Neurofeedback on children with Attention-Deficit/Hyperactivity Disorder: what factors influence its efficacy?

June 12, 2018

Aurore Bussalb[a], Richard Delorme[b], Eric Acquaviva[b], Marco Congedo[c], Jean-Arthur Micoulaud-Franchi[d], Louis Mayaud[a]

[a]: Mensia Technolgies, Paris France

[b]: Hôpital Robert Debré, Paris France

[c]: GIPSA-Lab, Grenoble France

[d]: Univ. Bordeaux, SANPSY, USR 3413, F-33000 Bordeaux, 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

Two authors, A. Bussalb and L. Mayaud, of this publication work for Mensia Technologies which is developing products related to the research being reported. An other author, 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).

Abstract

Neurofeedback is a noninvasive technique that aims to reduce the ADHD symptoms. Given the impact of meta-analysis on that subject, we first propose to replicate and update the most recent one. Then, we try to identify factors with an influence on Neurofeedback based on the heterogeneity of studies using three multivariate approaches which associate factors with the within subject effect size. The replication and update of the latest meta-analysis confirm the results obtained by its authors: effect sizes are not significant when probably blind ratings are the outcome whereas they are for most proximal raters. Analysis of factors identifies 3 elements which may have an impact on Neurofeedback efficacy: the length of the treatment, the quality of the acquisition of the signals and the person assessing the evolution of the symptoms. Besides these results, we introduce here a new way to look into the heterogeneity of clinical trials.

keywords: ADHD, Neurofeedback, influencing factors, linear regression, decision tree, meta-analysis.

1 Introduction

Attention deficit/hyperactivity disorder (ADHD) is a common psychiatric disorder of childhood characterized by impaired attention and/or hyperactivity/impulsivity, symptoms which 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 has an impact on the children well being because many of them may have low self esteem [Shaw-Zirt et al., 2005] and underachieve at school [Barry et al., 2002] but parents are also affected by this situation [Harpin, 2005]: they are often stigmatized due to the fact that for many, the behavior of children with ADHD is solely explained by bad parenting. Besides ADHD has a financial cost: it is estimated at between $12,005 and $17,458 per individual annually [Pelham et al., 2007].

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]. On the contrary, objective markers of brain function using electroencephalogram (EEG), functional Magnetic Resonance Imaging (fMRI), or Positron Emission Tomography (PET) could not successfully improve diagnosis [Neba Health] at the individual level but proved significantly different on the population. More specifically, these studies allowed to identify specific neurophysiological phenotypes of ADHD: this was particularly reported with EEG recordings [Loo et al., 2017]. For instance, ADHD patients were found to show an increase in theta waves (4-8Hz) in the frontal area whereas there are less beta waves (12-32Hz) and sensorimotor rhythm (SMR) (13-15Hz) in the central area [Monastra, 2005; Matoušek et al., 1984; Janzen et al., 1995].

Among all existing treatments, the most widely used is the psychostimulants, e.g. methylphenidate (MPH), which has been proven to be efficacious [Taylor, 2014; Storebo et al., 2015]. However, the long-term effects when taking psychostimulants are not established: it seems that the decrease of ADHD symptoms does not persist when the patient stops the treatment [DuPaul, 1998; Swanson et al., 2001; Jensen, 1999]. Moreover, ADHD children under medication commonly suffer from side effects such as loss of appetite and sleep problems but no serious adverse events have been reported [Storebo et al., 2015; Cooper et al., 2011]. These drawbacks make some parents and clinicians reluctant to choose such medications, so they turn to drug-free treatment options such as dietary changes [Bélanger et al., 2009] and behavioral therapy which are in most of cases less efficient [Sonuga-Barke et al., 2013].

Neurofeedback (NFB) is a noninvasive technique based on behavioral therapy that aims to reduce the ADHD symptoms [Arns et al., 2015; Steffert and Steffert, 2010]. 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 can be applied such as operant conditioning or voluntary control. To deliver this intervention, neurophysiological time series must be recorded and analyzed in real-time and implemented in serious games leveraging learning paradigms. To that effect, recorded brain signals are analyzed to extract a real-time representations of the activity of a population of neurons involved in attentional networks to which learning paradigms are applied, which is translated into a visual or auditory cues. The sensory feedback constitutes the rewards mechanism that promotes learning using a well-known operant conditioning protocol. The operant conditioning principle will enable the child to repeat more and more easily this task and thanks to the natural neuronal plasticity, a neuronal reorganization is observed [Van Doren et al., 2017].

In case of ADHD, several NFB protocols have been proposed and investigated to decrease the symptoms:

- protocols based on frequency band training: a child can be asked to enhance his SMR while suppressing theta or beta [Lubar and Shouse, 1976], or he can have to enhance beta while suppressing theta (this scenario is known as theta beta ratio (TBR)) [Arns et al., 2013];

- protocol based on the slow cortical potentials (SCPs) training which consists 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];

- protocol based on ERPs (P300) [Fouillen et al., 2017]: ADHD children have a reduced P300 amplitude so it can be considered as a specific neurophysiological marker of selective attention.

