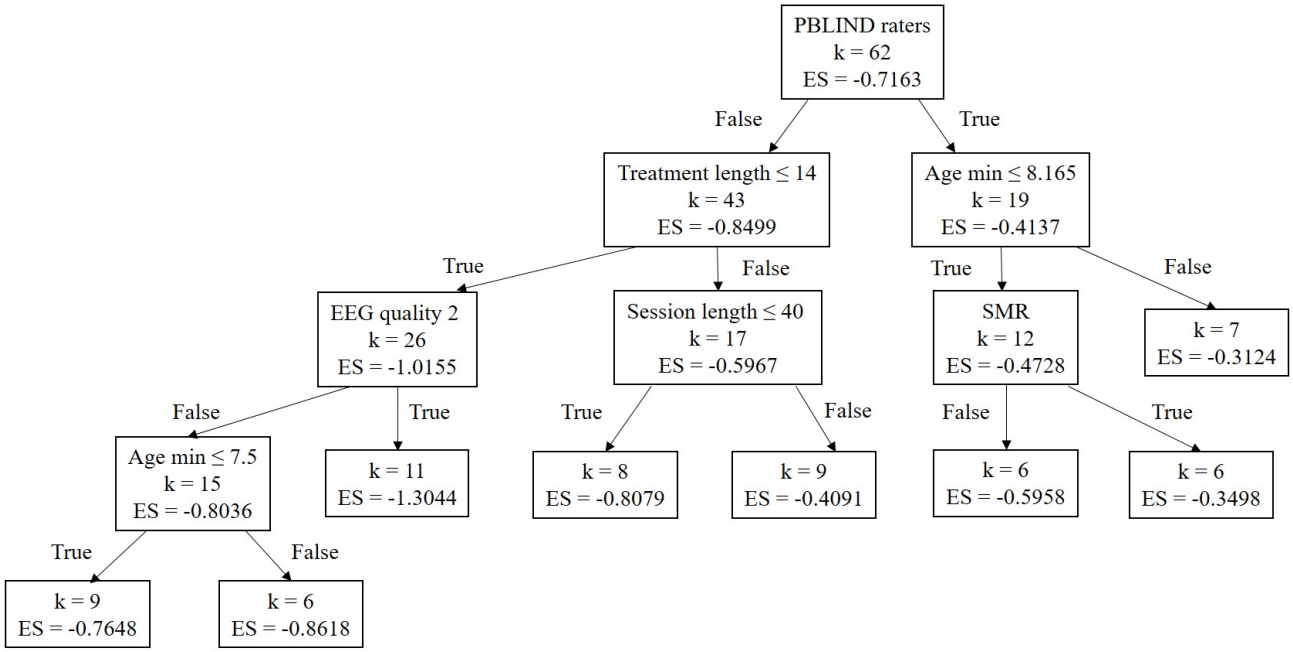


Figure 3: Decision Tree obtained: $ES$ corresponds to the within subject effect size and $k$ to the number of studies. The importance of variables is decreasing from top to bottom. from the root node.

