UNIVERSITY OF COPENHAGEN
FACULTY OR HEALTH AND MEDICAL SCIENCES

[Double click here to insert image]

Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depressive Disorder

PhD Thesis

Cheng-Teng Ip

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen October 30th 2020

PhD title: Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depressive Disorder

Author: Cheng-Teng Ip, MSc in Cognitive Psychology
Work place: Department of Experimental Medicine, H. Lundbeck A/S, Valby, Denmark & Neurobiology Research Unit, Department of Neurology, Copenhagen University Hospital Rigshospitalet, Denmark
Academic institution: Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

Principal supervisor: Professor Gitte Moos Knudsen, Department of Clinical Medicine, University of Copenhagen, Denmark
Primary co-supervisor: Senior Scientist Søren Christensen, Department of Experimental Medicine, H. Lundbeck A/S, Valby, Denmark
Co-supervisor: Professor Sandor Beniczky, Department of Clinical Neurophysiology, University of Aarhus, Denmark
External assessor: Associate professor Bastian Epp, Department of Health Technology, Technical University of Denmark

Date of submission: Oct 30th, 2020
Date of defense: Jan 29th, 2021

Assessment committee:
Chairperson: Poul Jennum
International assessors: Martin Brunovsky, Martijn Arns

Table of contents

THESIS SUMMARY (IN ENGLISH) I

THESIS SUMMARY (IN DANISH) III

LIST OF PUBLICATIONS V

ACKNOWLEDGEMENTS VII

ABBREVIATIONS IX

INTRODUCTION 1

MAJOR DEPRESSIVE DISORDER 2

Instruments for criteria in treatment response 4

Pathophysiology 5

Monoamine–deficiency hypothesis 5

Neurotrophin dysregulation 6

Hypothalamic–Pituitary–Adrenal axis dysfunction 7

Current treatment approach and its effectiveness 7

Pharmacological drugs 8

Psychotherapy 9

ECT 9

rTMS 10

EEG 11

The Physiological basis of EEG 12

EEG electrode positioning 14

Limitations of EEG 15

Pharmaco-EEG 16

EEG and the Serotonergic system 18

EEG diagnostic and predictive biomarkers in MDD 19

Alpha power and alpha asymmetry 20

Theta and other bands 20
VIGALL 21
ERP measures 22
Other measures 23

THE RATIONALE OF THE NEUROPHARM STUDY 23

AIMS AND HYPOTHESES 25

Aims and hypotheses of study Ⅰ 25

Aims and hypotheses of study Ⅱ 26

Aims and hypotheses of study Ⅲ 26

METHODS 27

Study Ⅰ: Pre-intervention test re-retest reliability of EEG/ERP 27

Participants 27
Procedures 27
EEG recording 28
EEG data prepossessing and analysis 29

Study Ⅱ and Study Ⅲ: EEG biomarkers from NeuroPharm trial 31

Study cohort and treatment 31
Clinical measures and treatment response 33
EEG recording 33
EEG data analysis 33
Study Ⅱ 33
Study Ⅲ 36

Statistical analysis 38

Study Ⅰ 38
Study Ⅱ 38
Study Ⅲ 39

RESULTS AND DISCUSSION 40

Study Ⅰ 40

Is there significant difference across different recording sessions? 40
Is EEG reproducible among four pre-intervention sessions? 44

Study Ⅱ **49**

Study Ⅲ **54**

Vigilance dysregulation in MDD 57

VIGALL as predictor for clinical outcome 58

Treatment effects on VIGALL parameter 58

METHODOLOGICAL CONSIDERATIONS **62**

Only male subjects were included when assessing the reliability of EEG parameters **62**

Inconsistent results when different scales or clinical criteria were assessed **62**

The variance of Alpha peak frequency between participants **63**

Drug effect or the condition of the patient **64**

CONCLUSION AND FUTURE RESEARCH PERSPECTIVES **65**

REFERENCES **67**

APPENDIX **81**

Thesis summary (in English)

Major depressive disorder (MDD) is a heterogeneous disorder with potentially diverse pathophysiological mechanisms. This diversity may account for the observation that far from all patients benefit from the same treatment. Consequently, identification of MDD subgroups would allow for treatment stratification which could improve clinical trial outcomes and lead to a precision medicine strategy in patients. However, previous studies examining whether electroencephalography (EEG) can predict the effect of antidepressant treatment have included low sample size, and independent replication are needed to confirm these findings.

The purpose of this PhD work was to a) examine the test-retest reliability of an EEG battery in healthy males who were given different antidepressants, b) to validate EEG candidate biomarkers that have previously shown some predictive value in our own cohort of unmedicated patients with moderate to severe MDD (NeuroPharm Trial) c) explore the effect on EEG measures after 8 weeks of selective serotonin reuptake inhibitor/ serotonin noradrenalin reuptake inhibitor (SSRI/SNRI) treatment.

In study I, we investigated the test-retest reliability of an EEG battery in 32 healthy males who were given four different antidepressant regimens. We compared baseline EEG recordings from the four interventions to assess whether EEG/ERP (event-related potentials) were stable over time. We found that middle frequency bands (θ, α and β) of continuous EEG were highly reliable while evoked power of task-related potentials was less stable. Furthermore, though the reliability of ERP measures in general was lower compared to power measures, large components such as P300 and Pe still exhibited fair to excellent reliability. Our results support that these EEG/ERP parameters are reliable over a three-week interval.

In study II, we aimed to replicate previously reported EEG predictive biomarkers for treatment outcome by using an independent cohort of 91 antidepressant-free outpatients and 35 healthy controls. We found that only 2 out of 6 chosen biomarkers could be partially validated; both of which involved alpha asymmetry. The results indicate that measurement of alpha asymmetry carries information that improves prediction of treatment efficacy.

In study III, used EEG data from the NeuroPharm data to determine the predictive value of vigilance regulation, we found that patients with MDD showed a hyperstable EEG-wakefulness regulation compared to healthy controls, replicating prior work. Treatment responders showed faster decline in vigilance regulation in comparison to non-responders at pretreatment. Furthermore, patients with good treatment response after 8 weeks of SSRI/SNRI treatment had their EEG-wakefulness regulation patterns reverted to look more like that of controls.

These findings support that EEG vigilance measures adds value diagnostically as well as in predicting treatment outcome in patients with MDD. Overall, the low cost of and methodological simplicity of EEG makes it a good tool for the optimization of patient stratification in future clinical trials and may even have value when choosing drug treatment for MDD patients.

Thesis summary (in Danish)

Depression er en klinisk heterogen sygdom som sandsynligvis forårsages af forskellige sygdomsmekanismer og tilgrundliggende forstyrrelser. Denne heterogenitet kan muligvis forklare, hvorfor langt fra alle patienter har effekt af den samme antidepressive behandling. Såfremt man an identificere undergrupper af depression vil man dels kunne målrette fremtidige kliniske forsøg bedre og dermed på sigt få en præcisionsmedicinsk tilgang til behandling af depression. Tidligere studier, der har undersøgt om elektroencefalografi (EEG) kan forudsige effekten af antidepressiv behandling har imidlertid kun inkluderet få patienter og der er behov for uafhængige studier til at bekræfte disse fund.

Formålet med dette PhD projekt var a) at undersøge test-retest reliabiliteten af EEG mål hos raske mænd, der modtog forskellige typer antidepressiv medicin, b) i en ny kohorte af patienter med en moderat til svær depression (NeuroPharm Trial) at validere EEG biomarkører, der tidligere er påvist at forudsige effekten af antidepressiv medicinog c) undersøge effekten af 8 ugers farmakologisk behandling med selektive serotonin reuptake inhibitors / serotonin noradrenalin reuptake inhibitor (SSRI/SNRI).

I Studie I undersøgte vi test-retest reliabiliteten af et EEG batteri i 32 raske mænd, som fik fire forskellige typer af antidepressiv medicin. Vi sammenholdt baseline EEG målinger fra de fire interventioner for at estimere, hvorvidt EEG og ERP (event-related potentials) målingerne var stabile over tid. Vi fandt at midt-frekvensbånd (θ, α og β) for kontinuerlig EEG havde høj reliabilitet mens evokeret power fra task-baserede potentialer var mindre stabil. Endvidere fandt vi at selvom ERP mål generelt havde lavere reliabilitet sammenholdt med EEG power mål, så havde store ERP komponenter som P300 og Pe moderat til høj reliabilitet. Disse EEG/ERP målinger er altså robuste over et tre-ugers interval.

I Studie II søgte vi at bekræfte tidligere fund af udvalgte EEG biomarkører som prediktorer for behandlingsrespons i en uafhængig kohorte af 91 medicin-fri patienter med depression og 35 raske forsøgspersoner. Vi fandt at 2 ud af de 6 potentielle EEG biomarkører delvist kunne bekræftes; begge markører involverede alpha asymmetri. Disse resultater tyder på at fund af alpha asymmetrikan forudsige behandlingseffekten af antidepressiv medicin.

I studie III undersøgte vi i NeuroPharm kohorten et EEG mål for regulering af *vigilance* (grad af vågenhed) for dets prediktive værdi af antidepressiv behandling. I tråd med tidligere fund fandt vi, at patienter med depression udviste over-stabil regulering af *vigilance* i forhold til raske kontroller. Patienter med god effekt af den medicinske behandling havde et hurtigere fald i *vigilance* regulering end dem, der ikke havde. Endvidere så vi, at EEG vågenhedsreguleringen normaliseredes hos de patienter der udviste god symptomlindring efter 8 ugers behandling med SSRI/SNRI .

Tilsammen understøtter disse fund at EEG kan bidrage med diagnostisk og prædiktiv værdi i behandlingen af patienter med depression og har EEG et godt potentiale til at underinddele patienter med depression i fremtidige behandlingsforsøg. Endvidere kan EEG måske i fremtiden anvendes til at guide den medicinske behandling af patienter med depression.

List of publications

The thesis is based on the following published article and manuscripts:

1. **Ip, C.T**., Ganz, M., Ozenne, B., Sluth, L.B., Gram, M., Viardot, G., l'Hostis, P., Danjou, P., Knudsen, G.M., Christensen, S.R., 2018. Pre-intervention test-retest reliability of EEG and ERP over four recording intervals. Int. J. Psychophysiol. https://doi.org/10.1016/j.ijpsycho.2018.09.007
2. **Ip, C.T**., Olbrich, S., Ganz, M., Ozenne, B., Köhler-Forsberg, K., Dam, V., Beniczky, S., Jørgensen, M., Frokjaer, V., Søgaard, B., Christensen, S., Knudsen., G., Pretreatment qEEG biomarkers for predicting pharmacological treatment outcome in Major Depressive Disorder: Independent validation from the NeuroPharm study. European Neuropsychopharmacology (submitted)
3. **Ip, C.T**., Ganz, M., Dam, V., Rüesch, A., Köhler-Forsberg, K., Jørgensen, M., Frokjaer, V., Søgaard, B., Christensen, S., Knudsen., G., Olbrich, S., NeuroPharm Study: EEG wakefulness regulation as a biomarker in MDD. Journal of Psychiatric Research (submitted)

Related articles not included in thesis:

1. Köhler-Forsberg, K., Jørgensen, A., Dam, V., Stenbæk, D., Fisher, P., **Ip, C.-T**., Melanie, G., Poulsen, H., Giraldi, A., Ozenne, B., Jørgensen, M., Knudsen, G., Frokjaer, V., 2020. Predicting treatment outcome in Major Depressive Disorder using serotonin 4 receptor PET brain imaging, functional MRI, cognitive-, EEG-based and peripheral biomarkers: a NeuroPharm open label clinical trial protocol. Front. Psychiatry. https://doi.org/https://doi.org/10.3389/fpsyt.2020.00641
2. Dam, V.H., Stenbæk, D.S., Köhler-Forsberg, K., **Ip, C**., Ozenne, B., Sahakian, B.J., Knudsen, G.M., Jørgensen, M.B., Frokjaer, V.G., 2020. Hot and cold cognitive disturbances in antidepressant-free patients with major depressive disorder: A NeuroPharm study. Psychol. Med. https://doi.org/10.1017/S0033291720000938
3. **Ip, C**., Wang, H., Fu, S., 2017. Relative expertise affects N170 during selective attention to superimposed face-character images. Psychophysiology. https://doi.org/10.1111/psyp.12862

4. Wang, H., **Ip, C**., Fu, S., Sun, P., 2017. Different underlying mechanisms for face emotion and gender processing during feature-selective attention: Evidence from event-related potential studies. Neuropsychologia. https://doi.org/10.1016/j.neuropsychologia.2017.03.017

5. Wang, H., Sun, P., **Ip, C**., Zhao, X., Fun, S., 2015. Configural and featural face processing are differently modulated by attentional resources at early stages: An event-related potential study with rapid serial visual presentation. Brain Res. https://doi.org/10.1016/j.brainres.2015.01.017

Acknowledgements

Above all, I would like to express my greatest appreciation to my supervisor professor Gitte Knudsen for her wisdom and mentorship throughout the project. She has always been very supportive and comprehensive of the sudden and sometimes odd thoughts I have had. I always admire her work efficiency and I am convinced that without her I would not have accomplished the work. Special thanks to professor Olaf Paulson who offered me the chance to come to this country and introduce me to this amazing lab. I would also like to thank my co-supervisor Søren Christensen for the underserved trusts he has on me and giving me the opportunity to bridge results between industrial and academy. I am thankful to Sandor Beniczky and Bastian Epp who helped me to initiate the project and set up the EEG experiment.

I am extremely grateful to Dr. Sebastian Olbrich for your assistance and guidance in many ways and your constructive feedbacks along these years. You have been so welcoming to me since the first day I knocked on your office's door and later introducing me to your lab, your family and also the tornado of the field. I really enjoy the many and productive discussions we have had. Your inspiration, enthusiasm and encouragement have strongly spurred me, and I am amazed to see the fruitful collaboration we have even when I once had little faith in myself.

I am thankful to my lab mates who are associated with NeuroPharm, Camilla Colding-Jørgensen, Ida Marie Brandt, Elizabeth Landman, Søren Larsen, Asbjørn Poulsen. Thank you for the coordination of every tests, scanning and recording for the participants. In addition, I would like to thank Vibe Frøkjær and Martin Jørgensen for their knowledge and expertise in this study. Special thanks to Gerda Thomsen, Svitlana Olsen and Lone Freyr for the recruitment of healthy controls and the preparation of the blood samples. In particular, I am grateful to Vibeke Dam, Kristin Forsberg and Brice Ozenne who accompanied me in the past four years at NRU, and also

being such nice friends to me. I must also thank Melanie Ganz who helped me with data processing and guided me though every step to become a dedicated researcher. To all other NRUs, I enjoy working with all of you and as well as all the small talks along the corridor and in the kitchen.

I would like to acknowledge Margrethe Wiehe, Dorthe Givard, Birgit Tang, Peter Jensen for your responsive help and always answering my inquiries.

I am grateful to my close friends, Xinyao, Emma, Weiran, Yingchuan and Xiaohang whom I shared many fond memories with. I dedicate this dissertation to my old professor Shimin, for embarking me to this journey and encouraging me to pursue my career in science.

Finally, I owe much to my parents, for their understanding of my decision to pursue a PhD in a far-away country. To my sister ChengIan and my brother PakLon who have been nice accompanies to grow up with. Your support and unconditional love have shaped who I am today.

Most of all, I am fully indebted to my husband Gert Hansen, who is very patient with all my complaints over so many things and being very supportive about almost all the decisions I made. Thanks to your tolerance and your dedication to our family. Last but not least, this work is also dedicated to my daughter, Nora Hansen who has been waking me up at night and who screamed during the day when I felt drowsy at work. I love you with all my heart and hope that lovely, innocent smile can always be on your face.

Abbreviations

5-HT	5-hydroxytryptamine
5-HTTLPR	Serotonin-transporter-linked polymorphic region
ADHD	Attention Deficit Hyperactivity Disorder
ANOVA	Analysis of variance
APF	Alpha peak frequency
ASSR	Auditory steady state response
ATR	Antidepressant treatment response
BDI	Beck Depression Inventory
BDNF	Brain-derived Neurotrophic Factor
BMI	Body mass index
CGI	Clinical Global Impression scale
CSD	Current source density
CSF	Cerebrospinal fluid
DA	Dopamine
ECT	Electroconvulsive treatment
EEG	Electroencephalography
EOG	Electrooculography
EPSP	Excitatory postsynaptic potential
ERP	Event-related Potentials
FAA	Frontal alpha asymmetry
FDA	Food and Drug Administration
fMRI	Functional magnetic resonance imaging
GAD	Generalized anxiety disorder
HD EEG	High-density EEG
HDRS	Hamilton Depressive scale
HPA	Hypothalamic-pituitary-adrenal
ICC	Intra-class Correlation
IPSP	Inhibitory postsynaptic potentials
IPT	Interpersonal psychotherapy

ISI	Interstimulus-interval
iSPOT-D	International Study to Predict Optimised Treatment - in Depression
LC	Locus coeruleus
LDAEP	Loudness dependence auditory evoked potential
LORETA	Low-resolution electromagnetic tomography
MDD	Major Depressive Disorder
MDRS	Montgomery Åsberg rating scale
MNI	Montreal Neurological Institute
MOAIs	Monoamine oxidase inhibitors
NE	Norepinephrine
NNT	Number-needed-no-treat
NRIs	Norepinephrine reuptake inhibitors
OCD	Obsessive-compulsive Disorder
OFC	Orbitofrontal cortex
PET	Positron emission tomography
pg/ad ACC	Perigenual and anterior dorsal Anterior Cingulate Cortex
QEEG	Quantitative EEG
rACC	Rostral anterior cingulate cortex
ROC	Receiver operator curve
ROIs	Regions of Interests
rTMS	Repetitive transcranial magnetic stimulation
SEMs	Slow eye movements
SNRI	Serotonin noradrenalin reuptake inhibitor
SOA	Stimulus onset asynchrony
SSRI	Selective serotonin reuptake inhibitor
STAR*D	Sequenced Treatment Alternatives to Relieve Depression
TCAs	Tricyclic antidepressants
TRD	Treatment-resistant Depression
VIGALL	Vigilance Algorithm Leipzig
WFSBP	World Federation of Societies of Biological Psychiatry
WHO	World Health Organization

Introduction

Major depressive disorder (MDD) is a heterogeneous disease which comprises wide-ranging biological causes. Currently, it often takes long time to settle the right treatment, given their trial-and-error method and as a result, patients often experience a protracted disease course (Baskaran, Milev, & McIntyre, 2012). They are thus far from satisfactory, especially for patients with a history of previous treatment failure. The treatment approaches we have available today include different classes of antidepressants, psychotherapy, electroconvulsive treatment (ECT) and repetitive transcranial magnetic stimulation (rTMS), we urgently need biomarkers to classify depressed patients and determine the best treatment approach for each patient based on their specific profile. Biomarkers are used to stratify patients into clinically meaningful subgroups based on biological characteristics, e.g. neurophysiological patterns. The idea is that patients who share similar profiles will also respond similarly to the same treatment, allowing clinicians to optimize treatment selection based on previous experience with other patients from the same subgroup.

During the past decades, researchers have attempted to use different neuroimaging tools to identify biomarkers that could facilitate the treatment selection in MDD. These tools include such as positron emission tomography (PET, McGrath et al., 2013), functional magnetic resonance imaging (fMRI, Miller et al., 2013), as well as genetic analysis (Tansey et al., 2012). Unfortunately, none of these measures have been adopted for clinical use due to the lack of validation, insufficient predictive value, and implementation difficulty (Baskaran et al., 2012; Widge et al., 2018). Meanwhile, neurophysiological measures have served as biomarkers in various disorders, including mood disorders (Lemere, 1936) as early as the first half of the 20th. Among them, electroencephalography (EEG), a monitoring technique for direct and ongoing neural activity, has great potential to work as a biomarker in a clinical environment due to its

low-costs, high temporal resolution of directly assessed mass-neuronal activity and its relatively ease of implementation.

Researchers recently have identified several EEG biomarkers to be associated with treatment outcome in MDD (Arns et al., 2016; Olbrich et al., 2016; Pizzagalli et al., 2018; Pizzagalli et al., 2001; Spronk et al., 2011; Wade & Iosifescu, 2016). While the application of EEG as predictive biomarkers seems promising, this application faces similar obstacles as with other imaging tools: low effect sizes, small samples, a lack of independent replication of the results in other study cohorts and research groups, its usage of different response criteria and the neglect of negative results (Widge et al., 2018). To address the different parameters in MDD prediction studies, the current study proposes to do the following in three parts:

a) To examine the pre-intervention test-retest reliability of an EEG battery in healthy males who were given different antidepressants.
b) To assess, using our own dataset (NeuroPharm trial), the effectiveness of EEG candidate biomarkers that have previously shown predictive power.
c) To explore the drug effect on EEG measures after 8 weeks of selective serotonin reuptake inhibitor/ serotonin noradrenalin reuptake inhibitor (SSRI/SNRI) treatment.

Major Depressive Disorder

Major Depressive Disorder (MDD) is a medical condition with which the patients shows a depressed mood and reduced interest in daily activities (see the DSM-IV manual, American Psychological Association (APA), 2013). It has been ranked by WHO as the world leading causes of disability (World Health Organisation, 2017) with a lifetime prevalence of 4.7% (Ferrari et al., 2013). Women are more vulnerable to the disease and with higher prevalence (World Health Organisation, 2017). For most patients, MDD is a chronic episodic disorder and

the recurrent course of MDD often require long-term clinical care (Fava & Kendler, 2000). Therefore, the loss of work capacity as a result of MDD and the consequent health-care costs constitute a tremendous social burden, which is made even more severe for treatment-resistant patients.

To be diagnosed with MDD, patients display a change of mood and loss of interest or pleasure in any activities, often accompanied by psychophysiological changes including a loss of appetite, disturbed sleep, impaired psychomotor speed and attention, fatigue, feelings of guilt, contemplation, and suicidal thoughts. However, given the heterogeneity in MDD, depressed patients often display wide variation in clinical symptoms (Fava & Kendler, 2000), which may be why not all patients benefit from the same treatment (Spronk et al., 2011). According to the International Study to Predict Optimised Treatment - in Depression (iSPOT-D), one of the largest multisite studies of predictive biomarkers in MDD, only 62% patients responded to the first-line drug and 46% of patients reached full remission (Saveanu et al., 2015). The large heterogeneity in MDD may explain why not all patients benefit from the same treatment (Spronk et al., 2011), and undoubtedly yet more challenges to the current treatment.

The current treatment strategy requires a lengthy observational period (commonly six to eight weeks), after which non-responders are shifted to augmented treatment or to other interventions, which also takes a long waiting period. The treatment cycle is repeated until the successful treatment is found, if any, or until the patient spontaneously remits. Worse still, patients who experience early treatment failure are less likely to achieve remission and would have higher relapse rates (Rush et al., 2006). Biomarkers have become a possible solution to the impasse since they could help tailor more accurate treatments for depressed patients with specific profiles. Biomarkers have thus been studied by recent researches that aim at increasing the efficacy of antidepressants (reviewed by Olbrich & Arns, 2013).

Instruments for criteria in treatment response

Depression rating scales were introduced in psychopharmacology in the 1960s and have been used for assessing the treatment response. The common instruments include Hamilton Depressive scale (HDRS) (Hamilton, 1967), Montgomery Åsberg Rating Scale (MDRS) (Montgomery & Asberg, 1979), self-rating Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and Clinical Global Impression scale (CGI) (Guy, 1976). The HDRS and MDRS are widely used in quantifying depressed severity with substantial similarity (Kearns et al., 1982) while BDI is the most commonly used self-rating scale with less agreement for core depressive symptoms compared to MDRS (Svanborg & Åsberg, 2001). CGI was designed to provide global impression of a depressed patient and has clinical advantage when illustrating the total score compared to HDRS and MDRS (Leucht et al., 2017). Although translations between depressive scales are available (Leucht, Fennema, Engel, Kaspers-Janssen, & Szegedi, 2018; Riedel et al., 2010), the use of different criteria and cut-offs by researchers limits the ability to generalize results between studies. For example, a patient could be categorized as a responder in one study but as a non-responder in a study that applies different response criteria. In addition, specific depressive symptoms might be neglected on the sum score on the depressive scales. Study has shown that the total score of HDRS does not provide sufficient discriminability to qualify a drug as an antidepressant as it is unable to differentiate between treatment and placebo (Bech, Kajdasz, & Porsdal, 2006). This implies that the treatment response may be overlooked by the gross score. As such, it could be that a subscale of HDRS as $HDRS_6$ could provide better sensitivity for capturing treatment changes compared to $HDRS_{17}$ for $HDRS_6$ focuses on the symptoms of sleep, anxiety and appetite (Bech et al., 2010, 2006; Østergaard, Bech, & Miskowiak, 2016). These rough depression rating scales could in turn hindered the process of biomarker development. As identified biomarkers rely on the sum scores from e.g. the $HDRS_{17}$ could be less sensitive since we can't be sure of what core symptoms have

the patients being response to. Another related issue is the cut-off value of treatment response. Responders are often defined as a 50% reduction from baseline to endpoint, though being questioned of how it became the golden standard (Bandelow, 2006), the cut-off value has been established as an efficient standard for HDRS, MADRS and BDI (Leucht et al., 2018; Riedel et al., 2010). From clinical experience, patients who were classified as responders, informed by the golden standard, still suffering from residual symptoms (Kennard et al., 2006). The development of clinical biomarkers relies heavily on these instruments, these inconsistent practices undoubtedly limit the possibilities for the biomarkers to be in clinical use. It remains unclear how the biomarkers relate to the specific depressive symptoms.

Pathophysiology

Many theories have been proposed to explain the pathophysiology of MDD including monoamine-deficiency hypothesis, neurotrophic dysregulation, hypothalamic-pituitary-adrenal (HPA) disruption, genetics, altered glutamatergic neurotransmission and dysfunction in specific brain structures (for a review see Belmaker & Agam, 2008). However, despite the growing understanding of neurobiological mechanism in MDD pathophysiology, none of the proposed mechanism can satisfactorily explain all aspects of MDD aetiology. In the following section, I focus on three of the most dominant theories: monoamine-deficiency hypothesis, neurotrophic dysregulation and HPA disruption.

Monoamine–deficiency hypothesis

The monoamine–deficiency hypothesis was proposed in 1950 based on clinical observations and animal studies (Freis, 1954). Patients who received reserpine, an antihypertensive agent that causes depletion in monoamine neurotransmitters, would develop depressive symptoms (Schildkraut, 1965). The induced symptoms could be reversed by the cessation of reserpine treatment (Carlsson, Shore, & Brodie, 1957). Moreover, treating patients with iproniazid, an

agent that produced increased levels of monoamine neurotransmitters, would improvement depressed mood (Crane, 1956). The hypothesis therefore became that there is a deficiency of the monoamine neurotransmitters norepinephrine (NE), 5-hydroxytryptamine (5-HT), and/or dopamine (DA) in the depressed brain (Coppen Alec, Shaw, Herzberg, & Maggs, 1967; Schildkraut, 1965). Based on the depletion theory, the first serotonin-based compounds tricyclic antidepressants (TCAs) and monoamine oxidase inhibitors (MAOIs) were developed to increase the concentrations of monoamine neurotransmitters in vivo and were used effectively to treat depression (Kuhn, 1958; Zeller, Ramachander, & Zeller, 1965). Nevertheless, the depletion theory cannot explain why drugs like buspirone and gepirone also work as antidepressants. These drugs act by decreasing 5-HT release in the brain yet still improve depressive symptoms (Hirschfeld, 2000). It was further challenged by the delay between administration of drugs and antidepressant effects. Typically, improvement of depressive symptoms takes days to weeks while the neurotransmitter concentrations are increased within hours after the intake (Hirschfeld, 2000).). To account for these discrepancies the monoamine-deficiency theory has continually been modified over decades (Goldberg, Bell, & Pollard, 2014; Kafka, 2003). it undoubtedly provides an important biological basis for the aetiology of depression and is still central in the development of pharmacological drugs.

Neurotrophin dysregulation

Neurotrophins constitutes four growth factors including nerve growth factor, brain-derived neurotrophic factor (BDNF), neurotrophin-3 and neurotrophin-4. In particular, downregulated BDNF has been associated with the genesis of depression and suicidal behaviour (Duman, Heninger, & Nestler, 1997; Dwivedi, 2009). Studies have demonstrated that patients with MDD had lower BDNF levels in comparison to healthy controls (see meta-analysis Molendijk et al., 2014; Sen, Duman, & Sanacora, 2008). BDNF levels have also been correlated to depressive

symptom severity; the lower BDNF levels the more severe the depressive symptoms (Karege et al., 2002) which might reflect failure of neuronal plasticity in depression (Gonul et al., 2005). Furthermore, a number of findings have indicated that antidepressants and electroconvulsive therapy (ECT) can increase serum BDNF in depressed patients (Chen, Dowlatshahi, MacQueen, Wang, & Young, 2001; Gonul et al., 2005; Nibuya, Morinobu, & Duman, 1995; Shimizu et al., 2003). One post-mortem study with MDD patients showed increased BDNF expression in hippocampus in patients treated with antidepressants (Chen et al., 2001). Although the role of BDNF pathophysiology has been debated (Egeland, Zunszain, & Pariante, 2015; Ernest Lyons et al., 1999), the implication of BDNF theory in neural plasticity and depression is encouraging (Brunoni, Boggio, & Fregni, 2008; Tao et al., 2020).

Hypothalamic–Pituitary–Adrenal axis dysfunction

Alterations to the hypothalamic–pituitary–adrenal (HPA) axis have been found in stress-related disorder, particularly depression (Pariante & Lightman, 2008). A hyperactive HPA axis in patients with MDD was been first documented in the 1950s (Board, Wadeson, & Persky, 1957)and evidence was accumulated later on (McKay & Zakzanis, 2010; Stetler & Miller, 2011). Furthermore, a hyperactive HPA axis has been found to normalize after successful treatment (Barden, Reul, & Holsboer, 1995; Hennings et al., 2009). It has been proposed that there is a potential link between HPA axis and the serotonergic hypothesis of depression (Tafet & Bernardini, 2003), and targeting the abnormalities of HPA axis is considered to be a promising strategy (Bosker et al., 2004).

Current treatment approach and its effectiveness

Based on the proposed genesis of depression, various treatment approaches are available such as different classes of pharmacological drugs, psychotherapy, and therapies including ECT and

rTMS. The goal of a successful treatment is to reduce depressive symptoms and preferably achieve remission as well as avoid later relapse. Antidepressants are facing two major challenges: the delayed onset of action and low response rates (Bosker et al., 2004). A substantial proportion of patients does not response to the antidepressants they first receive. According to the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, remission and response rates to first-line antidepressant drugs are lower than 30% and 50%, respectively (Rush et al., 2009). These challenges strongly affect patients' compliance and expectation which is why we face heightened need for clinically reliable biomarker and a shift from the current diagnostic approach to a prognostic approach to the prognostic approach (Olbrich & Conradi, 2016).

Pharmacological drugs

The first antidepressants TCAs and MAOIs act by increasing the concentration of NE and 5-HT in vivo; with these drugs have some success in alleviating depressive symptoms they unfortunately also have a series of undesirable side effects (Feighner & Cohn, 1985; Owens, 2004). The resulting side effects are caused by the TCAs and MAOIs acting on non-specific neurotransmitter receptor sites such as NE, serotonin reuptake, adrenergic and muscarinic receptors (Benkert, Gründer, & Wetzel, 1997). Overdose of TCAs even carry the risk of death. Therefore, the second wave of antidepressants, SSRIs and SNRIs, were developed to have more specific target profile (often with a single monoamine system) and thus fewer side effects. Although results from a meta-analysis did not demonstrate a better drug efficacy SSRIs over the "old" antidepressants TCAs, it showed that patients who were treated with TCAs withdrew more due to the significant side effects (MacGillivray et al., 2003). Since then, many different types of antidepressants have been introduced based on the various modes of actions on neurotransmitter levels and can be categorized as: TCA, tetracyclic antidepressants, SSRI, norepinephrine reuptake inhibitors (NRIs), SNRI, MAOIs and other antidepressants (for a pharmacotherapeutic

armamentarium please refer to Bauer et al., 2013). According to the World Federation of Societies of Biological Psychiatry (WFSBP) guidelines (Bauer et al., 2013), the first-line treatment options in treating depression include SSRI, followed by mirtazapine, SNRI and other "second generation" antidepressant. These types of medicine have the advantages of lower side-effects and higher tolerability compared to other pharmacological drugs. TCA and MAOIs are considered to be second- or third-line treatment due to their safety issues.