Shortly after the discovery of the brain's electric activity by Hans Berger in 1924, Durup and Fessard [1935] proved it could be voluntarily modulated. The first indication of its therapeutic potential came forty years later when Sterman et al. [1974] serendipitously found the training of SMR activity to reduce 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 and resulting in a relatively large body of scientific literature [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 subject to 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 most recent meta-analysis solely on the efficacy of NFB has been conducted by Cortese et al. [2016] in which 13 studies [illegible] included. Although only randomized controlled trials (RCTs) are selected, the authors of this meta-analysis have made some choices which have been debated by the community in particular by Micoulaud-Franchi et al. [2016] who 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] in the meta-analysis.

Because of the publication of new res[illegible]es meeting Cortese et al.'s inclusion criteria, we decided to update his work and take the opportunity to investigate some choices that later proved controversial. Eventually, we extended the analysis with a systematic analysis of biases (SAOB) taking advantage of studies technical and methodological high heterogeneity rather than suffering from it. Indeed, the NFB domain is characterized by a clinical literature that is tremendously heterogeneous: studies differ on a methodological point of view (randomization and presence of a blind assessor for instance), but also on the NFB implementation (number of sessions, session and treatment length and type of protocol for example) and on the acquisition and pre-processing of the EEG. Since we supposed that the methodological and technical choices made by authors may lead to various NFB results, we propose here to identify which of the factors independently influence the reported effect size (ES) thanks to adequate statistical tools.

2 Materials and Methods

2.1 Studies selection

Search terms were entered in Pubmed and studies included in previous meta-analysis were identified. Among these studies, those which satisfied each of these points were selected:

- assess NFB efficacy;
- subjects must be diagnosed 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 according to an expert psychiatrist;
- language is English, German or French;
- include at least 8 subjects in each group;

- subjects must be younger than 25 years old;
- provide enough data to compute required metrics for the following analysis.

The studies satisfying all these points were included in the SAOB, then in order to replicate and update Cortese et al. meta-analysis, we apply the inclusion criteria of this meta-analysis to the found studies.

2.2 Outcome definition

In included studies, the severity of ADHD symptoms have been assessed by parents and, when available, by teachers. Cortese et al. [2016] and Micoulaud-Franchi et al. [2014] defined parents as most proximal (MProx) raters who are not blind to the treatment of their child, as opposed to teachers who are 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 by their perception of it. Efficacy of NFB was given for the following outcomes on clinical scales when available: inattention, hyperactivity/impulsivity and total scores. The factors analysis was performed using the total score only.

2.3 Meta-analysis

Meta-analysis are typically used to aggregate results from different clinical investigations and offer a consolidated state of the evidence. To aggregate results from different studies, it is necessary to assume some level of homogeneity in their design relative to: the inclusion criteria, the intervention, the presence, and type of control. Because studies occasionally use slight variations of a clinical scales and because the populations and control groups might vary in their nature, the scores are typically standardized before being pooled. The between ES is one of such standardized metrics, which we implemented in this paper as described in the Supplemental Materials [?]. The work was carried with an open-source Python package developed for this work that offers a more transparent approach to the choice or parameters and increases replicability. This package was benchmark against RevMan v5.3 [Cochrane Collaboration, 2011] by replicating Cortese et al. [2016]'s work and obtaining the same results. The code is made fully available on a GitHub repository ? including raw data for everyone to review its implementation, update it, and use it for different projects.

Before updating the Cortese et al. [2016] work with recently published clinical work meeting his inclusion criteria [Strehl et al., 2017; Baumeister et al., 2016], we decided to run a sensitivity analysis investigating choices that later proved controversial [Micoulaud-Franchi et al., 2016]. Altogether, the changes investigated in our update included the following:

- the ES of Arnold et al. study is computed from the post-test clinical values taken after the completion of the 40 sessions contrary to Cortese et al. [2016] who used the results after 12 sessions of NFB because final values were not available;

- the ES computed from teachers' assessments for Steiner et al. [2014] rely on the BOSS Classroom Observation [Shapiro, 2010] whereas another scale more often used [Christiansen et al., 2014; Bluschke et al., 2016] and which is the revision of the Conners Rating Scale Revised [Conners et al., 1998] and whose reliability has been studied [Collett et al., 2003] was provided. Thus we decided to compute the ES based on the results from the Conners.

Eventually, the new studies meeting the inclusion criteria defined by Cortese et al. were added to the replication of the meta-analysis.

As initially suggested we performed two subgroups analysis: first, summary effect (SE), the weighted average of all the ES, was calculated with only studies following standard protocol as defined by Arns et al. [2014] and second with studies whose participants take low-dose or no medication during the trial. These analysis were performed with the choices described above.

2.4 Detect factors influencing the Neurofeedback

While revisiting the work carried on meta-analysis, it became apparent that the studies pooled together where highly heterogeneous in terms of methodological and practical implementation. For instance, all NFB interventions were pooled together irrelevant to the quality of the acquisition, the level excreted on real time data quality, and the trained neuromarker. Likewise, the methodological implementations varied significantly, requiring the 'subgroup' analysis (low drug, standard protocols) that are somewhat arbitrary. To circumvent these limitations, we implemented a novel approach taking advantage of the studies heterogeneity rather than suffering from it. In this setting, the within-ES of each intervention is considered as a dependent variable that methodological and technical biases try to explain using standard statistical tools. The results of such analysis should enable us to identify known methodological biases (e.g. blind assessments negatively associates 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 classify the factors influencing the efficacy of NFB in 5 categories: methodological, technical, characteristics of the included population, representative of the quality of acquisition and of the signal. Factors were chosen based on what was typically reported in the literature and presumed to influence ES.