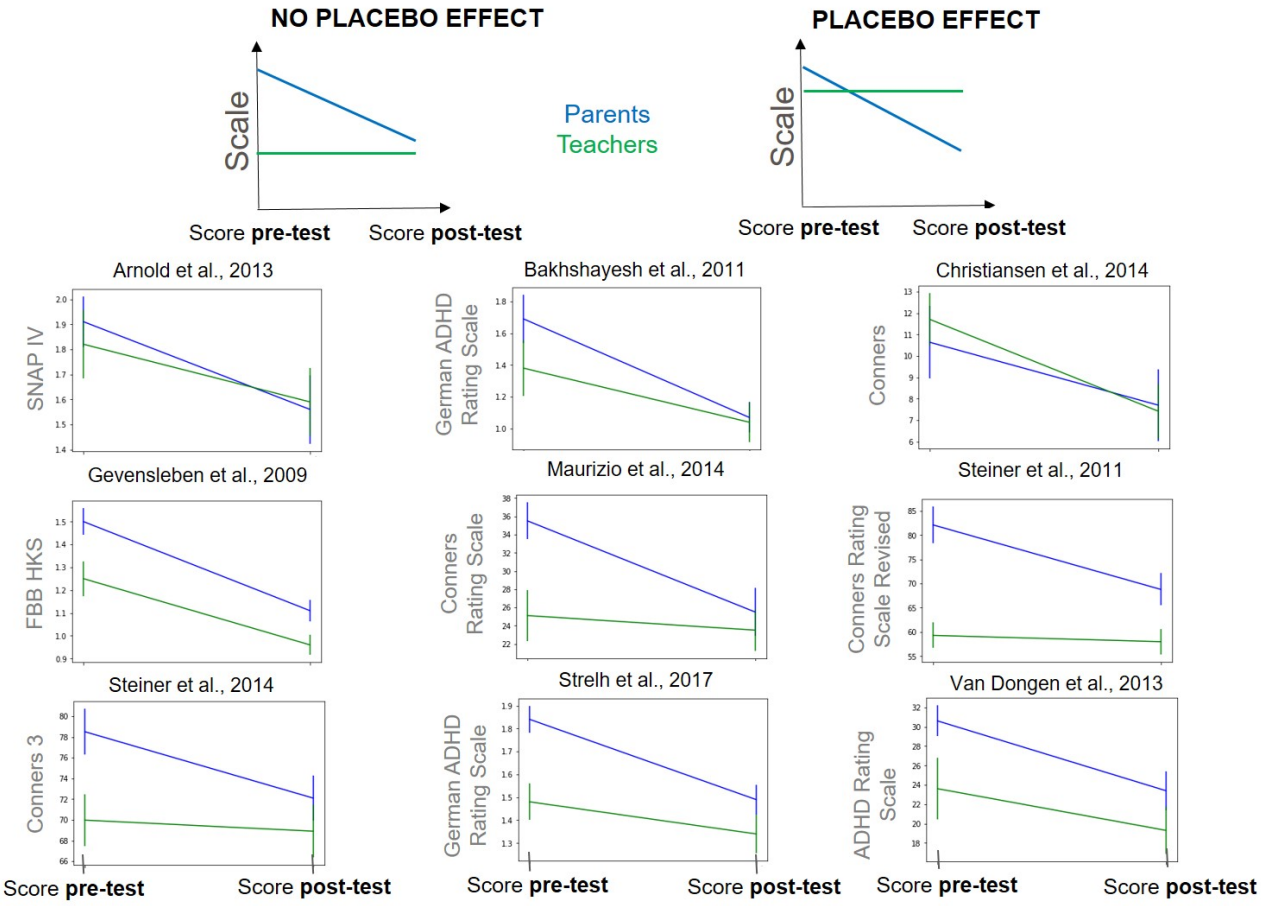


Figure 4: Pre-test and post-test scores ($\pm$ standard error) given by Parents (MPROX) in blue and teachers (PBLIND) in green. Data hypothesized under 2 different hypothesis: **(A)** teachers see less symptoms altogether so that difference pre-post is low and **(B)** Placebo: teacher see as much baseline symptoms as parents but don't see as much improvement as they do. **(C)** Real data: evolution of parents and teachers' scores between pre and post-test on studies that satisfy Cortese et al.'s inclusion criteria and that provide teachers and parents scores on the same scale.

# Table captions

Table 1: List of all studies included in the three different analysis.

| Dataset | | Study | Year | Size of the NFB group |
|---|---|---|---|---|
| | | Arnold et al. | 2014 | 26 |
| | | Bakhshayesh et al. | 2011 | 18 |
| | | Beauregard and Levesque | 2006 | 15 |
| | | Bink et al. | 2014 | 45 |
| | | Christiansen et al. | 2014 | 14 |
| | | Gevensleben et al. | 2009 | 59 |
| | | Heinrich et al. | 2004 | 13 |
| | | Holtmann et al. | 2009 | 20 |
| | | Linden et al. | 1996 | 9 |
| | | Maurizio et al. | 2014 | 13 |
| | | Steiner et al. | 2011 | 9 |
| | | Steiner et al. | 2014 | 34 |
| | | van Dongen-Boomsma et al. | 2013 | 22 |
| | Replicate Cortese et al.[a] | 13 studies | | 297 |
| | | Baumeister et al. | 2016 | 8 |
| | | Strehl et al. | 2017 | 72 |
| Update Cortese et al.[b] | | 15 studies | | 377 |
| | | Bluschke et al. | 2016 | 19 |
| | | Deilami et al. | 2016 | 12 |
| | | Drechsler et al. | 2007 | 17 |
| | | Duric et al. | 2012 | 23 |
| | | Escolano et al. | 2014 | 20 |
| | | Fuchs et al. | 2003 | 22 |
| | | Kropotov et al. | 2005 | 86 |
| | | Lee and Jung | 2017 | 18 |
| | | Leins et al. | 2007 | 19 |
| | | Li et al. | 2013 | 32 |
| | | Meisel et al. | 2014 | 12 |
| | | Mohagheghi et al. | 2017 | 30 |
| | | Mohammadi et al. | 2015 | 16 |
| | | Monastra et al. | 2002 | 51 |
| | | Ogrim and Hestad | 2013 | 13 |
| | | Strehl et al. | 2006 | 23 |
| SAOB[c] | | 31 studies | | 790 |