Psychotherapy

There are seven major types of psychotherapy in treating depression, including cognitive–behaviour therapy, nondirective supportive treatment, behavioural activation treatment, psychodynamic treatment, problem-solving therapy, interpersonal psychotherapy, and social skills training (Cuijpers, van Straten, Andersson, & van Oppen, 2008). These different psychological treatments are equally efficient (see a meta-analysis, Cuijpers, van Straten, Andersson, & van Oppen, 2008). Moreover, studies have shown that depressed patients who received interpersonal psychotherapy (IPT) perform significant better on social functioning with or without pharmacological intervention (Cuijpers et al., 2011; Hollon et al., 2005; Weissman, Klerman, Prusoff, Sholomskas, & Padian, 1981). Also, IPT improves depressive symptoms in a large scale and was suggested to include in the treatment guidelines (Cuijpers et al., 2011).

ECT

ECT is a medical procedure where the brain is electrically stimulated while the patient is under anaesthesia. It is used mainly when pharmacological drugs fail to treat severe depression (Kellner et al., 2012; Lisanby, 2007), but it can also be the first-line treatment choice for severely depressed patients with psychotic features or psychomotor retardation, treatment-resistant depression (TRD) and some patients who require a rapid relief from depression (WFSBP guidelines, Bauer et al., 2013). Although ECT is a blunt procedure, it has better clinical outcome

than other antidepressants (Pagnin, De Queiroz, Pini, & Cassano, 2004). A meta-analytic study compared effects of ECT to other treatments including simulated ECT, placebo, and various antidepressants, demonstrating that ECT is clearly more effective than all other treatments (Pagnin et al., 2004). Moreover, patients who received ECT are found to have a remission rate as high as 75% (Husain et al., 2004). It was suggested that ECT improves depressive symptoms by altering patient's biological entities in many aspects (Wahlund & von Rosen, 2003) such as an increased plasma BDNF (Haghighi et al., 2013). In general, ECT is a highly effective treatment especially for TRD patients even it requires anaesthesia and involves a risk of cognitive impairment (Mathew et al., 2019).

rTMS

rTMS is a focal and non-invasive technology that induces patient's electrical pulses through a TMS coil on the scalp. Without involving a surgical operation, it is an alternative stimulation technique to treat a board range of psychiatric disorders, especially depression (McNamara, Ray, Arthurs, & Boniface, 2001; Slotema, Blom, Hoek, & Sommer, 2010). An increasing amount of evidence has demonstrated that treating patients with prefrontal rTMS could induce antidepressive effects (George et al., 2014; George, Taylor, & Short, 2013; Padberg et al., 2002). Previous studies have investigated the therapeutic effect of rTMS applied to different stimulate regions of prefrontal cortices (Downar & Daskalakis, 2013) and the right lateral orbitofrontal cortex (OFC) (Feffer et al., 2018). The techniques of rTMS have been steadily improved via dosing augmentation, extension of treatment courses, and designing personalized stimulation frequencies for patients, etc. (see review Downar & Daskalakis, 2013). As such, patient's remission and response rate can be comparable to those adopting pharmacological interventions (Downar & Daskalakis, 2013). In general practice, rTMS is typically prescribed for patients who are taking adjunctive pharmacological medicine (Carpenter, Laney, & Mezulis, 2012). A study

showed that there exists an interaction between concurrent medication and the rTMS therapy, indicating that the combination of rTMS and psychostimulants such as amphetamine and armodafinil may result in improved clinical outcome (Hunter et al., 2019). Furthermore, rTMS was approved by the US Food and Drug Administration (FDA) in October 2008 as a treatment of depression and has since then become a promising new solution for treating depression (see meta-analysis Slotema et al., 2010).

EEG

Electroencephalography is a monitoring technique for direct, ongoing mass-neuronal activities. It acquires information through the electrodes placed on the scalp. When cortical tissues are activated, it produces electric currents flowing from different tissues to the brain, which also pass the scalp into the skin. If electrodes are applied to the skin at that time, the currents—only tiny fractions of the flowed currents—can be detected and recorded by the applied electrodes as a potential distribution. The recorded voltage is directly reflective of neural oscillations in the cortex. In general, the recorded signal can be quantified in the dimensions of time, space, and frequency. The dimension of time most directly reflects the oscillating neuronal activity. The EEG signal is in the time frame of tens to hundreds of milliseconds to a few seconds, where most of the cognitive processes occur. EEG has therefore been used as a neurofeedback technique (Groen et al., 2008) as well as in ERP studies (Hillyard, Luck, & Mangun, 1994). With a Fourier transform, EEG signals can be parsed into the frequency domain, which have three characteristics: frequency, power (the strength) and the phase (the timing of the frequency activity). EEG data could also address the ‘where’ question with its localization methods, for instances, instead of relying on the two-dimensional EEG electrode array, the low-resolution electromagnetic tomography (LORETA) provides the distribution of the current source density

(CSD) throughout the full brain volume (Pascual-Marqui, Michel, & Lehmann, 1994; Pascual-Marqui et al., 1999). With neurophysiological rooted hypotheses, this multidimensional information allows us to examine the brain-behaviour relationship.

The first report of human EEG dates to 1929, when a German physician Hans Berger first demonstrated that rhythmic brain activities could be observed on human scalp without craniotomy and he also pioneered Fourier transform to quantify the EEG. He described the observed oscillations of the electrical currents on the scalp as two wave types, one larger with the characteristics of slower time course and one shorter with more rapid one (La Vaque, 1999). Berger later identified them as alpha and beta waves, which, he found, appear alternatively depending on changing arousal levels. Moreover, Berger used EEG as a tool to examine the neuropathology, pharmacological effects, sleep and some psychiatry disorders (La Vaque, 1999). Berger's early research has laid a solid foundation for EEG studies. Today, EEG has been intensively used in exploring the cognitive processes and has also been commonly used in clinical environments to study different psychiatric disorders (Brassen & Adler, 2003; Loo et al., 2003; Mulert et al., 2007).

The Physiological basis of EEG

EEG represents the postsynaptic potentials of the activated cells flow from the summated activity of excitatory postsynaptic potential (EPSP) and inhibitory postsynaptic potentials (IPSP) with similar geometric orientation (Figure 1). The action potential, because of its rapidly deceased propagation from the source to the scalp and short-lasting duration, is less likely to be detected by scalp EEG (Brienza & Mecarelli, 2019). However, it can occur in epileptic patients, where many action potentials activated simultaneously and be detected as a "spike" (M. Holmes, 2008).

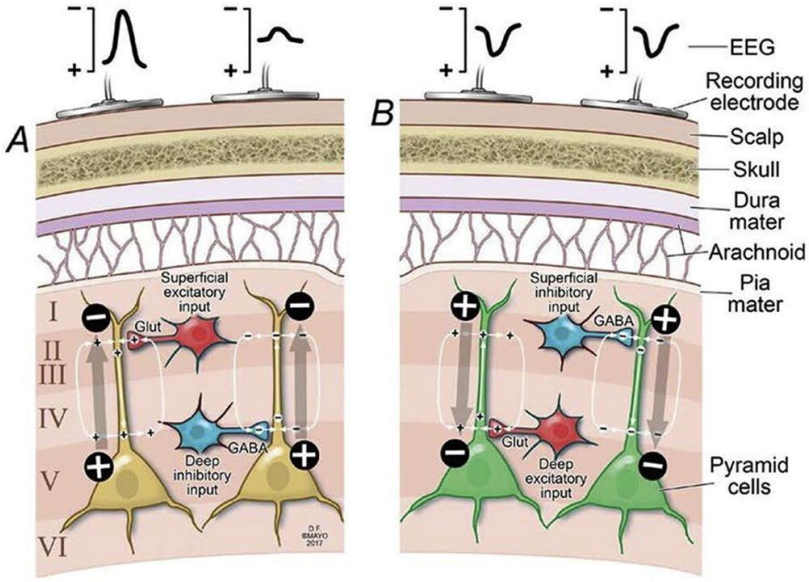


Figure 1 The scalp EEG recorded from the summation of EPSP (A) and IPSPs (B) at postsynaptic sites from a large numbers of pyramid cells with ensembled orientation. The figure is adapted from Feyissa & Tatum (2019).

Scalp EEG signals are generated by the large population of neurons dominated by the activity of pyramidal cells with assembled orientation in the cortical beneath the placed electrode (Murakami & Okada, 2006). Thus, the location and the geometric orientation of the generator play a critical role in shaping EEG signals. Field potentials with radial orientation generated from cortical surface constitute most to the EEG signals while dipoles (generators) with tangential orientation and dipoles in the sulci are less likely to be recorded by a scalp EEG. Therefore, sources from deep brain structures like thalamus, basal ganglia, hippocampus, and brainstem (Holmes & Khazipov, 2007) are difficult to be captured by the electrode placed on the scalp. Another related issue is volume conduction. Volume conduction is mainly caused by cerebrospinal fluid (CSF), which results in a tangential spread of electrical fields (Van Den Broek, Reinders, Donderwinkel, & Peters, 1998). As such, a placed electrode would record ripple signals from the neuronal activity just beneath another electrode. These factors challenge the accuracy of the ‘inverse’ problem which calculates the estimated sources from the observed activities.

The periodic activity observed in raw EEG, also called oscillations, can be considered as the result of the interaction between excitatory and inhibitory neurons. According to the excitatory-inhibitory feedback loop synchronization mechanism, an oscillatory cycle begins with the excitation of a population of neurons which then generates further excitation, until the inhibitory networks within these population are also activated to bring down the population activity. The activity of the interneurons then decreases, allowing the excitation to recover and generate the next 'waxing-and-waning' cycle. This reciprocal behaviour between EPSPs and IPSPs produces the oscillatory process (Wang, 2010).

EEG electrode positioning

The physiological characteristics of the scalp EEG signal has implied the importance of the electrode positions. In order to provide a standardized measurement from the anatomical electrical current distribution on the skull, the 10-20 International electrode system was established (Jasper, 1958), resulting in a standardized system for clinical EEG (Klem, Lüders, Jasper, & Elger, 1999). The 10-20 International electrode system means that, taking the circumferential line as an example, T7 is 10% distance of the whole line from T9, C3 is 20 % from T7, and 20 % for Cz-C3, C4-Cz, T8-C4 and lastly, 10% for T10-T8 (Figure 2, more details please refer to Klem et al., (1999). It also provides a naming system to the electrode positions which makes the communication across EEG studies available (Seeck et al., 2017). Later, the extended 10% and high-density array of 5% were developed, providing electrode numbers up to 345 (Oostenveld & Praamstra, 2001). A high-density EEG (EEG) array of 256 channels is shown in Figure 2. Such an electrode cap is mainly used in electrical source imaging for presurgical evaluation, especially when estimating the sources of epileptic activity is in focus (Seeck et al., 2017). However, the nomenclature of scalp electrode in the HD EEG array is in strong needs for standardization (Heine, Dobrota, Schomer, Wigton, & Herman, 2020).

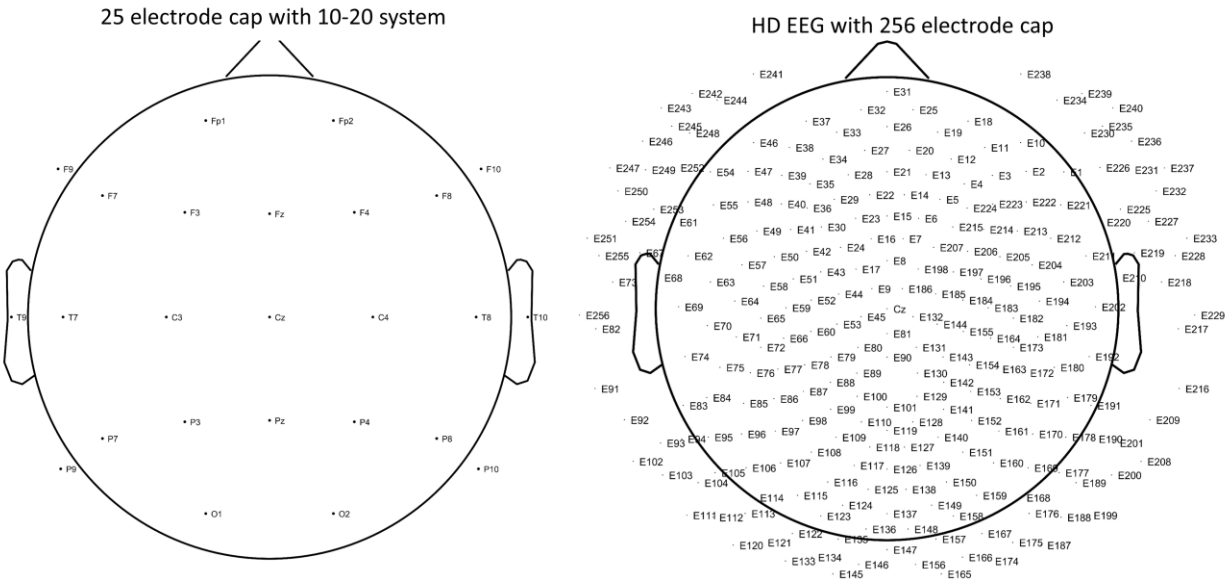


Figure 2 EEG electrode system from different position montages. The left panel refers to the standardized International 10-20 electrode system while the right panel demonstrates the high-density scalp EEG array from HydroCel Sensor Net system (EGI, Inc., Eugene, OR).

Limitations of EEG

Low spatial information and EEG artefacts are considered the major limitations of EEG. As mentioned earlier, due to the nature of volume conduction in the brain, EEG has a limited capacity in explaining activities coming from deep brain sources and separating the brain's functional activities from other sources. Moreover, the accuracy of EEG signals may also be affected by the number of electrodes applied to the scalp. This number ranges from a basic array containing 25 electrodes to a HD EEG arrangement of up to 256 electrodes (Figure 2)—and even the latter number may not be sufficient (Seeck et al., 2017). Electrical activities between applicable electrodes could only be estimated using the spline interpolation method from neighbouring electrodes (Perrin, Pernier, Bertrand, & Echallier, 1989). Therefore, the signal projected to topographical presentation may not reflect the true signals generated by the particular areas of the cortices. The exact location of the neuronal generator is affected by factors

such as the different conductivity of brain tissues, non-homogeneity properties of the skull, and different geometric orientation of cortical sources (Teplan, 2002). Although there are several proposed methods, such as dipole fitting, adaptive and nonadaptive distributed-source imaging method, to address the 'inverse' problem, the accurate estimation of the location, orientation, and magnitude of the signals could still be difficult.

Artefacts in EEG signals are defined as unwanted and non-neurophysiological signals (Tandle & Jog, 2015). The most common artefacts contaminating EEG signals are those generated by eye movements, saccades and blinks, muscular movements, electrocardiogram, perspiration, and other addictive random noise arising from instrumental noise and other physiological generators (Romo-Vázquez et al., 2012). Study on epilepsy have demonstrated that the contamination from muscle artefacts have hindered the identification of epileptic tissues (Ren, Gliske, Brang, & Stacey, 2019). These artefacts pose challenges to the analysis and interpretation of the EEG signal.

Pharmaco-EEG

Pharmaco-EEG provides a non-invasive tool for CNS intoxications by monitoring neurophysiological signals to aid treatment selection. In psychiatric and industrial settings, Pharmaco-EEG often measures resting-state EEG and is quantified by the spectral power. It has two main implications: monitoring drug-induced EEG alterations and predicting clinical response to a pharmacological medicine (see review by Mucci, Volpe, Merlotti, Bucci, & Galderisi, 2006). Pharmaco-EEG has been intensively used in studying the diagnosis and treatment follow-ups including immediate (Iinuma, Tamahashi, Otomo, Onuma, & Takamatsu, 1978) and long-term drug effects (Mikati, Trevathan, Krishnamoorthy, & Lombroso, 1991) on epilepsy patients (Meador et al., 2016; Puspita, Soemarno, Jaya, & Soewono, 2017; Smith, 2005). In a review of Höller, Helmstaedter, & Lehnertz (2018), most of the antiepileptic agents function by supressing the paroxysmal epileptic form of discharges and through an overall

attenuation of power. The most frequently observed antiepileptic effects involves the slowing down of EEG signals, including an increased power of low frequencies (δ and θ) and a decrease in higher frequencies (Höller et al., 2018). These EEG changes were also observed in the medication treatment of schizophrenia (Koenig, Hernandez, & Rieger, 2016). Recently, the focus has been drawn to the study of depression and antidepressants (reviewed by Alhaj, Wisniewski, & McAllister-Williams, 2011). A number of studies revealed an increase in the θ and a decrease in α band on drug-free MDD patients who were administered with buspirone, an agonist of 5-HT_{1A} receptor (Anderer, Saletu, & Pascual-Marqui, 2000; McAllister-Williams & Massey, 2003; Hamish McAllister-Williams, Massey, & Fairchild, 2007). Moreover, in a four-way crossover study on the effects of desipramine and two regimens of duloxetine on healthy male compared to placebo, results showed that duloxetine prolonged the onset latency of rapid eye movement sleep (Chalon et al., 2005). This provides direct evidence through EEG that the sleep disturbances in MDD patients can be improved by antidepressants.

According to Mucci et al., (2006), EEG changes after medication is even a better clinical manifestation than serum levels because the drug remnant stays in cells even after the patient's serum concentration drops to zero. However, the EEG measures indexing drug activity have not been established. Traditional pharmaco-EEG indices such as power analysis might not be sufficient to capture the highly complex interactions between neuromodulatory systems in the brain. The use of more sophisticated and sensitive measures is required. For instance, combining absolute and relative powers, topographic parameters, and sources localization, the Vigilance Algorithm Leipzig (VIGALL) has been developed to identify and quantify different functional brain states during wakefulness (Ulrich Hegerl, Wilk, Olbrich, Schoenknecht, & Sander, 2012; Olbrich et al., 2015). Several studies have documented that this theoretical framework has predictive value for the treatment efficacy of SSRI antidepressants (Olbrich et al., 2016; Schmidt et al., 2017). Overall, a mechanism measuring changes in EEG responses to drug use is yet to be

established before it can be used for aiding treatment selection and for assessing drug efficacy and safety.

EEG and the Serotonergic system

The serotonergic system plays a role in various biological functions including mood, appetite, sleep-wake regulation, and cognition (Bravo et al., 2013). Using EEG, preclinical studies have found evidences showing the effect of serotonergic modulation of the cholinergic neurons on cortical activity (Fumoto et al., 2010). Serotonin medication produced a dose-dependent increase of the low frequency α activity while decreased the high frequency γ activity in the cortical EEG (Bronzino, Brusseau, Morgane, & Stern, 1972; Cape & Jones, 1998). Primary auditory cortices were found to have the richest concentration of serotonin levels as well as the highest synthesis rates (Hegerl, Gallinat, & Juckel, 2001). As such, loudness dependence of auditory evoked potential (LDAEP) has been suggested to reflect the serotonin levels in vivo (Gallinat, Bottlender, Juckel, & Stotz, 2000; U. Hegerl et al., 2001). In animal studies, LDAEP has been consistently shown to be inversely linked to the brain serotonin levels, with a higher LDAEP reflecting a lower level of serotonin and vice versa (Juckel, Hegerl, Molnár, Csépe, & Karmos, 1999; Juckel, Molnár, Hegerl, Csépe, & Karmos, 1997). However, in human studies there was less agreement (Nathan, Segrave, Phan, O'Neill, & Croft, 2006; O'Neill, Croft, & Nathan, 2008). In a recent multimodal study combining EEG and PET results, albeit conducted on a small sample (n = 23), demonstrated a correlation between LDAEP and serotonin-1A and serotonin transporter in the temporal cortex (Pillai et al., 2020). A number of studies have also reported serotonin dysfunction using LDAEP as an indicators in patients with depression (Park, Lee, Kim, & Bae, 2010), schizophrenia (Wyss et al., 2013), anxiety disorder (Park et al., 2010) and OCD (Stein, 2002). However LDAEP might also be modulated by other neuromodulatory systems (O'Neill et al., 2008). In the study of (Carrillo-De-La-Peña et al., 2000), researchers found no

difference in the LDAEP of patients with OCD compared to healthy controls. Moreover, citalopram, an antidepressant that increases the amount of serotonin with SSRI have shown no alterations on LDAEP (Guille et al., 2008). Furthermore, due to the close connection between serotonergic system and arousal, alpha asymmetry has been considered to be associated with the serotonergic system through a right temporoparietal and subcortical pathway (Bruder et al., 2008; Tenke et al., 2011). The increased alpha and alpha asymmetry in SSRI responders might be explained by the low level of serotonin in raphe nuclei and cortical afferents (Bruder et al., 2008). However, direct evidence showing a connection between cortical alpha activity and serotonergic is yet to be found. Future studies with multimodal comparison would be beneficial for investigating if the EEG technique could serve as a surrogate of serotoninergic activity in vivo.

EEG diagnostic and predictive biomarkers in MDD

Even with the same diagnosis, patients with MDD may display widely different symptoms (Fava & Kendler, 2000), and the panacea for this heterogeneous disease is not available yet. The treatment of MDD is often a shared decision made by the patient and the physician based on his/her subjective experiences. Although many different treatment options are available for MDD, the non-response rates to first-line treatment remain high. This calls for objective, reliable and accessible markers to help stratify treatment decisions. So far, systematic efforts have been made to find such predictive biomarkers, including EEG measures, in order to assess the efficacy of antidepressants (Olbrich & Arns, 2013; Spronk et al., 2011; Trivedi et al., 2016; Wade & Iosifescu, 2016; Leanne M. Williams et al., 2011). The following section summarizes findings of studies that measure the effectiveness using electrophysiological profiles as predictive biomarkers to assess different treatments for MDD.

Alpha power and alpha asymmetry

Inconsistent results were reported in previous studies when the absolute and relative alpha powers were being assessed in patient's treatment response and non-response (Mumtaz, Malik, Yasin, & Xia, 2015; Ulrich, Renfordt, Zeller, & Frick, 1984). Clinically defined responders are found to have lower absolute occipital alpha power compared to non-responders (Ulrich et al., 1984). However, the study of Bruder et al., (2008) indicated that fluoxetine responders showed greater pretreatment occipital absolute alpha power compared to non-responders and healthy controls. Also results from the same group, Tenke et al., (2011) assessed the CSD and demonstrated that patients with less alpha CSD were less likely to respond to serotonergic drugs. The hemispheric asymmetry of alpha power (known as "alpha asymmetry") and has been consistently reported to have prognostic value for improving treatment outcome. Intriguingly, alpha asymmetry was found to be gender-specific. Evidence has demonstrated that the response of female non-responders to fluoxetine tended to be characterized by right hemisphere hyperactivity (right alpha less than left) (Bruder et al., 2001). In the same study setting, Bruder et al., (2008) found that responders showed greater activation of the right alpha than the left while non-responders showed an opposite asymmetry. Similarly, alpha asymmetry was also reported by (Arns et al., 2016) as gender-specific biomarker. Female responders with greater right frontal alpha tended to have favourable clinical response to SSRI treatments (Arns et al., 2016).

Theta and other bands

By analysing the conventional power spectrum, decreased relative delta, relative theta, and both absolute and relative beta powers were associated with better treatment response to paroxetine (V. Knott, Mahoney, Kennedy, & Evans, 2000). In particular, patients with decreased pretreatment theta have been reported to have favourable treatment outcomes (rTMS: Arns, Drinkenburg, Fitzgerald, & Kenemans, 2012; Iosifescu et al., 2009; V. J. Knott, Telner, Lapierre,

Browne, & Horn, 1996), though no theta differences in clinical outcomes were reported (Cook et al., 1999). Furthermore, treatment response of MDD patients has also been examined using quantitative EEG (QEEG) with LORETA. Studies that utilized qEEG with LORETA showed that medication responders tended to have elevated theta power in rostral anterior cingulate cortex (rACC) (Korb, Hunter, Cook, & Leuchter, 2009; Mulert et al., 2007; Pizzagalli et al., 2018; Pizzagalli et al., 2001) and elevated medial OFC theta power (Korb et al., 2009) compared to non-responders. Arns & Olbrich, 2014; Jaworska & Protzner, (2013) suggested that these seemingly contradictory findings on theta activities (from scalp EEG and from source localization) came from different signal sources: decreased scalp theta (tonic theta) was associated to treatment non-responders while increased theta ACC (phasic theta) was linked to responders to medication (Arns et al., 2015; Arns & Olbrich, 2014). However, in the largest EEG study on MDD (n = 655) so far, Arns et al., (2015) reported that lower theta is an indicator for venlafaxine-XR responders, but not those to SSRI treatment. It is hard to conclude the predictive value of theta ACC since different interventions have been used in the studies of theta ACC. There might exist treatment specificity in theta activity (Arns et al., 2015).

VIGALL

Dysregulation of sleep and wakefulness is a core symptom of MDD (Nutt, Wilson, & Paterson, 2008; Seifritz, 2001). However, conventional EEG measures have focused on either state of sleep or wakefulness, but these measures are not sensitive to sleep-awake dysregulation, an important symptom of MDD. VIGALL is a powerful tool allowing researchers to evaluate the resting EEG and electrooculogram data from full wakefulness to sleep onset (Ulrich Hegerl et al., 2012; Olbrich et al., 2015). According to the theoretical framework of VIGALL, hyperstable vigilance regulation with less declines toward relaxation is found in patients with MDD and obsessive-compulsive disorder (OCD), while patients with mania and attention deficit

hyperactivity disorder (ADHD) tended to show unstable wakefulness regulation with a rapid drop of vigilance levels (Ulrich Hegerl & Hensch, 2014; Ulrich Hegerl et al., 2012; Olbrich et al., 2013; Olbrich, Sander, Jahn, et al., 2012). Furthermore, Olbrich et al., (2016) reported that treatment responders had a faster decline of wakefulness while Schmidt et al., (2017) reported that patients with a pronounced hyperstable wakefulness regulation were more likely to respond to SSRI treatment. These studies, however, used different methods of recording time, which probably impacted on patients' wakefulness regulation. It is therefore inadequate to draw conclusions from only two studies on the prognostic power of VIGALL on the treatment outcome.

ERP measures

Auditory ERP components such as P300 and LDAEP were promising candidates for predicting patients' response to antidepressants (Iosifescu, 2011). P300 refers to a positivity at 300 ms after stimulus onset and it could be evoked by odd, infrequent stimuli in an auditory oddball paradigm. Patients suffering from depression were found to have smaller amplitude and prolonged latency of P300 compared to healthy controls (Mumtaz et al., 2015). With regard to the treatment response on P300, prior reports showed that patients with smaller pretreatment P300 amplitude tended to be non-responders to bupropion and escitalopram treatments (Jaworska & Protzner, 2013). Although there were some inconsistencies, P300 latency has also been reported to have some value in treatment prediction (Işintaş, Ak, Erdem, Öz, & Özgen, 2012; Jaworska & Protzner, 2013). LDAEP, a means to measure the cortical activity arising from one's auditory cortex, is acquired from the slope of auditory tones of increasing loudness. It has been reversely linked to serotonergic system in vivo (Juckel et al., 1999, 1997). Patients with higher pretreatment LDAEP (lower serotonin levels) were more likely to be treatment responders to SSRI drugs (Gallinat et al., 2000; Juckel et al., 2007; Lee, Yu, Chen, & Tsai, 2005) while the

ones with lower LDAEP tended to have better clinical outcome to SNRI medication (Juckel et al., 2007).

Other measures

Other measures such as antidepressant treatment response (ATR) index and theta cordance have also been studied as predictors for patients' response to antidepressants. ATR consists of prefrontal theta and alpha power from baseline and 1 week after treatment. It was reported to have good accuracy (> .70) in predicting response and remission (see review Iosifescu, 2011). Theta cordance is an EEG measure that combines absolute and relative power. Results demonstrated that patients with a higher theta cordance tend to respond to SSRI and venlafaxine (Bares et al., 2008; Cook, Hunter, Abrams, Siegman, & Leuchter, 2009). However, ATR and theta cordance are both "treatment emergent markers" which reflect changes in EEG measures one week after the treatment and cannot be assessed at baseline, and thus they have limited value in clinical use. More recently, connectivity was found to be associated with patients' clinical response to antidepressants (Lee, Wu, Yu, Chen, & Chen, 2011). The latest findings in combing EEG and deep learning approaches have shown great promise in predicting the efficacy of antidepressants (Uyulan et al., 2020; Wu et al., 2020).

The rationale of the NeuroPharm study

Patients diagnosed with MDD display a wide variety of symptoms in clinical practices (Fava & Kendler, 2000), and the heterogeneity in MDD thus poises great challenge for treating the disease. The current 'trial-and-error' treatment approach is far from satisfactory, giving that only less than 50% of MDD patients achieve sufficient remission from depressive symptoms (Rush et al., 2006). These challenges have motivated many research teams to develop biomarkers in collaboration with psychiatry environments (Leuchter et al., 2009; Trivedi et al., 2016; Williams

et al., 2011). A number of biomarkers have been proposed by these multicentre studies, whilst the direct brain mechanism behind depressive symptoms, antidepressants, and serotonergic system as well as their interaction still remains unclear. Therefore, NeuroPharm, as a novel initiative, aims to not just study patients on a relatively large scale but also to provide insights on depression on the level of neurotransmitter (Köhler-Forsberg et al., 2020). Adopting a cross-disciplinary design, NeuroPharm trial covers topics ranging from neurotransmitter, neuroimaging, cortical activation to cognitive levels to provide opportunities for cross-modal comparisons and thus help understand the mechanism behind depression. EEG, having been proved be a promising biomarker (see review Olbrich & Arns, 2013), has been associated with the serotonergic transmitter system (Pillai et al., 2020). Therefore, this thesis examines the effectiveness of pretreatment EEG as a biomarker and its clinical utility as a part of the NeuroPharm Trial.

Aims and hypotheses

The overall aim of the study is to examine the effectiveness of EEG/ERP as biomarkers for predicting treatment outcome for MDD patients. The study consists of three parts: 1. We first evaluated the general reliability of EEG/ERP parameters on healthy controls to construct the methodological basis of EEG parameters. The existing dataset (reference No. 15835A) was designed for this purpose. 2. We collected an independent dataset (NeuroPharm: Köhler-Forsberg et al., 2020) to directly validate the efficacy of published QEEG biomarkers, chosen from a recent meta-analysis (Widge et al., 2018). The dataset consisted of EEG data of 91 MDD patients and 35 healthy controls. 3. We further examined the diagnostic value of vigilance regulation and, as an exploratory attempt, the treatment effect on the parameters of vigilance regulation in the same dataset.

Aims and hypotheses of study I

In study I, we aimed at examining the test-retest reliability of EEG/ERP parameters on healthy males at four preintervention recording intervals. The purpose of study I was to assess whether the drug's impact on EEG measurements recede after a washout period (18–22 days). Blood serum concentrations were measured to monitor the possible carry-over effect of previous pharmacological drug. We hypothesized that: 1. The test-retest reliability of EEG and ERP measures will demonstrate at least moderate scores across four preintervention recording intervals. 2. The power spectrum of resting EEG will exhibit higher test-retest reliability than the traditional peak-picking ERP measures. 3. Amplitude measures will have higher test-retest reliability compared to peak latency measures.