- *signal quality*: the ocular artifacts correction and the artifact correction based on amplitude;
- *population*: the psychostimulants intake during NFB treatment and the age bounds of children;
- *methodological biases*: the presence of a control group, the blinding of assessors, the randomization of subjects, and the approval by an Institutional Review Board (IRB);

- *NFB implementation*: the protocol used (SCP, SMR, theta up, beta up in frontal 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 reminding of the NFB session, the type of thresholding reward, the number of NFB sessions, the sessions frequency during a week, the session length and the treatment length;

- *quality of acquisition*: the presence of one or more active electrode and the EEG quality. This last factor was an indicator between 1 and 3: if EEG was recorded and processed in poor conditions then the indicator would be 1. Besides, if the article didn't precise the recording conditions, the factor would be set to 1. To get an indicator bigger than 1, several points had to be satisfied:

  - *the type of electrodes used*: Silver Chloride (AgCl)/Gel and Gold (Au)/Gel;
  - *check of the electrode contact quality trough the amplifier impedance acquisition mode*: inter-electrode impedance must be smaller than $40k\Omega$;
  - *the amplifier used*: those that are conformed to European norms (such as Thera Prax ®Neuroconn and Eemagine EE-430) are preferable or whose reliability is known.

We provide in the Github repository **??** the raw data extracted from the publications. To prevent any bias, the names of these factors were hidden during the entire analysis so that the data scientist (AB) was fully blind to them. The variable names were only revealed 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 by a 1 and its absence by 0. All variables were standardized, except when the decision tree was performed.

2.4.2 Associate independent factors to effect sizes

To compute this ES, means of total ADHD score given by parents and teachers were used. Besides, in case studies provided results for more than one behavioral scale, ES were computed for each one as

$$\mathrm{ES} = \frac{M_{\text{post},T} - M_{\text{pre},T}}{\sqrt{\frac{\sigma^2_{\text{pre},T} + \sigma^2_{\text{post},T}}{2}}}, \tag{1}$$

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

The ES computed in this analysis was different from the one used previously for the replication and updating of Cortese et al. [2016]. Indeed, here we focused on the effect of the treatment within a group as defined by Cohen [1988], definition of the ES that was already used 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 identified using the following three methods, which were implemented in the Scikit-Learn Python [Pedregosa et al., 2011, 0.18.1] and the Statsmodels Python [Seabold and Perktold, 2010, 0.8.0] libraries:

- 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 most often used linear regression analysis is the Ordinary Least Squares (OLS) but here we applied the WLS as described in eq. (2): a weight was assigned to each observation in order to take into account the fact that some observations came from the same study because studies may provide several scales. Besides, the weight was a function of the sample size as well: because of their different sample sizes, studies were not equivalent and should be analyzed accordingly. That's why the weight corresponded to the ratio between the experiment group's sample size of the study and the number of behavioral scales available in the study. We also ran the analysis with OLS method to assess the impact of the weights on the results.

$$\mathbf{W}y = \mathbf{W}\mathbf{X}\beta + \epsilon. \tag{2}$$

$\mathbf{X}$ 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)

$$\text{WRSS} = \sum_{i=1}^{n} w_i \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2. \tag{3}$$

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, which naturally incorporates variable selection in the linear model thanks to $\ell - 1$-norm applied on the coefficients. The coefficients $\hat{\beta}_j$ are obtained by minimizing the term

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \Big)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{4}$$

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. 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 Mean Square Error (MSE) criteria

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2, \tag{5}$$

with $\hat{y}$ the predicted values. Eventually, a tree is composed of several nodes and leafs, the importance of which is decreasing from the top node.

These methods are intrinsically different from each others, so we compared their results.

3 Results

3.1 Systematic review

Search terms entered in Pubmed returned 152 results during the last check on December 14, 2017 and 28 articles included in previous meta-analysis on NFB were identified. After the selection process illustrated in fig. 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 correspond to Cortese et al.'s criteria to the exception of a need for a control group since we are computing the within subject ES.

3.2 Perform the meta-analysis

The Python module created in order to perform a meta-analysis was successfully validated as describes in the Supplemental Materials and made available online. So all the following results were computed with the Python code.

The replication and the following update of Cortese et al. was conducted by applying the following choices and the results obtained are presented in table 2:

- to compute the ES for Arnold et al. [2014], Cortese et al. [2016] took as post-test values the assessments after 12 NFB sessions because results at post-test were not available. In our case, we use the values at post-test (i.e after the 40 sessions). With these values, we find smaller ES than Cortese et al. [2016];

- different results for teachers' assessments are found for Steiner et al. [2014] because we decided not to use the same scale as Cortese et al.. Indeed, Cortese et al. relied on the BOSS Classroom Observation [Shapiro, 2010] to compute ES

