

Université de Paris
Ecole Doctorale Bio Borbonne Paris Cité ED 562
Hôpital Robert Debré

Etude du Neurofeedback électroencéphalographique appliqué au trouble du déficit de l'attention avec ou sans hyperactivité

Par Aurore Bussalb

Thèse de doctorat de neurosciences

Dirigée par Richard Delorme

Présentée et soutenue publiquement le

Devant un jury composé de :

Prénom Nom, titre [ex. : Maître de conférence], établissement, rôle par rapport à la thèse
[ex : Rapporteur, Président ou encore directeur de thèse]



Except where otherwise noted, this work is licensed under
<https://creativecommons.org/licenses/by-nd/3.0/fr/>

A mon cher grand-père, René Marty.

REMERCIEMENTS

RESUME DE THESE

Mots-clés

ABSTRACT

Keywords

TABLE DES MATIÈRES

| | | |
|----------|---|-----------|
| 1 | Introduction | 10 |
| 2 | Evaluation de l'efficacité du Neurofeedback par la méta-analyse | 11 |
| 2.1 | Principe d'une méta-analyse | 12 |
| 2.1.1 | Buts d'une méta-analyse | 12 |
| 2.1.2 | Choix du modèle | 13 |
| 2.1.3 | Calcul de la taille d'effet | 14 |
| 2.1.4 | Calcul de la fiabilité de chaque taille d'effet | 16 |
| 2.1.5 | Calcul de l'effet total du traitement | 18 |
| 2.1.6 | Calcul de la fiabilité de la taille d'effet totale | 19 |
| 2.1.7 | Calcul de la significativité statistique de la taille d'effet totale | 19 |
| 2.1.8 | Analyse des biais | 20 |
| 2.1.9 | Répresentation graphique des résultats d'une méta-analyse | 21 |
| 2.2 | Replication et mise à jour de la méta-analyse de Cortese et al., 2016 | 21 |
| 2.2.1 | Evaluation des résultats obtenus avec le package Python | 21 |
| 2.2.2 | Réplication de la méta-analyse de Cortese et al., 2016 . . | 23 |
| 2.2.3 | Mise à jour de la méta-analyse de Cortese et al., 2016 . . | 25 |
| 2.3 | Discussion | 29 |
| 2.3.1 | Discussion sur les résultats obtenus | 30 |
| 2.3.2 | Hétérogénéité des études | 32 |
| 2.4 | Importance de la mise à jour des méta-analyses | 32 |
| 2.5 | Conclusion | 40 |

- 3 Identification des facteurs influençant l'efficacité du Neuro-feedback 41
- 4 Analyse de la distribution d'un marqueur de l'attention au sein d'une population d'enfants TDAH 42

Table des figures

| | | |
|-----|---|----|
| 2.1 | Résumé des étapes d'une méta-analyse. | 13 |
| 2.2 | <i>Forest plots</i> de la mise à jour de Cortese et al. [2016]. | 28 |
| 2.3 | <i>Funnel plots</i> de la mise à jour de Cortese et al. [2016]. | 29 |
| 2.4 | Exemples d'évolution des <i>p</i> -value des <i>Effect Size</i> Total (EST) factices au fur et à mesure que la taille de l'échantillon augmente. | 33 |
| 2.5 | Evolution des <i>Effect Size</i> Total (EST) au fur et à mesure de l'ajout de nouvelles études. | 35 |
| 2.6 | Evolution des <i>Effect Size</i> Total (EST) au fur et à mesure de l'ajout de nouvelles études suivant un protocole standard. | 36 |
| 2.7 | Evolution de la <i>p</i> -value des <i>Effect Size</i> Total (EST) au fur et à mesure de l'ajout de nouvelles études. | 37 |
| 2.8 | Evolution de la <i>p</i> -value des <i>Effect Size</i> Total (EST) au fur et à mesure de l'ajout de nouvelles études suivant un protocole standard. | 38 |

Liste des tableaux

| | | |
|-----|---|----|
| 2.1 | Comparaison entre les résultats de Cortese et al. [2016] obtenus avec RevMan [Cochrane Collaboration and others, Copenhagen, Denmark, 2011] et ceux obtenus avec le package Python. | 22 |
| 2.2 | Comparaison entre les résultats de Cortese et al. [2016] et ceux de la réplication avec nos choix de modifications. | 24 |
| 2.3 | Détail des études satisfaisant les critères d'inclusion de [Cortese et al., 2016]. | 27 |

Chapitre 1

Introduction

Chapitre 2

Evaluation de l'efficacité du Neurofeedback par la méta-analyse

Introduction

Les méta-analyses ont pour but de combiner les données de plusieurs études visant à démontrer l'efficacité d'un traitement. Cette méthode est particulièrement intéressante lorsque les études comportent un faible nombre de sujets ou que les résultats des études se contredisent, comme c'est notamment le cas pour le neurofeedback (NFB) appliqué aux enfants Trouble du Déficit de l'Attention avec ou sans Hyperactivité (TDAH), qui est l'un des principaux usages du NFB.

Les différentes étapes à suivre pour réaliser une méta-analyse sont détaillées précisément dans ce chapitre. Ces étapes sont ensuite appliquées à la réplique et à la mise à jour d'une récente méta-analyse sur l'efficacité du NFB appliqué aux enfants TDAH : celle de Cortese et al. [2016] dont certains résultats ont été débattus par la communauté scientifique [Micoulaud-Franchi et al., 2016]. Ainsi, cette analyse a pour but de relever l'éventuel impact de certains choix méthodologiques de Cortese et al. [2016] sur ses conclusions quant à l'efficacité du NFB et de mettre à jour ces résultats en incluant de nouvelles études.

A travers le travail présenté ici, la performance du NFB sur les enfants TDAH est alors évaluée et la revue de littérature qui a été menée a permis de se familiariser avec les études sur l'efficacité de ce traitement et d'en noter les éventuelles faiblesses.

2.1 Principe d'une méta-analyse

Les différentes étapes pour réaliser une méta-analyse sont décrites dans cette partie et résumées à la Figure 2.1. Bien qu'il existe des logiciels permettant de réaliser une méta-analyse, ces étapes ont été implémentées en Python par souci de transparence et de reproductibilité. En effet, les logiciels généralement employés proposent une interface graphique où les formules mathématiques utilisées ne sont pas énoncées clairement. Le code source de ce package Python est disponible sur un dépôt GitHub [Bussalbé et al., 2019] avec sa documentation associée générée par Sphinx (version 1.5.6.).

2.1.1 Buts d'une méta-analyse

Les méta-analyses rassemblent les résultats de plusieurs études, satisfaisant des critères d'inclusion préalablement établis, dans le but d'analyser sur un plus grand nombre de sujets provenant de populations différentes, l'efficacité d'un traitement.

Alors qu'avant les années 1990 les revues narratives (*narrative reviews* en anglais) étaient le plus couramment utilisées pour cette tâche, elles ont perdu leur popularité au profit des méta-analyses. En effet, les revues narratives souffrent de la subjectivité des auteurs qui choisissent notamment le poids à donner à telle ou telle étude : alors que certains vont donner plus d'importance aux études incluant de nombreux sujets, d'autres vont favoriser celles qu'ils jugent de bonne qualité. La méta-analyse permet de réduire cette subjectivité en utilisant par exemple des critères mathématiques définis à l'avance pour calculer le poids à attribuer à chaque étude incluse [Borenstein et al., 2009, Chapitre 1].

Par ailleurs, des recommandations précises quant à la conduite d'une méta-analyse existent [Moher et al., 2009; Cochrane Group]. Ces recommandations appellent notamment à apporter un soin particulier à la sélection des études qui doivent répondre à des critères d'inclusion fixés *a priori* et à évaluer le risque de biais intra et inter-études, comme le biais de publication [Higgins and Green, 2011] qui consiste, pour un chercheur, à avoir tendance à publier des expériences présentant des résultats statistiquement significatifs.

Réaliser une méta-analyse permet de confronter les résultats de toute étude incluse à ceux des autres études intégrées dans l'analyse. L'efficacité du traitement observée pour chaque étude est mesurée à l'aide d'une valeur appelée taille d'effet (ou *Effect Size* (ES) en anglais) présentée en 2.1.3 qui est, le plus souvent, standardisée du fait du regroupement de populations et de mesures relativement

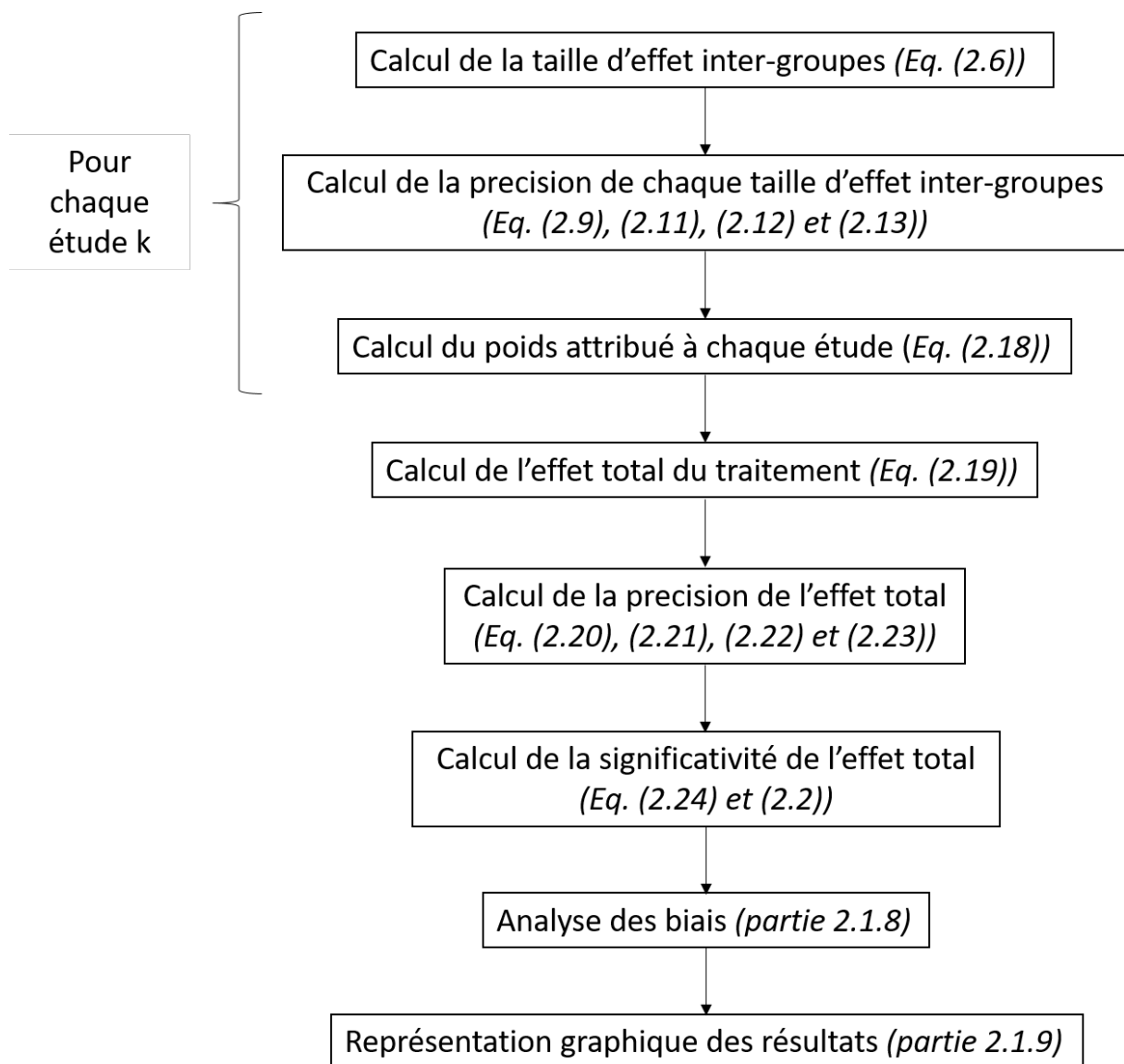


FIGURE 2.1 – Résumé des étapes à suivre pour effectuer une méta-analyse dans le cadre d'un modèle à effets aléatoires.

hétérogènes [Cortese et al., 2016].

2.1.2 Choix du modèle

La première étape consiste à choisir le modèle statistique de la méta-analyse. La plupart des méta-analyses sont basées sur l'un des deux modèles suivants qui reposent sur des hypothèses scientifiques différentes [Borenstein et al., 2009, Chapitre 10] :

- le modèle à effet fixe (*fixed-effect model* en anglais),
- le modèle à effets aléatoires (*random-effects model* en anglais).