Aims and hypotheses of study II

Since a recent meta-analysis (Widge et al., 2018) points out that direct replication of the published QEEG biomarkers is missing, we validated those biomarkers with our own dataset (NeuroPharm, Forsberg, 2020). Therefore, the aims of study II were to replicate previous studies as closely as possible, in terms of the methodology, data analysis, criteria of treatment response and the statistical approach. We hypothesized that 1) biomarkers are possible to be validated when utilizing the same response criteria. 2) Biomarkers based on a large sample size is possible to be replicated in the current dataset.

Aims and hypotheses of study III

Vigilance regulation was not analysed in the meta-analysis (Widge et al., 2018) and it has been deemed relevant for the diagnosis of MDD and has an impact on predicting the responses to depression treatment (Olbrich et al., 2016; Schmidt et al., 2017). The aims of the study III were to use the data from the NeuroPharm trial (Köhler-Forsberg et al., 2020), and 1) to replicate the findings of a hyperstable EEG wakefulness regulation in MDD in comparison to healthy controls. 2) to replicate the predictive properties of the VIGALL algorithm with respect to treatment outcome for psychopharmacological interventions using SSRIs and SNRIs. It was hypothesized that 1) MDD patients will show more high vigilance stages and fewer declines toward sleep stages in comparison to healthy controls and 2) that as assessed with the VIGALL algorithm, remitters/responders will show a less stable EEG wakefulness regulation over time.

Methods

Study I was based on an existing dataset (reference No. 1583A) and study II and III were based on the data from NeuroPharm trial (registration number: NCT02869035; Trial paper: Köhler-Forsberg et al., (2020)). Details of the methodology is given in the paper; below is a brief summary of the methods used in the three papers.

Study I: Pre-intervention test re-retest reliability of EEG/ERP

Participants

Thirty-two healthy male participants (mean age 33.1 ± 6.8) were recruited in the study. Women were excluded to eliminate the confounder effect of menstrual cycle. The inclusion criteria for participants were: 18–45 years of age, 18.5–30 kg/m^2 body mass index (BMI) and 50–100 bpm resting pulse. Exclusion criteria included use of psychoactive medication, drug or alcohol abuse, severe drug allergy or hypersensitivity and history of any medical, psychiatric, and neurological (such as immunological, cardiovascular, respiratory, metabolic neurological, or psychiatric) disease. Written informed consent was obtained from each participant prior the study. All the collected data were included and analysed.

Procedures

Study I was an interventional, randomized, double-blinded, four-way crossover and placebo-controlled study. Three interventions 10 mg vortioxetine (A), 20 mg vortioxetine (B),
15 mg escitalopram (C) and a placebo (D) were included. Each participant was randomly allocated into a sequence group (ABDC, BCAD, CDBA or DACB) with a washout period (20-22 days) between interventions, resulting eight participants in each group (Figure 3). Prior each intervention, a bioanalysis was performed to monitor if there is leftover effect from the previous intervention. EEG data was recorded at the beginning of each session, before the administration

of intervention. Continuous EEG, auditory steady state response, auditory oddball and hybrid flanker Go/Nogo tasks were included as an EEG battery.

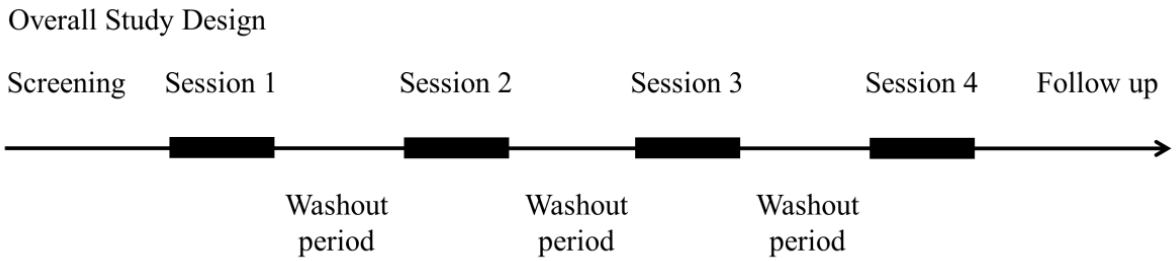


Figure 3 Overall study design of study I. Three interventions 10 mg vortioxetine (A), 20 mg vortioxetine (B), 15 mg escitalopram (C) and a placebo (D) were included. Each participant was randomly allocated to one session sequence including ABDC, BCAD, CDBA and DACB. After each session, there was a washout period (ranging from 18–22 days) between each session, and plasma concentrations was assessed to make sure a completely washed out from the previous intervention.

EEG recording

EEG was recorded from 28 scalp sites using a 10-20 electrode system, with a sample rate of 400 Hz (Comet EEG system, Grass Technologies, West Warwick, RI, USA). AFz served as the ground and POz served as the reference. Electrooculography (EOG) and electromyogram (EMG) were recorded at bipolar channels for later artifact removals. Resistance across all electrodes was maintained at less than 5 kΩ.

Continuous EEG

Resting and vigilance-controlled EEG were each recorded for at least 3 min. Participants were told to relax, keep their eyes closed and stay awake in both conditions. They were instructed to press two buttons using their thumbs during the vigilance-controlled recording. There would be an alert sound if the participant let go of the button.

Auditory steady state response (ASSR)

In order to assess gamma activity, participants were presented with a 40 Hz impulse trains sound at 89 dB binaurally through a headset (Sennheiser HD 25-1 II pro). Each train was composed of

20 biphasic 1 ms clicks over a 500 ms period followed by 700 ms of silence. These trains were repeated for 5 min.

Auditory Oddball

The auditory oddball paradigm consisted of two acoustic stimuli with different frequencies: standard tones (500 Hz) and deviant tones (2000 Hz). Participants were asked to count the deviant sounds. Each session consisted of on average of 35 deviants (randomized between 30 and 40) and 198 standards (randomized between 170 and 226). Deviant tones made up 15% of the presentations. The sound level for each tone was 85 dB, with duration of 100 ms and interstimulus-interval (ISI) of on average 1550 ms (randomized between 1200 and 1900 ms). The test lasted approximately 7 min.

Hybrid Flanker Go/Nogo (Hybrid FT)

Stimuli consisted of one of the following letter strings (BBBBB, DDDDD, VVVVV, UUUUU, BBDBB, DDBDD, UUVUU, or VVUVV) and were presented on a computer screen for 300 ms in randomized order. Participants were instructed to focus on the center letter and to press a button whether it was a B or a U (Go condition), and to withhold a button press when appearance of a D or V (NoGo condition). Each condition consisted of 420 trials, resulting 840 trials in total. Strings with congruent letters made up 40% of presentations, while strings with incongruent letters were shown in 60% of all trials. The presentation of the letter string was followed by 750 ms for stimulus onset asynchrony (SOA) and 500 ms for feedback in response to the participants' performance: ‘true’ (i.e. correct and in time), ‘faster’ (i.e. correct but out of time) or ‘false’. The ISI was 800 ms (randomized between 600 and 1000 ms). The test lasted approximately 45 min.

EEG data prepossessing and analysis

Details were described in paper I and were briefly summarized here. Eye blink and other ocular corrections were conducted by the ocular artefact reduction option of NeuroScan 4.1 software

and then were processed in the following steps: 1. Butterworth notch filtering (50 Hz) with a filter order of 6; 2. Continuous EEG: Butterworth band-pass (1–80 Hz) filtering with a filter order of 2. ERP: Butterworth band-pass (0.1–30 Hz) filtering with a filter order of 2; 3) Continuous EEG: re-referencing to the average electrode for continuous EEG recordings. ERP: re-referenced offline to the linked mastoids.

Time-frequency analysis of all data: A complex Morlet wavelet with a bandwidth of 10 Hz and a center frequency of 1 Hz was applied. The wavelet frequencies ranging from 1–80 Hz with a 0.5 Hz between-scale frequency interval. Power was acquired from the following standardized bands: δ (1–4 Hz), θ (4–8 Hz), α (8–12 Hz), β (12–30 Hz), $\gamma 1$ (30–45 Hz) and $\gamma 2$ (45–80 Hz). The γ band was divided into two bands to avoid muscle artefacts. Specifically, absolute power was calculated for the continuous EEG and evoked power was calculated for ASSR, auditory oddball and hybrid Flanker tasks, in which only the error commission was examined. Power from the three midline sites (Fz, Cz, Pz) were extracted for further analysis. Logarithm was applied for the normalization purpose.

Grand average analysis of ERP data: EEG data was segmented into epochs of 1000 ms and 600 ms for auditory oddball and hybrid Flanker task respectively. All epochs include 200 ms pre-stimulus baseline. Epochs were rejected if the voltage in EOG channels, Fp1, Fp2 exceeded ±75 μV. In the auditory oddball task, the selected components and the corresponding latency windows for peak identification included: standard: N100 (80–140 ms), P200 (140–270 ms); deviant: N100 (80–140 ms) and P300 (270–550 ms). Baseline to peak measures were determined on three midline sites (Fz, Cz, Pz). In the hybrid Flanker task, ERN (0–250 ms) was analyzed at sites in the fronto-central area (Fz, Cz) and Pe (100–350ms) was analyzed at sites in the centro-parietal area (Cz, Pz). The number of accepted epochs is shown in Table 1.

Table 1 The number of accepted epochs for different tasks.

Condition	BL1	BL2	BL3	BL4	p values

Task						
Auditory oddball	Standard	180±23(117-212) [a]	169±37(101-227)	167±27(97-215)	174±34(80-227)	F (3,92) = 1.884, p = 0.138
	Deviant	31±5(22-38)	30±7(16-40)	29±6(12-38)	30±7(15-40)	F (3,92) = 1.270, p = 0.289
Hybrid Flanker	Error	85±35(28-180)	70±28(8-133)	67±36(8-183)	72±31(3-141)	F (3,92) = 3.981, p = 0.010

Notes. [a] Mean and standard deviation are reported. Minimum and maximum of epochs are provided in the brackets.

Study II and Study III: EEG biomarkers from NeuroPharm trial

Study cohort and treatment

A hundred outpatients diagnosed with MDD were recruited in the NeuroPharm trial (see Figure 4 for a flowchart of the study). The inclusion criteria included: 18–65 years of age, with moderate to severe, first or recurrent major depressive episode and with a minimum score of 18 on Hamilton Depression Rating Scale 17 items ($HDRS_{17}$). Exclusion criteria included clinically significant psychosis, severe somatic co-morbidity, current or previous psychiatric severe co-morbidity, acute suicidal ideation and more than 2 years of duration of the current depressive episode. Thirty-five healthy controls were included as controls in the EEG analysis. All participants provided written informed consent prior to participation. Ethical approval was obtained by the National Committee on Health Research Ethics (protocol: H-15017713).

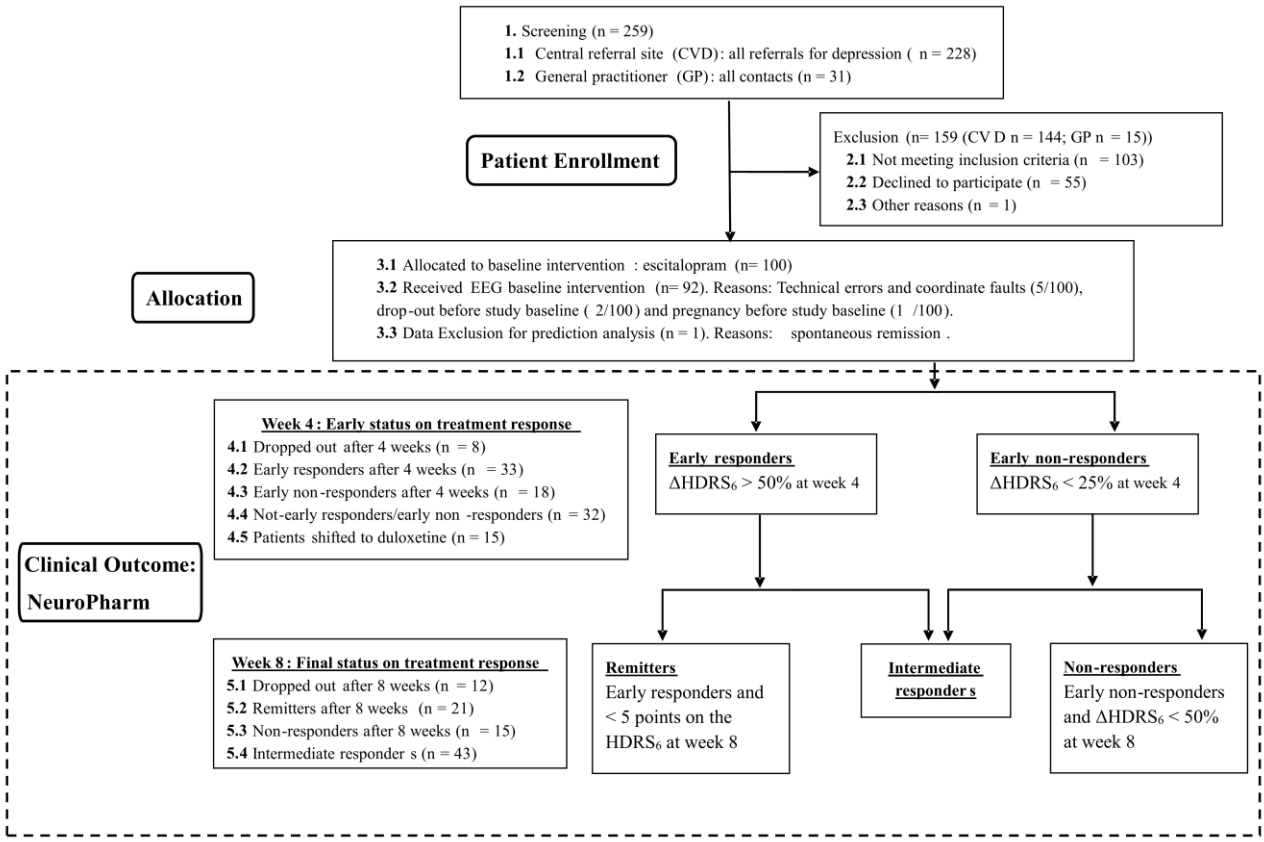


Figure 4 The flowchart of the study design of NeuroPharm trial. The definition of treatment responses of NeuroPharm was framed with dashed line. The treatment responses were defined as follow: Patients with $\Delta HDRS_6 > 50\%$ at week 4 were considered early responders; Patients with $\Delta HDRS_6 < 25\%$ at week 4 were considered early non-responders; Patients with $\Delta HDRS_6 > 50\%$ at week 4, and < 5 points on the $HDRS_6$ scale at week 8 were considered remitters; Patients with $\Delta HDRS_6 < 25\%$ at week 4 and $\Delta HDRS_6 < 50\%$ at week 8 were considered non-responders. Patients who were neither remitter nor non-responder criteria were classified as intermediate responders.

EEG recording was obtained at a pretreatment visit (before medication) and 40 of the patients were recorded again 8 weeks after treatment administration. EEG data were acquired prior treatment in 79 patients and 35 controls (please refer to Figure 4 for the exclusion reason). At 8 weeks follow-up, EEG data from 39 patients was acquired.

Patients were first treated with the SSRI escitalopram at flexible doses of 5–20 mg/day adjusted depending on effects and side effects by trained physicians. Patients with no response to escitalopram after 4 weeks were shifted to the SSRI/SNRI duloxetine (n = 15). Compliance,

side-effects to antidepressant treatment, and depressive symptoms were monitored at each visit: week 1, 2, 4, 8 and 12.

Clinical measures and treatment response

Changes in clinical symptoms were assessed by comparing the follow-up sessions with pretreatment depressive scores. The $HDRS_6$ subscale was extracted from the $HDRS_{17}$ for defining the treatment response in the NeuroPharm trial. In detail, the treatment responses were defined as follow: Patients with $\Delta HDRS_6 > 50\%$ at week 4 were considered early responders; Patients with $\Delta HDRS_6 < 25\%$ at week 4 were considered early non-responders; Patients with $\Delta HDRS_6 > 50\%$ at week 4, and < 5 points on the $HDRS_6$ scale at week 8 were considered remitters; Patients with $\Delta HDRS_6 < 25\%$ at week 4 and $\Delta HDRS_6 < 50\%$ at week 8 were considered non-responders. Patients who were neither remitter nor non-responder criteria were classified as intermediate responders.

EEG recording

Participants were instructed to remain still and relaxed, avoid eye-blinks and movements and to relax chin muscles during recording. Resting EEG was recorded during four 3 min periods with a counterbalanced order of OCOC (O for eyes open, C for eyes closed) or COCO between subjects. EEG data was recorded using a 256-channel HydroCel Sensor Net system (EGI, Inc., Eugene, OR) at 1000 Hz with 0.1–100 Hz analog filtering where vertex electrode served as the reference. Impedances of all electrodes were kept below 50 kΩ.

EEG data analysis

Study II

Study II intended to replicate pretreatment QEEG biomarkers that have been previously found identified to successfully differentiate treatment responses. We selected candidate biomarkers

based on pretreatment QEEG biomarkers that were associated to medication and resting EEG from the meta-analysis (Widge et al., 2018) for validation purposes (Table 2). The study of (Pizzagalli et al., 2018a) was also included since it was the largest multicenter randomized placebo controlled clinical trial (Trivedi et al., 2016) and was published after the meta-analysis (Widge et al., 2018). Treatment emergent biomarkers such as theta cordance and antidepressant treatment response index were not applicable in NeuroPharm trial and thus were excluded from this analysis. Studies with alert EEG (Cook et al., 2009; Dan V. Iosifescu et al., 2009; Korb et al., 2009) were also excluded due to different subject vigilance states were recorded.

Table 2 The included candidate biomarkers for replication

Studies	Sig. QEEG biomarkers	Recording duration[1]	R	NR	Clinical evaluation	Medication	Response criteria
Arns et al. (2016)	Greater FAA in R and better response to SSRI in female	Two 2 min (CO or OC)	427	240	**$HDRS_{17}$**; Baseline, week 8	escitalopram, sertraline, venlafaxine	R: ≥ 50% improved on $HDRS_{17}$. Remission: score of ≤ 7 on the $HRSD_{17}$
Bruder et al. (2001)	Less right alpha than left in female NR	Four 2 min (COOC or OCCO)	34	19	**CGI-I**; Baseline, week 12	fluoxetine	R: "Much improved" or "very much improved" on CGI-I
Bruder et al. (2008)	1. Greater occipital alpha in R; 2. Greater right hemispheric alpha in R.	Four 2 min (COOC or OCCO)	11	7	**CGI-I**; Baseline, week 12	fluoxetine	R: "Much improved" or "very much improved" on CGI-I
Pizzagalli et al. (2001)	Higher rACC theta activity, better response	Ten 3 min (COCOCOCOCO or OCOCOCOCOC)	9	9	**BDI**; Baseline, 4-6 months	nortriptyline	Median split of BDI scores
Pizzagalli et al. (2018)	Higher rACC theta activity, better response	Four 2 min (COOC or OCCO)	248[2]		**$HDRS_{17}$**; Baseline, weeks 1, 2, 3, 4, 6, and 8	sertraline, placebo	Absolute score on $HDRS_{17}$
Rentzsch et al. (2014)	Higher right pg/adACC delta in R	≥ 10 min (C)	11	20	**$HDRS_{21}$**; Baseline, weeks 2 and 4	various SSRIs	R: ≥ 50% improved on $HDRS_{21}$ at week 4.

Notes: [1] C refers to eyes closed, and O refers to eyes open. [2] The study did not provide information on the numbers for treatment responders/non-responders.

Abbreviation: R, responders; NR, non-responders; SSRI, selective serotonin reuptake inhibitor; FAA, frontal alpha asymmetry; rACC, rostral anterior cingulate cortex; pg/ad ACC, perigenual and anterior dorsal anterior cingulate cortex; HDRS, Hamilton Depression Rating Scale; CGI-I, Clinical Global Impression Improvement scale; BDI, Beck Depression Inventory.

These chosen biomarkers were then validated as close as possible to the resample studies, in terms of EEG data analysis, clinical criteria and statistical approaches.

Study III

VIGALL was adopt to the eyes closed resting EEG in order to identify and different functional brain states from wakefulness to sleep onset. VIGALL was an algorithm-based method that could automatically classified EEG epochs into the following arousal states: stage 0 (highest arousal), A1, A2, A3, B1, B2/3, C (lowest arousal and indicating the sleep onset). For technical details please refer to VIGALL manual (VIGALL 2.1 manual; https://research.uni-leipzig.de/vigall) and here we briefly summarized. The classification method mainly based on the distribution of alpha cortical current density over our regions of interests (ROIs): frontal, central, temporal and occipital. After closing eyes, desynchronized alpha activity with the absence of slow eye movements (SEMs) would dominate the cortical activity (state 0). Along with the relaxation of the participant, synchronized alpha activity would dominate progressively from occipital (A1), central and frontal (A2), and to mainly frontal area (A3). Next, desynchronized alpha activity would appear again and replaced by low amplitude with SEM (B1) then dominate by delta and theta activity (B2/3). With the appearance of K-complexes and sleep spindles, the epochs would be an indicative of sleep onset (C). The resulting classification was then assigned numerically with a range from 7 (stage 0, full wakefulness) to 1 (stage C, sleep onset) for further calculation.

Twenty-five channels were selected from our high-density EEG net prior calculation, which included: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP6, P7, P3, Pz, P4, P8, O1 and O2. HEOG and VEOG were chosen as close as possible to the electrode locations according to the VIGALL manual. The EEG data was then down-sampled to 250 Hz and re-referenced to an average reference. The data was cut into 1s epochs for manually check

for artefacts. Eye movement artefacts were removed through independent component analysis. A zero-phase digital IIR Butterworth bandpass filter with cut-off frequencies (0.5–70 Hz), and an additional 50 Hz notch filter were applied to the data. Since no subject had any sign of sleep stage (stage C) and the prevalence of stages A2 and A3 have been quite low in previous studies, we followed the usual procedure of pooling two A stages (A2/3), resulting in five different vigilance stages (0, A1, A2/3, B1, B2/3). Median vigilance of each 1 min block and the vigilance slope of each 3 min block were reported. A positive vigilance slope indicated less decay in vigilance while a negative slope indicated more pronounced decrease of vigilance toward drowsiness.

Statistical analysis

Study I

The statistics were divided into two parts. First, all EEG and ERP measures were analyzed with a linear mixed model with assigned sequence (ABDC, BCAD, CDBA or DACB) and baseline recordings of each session (BL1, BL2, BL3, BL4) as fixed factors. An unstructured covariate matrix was employed. In all the mixed models, age was included as covariate and least-square means were used in post hoc analyses. Main effects of session and sequence were tested using *F*-tests.

Second, the test-retest reliability was evaluated by intra-class correlation (*ICC*) with absolute agreement (Brunner et al., 2013; Hämmerer, Li, Völkle, Müller, & Lindenberger, 2013). Single measure *ICC* (A, 1) was calculated respectively by a two-way mixed random model (McGraw & Wong, 1996), where participant served as random variable and session served as mixed variable. *ICC* of adjacent time points and overall time variance are reported. In accordance with the classification of ICC levels in a previous study (Rentzsch, Jockers-Scherübl, Boutros, & Gallinat, 2008), *ICC* < 0.39 would be considered poor, 0.4–0.59 fair, 0.6–0.75 good and > 0.75 would be considered excellent.

Study II

All the statistical analyses were performed followed as they are in the resampled study, including the selection of models and included covariates. The included biomarkers were examined with both the criteria of NeuroPharm and of the resample studies. Since the rating scales of clinical measures from previous studies might not be included in the current study, a transformation between scales for treatment response was applied whenever possible (Riedel et al., 2010); otherwise responders were defined by at least 50% improvement of depressive symptoms as assessed by the $HDRS_{17}$ score. One-sided *p* values were chosen for validation purpose.

Study III

The main outcome of VIGALL included percentages of each vigilance stage (5 states × 6 blocks), median vigilance of each block (6 blocks) and vigilance slope (two separate eyes closed recordings from one session: 1st and 2nd recordings). They were correspondingly subjected to repeated analysis of variance (ANOVA) to determine whether there were temporal difference in EEG vigilance patterns 1) between pretreatment MDD and healthy controls, 2) between remitters and non-responders (NeuroPharm criteria) in pretreatment vigilance; between responders and non-responders, and remitters and non-remitters in pretreatment vigilance (iSPOT-D criteria), 3) between pretreatment EEG and the EEG after 8 weeks of treatment (treatment), and whether these vigilance patterns differ between different treatment responses with both criteria of NeuroPharm and iSPOT-D. The analysis of Receiver Operator Curve (ROC) was performed for significant discriminant on the clinical outcome. Furthermore, a correlation analysis on ΔHDRS scores (both $\Delta HDRS_6$ and $\Delta HDRS_{17}$) and vigilance slope was performed to investigate the vigilance effect on clinical outcome.

Results and discussion

Study I

Is there significant difference across different recording sessions?

Absolute power of continuous EEG

Absolute γ1 power at BL1 was larger than the last session BL4 (17.83 vs 16.39 μV, p=.006, Figure 5) under vigilance-controlled recording (p values > .05). No other significant effect was found (p values > .05).

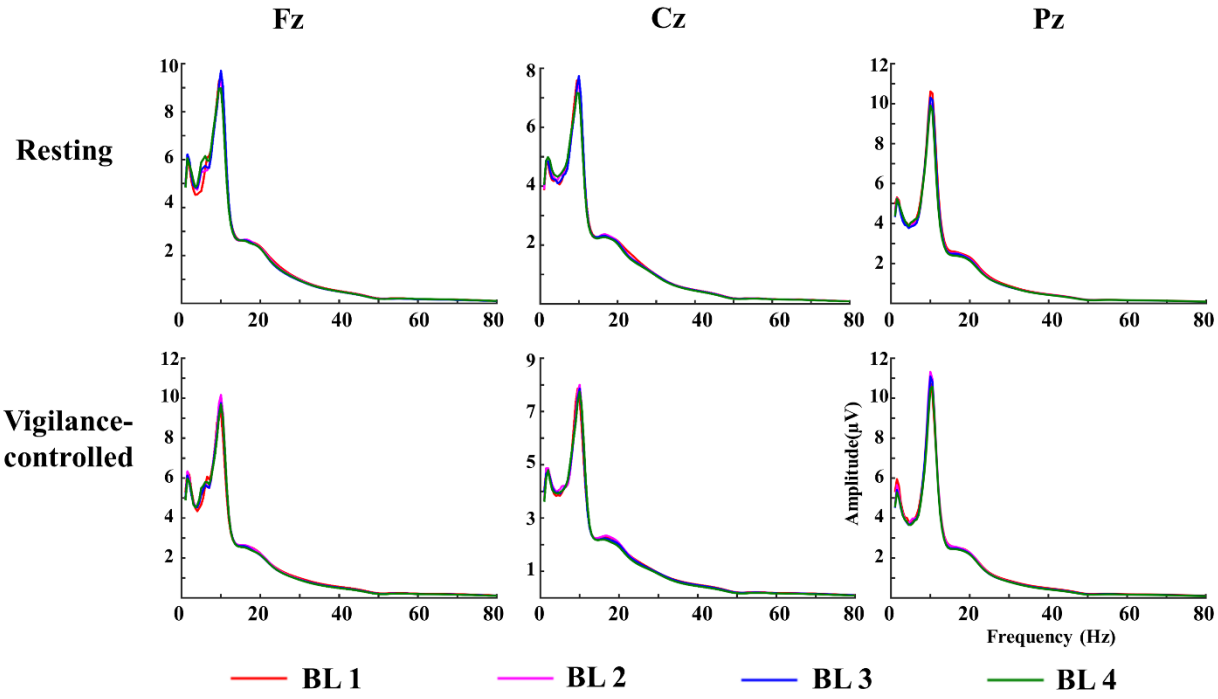


Figure 5 Spectral analysis for resting state and vigilance-controlled across four pre-intervention sessions. Three midline electrodes (Fz, Cz and Pz) are shown for each condition. Four pre-intervention sessions (BL1, BL2, BL3 and BL4) are shown in different colors.

Evoked power of ASSR, auditory oddball and hybrid flanker task

In ASSR task (Figure 6a), smaller frontal δ and larger γ1 power was observed at BL1 compared to BL3 and BL4 (**δ**: F (3, 31) = 3.919, p = .018; **γ1**: F (3, 31) = 3.567, p = .025). Similarly, a smaller parietal δ was observed at BL1 compared to BL4 (F (3, 31) = 3.179, p = .038). Central θ

at BL1 was the smallest compared to other recording sessions (F (3, 31) = 4.352, p = .011), and central α at BL1 and BL2 were smaller than that of BL4 (F (3, 31) = 3.409, p = .03).

In auditory oddball task (Figure 6b), standard and deviant stimuli were analyzed separately. Standard tone: smallest frontal δ and largest parietal δ at BL1 were observed among all recording sessions (F (3, 31) = 5.651, p = .003; F (3, 31) = 3.844, p = .02). Deviant tone: the frontal δ was larger at BL2 compared to BL3 and BL4 (F (3, 31) = 3.111, p =.04). Frontal θ at BL1 were smaller than BL2 and BL 4 (F (3, 31) = 3.327, p = .032). Parietal θ at BL1 was smaller than BL2 (F (3, 31) = 3.185, p = .038).

In hybrid Flanker task (Figure 6c), smaller central θ at BL1 and BL2 were recorded compared to BL3 (F (3, 31) = 3.52, p = .027). Smallest fronto-central β and $\gamma 2$ was observed at BL1 compared to other recording sessions (F (3, 31) = 3.679, p = .023; F (3, 31) = 3.219, p = .036). Central β, γ1 and γ2 at BL1 was smaller than BL3 (F (3, 31) = 3.297, p = .033; F (3, 31) = 4.804, p = .007; F (3, 31) = 3.640, p = .023).

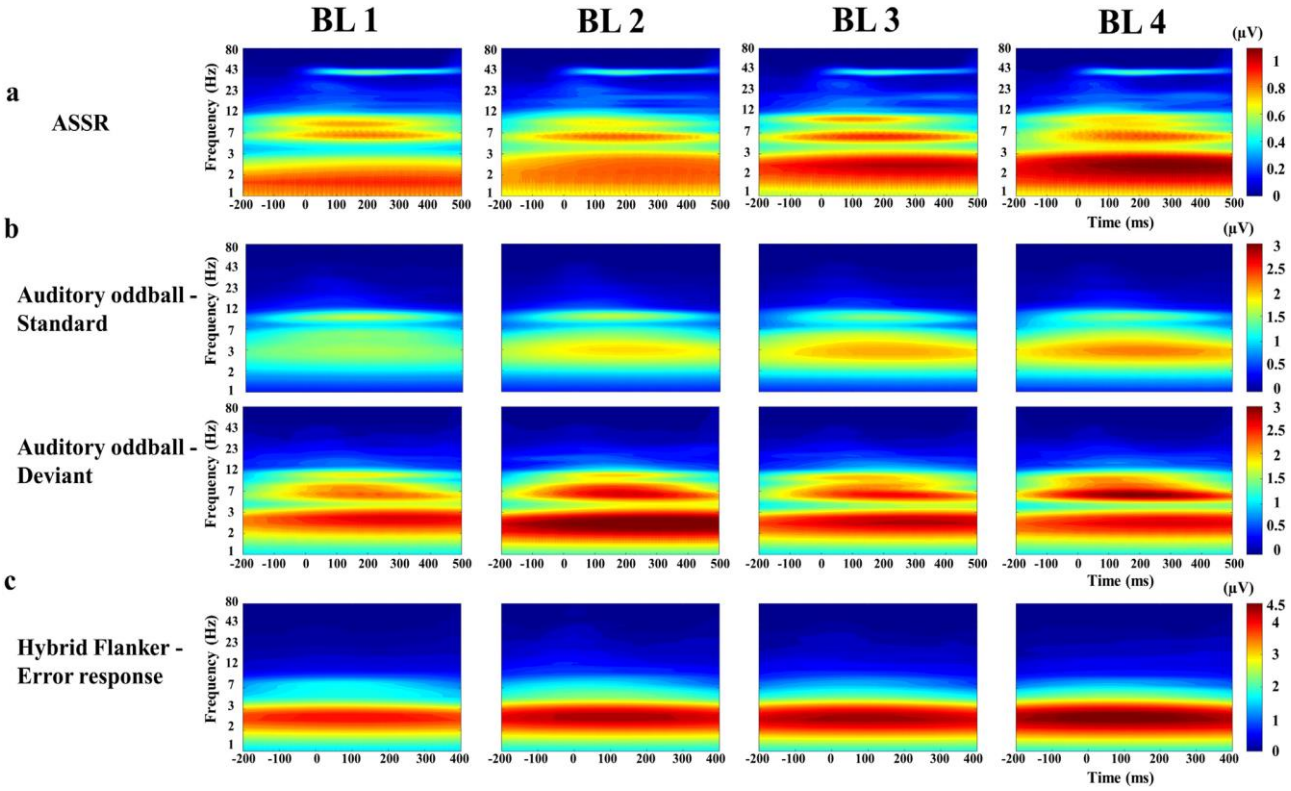


Figure 6 Time-frequency demonstration for ASSR, auditory oddball and hybrid flanker tasks. Results from Fz are shown here since it reveals the most significant differences. Four pre-intervention sessions (BL1, BL2, BL3 and BL4) are shown in columns. Log-scale is shown for the frequency range.