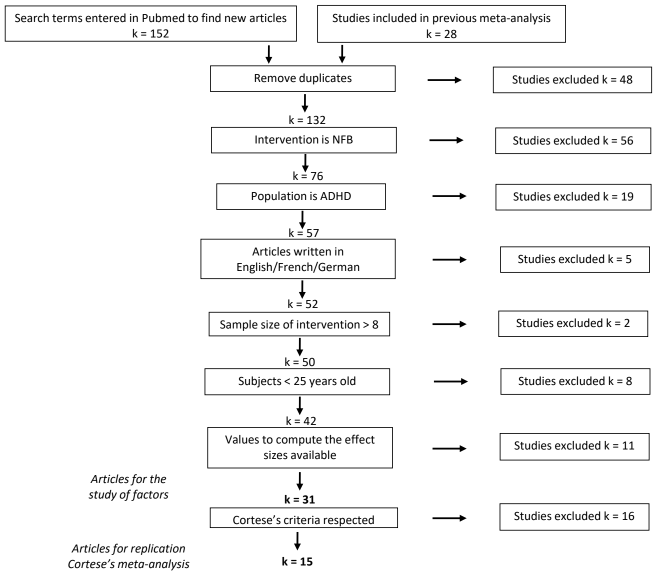

Figure 1: Flow diagram of selection of studies (last search on December 14, 2017). The forlast subset of study exactly corresponds to the studies included in Cortese et al. [2016] and more recent work meeting the same criteria. The last subset (k=31) corresponds to the same criteria without the requirement for the presence of a control group.

for teachers' ratings even if this scale is not as used as other scales provided in this study. That's why we decided to conduct our analysis with a more common scale which has been part of studies assessing the pros and cons of different ADHD scales [Epstein and Weiss, 2012; Collett et al., 2003]: The Conners. Besides, this scale is already used in this study to compute the ES for the parents' ratings. Using this scale leads to higher ES in attention but not in total and hyperactivity. Moreover, this different choice of scale does not affect the significance of the ESs.

Consequently, the rest of this work was carried with the following choices:

- values used to compute ES from Arnold et al. correspond to the scores at post-test (after the 40 sessions of NFB);
- to compute the ES based on teachers' assessments, we used the Conners because this scale is more common.

The next step consists in extend Cortese et al. meta-analysis by adding the two new articles [Strehl et al., 2017; Baumeister et al., 2016] found during the systematic review. Baumeister et al. [2016] provides results only for parents total outcome whereas Strehl et al. [2017] gives teachers and parents' assessments for all outcomes. In spite of favorable results for NFB, particularly on parents' assessments, adding these two

Table 1: List of all studies included in the three different analysis. [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 December 14, 2017), [c] studies satisfying Cortese et al. [2016]'s criteria to the exception of the part relative to the control group (search on December 14, 2017).

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
Replicate Cortese et al.[a]	van Dongen-Boomsma et al.	2013	22
	Baumeister et al.	2016	8
Update Cortese et al.[b]	Strehl et al.	2017	72
	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	20
	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
SAOB[c]	Strehl et al.	2006	23

new studies does not change either the magnitude or the significance of the SE for any outcome whatever the raters as illustrated in fig. 2.

Next, we run the analysis on two specific subgroups: on the one hand only studies following standard protocol defined by Arns et al. [2014] are selected and on the other hand only studies forbidding participants to take medication during the clinical trial are included.

Regarding the standard protocol subgroup, Cortese et al. [2016] found all the outcomes significant except for the hyperactivity symptoms rated by teachers. However, when adding [Strehl et al., 2017] results, we find no significance for the inattention symptoms assessed by teachers as well (p-value = 0.11). As for the low/no medication

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.

Input data		Results from Cortese et al. [2016]	Means and standard deviations from articles included in Cortese et al. [2016]
Implementation		RevMan Cochrane Collaboration [2011]	Python program
Hypothesis		Same as Cortese et al. [2016]	Our choices
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)

subgroup, SEs are not significant except for the inattention symptoms assessed by parents (p-value = 0.013). Besides, the differences in Arnold et al. [2014] values causes 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 are not included in this subgroup because subjects are taking psychostimulants during the trial.

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

3.3 Detect factors influencing the Neurofeedback

This analysis is performed on 31 trials assessing the efficacy of NFB as presented in table 1. Among the 25 factors selected, 6 are removed because there are either too many missing observations or they were too homogeneous: beta up frontal, the use of a transfer card, the type of threshold for the rewards (incremental or fixed), the EEG quality equal to 3 and presence of a control group.

The ES within subjects is computed for all available clinical scales of each study and then factors are associated with the computed ES thanks to three different methods: the WLS, the LASSO and the decision tree. The global results are presented in table 3 where we observe that the techniques used do not return exactly the same factors. Thus, the more methods identified a factor, the more confident we can be in the results.