Dans le cas du modèle à effet fixe, il est supposé qu'il existe un ES réel (*true ES* en anglais), c'est à dire l'ES qui serait observé avec un nombre de sujets infiniment grand, qui serait le même pour l'ensemble des études incluses dans

la méta-analyse. Les différences entre les ES observés pour chaque étude sont dues à des erreurs d'échantillonnage. Au contraire, dans le cas du modèle à effets aléatoires, l'ES réel peut varier entre les études. Cette variabilité s'explique non seulement par des erreurs d'échantillonnage, mais aussi par les différentes conceptions des études et/ou par les différences entre les sujets inclus.

Les ES obtenus pour chaque étude sont moyennés pour mener à un *Effect Size Total* (EST) (*summary effect* en anglais dont le calcul est décrit en 2.1.5) sur lequel une hypothèse nulle est testée, qui diffère selon le modèle choisi :

- pour le modèle à effet fixe : H_0 : le traitement n'a aucun effet dans chaque étude,
- pour le modèle à effets aléatoires : H_0 : l'effet moyen du traitement est nul.

Le modèle à effets aléatoires est souvent plus approprié du fait de la variabilité des études. En effet, même si les études incluses dans la méta-analyse répondent toutes aux critères d'inclusion fixés au préalable, rien ne peut généralement permettre de supposer que ces études sont identiques et qu'elles partagent donc toutes le même ES réel. Le modèle à effet fixe est ainsi rarement utilisé, on peut cependant y avoir recours lorsque le nombre d'études incluses est très petit. En effet, dans le cas du modèle à effet fixe, les poids associés à chaque étude sont moins équilibrés : les études incluant un grand nombre de sujets se voient attribuer un poids plus important que dans le cas du modèle à effets aléatoires, et les études aux petits échantillons un plus faible poids.

Au sein du domaine du NFB appliqué aux enfants TDAH, les méta-analyses suivent le modèle à effets aléatoires [Cortese et al., 2016; Micoulaud-Franchi et al., 2014]. Ainsi, étant donné l'hétérogénéité des populations étudiées et des critères d'évaluation, et pour être en accord avec la littérature existante, c'est le modèle à effets aléatoires qui est utilisé par la suite.

2.1.3 Calcul de la taille d'effet

Une fois le modèle choisi, l'étape suivante est de quantifier l'efficacité de chaque étude incluse dans la méta-analyse en calculant son ES. Il existe différents ES [Borenstein et al., 2009, Chapitre 3] :

- ES basés sur des *moyennes* :
 - la différence moyenne non standardisée (*unstandardized mean difference* en anglais), D :

$$D = \mu_1 - \mu_2, \quad (2.1)$$

avec μ_1 et μ_2 les moyennes de deux groupes indépendants, ou du même groupe à pré- et à post-test,

- la différence moyenne standardisée (*standardized mean difference* en

anglais), d :

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}}}, \quad (2.2)$$

avec n_1 et n_2 le nombre de sujets dans les groupes 1 et 2, et σ_1 et σ_2 les écarts-types des deux groupes. Le dénominateur correspond à l'écart type intra-groupe calculé à travers les deux groupes.

- ES basés sur des *données binaires* :

- le taux de risque (*risk ratio* en anglais), RR :

$$RR = \frac{A/n_1}{C/n_2}, \quad (2.3)$$

avec A et C le nombre d'évènements observés (par exemple présence d'un effet secondaire) respectivement dans les groupes 1 et 2,

- le taux de chance (*odds ratio* en anglais), OR :

$$OR = \frac{AD}{BC}, \quad (2.4)$$

avec D et B le nombre de non-évènements observés (par exemple absence d'un effet secondaire) respectivement dans les groupes 2 et 1,

- la différence de risque (*risk difference* en anglais), RD :

$$RD = \frac{A}{n_1} - \frac{C}{n_2}. \quad (2.5)$$

Etant donné que les données que nous allons utiliser pour la réplication et la mise à jour de Cortese et al. [2016] sont les moyennes des scores cliniques obtenus par les sujets sur des échelles évaluant les symptômes du TDAH avant le traitement (pré-test) et après le traitement (post-test) et leur écart-type, nous nous concentrons sur les ES basés sur des moyennes. Par ailleurs, les échelles cliniques variant d'une étude à l'autre, les moyennes et les écarts-types ne sont pas comparables : il faut donc standardiser l'ES. Ainsi, nous allons utiliser la différence moyenne standardisée [Cortese et al., 2016; Micoulaud-Franchi et al., 2014].

Enfin, lorsqu'un groupe contrôle est disponible, on peut calculer l'ES-inter-groupes (*between-ES*) comme défini par Morris [2008]. Cet ES est utilisé par Cortese et al. [2016]; Micoulaud-Franchi et al. [2014] et implémenté dans Bussalbi et al. [2019] :

$$\text{ES-inter-groupes} = c_p \left(\frac{(M_{\text{post},T} - M_{\text{pre},T}) - (M_{\text{post},C} - M_{\text{pre},C})}{\sigma_{\text{pre}}} \right). \quad (2.6)$$

L'ES-inter-groupes est équivalent au Z -score d'une distribution normale qui est une mesure de combien d'écarts-types un score brut est au-dessus ou en-dessous de la moyenne d'une population.

L'ES-inter-groupes correspond à la différence entre la moyenne à post-test et à pré-test dans le groupe qui reçoit le traitement ($M_{\text{pre},T}$, $M_{\text{post},T}$) moins la différence entre la moyenne du score à post-test et à pré-test dans le groupe contrôle ($M_{\text{pre},C}$, $M_{\text{post},C}$), divisée par la *pooled standard deviation* à pré-test (σ_{pre}) :

$$\sigma_{\text{pre}} = \sqrt{\frac{(n_T - 1)\sigma_{\text{pre},T}^2 + (n_C - 1)\sigma_{\text{pre},C}^2}{n_T + n_C - 2}}, \quad (2.7)$$

où $\sigma_{\text{pre},G}$ correspond à l'écart-type du groupe G à pré-test et n_G indique le nombre de sujets ayant suivi l'intégralité du traitement dans chaque groupe.

La normalisation de l'ES-inter-groupes (équation Eq. (2.6)) à l'aide de σ_{pre} (équation Eq. (2.7)) se fait seulement avec les données obtenues en pré-test pour ne pas y inclure la variabilité induite par le traitement.

Il s'avère que l'ES-inter-groupes présente un léger biais qui tend à sous-estimer la valeur absolue de la différence moyenne standardisée de la population lorsque les études incluent peu de sujets. Cette valeur est obtenue en calculant la différence entre les vraies moyennes de la population de chaque groupe divisée par l'erreur type de la vraie population qui est supposée être identique pour les deux groupes [Borenstein et al., 2009, Chapitre 4]. Ce biais peut être supprimé grâce à un facteur de correction c_p qui est utilisé pour les petites études (c'est à dire pour lesquelles $n_T + n_C - 2 < 10$) :

$$c_p = 1 - \frac{3}{4(n_T + n_C - 2) - 1}, \quad (2.8)$$

avec $n_T + n_C - 2$ qui correspond au degré de liberté utilisé pour le calcul de σ_{pre} à l'équation Eq. (2.7).

Plus la valeur absolue de l'ES-inter-groupes est élevée, plus l'efficacité du traitement est importante.

2.1.4 Calcul de la fiabilité de chaque taille d'effet

Le terme précision englobe trois valeurs statistiques liées les unes aux autres : la variance, l'Erreur Type (ET) (à ne pas confondre avec l'écart type), et l'intervalle de confiance. Ces trois facteurs de précision définissent un intervalle de valeurs probables pour l'ES réel.

Tout d'abord, la variance de chaque ES-inter-groupes est calculée [Morris, 2008] :

$$\sigma^2(\text{ES}) = c_p^2 \left(\frac{n_T + n_C - 2}{n_T + n_C - 4} \right) \left(\frac{2(1 - r)(n_T + n_C)}{n_T n_C} + \text{ES}^2 \right) - \text{ES}^2, \quad (2.9)$$

où ES désigne l'ES-inter-groupes et r la corrélation de Pearson groupée intra-groupes [James et al., 2013] :

$$r = \frac{\sum_{i=1}^n (s_{pre,i} - \mu_{pre})(s_{post,i} - \mu_{post})}{\sqrt{\sum_{i=1}^n (s_{pre,i} - \mu_{pre})^2} \sqrt{\sum_{i=1}^n (s_{post,i} - \mu_{post})^2}}, \quad (2.10)$$

où n est le nombre de patients inclus dans une étude, $s_{pre,i}$, $s_{post,i}$ les valeurs de scores cliniques pour le sujet i respectivement à pré- et post-test, et μ_{pre} , μ_{post} les scores moyens calculés sur tous les sujets. Il s'agit d'une mesure de corrélation linéaire entre deux variables : une valeur de 1 signifie une corrélation positive entre ces variables, une valeur de -1 une corrélation négative, et une valeur de 0 une absence de corrélation linéaire. Ainsi, si le traitement n'a aucun effet, que l'état des sujets n'évolue pas naturellement, et en l'absence de bruit, r sera égal à 1. On peut en déduire alors qu'une étude composée d'enfants plus jeunes qui sont davantage susceptibles de changer naturellement va présenter une valeur de r plus faible, il en va de même si le traitement est efficace. Enfin la présence de bruit va également avoir un impact sur cette valeur. Par conséquent, une valeur de r doit être calculée pour chaque étude.

Dans notre cas, cette corrélation étant inconnue et les données brutes n'étant pas disponibles pour la majorité des études cliniques, nous approximations la valeur de r en accord avec Balk et al. [2012], qui a trouvé qu'une valeur de 0.5 conduit à des résultats proches de ceux obtenus avec la véritable valeur de la corrélation.

Une fois la variance obtenue, il est aisé de calculer l'Erreur Type (ET) (*standard error* en anglais) de l'ES-inter-groupes [Borenstein et al., 2009, Chapitre 8] :

$$ET = \sqrt{\sigma^2(ES)} = \sigma(ES), \quad (2.11)$$

où ES désigne l'ES-inter-groupes. Alors que la variance est intéressante pour les calculs statistiques, l'ET est quant à elle un index plus aisé à comprendre car elle est sur la même échelle que l'ES.

Enfin, si on suppose que l'ES-inter-groupes suit une distribution normale, son intervalle de confiance à 95% peut être calculé [Borenstein et al., 2009, Chapitre 8] :

$$LL = \text{ES-inter-groupes} - 1.96 * ET, \quad (2.12)$$

et

$$UL = \text{ES-inter-groupes} + 1.96 * ET, \quad (2.13)$$

où LL correspond à la limite inférieure de l'intervalle et UL à la limite supérieure. La valeur de 1.96 correspond au Z -score d'une distribution normale standard pour un intervalle de confiance à 95%. La probabilité de trouver dans cet intervalle la vraie valeur de l'ES-inter-groupes de la population est de 95%.

La précision est affectée, dans une large mesure, par le nombre de sujets inclus dans l'étude : les échantillons plus grands mènent à des variances plus faibles et donc à des estimations d'ES-inter-groupes plus précises [Borenstein et al., 2009, Chapitre 8], c'est pourquoi un plus grand poids leur est attribué dans la méta-analyse comme détaillé par la suite.

2.1.5 Calcul de l'effet total du traitement

Afin d'obtenir l'estimation la plus précise possible de l'effet du traitement sur la population, une moyenne pondérée des ES-inter-groupes des études incluses est calculée.

Si le modèle à effet fixe est choisi, le poids $w_{\text{fixe},k}$ assigné à chaque étude k correspond à l'inverse de la variance de son ES-inter-groupes ($\sigma^2(\text{ES})$, la variance intra-étude) [Borenstein et al., 2009, Chapitre 12] :

$$w_{\text{fixe},k} = \frac{1}{\sigma^2(\text{ES}_k)}. \quad (2.14)$$

Cette étape a pour but de minimiser l'importance des études avec une large variabilité intra-étude.

Dans notre cas nous employons le modèle à effets aléatoires, qui inclut également la variance inter-études τ^2 conduisant à des poids $w_{\text{aléatoires},k}$, simplement notés w_k par la suite, associés aux études différents. Toutefois, afin d'obtenir les poids w_k , il est nécessaire de calculer d'abord les $w_{\text{fixe},k}$ grâce à l'équation Eq. (2.14).