ERP analysis of auditory oddball and hybrid Flanker tasks

For the standard stimuli in the auditory oddball task (Figure 7), fronto-central N100 exhibited larger amplitude at BL1 and BL2 than BL4 (F (3, 31) = 5.21, p = .005; F (3, 31) = 6.93, p = .001). Moreover, parietal N100 latency showed longer latency at BL1 than other recording sessions (F (3, 31) = 3.36, p = .034). No sessions effect was found for P200 amplitude and latency. For the deviant ERPs (Figure 7), shortest central N100 latency was found at BL4 (F (3, 31) = 3.26, p = .034) while a shorter P300 latency was observed at BL1 compared to BL4 (F (3, 31) = 4.13, p = .014).

In hybrid Flanker task (Figure 7), fronto-central ERN exhibited longer latency at BL1 than at BL3 and BL4 (F (3, 31) = 3.78, p = .02; F (3, 31) = 6.91, p = .001). The centro-parietal Pe showed longer latency at BL1, BL2 than at BL3 and BL4 (F (3, 31) = 29.34, p < .001; F (3, 31) = 22.66, p < .001).

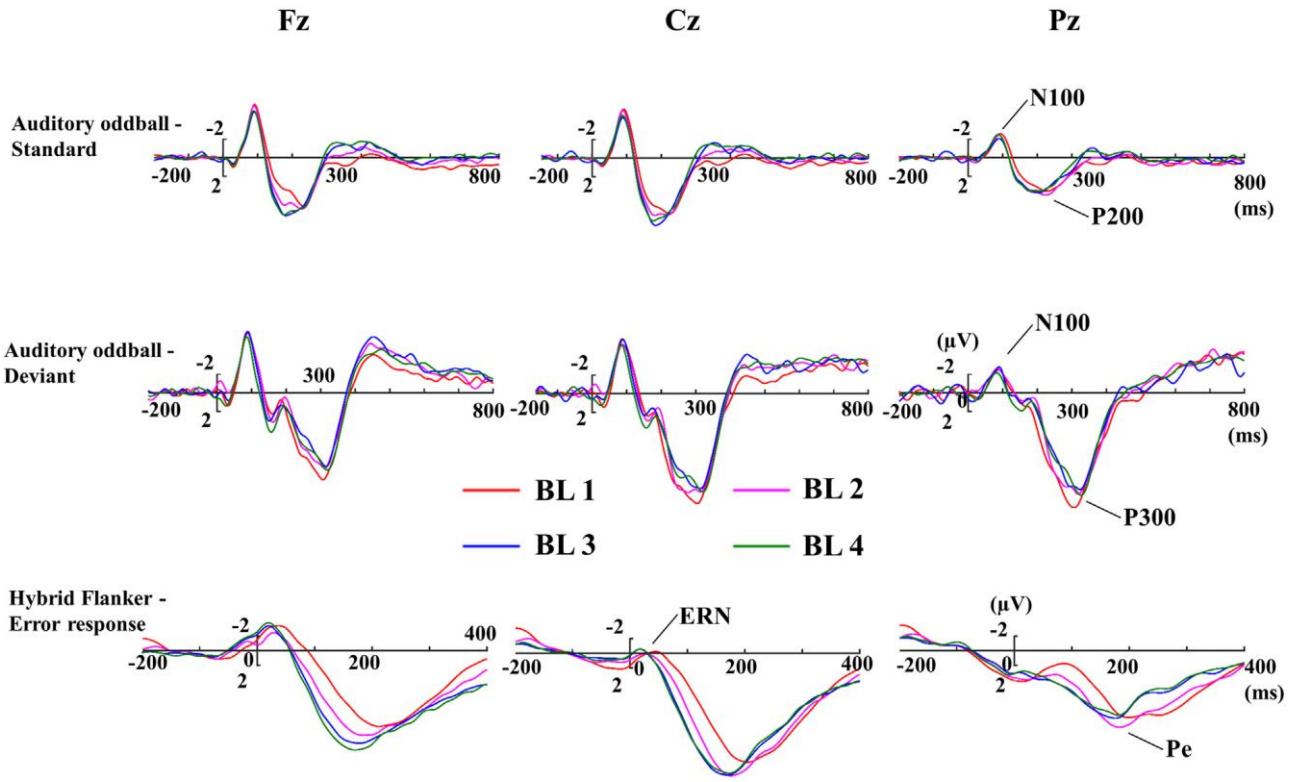


Figure 7 The mean ERP waveforms for the auditory oddball and hybrid flanker task (epoched by the error commission). Three midline electrodes (Fz, Cz and Pz) are shown for each task/component. Four pre-intervention sessions (BL1, BL2, BL3 and BL4) are shown in different colors.

Our results indicated that the absolute power of continuous EEG showed less variation between sessions compared to task-related EEG and ERP. Also, the EEG activity at the first recording (BL1) contributed the most to the variations. More discussion would come after the results of test-retest reliability.

Repeated ANOVA was mainly used when testing difference between two recording sessions, however, this traditional approach is not suitable for four sessions design. The reason is that ANOVA uses a compound symmetry structure to assume the covariance matrix between pairs of within-subject variable (Figure 8 left panel), while this is not the ideal case. Linear mixed model with unstructured covariance matrix is preferable because the variances of the differences between all possible pairs of conditions do not necessarily to be equal (Figure 8 right panel). For instances, the variance between BL1 and BL2 doesn't require to be the same as BL1 and BL3 on

the same subject. This approach is reasonable to use given the fact that we cannot be sure if EEG/ERP parameters decay, increase or remain still between different recording sessions.

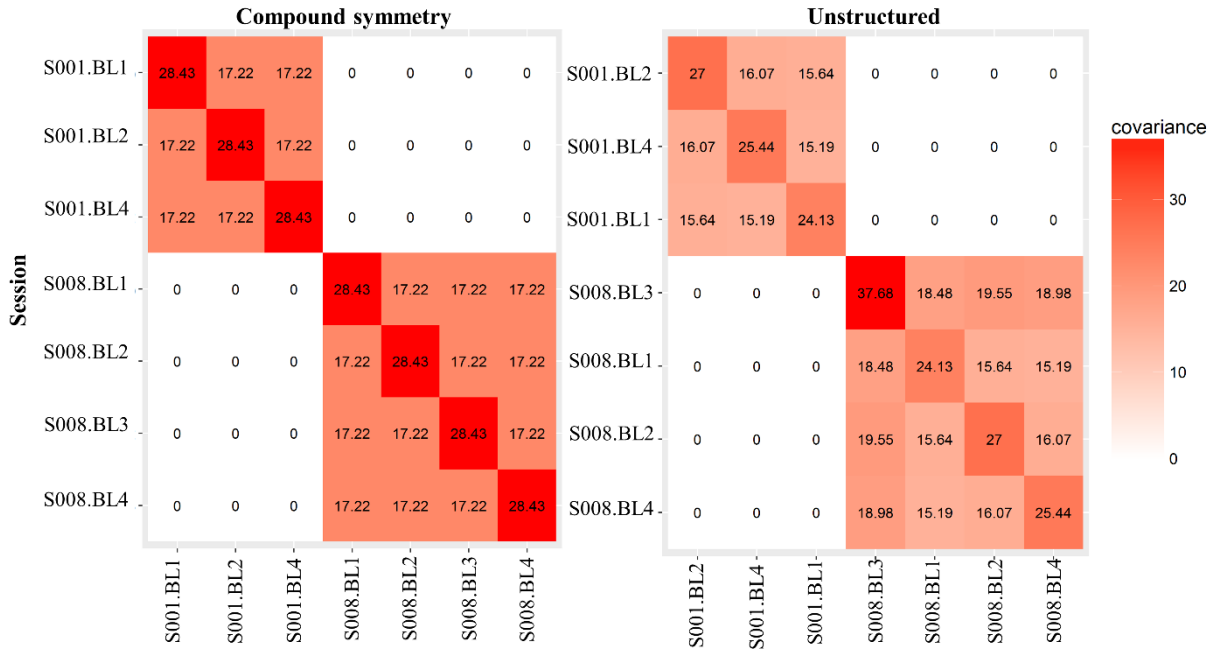


Figure 8 A comparison of compound symmetry and unstructured covariance matrix.

Is EEG reproducible among four pre-intervention sessions?

Absolute power of continuous EEG

The results of *ICCs* showed excellent test-retest reliability (0.84–0.97, Figure 9) at adjacent sessions in the frequency bands of θ, α and β at midline electrodes. However, the lower and higher ends of the frequency bands demonstrated more variances (Figure 9).

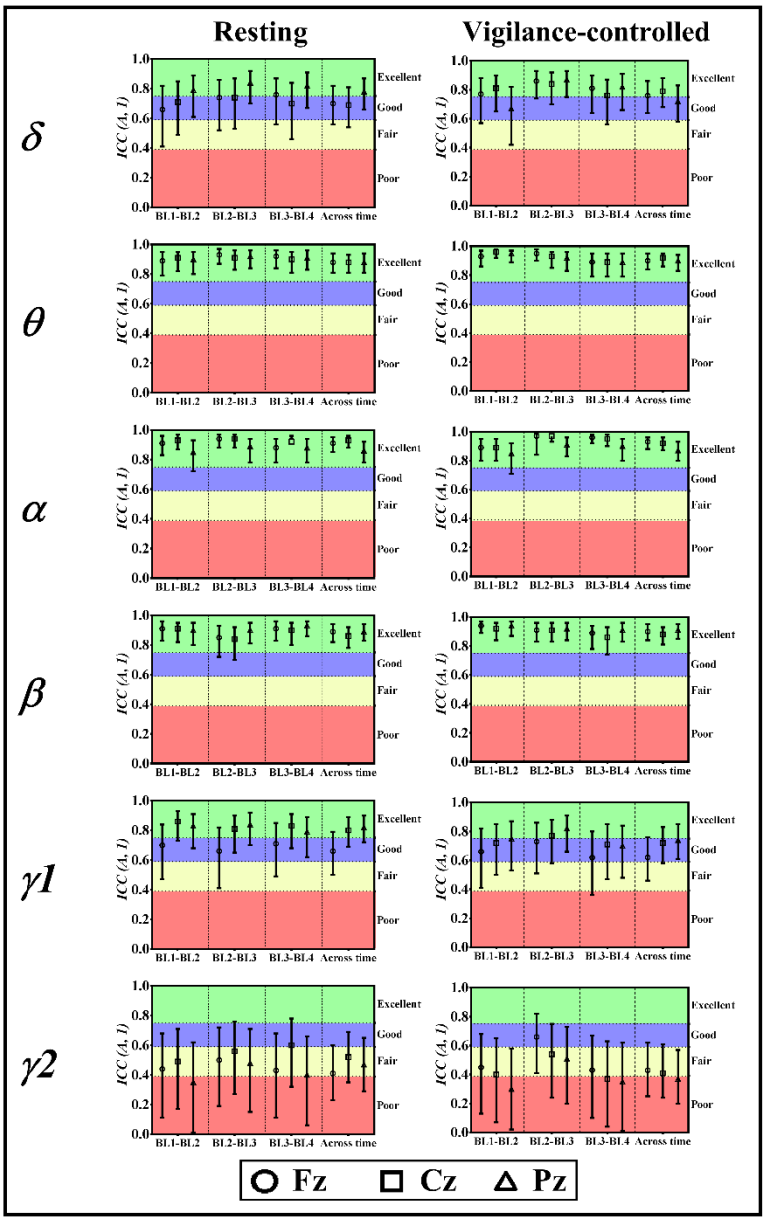


Figure 9 Intra-class correlation coefficient (*ICC*) for continuous EEG across four pre-intervention sessions. Midline θ, α and β show excellent test-retest reliability (0.84–0.97). Mean and confident intervals for *ICC*s are shown in the figure.

Evoked power of ASSR, auditory oddball and hybrid flanker task

Compared to continuous EEG (Figure 9), the test-retest reliability of ERP tasks was clearly less stable over time.

For the ASSR task, midline γ1 –which contains the stimulation frequency– exhibited fair to excellent reliability for both adjacent sessions and across time (.57–.86, Figure 10). Standard tone was more robust than deviant tones, but also with large variation between sessions. Midline β and γ1 revealed fair to excellent levels of reliability for both adjacent sessions and across time

(0.44–0.85). For the deviant tones, the *ICCs* of δ were in the range of good to excellent (0.63–0.83). For the error response of the hybrid flanker task, the *ICCs* of θ exhibit better reliability compared to other bands (0.50–0.85).

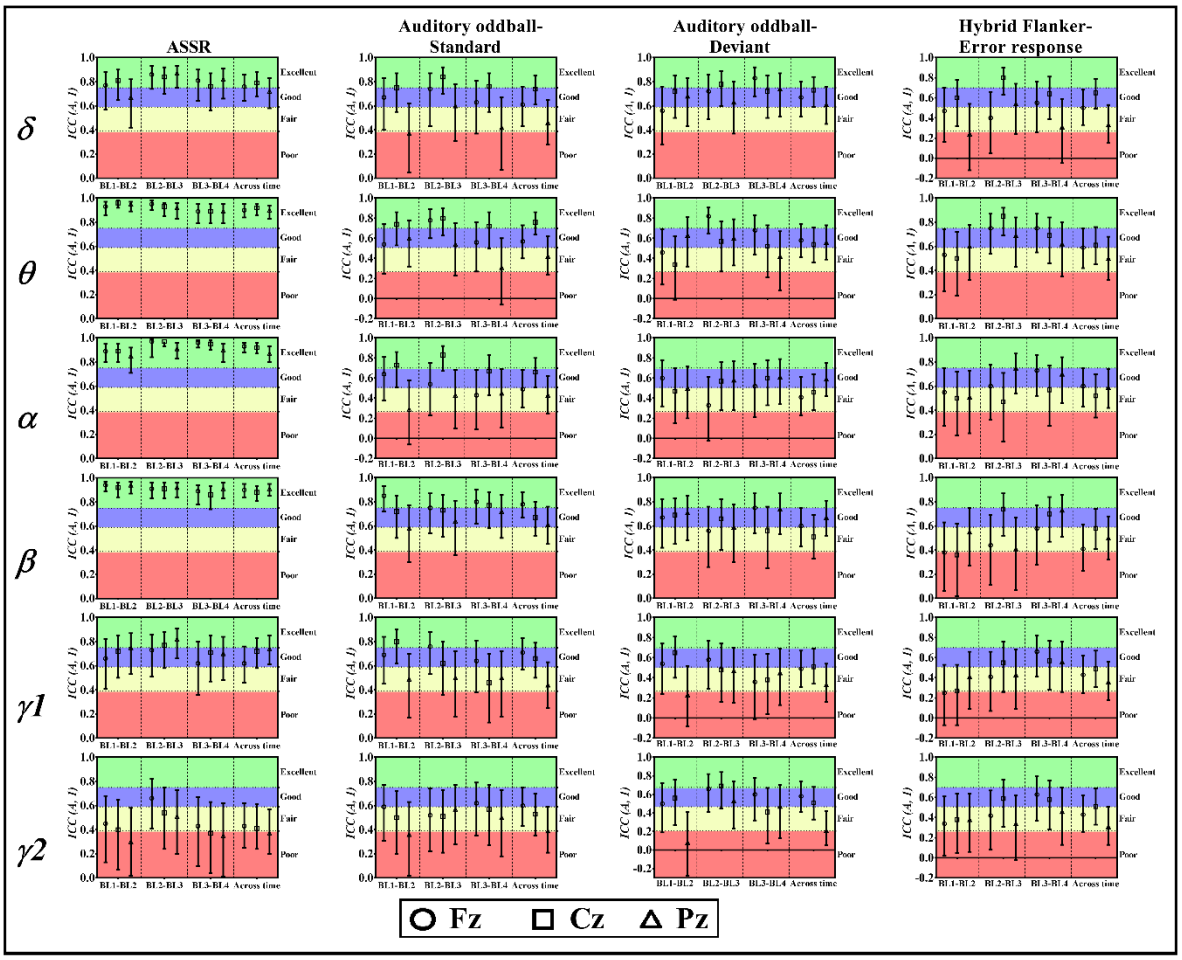


Figure 10 Intra-class correlation coefficient (*ICC*) for ASSR, auditory oddball and hybrid flanker tasks across four sessions. Power measures of tasks related EEG were less robust compared to continuous EEG. Mean and confident intervals for *ICC*s are shown in the figure.

ERP analysis of auditory oddball and hybrid Flanker tasks

Late components such as P300 amplitude (0.55–0.80), P200 (0.49–0.83) and Pe (0.60–0.82) had good to excellent reliability (Figure 11). Latency measures were less stable compared to amplitude measures in general.

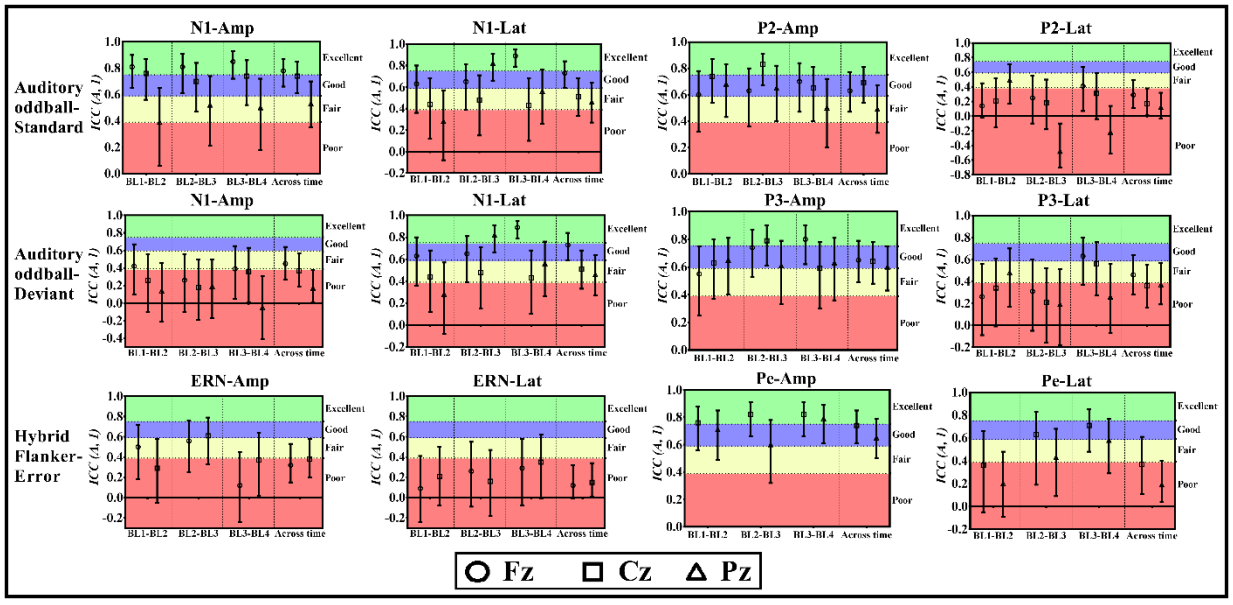


Figure 11 Intra-class correlation coefficient (*ICC*) for peak amplitude and latency measures of auditory oddball and hybrid flanker tasks across four pre-intervention sessions. Later components such as P300, P2 and Pe have higher test-retest reliability compared to other early components. Amplitude (Amp) measures are more reliable than latency (Lat) measures. Mean and confident intervals for *ICC*s are shown in the figure.

In study I, we are first to report the test-retest reliability of continuous EEG, ASSR, auditory oddball and hybrid Flanker tasks over *four* pre-intervention sessions while previous studies have mainly conducted with two sessions. We found that the absolute power of continuous EEG demonstrated excellent reliability in the middle frequency bands (θ, α and β), while evoked power of ERP tasks exhibited more variances over time. Regarding to the peak-picking and latency measures, we found that late components (P300, P200, Pe) had higher reliability compared to those early components (N1, ERN). Latency measures was unstable over time compared to other measures. Our results provide evidence that these EEG/ERP parameters are reliable across three-week intervals and thus are sufficiently reliable for future investigations.

For the power analysis, our results suggest that the middle frequency bands have higher reliability than the edge bands in both EEG and ERP data. Our findings duplicated the finding of Gudmundsson, Runarsson, Sigurdsson, Eiriksdottir, & Johnsen, (2007), which also showed that the δ and γ bands were less reliable than the other bands. On the contrary the study of Williams

et al., (2005), researchers reported that lower frequency bands, as δ, θ, tend to have higher reliability than α and β bands. They suggested that the discrepancies might be due to different time intervals between recordings. We argue that the length of session intervals might not be the main reason, but that the noise level is. The γ band has been divided into two bands in our analysis in order to avoid the noise coming from the muscle activity and the main power source. This results in $\gamma 1$ (with less noise blended in) having higher test-retest reliability than $\gamma 2$ (Figure 9). Since these two bands share the same session intervals but have different reliability, we hypothesize that the different noise levels influence the reliability. With regard to the reliability of ASSR, there have been only a few studies that have investigated it (80 Hz-ASSR: (Kaf, Sabo, Durrant, & Rubinstein, 2006); 40 Hz-ASSR: (McFadden et al., 2014)). Little is known about the reliability of power spectra, even though it is the main purpose of this paradigm to see the changes in the γ band. The results suggest that ASSR measures are highly stable over time. Our results on the reproducibility of ASSR extend the current findings by presenting the power spectra of ASSR measure.

The reliability of ERP data is affected by various factors (Brunner et al., 2013; Larson, Baldwin, Good, & Fair, 2010), resulting in less stable reliability compared to power measures. We found that the reliability of ERP measures is affected by the size of the components. Smaller-sized components such as N100 and ERN exhibit lower reliability relative to larger-sized components, P300 and Pe (Figure 11). This discrepancy could be caused by the different SNRs existing in different sizes of ERP component (Luck, 2005). Increasing the number of averaged trials and a better control of artefacts could increase the SNR for ERP components (See exploratory analysis in paper I). Higher SNR for small components could notice future studies for reaching high or comparable reliability as large components.

Study II

Of those who completed the study, the remission rate (assessed using the criteria from NeuroPharm) and response rate (≥ 50% improvement of depressive symptoms on $HDRS_{17}$ score) were 26.58% and 55.70%.

We found that only two out of six candidate biomarkers could be partially replicated (Table 3). These two biomarkers both involoved alpha asymmetries: frontal alpha asymmetry (FAA) and alpha asymmetry.

FAA: After the exclusion of patients with low serum concentrations and patients with duloxetine (remaining n = 35), we found a significant gender-specific partial correlations between FAA score and $HDRS_6$ scores were found at week 8 (r (30) = -0.29, $p_{\text{one-tailed}}$ = .048) and for the improvement at week 8 (r (30) = 0.32, $p_{\text{one-tailed}}$ = .036) in eyes closed condition (Figure 12a), as well as in eyes open condition ($HDRS_6$: r(30) = -0.28, $p_{\text{one-tailed}}$ = .060; $\Delta HDRS_6$: r (30) = -0.31, $p_{\text{one-tailed}}$ = .041). Consistent with prior report of FAA (Arns et al., 2016) , this association between treatment response and FAA was only found in female patients but not male.

Mean alpha asymmetry: When testing the mean alpha asymmetry (Bruder et al., 2001), log alpha powers at three regions (frontal, central and posterior) were subjected to repeated ANOVA without taking the average across three regions. We found a greater right posterior alpha in male non-responders (both NeuroPharm and criteria of the previous study, F (2, 64) = 3.87, p = .041; F (2, 150) = 4.31, p = .025) and a less right central alpha in female non-responders (only found when the criterion of the previous study was applied (Figure 12b).

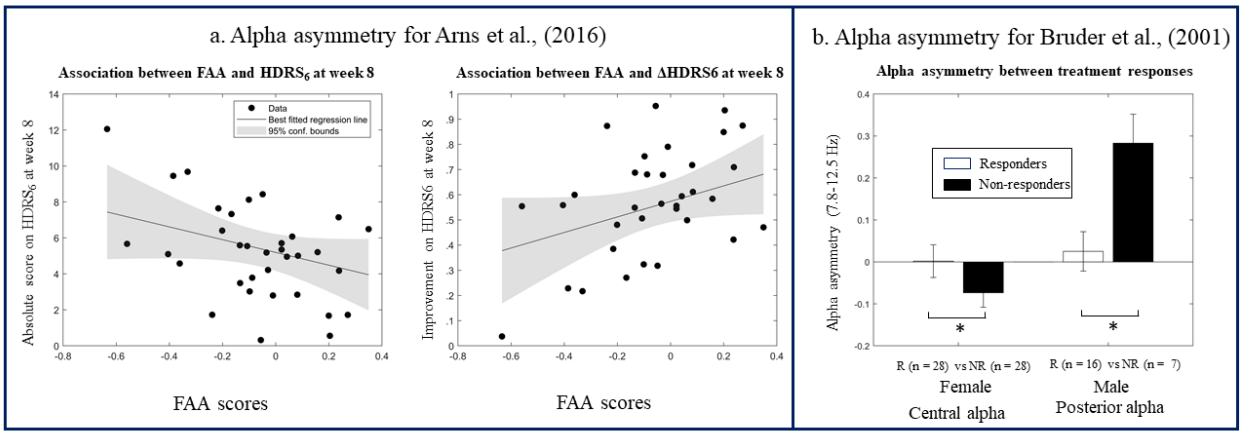


Figure 12 a. Female patients who have higher pretreatment FAA score (right dominated FAA) are associated with lower depressive symptoms at week 8 (Figure 12a left panel). Also, pretreatment FAA scores are positively (right dominated FAA) associated with patients' improvement on ΔHDRS6 at week 8 (Figure 12a right panel). Grey areas indicate 95% confident intervals for the fitted lines. b. Mean alpha asymmetry shows a gender difference on the treatment response when the criterion of the study of Bruder et al. (21) was examined. Error bars indicate standard deviations. Female responders have greater central alpha asymmetry (right > left) than female non-responders while male responders have lower posterior alpha asymmetry (right < left) than male non-responders.

Note: *: $p < .05$.

We observed a gender-specific effect for alpha asymmetry, in line with the literature (Arns et al., 2016), demonstrating that greater right FAA indicates lower $HDRS_6$ score and better improvement of depressive symptoms at week 8 in female patients. Furthermore, female non-responders showed an opposite asymmetry at the central sites. Our results might indicate that there exists a gender-related lateralization in the serotonergic neurotransmitter system, which modulates the effect of an antidepressant. A recent study reported that the sex-specific cortical lateralization is associated with 5-HTTLPR (serotonin-transporter-linked polymorphic region) genotype (Volf, Belousova, Knyazev, & Kulikov, 2015). Moreover, in the study of Volf et al., (2015), researchers revealed that the 5-HTTLPR polymorphism affects lateralization in healthy female. Furthermore, previous studies, measured by PET, have also found sex differences in cortical asymmetry of both the serotonin transporter (Kranz et al., 2014) and the serotonin 1A binding (Fink et al., 2009). Future investigations on sex-related asymmetry on serotonergic system could help the understanding of this observation.

Theta current source density: The current results revealed an opposite direction of ACC theta (see Figure 13a for the chosen Montreal Neurological Institute (MNI) coordinates) compared to previous study (Pizzagalli et al., 2001). We found a higher ACC theta in non-responders compared to remitters (F (1, 33) = 5.10, p = .03, Figure 13b). A negative correlation was found when correlating ACC theta and the improvement of depressive scores at week 8 ($\Delta HDRS_6$: r (44) = -0.21, p = .085). No such effect was found when the criteria of the resampled study was being assessed (p values > .05).

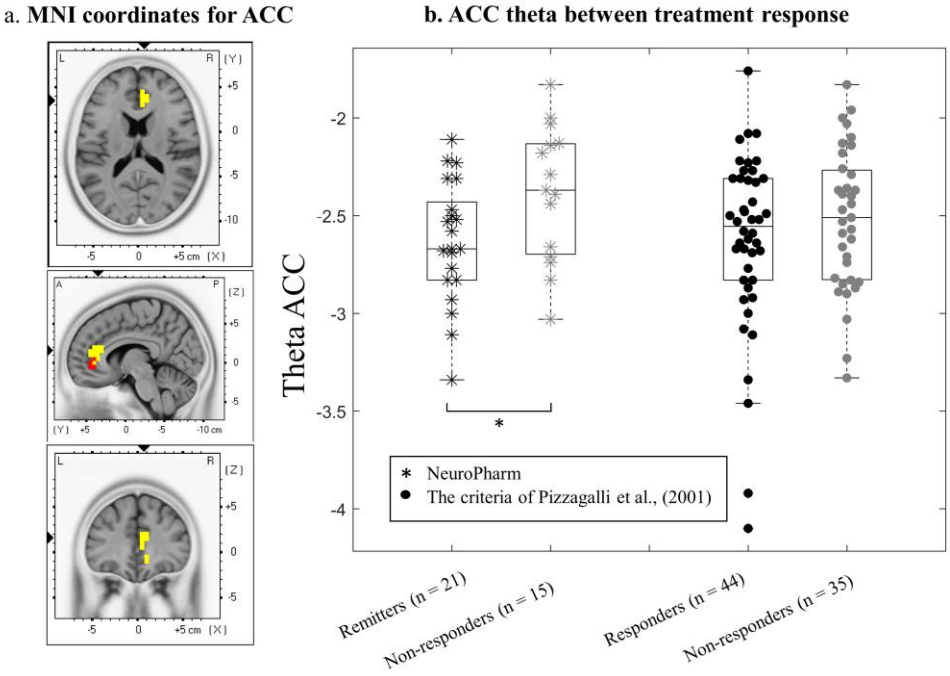


Figure 13 a. The visualization of the chosen MNI-coordinates (view angle: [X, Y, Z] = [7, 35, 16] mm). Theta power was extracted from a rACC cluster (identical 14 voxels from (Pizzagalli et al., 2018b; Pizzagalli et al., 2001), highlighted as red) and delta power was extracted from pg/adACC (identical 22 voxels from (Rentzsch, Adli, Wiethoff, Gómez-Carrillo De Castro, & Gallinat, 2014), highlighted as yellow). b. Theta activity at rACC cluster (red blocks) for both response criteria: NeuroPharm and the previous study (Pizzagalli et al., 2001). Logarithmic theta current source density was extracted from rACC using eLORETA. When the criteria of NeuroPharm was assessed, the results showed a significant higher ACC theta in non-responders compared to remitters. No such difference was observed when the criteria of the resampled study was assessed. Data were visualized the same way as the resampled study (Pizzagalli et al., 2001).
Note: *: p < .05.