First, after applying the WLS, we find that 8 factors are significantly different from zero for an adjusted R-squared of 0.74. When applying the OLS, the same factors are returned as significant except the transfer phase, the protocol theta down and the

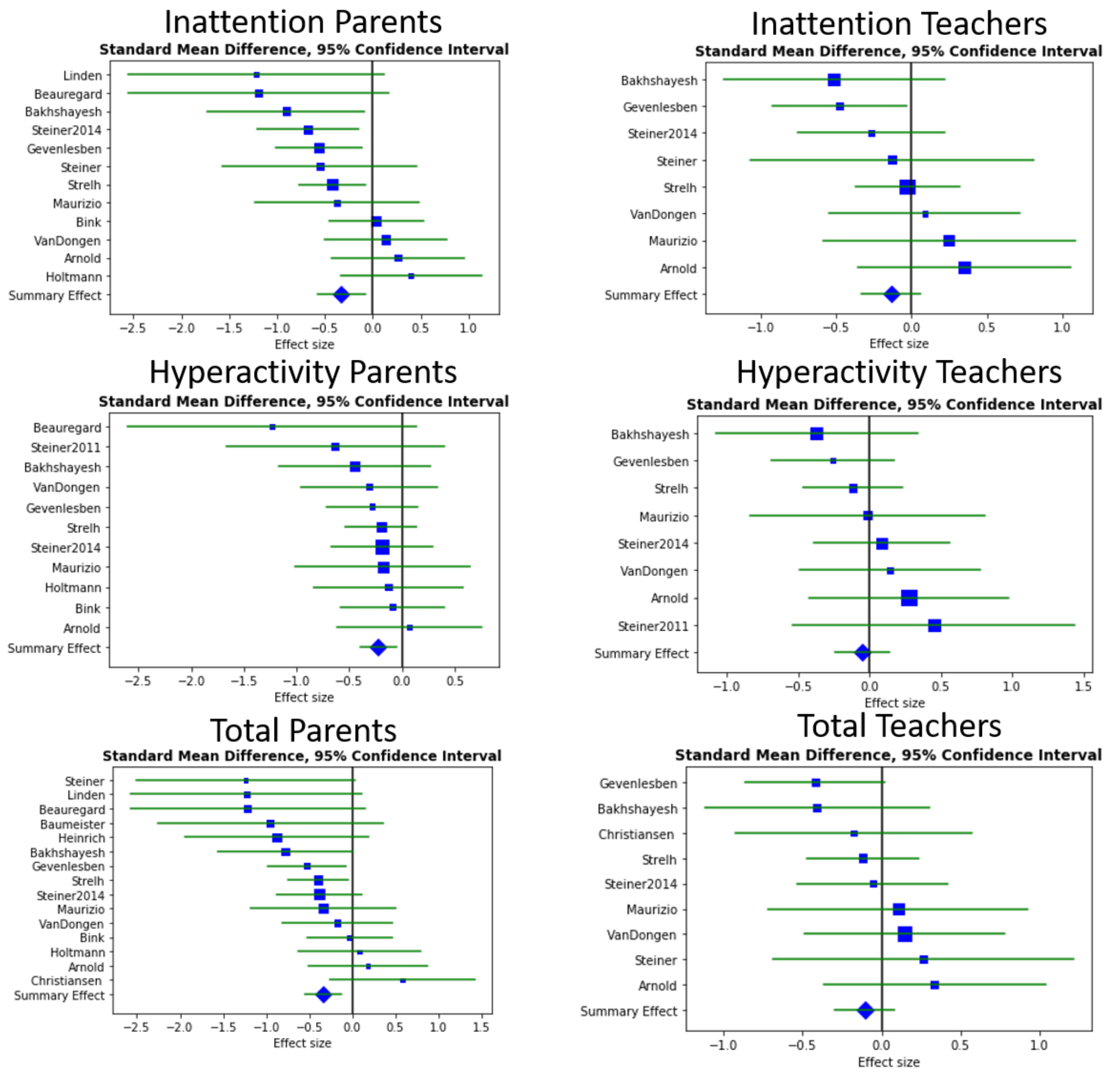

Figure 2: Forest plots obtained on the dataset "Update meta-analysis" with the Python code. A negative ES is in favor of NFB. The blue squares correspond to the ES, the blue diamond to the SE and the green line to the confidence interval.

artifact correction based on amplitude with an adjusted R-squared of 0.42. The LASSO regression selects significant factors by setting to 0 the others: 12 factors are different from 0 here. With these two methods, a negative coefficient means that the factor is in favor of the NFB.

Eventually, the decision tree presented in fig. 3 splits the dataset based on the factor leading to the smallest MSE. The best predictor is the one at the top of the tree: in our case it is the Pblind. Five other factors also splits the subsets.

We can notice that several factors are common to the three methods, in particular the treatment length, the assessment by a blind rater and EEG quality of 2 which are returned by the three methods. Besides, the methods agree on the direction of the influence of these factors. However, it is more difficult to interpret the influence of the

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. When the value of the coefficient is negative, the corresponding factor may lead to better NFB results.

Independent variables (factors)		WLS (p-value)	LASSO	Decision Tree
Signal quality	electro-oculogram (EOG) correction	-0.078 (0.42)	0.00	/
	artifact correction based on amplitude	**0.15(0.040)**	**0.047**	/
Methodological	Pblind	**0.10 (0.043)**	**0.11**	**selected**
	randomization	0.0069 (0.92)	**0.032**	/
	IRB	**-0.29 (0.00)**	**-0.15**	/
Population	age max	-0.090 (0.16)	0.00	/
	age min	-0.055 (0.37)	0.00	**selected**
	on drugs	0.069 (0.42)	**0.032**	/
NFB implementation	number of sessions	-0.0075 (0.92)	0.00	/
	session length	0.17 (0.17)	0.00	**selected**
	treatment length	**0.57 (0.00)**	**0.33**	**selected**
	session pace	**-0.25 (0.00)**	**-0.14**	/
	SMR	-0.063 (0.41)	**0.061**	**selected**
	beta up central	-0.027 (0.72)	0.00	/
	theta down	**-0.29 (0.014)**	**-0.051**	/
	SCP	-0.099 (0.50)	**0.10**	/
	transfer phase	**0.27 (0.032)**	**0.11**	/
Quality of acquisition	more than one active electrode	0.064 (0.36)	0.00	/
	EEG quality 2	**-0.36 (0.00)**	**-0.23**	**selected**