Calculer la variance inter-études se fait en trois étapes décrites par les équations Eq. (2.15), Eq. (2.16) et Eq. (2.17) [Borenstein et al., 2009, Chapitre 12] :

$$Q = \sum_{k=1}^K w_{\text{fixe},k} * \text{ES}_k^2, \quad (2.15)$$

$$C = \sum_{k=1}^K w_{\text{fixe},k} - \frac{\sum_{k=1}^K (w_{\text{fixe},k})^2}{\sum_{k=1}^K w_{\text{fixe},k}}, \quad (2.16)$$

avec K le nombre total d'études incluses, et :

$$\tau^2 = \frac{Q - \text{df}}{C}, \quad (2.17)$$

avec $\text{df} = K - 1$ le degré de liberté.

Le modèle à effets aléatoires prenant en compte les différences entre les études, les poids sont égaux à l'inverse de la somme entre la variance intra-étude

$\sigma^2(ES_k)$ et la variance inter-études τ^2 [Borenstein et al., 2009, Chapitre 12] :

$$w_k = \frac{1}{\sigma^2(ES_k) + \tau^2}. \quad (2.18)$$

Enfin, la moyenne pondérée des K ES-inter-groupes est calculée pour obtenir l'EST comme décrit dans l'équation Eq. (2.19) [Borenstein et al., 2009, Chapitre 12] :

$$EST = \frac{\sum_{k=1}^K w_k * ES_k}{\sum_{k=1}^K w_k}. \quad (2.19)$$

2.1.6 Calcul de la fiabilité de la taille d'effet totale

Une fois l'EST obtenu, on peut calculer ses trois valeurs de précision. Tout d'abord, sa variance $\sigma^2(EST)$:

$$\sigma^2(EST) = \frac{1}{\sum_{k=1}^K w_k}, \quad (2.20)$$

puis son erreur type $ET(EST)$:

$$ET(EST) = \sqrt{\sigma^2(EST)} = \sigma(EST), \quad (2.21)$$

et enfin son intervalle de confiance à 95% :

$$LL(EST) = EST - 1.96 * ET(EST), \quad (2.22)$$

et

$$UL(EST) = EST + 1.96 * ET(EST), \quad (2.23)$$

avec LL la limite inférieure de cet intervalle de confiance et UL sa limite supérieure.

2.1.7 Calcul de la significativité statistique de la taille d'effet totale

Une Z -value est calculée pour tester l'hypothèse nulle pour le modèle à effets aléatoires énoncée en 2.1.2 :

$$Z = \frac{EST}{ET(EST)}. \quad (2.24)$$

Pour un test bilatéral (utilisé par Micoulaud-Franchi et al. [2014]; Cortese et al. [2016]), la p -value est ensuite calculée avec :

$$p = 2(1 - \Phi(|Z|)), \quad (2.25)$$

où Φ est la fonction de distribution cumulative de la loi normale standardisée (*i.e.* centrée et normée).

Si la p -value obtenue est inférieure à 0.05, alors il y a une différence statistiquement significative entre l'efficacité du traitement étudié et celle du groupe contrôle.

D'autres valeurs peuvent ensuite être calculées pour compléter l'analyse, comme par exemple I^2 qui estime l'hétérogénéité des ES-inter-groupes.

2.1.8 Analyse des biais

Les résultats d'une méta-analyse peuvent être impactés par des biais, comme par exemple un biais de publication qui consiste en la tendance qu'ont les auteurs d'études scientifiques de plutôt publier des résultats statistiquement significatifs et en faveur de ce qu'ils veulent démontrer. De plus, d'autres facteurs peuvent influencer les résultats de la méta-analyse : une faible qualité méthodologique conduisant à des effets faussement gonflés dans les petites études, l'hétérogénéité des ES selon la taille des études, et le hasard [Sterne et al., 2011].

L'analyse visuelle de la symétrie d'un *funnel plot* est une méthode couramment utilisée pour détecter un éventuel biais de publication et l'influence des facteurs cités précédemment [Sterne et al., 2011].

Il s'agit d'un nuage de points de la précision de chaque ES-inter-groupes en fonction des ES-inter-groupes. L'ET est couramment utilisée comme estimation de la précision et de la taille d'une étude et est placée sur un axe des abscisses inversé de façon à ce que les plus grandes études soient au sommet et que les plus petites se retrouvent dispersées en bas. En l'absence de biais et d'hétérogénéité entre les études, la répartition des points est seulement due à la variabilité de la taille des études : le graphique est symétrique. Le triangle centré sur l'EST, obtenu avec un modèle à effet fixe et s'étendant de 1.96 ET de chaque côté, inclut 95% des études s'il n'y a pas de biais et si l'hypothèse du modèle à effet fixe est valide.

Déterminer l'asymétrie d'un *funnel plot* peut se faire visuellement mais aussi mathématiquement en utilisant, par exemple, le test d'Egger et al. [1997]. Il s'agit de régresser les ES-inter-groupes divisés par leur ET sur l'inverse des ET. Si l'intercept diffère significativement de zéro (seuil de significativité statistique à 5%), alors le *funnel plot* est asymétrique.

2.1.9 Représentation graphique des résultats d'une méta-analyse

Afin de faciliter la lecture des résultats d'une méta-analyse, ceux-ci sont résumés dans un *forest plot* [Borenstein et al., 2009, Chapitre 1]. Les études incluses sont en ordonnées et les ES-inter-groupes en abscisses. Chaque ES-inter-groupes est représenté par un carré dont la taille est proportionnelle au poids w_k attribué à l'étude k . Les intervalles de confiance à 95% pour chaque ES-inter-groupes sont représentés. En bas du graphique, l'EST est symbolisé par un diamant avec son intervalle de confiance à 95%. Une droite verticale d'équation $x = 0$ est tracée pour délimiter la partie du graphique où les ES sont en faveur du traitement de celle où ils ne le sont pas. Dans notre cas, la partie à gauche de cette droite est en faveur du NFB.

2.2 Replication et mise à jour de la méta-analyse de Cortese et al., 2016

La méta-analyse de Cortese et al. [2016] a eu un fort impact dans la communauté du NFB, du fait du nombre d'études incluses (13) et du soin apporté à leur sélection. Cependant, certains choix méthodologiques ont fait l'objet de débats [Micoulaud-Franchi et al., 2016] et méritent une attention particulière. C'est pourquoi dans un premier temps, la méta-analyse de Cortese et al. [2016] est répliquée en modifiant les points débattus pour évaluer leur impact sur les résultats. Ensuite, étant donné que la recherche dans le domaine du NFB appliqué aux enfants TDAH est très active, de nouvelles études répondant au critère d'inclusion de Cortese et al. [2016] ont pu être publiées depuis cette méta-analyse. La mise à jour de ce genre d'analyse est importante car plus des études y seront incluses, plus les résultats vont se stabiliser, c'est pourquoi la méta-analyse décrite en 2.2.3 comprend, en plus des 13 études originelles, celles qui satisfaisaient les critères d'inclusion de Cortese et al. [2016] depuis sa publication.

Dans un souci de transparence et pour faciliter la réplication de ce travail, ces deux analyses sont effectuées en utilisant le package Python (version de Python 3.6.1) dont le contenu est décrit en 2.1 [Bussalib, 2019]. Par conséquent, la première étape est d'évaluer la fiabilité des résultats obtenus avec ce package.

2.2.1 Evaluation des résultats obtenus avec le package Python

Pour effectuer une méta-analyse, les auteurs utilisent en général le logiciel RevMan [Cochrane Collaboration and others, Copenhague, Danemark, 2011; Cortese et al., 2016; Micoulaud-Franchi et al., 2014]. Ici, les résultats obtenus avec le package Python vont être comparés à ceux donnés par la version 5.1 de

RevMan, présentés dans Cortese et al. [2016], afin de valider l'utilisation de ce package.

Matériel et méthodes

La première étape a été d'entrer dans un fichier .csv, pour chaque étude incluse, les moyennes des scores cliniques à pré-test et post-test avec leur écart-type utilisés par Cortese et al. [2016] pour calculer les ES-inter-groupes. Lorsqu'ils sont disponibles, les scores cliniques évaluant l'inattention, l'hyperactivité et la totalité des symptômes sont extraits afin d'obtenir une valeur d'efficacité pour chacune de ces composantes. Par ailleurs, les scores donnés par les parents (*Most Proximal* (MProx)) mais aussi par les enseignants (*Probably Blind* (PBlind)) sont utilisés.

Une fois cette étape remplie, le package Python lit ce fichier et retourne l'EST ainsi que sa p -value associée en suivant les étapes présentées en 2.1, ce qui permet de conclure quant à l'efficacité du NFB sur chaque composante.

Résultats

En accord avec les résultats de précédentes méta-analyses [Sonuga-Barke et al., 2013; Micoulaud-Franchi et al., 2014], Cortese et al. [2016] met en évidence un effet significativement favorable du NFB lorsque son efficacité est calculée grâce aux évaluations des parents, mais aucune amélioration significative des symptômes n'est reportée par les enseignants. Ces résultats sont retrouvés lorsque l'analyse est effectuée par le programme Python, les valeurs d'EST et leur p -value obtenues sont très proches voire égales à celles de Cortese et al. [2016] comme résumé dans la Table 2.1.

TABLE 2.1 – Comparaison entre les résultats de Cortese et al. [2016] obtenus avec RevMan [Cochrane Collaboration and others, Copenhagen, Danemark, 2011] et ceux obtenus avec le package Python [Bussalib et al., 2019]. Avec le package Python, un ES négatif est en faveur du NFB. Le seuil de significativité statistique est fixé à 5%.

| Données | | Résultats de Cortese et al. [2016] | Résultats de Bussalib et al. [2019] |
|--------------------|---------------|---------------------------------------|--|
| Implémentation | | RevMan | Package meta-analysis |
| <i>Parents</i> | Total | 0.35 (0.004) | −0.34 (0.004) |
| | Inattention | 0.36 (0.009) | −0.35 (0.011) |
| | Hyperactivité | 0.26 (0.004) | −0.24 (0.02) |
| <i>Enseignants</i> | Total | 0.15 (0.20) | −0.13 (0.25) |
| | Inattention | 0.06 (0.70) | −0.09 (0.50) |
| | Hyperactivité | 0.17 (0.13) | −0.15 (0.21) |

Les différences de signe entre les EST de Cortese et al. [2016] et ceux de

Bussalbi et al. [2019] s'expliquent par le choix qu'il a été fait de ne pas multiplier par -1 les ES-inter-groupes calculés avec l'équation Eq. (2.6), cette décision a également été prise par Micoulaud-Franchi et al. [2014].

Les petites différences observées notamment au niveau des p -values sont dues à notre décision d'utiliser systématiquement une *pooled* corrélation de Pearson intra-groupes toujours égale à 0.5 [Balk et al., 2012] lors du calcul de la variance de chaque ES-inter-groupes à l'équation Eq. (2.9). Une analyse de sensibilité a été menée pour s'assurer de l'impact mineur de la valeur de la *pooled* corrélation de Pearson intra-groupes : lorsqu'elle varie entre 0.2 et 0.8, la significativité statistique de l'EST ne change pas.

Au vu de la faible différence entre ce qui est retourné par RevMan et par le package Python, la validité de ce dernier est confirmée : il sera donc utilisé pour mener les analyses suivantes.

2.2.2 Réplication de la méta-analyse de Cortese et al., 2016

Certains choix de la méta-analyse de Cortese et al. [2016] ont été discutés par la communauté scientifique [Micoulaud-Franchi et al., 2016]. Le but de cette section est d'étudier leur impact sur les conclusions de cette méta-analyse.

Matériau et méthodes

Les changements suivants ont été mis en place afin d'être étudiés :

- l'ES-inter-groupes de Arnold et al. [2014] est calculé en utilisant les valeurs cliniques à post-test, c'est à dire obtenues lorsque la totalité des 40 sessions est effectuée, à la différence de Cortese et al. [2016] qui a utilisé les scores cliniques donnés après 12 sessions de NFB car les valeurs finales n'étaient pas encore disponibles,
- Cortese et al. [2016] a calculé les ES-inter-groupes de Steiner et al. [2014] pour les enseignants à partir de leurs évaluations reportées sur la BOSS Classroom Observation [Shapiro, 2010]. Ce choix interpelle car il s'agit d'une échelle peu commune pour quantifier les symptômes du TDAH, d'autant qu'une échelle clinique plus connue et bien définie [Collett et al., 2003; Epstein and Weiss, 2012; Bluschke et al., 2016], la Conners-3 Teachers [Conners et al., 1998; Conners, 2008], est disponible dans cette étude. Ainsi, nous avons utilisé les scores cliniques donnés par les enseignants sur la Conners-3, échelle qui a d'ailleurs été utilisée pour calculer l'ES-inter-groupes basé sur l'évaluation des parents de cette étude.