Inconsistent with prior studies (Korb et al., 2009; Mulert et al., 2007; Pizzagalli et al., 2018; Pizzagalli et al., 2001; Rentzsch et al., 2014), we did not find that higher slow-frequency activity

at ACC was associated with response. Conversely, the results demonstrated that remitters had lower theta activity at ACC compared to non-responders (Figure 13b). The inconsistency could not be caused by the different analysis approaches, i.e. whole brain vs. ROI analysis, for which the latter approach would maximize the possibility to detect any group effect. Both measures have been used previously to report the effects of theta ACC (whole brain: (Pizzagalli et al., 2001); ROI-based analysis: (Pizzagalli et al., 2018)). Moreover, it would also be unlikely to show the opposite direction by adapting whole brain analysis in the current dataset. Despite this, we reran the analysis with whole brain voxel-by-voxel comparison with the same statistical method as earlier studies (Pizzagalli et al., 2001; Rentzsch et al., 2014)for both delta and theta bands. No significant result was observed for any group effects (*p* values > .05). The results indicated that theta activity at ACC might not serve as a reliable biomarker, resulting from the failure in the current replication. Furthermore, the fact that theta ACC could not differentiate responders of actual antidepressant and placebo receivers (Korb et al., 2009; Pizzagalli et al., 2018)dampened the effectiveness of using it as a prognostic biomarker with clinical value.

Nevertheless, our results are consistent with the finding from one of the largest qEEG studies on depressed individuals so far, the iSPOT-D, in which they also reported a lower theta activity for treatment responders. One possibility could be that lower theta ACC is an indicator to SNRI antidepressant responder. In the study of Arns et al., (2015), they reported significant lower theta for venlafaxine-XR responders but not for SSRI treatments. This was confirmed by our data, theta ACC was found for responders only when duloxetine receiver was included, the effect disappeared when excluding this part of patients. There seems to exist treatment specificity on theta activity in relation to treatment response, as proposed by Arns et al., (2015). Future studies are needed to investigate whether pretreatment theta ACC could be a specific differential indicator to different types of antidepressants.

Table 3 Summary of replication results of study II

Study	Statistical analysis[1]	Included covariates	Results[2]	
			Criteria of NeuroPharm	Criteria of the previous study
Arns et al. (2016)	**Primary**: ANOVA with FAA score under different conditions and responses' groups. Sex-specific effect on female patients. **Secondary**: Partial correlation between FAA and HDRS scores. **Exploratory:** ANOVA with hemisphere and groups.	Age, sex, pretreatment anxiety level and pretreatment HDRS score. Age and sex were included in the exploratory analysis.	Partial replication: Greater right FAA, better response in female	No replication
Bruder et al. (2001)	**Primary:** ANOVA with alpha power (overall and high alpha) on both hemisphere, groups and sex were tested. **Exploratory**: ANOVA with regions (anterior, central, posterior), hemisphere, groups and sex were tested.	Not applicable	No replication	Partial replication: Less right central alpha in female NR
Bruder et al. (2008)	**Primary:** ANOVA with hemisphere and condition and groups. Possible handedness effect on right-handed patients were examined.	Number of years of education	No replication	No replication
Pizzagalli et al. (2001)	**Primary:** ANOVA with narrow ACC theta (14 voxels) and groups. **Secondary:** Pearson correlation between ACC theta and HDRS scores.	Not applicable	No replication[3]	No replication
Pizzagalli et al. (2018)	**Primary:** Partial correlation between ACC theta (narrow and broad theta) and the ΔHDRS at week 8.	Age, sex, race, marital status, employment status, pretreatment anxiety level and pretreatment HDRS score[3]	No replication[3]	No replication[3]
Rentzsch et al. (2014)	**Primary:** ANOVA with ACC delta (22 voxels) and groups. **Secondary:** Pearson correlation between ACC delta and ΔHDRS at week 4.	Not applicable	No replication	No replication

Notes: [1] The analysis was kept as close to the resampled study as possible. [2] To allow comparison with previous study, both results from the criteria of NeuroPharm and of the previous studies were reported. [3]An opposite effect was found.

Abbreviation: ANOVA, repeated-measures analysis of variance; FAA, Frontal alpha asymmetry; ACC, anterior cingulate cortex; HDRS, Hamilton Depression Rating Scale.

Study III

Sociodemographic characteristics between patients and controls, and clinical outcome at week 8 with criteria of NeuroPharm and iSPOT-D were shown in Table 4. Numerical results of vigilance parameters at pretreatment visit and at week 8 can be found in Table 5.

Table 4 Descriptive characteristics at pretreatment visit and the clinical outcome at week 8

	Pretreatment visit		**Clinical outcome at week 8**			
			NeuroPharm[1]		iSPOT-D[2]	
	Healthy controls	MDD	Remitters	Non-responders	Responders[4]	Non-responders[4]
N	35	91	21	15	44	35
Sex (M/F)	10/25	25/66	11/10	4/11	16/28	6/29
Age (Mean±SD)	29.0±9.7	27.4±8.3	29.4±9.7	25.7±9.3	28.5±8.8	25.5±7.2
Education	16.0±1.4	14.9±2.2[3]	15.8±1.3	14.8±2.3	15.2±1.9	14.6±2.5
GAD_{10}		22.9±9.7	22.3±8.5	22.1±8.6	23.4±7.9	22.1±10.3
Pretreatment HDRS		12.4±1.7[1] (22.9±3.4[2])	11.8±1.6	11.4±1.8	23.0±3.5	22.1±3.0
Week 8 HDRS			2.4±1.3	10.1±2.4[5]	7.0±3.3	17.5±4.5[5]

Notes: [1] $HDRS_6$ scores are shown in both pretreatment and week 8 $HDRS_6$ scores
[2] $HDRS_{17}$ scores are shown in both pretreatment and week 8 $HDRS_{17}$ scores
[3] Healthy controls had a significant higher education score compared to MDD (t (98) = -2.607, p = .011)
[4] Responders were defined by at least 50% improvement of depressive symptoms assessed by $HDRS_{17}$ score. $HDRS_{17}$ scores are shown in both pretreatment and week 8 $HDRS_{17}$ scores
[5] Between group comparisons showed that remitters had significant lower $HDRS_6$ score at week 8 compared to non-responders (t (34) = -12.71, p < .001); responders had significant lower $HDRS_{17}$ score at week 8 compared to non-responders (t (77) = -11.90, p < .001). No significant was found for other demographic characteristics (p values >.05).
Abbreviation: GAD_{10}, generalized anxiety disorder-10 score; HDRS, Hamilton Depression Rating Scale.

Table 5 Vigilance parameters at pretreatment visit and at week 8

	Pretreatment visit		**Clinical outcome at week 8**			
			NeuroPharm[1]		iSPOT-D[2]	
	Healthy controls	MDD	Remitters	Non-responders	Responders	Non-responders
Stage 0 (%, Mean±SD)	13.7±3.3	15.3±2.1	14.2±5.1	22.6±6.0	16.2±3.4	15.3±3.6
Stage A1	32.6±5.7	36.3±3.5	36.1±7.7	35.4±9.2	35.2±5.3	37.0±5.7
Stage A23	10.1±3.2	14.1±2.0	14.1±3.9	6.1±4.6	13.0±3.3	17.1±3.5
Stage B1	34.8±4.4	26.4±2.7	23.2±5.5	23.6±6.6	27.7±3.9	22.0±4.1
Stage B23	8.8±2.4	7.8±1.5	12.4±4.8	12.4±5.7	7.9±2.4	8.5±2.6
Median vigilance (Mean±SD)	4.07±0.15	4.27±0.93	4.24±0.25	4.31±0.30	4.28±0.15	4.30±0.16
Vigilance slope at 1st recording (Mean±SD)	-0.26±0.43	-0.02±0.50*	-0.09±0.39	0.03±0.68	-0.11±0.48	0.13±0.54*
Vigilance slope at 2nd recording (Mean±SD)	-0.01±0.51	0.05±0.43	0.10±0.24	0.16±0.56	0.10±0.38	0.06±0.42

Notes: * $p < .05$, age was included as covariate in all models

Vigilance dysregulation in MDD

As expected, we observed a less declines in vigilance slope in pretreatment MDD compared to healthy controls (F (1, 123) = 4.59, p = .034, Figure 14).

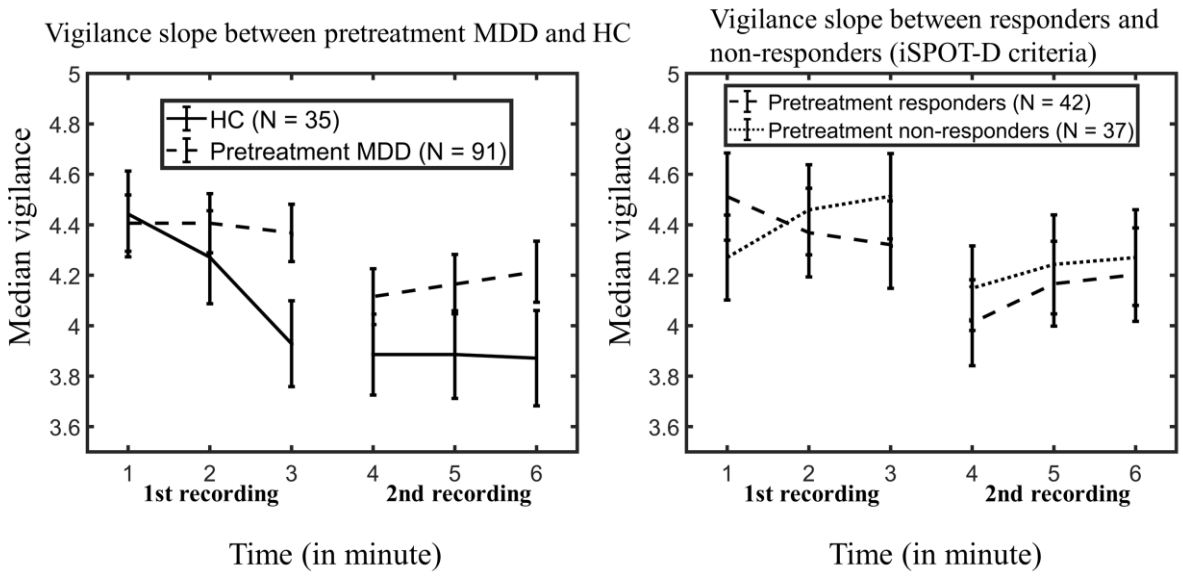


Figure 14 Vigilance slope for two separate eyes closed recordings. Left panel: pretreatment MDD had less decline toward sleep stage compared to healthy controls. Right panel: pretreatment responders defined by iSPOT-D criteria showed faster declines on vigilance slope compared non-responders in the first 3 min recording.

Our results confirm that patients suffering from MDD show a more rigid vigilance regulation during rest (Hegerl et al., 2012; Olbrich, Sander, Minkwitz, et al., 2012). The hyperstable wakefulness may reflect the difficulties of depressed patients to fall asleep. Because of consistently having high inner tension, MDD patients may try to counteract it by avoiding arousing activities (Hegerl et al., 2012) resulting in at behavioural level a lack of drive. Losing the ability to experience pleasure in activities is also a key symptom in depression. On the other hand, in comparison to prior long recording resting EEG (15 min), our results support that the vigilance dysregulation can be assessed in a relatively short EEG recording window, i.e. 3 min.

This short window opens op the feasibility of vigilance-informed prescription in future clinical use.

VIGALL as predictor for clinical outcome

We found that responders defined by iSPOT-D criteria had more pronounced declines toward sleep stage compared to non-responders in the first 3 min recording (F (1, 76) = 4.16, p = .045, Figure 14). However, the area under curve of ROC was only .60 (p = .12). No significant result was found when NeuroPharm criteria was being assessed (p values > .51).

This is the first independent replication of treatment prediction using EEG vigilance measures and found that faster declines of wakefulness regulation are linked to better clinical outcome. This is consistent with the largest multisite EEG study on treatment prediction on MDD (n = 599) (Olbrich et al., 2016). However, the current study could only replicate the association between the slope of vigilance and the response to SSRI treatment when using the same criteria of clinical outcome as they did in the previous study (Olbrich et al., 2016). The replication did not apply when the NeuroPharm criteria was being assessed. This heightens the importance of using the same outcome measures when comparing studies (Widge et al., 2018).

Treatment effects on VIGALL parameter

We compared the wakefulness before and after drug administration and found that depressed patients in general had a higher amount of stage B1 (F (4, 152) = 3.11, p = .026, Figure 15) after 8 weeks of SSRI/SNRI treatment compared to pretreatment EEG, regardless of their clinical outcome.

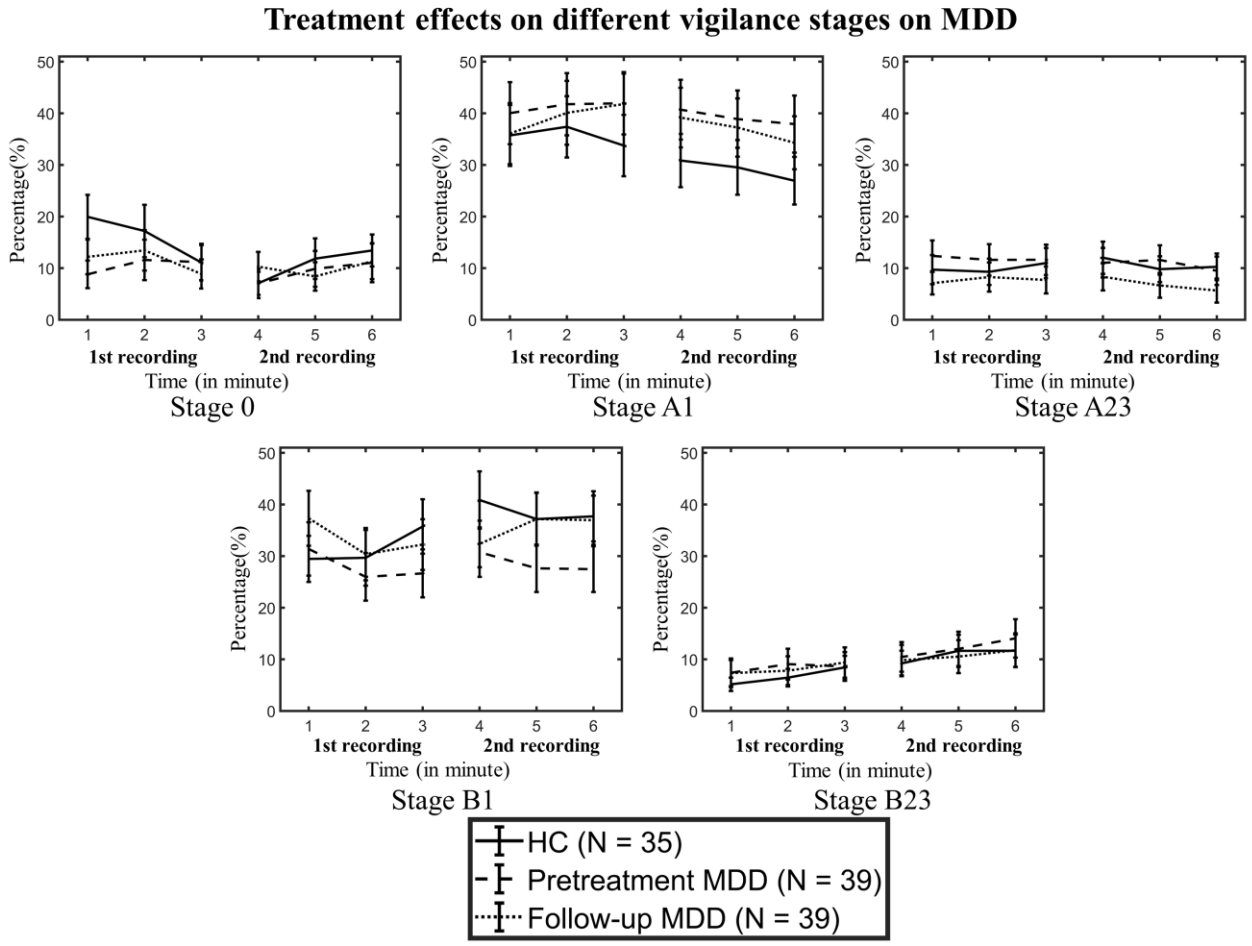


Figure 15 Vigilance regulation changes after 8 weeks on SSRI/SNRI treatment. Depressed patients had a higher amount of B1 stage 8 weeks after treatment.

Next, we investigated whether the treatment effects on patients would perform differently between different clinical outcomes. Interestingly, we found that only patients with good treatment response had higher amount of stage B1 after 8 weeks of treatment (Figure 16). This treatment effect further revealed a normalization pattern towards healthy controls. Before drug administration, there was lower percentage of stage B1 between depressed patients and controls, as well as patients with good response and controls. While the difference at stage B1 was disappeared after 8 weeks of treatment (Figure 16).

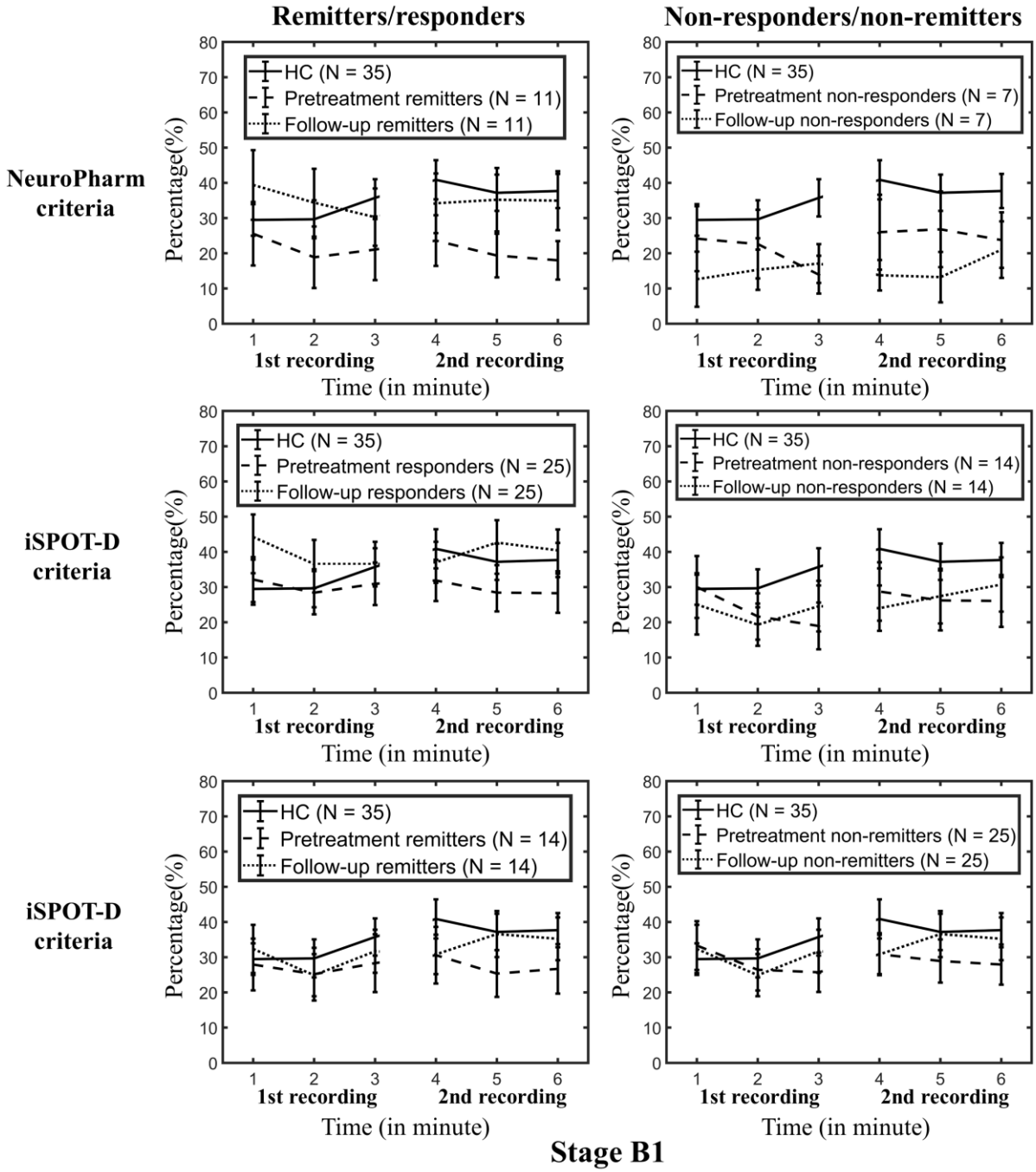


Figure 16 Vigilance regulation changes at stage B1 after 8 weeks on SSRI/SNRI treatment on different clinical outcome. Only patients with good clinical response (remitters/responders) showed a higher amount of stage B1 after 8 weeks of treatment. This improvement was not found on non-remitters and non-responders.

This is the first evidence to show the SSRI/SNRI treatment effect on patients suffering from MDD. Our results demonstrated that depressed patients with good clinical response had increased downregulation of EEG-vigilance after 8 weeks of SSRI/SNRI treatment, regardless of clinical criteria of treatment responses. The effect was normalized towards the vigilance patterns

of healthy controls. This was not found on non-responders and non-remitters (neither the NeuroPharm nor iSPOT-D criterion). Intriguingly, the treatment effect was only restricted to the stage B1, a brain state between alpha-dominated period and the occurrence of slow wave activity just before sleep onset. It may hypothesize that SSRI/SNRI treatment helps patients to pass through these desynchronized stages and to achieve recreational rest.

Methodological considerations

Only male subjects were included when assessing the reliability of EEG parameters

We excluded women when testing the test-retest reliability of EEG parameters in study I. The main reason was to avoid the possible confounder of the menstrual cycle, which might result in lower reliability of data. Previous studies have demonstrated changes in ERPs across menstrual cycles (O'Reilly, Cunningham, Lawlor, Walsh, & Rowan, 2004; Walpurger, Pietrowsky, Kirschbaum, & Wolf, 2004), but the effects of different menstruation phases and the relation between phases and individual ERPs remain unclear. For instance, no significant effect of the menstrual cycle could be detected on the P300 in the study of Walpurger et al., (2004), while greater P300 amplitude during menses was observed in the study of O'Reilly et al., (2004). To control the variables in our research, we therefore excluded women from our recruited population. The exclusion, on the other hand, could lower the generalizability of our data.

Inconsistent results when different scales or clinical criteria were assessed

When examining the association between $HDRS_6$ and the treatment response at week 8, we could not see the same effects as when $HDRS_{17}$ was applied. Null effects were reported for $HDRS_{17}$ when assessing the association between FAA and symptom improvement at week 8 (study II). The inconsistency might indicate that the $HDRS_6$ subscale is more sensitive to frontal alpha asymmetry and the treatment outcome. As pointed out in earlier reports, $HDRS_6$ subscale has an advantage for excluding three negative side effects (item 12–somatic symptoms, gastrointestinal; item 14–genital symptoms; and item 16–loss of weight), and thus, it could result in a larger power compared to when $HDRS_{17}$ was applied on different antidepressant responses (Bech et al., 2010, 2006; Østergaard et al., 2016). The exclusion of the less relevant items should enlarge the detection of the treatment effect. Likewise, the results on theta ACC demonstrated

that a significant difference in treatment responses was only present when $HDRS_6$ was used, but not when $HDRS_{17}$ was used. To sum up, $HDRS_6$ subscale might be capturing the key items reflected in alpha asymmetry and theta ACC and including more items could obscure the effect. Therefore, $HDRS_6$ should be favoured in detecting the subtle differences in physiological signals of treatment responses and it could be beneficial to studies with a small sample size.

The variance of Alpha peak frequency between participants

Alpha frequency seems to play an important role in predicting clinical outcome in MDD. Patients with right dominant alpha are more likely to benefit from SSRI treatment (study II) and alpha is also a crucial factor that contributes to VIGALL classification (study III). The classic alpha defined range is 8–12 Hz. However, it is noteworthy that alpha peak frequency (APF) could vary from person to person and in some cases, APF could happen below 8 Hz (Van Beijsterveldt & Van Baal, 2002). In those cases, traditional power extraction would miss the synchronized alpha activity. Besides APF's potential as an endophenotype, previous studies have linked APF to psychiatric disorders and found that it tends to have serious medical implications for patients with slow alpha activity (Boutros, 1996). Furthermore, prior evidence has shown that patients with pretreatment for slow APF are more likely to be clinical responders to sertraline (Arns, Gordon, & Boutros, 2017) and APF seems to be faster in responders after 4 weeks of pharmacotherapy (Ulrich et al., 1984). Therefore, individual adjustment should be considered when alpha frequency is being used as a predictive biomarker. In our application, APF was identified and adjusted on the individual level in VIGALL measures and the defined range of alpha spanned below 8 Hz in the calculation of frontal alpha asymmetry. However, the adjustment was not performed when general alpha frequency was assessed, thus the present results could have overlooked the synchronized alpha activity in patients with different clinical outcomes.

Drug effect or the condition of the patient

Since there was no placebo included in the NeuroPharm trial, we do not know for sure whether the treatment effect at vigilance regulation (study III) is due to the use of the drug itself or simply to the condition of the patient. Patients could have lowered vigilance and an increase of downregulated vigilance after the 8-week time course, or the clinical condition (remitters or responders) could result in an improvement in vigilance measures. However, preclinical studies showed that depressed symptoms at behavioural level are associated with an increased firing rate of the locus coeruleus (LC), where most of the noradrenergic neurons are located (reviewed by Hegerl & Hensch, 2014). SSRI treatment and most of other antidepressive measures could decrease the firing rate of the LC and thus decrease the vigilance states. Future investigations on the drug's real effect on vigilance measures could help bring some certainty.

Conclusion and future research perspectives

Our findings underscore the importance of reliability and replicability of EEG biomarker before it can be used in clinical practise. Spontaneous (resting) EEG is highly reliable over time and could be used for assessing pharmacological intervention with CNS effects. Its reliability is a useful feature of a candidate biomarker. Alpha asymmetry and vigilance measures, acquired from resting EEG, are associated with the clinical response to SSRI/SNRI treatment. In study II, we found that female patients with a greater left frontal activation seem to benefit more from pharmacological treatments. Patients with rigid vigilance regulation are less likely to respond to escitalopram/duloxetine treatments. Furthermore, vigilance regulations of responders and remitters normalize towards the profiles of healthy controls after successful treatment.

Given that the EEG studies included here are replication studies, the use of EEG as biomarkers is thus reliable enough to be included in the treatment process and could aid its selection. Our findings therefore encourage an EEG-informed treatment strategy in a shared decision process, including:

1) Apply EEG recording (following the International Pharmaco-EEG Society guidelines: Jobert et al., 2012) on patients prior a treatment decision.
2) Identify alpha asymmetry and vigilance measures on the patients.
3) Female patients with right dominate frontal alpha should be given escitalopram.
4) Patients with faster decline toward lower vigilance stages should be given escitalopram.
5) Initiate other treatments than SSRI for female patients with greater left central alpha.
6) Initiate other treatments than SSRI for male patients with greater right posterior alpha.
7) Initiate other treatments than SSRI for patients with rigid vigilance profiles.

8) Assess treatment efficacy of novel drugs for patients without the profile of serotonin dysfunction, such as drugs that do not target the serotonergic system.

Future prospective trials with larger samples are needed to develop a more complete drug stratification program. These samples, especially, should include patients who are non-responders to escitalopram, which is often considered as the first-line treatment. Better treatment approaches should be developed to improve patients' response and remission rates. Moreover, the treatments included in the current research had mainly focused on standardized antidepressants such as sertraline, escitalopram, duloxetine and venlafaxine. The future strategy of EEG-informed prescription would benefit from involving more antidepressants as prior evidence of biomarkers have shown a certain level of treatment specificity. Furthermore, it is possible that biomarkers originated from cross-modalities could provide a more precise treatment selection strategy. Such integration, as a yet unexplored advantage of NeuroPharm trial, could provide us the opportunity to better understand the underpinning neural correlates of depressed symptoms and the treatment response prediction in MDD. Last but not least, several biomarkers have been investigated using the state-of-the-art tools such as MR, PET and EEG. However, the clinical outcome of treatments is still based on multidimensional scales (Faries et al., 2000; Gibbons, Clark, & Kupfer, 1993). This may reduce the validity to detect the true treatment response of symptoms. For instance, patients could have adequate improvement on specific depressive symptoms (even core symptoms) while the responses might be neglected by a practitioner giving a judgement based on overall scores. Future development of biomarker should rely on specific depressive symptoms rather than the gross outcome of current diagnosis instruments.

References

Alhaj, H., Wisniewski, G., & McAllister-Williams, R. H. (2011). The use of the EEG in measuring therapeutic drug action: Focus on depression and antidepressants. *Journal of Psychopharmacology*. https://doi.org/10.1177/0269881110388323

American Psychological Association (APA). (2013). Diagnostic and Statistical Manual of Mental Disorders: Depressive Disorders. In *Diagnostic and Statistical Manual of Mental Disorders,*. https://doi.org/10.1176/appi.books.9780890425596.dsm04

Anderer, P., Saletu, B., & Pascual-Marqui, R. D. (2000). Effect of the 5-HT(1A) partial agonist buspirone on regional brain electrical activity in man: A functional neuroimaging study using low-resolution electromagnetic tomography (LORETA). *Psychiatry Research - Neuroimaging*. https://doi.org/10.1016/S0925-4927(00)00066-4

Arns, M., Bruder, G., Hegerl, U., Spooner, C., Palmer, D. M., Etkin, A., … Gordon, E. (2016). EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized iSPOT-D study. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2015.05.032

Arns, M., Drinkenburg, W. H., Fitzgerald, P. B., & Kenemans, J. L. (2012). Neurophysiological predictors of non-response to rTMS in depression. *Brain Stimulation*, *5*(4), 569–576. https://doi.org/10.1016/j.brs.2011.12.003

Arns, M., Etkin, A., Hegerl, U., Williams, L. M., DeBattista, C., Palmer, D. M., … Gordon, E. (2015). Frontal and rostral anterior cingulate (rACC) theta EEG in depression: Implications for treatment outcome? *European Neuropsychopharmacology*, *25*(8), 1190–1200. https://doi.org/10.1016/j.euroneuro.2015.03.007

Arns, M., Gordon, E., & Boutros, N. N. (2017). EEG Abnormalities Are Associated with Poorer Depressive Symptom Outcomes with Escitalopram and Venlafaxine-XR, but Not Sertraline. *Clinical EEG and Neuroscience*. https://doi.org/10.1177/1550059415621435

Arns, M., & Olbrich, S. (2014). Personalized medicine in ADHD and depression: Use of pharmaco-EEG. *Current Topics in Behavioral Neurosciences*. https://doi.org/10.1007/7854_2014_295

Bandelow, B. (2006). Defining response and remission in anxiety disorders: Toward an integrated approach. *CNS Spectrums*. https://doi.org/10.1017/s1092852900025815

Barden, N., Reul, J. M. H. M., & Holsboer, F. (1995). Do antidepressants stabilize mood through actions on the hypothalamic-pituitary-adrenocortical system? *Trends in Neurosciences*. https://doi.org/10.1016/0166-2236(95)93942-Q

Bares, M., Brunovsky, M., Kopecek, M., Novak, T., Stopkova, P., Kozeny, J., … Höschl, C. (2008). Early reduction in prefrontal theta QEEG cordance value predicts response to venlafaxine treatment in patients with resistant depressive disorder. *European Psychiatry*. https://doi.org/10.1016/j.eurpsy.2008.03.001

Baskaran, A., Milev, R., & McIntyre, R. S. (2012). The neurobiology of the EEG biomarker as a predictor of treatment response in depression. *Neuropharmacology*, *63*(4), 507–513. https://doi.org/10.1016/j.neuropharm.2012.04.021

Bauer, M., Pfennig, A., Severus, E., Whybrow, P. C., Angst, J., Möller, H. J., … Yamada, K. (2013). World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Unipolar Depressive Disorders, Part 1: Update 2013 on the acute and continuation treatment of unipolar depressive disorders. *World Journal of Biological Psychiatry*. https://doi.org/10.3109/15622975.2013.804195

Bech, P., Boyer, P., Germain, J. M., Padmanabhan, K., Haudiquet, V., Pitrosky, B., & Tourian, K. A. (2010). HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. *Pharmacopsychiatry*. https://doi.org/10.1055/s-0030-1263173

Bech, P., Kajdasz, D. K., & Porsdal, V. (2006). Dose-response relationship of duloxetine in placebo-controlled clinical trials in patients with major depressive disorder. *Psychopharmacology*, *188*(3), 273–280. https://doi.org/10.1007/s00213-006-0505-1

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*. https://doi.org/10.1001/archpsyc.1961.01710120031004

Belmaker, R. H., & Agam, G. (2008). Major depressive disorder. *New England Journal of Medicine*. https://doi.org/10.1056/NEJMra073096

Benkert, O., Gründer, G., & Wetzel, H. (1997). Is there an advantage to venlafaxine in comparison with other antidepressants? *Human Psychopharmacology*. https://doi.org/10.1002/(SICI)1099-1077(199701/02)12:1<53::AID-HUP837>3.0.CO;2-5

Board, F., Wadeson, R., & Persky, H. (1957). Depressive Affect and Endocrine Functions: Blood Levels of Adrenal Cortex and Thyroid Hormones in Patients Suffering from Depressive Reactions. *Archives of Neurology And Psychiatry*. https://doi.org/10.1001/archneurpsyc.1957.02330420072015

Bosker, F. J., Westerink, B. H. C., Cremers, T. I. F. H., Gerrits, M., Van Der Hart, M. G. C., Kuipers, S. D., … Korf, J. (2004). Future antidepressants: What is in the pipeline and what is missing? *CNS Drugs*. https://doi.org/10.2165/00023210-200418110-00002

Boutros, N. N. (1996). Diffuse electroencephalogram slowing in psychiatric patients: A preliminary report. *Journal of Psychiatry and Neuroscience*.