factors returned by only one or two methods. For instance, both WLS and LASSO find that relying on the amplitude of the signal to correct artifacts, and including a transfer phase seem not to improve the ADHD symptoms. Conversely, the IRB approval, a theta down protocol, and a high number of sessions per week appear to positively influence the results. The decision tree and LASSO have in common the protocol SMR: it is associated with lower ES. Five factors are returned by only one of the methods: the

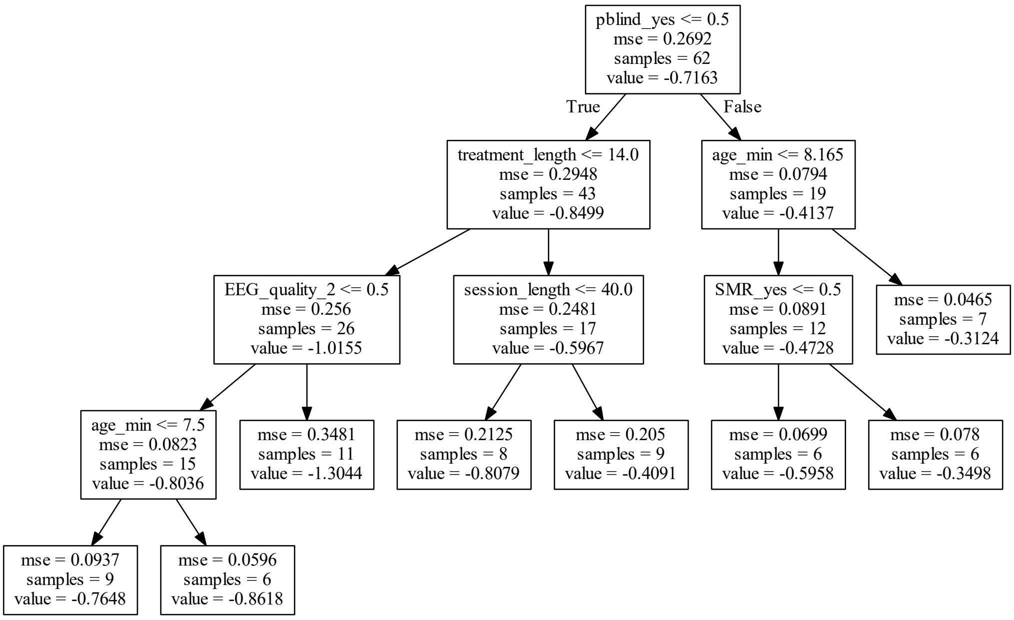

Figure 3: Decision Tree obtained with the factors. The criteria to minimize was the MSE. For the dummy variables, a value of 1 means that the factor is observed in the study. The term value corresponds to the dependent variable.

minimal age of the children, being on drugs during NFB treatment, randomizing the groups and the SCPs protocol.

4 Discussion

4.1 Perform the 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 final results weren't 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, 2008] instead of the BOSS Classroom Observation [Shapiro, 2010] for teachers ratings seems preferable because this scale is more often 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 does not change the significance of the ES whatever the outcome studied.

Second, to compute the ES of Arnold et al. [2014] we used the values at post-test that is to say when all sessions were completed. Some studies suggest that the number of sessions positively correlates with the changes in the EEG [Vernon et al., 2004] so 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] are smaller than those obtained after 12 sessions but these differences do not lead to a change of significance of the SE.

To conclude on that meta-analysis, though some points from the original were controversial and the fact that - for the reasons mentioned earlier - different choices could reasonably be made, it turns out that the impact on the meta-analysis results are minimal and do not change the statistical significance of any outcome. Consequently, the completion of the meta-analysis with studies published since the publication of his work were done with the choices:

- to compute the ES of Arnold et al. [2014] the values at post-test were used;
- the scores reported by teachers on the Conners-3 in Steiner et al.'s study were taken into account instead of these of the BOSS Classroom Observation.

The addition of the two new studies [Strehl et al., 2017; Baumeister et al., 2016] further confirms those results. Indeed, the significance does not change for any outcome: SE found remains significant for MProx raters and non-significant for Pblind.

Adding two more studies increases the significance of the sensitivity analysis ran by Cortese et al., most interestingly, the SE from the subset of studies corresponding to standard protocols of NFB as defined by Arns et al. [2014]. While Cortese et al. found that this subset tends to perform better, particularly on the Pblind outcome, adding two studies confirms this result on the total clinical score (p-value < 0.05). Despite the obvious heterogeneity of the studies included in this subset (particularly in terms of protocol used), this result suggests a positive relation between the features of this *standard* design and NFB performance.