Résultats

Les résultats obtenus avec ces changements sont résumés dans la Table 2.2. La significativité statistique des EST est inchangée aussi bien pour les parents que pour les enseignants. Les variations de p -values sont expliquées par le fait que les données de départ utilisées sont différentes de celles de Cortese et al. [2016] compte tenu des choix suivis ici.

TABLE 2.2 – Comparaison entre les résultats de Cortese et al. [2016] obtenus avec RevMan [Cochrane Collaboration and others, Copenhagen, Danemark, 2011] et ceux obtenus avec le package Python [Bussalib et al., 2019] avec nos choix de modifications (^a valeurs à post-test de Arnold et al. [2014] sont prises après 40 sessions de NFB et l'efficacité du NFB évaluée par les enseignants dans Steiner et al. [2014] se base sur la Conners-3 Teachers). Avec le package Python, un ES négatif est en faveur du NFB. Le seuil de significativité statistique est fixé à 5%.

| Hypothèses de travail | | Celles de Cortese et al. [2016] | Celles de Bussalib et al. [2019] ^a |
|-----------------------|---------------|------------------------------------|--|
| <i>Parents</i> | Total | 0.35 (0.004) | −0.32 (0.013) |
| | Inattention | 0.36 (0.009) | −0.31 (0.036) |
| | Hyperactivité | 0.26 (0.004) | −0.24 (0.02) |
| <i>Enseignants</i> | Total | 0.15 (0.20) | −0.11 (0.37) |
| | Inattention | 0.06 (0.70) | −0.17 (0.16) |
| | Hyperactivité | 0.17 (0.13) | −0.022 (0.85) |

L'influence de ces changements est examinée plus précisément en s'intéressant à leur impact individuel :

- les ES-inter-groupes obtenus pour l'étude de Arnold et al. [2014] après 40 sessions de NFB sont plus faibles que ceux obtenus par Cortese et al. [2016]. Ces ES-inter-groupes plus faibles vont légèrement diminuer les EST mais sans impacter leur significativité statistique (cf. les trois premières lignes de la Table 2.2),
- le calcul des ES-inter-groupes basé sur les évaluations des enseignants dans Steiner et al. [2014] avec la Conners-3 Teachers conduit à un ES-inter-groupes plus élevé pour la composante inattention, mais plus faible pour les composantes hyperactivité et totale. Toutefois, ces différences n'influencent pas la significativité statistique des EST (cf. les trois dernières lignes de la Table 2.2).

Par conséquent, bien que ces points aient été débattus, leur impact sur les conclusions de la méta-analyse sont minimes et ne changent la significativité statistique d'aucun EST.

2.2.3 Mise à jour de la méta-analyse de Cortese et al., 2016

L'étape suivante consiste à mettre à jour la méta-analyse de Cortese et al. [2016] avec les choix faits lors de la réplication décrite en 2.2.2. Pour ce faire, une recherche sur PubMed a été effectuée puis la performance du NFB a été évaluée d'abord sur l'ensemble des études et ensuite sur des sous-groupes (un qui regroupe les études suivant un protocole standard [Arns et al., 2014] et un autre qui comprend celles où les enfants se voient interdire la prise de psychostimulants durant le traitement par NFB) comme ce qui a été réalisé dans Cortese et al. [2016]. Ces analyses sont effectuées avec les choix définis en 2.2.2.

Matériel et méthodes

Un soin particulier a été mis en oeuvre par Cortese et al. [2016] pour sélectionner les études bien conduites. Les principaux critères d'inclusion et d'exclusion sont [Cortese et al., 2016] :

- seules les études randomisées et contrôlées sont retenues,
- les participants doivent avoir entre 3 et 18 ans et être diagnostiqués TDAH,
- les participants présentant une comorbidité rare sont exclus,
- les contrôles acceptés sont : le traitement habituel, la liste d'attente, un traitement actif ou un placebo/*sham*-NFB,
- les études comparant le NFB au traitement médicamenteux dont la dose est optimisée ou bien où le NFB est couplé au traitement médicamenteux dont la dose est optimisée sont exclues,
- la prise de médicaments en tant que traitement en arrière-plan dans le groupe NFB ou contrôle est acceptée.

Afin de trouver les études à inclure, Cortese et al. [2016] ont cherché sur plusieurs bases de données (dont la dernière vérification date du 30 août 2015) telles qu'ERIC, OVID et PubMed. Dans notre cas seule PubMed, qui est la principale base de données, a été questionnée en entrant les mêmes termes de recherche que Cortese et al. [2016] (disponibles dans le matériel supplémentaire de leur méta-analyse).

La mise à jour de Cortese et al. [2016] se fait dans un premier temps avec l'ensemble des études (celles initialement incluses et celles nouvellement identifiées), puis sur des sous-groupes.

La première sous-population étudiée correspond aux sujets ayant suivi un entraînement par NFB dit *standard*, c'est à dire répondant aux critères établis par Arns et al. [2014] :

- le protocole de NFB utilisé a fait l'objet de plusieurs études cliniques (diminution du Theta-Beta Ratio (TBR), augmentation du Rythme Sensori-

moteur (SMR) et *Slow Cortical Potentials* (SCP)),

- une phase de transfert doit être proposée durant l'entraînement pour aider la transposition dans la vie de tous les jours,
- le lieu où sont effectuées les sessions doit être précisé,
- le traitement par NFB doit satisfaire les critères de la théorie de l'apprentissage, c'est à dire ne pas utiliser de seuil de récompense automatique.

La deuxième sous-population étudiée comprend les sujets ne prenant aucun traitement médicamenteux durant le traitement par NFB.

Résultats de la sélection des études

Finalement, 7 études ont été identifiées lors de la recherche menée le 2 septembre 2019, dont le détail est présenté à la Table 2.3 :

- [Baumeister et al., 2016] pour qui seuls les résultats pour les symptômes totaux évalués par les parents sont disponibles,
- [Strehl et al., 2017] pour qui les évaluations des parents et des enseignants sont données pour toutes les composantes,
- [Bazanov et al., 2018] qui donne les évaluations des parents pour les composantes inattention, hyperactivité et totale (pour être cohérent avec la *Systematic Analysis of Biases* (SAOB) présentée dans le chapitre ??, seul le groupe NFB est retenu),
- [Minder et al., 2018] pour qui les évaluations des parents et des enseignants sont données pour toutes les composantes ; un groupe NFB et un groupe contrôle suivent leur traitement à l'école, alors que les autres groupes NFB et contrôle l'effectue en clinique,
- [Moreno-García et al., 2019] pour qui seuls les résultats pour les symptômes totaux évalués par les parents sont disponibles,
- [Shereena et al., 2019] pour qui les évaluations des parents et des enseignants sont données pour toutes les composantes,
- [Aggensteiner et al., 2019] pour qui les évaluations des parents sont disponibles pour la composante totale et inattention.

Avec cette mise à jour, la méta-analyse comprend désormais 20 articles au lieu de 13 initialement.

En ce qui concerne l'étude des sous-groupes :

protocole standard : quatre études sont ajoutées [Strehl et al., 2017; Baumeister et al., 2016; Aggensteiner et al., 2019; Minder et al., 2018] aux 7 initiales [Bakhshayesh et al., 2011; Christiansen et al., 2014; Gevensleben et al., 2009; Beauregard and Levesque, 2006; Holtmann et al., 2009; Heinrich et al., 2004; Linden et al., 1996],

pas de traitement médicamenteux en simultané : deux études sont

TABLE 2.3 – Détail des études satisfaisant les critères d'inclusion de [Cortese et al., 2016] après la recherche PubMed du 2 septembre 2019.

| Articles | Nombre de sujets | Résultats disponibles |
|----------------------------------|------------------------|--|
| Baumeister et al. [2016] | 8 NFB 8 contrôles | parents : total enseignants : / |
| Strehl et al. [2017] | 72 NFB 68 contrôles | parents : total, inattention, hyperactivité enseignants : total, inattention, hyperactivité |
| Bazanov et al. [2018] | 17 NFB 14 contrôles | parents : total, inattention, hyperactivité enseignants : / |
| Aggensteiner et al. [2019] | 75 NFB 69 contrôles | parents : total et inattention enseignants : / |
| Minder et al. [2018] à l'école | 19 NFB 19 contrôles | parents : total, inattention, hyperactivité enseignants : total, inattention, hyperactivité |
| Minder et al. [2018] en clinique | 19 NFB 20 contrôles | parents : total enseignants : total |
| Moreno-García et al. [2019] | 19 NFB 19 contrôles | parents : total enseignants : total |
| Shereena et al. [2019] | 15 NFB 19 contrôles | parents : total, inattention, hyperactivité enseignants : total, inattention, hyperactivité |

ajoutées [Bazanov et al., 2018; Moreno-García et al., 2019] aux 7 initiales [Beauregard and Levesque, 2006; Gevensleben et al., 2009; Bakhshayesh et al., 2011; Arnold et al., 2014; Linden et al., 1996; Christiansen et al., 2014; Maurizio et al., 2014].

Résultats de la mise à jour sur toutes les études

Parmi les nouvelles études ajoutées, certaines telles que Shereena et al. [2019]; Strehl et al. [2017] concluent à des résultats favorables du NFB, contrairement à d'autres [Moreno-García et al., 2019; Minder et al., 2018].

Les résultats de la mise à jour sont illustrés par les *forest plots* à la Figure 2.2 : les évaluations des parents mènent à des résultats statistiquement significatifs et en faveur de l'efficacité du NFB, à l'inverse des enseignants pour qui aucun EST n'est significatif.

Afin de détecter un biais de publication, une hétérogénéité parmi les études incluses ou la faible qualité méthodologique de certaines études, deux *funnel plot* basés sur les ES-inter-groupes calculés sur les évaluations des symptômes totaux par les parents et les enseignants sont tracés à la Figure 2.3.

Visuellement, alors que le *funnel plot* correspondant aux EST des enseignants

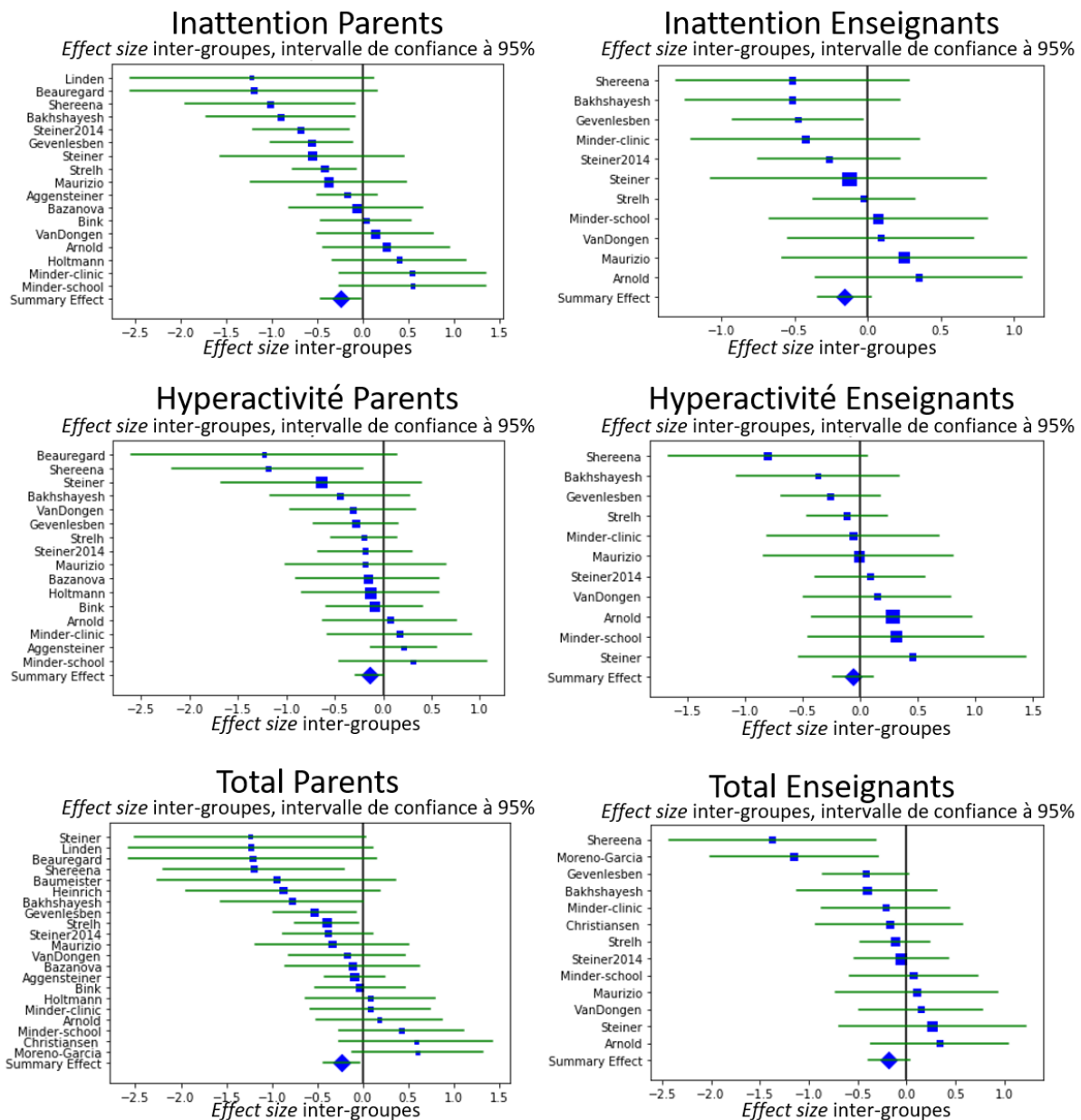


FIGURE 2.2 – *Forest plots* des ES-inter-groupes (les carrés bleus) avec leur intervalle de confiance à 95% (en vert) obtenus après la mise à jour de Cortese et al. [2016]. Le losange bleu correspond à l'EST. Un ES négatif est en faveur du NFB.