Brassen, S., & Adler, G. (2003). Short-term Effects of Acetylcholinesterase Inhibitor Treatment on EEG and Memory Performance in Alzheimer Patients: An Open, Controlled Trial. *Pharmacopsychiatry*. https://doi.org/10.1055/s-2003-45118

Bravo, R., Matito, S., Cubero, J., Paredes, S. D., Franco, L., Rivero, M., … Barriga, C. (2013). Tryptophan-enriched cereal intake improves nocturnal sleep, melatonin, serotonin, and total antioxidant capacity levels and mood in elderly humans. *Age*. https://doi.org/10.1007/s11357-012-9419-5

Brienza, M., & Mecarelli, O. (2019). Neurophysiological Basis of EEG. In *Clinical Electroencephalography* (pp. 9–21). https://doi.org/10.1007/978-3-030-04573-9_2

Bronzino, J. D., Brusseau, J., Morgane, P. J., & Stern, W. C. (1972). Power spectrum analysis of EEG synchronization following application of serotonin to area postrema. *Annals of Biomedical Engineering*. https://doi.org/10.1007/BF02584210

Bruder, G. E., Sedoruk, J. P., Stewart, J. W., McGrath, P. J., Quitkin, F. M., & Tenke, C. E. (2008). Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: Pre- and post-treatment findings. *Biological Psychiatry*, *63*(12), 1171–1177. https://doi.org/10.1016/j.biopsych.2007.10.009

Bruder, G. E., Stewart, J. W., Tenke, C. E., McGrath, P. J., Leite, P., Bhattacharya, N., & Quitkin, F. M. (2001). Electroencephalographic and perceptual asymmetry differences between responders and nonresponders to an SSRI antidepressant. *Biological Psychiatry*, *49*(5), 416–425. https://doi.org/10.1016/S0006-3223(00)01016-7

Brunner, J. F., Hansen, T. I., Olsen, A., Skandsen, T., Håberg, A., & Kropotov, J. (2013). Long-term test-retest reliability of the P3 NoGo wave and two independent components decomposed from the P3 NoGo wave in a visual Go/NoGo task. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2013.06.005

Brunoni, A. R., Boggio, P. S., & Fregni, F. (2008). Can the “yin and yang” BDNF hypothesis be

used to predict the effects of rTMS treatment in neuropsychiatry? *Medical Hypotheses*. https://doi.org/10.1016/j.mehy.2008.02.011
Cape, E. G., & Jones, B. E. (1998). Differential modulation of high-frequency γ-electroencephalogram activity and sleep-wake state by noradrenaline and serotonin microinjections into the region of cholinergic basalis neurons. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.18-07-02653.1998
Carlsson, A., Shore, P. A., & Brodie, B. B. (1957). Release of serotonin from blood platelets by reserpine in vitro. *The Journal of Pharmacology and Experimental Therapeutics*.
Carpenter, T. P., Laney, T., & Mezulis, A. (2012). Religious coping, stress, and depressive symptoms among adolescents: A prospective study. *Psychology of Religion and Spirituality*. https://doi.org/10.1037/a0023155
Carrillo-De-La-Peña, M. T., Mavrogiorgou, P., Juckel, G., Hauke, W., Gallinat, J., Frodl, T., … Hegerl, U. (2000). Loudness dependence of auditory evoked potentials in obsessive-compulsive disorder: A pilot study. *Psychiatry Research*. https://doi.org/10.1016/S0165-1781(00)00106-2
Chalon, S., Pereira, A., Lainey, E., Vandenhende, F., Watkin, J. G., Staner, L., & Granier, L. A. (2005). Comparative effects of duloxetine and desipramine on sleep EEG in healthy subjects. *Psychopharmacology*. https://doi.org/10.1007/s00213-004-1961-0
Chen, B., Dowlatshahi, D., MacQueen, G. M., Wang, J. F., & Young, L. T. (2001). Increased hippocampal BDNF immunoreactivity in subjects treated with antidepressant medication. *Biological Psychiatry*. https://doi.org/10.1016/S0006-3223(01)01083-6
Cook, I. A., Hunter, A. M., Abrams, M., Siegman, B., & Leuchter, A. F. (2009). Midline and right frontal brain function as a physiologic biomarker of remission in major depression. *Psychiatry Research - Neuroimaging*, *174*(2), 152–157. https://doi.org/10.1016/j.pscychresns.2009.04.011
Cook, I. A., Leuchter, A. F., Witte, E., Abrams, M., Uijtdehaage, S. H. J., Stubbeman, W., … Anderson-Hanley, C. (1999). Neurophysiologic predictors of treatment response to fluoxetine in major depression. *Psychiatry Research*. https://doi.org/10.1016/S0165-1781(99)00010-4
Coppen Alec, Shaw, D., Herzberg, B., & Maggs, R. (1967). Tryptophan in the treatment of depression. *The Lancet*, *290*(7527), 1178–1180.
Crane, G. E. (1956). The psychiatric side-effects of iproniazid. *The American Journal of Psychiatry*. https://doi.org/10.1176/ajp.112.7.494
Cuijpers, P., Geraedts, A. S., Van Oppen, P., Andersson, G., Markowitz, J. C., & Van Straten, A. (2011). Interpersonal psychotherapy for depression: A meta-analysis. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2010.10101411
Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for Depression in Adults: A Meta-Analysis of Comparative Outcome Studies. *Journal of Consulting and Clinical Psychology*. https://doi.org/10.1037/a0013075
Downar, J., & Daskalakis, Z. J. (2013). New targets for rTMS in depression: A review of convergent evidence. *Brain Stimulation*. https://doi.org/10.1016/j.brs.2012.08.006
Duman, R. S., Heninger, G. R., & Nestler, E. J. (1997). A molecular and cellular theory of depression. *Archives of General Psychiatry*. https://doi.org/10.1001/archpsyc.1997.01830190015002
Dwivedi, Y. (2009). Brain-derived neurotrophic factor: Role in depression and suicide. *Neuropsychiatric Disease and Treatment*. https://doi.org/10.2147/ndt.s5700
Egeland, M., Zunszain, P. A., & Pariante, C. M. (2015). Molecular mechanisms in the regulation of adult neurogenesis during stress. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3855

Ernest Lyons, W., Mamounas, L. A., Ricaurte, G. A., Coppola, V., Reid, S. W., Bora, S. H., … Tessarollo, L. (1999). Brain-derived neurotrophic factor-deficient mice develop aggressiveness and hyperphagia in conjunction with brain serotonergic abnormalities. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.96.26.15239

Faries, D., Herrera, J., Rayamajhi, J., Debrota, D., Demitrack, M., & Potter, W. Z. (2000). The responsiveness of the Hamilton Depression Rating Scale. *Journal of Psychiatric Research*. https://doi.org/10.1016/S0022-3956(99)00037-0

Fava, M., & Kendler, K. S. (2000). Major depressive disorder. *Neuron*. https://doi.org/10.1016/S0896-6273(00)00112-4

Feffer, K., Fettes, P., Giacobbe, P., Daskalakis, Z. J., Blumberger, D. M., & Downar, J. (2018). 1 Hz rTMS of the right orbitofrontal cortex for major depression: Safety, tolerability and clinical outcomes. *European Neuropsychopharmacology*. https://doi.org/10.1016/j.euroneuro.2017.11.011

Feighner, J. P., & Cohn, J. B. (1985). Double-blind comparative trials of fluoxetine and doxepin in geriatric patients with major depressive disorder. *Journal of Clinical Psychiatry*.

Ferrari, A. J., Somerville, A. J., Baxter, A. J., Norman, R., Patten, S. B., Vos, T., & Whiteford, H. A. (2013). Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychological Medicine*. https://doi.org/10.1017/S0033291712001511

Feyissa, A. M., & Tatum, W. O. (2019). Adult EEG. In *Handbook of Clinical Neurology*. https://doi.org/10.1016/B978-0-444-64032-1.00007-2

Fink, M., Wadsak, W., Savli, M., Stein, P., Moser, U., Hahn, A., … Lanzenberger, R. (2009). Lateralization of the serotonin-1A receptor distribution in language areas revealed by PET. *NeuroImage*, *45*(2), 598–605. https://doi.org/10.1016/j.neuroimage.2008.11.033

Freis, E. D. (1954). Mental depression in hypertensive patients treated for long periods with large doses of reserpine. *The New England Journal of Medicine*. https://doi.org/10.1056/NEJM195412162512504

Fumoto, M., Oshima, T., Kamiya, K., Kikuchi, H., Seki, Y., Nakatani, Y., … Arita, H. (2010). Ventral prefrontal cortex and serotonergic system activation during pedaling exercise induces negative mood improvement and increased alpha band in EEG. *Behavioural Brain Research*. https://doi.org/10.1016/j.bbr.2010.04.017

Gallinat, J., Bottlender, R., Juckel, G., & Stotz, A. M. G. (2000). *The loudness dependency of the auditory evoked N1/P2-component as a predictor of the acute SSRI response in depression*. *148*(4), 404–411. https://doi.org/https://doi.org/10.1007/s002130050070

George, M. S., Raman, R., Benedek, D. M., Pelic, C. G., Grammer, G. G., Stokes, K. T., … Stein, M. B. (2014). A two-site pilot randomized 3 day trial of high dose left prefrontal repetitive transcranial magnetic stimulation (rTMS) for suicidal inpatients. *Brain Stimulation*. https://doi.org/10.1016/j.brs.2014.03.006

George, M. S., Taylor, J. J., & Short, E. B. (2013). The expanding evidence base for rTMS treatment of depression. *Current Opinion in Psychiatry*. https://doi.org/10.1097/YCO.0b013e32835ab46d

Gibbons, R. D., Clark, D. C., & Kupfer, D. J. (1993). Exactly what does the Hamilton depression rating scale measure? *Journal of Psychiatric Research*. https://doi.org/10.1016/0022-3956(93)90037-3

Goldberg, J. S., Bell, C. E., & Pollard, D. A. (2014). Revisiting the monoamine hypothesis of depression: A new perspective. *Perspectives in Medicinal Chemistry*. https://doi.org/10.4137/PMc.s11375

Gonul, A. S., Akdeniz, F., Taneli, F., Donat, O., Eker, Ç., & Vahip, S. (2005). Effect of

treatment on serum brain-derived neurotrophic factor levels in depressed patients. *European Archives of Psychiatry and Clinical Neuroscience*. https://doi.org/10.1007/s00406-005-0578-6

Groen, Y., Wijers, A. A., Mulder, L. J. M., Waggeveld, B., Minderaa, R. B., & Althaus, M. (2008). Error and feedback processing in children with ADHD and children with Autistic Spectrum Disorder: An EEG event-related potential study. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2008.08.004

Gudmundsson, S., Runarsson, T. P., Sigurdsson, S., Eiriksdottir, G., & Johnsen, K. (2007). Reliability of quantitative EEG features. *Clinical Neurophysiology*, *118*(10), 2162–2171. https://doi.org/10.1016/j.clinph.2007.06.018

Guille, V., Croft, R. J., O'Neill, B. V., Illic, S., Phan, K. L., & Nathan, P. J. (2008). An examination of acute changes in serotonergic neurotransmission using the loudness dependence measure of auditory cortex evoked activity: Effects of citalopram, escitalopram and sertraline. *Human Psychopharmacology*. https://doi.org/10.1002/hup.922

Guy, W. (1976). CGI Clinical Global Impressions. *ECDEU Assessment Manual*.

Haghighi, M., Salehi, I., Erfani, P., Jahangard, L., Bajoghli, H., Holsboer-Trachsler, E., & Brand, S. (2013). Additional ECT increases BDNF-levels in patients suffering from major depressive disorders compared to patients treated with citalopram only. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2013.03.006

Hamilton, M. (1967). Development of a Rating Scale for Primary Depressive Illness. *British Journal of Social and Clinical Psychology*. https://doi.org/10.1111/j.2044-8260.1967.tb00530.x

Hämmerer, D., Li, S. C., Völkle, M., Müller, V., & Lindenberger, U. (2013). A lifespan comparison of the reliability, test-retest stability, and signal-to-noise ratio of event-related potentials assessed during performance monitoring. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.2012.01476.x

Hegerl, U., Gallinat, J., & Juckel, G. (2001). Event-related potentials: Do they reflect central serotonergic neurotransmission and do they predict clinical response to serotonin agonists? *Journal of Affective Disorders*. https://doi.org/10.1016/S0165-0327(00)00353-0

Hegerl, Ulrich, & Hensch, T. (2014). The vigilance regulation model of affective disorders and ADHD. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2012.10.008

Hegerl, Ulrich, Wilk, K., Olbrich, S., Schoenknecht, P., & Sander, C. (2012). Hyperstable regulation of vigilance in patients with major depressive disorder. *World Journal of Biological Psychiatry*. https://doi.org/10.3109/15622975.2011.579164

Heine, W., Dobrota, M. A., Schomer, D., Wigton, R., & Herman, S. (2020). High-Density Adaptive Ten Ten: Proposal for Electrode Nomenclature for High-Density EEG. *Journal of Clinical Neurophysiology : Official Publication of the American Electroencephalographic Society*. https://doi.org/10.1097/WNP.0000000000000632

Hennings, J. M., Owashi, T., Binder, E. B., Horstmann, S., Menke, A., Kloiber, S., … Lucae, S. (2009). Clinical characteristics and treatment outcome in a representative sample of depressed inpatients - Findings from the Munich Antidepressant Response Signature (MARS) project. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2008.05.002

Hillyard, S. A., Luck, S. J., & Mangun, G. R. (1994). The Cuing of Attention to Visual Field Locations: Analysis with ERP Recordings. In *Cognitive Electrophysiology*. https://doi.org/10.1007/978-1-4612-0283-7_1

Hirschfeld, R. M. A. (2000). History and evolution of the monoamine hypothesis of depression. *Journal of Clinical Psychiatry*.

Höller, Y., Helmstaedter, C., & Lehnertz, K. (2018). Quantitative Pharmaco-Electroencephalography in Antiepileptic Drug Research. *CNS Drugs*. https://doi.org/10.1007/s40263-018-0557-x

Hollon, S. D., Jarrett, R. B., Nierenberg, A. A., Thase, M. E., Trivedi, M., & Rush, A. J. (2005). Psychotherapy and medication in the treatment of adult and geriatric depression: Which monotherapy or combined treatment? *Journal of Clinical Psychiatry*. https://doi.org/10.4088/JCP.v66n0408

Holmes, G. L., & Khazipov, R. (2007). Basic neurophysiology and the cortical basis of EEG. In *The Clinical Neurophysiology Primer*. https://doi.org/10.1007/978-1-59745-271-7_2

Holmes, M. (2008). Dense array EEG: Methodology and new hypothesis on epilepsy syndromes. *Epilepsia*. https://doi.org/10.1111/j.1528-1167.2008.01505.x

Hunter, A. M., Minzenberg, M. J., Cook, I. A., Krantz, D. E., Levitt, J. G., Rotstein, N. M., … Leuchter, A. F. (2019). Concomitant medication use and clinical outcome of repetitive Transcranial Magnetic Stimulation (rTMS) treatment of Major Depressive Disorder. *Brain and Behavior*. https://doi.org/10.1002/brb3.1275

Husain, M. M., Rush, A. J., Fink, M., Knapp, R., Petrides, G., Rummans, T., … Kellner, C. H. (2004). Speed of response and remission in major depressive disorder with acute electroconvulsive therapy (ECT): A consortium for research in ECT (CORE) report. *Journal of Clinical Psychiatry*. https://doi.org/10.4088/JCP.v65n0406

Iinuma, K., Tamahashi, S., Otomo, H., Onuma, A., & Takamatsu, N. (1978). Immediate Changes of the Electroencephalograms After Intravenous Injection of Clonazepam and their Relation to its Effect on Clinical Fits in Children with Minor Seizures. *The Tohoku Journal of Experimental Medicine*. https://doi.org/10.1620/tjem.125.223

Iosifescu, D. (2011). Electroencephalography-derived biomarkers of antidepressant response. *Harvard Review of Psychiatry*, *19*(3), 144–154. https://doi.org/10.3109/10673229.2011.586549

Iosifescu, Dan V., Greenwald, S., Devlin, P., Mischoulon, D., Denninger, J. W., Alpert, J. E., & Fava, M. (2009). Frontal EEG predictors of treatment outcome in major depressive disorder. *European Neuropsychopharmacology*, *19*(11), 772–777. https://doi.org/10.1016/j.euroneuro.2009.06.001

Işintaş, M., Ak, M., Erdem, M., Öz, O., & Özgen, F. (2012). Event-related potentials in major depressive disorder: The relationship between P300 and treatment response. *Turk Psikiyatri Dergisi*. https://doi.org/10.5080/u6650

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*.

Jaworska, N., & Protzner, A. (2013). Electrocortical features of depression and their clinical utility in assessing antidepressant treatment outcome. *Canadian Journal of Psychiatry*, *58*(9), 509–514. https://doi.org/10.1177/070674371305800905

Jobert, M., Wilson, F. J., Ruigt, G. S. F., Brunovsky, M., Prichep, L. S., & Drinkenburg, W. H. I. M. (2012). Guidelines for the recording and evaluation of pharmaco-eeg data in man: The international pharmaco-EEG society (IPEG): The IPEG pharmaco-EEG guideline committee. *Neuropsychobiology*. https://doi.org/10.1159/000343478

Juckel, G., Hegerl, U., Molnár, M., Csépe, V., & Karmos, G. (1999). Auditory evoked potentials reflect serotonergic neuronal activity - A study in behaving cats administered drugs acting on 5-HT(1A) autoreceptors in the dorsal raphe nucleus. *Neuropsychopharmacology*. https://doi.org/10.1016/S0893-133X(99)00074-3

Juckel, G., Molnár, M., Hegerl, U., Csépe, V., & Karmos, G. (1997). Auditory-evoked potentials as indicator of brain serotonergic activity - First evidence in behaving cats. *Biological Psychiatry*. https://doi.org/10.1016/S0006-3223(96)00240-5

Juckel, G., Pogarell, O., Augustin, H., Mulert, C., Müller-Siecheneder, F., Frodl, T., … Hegerl, U. (2007). Differential Prediction of First Clinical Response to. *Psychiatry: Interpersonal and Biological Processes*, *68*(August), 1206–1212. https://doi.org/Authors: Georg Juckel, Oliver Pogarell, Holger Augustin, Christoph Mulert, Florian Müller-Siecheneder, Thomas Frodl, Paraskevi Mavrogiorgou, Ulrich Hegerl https://doi.org/10.4088/jcp.v68n0806

Kaf, W. A., Sabo, D. L., Durrant, J. D., & Rubinstein, E. (2006). Reliability of electric response audiometry using 80 Hz auditory steady-state responses. *International Journal of Audiology*. https://doi.org/10.1080/14992020600753197

Kafka, M. P. (2003). The monoamine hypothesis for the pathophysiology of paraphilic disorders: An update. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/j.1749-6632.2003.tb07295.x

Karege, F., Perret, G., Bondolfi, G., Schwald, M., Bertschy, G., & Aubry, J. M. (2002). Decreased serum brain-derived neurotrophic factor levels in major depressed patients. *Psychiatry Research*. https://doi.org/10.1016/S0165-1781(02)00005-7

Kearns, N. P., Cruickshank, C. A., McGuigan, K. J., Riley, S. A., Shaw, S. P., & Snaith, R. P. (1982). A comparison of depression rating scales. *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.141.1.45

Kellner, C. H., Greenberg, R. M., Murrough, J. W., Bryson, E. O., Briggs, M. C., & Pasculli, R. M. (2012). ECT in treatment-resistant depression. *American Journal of Psychiatry*. https://doi.org/10.1176/appi.ajp.2012.12050648

Kennard, B., Silva, S., Vitiello, B., Curry, J., Kratochvil, C., Simons, A., … March, J. (2006). Remission and residual symptoms after short-term treatment in the Treatment of Adolescents with Depression Study (TADS). *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1097/01.chi.0000242228.75516.21

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology. Supplement*.

Knott, V. J., Telner, J. I., Lapierre, Y. D., Browne, M., & Horn, E. R. (1996). Quantitative EEG in the prediction of antidepressant response to imipramine. *Journal of Affective Disorders*. https://doi.org/10.1016/0165-0327(96)00003-1

Knott, V., Mahoney, C., Kennedy, S., & Evans, K. (2000). Pre-treatment EEG and it's relationship to depression severity and paroxetine treatment outcome. *Pharmacopsychiatry*, *33*(6), 201–205. https://doi.org/10.1055/s-2000-8356

Koenig, T., Hernandez, L. D., & Rieger, K. (2016). Quantitative EEG in schizophrenia: Current state and future direction. *Epileptologie*, *33*(3), 183–188.

Köhler-Forsberg, K., Jørgensen, A., Dam, V., Stenbæk, D., Fisher, P., Ip, C.-T., … Frokjaer, V. (2020). Predicting treatment outcome in Major Depressive Disorder using serotonin 4 receptor PET brain imaging, functional MRI, cognitive-, EEG-based and peripheral biomarkers: a NeuroPharm open label clinical trial protocol. *Frontiers in Psychiatry*. https://doi.org/https://doi.org/10.3389/fpsyt.2020.00641

Korb, A. S., Hunter, A. M., Cook, I. A., & Leuchter, A. F. (2009). Rostral anterior cingulate cortex theta current density and response to antidepressants and placebo in major depression. *Clinical Neurophysiology*, *120*(7), 1313–1319. https://doi.org/10.1016/j.clinph.2009.05.008

Kranz, G. S., Hahn, A., Baldinger, P., Haeusler, D., Philippe, C., Kaufmann, U., … Lanzenberger, R. (2014). Cerebral serotonin transporter asymmetry in females, males and male-to-female transsexuals measured by PET in vivo. *Brain Structure and Function*, *219*(1), 171–183. https://doi.org/10.1007/s00429-012-0492-4

Kuhn, R. (1958). The treatment of depressive states with G 22355 (imipramine hydrochloride).

The American Journal of Psychiatry. https://doi.org/10.1176/ajp.115.5.459
La Vaque, T. J. (1999). The history of eeg hans berger: Psychophysiologist. a historical vignette. *Journal of Neurotherapy*. https://doi.org/10.1300/J184v03n02_01
Larson, M. J., Baldwin, S. A., Good, D. A., & Fair, J. E. (2010). Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): The role of number of trials. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.2010.01022.x
Lee, T. W., Wu, Y. Te, Yu, Y. W. Y., Chen, M. C., & Chen, T. J. (2011). The implication of functional connectivity strength in predicting treatment response of major depressive disorder: A resting EEG study. *Psychiatry Research - Neuroimaging*. https://doi.org/10.1016/j.pscychresns.2011.02.009
Lee, T. W., Yu, Y. W. Y., Chen, T. J., & Tsai, S. J. (2005). Loudness dependence of the auditory evoked potential and response to antidepressants in Chinese patients with major depression. *Journal of Psychiatry and Neuroscience*.
Lemere, F. (1936). The significance of individual differences in the Berger rhythm. *Brain*, *59*(3), 366–375. https://doi.org/https://doi.org/10.1093/brain/59.3.366
Leucht, S., Fennema, H., Engel, R. R., Kaspers-Janssen, M., Lepping, P., & Szegedi, A. (2017). What does the MADRS mean? Equipercentile linking with the CGI using a company database of mirtazapine studies. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2016.12.041
Leucht, S., Fennema, H., Engel, R. R., Kaspers-Janssen, M., & Szegedi, A. (2018). Translating the HAM-D into the MADRS and vice versa with equipercentile linking. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2017.09.042
Leuchter, A. F., Cook, I. A., Marangell, L. B., Gilmer, W. S., Burgoyne, K. S., Howland, R. H., … Greenwald, S. (2009). Comparative effectiveness of biomarkers and clinical indicators for predicting outcomes of SSRI treatment in Major Depressive Disorder: Results of the BRITE-MD study. *Psychiatry Research*. https://doi.org/10.1016/j.psychres.2009.06.004
Lisanby, S. H. (2007). Electroconvulsive therapy for depression. *New England Journal of Medicine*. https://doi.org/10.1056/NEJMct075234
Loo, S. K., Specter, E., Smolen, A., Hopper, C., Teale, P. D., & Reite, M. L. (2003). Functional effects of the DAT1 polymorphism on EEG measures in ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1097/01.CHI.0000046890.27264.88
Luck, S. J. (2005). Ten simple rules for designing and interpreting ERP experiments. In *Event-Related Potentials: A Methods Handbook*.
MacGillivray, S., Arroll, B., Hatcher, S., Ogston, S., Reid, I., Sullivan, F., … Crombie, I. (2003). Efficacy and tolerability of selective serotonin reuptake inhibitors compared with tricyclic antidepressants in depression treated in primary care: Systematic review and meta-analysis. *British Medical Journal*. https://doi.org/10.1136/bmj.326.7397.1014
Mathew, S. J., Wilkinson, S. T., Altinay, M., Asghar-Ali, A., Chang, L. C., Collins, K. A., … Anand, A. (2019). ELEctroconvulsive therapy (ECT) vs. Ketamine in patients with Treatment-resistant Depression: The ELEKT-D study protocol. *Contemporary Clinical Trials*. https://doi.org/10.1016/j.cct.2018.12.009
McAllister-Williams, R. H., & Massey, A. E. (2003). EEG effects of buspirone and pindolol: A method of examining 5-HT1A receptor function in humans. *Psychopharmacology*. https://doi.org/10.1007/s00213-002-1339-0
McAllister-Williams, R. Hamish, Massey, A. E., & Fairchild, G. (2007). Repeated cortisol administration attenuates the EEG response to buspirone in healthy volunteers: Evidence for desensitization of the 5-HT1A autoreceptor. *Journal of Psychopharmacology*. https://doi.org/10.1177/0269881107078292

McFadden, K. L., Steinmetz, S. E., Carroll, A. M., Simon, S. T., Wallace, A., & Rojas, D. C. (2014). Test-retest reliability of the 40 Hz EEG auditory steady-state response. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0085748

McGrath, C. L., Kelley, M. E., Holtzheimer, P. E., Dunlop, B. W., Craighead, W. E., Franco, A. R., … Mayberg, H. S. (2013). Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry*, *70*(8), 821–829. https://doi.org/https://doi.org/10.1001/jamapsychiatry.2013.143

McGraw, K. O., & Wong, S. P. (1996). Forming Inferences about Some Intraclass Correlation Coefficients. *Psychological Methods*. https://doi.org/10.1037/1082-989X.1.1.30

McKay, M. S., & Zakzanis, K. K. (2010). The impact of treatment on HPA axis activity in unipolar major depression. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2009.07.012

McNamara, B., Ray, J. L., Arthurs, O. J., & Boniface, S. (2001). Transcranial magnetic stimulation for depression and other psychiatric disorders. *Psychological Medicine*. https://doi.org/10.1017/S0033291701004378

Meador, K. J., Loring, D. W., Boyd, A., Echauz, J., LaRoche, S., Velez-Ruiz, N., … Webster, E. (2016). Randomized double-blind comparison of cognitive and EEG effects of lacosamide and carbamazepine. *Epilepsy and Behavior*. https://doi.org/10.1016/j.yebeh.2016.07.007

Mikati, M. A., Trevathan, E., Krishnamoorthy, K. S., & Lombroso, C. T. (1991). Pyridoxine-dependent epilepsy: EEG investigations and long-term follow-up. *Electroencephalography and Clinical Neurophysiology*. https://doi.org/10.1016/0013-4694(91)90035-3

Miller, J. M., Schneck, N., Siegle, G. J., Chen, Y., Ogden, R. T., Kikuchi, T., … Parsey, R. V. (2013). fMRI response to negative words and SSRI treatment outcome in major depressive disorder: A preliminary study. *Psychiatry Research - Neuroimaging*, *214*(3), 296–305. https://doi.org/10.1016/j.pscychresns.2013.08.001

Molendijk, M. L., Spinhoven, P., Polak, M., Bus, B. A. A., Penninx, B. W. J. H., & Elzinga, B. M. (2014). Serum BDNF concentrations as peripheral manifestations of depression: Evidence from a systematic review and meta-analyses on 179 associations (N=9484). *Molecular Psychiatry*. https://doi.org/10.1038/mp.2013.105

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.134.4.382

Mucci, A., Volpe, U., Merlotti, E., Bucci, P., & Galderisi, S. (2006). Pharmaco-EEG in psychiatry. *Clinical EEG and Neuroscience*. https://doi.org/10.1177/155005940603700206

Mulert, C., Juckel, G., Brunnmeier, M., Karch, S., Leicht, G., Mergl, R., … Pogarell, O. (2007). Prediction of treatment response in major depression: Integration of concepts. *Journal of Affective Disorders*, *98*(3), 215–225. https://doi.org/10.1016/j.jad.2006.07.021

Mumtaz, W., Malik, A. S., Yasin, M. A. M., & Xia, L. (2015). Review on EEG and ERP predictive biomarkers for major depressive disorder. *Biomedical Signal Processing and Control*, *22*, 85–98. https://doi.org/10.1016/j.bspc.2015.07.003

Murakami, S., & Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *Journal of Physiology*. https://doi.org/10.1113/jphysiol.2006.105379

Nathan, P. J., Segrave, R., Phan, K. L., O'Neill, B., & Croft, R. J. (2006). Direct evidence that acutely enhancing serotonin with the selective serotonin reuptake inhibitor citalopram modulates the loudness dependence of the auditory evoked potential (LDAEP) marker of central serotonin function. *Human Psychopharmacology*. https://doi.org/10.1002/hup.740

Nibuya, M., Morinobu, S., & Duman, R. S. (1995). Regulation of BDNF and trkB mRNA in rat brain by chronic electroconvulsive seizure and antidepressant drug treatments. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.15-11-07539.1995