Eventually, concerning the raters, we considered teachers as Pblind raters as Cortese et al. and Micoulaud-Franchi et al. did although they may be aware of the treatment followed thanks to the parents. Besides, the amplitude of the clinical scale at baseline suggests that teachers do not capture the full picture of the condition or see it differently and are therefore less likely to see a change (prone to type II error) [Sollie et al., 2013; Narad et al., 2015]. So using Pblind as an estimate for placebo amplitude assessment may be wrong.

Along with this article, the Python code and raw data are provided in order to facilitate a potential replication of this work (available on the GitHub repository).

4.2 Identify factors influencing the Neurofeedback

Description and analysis 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 detect the influence of methodological, clinical and technical factors on such a wide range of studies.

A somewhat puzzling result is the fact that the three methods which offer to identify factors contributing to the NFB performance do not lead to the exact same results. These discrepancies are clearly explained by the varying hypothesis of these models and actually offer interesting insight into the results and their significance. For instance, the decision tree method is non linear and accounts for variables interaction which is not the case for the two others methods. Moreover, the decision tree is unstable [Dwyer and Holte, 2007], meaning that a small change in the data can cause an important change in the structure of the optimal decision tree.

Nevertheless, despite these differences between the methods, several factors are selected by at least two methods and among them 3 are consistently identified by all the methods with the same direction of influence: if the rater is probably blind to the treatment, the treatment length, and the EEG quality.

Surprisingly, the number of sessions is not found as a significant influencing factor by any method, which is somewhat in contradiction with existing literature. For instance, Enriquez-Geppert et al. [2017] insisted on the fact that the number of sessions should be chosen carefully to avoid "overtraining". Moreover, Arns et al. [2014] stated that performing less than 20 NFB sessions lead to smaller effects. Indeed Vernon et al. [2004] observed that positive changes in the EEG and behavioral performance occurred after a minimum of 20 sessions, but he also points out the fact that the location of the NFB training may had an important influence. Nevertheless, in our study, regardless of the significance of the number of sessions, the coefficient found by the WLS is negative, meaning that as expected, the more sessions performed the more efficient the NFB seems to be.

The type of protocol is not identified by all the three methods but it seems to influence the NFB results according to 2 methods, in particular the theta down protocol which appears more efficient and SMR protocol which conversely seems associated with lower ES. We expected more precised results on the protocols criteria because this point is central in NFB as pointed out by Vernon et al. [2004]. A possible explanation is that all these protocols are equally efficacious to the populations they were offered to and thereby do not constitute a significant explanatory factor. This result, however, does not preclude a combined and personalized strategy (offer the right protocol to the right kid) to further improve performance.

As precised earlier, the three methods agree on three factors.

First, this analysis points out the fact that recording EEG in good conditions seems to lead to better results, which can be explained by the fact that better signal quality enables to extract more correctly the EEG patterns linked to ADHD and henceforth leads to better learning and therapeutic efficacy. However, it remains difficult to really assess the quality of the hardware because little information is provided in the studies.

Next, it appears here that the longer the treatment the less efficient it becomes. Arguably, the treatment length is a proxy for treatment intensity, which means that a treatment that is short in length (and consequently intense in pace) is more likely to succeed. This hypothesis is 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.

As expected, the assessment of symptoms by non-blind raters leads to more favorable results than by Pblind raters - result observed in several meta-analysis [Cortese et al., 2016; Micoulaud-Franchi et al., 2014].

We decided to investigate more precisely this factor, in order to determine if this observation can be solely explained by the placebo effect. First, to assess the impact of the placebo effect, we rely on the results of the decision tree presented in fig. 3 which splits the dataset in two subsets: on one hand 43 observations corresponding to MProx raters assessments and on the other hand 19 observations corresponding to Pblind. If the differences observed between Pblind and MProx raters are due to the placebo effect, we can expect to find in the MProx part of the decision tree factors linked to the perception of the treatment. It is, indeed the case: session and treatment length are returned but if these factors are really linked to the placebo effect, we can expect that the longer the session and the treatment, the higher the ES. Yet, we observe the opposite: it is not in favor of the presence of the placebo effect.

Another way to highlight a possible placebo effect, is to focus on the difference between Pblind and MProx raters first at pre-test and then at post-test. The differences of ratings between teachers and parents have been prone to different studies [Sollie et al., 2013; Narad et al., 2015] which note that teachers are more susceptible to give lower scores especially on younger children. We also observe that fact on most of the studies included in our analysis as illustrated in fig. 4.

Thus, these results suggest that Pblind assessments can hardly be used to assess placebo as they seem to be more blind to symptoms rather than to intervention. In the absence of ethically and technically feasible sham NFB protocols, it seems preferable that evidence of control, learning, and specific changes (correlation with symptoms) are used to determine the mode of action further (neuromarker analysis).

It would have been interesting to study the influence of some other factors such as the delay between brain state and feedback signal as well as the type of NFB game used, but these information are rarely available in studies. Besides, to add more reliability to these results it should be preferable to add more studies, particularly studies with teachers assessments (considered as Pblind).

5 Conclusion

This work offers an open source tool for running meta-analysis and SAOB: the code and data used are available, assuring the transparency and replicability of these analysis.

The results of the first part of the work confirm Cortese et al. [2016]'s findings in the light of recent published clinical work.

Besides this meta-analysis, we propose here a new method for tackling the heterogeneity of clinical data on NFB. This method aims at identify factors as positively or negatively contributing to NFB efficacy.