[a] Studies originally included in Cortese et al. [2016] (search on August 30, 2015), [b] studies satisfying Cortese et al. [2016]'s criteria (search on February 12, 2018), [c] studies satisfying Cortese et al. [2016]'s criteria to the exception of the part relative to the control group (search on February 12, 2018).

Table 2: Comparison between Cortese et al. [2016] results obtained with RevMan [Cochrane Collaboration, 2011] and those obtained with the Python code with our choices applied. $SEs$ and their corresponding p-value (in parenthesis) are presented. With the Python program, a negative $SE$ is in favor of NFB unlike Cortese et al..

| Working hsypothesis | | Same as Cortese et al. [2016] | Our choices[a] |
|---|---|---|---|
| *Parents* | Total | $0.35\ (0.004)$ | $-0.32\ (0.013)$ |
| | Inattention | $0.36\ (0.009)$ | $-0.31\ (0.036)$ |
| | Hyperactivity | $0.26\ (0.004)$ | $-0.24\ (0.02)$ |
| *Teachers* | Total | $0.15\ (0.20)$ | $-0.11\ (0.37)$ |
| | Inattention | $0.06\ (0.70)$ | $-0.17\ (0.16)$ |
| | Hyperactivity | $0.17\ (0.13)$ | $-0.022\ (0.85)$ |

[a] post-test values for Arnold et al. are obtained after 40 sessions of NFB and Conners scale is used for Steiner et al. teachers' outcomes.

Table 3: Results of the WLS, LASSO and decision tree. For the WLS, a p-value $< 0.05$ (in bold) means that the coefficient of the corresponding factor is significantly different from 0. For the LASSO, factors not set to 0 (in bold) are selected. For the decision tree, the place of the factor in the tree is precised. When the value of the coefficient is negative, the corresponding factor may lead to better NFB results.

| Independent variables (factors) | | Coefficients found by WLS (p-value) | Coefficients found by LASSO | Place on the Decision Tree |
|---|---|---|---|---|
| *Signal quality* | Electro-Oculogram (EOG) correction | -0.078 (0.42) | 0.00 | / |
| | artifact correction based on amplitude | **0.15(0.040)** | **0.049** | / |
| *Methodological* | PBLIND | **0.10 (0.043)** | **0.11** | **root node** |
| | randomization | 0.0069 (0.92) | **0.033** | / |
| | IRB | **-0.29 (0.00)** | **-0.15** | / |
| *Population* | age max | -0.090 (0.16) | 0.00 | / |
| | age min | -0.055 (0.37) | 0.00 | **2$^{nd}$ and 4$^{th}$ nodes** |
| | on drugs | 0.069 (0.42) | **0.033** | / |
| *NFB implementation* | number of sessions | -0.0075 (0.92) | 0.00 | / |
| | session length | 0.17 (0.17) | 0.00 | **3$^{rd}$ node** |
| | treatment length | **0.57 (0.00)** | **0.34** | **2$^{nd}$ node** |
| | session pace | **-0.25 (0.00)** | **-0.14** | / |
| | SMR | -0.063 (0.41) | **0.063** | **3$^{rd}$ node** |
| | beta up central | -0.027 (0.72) | 0.00 | / |
| | theta down | **-0.29 (0.014)** | **-0.055** | / |
| | SCP | -0.099 (0.50) | **0.10** | / |
| | transfer phase | **0.27 (0.032)** | **0.12** | / |
| *Quality of acquisition* | more than one active electrode | 0.064 (0.37) | 0.00 | / |
| | EEG quality 2 | **-0.36 (0.00)** | **-0.24** | **3$^{rd}$ node** |