Nutt, D. J., Wilson, S., & Paterson, L. (2008). Sleep disorders as core symptoms of depression. *Dialogues in Clinical Neuroscience*.
O'Neill, B. V., Croft, R. J., & Nathan, P. J. (2008). The loudness dependence of the auditory evoked potential (LDAEP) as an in iivo biomarker of central serotonergic function in humans: Rationale, evaluation and review of findings. *Human Psychopharmacology*. https://doi.org/10.1002/hup.940
O'Reilly, M. A., Cunningham, C. J., Lawlor, B. A., Walsh, C. D., & Rowan, M. J. (2004). The effect of the menstrual cycle on electrophysiological and behavioral measures of memory and mood. *Psychophysiology*. https://doi.org/10.1111/j.1469-8986.2004.00194.x
Olbrich, S., & Arns, M. (2013). EEG biomarkers in major depressive disorder: Discriminative power and prediction of treatment response. *International Review of Psychiatry*. https://doi.org/10.3109/09540261.2013.816269
Olbrich, S., & Conradi, J. (2016). Future of clinical EEG in psychiatric disorders: Shifting the focus from diagnosis to the choice of optimal treatment. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2015.06.018
Olbrich, S., Fischer, M. M., Sander, C., Hegerl, U., Wirtz, H., & Bosse-Henck, A. (2015). Objective markers for sleep propensity: Comparison between the Multiple Sleep Latency Test and the Vigilance Algorithm Leipzig. *Journal of Sleep Research*. https://doi.org/10.1111/jsr.12290
Olbrich, S., Olbrich, H., Adamaszek, M., Jahn, I., Hegerl, U., & Stengler, K. (2013). Altered EEG lagged coherence during rest in obsessive-compulsive disorder. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2013.05.031
Olbrich, S., Sander, C., Jahn, I., Eplinius, F., Claus, S., Mergl, R., … Hegerl, U. (2012). Unstable EEG-vigilance in patients with cancer-related fatigue (CRF) in comparison to healthy controls. *World Journal of Biological Psychiatry*. https://doi.org/10.3109/15622975.2010.545434
Olbrich, S., Sander, C., Minkwitz, J., Chittka, T., Mergl, R., Hegerl, U., & Himmerich, H. (2012). EEG vigilance regulation patterns and their discriminative power to separate patients with major depression from healthy controls. *Neuropsychobiology*. https://doi.org/10.1159/000337000
Olbrich, S., Tränkner, A., Surova, G., Gevirtz, R., Gordon, E., Hegerl, U., & Arns, M. (2016). CNS- and ANS-arousal predict response to antidepressant medication: Findings from the randomized iSPOT-D study. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2015.12.001
Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*. https://doi.org/10.1016/S1388-2457(00)00527-7
østergaard, S. D., Bech, P., & Miskowiak, K. W. (2016). Fewer study participants needed to demonstrate superior antidepressant efficacy when using the Hamilton melancholia subscale (HAM-D6) as outcome measure. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2014.10.047
Owens, M. J. (2004). Selectivity of antidepressants: From the monoamine hypothesis of depression to the SSRI revolution and beyond. *Journal of Clinical Psychiatry*.
Padberg, F., Di Michele, F., Zwanzger, P., Romeo, E., Bernardi, G., Schüle, C., … Rupprecht, R. (2002). Plasma concentrations of neuroactive steroids before and after repetitive transcranial magnetic stimulation (rTMS) in major depression. *Neuropsychopharmacology*. https://doi.org/10.1016/S0893-133X(02)00355-X
Pagnin, D., De Queiroz, V., Pini, S., & Cassano, G. B. (2004). Efficacy of ECT in Depression: A Meta-Analytic Review. *Journal of ECT*. https://doi.org/10.1097/00124509-200403000-

00004
Pariante, C. M., & Lightman, S. L. (2008). The HPA axis in major depression: classical theories and new developments. *Trends in Neurosciences*. https://doi.org/10.1016/j.tins.2008.06.006
Park, Y. M., Lee, S. H., Kim, S., & Bae, S. M. (2010). The loudness dependence of the auditory evoked potential (LDAEP) in schizophrenia, bipolar disorder, major depressive disorder, anxiety disorder, and healthy controls. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. https://doi.org/10.1016/j.pnpbp.2009.12.004
Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*(1), 49–65. https://doi.org/10.1016/0167-8760(84)90014-X
Pascual-Marqui, R., Lehmann, D., Koenig, T., Kochi, K., Merlo, C. G. M., Hell, D., & Koukkou, M. (1999). Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. *Psychiatry Research - Neuroimaging*. https://doi.org/10.1016/S0925-4927(99)00013-X
Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*. https://doi.org/10.1016/0013-4694(89)90180-6
Pillai, R. L. I., Bartlett, E. A., Ananth, M. R., Zhu, C., Yang, J., Hajcak, G., … DeLorenzo, C. (2020). Examining the underpinnings of loudness dependence of auditory evoked potentials with positron emission tomography. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2020.116733
Pizzagalli, D. A., Webb, C. A., Dillon, D. G., Tenke, C. E., Kayser, J., Goer, F., … Trivedi, M. H. (2018). Pretreatment rostral anterior cingulate cortex theta activity in relation to symptom improvement in depression: A randomized clinical trial. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2018.0252
Pizzagalli, D., Pascual-Marqui, R. D., Nitschke, J. B., Oakes, T. R., Larson, C. L., Abercrombie, H. C., … Davidson, R. J. (2001). Anterior cingulate activity as a predictor of degree of treatment response in major depression: Evidence from brain electrical tomography analysis. *American Journal of Psychiatry*, *158*(3), 405–415. https://doi.org/10.1176/appi.ajp.158.3.405
Puspita, J. W., Soemarno, G., Jaya, A. I., & Soewono, E. (2017). Interictal Epileptiform Discharges (IEDs) classification in EEG data of epilepsy patients. *Journal of Physics: Conference Series.*, *943*(1), 1–6. https://doi.org/10.1088/1742-6596/943/1/012030
Ren, S., Gliske, S. V., Brang, D., & Stacey, W. C. (2019). Redaction of false high frequency oscillations due to muscle artifact improves specificity to epileptic tissue. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2019.03.028
Rentzsch, J., Adli, M., Wiethoff, K., Gómez-Carrillo De Castro, A., & Gallinat, J. (2014). Pretreatment anterior cingulate activity predicts antidepressant treatment response in major depressive episodes. *European Archives of Psychiatry and Clinical Neuroscience*, *264*(3), 213–223. https://doi.org/10.1007/s00406-013-0424-1
Rentzsch, J., Jockers-Scherübl, M. C., Boutros, N. N., & Gallinat, J. (2008). Test-retest reliability of P50, N100 and P200 auditory sensory gating in healthy subjects. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2007.10.006
Riedel, M., Möller, H. J., Obermeier, M., Schennach-Wolff, R., Bauer, M., Adli, M., … Seemüller, F. (2010). Response and remission criteria in major depression - A validation of current practice. *Journal of Psychiatric Research*, *44*(15), 1063–1068. https://doi.org/10.1016/j.jpsychires.2010.03.006
Romo-Vázquez, R., Vélez-Pérez, H., Ranta, R., Louis Dorr, V., Maquin, D., & Maillard, L. (2012). Blind source separation, wavelet denoising and discriminant analysis for EEG

artefacts and noise cancelling. *Biomedical Signal Processing and Control*. https://doi.org/10.1016/j.bspc.2011.06.005

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., … Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *American Journal of Psychiatry*. https://doi.org/10.1176/ajp.2006.163.11.1905

Rush, A. J., Warden, D., Wisniewski, R. S., Fava, M., Trivedi, H. M., Gaynes, N. B., & Nierenberg, A. A. (2009). STAR*D: Revising conventional wisdom. *CNS Drugs*. https://doi.org/10.2165/00023210-200923080-00001

Saveanu, R., Etkin, A., Duchemin, A. M., Goldstein-Piekarski, A., Gyurak, A., Debattista, C., … Williams, L. M. (2015). The International Study to Predict Optimized Treatment in Depression (iSPOT-D): Outcomes from the acute phase of antidepressant treatment. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2014.12.018

Schildkraut, J. J. (1965). The catecholamine hypothesis of affective disorders: a review of supporting evidence. *The American Journal of Psychiatry*. https://doi.org/10.1176/ajp.122.5.509

Schmidt, F. M., Sander, C., Dietz, M. E., Nowak, C., Schröder, T., Mergl, R., … Hegerl, U. (2017). Brain arousal regulation as response predictor for antidepressant therapy in major depression. *Scientific Reports*. https://doi.org/10.1038/srep45187

Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., … Beniczky, S. (2017). The standardized EEG electrode array of the IFCN. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2017.06.254

Seifritz, E. (2001). Contribution of sleep physiology to depressive pathophysiology. *Neuropsychopharmacology*. https://doi.org/10.1016/S0893-133X(01)00319-0

Sen, S., Duman, R., & Sanacora, G. (2008). Serum Brain-Derived Neurotrophic Factor, Depression, and Antidepressant Medications: Meta-Analyses and Implications. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2008.05.005

Shimizu, E., Hashimoto, K., Okamura, N., Koike, K., Komatsu, N., Kumakiri, C., … Iyo, M. (2003). Alterations of serum levels of brain-derived neurotrophic factor (BDNF) in depressed patients with or without antidepressants. *Biological Psychiatry*. https://doi.org/10.1016/S0006-3223(03)00181-1

Slotema, C. W., Blom, J. D., Hoek, H. W., & Sommer, I. E. C. (2010). Should we expand the toolbox of psychiatric treatment methods to include repetitive transcranial magnetic stimulation (rTMS)? A meta-analysis of the efficacy of rTMS in psychiatric disorders. *Journal of Clinical Psychiatry*. https://doi.org/10.4088/JCP.08m04872gre

Smith, S. J. M. (2005). EEG in the diagnosis, classification, and management of patients with epilepsy. *Neurology in Practice*. https://doi.org/10.1136/jnnp.2005.069245

Spronk, D., Arns, M., Barnett, K. J., Cooper, N. J., & Gordon, E. (2011). An investigation of EEG, genetic and cognitive markers of treatment response to antidepressant medication in patients with major depressive disorder: A pilot study. *Journal of Affective Disorders*, *128*(1–2), 41–48. https://doi.org/10.1016/j.jad.2010.06.021

Stein, D. J. (2002). Obsessive-compulsive disorder. *Lancet*. https://doi.org/10.1016/S0140-6736(02)09620-4

Stetler, C., & Miller, G. E. (2011). Depression and hypothalamic-pituitary-adrenal activation: A quantitative summary of four decades of research. *Psychosomatic Medicine*. https://doi.org/10.1097/PSY.0b013e31820ad12b

Svanborg, P., & Åsberg, M. (2001). A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Åsberg Depression Rating Scale (MADRS). *Journal of Affective Disorders*. https://doi.org/10.1016/S0165-0327(00)00242-1

Tafet, G. E., & Bernardini, R. (2003). Psychoneuroendocrinological links between chronic stress and depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. https://doi.org/10.1016/S0278-5846(03)00162-3

Tandle, A., & Jog, N. (2015). Classification of Artefacts in EEG Signal Recordings and Overview of Removing Techniques. *International Journal of Computer Applications*.

Tansey, K. E., Guipponi, M., Perroud, N., Bondolfi, G., Domenici, E., Evans, D., … Uher, R. (2012). Genetic predictors of response to serotonergic and noradrenergic antidepressants in major depressive disorder: a genome-wide analysis of individual-level data and a meta-analysis. *PLoS Medicine*, *9*(10). https://doi.org/10.1371/journal.pmed.1001326

Tao, Y., Zheng, N., Haifeng, S., Yongqin, K., Xin, C., Jingmin, C., … Huiying, L. (2020). The Role of BDNF on Neural Plasticity in Depression. *Frontiers in Cellular Neuroscience*, *14*, 82. https://doi.org/10.3389/fncel.2020.00082

Tenke, C. E., Kayser, J., Manna, C. G., Fekri, S., Kroppmann, C. J., Schaller, J. D., … Bruder, G. E. (2011). Current source density measures of electroencephalographic alpha predict antidepressant treatment response. *Biological Psychiatry*, *70*(4), 388–394. https://doi.org/10.1016/j.biopsych.2011.02.016

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*.

Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V., Kurian, B. T., Phillips, M. L., … Weissman, M. M. (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *Journal of Psychiatric Research*, *78*(2016), 11–23. https://doi.org/10.1016/j.jpsychires.2016.03.001

Ulrich, G., Renfordt, E., Zeller, G., & Frick, K. (1984). Interrelation between changes in the EEG and psychopathology under pharmacotherapy for endogenous depression. A contribution to the predictor question. *Pharmacopsychiatry*. https://doi.org/10.1055/s-2007-1017433

Uyulan, C., Ergüzel, T. T., Unubol, H., Cebi, M., Sayar, G. H., Asad, M. N., & Tarhan, N. (2020). Major Depressive Disorder Classification Based on Different Convolutional Neural Network Models: Deep Learning Approach. *Clinical EEG and Neuroscience*, 1–14. https://doi.org/https://doi.org/10.1177/1550059420916634

Van Beijsterveldt, C. E. M., & Van Baal, G. C. M. (2002). Twin and family studies of the human electroencephalogram: A review and a meta-analysis. *Biological Psychology*. https://doi.org/10.1016/S0301-0511(02)00055-8

Van Den Broek, S. P., Reinders, F., Donderwinkel, M., & Peters, M. J. (1998). Volume conduction effects in EEG and MEG. *Electroencephalography and Clinical Neurophysiology*. https://doi.org/10.1016/S0013-4694(97)00147-8

Volf, N. V., Belousova, L. V., Knyazev, G. G., & Kulikov, A. V. (2015). Gender differences in association between serotonin transporter gene polymorphism and resting-state EEG activity. *Neuroscience*, *284*, 513–521. https://doi.org/10.1016/j.neuroscience.2014.10.030

Wade, E. C., & Iosifescu, D. V. (2016). Using Electroencephalography for Treatment Guidance in Major Depressive Disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(5), 411–422. https://doi.org/10.1016/j.bpsc.2016.06.002

Wahlund, B., & von Rosen, D. (2003). ECT of major depressed patients in relation to biological and clinical variables: A brief overview. *Neuropsychopharmacology*. https://doi.org/10.1038/sj.npp.1300135

Walpurger, V., Pietrowsky, R., Kirschbaum, C., & Wolf, O. T. (2004). Effects of the menstrual cycle on auditory event-related potentials. *Hormones and Behavior*. https://doi.org/10.1016/j.yhbeh.2004.07.002

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*. https://doi.org/10.1152/physrev.00035.2008

Weissman, M. M., Klerman, G. L., Prusoff, B. A., Sholomskas, D., & Padian, N. (1981). Depressed Outpatients: Results One Year after Treatment with Drugs and/or Interpersonal Psychotherapy. *Archives of General Psychiatry*. https://doi.org/10.1001/archpsyc.1981.01780260053005

Widge, A. S., Bilge, M. T., Montana, R., Chang, W., Rodriguez, C. I., Deckersbach, T., … Nemeroff, C. B. (2018). Electroencephalographic biomarkers for treatment response prediction in major depressive illness: A meta-analysis. *American Journal of Psychiatry*, *176*(1), 44–56. https://doi.org/10.1176/appi.ajp.2018.17121358

Williams, L. M., Simms, E., Clark, C. R., Paul, R. H., Rowe, D., & Gordon, E. (2005). The test-retest reliability of a standardized neurocognitive and neurophysiological test battery: "Neuromarker." *International Journal of Neuroscience*, *115*(12), 1605–1630. https://doi.org/10.1080/00207450590958475

Williams, Leanne M., Rush, A. J., Koslow, S. H., Wisniewski, S. R., Cooper, N. J., Nemeroff, C. B., … Gordon, E. (2011). International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: Rationale and protocol. *Trials*. https://doi.org/10.1186/1745-6215-12-4

World Health Organisation. (2017). Depression and other common mental disorders: global health estimates. *World Health Organization*.

Wu, W., Zhang, Y., Jiang, J., Lucas, M. V., Fonzo, G. A., Rolle, C. E., … Etkin, A. (2020). An electroencephalographic signature predicts antidepressant response in major depression. *Nature Biotechnology*. https://doi.org/10.1038/s41587-019-0397-3

Wyss, C., Hitz, K., Hengartner, M. P., Theodoridou, A., Obermann, C., Uhl, I., … Kawohl, W. (2013). The Loudness Dependence of Auditory Evoked Potentials (LDAEP) as an Indicator of Serotonergic Dysfunction in Patients with Predominant Schizophrenic Negative Symptoms. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0068650

Zeller, V., Ramachander, G., & Zeller, E. A. (1965). Amine Oxidases. XXI. A Rapid Method for the Determination of the Activity of Monoamine Oxidase and Monoamine Oxidase Inhibitors. *Journal of Medicinal Chemistry*. https://doi.org/10.1021/jm00328a006

Appendix

International Journal of Psychophysiology 134 (2018) 30–43

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Pre-intervention test-retest reliability of EEG and ERP over four recording intervals

Cheng-Teng Ip[a,b], Melanie Ganz[b,c], Brice Ozenne[b,d], Lasse B. Sluth[a], Mikkel Gram[e,f], Geoffrey Viardot[g], Philippe l'Hostis[g], Philippe Danjou[g], Gitte M. Knudsen[b], Søren R. Christensen[a,]

[a] Department of Clinical Pharmacology, H. Lundbeck A/S, Ottiliavej 9, DK-2500 Valby, Denmark
[b] Neurobiology Research Unit, University Hospital of Copenhagen, Rigshospitalet, N9201, 9 Blegdamsvej, DK-2100 Copenhagen, Denmark
[c] Department of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen, Denmark
[d] Section of Biostatistics, Department of Public Health, University of Copenhagen, Østerfarimagsgade 5, 1014 Copenhagen, Denmark
[e] Mech-Sense, Department of Gastroenterology and Hepatology, Aalborg University Hospital, Hobrovej 18-22, DK-9000, Denmark
[f] Coaze IVS, Kronhjorten 129, DK-9530 Støvring, Denmark
[g] Biotrial, 35000 Cedex, Rennes, France

ARTICLE INFO

Keywords:
Test-retest reliability
EEG
ERP
Repeated measurements
Crossover design

ABSTRACT

In this study we present the test-retest reliability of pre-intervention EEG/ERP (electroencephalogram/event-related potentials) data across four recording intervals separated by a washout period (18–22 days). POz-re-cording-reference EEG/ERP (28 sites, average reference) were recorded from thirty-two healthy male partici-pants. Participants were randomly allocated into different intervention sequences, each with four intervention regimens: 10 mg vortioxetine, 20 mg vortioxetine, 15 mg escitalopram and Placebo. We report classical EEG spectra: δ (1–4 Hz), θ (4–8 Hz), α (8–12 Hz), β (12–30 Hz), γ1 (30–45 Hz) and γ2 (45–80 Hz) of resting state and vigilance-controlled, and of auditory steady state response, as well as ERP components N100, P200 and P300 in auditory oddball task and error related negativity (ERN) and error positivity (Pe) in hybrid flanker task. Reliability was quantified using intra-class correlation coefficient (ICC). We found that θ, α and β of continuous EEG were highly reliable (ICCs ≥ 0.84). Evoked power of other tasks demonstrated larger variability and less reliability compared to the absolute power of continuous EEG. Furthermore, reliabilities of ERP measures were lower compared to those of the EEG spectra. We saw fair to excellent reliability of the amplitude of the com-ponents such as Pe (0.60–0.82) and P300 (0.55–0.80). Moreover, blood tests confirmed that there was no measurable drug carry-over from the previous intervention. The results support that EEG/ERP is reliable across four recording intervals, thus it can be used to assess the effect of different doses and types of drugs with CNS effects.

1. Introduction

Electroencephalography (EEG) provides a noninvasive method to measure electrical activity of the brain with high temporal resolution. The technique has shown great potential in clinical practice to monitor and access the intervention effects in diagnoses such as depression (Mulert et al., 2007; Tenke et al., 2011), Alzheimer (Brassen and Adler, 2003; Yener et al., 2007) and attention-deficit/hyperactivity disorder (ADHD) (Loo et al., 2000).

With the increased use of EEG and ERP in clinical practice, a sys-tematic investigation of EEG and ERP reliability becomes more im-portant, especially for commonly-used paradigms (e.g. resting state EEG and an auditory oddball task). Previous studies have investigated the reliability of EEG and ERP in various paradigms including resting state EEG with eyes-closed (Corsi-Cabrera et al., 2007) and eyes-opened (Williams et al., 2005), ERP components in an auditory oddball task (Williams et al., 2005), a working memory task (McEvoy et al., 2000) and a Sternberg task (Cassidy et al., 2012). These studies showed that a fair reliability of EEG and ERP could be obtained but that reliability could also be affected by various factors. For example, the reliability of EEG is affected by the epoch length of resting EEG (Gudmundsson et al., 2007), recording intervals (Sandman and Patterson, 2000), different reference schemes (Towers and Allen, 2009), and different aspects of the same EEG indicator (Tenke et al., 2018). In the study of Towers and

Corresponding author.
E-mail address: SRAC@Lundbeck.com (S.R. Christensen).
https://doi.org/10.1016/j.ijpsycho.2018.09.007
Received 9 February 2018; Received in revised form 20 September 2018; Accepted 21 September 2018
Available online 22 September 2018

Allen (2009), different reference schemes including online reference, re-referencing to linked-mastoids and average were compared for the reliability of frontal α asymmetry. Their results showed that linked-mastoids demonstrated greater reliability than other reference schemes, while other reference schemes still exhibited excellent split-half relia-bility (> 0.9). Different spectral parameters were compared, and showed that both absolute and relative power are reliable parameters (Fernandez et al., 1993). In a recent study, researchers assessed the temporal stability of different aspects of posterior EEG α over twelve years (Tenke et al., 2018). They suggested that lower reliability of net α (eyes closed-plus-open) and α asymmetry might result from additive errors when separating the α estimates. For ERP studies, there is ac-cumulating evidence showing that ERP amplitudes have higher relia-bility than ERP peak-latency measures (Cassidy et al., 2012; Walhovd and Fjell, 2002; Weinberg and Hajcak, 2011), which might be a result of the considerable variations in peak-latency detection. These varia-tions could be due to individual differences in information processing efficiency or induced by the appearance time of the peak amplitude, thus lowering the test-retest reliability. Since the replication of results is not always guaranteed within the field, it is essential to assess the re-liability of EEG and ERP measurements.

Among all the factors that could affect reliability, the number of recording sessions bring the biggest challenge to clinical application as it is almost impossible to maintain consistency between or within subjects. So far, a number of studies have investigated the reliability of EEG and ERP over both shorter (days: (McEvoy et al., 2000); weeks: (Cassidy et al., 2012; Hämmerer et al., 2013; Huffmeijer et al., 2014)) and longer recording intervals (months: (Brunner et al., 2013; Näpflin et al., 2007); years: (Sandman and Patterson, 2000; Tenke et al., 2018)). Sandman and Patterson (2000), evaluated ERP reliability in the para-digm of a dual rare-event over a three-year period and found that ERP measurements of adjacent years (e.g. Year 1 & 2) are more similar than ERP measurements of nonadjacent years (Year 1 & 3). Meanwhile, the test-retest reliability of resting EEG was not affected by the recording intervals (Corsi-Cabrera et al., 2007). One might argue that this in-consistency could be a result of different lengths of time (3 years vs. 9 months) during which the results were evaluated. Another possibility could be that different quantifications were investigated, i.e. EEG vs ERP. It could be possible that measures of EEG are more reliable than ERP measures, thus manifesting higher reliability over time. In the study of Williams et al. (2005), they reported high to excellent relia-bility for EEG power while only fair to excellent reliability for ERP measures. Furthermore, it is unclear how EEG and ERP vary across multiple recording intervals since only a few studies have reported the reliability across more than two sessions (Corsi-Cabrera et al., 2007; Kinoshita et al., 1996; Sandman and Patterson, 2000). In order to ad-dress this issue, the current study included four-time points to assess the reliability of both EEG and ERP measures.

In addition to recording intervals, the age of the participants is also known to contribute to the variations in ERP reliability (Alperin et al., 2014). Older adults show higher reliability of the P3 amplitude at the fronto-central site (Cz) while young adults have higher reliability at the centro-parietal area site (Pz) (Walhovd and Fjell, 2002). Hämmerer et al. (2013) suggested that age differences might be a result of different people's signal-to-noise ratio (SNRs), with children and older adults having lower SNRs than other age groups. Despite these variations, ERP measures still exhibit moderate to high reliability when evaluated with varying recording intervals and when participants of different age groups are selected (Hämmerer et al., 2013; Walhovd and Fjell, 2002). Therefore, age was used as a covariate throughout all our analyses.

Besides the signal itself, methodological differences in the statistical analysis have also led to discrepancies in test-retest reliability in EEG/ ERP studies. Different statistical methods have been adopted by studies that investigated the correlations between different recording sessions, and the test-retest reliability within sessions, such as ICC (Gudmundsson et al., 2007), Pearson's r (Walhovd and Fjell, 2002) and Spearman-Brown-corrected coefficients (Cassidy et al., 2012; Hämmerer et al., 2013; Walhovd and Fjell, 2002). Furthermore, there exist various types of Intra-class correlation coefficients (ICC) (Mcgraw and Wong, 1996) and previous studies have investigated the test-retest reliability by using different ICC measures. For instance, researchers have used a one-way random model of ICC (Gudmundsson et al., 2007), a two-way mixed model with absolute agreement (Brunner et al., 2013; Hämmerer et al., 2013) and a two-way mixed model with consistency (Rentzsch et al., 2008). When assessing EEG reliability between ses-sions, we define reliability as having both accuracy (i.e. no systematic bias) and precision (i.e. small variance caused by subject variability). We will therefore favor the ICC for absolute agreement over correlation coefficients or ICC for consistency, since the latter two only measure precision and will overestimate the reliability in presence of systematic biases.

Since the present study aims at assessing the reliability of EEG/ERP parameters, it can serve as a reference for investigating intervention effects. Therefore, we included spontaneous EEG, auditory steady state response, auditory oddball and hybrid flanker Go/Nogo tasks which are common measures in human cognition and executive function. In the present study, we incorporated the baseline data from four different sessions of an intervention study into one model. The carry-over drug effect from the previous session was evaluated through blood tests. The interventions included two different dosing levels of vortioxetine, one dosing level of escitalopram and placebo. The reliability of baseline data across different doses and types of antidepressants was evaluated through a linear mixed model with unstructured covariance matrix and was quantified by absolute agreement ICC. We hypothesized that: 1. The ICC of EEG and ERP measures will show at least moderate test-retest reliability across four recording intervals. 2. The power spectrum of continuous EEG will exhibit higher test-retest reliability than peak-picking ERP measures. 3. Amplitude measures will have higher test-retest reliability compared to peak latency measures.

2. Method

The study was conducted at the clinical site of Biotrial, Rennes, France. The research protocol was approved by the local ethics com-mittee (reference No. 15835A).

2.1. Participants

Participants were recruited in this study through advertisements and were screened by a trained investigator. To minimize the varia-bility, women were excluded to eliminate the menstrual cycle as a covariate. Thirty-two healthy male participants were enrolled in the study and were compensated for participation. Enrolled participants were aged 22 to 45 years (mean age 33.1 ± 6.8), their body mass index (BMI) ranged from 19.5 to 27.9 kg/m^2 (mean BMI 23.9 kg/m^2 ± 2.24), 94% of participants were Caucasian and 6% were African American. Exclusion criteria included use of psychoactive medication, drug or alcohol abuse, severe drug allergy or hypersensitivity and history of any medical, psychiatric, and neurological (such as immunological, cardi-ovascular, respiratory, metabolic neurological, or psychiatric) disease. Informed consent was obtained from all the participants before the study. All participants conducted the experiment except for one parti-cipant who has missing baseline data for three tasks (auditory steady state response (ASSR), auditory oddball and hybrid flanker task) in the 3rd session. All the collected data were included and analyzed.

2.2. Experimental protocol

This was an interventional, randomized, double-blind, placebo-controlled and four-way crossover study. The four included intervention regimens were: 10 mg vortioxetine (A), 20 mg vortioxetine (B), 15 mg escitalopram (C) and Placebo (D). Each participant was

randomly allocated into one sequence group (ABDC, BCAD, CDBA or DACB) with 8 participants in each group and was investigated under all intervention regimens separated by a washout period (20–22 days[1], median of all between sessions were 21 days) (Fig. 1). Bioanalysis was conducted before the administration of the next intervention to assess the leftover effects from the previous intervention. Within each session, an EEG battery was recorded on Day −1 (pre-intervention), Day 1 (the 1st day after intervention) and Day 3 (the 3rd day after intervention). The EEG battery included continuous EEG with resting and with vigi-lance-controlled, ASSR, auditory oddball and hybrid flanker tasks. Since the main purpose of this study was to assess the test-retest reliability, only the EEG recording of the four pre-interventions was considered in the subsequent analysis.

2.3. EEG battery

A previous study of antidepressants on rodents has shown a dis-sociation marker on different treatments, especially on the γ band (Leiser et al., 2014). Moreover, ERP components like P300 and ERN provide physiological measures associated with attentional engagement (Olbrich and Arns, 2013) and early error processing (Olvet and Hajcak, 2009a). The initiative of this study is whether the similar findings could be replicated in humans, as well as how antidepressants would affect human cognition and executive function. Therefore, we included spontaneous EEG, auditory steady state response, auditory oddball and hybrid flanker Go/Nogo tasks.

2.3.1. Continuous EEG

Continuous EEG data were acquired under two conditions: resting and vigilance-controlled. Participants were instructed to relax, keep their eyes closed and stay awake in both conditions. They were in-structed to keep pressing two buttons using their thumbs of each hand under the vigilance-controlled condition. A sound would play if the participant let go of the button. Each condition was recorded at least 3 min.

2.3.2. Auditory steady state response (ASSR)

Participants were presented with a 40 Hz impulse trains sound at 89 dB binaurally through a headset (Sennheiser HD 25-1 II pro) (McFadden et al., 2014; Van Deursen et al., 2011). Each train was composed of 20 biphasic 1 ms clicks, and each click was followed by silences lasting 24 ms. There was a silent period of 700 ms after each train. These trains were repeated for 5 min.

2.3.3. Auditory oddball

The auditory oddball paradigm consisted of two acoustic stimuli with different frequencies. Participants were presented with a series of standard tones (500 Hz) and deviant tones (2000 Hz) binaurally through a headset (Sennheiser HD 25-1 II pro). They were asked to count the deviant sounds. To make sure participants performed the task, the presentations of deviant and standard tones were different in sessions. Each session consisted of on average of 35 deviants (rando-mized between 30 and 40) and 198 standards (randomized between 170 and 226). Deviant tones made up 15% of the presentations. The sound level for each tone was 85 dB, with duration of 100 ms and inter-stimulus-interval (ISI) of on average 1550 ms (randomized between 1200 and 1900 ms). The test lasted approximately 7 min.

2.3.4. Hybrid flanker go/Nogo

Participants performed a hybrid flanker Go/Nogo paradigm (Ruchsow et al., 2006, 2005) with a monitor approximately 100 cm from them. Stimuli consisted of one of the following letter strings (BBBBB, DDDDD, VVVVV, UUUUU, BBDBB, DDBDD, UUVUU, or VVUVV) and were presented on a computer screen for 300 ms in ran-domized order. Participants were required to focus on the center letter and to press a button whether it was a B or a U (Go condition), and to withhold a button press upon appearance of a D or V (NoGo condition). Each condition consisted of 420 trials. There were 840 trials overall. Strings with congruent letters made up 40% of presentations, while strings with different letters were shown in 60% of all trials. Each trial was followed by 750 ms for stimulus onset asynchrony (SOA) and 500 ms for feedback in response to the participants' performance: 'true' (i.e. correct and in time), 'faster' (i.e. correct but out of time) or 'false'. The deadline for response time was 300 ms after stimulus onset. The ISI was 800 ms (randomized between 600 and 1000 ms). Test duration was approximately 45 min.