semble plutôt symétrique, celui des parents paraît asymétrique. Pour plus de précision, le test d'Egger a également été utilisé pour déterminer statistiquement si un biais s'est glissé dans l'analyse [Egger et al., 1997] :

pour les parents : contrairement à ce qui est observé, le *funnel plot* n'est pas asymétrique, l'intercept ne diffère pas significativement de 0 (p -value = 0.264),

pour les enseignants : le *funnel plot* n'est pas non plus asymétrique (p -value = 0.543).

Les études en dehors des pseudo limites de confiance à 95% de l'EST sous le modèle à effet fixe sont :

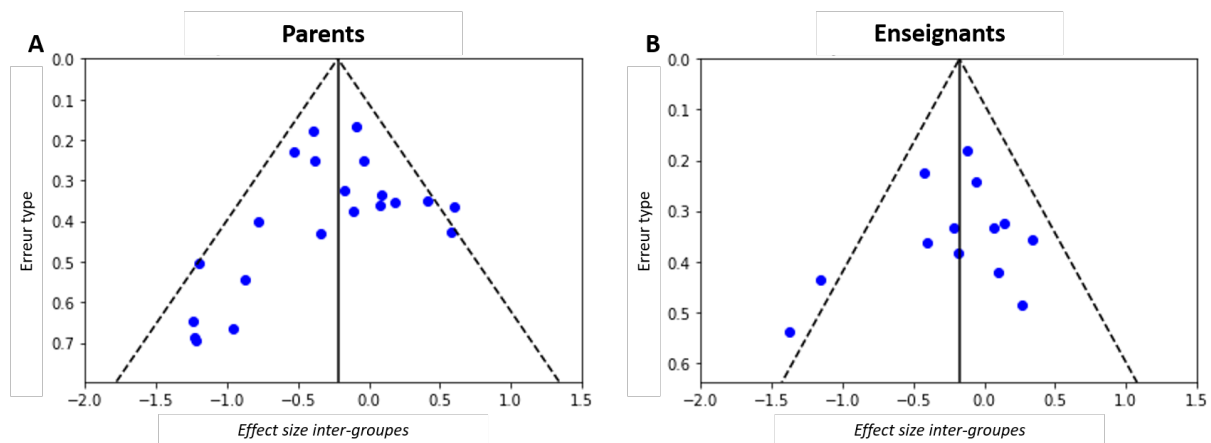


FIGURE 2.3 – *Funnel plots* obtenus pour les ES-inter-groupes calculés sur les évaluations des symptômes totaux par les parents (en **A**) et les enseignants (en **B**) avec un intervalle de confiance à 95%.

- pour les parents : Christiansen et al. [2014],
- pour les enseignants : Moreno-García et al. [2019]; Shereena et al. [2019].

Résultats de la mise à jour sur les sous-groupes

En ce qui concerne le sous-groupe *protocole standard*, seul l'EST de la composante totale évaluée par les parents est significatif (p -value = 0.047). L'EST calculé pour la composante totale évaluée par les enseignants est limite significatif (p -value = 0.053).

Le sous-groupe *pas de traitement médicamenteux en simultané* ne présente que l'EST de la composante inattention évaluée par les parents de significatif (p -value = 0.017).

2.3 Discussion

Les méta-analyses doivent être menées rigoureusement, ainsi pour guider les auteurs des recommandations existent comme celles de PRISMA [Moher et al., 2009]. La réplication et la mise à jour présentées ici remplissent la majorité des points de cette checklist, sauf notamment l'évaluation du risque de biais dans chaque étude. Par ailleurs, elles ont été effectuées avec le package Python dont la fiabilité n'a été testée que sur une méta-analyse.

Le travail décrit précédemment a pour but d'explorer l'impact de certains choix de Cortese et al. [2016] qui ont été débattus dans la communauté scientifique [Micoulaud-Franchi et al., 2016]. Nous résumons ici la liste des changements, leur justification et leurs conséquences sur les résultats.

2.3.1 Discussion sur les résultats obtenus

Les résultats obtenus suite à la réplication et à la mise à jour de Cortese et al. [2016] sont analysés et mis en perspective avec la littérature existante.

Réplication

Un des choix qui a été fait ici est d'utiliser la Conners-3 Teachers [Conners, 2008] plutôt que la BOSS Classroom [Shapiro, 2010] pour calculer les ES-inter-groupes obtenus par les évaluations des enseignants du fait de son utilisation plus commune [Christiansen et al., 2014; Bluschke et al., 2016]. Toutefois, l'utilisation de l'une ou l'autre de ces échelles ne change pas la significativité statistique des EST calculés pour les enseignants.

La seconde différence entre [Cortese et al., 2016] et la réplication effectuée ici est que le calcul des ES-inter-groupes de Arnold et al. [2014] se base sur les valeurs à post-test obtenues après 40 sessions de NFB au lieu de valeurs temporaires obtenues après 12 sessions. Des études montrent que le nombre de sessions est corrélé positivement avec les changements observés sur l'électroencéphalogramme (EEG) [Vernon et al., 2004], ainsi un faible nombre de sessions mènerait à des ES-inter-groupes plus petits. Or, ici étonnamment les ES-inter-groupes calculés après la réalisation de toutes les sessions sont plus faibles que ceux obtenus après 12 sessions, ce qui n'est pas favorable à l'efficacité du NFB. Toutefois, ces différences ne changent pas la significativité statistique des EST.

En conclusion, sur la base de cette étude de sensibilité, nous suggérons que les points de débat soulevés par Micoulaud-Franchi et al. [2016] ne sont pas majeurs quant à l'analyse et l'interprétation des résultats obtenus.

Mise à jour

La méta-analyse de [Cortese et al., 2016] a déjà fait l'objet d'une mise à jour présentée dans Bussalbi et al. [2019] qui a intégré les études de [Bazanov et al., 2018; Baumeister et al., 2016] et Strehl et al. [2017]. Cette mise à jour confirmait les résultats de [Cortese et al., 2016] : les EST calculés à partir des évaluations des parents sont en faveur de l'efficacité du NFB, au contraire des évaluations des enseignants.

Les résultats obtenus ici sont globalement les mêmes qu'après la mise à jour de Bussalbi et al. [2019] au niveau de la significativité statistique : la seule différence importante est la perte de la significativité statistique de l'EST pour la composante hyperactivité évaluée par les parents (p -value = 0.069). Les EST évalués par les enseignants sont, quant à eux, toujours non significatifs (p -value

pour la composante totale = 0.09).

Cette mise à jour conduit à des EST plus faibles en valeur absolue que ceux obtenus par Cortese et al. [2016] comme l'illustrent les *forest plots* à la Figure 2.2 : par exemple l'EST calculé pour la composante totale évaluée par les parents est de -0.23 alors que la réplication de Cortese et al. [2016] résumée à la Table 2.2 menait à un EST de -0.32.

La détection de biais dans cette mise à jour s'est déroulée en deux temps : une analyse visuelle des *funnel plots* présentés à la Figure 2.3 puis le test d'Egger. Aucun biais ne s'est immiscé dans les évaluations des enseignants (p -value du test d'Egger = 0.543) contrairement à ce qu'avaient trouvé Cortese et al. [2016] (p -value = 0.042). Cortese et al. [2016] avaient expliqué ce résultat non pas par la présence d'un biais de publication, mais par le fait que les plus petites études sont de moins bonne qualité et donc n'arrivent pas à montrer l'efficacité du NFB. En effet, en cas de biais de publication, les résultats des petites études auraient été en faveur du NFB.

En ce qui concerne les évaluations des parents, le *funnel plot* paraît asymétrique, or le test d'Egger contredit cette observation (p -value = 0.264). Cela pourrait s'expliquer par le faible poids associé à ces études.

Enfin, l'ajout de ces sept nouvelles études donne plus de puissance statistique aux résultats, notamment pour l'analyse des sous-groupes menée par Cortese et al. [2016].

En effet, pour la sous population suivant un protocole NFB standard [Arns et al., 2014], Cortese et al. [2016] avait noté que les PBlind observaient une différence statistiquement significative entre le NFB et les groupes contrôles en faveur du NFB. Alors que cette tendance a été confirmée avec l'ajout de Baumeister et al. [2016] et Strehl et al. [2017] dans Bussalbi et al. [2019], la mise à jour présentée ici rend presque tous les EST initialement significatifs non significatifs ; désormais seul l'EST de la composante totale évaluée par les parents l'est encore (p -value = 0.047). L'EST calculé pour la composante totale évaluée par les enseignants est limite significatif (p -value = 0.053). Ces résultats remettent en question l'efficacité supérieure des protocoles standards.

En ce qui concerne le sous-groupe constitué d'études qui interdisent la prise de médicaments, la seule différence avec [Cortese et al., 2016] est la perte de significativité statistique de l'EST de la composante hyperactivité évaluée par les parents (p -value = 0.062).

2.3.2 Hétérogénéité des études

Même si les méta-analyses regroupent les études répondant toutes aux critères d'inclusion comme, par exemple, ceux énoncés à 2.2.3, elles diffèrent tout de même sur de nombreux points, comme le nombre de sessions et la durée du traitement, limitant la fiabilité de leurs résultats comme le souligne [Alkoby et al., 2017]. Afin de pallier cette hétérogénéité, l'analyse peut cibler une population plus précise, comme ce qui a été effectué lors de l'analyse des sous populations. Cependant, ce genre de restriction souffre d'une puissance statistique plus faible et regroupe encore des études fortement hétérogènes. En effet, même si seuls les protocoles SCP, TBR et SMR sont utilisés, ils restent intrinsèquement différents et il est probable que leur efficacité ne soit pas la même.

Par ailleurs, le matériel d'acquisition utilisé, ainsi que le traitement du signal varient d'une étude à l'autre, or ces points sont sans doute centraux dans la performance du traitement.

De plus, aucun consensus n'existe quant au nombre de sessions ou à la durée du traitement, ainsi ces choix sont très variables d'une étude à l'autre et peu souvent questionnés, bien qu'ils soient centraux dans la théorie de l'apprentissage.

2.4 Importance de la mise à jour des méta-analyses

Afin d'éviter de tomber dans le *p-hacking*, c'est à dire faire en sorte d'obtenir un résultat significatif, il est important de ne pas arrêter la collecte de données et donc de figer les résultats une fois que la *p*-value de l'EST passe sous le seuil de signicativité de 0.05 [Head et al., 2015; Coffman and Niederle, 2015]. En effet, l'arrêt prématuré des résultats peut générer un biais : par exemple la *p*-value peut être significative à un instant t avec un certain nombre de sujets inclus, puis perdre sa significativité statistique à $t + 1$ et ne plus la retrouver. Ces deux configurations sont représentées avec des données factices à la Figure 2.4.

Les conclusions d'une méta-analyse pourraient donc souffrir de l'arrêt prématuré des analyses lorsque la *p*-value de l'EST n'est pas stable. Ainsi, la méta-analyse de Cortese et al. [2016] a déjà fait l'objet de deux mises à jour, dont les différences vont être comparées par la suite : la première a été publiée en 2019 [Bussalib et al., 2019] et la deuxième est présentée en 2.2.3.