Among the probable influencing factors, one of them is linked to the quality of acquisition of the EEG. This result indirectly confirms a mode of action through specific EEG training. Likewise, treatment intensity corroborates what is known of learning theory (memory consolidation) and indirectly validates it as part of the mode of action.

These elements converge to the conclusion that existing methodologies traditionally used to assess pills are not adapted to the evaluation of medical devices. Consequently, NFB is probably efficacious and the risk/benefit ratio is certainly in favor of its use as the clinical evidence stands today.

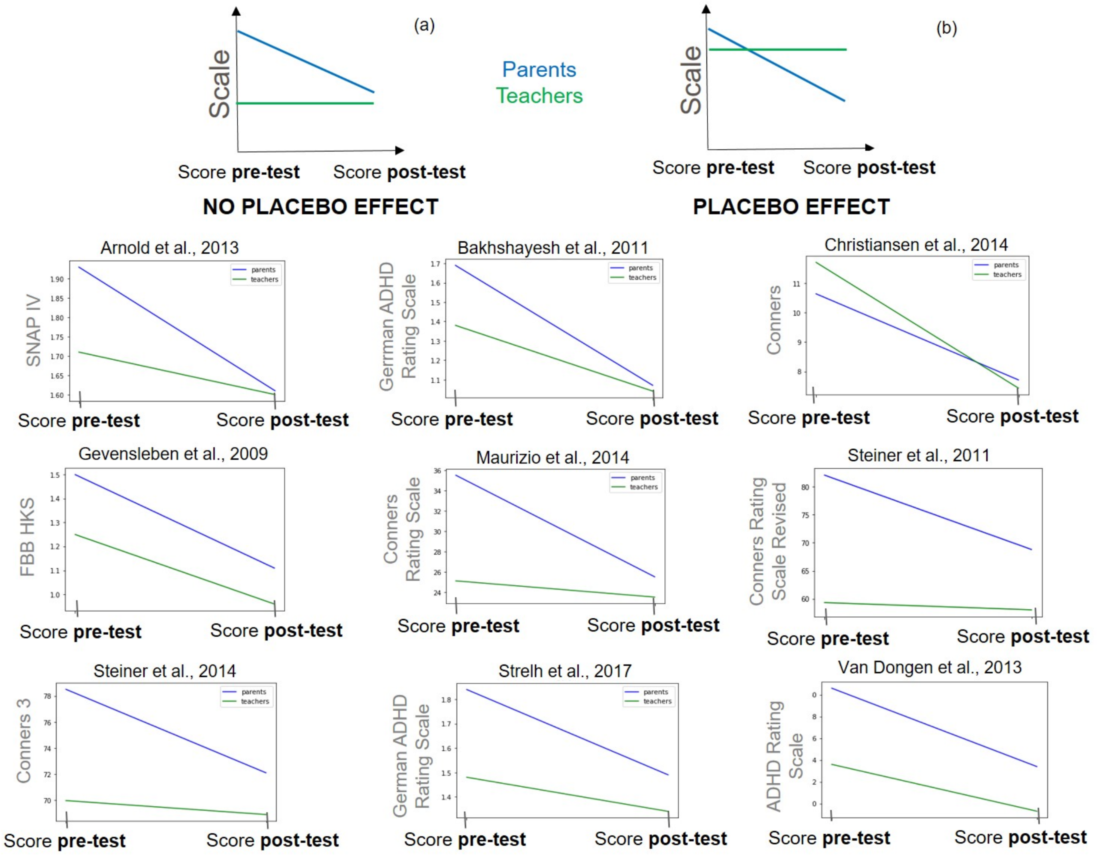

Figure 4: Pre-test and post-test scores given by Parents (MProx) in blue and teachers (Pblind) in green. Studies have to satisfy Cortese et al.'s inclusion criteria and provide teachers and parents scores on the same scale. Two configurations: (a) teachers don't see the symptoms at pre-test so they can't see any improvement at post-test, (b) teachers see the symptoms at pre-test and don't see any improvement at post-test.

6 Disclosure statement

Two authors, A. Bussalb and L. Mayaud, of this publication work for Mensia Technologies which is developing products related to the research being reported. An other author, M. Congedo, served as an advisor for Mensia Technologies when this work was conducted.

7 Funding

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

8 Acknowledgments

We would like to thank Quentin Barthelemy, PhD and David Ojeda, PhD for their helpful comments and ideas on that work.

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.

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.

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.

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.

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, 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.

C. K. Conners. Conners 3. MHS, 2008.

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.

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 psycho-physiologiques relatives à l'action des stimuli visuels et auditifs. *L'année psychologique*, 36(1):1–32, 1935.

K. Dwyer and R. Holte. Decision tree instability and active learning. In *European Conference on Machine Learning*, pages 128–139. Springer, 2007.

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. 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.

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. 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.

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.

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.

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.

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. John Wiley & Sons, 2012.

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. 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.

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.

W. E. Pelham, E. M. Foster, and J. A. Robb. The economic impact of attention-deficit/hyperactivity disorder in children and adolescents. *Journal of pediatric psychology*, 32(6):711–727, 2007. doi: https://doi.org/10.1093/jpepsy/jsm022.

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

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.

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.

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.

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