Mise en évidence de l'importance des mises à jour

L'importance de mettre à jour les méta-analyses est étudiée en deux temps :

- par l'évolution de l'EST calculé sur la composante totale au fur et à mesure de l'inclusion des études selon leur année de publication ; l'intervalle de

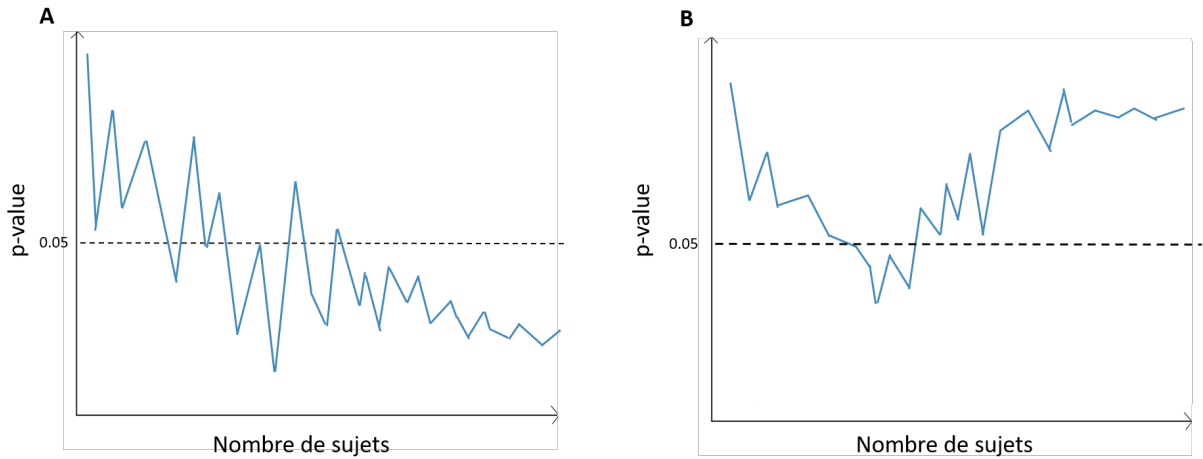


FIGURE 2.4 – Exemples d'évolution des p -value des EST factices au fur et à mesure que la taille de l'échantillon augmente. En **A**, la p -value se stabilise sous le seuil de significativité statistique suite à l'ajout de nouveaux sujets ; en **B**, la p -value passe temporairement sous le seuil de significativité puis s'en éloigne suite à l'ajout de nouveaux sujets. Le seuil de significativité statistique à 5% est représenté en pointillés noirs.

confiance à 95% est aussi représenté,

- par l'évolution de la p -value de ces EST ; l'intervalle de confiance à 95% des p -values est obtenu avec les p -values de chaque étude incluse. Une fois cet intervalle, noté CI, obtenu, ses bornes sont ajustées de façon à prendre en compte le fait que la p -value est une valeur bornée entre 0 et 1 [Mandelkern et al., 2002] :

$$CI = \begin{cases} [max(0; \text{borne inférieure}); max(0; \text{borne supérieure})], \\ [min(1; \text{borne inférieure}); min(1; \text{borne supérieure})]. \end{cases} \quad (2.26)$$

Les courbes sont tracées pour les évaluations des parents et des enseignants. Le nombre de sujets par étude (NS) et cumulatif (NSC) sont précisés sur les axes des abscisses en haut de chaque graphique. Les NSC sont obtenus de la façon suivante : $NSC(k) = NSC(k - 1) + NS(k)$, avec k l'indice de l'étude incluse allant de 1 à K , avec K le nombre total d'études incluses.

Par ailleurs, la méta-analyse de Cortese et al. [2016] ainsi que ses mises à jour sont symbolisées sur les tracés par une étoile si l'EST est statistiquement significatif ou par un carreau sinon : les symboles rouges correspondent aux résultats de Cortese et al. [2016], les verts à ceux de Bussalbi et al. [2019] et les violets à ceux de la mise à jour décrite en 2.2.3. Les méta-analyses étant généralement publiées quelque temps après la fin de sélection des études, leur année de publication est représentée grâce à un point de la même couleur que le carreau ou l'étoile relié au résultat de la méta-analyse par une droite en pointillés.

L'axe des ordonnées a été inversé de façon à ce que les points en hauteur correspondent aux méta-analyses les plus favorables au traitement.

Analyse des courbes de l'évolution de la taille d'effet totale

La Figure 2.5 représente l'évolution de l'EST calculé grâce aux évaluations des parents (en **A**) et des enseignants (en **B**) sur la composante totale en fonction de l'année de publication des études satisfaisant le critère d'inclusion établi par [Cortese et al., 2016] et énoncé en 2.2.3.

Les EST obtenus grâce aux évaluations des parents (**A** de la Figure 2.5) varient de façon importante lorsque peu d'études sont incluses notamment entre 2006 et 2009, puis commencent à se stabiliser à partir de 2015, même si la valeur absolue des EST continue à diminuer. Les intervalles de confiance soulignent également cette stabilisation : ils diminuent avec l'augmentation du nombre d'études incluses.

Cette tendance est également visible du côté des enseignants (**B** de la Figure 2.5). Toutefois, à partir de fin 2018 la valeur absolue des EST augmente.

Les mêmes graphiques sont tracés à la Figure 2.6 mais cette fois en incluant seulement les études suivant un protocole standard [Arns et al., 2014].

Dans le cas des études satisfaisant la définition du protocole standard, les EST calculés à partir des évaluations des parents (**A** de la Figure 2.6) diminuent avec le temps, tout comme leur intervalle de confiance à 95%. En ce qui concerne les EST obtenus grâce aux enseignants (**B** de la Figure 2.6), un faible nombre d'études est inclus et alors que dans les cas précédents la diminution de l'EST ne menait pas à une perte de significativité statistique parmi les méta-analyses, ici la mise à jour de Cortese et al. [2016] présentée en 2.2.3 conduit à un EST non significativement en faveur du NFB.

Analyse des courbes de l'évolution de la p -value des tailles d'effet totales

L'évolution des p -value de ces EST en fonction des études incluses est ensuite étudiée et présentée à la Figure 2.7 pour l'intégralité des études et à la Figure 2.8 pour les études suivant un protocole standard [Arns et al., 2014].

Tout d'abord, en ce qui concerne l'évolution de la p -value en incluant l'intégralité des études pour les évaluations des parents (**A** de la Figure 2.7), on remarque que les p -values sont toutes en-dessous du seuil de significativité à 5% sauf en 2009. Après 2009, les p -values sont plutôt stables et leur intervalle de confiance à 95% diminue, ce qui laisse penser que nous sommes plutôt dans le cas **A** illustré à la Figure 2.4.

Quant à l'évolution des p -values présentée en **A** de la Figure 2.7, bien que toujours non statistiquement significative, elle n'est pas stable.

Enfin, en ce qui concerne le groupe standard, les fluctuations des p -values

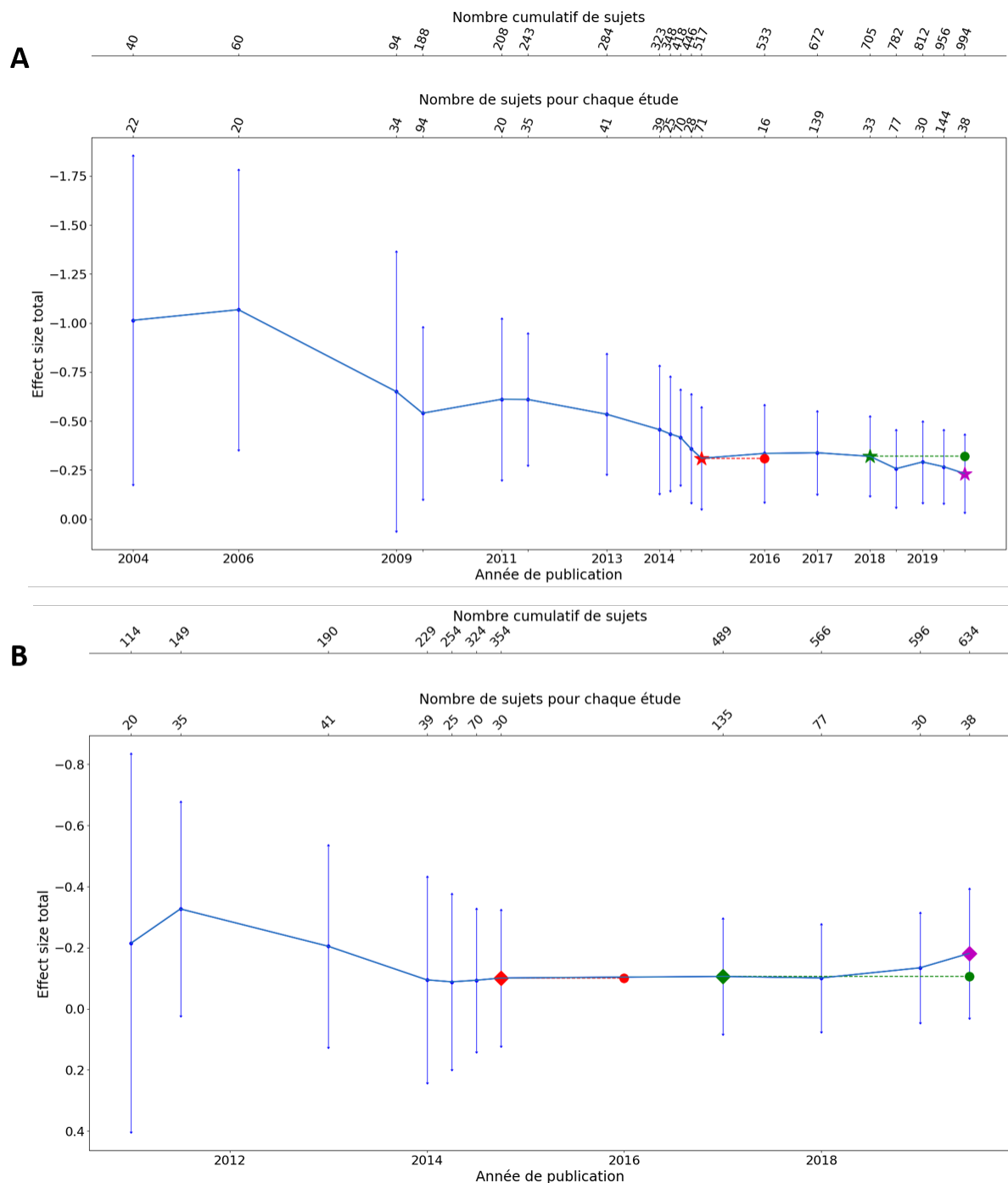


FIGURE 2.5 – Evolution des EST et de leur intervalle de confiance à 95% au fur et à mesure de l'ajout des études satisfaisant le critère d'inclusion de Cortese et al. [2016] pour les évaluations des parents en **A** et des enseignants en **B** sur la composante totale. Les résultats de [Cortese et al., 2016] sont représentés par des symboles rouges, ceux de ? en vert et ceux présentés en 2.2.3 en violet. Les étoiles indiquent un EST statistiquement significatif, un losange un EST non statistiquement significatif. L'année de publication de ces méta-analyses est représentée grâce à un point de la couleur d'intérêt relié au résultat de la méta-analyse par une droite en pointillés. Plus la valeur absolue de l'EST est élevée plus le traitement est efficace. Le nombre de sujets inclus par étude et cumulatif est indiqué sur les axes des abscisses supérieurs. La manière dont est calculé ce dernier est décrite en 2.4

sont moins importantes aussi bien pour les évaluations des parents que celles des enseignants. Cependant, les p -values sont à la limite du seuil de significativité

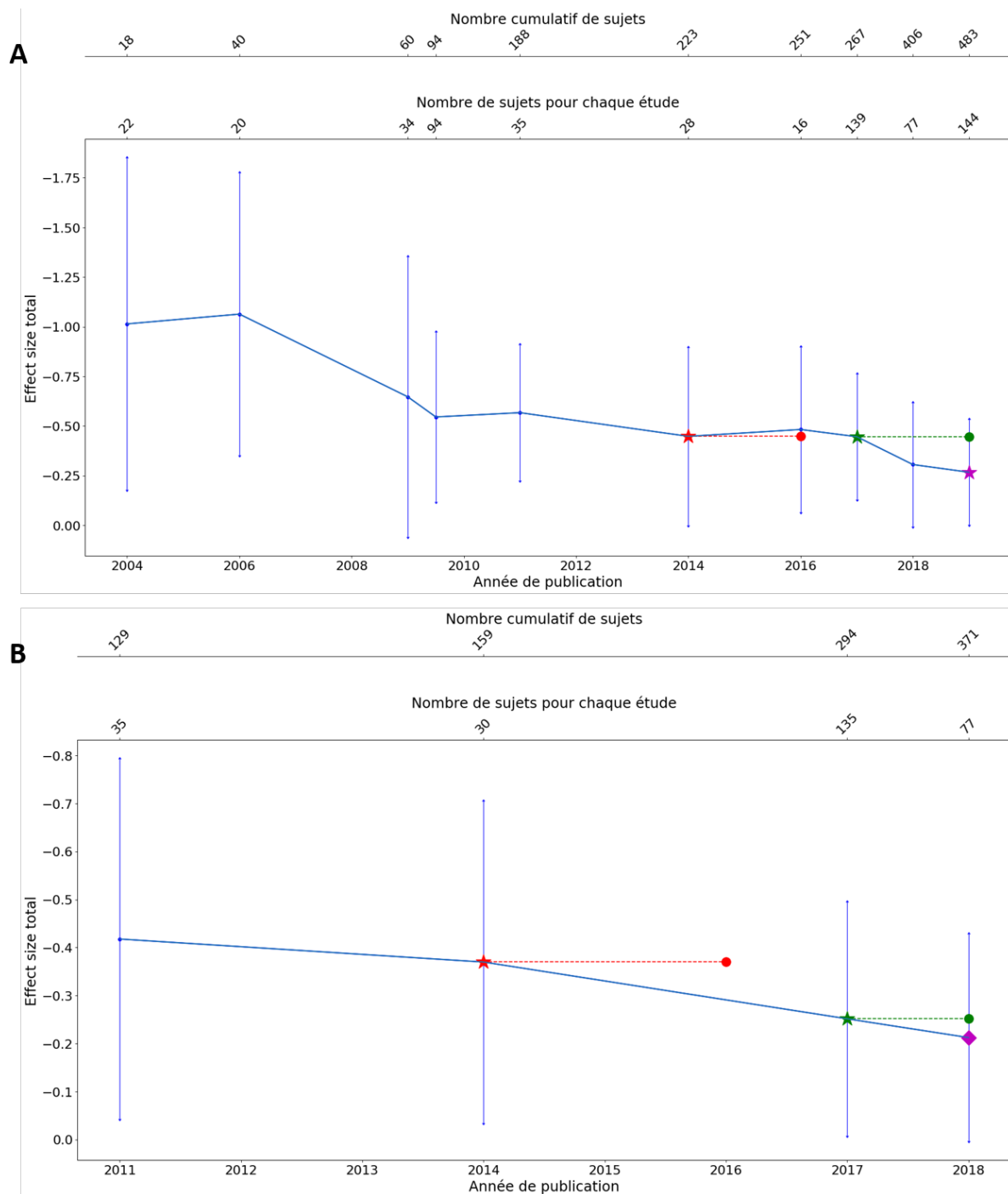


FIGURE 2.6 – Evolution des EST et de leur intervalle de confiance à 95% au fur et à mesure de l'ajout des études suivant un protocole standard pour les évaluations des parents en **A** et des enseignants en **B** sur la composante totale. Les résultats de [Cortese et al., 2016] sont représentés par des symboles rouges, ceux de Bussalbi et al. [2019] en vert et ceux présentés en 2.2.3 en violet. Les étoiles indiquent un EST statistiquement significatif, un losange un EST non statistiquement significatif. L'année de publication de ces méta-analyses est représentée grâce à un point de la couleur d'intérêt relié au résultat de la méta-analyse par une droite en pointillés. Plus la valeur absolue de l'EST est élevée plus le traitement est efficace. Le nombre de sujets inclus par étude et cumulatif est indiqué sur les axes des abscisses supérieurs. La manière dont est calculé ce dernier est décrite en 2.4

dans les deux cas lors de la dernière mise à jour.

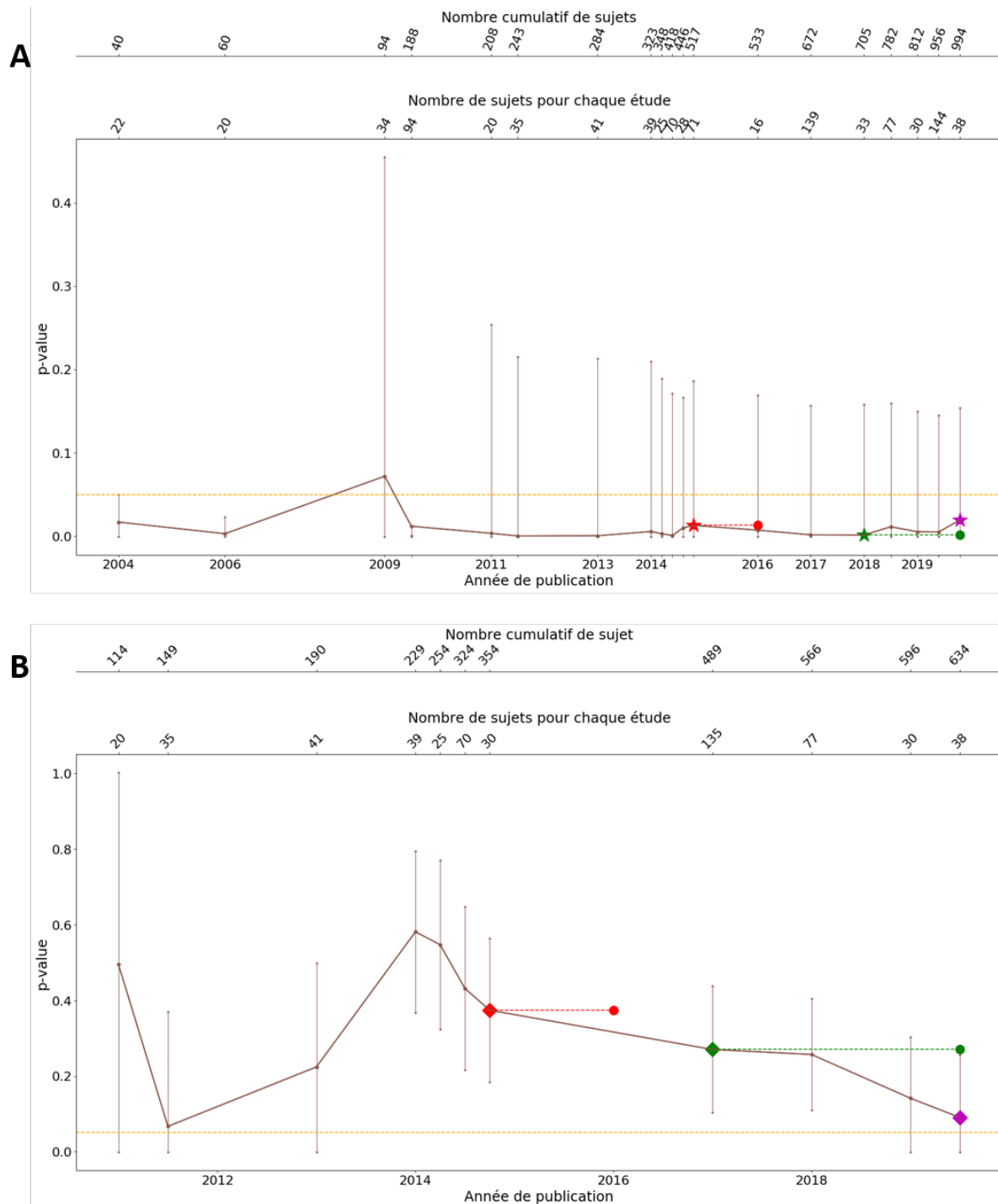


FIGURE 2.7 – Evolution de la p -value des EST et de leur intervalle de confiance à 95% au fur et à mesure de l'ajout des études satisfaisant le critère d'inclusion de Cortese et al. [2016] pour les évaluations des parents en **A** et des enseignants en **B** sur la composante totale. Les résultats de [Cortese et al., 2016] sont représentés par des symboles rouges, ceux de ? en vert et ceux présentés en 2.2.3 en violet. Les étoiles indiquent un EST statistiquement significatif, un losange un EST non statistiquement significatif. L'année de publication de ces méta-analyses est représentée grâce à un point de la couleur d'intérêt relié au résultat de la méta-analyse par une droite en pointillés. Le nombre de sujets inclus par étude et cumulatif est indiqué sur les axes des abscisses supérieurs. La manière dont est calculé ce dernier est décrite en 2.4 Le seuil de significativité à 5% est représenté par des pointillés orange.

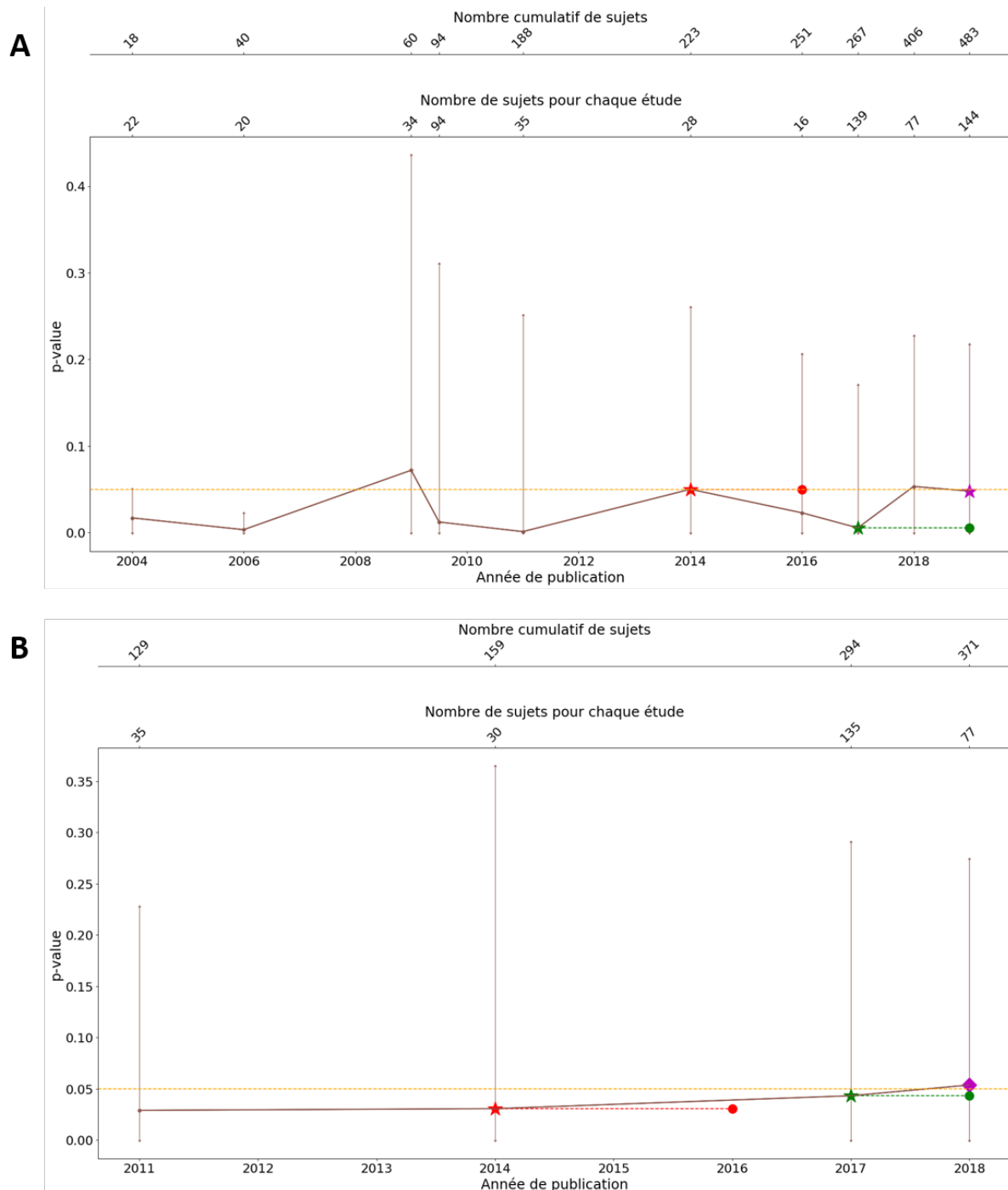


FIGURE 2.8 – Evolution de la p -value des EST et de leur intervalle de confiance à 95% au fur et à mesure de l'ajout des études répondant à la définition du protocole standard pour les évaluations des parents en **A** et des enseignants en **B** sur la composante totale. Les résultats de [Cortese et al., 2016] sont représentés par des symboles rouges, ceux de Bussalbé et al. [2019] en vert et ceux présentés en 2.2.3 en violet. Les étoiles indiquent un EST statistiquement significatif, un losange un EST non statistiquement significatif. L'année de publication de ces méta-analyses est représentée grâce à un point de la couleur d'intérêt relié au résultat de la méta-analyse par une droite en pointillés. Le nombre de sujets inclus par étude et cumulé est indiqué sur les axes des abscisses supérieurs. La manière dont est calculée ce dernier est décrite en 2.4. Le seuil de significativité à 5% est représenté par des pointillés orange.

Discussion sur ces courbes

Les courbes d'évolution des EST au fur et à mesure de l'inclusion de l'ensemble des études (Figure 2.5) illustrent le fait que l'EST devient de plus en plus stable avec le temps. Par ailleurs, la diminution des intervalles de confiance à 95% montre que les résultats deviennent de plus en plus précis.

Les courbes correspondant à l'évolution de l'EST dans le sous-groupe des études suivant un protocole standard (Figure 2.5) sont moins stables du fait du nombre limité de sujets et d'études incluses.

Alors que les EST commencent dans l'ensemble à se stabiliser, leur p -value continue de fluctuer (Figure 2.7 et Figure 2.8) même si sur la totalité des études elles s'accordent depuis 2011 sur la conclusion de Cortese et al. [2016] : les évaluations des parents (**A** de la Figure 2.7) sont significativement en faveur du NFB au contraire des évaluations des enseignants (**B** de la Figure 2.7). Toutefois, la p -value de la dernière mise à jour de Cortese et al. [2016] tend vers la significativité statistique, il serait donc intéressant de continuer à mettre à jour ce travail pour déterminer si cette p -value devient significative et le reste.

Dans le cas de l'évolution des p -values dans le sous-groupe des études suivant un protocole standard (Figure 2.8), la dernière mise à jour de Cortese et al. [2016] conduit à des résultats limite significatifs pour les évaluations des parents (**A**) et à une perte de significativité statistique pour celles des enseignants (**B**), remettant en question les conclusions de Cortese et al. [2016] et Bussalbi et al. [2019] quant à l'efficacité des protocoles standards.

La p -value est très largement utilisée en méta-analyse, mais est également de plus en plus contestée du fait qu'elle soit trop souvent mal interprétée [Halsey, 2019]. En effet, la p -value mesure le degré de compatibilité entre les données et l'hypothèse nulle, et non la probabilité que l'hypothèse nulle soit vraie.

Ainsi, les chercheurs se tournent de plus en plus vers les statistiques bayésiennes, avec des outils comme le rapport de vraisemblance (*likelihood ratio* en anglais) [Dormuth et al., 2016], ou le facteur de Bayes [Morey et al., 2016; Rouder and Morey, 2011]. L'évolution de ces valeurs avec l'inclusion de nouvelles études serait intéressante à étudier.

Pour conclure, la mise à jour des méta-analyses peut avoir une incidence sur la significativité statistique de la p -value, notamment dans le cas où peu d'études et donc de sujets sont inclus. Continuer à inclure de nouvelles études permet d'augmenter la puissance statistique des calculs : en particulier les études les plus récentes sur l'efficacité du NFB appliqué aux enfants TDAH comportent de plus en plus de sujets.

Enfin, étant donné l'essor de l'utilisation du NFB appliqué aux enfants TDAH,

au moment de la publication d'une méta-analyse de nouvelles études satisfaisant ses critères d'inclusion sont déjà disponibles comme illustré aux figures 2.5 et 2.6. Ainsi, à peine publiés, les résultats de la méta-analyse sont déjà obsolètes, ce qui appelle à leur mise à jour.

Cette approche qui consiste à affiner les résultats à l'aide de nouvelles observations est l'une des bases de l'inférence bayésienne.

2.5 Conclusion

Cette partie s'est concentrée sur un outil couramment utilisé pour évaluer l'efficacité du NFB : la méta-analyse. Ce travail a été l'occasion de mettre au point un package *open-source* pour effectuer une méta-analyse en toute transparence et aider à la réplication ou à la mise à jour de ses résultats.

Ce package a justement été utilisé pour répliquer la méta-analyse de Cortese et al. [2016] dont certains points ont été discutés par la communauté scientifique. Il se trouve que les choix effectués par Cortese et al. [2016] s'avèrent anodins mais illustrent la complexité de l'émergence de la preuve clinique.

Cette réplication a également mis en évidence l'importance de la disponibilité des données cliniques afin d'effectuer des calculs précis et de conclure de façon plus fiable quant à l'efficacité d'un traitement.

Il est important de noter qu'une méta-analyse ne rend compte de l'efficacité d'un traitement qu'à un moment donné : étant donné le nombre important d'études évaluant l'efficacité du NFB sur les enfants TDAH, mettre à jour les méta-analyses est nécessaire, c'est pourquoi un tel travail a été entrepris ici. Par ailleurs, l'analyse de l'évolution de l'EST et des p -values associées en fonction de l'ajout des études montrent que ces métriques tendent à se stabiliser depuis quelques années mais ne convergent pas encore, appelant à de nouvelles mises à jour.

Les résultats des méta-analyses donnent une idée de l'efficacité du traitement étudié, même s'il faut garder à l'esprit que les études incluses diffèrent sur différents points, ce qui est notamment vrai pour le NFB. Afin de pallier cette faiblesse, des sous-groupes peuvent être étudiés comme ici avec le groupe d'études suivant un protocole standard et celui d'études où les enfants ne prennent pas de médicaments.

Cependant cette approche ne gomme pas toutes les différences entre les études. Or, ces différences ont sans doute une influence sur l'efficacité du NFB, ainsi nous proposons dans le chapitre suivant de tirer avantage de cette hétérogénéité entre les études pour tenter de mettre en évidence les facteurs techniques

et/ou cliniques qui pourraient avoir une influence sur la performance du NFB.

Chapitre 3

Identification des facteurs influençant l'efficacité du Neurofeedback

Chapitre 4

Analyse de la distribution d'un
marqueur de l'attention au sein
d'une population d'enfants TDAH

Bibliographie

- P.-M. Aggensteiner, D. Brandeis, S. Millenet, S. Hohmann, C. Ruckes, S. Beuth, B. Albrecht, G. Schmitt, S. Schermuly, S. Wörz, et al. Slow cortical potentials neurofeedback in children with adhd : comorbidity, self-regulation and clinical outcomes 6 months after treatment in a multicenter randomized controlled trial. *European child & adolescent psychiatry*, 28(8) :1087–1095, 2019.
- 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 : 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 Attention Disorder*, 2014. doi : 10.1177/1087054712446173.
- M. Arns, H. Heinrich, and U. Strehl. Evaluation of neurofeedback in adhd : the long and winding road. *Biological psychology*, 95 :108–115, 2014. doi : 10.1016/j.biopsycho.2013.11.013.
- 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 : 10.1007/s00787-011-0208-y.
- E. M. Balk, A. Earley, K. Patel, T. A. Trikalinos, and I. J. Dahabreh. Empirical assessment of within-arm correlation imputation in trials of continuous outcomes. 2012.
- S. Baumeister, I. Wolf, N. Holz, R. Boecker-Schlier, N. Adamo, M. Holtmann, M. Ruf, T. Banaschewski, S. Hohmann, and D. Brandeis. Neurofeedback training effects on inhibitory brain activation in adhd : A matter of learning? *Neuroscience*, 2016. doi : 10.1016/j.neuroscience.2016.09.025.
- O. M. Bazanova, T. Auer, and E. A. Sapina. On the efficiency of individualized theta/beta ratio neurofeedback combined with forehead emg training in adhd children. *Frontiers in human neuroscience*, 12 :3, 2018.

- 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 : 10.1007/s10484-006-9001-y.
- 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 : 10.1038/srep31178.
- M. Borenstein, L. V. Hedges, J. Higgins, and H. R. Rothstein. *Introduction to meta-analysis*. Wiley, 2009.
- A. Bussalb. Meta-analysis statistical tools. <https://github.com/AuroreBussalb/meta-analysis-statistical-tools>, 2019.
- A. Bussalb, M. Congedo, Q. Barthélemy, D. Ojeda, E. Acquaviva, R. Delorme, and L. Mayaud. Clinical and experimental factors influencing the efficacy of neurofeedback in adhd : a meta-analysis. *Frontiers in psychiatry*, 10 :35, 2019.
- 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 : 10.3389/fnhum.2014.00943.
- Cochrane Collaboration and others, Copenhagen, Danemark. Review manager (revman)[computer program]. version [5.1] copenhagen : The nordic cochrane centre, 2011.
- Cochrane Group. <https://www.cochrane.org/>.
- L. C. Coffman and M. Niederle. Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives*, 29(3) :81–98, 2015.
- 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 : 10.1097/01.CHI.0000070245.24125.B6.
- C. Conners. Conners 3rd edition : Manual : Multi-health systems. *North Tona-wanda, NY, MHS Assessments*, 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.

- 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 : 10.1016/j.jaac.2016.03.007.
- C. R. Dormuth, K. B. Fillion, and R. W. Platt. Likelihood ratio meta-analysis : new motivation and approach for an old method. *Contemporary clinical trials*, 47 :259–265, 2016.
- M. Egger, G. D. Smith, M. Schneider, and C. Minder. Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109) :629–634, 1997.
- 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 : 10.4088/PCC.11r01336.
- 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 : 10.1111/j.1469-7610.2008.02033.x.
- L. G. Halsey. The reign of the p-value is over : what alternative analyses could we employ to fill the power vacuum ? *Biology letters*, 15(5) :20190174, 2019.
- M. L. Head, L. Holman, R. Lanfear, A. T. Kahn, and M. D. Jennions. The extent and consequences of p-hacking in science. *PLoS biology*, 13(3) :e1002106, 2015.
- 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 : 10.1016/j.biopsych.2003.11.013.
- J. P. Higgins and S. Green. *Cochrane handbook for systematic reviews of interventions*, volume 4. John Wiley & Sons, 2011.
- 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 : 10.1026/0942-5403.18.2.95.
- G. James, D. Witten, T. Hastie, and R. Tibshirani. An introduction to statistical learning. volume 112, chapter 8, pages 303–332. Springer, 2013.

- M. Linden, T. Habib, and V. Radojevic. A controlled study of the effects of eeg biofeedback on cognition and behavior of children with attention deficit disorder and learning disabilities. *Applied Psychophysiology and Biofeedback*, 21(1) :35–49, 1996.
- M. Mandelkern et al. Setting confidence intervals for bounded parameters. *Statistical Science*, 17(2) :149–172, 2002.
- 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 : 10.1016/j.biopsycho.2013.10.008.
- 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 : 10.3389/fnhum.2014.00906.
- J.-A. Micoulaud-Franchi, F. Salvo, S. Bioulac, and T. Fovet. Neurofeedback in attention-deficit/hyperactivity disorder : Efficacy. *Journal of the American Academy of Child & Adolescent Psychiatry*, 2016. doi : 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 : 10.1007/s00787-018-1116-1.
- 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.
- I. Moreno-García, S. Meneres-Sancho, C. Camacho-Vara de Rey, and M. Servera. A randomized controlled trial to examine the posttreatment efficacy of neurofeedback, behavior therapy, and pharmacology on adhd measures. *Journal of attention disorders*, 23(4) :374–383, 2019.
- R. D. Morey, J.-W. Romeijn, and J. N. Rouder. The philosophy of bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, 72 :6–18, 2016.
- S. B. Morris. Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2) :364–386, 2008. doi : 10.1177/1094428106291059.

- J. N. Rouder and R. D. Morey. A bayes factor meta-analysis of bem's esp claim. *Psychonomic Bulletin & Review*, 18(4) :682–689, 2011.
- E. S. Shapiro. Academic skills problems fourth edition workbook. Guilford Press, 2010.
- E. Shereena, R. Gupta, C. Bennett, K. Sagar, and J. Rajeswaran. Eeg neurofeedback training in children with attention deficit/hyperactivity disorder : A cognitive and behavioral outcome study. *Clinical EEG and neuroscience*, page 1550059418813034, 2019.
- 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 : 10.1176/appi.ajp.2012.12070991.
- 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 : 10.1097/DBP.0000000000000009.
- J. A. Sterne, A. J. Sutton, J. P. Ioannidis, N. Terrin, D. R. Jones, J. Lau, J. Carpenter, G. Rücker, R. M. Harbord, C. H. Schmid, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj*, 343 :d4002, 2011.
- 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 : 10.3389/fnhum.2017.00135.
- 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.