2.4. Electrophysiological recording

All participants were seated on a comfortable armchair in a quiet room. During data acquisition, participants were instructed to keep their eyes closed during continuous EEG, auditory oddball, and ASSR recordings. Participants conducted hybrid flanker task with open eyes and were told to refrain from eyes blinking and movement. EEG was recorded from 28 scalp sites using a 10–20 electrode system, with a sample rate of 400 Hz (Comet EEG system, Grass Technologies, West Warwick, RI, USA). AFz served as the ground and POz served as the reference electrode. In order to remove ocular and muscle artifacts electrooculography (EOG) and electromyogram (EMG) were recorded at bipolar channels. Impedances across all electrodes were maintained at < 5 kΩ.

2.5. Preprocessing of all data

Eye-blink and other ocular corrections were conducted for all the collected data by the ocular artifact reduction option of NeuroScan 4.1 software. It computes a linear regression of covariance between EEG and EOG, and then performs a point-by-point proportional subtraction of the blinks (Semlitsch et al., 1986). The data were further processed in Matlab 2012a (The Matworks, Inc., Natick, MA, USA).

2.5.1. Data preprocessing for spectral analysis

A zero-phase digital IIR Butterworth bandpass filter was applied to all data. The cut-off frequencies of the filter were 1 and 80 Hz, with an order of 2. In addition, a 50 Hz notch filter with the order of 6 was applied. All data (including continuous EEG, ASSR, auditory oddball and hybrid flanker tasks) were re-referenced to the average electrode for later time-frequency analysis. Continuous EEG was cleaned by cut-ting sections of noisy EEG from the signal by manual inspection.

2.5.2. Data preprocessing for ERP analysis

A zero-phase digital IIR Butterworth bandpass filter was applied to auditory oddball and hybrid flanker tasks. The cut-off frequencies of the filter were 0.1 and 30 Hz, with an order of 2. ERP data were re-refer-enced to the averages of linked mastoid electrodes (Segalowitz et al., 2010; Weinberg and Hajcak, 2011; Williams et al., 2005).

2.6. Data analysis

2.6.1. Time-frequency analysis of all data

Since EEG data have non-stationary characteristic, all data were analyzed using a wavelet transform as this has a better time-frequency resolution than the more common Fourier transform (Akin, 2002). The continuous wavelet transform was applied using the complex Morlet wavelet as a mother wavelet function with a bandwidth of 10 Hz and a center frequency of 1 Hz. The scales for the mother wavelet were chosen to match frequencies ranging from 1 to 80 Hz with a 0.5 Hz

[1] There was one outlier (91 days) in the last washout period due to recording cancellation. This recording was rescheduled after all participants were re-corded.

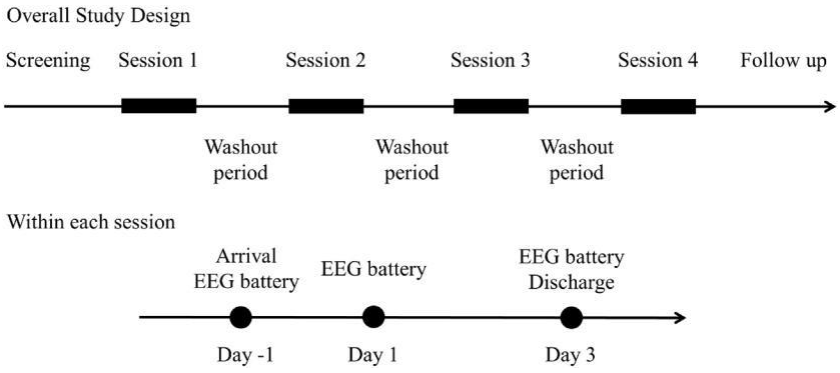


Fig. 1. Overall study design. Three interventions (A, B, C) and one placebo (D) were included in the study. Each participant was randomly allocated to one session sequence including ABDC, BCAD, CDBA and DACB. There were washout periods (median intervals were 21 days) between two sessions, and pharmacokinetic assessments were conducted to assess the carry-over drug effect from the previous intervention. Three EEG recordings were conducted within each session, in-cluding Day −1, Day 1 and Day 3. In this study, only the data from Day −1 was analyzed.

between-scale frequency interval. In the current study we worked on absolute power only, thus, the absolute values of the obtained wavelet coefficients were used for the following analysis: First, the wavelet coefficients were divided into the following standardized bands: δ (1–4 Hz), θ (4–8 Hz), α (8–12 Hz), β (12–30 Hz), γ1 (30–45 Hz) and γ2 (45–80 Hz). Then, the γ band was divided into two bands to deal with artifacts from muscle activity. Next, the wavelet coefficients were averaged over time and summed within each frequency band.

We applied different approaches for the continuous EEG and all the other tasks. The wavelet transform was applied on the noise-free con-tinuous EEG data without segmentation, including resting state and vigilance-controlled. All other tasks were segmented prior to time-fre-quency analysis and then evoked power was calculated for each task. ASSR and auditory oddball were segmented into stimulus-locked epochs of 500 ms according to the onset of the stimulus. For the audi-tory oddball task, evoked power was calculated for standards and de-viants separately. The hybrid flanker task was segmented from 0 to 400 ms according to the onset of error response.

The spectral analysis focused on three midline sites (Fz, Cz, Pz), therefore, the values represent the absolute values contained in each frequency band at these channels. Data were log-transformed prior to statistical analysis.

2.6.2. Grand average analysis of ERP data

In the auditory oddball task, EEG data were segmented into sti-mulus-locked epochs of 1000 ms (including a 200 ms pre-stimulus baseline) according to the onset of the sounds. Averaging was per-formed for standards and deviants separately. Epochs were rejected if the voltage in EOG channels, Fp1, Fp2 exceeded ± 75 μV. Based on prior studies investigating auditory oddball key components (Kemp et al., 2010; Poyraz et al., 2017), both peak latency and amplitude (baseline to peak) were determined on midline channels (Fz, Cz, Pz). The selected components and the corresponding latency windows for peak identification included: standard: N100 (80–140 ms), P200 (140–270 ms); deviant: N100 (80–140 ms) and P300 (270–550 ms). All epochs were manually inspected for other artifacts. A similar approach was applied to the hybrid flanker task. The main interest of the hybrid flanker task was the false positive response (Ruchsow et al., 2006, 2005), thus only responses of error commission were reported. EEG data were then segmented into response-locked epochs of 600 ms (in-cluding 200 ms pre-response baseline) according to the onset of error response. ERN (0–250) was analyzed at sites in the fronto-central area (Fz, Cz) and Pe (100–350) was analyzed at sites in the centro-parietal area (Cz, Pz) (Falkenstein et al., 2000). The number of accepted epochs is shown in Table 1.

2.7. Blood sampling

The blood samples (2 mL for each regimen) were analyzed for the plasma concentrations of vortioxetine and escitalopram. Plasma con-centrations were determined by using protein precipitation followed by liquid chromatography with tandem mass spectrometric detection. The purpose of these assessments was to ensure that previous intervention was completely washed out so that it would not interfere with the current intervention administration.

2.8. Statistics

The statistics were divided into two parts and performed in SPSS version 24 (IBM Corp., Armonk, NY). First, all EEG and ERP measures were analyzed with a linear mixed model (restricted maximum like-lihood estimation) using an unstructured covariance matrix with as-signed sequence (ABDC, BCAD, CDBA or DACB) and pre-intervention recordings of each session (BL1, BL2, BL3, BL4) as fixed factors. This was done in order to investigate if there was an effect of session or assigned sequence on our measurements. Participant served as a random variable to account for the correlation between measurements from the same patient. An unstructured covariance matrix was em-ployed to make minimal assumption on the covariance structure - meaning we relax the assumption of homogeneity of variance by modeling a different variance at each session and allow the correlation to vary between pairs of sessions. The structure of the covariance matrix used in the mixed models was decided upon inspection of the model fit. Using likelihood ratio tests, we found a significantly worse fit for the compound symmetry structure (i.e. assuming constant variance over time and constant correlation between any two timepoints) compared to a compound symmetry structure for some of the power measures of resting EEG and some of the ERP measures of flanker hybrid task. Therefore, an unstructured covariance matrix was employed. In all mixed models, age was included as covariate. Main effects of session and sequence were tested using F-tests. In post hoc analyses, regression coefficients of the different levels of the main effects were compared using Wald tests with Tukey contrasts. This was performed using the module EM Means for Linear Mixed Model in SPSS. Neither the p-values from the F-tests nor the post hoc analyses were adjusted for multiple comparisons in order to not reduce power. In this fashion we are maximizing our chance to detect any session or sequence effect despite detecting possible false positives. Second we assessed the reliability of our measurement using the intra-class correlation (ICC) with absolute agreement (Brunner et al., 2013; Hämmerer et al., 2013). Single mea-sure ICC (A, 1) was calculated by a two-way mixed random model (Mcgraw and Wong, 1996), where participant served as random vari-able and session served as fixed variable. ICC of adjacent time points, BL1 & BL2, BL2 & BL3 and BL3 & BL4 are reported. In accordance with the classification of ICC levels in a previous study (Rentzsch et al., 2008), ICC < 0.39 would be considered poor, 0.4–0.59 fair, 0.6–0.75 good and > 0.75 would be considered excellent. Overall, time var-iances are reported in the supplement and were computed by the structure of compound symmetry. To provide a synthetic measure of the ICC over time, we computed “average ICCs” using a mixed model with a compound symmetry covariance matrix instead of an unstructured covariance matrix. This enables us to provide a graphical representation

Table 1
The number of accepted epochs for different tasks.

Task	Condition	BL1	BL2	BL3	BL4	p values
Auditory oddball	Standard	180 ± 23(117–212)[a]	169 ± 37(101–227)	167 ± 27(97–215)	174 ± 34(80–227)	F (3,92) = 1.884, p = .138
	Deviant	31 ± 5(22–38)	30 ± 7(16–40)	29 ± 6(12–38)	30 ± 7(15–40)	F (3,92) = 1.270, p = .289
Hybrid Flanker	Error	85 ± 35(28–180)	70 ± 28(8–133)	67 ± 36(8–183)	72 ± 31(3–141)	F (3,92) = 3.981, p = .01

Notes. [a] The minimum and maximum of epochs are provided in the brackets. The mean and standard deviation are reported.

of the ICC as a function of the percentage of accepted trials of across time (Fig. 8).

3. Results

Blood tests were performed to assess the carry-over drug effect of previous interventions. The blood concentration of the previous treat-ment, C_{max} for all participants across sessions was below 5%, which was considered as complete washout.

3.1. Behavioral results

In the hybrid flanker Nogo trials, participants demonstrated a mean false positive alarm rate of 21% (SD: 7.8) for BL1, 17% (± 7.3) for BL2, 16% (± 9.1) for BL3 and 18% (± 7.7) for BL4. A linear mixed model revealed that there were no significant effects for session and assigned sequence in error rate (p values > .05). Considering the mean reaction time, participants demonstrated a mean false positive reaction time of 283 ms (± 18) for BL1, 282 ms (± 22) for BL2, 276 ms (± 17) for BL3 and 274 ms (± 22) for BL4. There were no significant effects of session and assigned sequence in Nogo reaction time (p values > .05).

3.2. Absolute power of resting EEG

Since there was no segmentation for continuous EEG, spectra were used for presentation instead of time-frequency plots (Fig. 2). There were no significant effects of session and assigned sequence in resting condition for all frequency bands (p values > .05). In the vigilance-controlled task, γ1 at the central site exhibited a significant main effect of session (F (3, 31) = 3.41, p = .029). Post hoc analyses revealed that absolute γ1 power at the first recording session BL1 was larger than the last session BL4 (17.83 vs 16.39 μV, p = .006). No other significant effect was found.

3.3. Evoked power of ASSR, auditory oddball and hybrid flanker task

Fig. 3 shows the absolute evoked power for ASSR, auditory oddball and hybrid flanker tasks for all four recording sessions. Compared to the absolute power of continuous EEG, evoked power demonstrated more variations between sessions. Specifically, the absolute evoked power at the first recording (BL1) contributed the most to the significance.

For the ASSR task, no sequence effect was found for all frequency bands. Significant main effects of session were found for δ and γ1 at the frontal site (F (3, 31) = 3.919, p = .018; F (3, 31) = 3.567, p = .025, Fig. 3a). Post hoc analyses revealed that a smaller absolute δ power was observed at BL1 compared to BL3 and BL4 (1.35 vs. 1.51 μV, p = .02; 1.35 vs. 1.58 μV, p = .004), and larger γ1 was observed at BL1 com-pared to BL3 and BL4 (1.91 vs. 1.79 μV, p = .02; 1.91 vs. 1.84 μV, p = .03). Similarly, δ at the parietal site indicated a significant session effect (F (3, 31) = 3.179, p = .038), showing that the absolute δ power at BL1 was smaller than BL4 (0.61 vs. 0.83 μV, p = .014). Moreover, θ and α at the central site exhibited significant session effects (F (3,
31) = 4.352, p = .011; F (3, 31) = 3.409, p = .03). The absolute θ power of BL1 was the smallest compared to other recording sessions (p values < .05) and the absolute α of BL1 and BL2 were smaller than that of BL4 (1.13 vs. 1.30 μV, p = .03; 1.08 vs. 1.30 μV, p = .013).

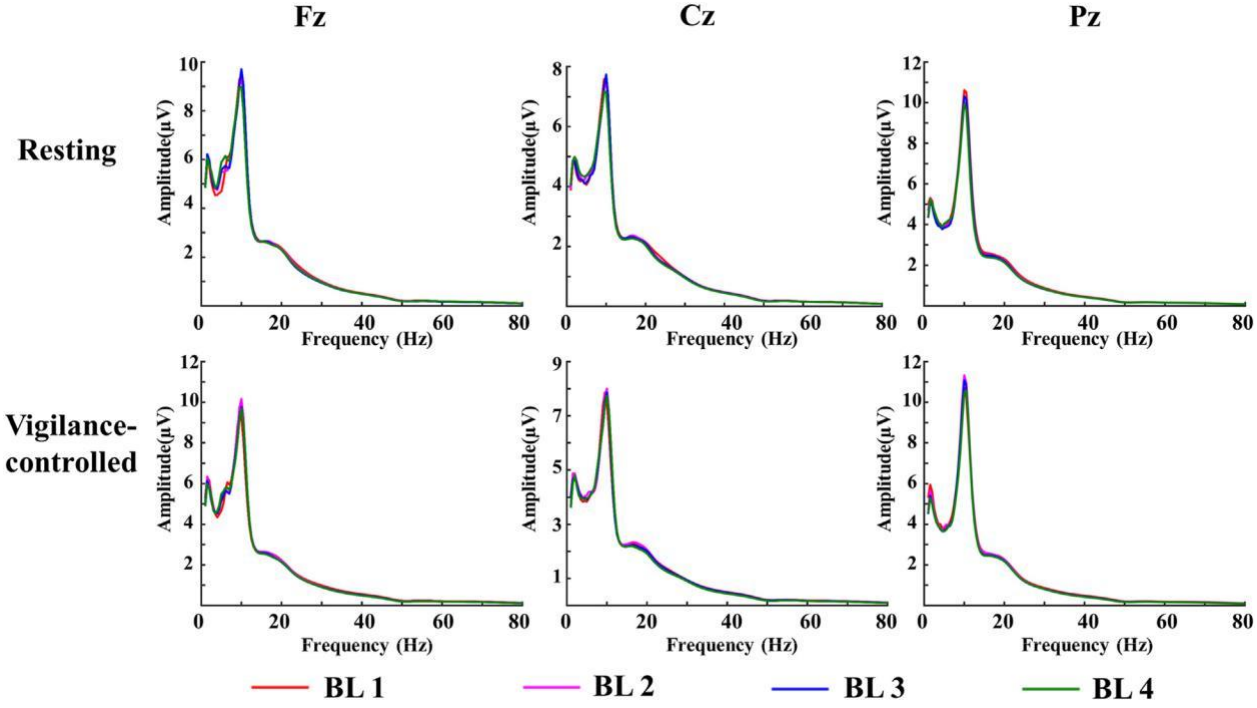


Fig. 2. Spectral results for continuous EEG including conditions of resting state and vigilance-controlled. Three midline electrodes (Fz, Cz and Pz) are shown for each condition. Four recording sessions (BL1, BL2, BL3 and BL4) are shown in different colors.

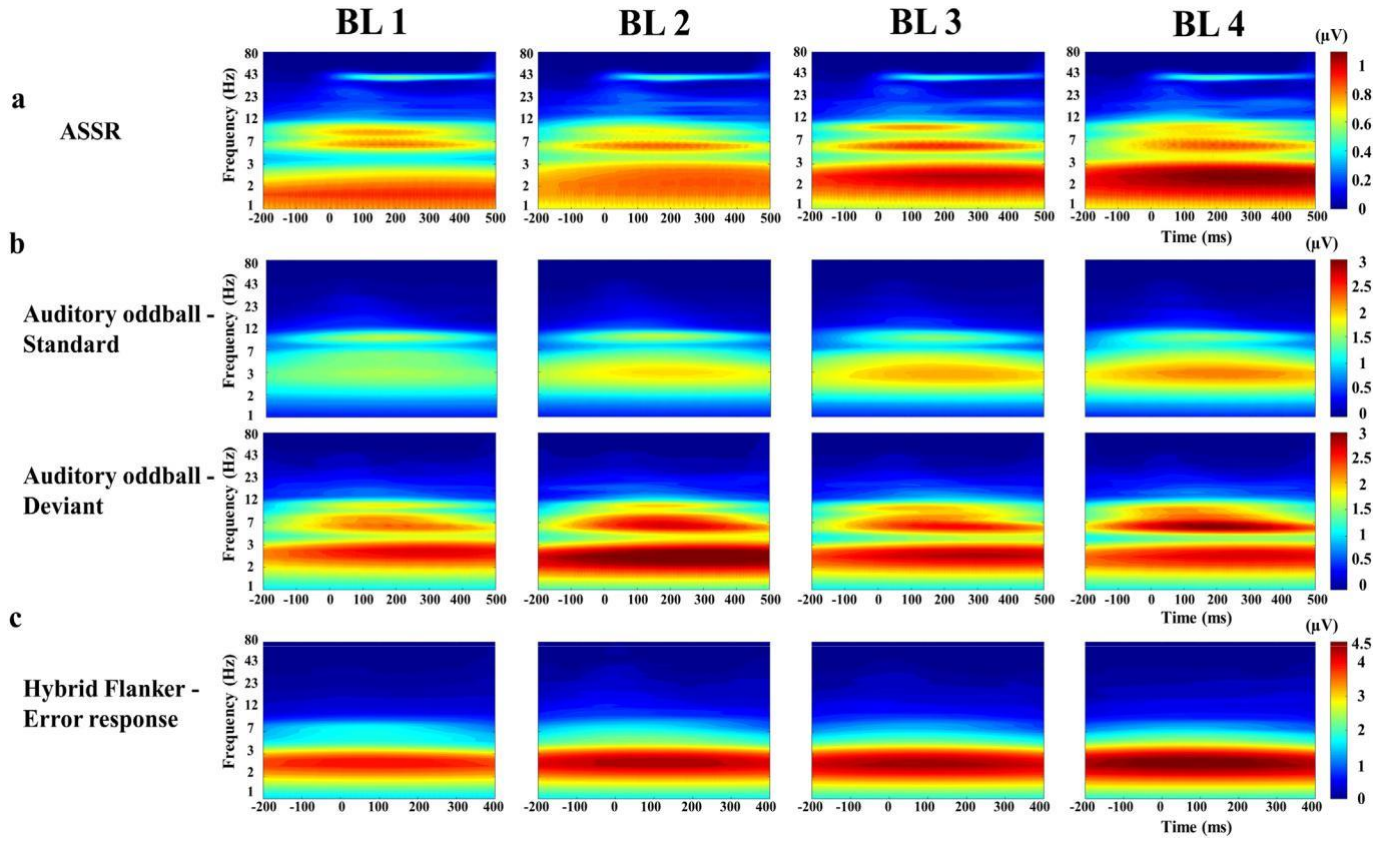


Fig. 3. Time-frequency results for ASSR, auditory oddball and hybrid flanker tasks. Only results at electrode Fz are shown here since most of the significant results were found on this electrode. Four recording sessions (BL1, BL2, BL3 and BL4) are shown in columns. Log-scale is shown for the frequency range.

There was no significant effect of session for β and γ2 bands (p va-lues > .05).

For the standard tones in the auditory oddball task, significant session effects of the frontal and the parietal sites were observed in the δ band (F (3, 31) = 5.651, p = .003; F (3, 31) = 3.844, p = .02; Fig. 3b). Post hoc analyses revealed that the absolute frontal δ power of BL1 was the smallest (p values < .05), while absolute parietal δ power was the largest among all recording sessions (p values < .05). No other sig-nificant effect was found. For the deviant tones in the auditory oddball task, absolute frontal δ power showed a significant session effect (F (3, 31) = 3.111, p = .04), indicating that the absolute frontal δ of BL2 was larger than BL3 and BL4 (2.54 vs. 2.37 μV, p = .02; 2.54 vs. 2.34 μV, p = .011). Notably, absolute θ power of BL1 was significantly smaller than BL2 (2.54 vs. 2.71 μV, p = .02; 2.13 vs. 2.32 μV, p = .005), in-dicated by a significant main session effect at frontal and parietal sites (F (3, 31) = 3.327, p = .032; F (3, 31) = 3.185, p = .038). Moreover, absolute frontal θ power of BL1 was smaller than BL4 (2.54 vs. 2.73 μV, p = .02). No sequence effect was found (p values < .05).

There were no session effects in the bands of δ, α and γ1 for the error response of the hybrid flanker task. A significant session effect was observed for the absolute θ power at the central site (F (3, 31) = 3.52, p = .027), due to smaller absolute θ during the first two recording sessions than BL3 (Fig. 3c). Significant main effects of session were found for β and γ2 at fronto-central sites. Post hoc analyses indicated that absolute frontal β and γ2 powers of BL1 were the smallest (F (3,
31) = 3.679, p = .023; F (3, 31) = 3.219, p = .036) among other re-cording sessions. Absolute central β, γ1 and γ2 powers of BL1 were
smaller than BL3 (F (3, 31) = 3.297, p = .033; F (3, 31) = 4.804, p = .007; F (3, 31) = 3.640, p = .023). Moreover, absolute γ2 of BL1 at the central site was smaller than BL4. Sequence effects were observed in γ2 at the frontal site (F (3, 24) = 4.381, p = .014), due to the greatest power observed in the sequence of ABDC among others.

3.4. Amplitude and latency analysis of auditory oddball and hybrid flanker tasks

Table 1 shows the number of accepted epochs for both auditory oddball and hybrid flanker tasks. The results of a linear mixed model indicated that there was a significant session effect (F (3,92) = 3.981, p = .01) for the number of accepted epochs in the hybrid flanker task. BL1 demonstrated a significant higher number in accepted epochs than BL2 (p = .043), and BL3 (p = .014). No session effect was observed for the auditory oddball task (p values > .05).

Fig. 4 shows the mean ERP waveform for auditory oddball and hybrid flanker tasks for all four recording sessions. For the standard ERPs in the auditory oddball task, fronto-central N100 amplitude ex-hibited a significant session effect (F (3, 31) = 5.21, p = .005; F (3,
31) = 6.93, p = .001, Fig. 4). BL1 and BL2 showed larger fronto-central N100 amplitude than BL3 and BL4 (p values < .05). No session effect was found for fronto-central N100 latency. However, parietal N100
latency indicated a significant session effect (F (3, 31) = 3.36, p = .034), showing that BL1 had longer latency than all other recording sessions (p values ≤ .052). There was no session effect on the P200 amplitude. No assigned sequence effect was found for standard ERPs. For the deviant ERPs in the auditory oddball task, there were no sig-nificant effects of session and assigned sequence on the N100 ampli-tude. The central N100 latency was shortest for the last recording (BL4,
p values < .05), as suggested by a significant session effect (F (3, 31) = 3.26, p = .034). The fronto-central P300 amplitude seemed not to be affected by session or assigned sequence (Fig. 4). However, a significant session effect was also found for the parietal P300 latency (F (3, 31) = 4.13, p = .014), showing that a shorter P300 latency was observed at BL1 compared to BL4 (310 vs. 325 ms, p = .009).

For the error ERPs in the hybrid flanker task, there were no sig-nificant effects of session and assigned sequence for ERN and Pe

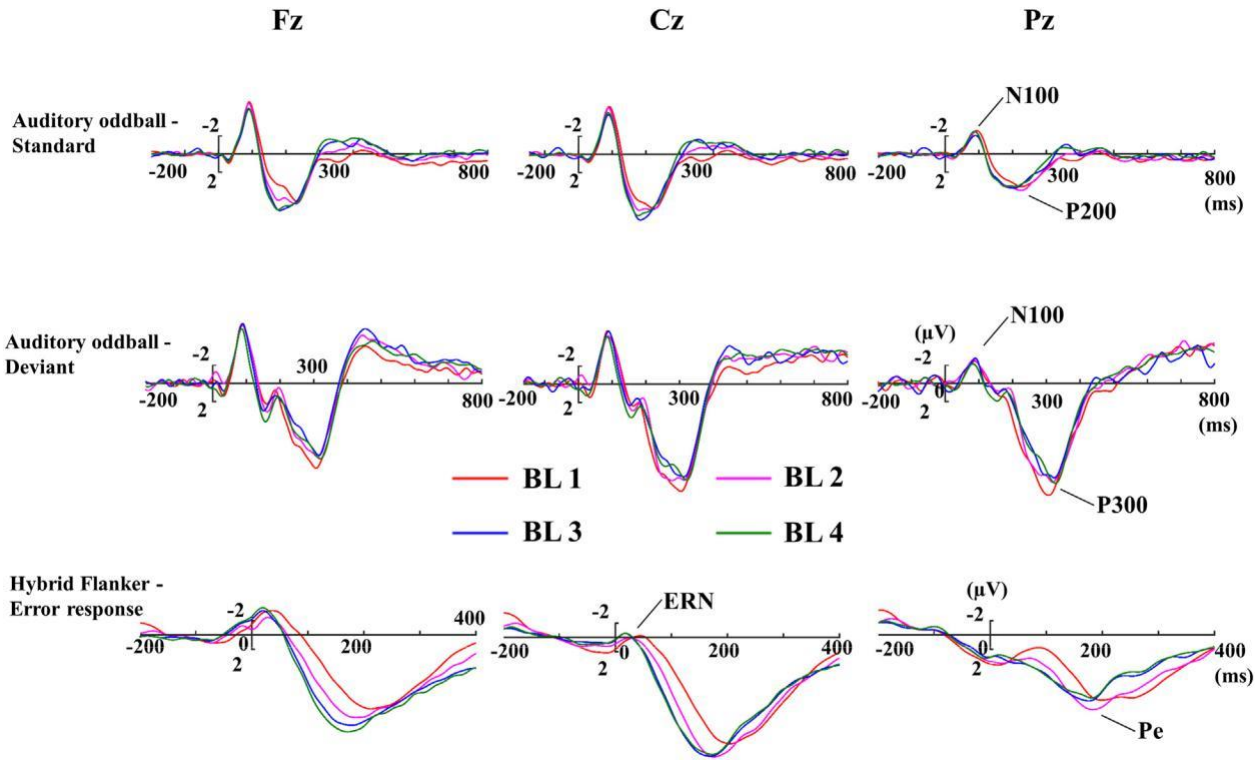


Fig. 4. The grand-averaged ERP waveforms for the auditory oddball (epoched by the stimuli) and hybrid flanker task (epoched by the error response). Three midline electrodes (Fz, Cz and Pz) are shown for each task/component. Four recording sessions (BL1, BL2, BL3 and BL4) are shown in different colors.

amplitudes (see Fig. 4). However, session effects were observed for the latency measures. The fronto-central ERN latency exhibited a sig-nificant session effect ($F(3, 31) = 3.78$, $p = .02$; $F(3, 31) = 6.91$, $p = .001$), showing that a longer ERN latency was observed at BL1 than at BL3 and BL4 (p values < .05). The centro-parietal Pe latency was longer at BL1 and BL2 than BL3 and BL4 (p values < .05), as indicated by a significant session effect ($F(3, 31) = 29.34$, $p < .001$; $F(3, 31) = 22.66$, $p < .001$).

3.5. Test-retest reliability

3.5.1. Absolute power of resting EEG

Between session ICCs for continuous EEG (both resting state and vigilance-controlled) are presented in Fig. 5. The test-retest reliabilities were similar in both conditions. The ICCs of adjacent sessions showed excellent test-retest reliability (0.84–0.97) in the frequency bands of θ, α and β. Midline δ and γ1 bands were less robust but still indicated good to excellent levels of reliability (0.62–0.87). ICCs for midline γ2 ex-hibited the least reliability among all other bands (0.30–0.66). Across time, ICC showed similar results to adjacent time points. Midline θ, α and β had excellent reliability (0.86–0.93) while δ and γ1 bands showed good to excellent reliability (0.66–0.82). Compared to adjacent time points, γ2 ICC across time performed worse with poor to fair levels of reliability (0.37–0.52).

3.5.2. Evoked power of ASSR, auditory oddball and hybrid flanker task

Between session ICCs for evoked power of ASSR, auditory oddball and hybrid flanker tasks are presented in Fig. 6. Across time ICC showed similar results to adjacent time points.

For the ASSR task, midline γ1 –which contains the stimulation fre-quency- exhibited good to excellent reliability for both adjacent ses-sions and across time (0.66–0.86), except for the fair ICC measured between the last two sessions at the frontal site (0.57). The ICCs of δ, β and γ1 were less robust but still indicative of fair to good levels of re-liability (0.44–0.76). Midline θ exhibited larger variations in different recording sessions, where the reliability varied from poor to excellent (0.37–0.83). The ICCs of the α band demonstrated poor to fair levels of reliability in the ASSR task (0.19–0.56).

For the standard tones of the auditory oddball task, midline β and γ1 revealed fair to excellent levels of reliability for both adjacent sessions and across time (0.44–0.85). The ICCs of the δ, θ and α bands exhibited larger variation between sessions compared to the β and γ1 bands, in the range of poor to excellent (0.29–0.84). Compared to other fre-quency bands, midline γ2 of standard tones showed less robust relia-bility with poor to good levels of ICC (0.36–0.62). In general, deviant tones were less robust compared to standard tones. The ICCs of δ were in the range of good to excellent (0.63–0.83). Midline θ had poor to excellent reliability (0.34–0.82) while the ICCs of other bands were in the range of poor to good (0.08–0.75).

For the error response of the hybrid flanker task, midline θ tended to exhibit the best reliability among other bands for both adjacent sessions and across time (0.50–0.85). The ICCs of α were fair to good (0.47–73) while the ICCs of δ showed more variability, in the range of poor to excellent (0.24–0.80). Midline β, γ1 and γ2 bands demonstrated similar reliability, in the range of poor to good levels of reliability (0.25–74).

3.5.3. Amplitude and latency analysis of ERP task

Between session ICCs for peak amplitude and latency measures of the auditory oddball and hybrid flanker tasks are presented in Fig. 7. Generally, amplitude and latency analysis of ERP showed lower relia-bility compared to the power spectrum analysis of ERP data. Further-more, latency measures were less stable than amplitude measures.

For standard ERPs in the auditory oddball task, the fronto-central N100 amplitude showed good to excellent reliability for adjacent ses-sions and across time (ICC > 0.70), and the parietal N100 amplitude demonstrated poor to fair levels of reliability (0.39–0.53). The P200 amplitude exhibited similar reliability, with good to excellent levels of reliability at fronto-central sites (0.65–0.83) and less stable perfor-mance at parietal site (0.49–0.68). Latency measures exhibited more variations from session to session. The ICCs of the N100 latency were in

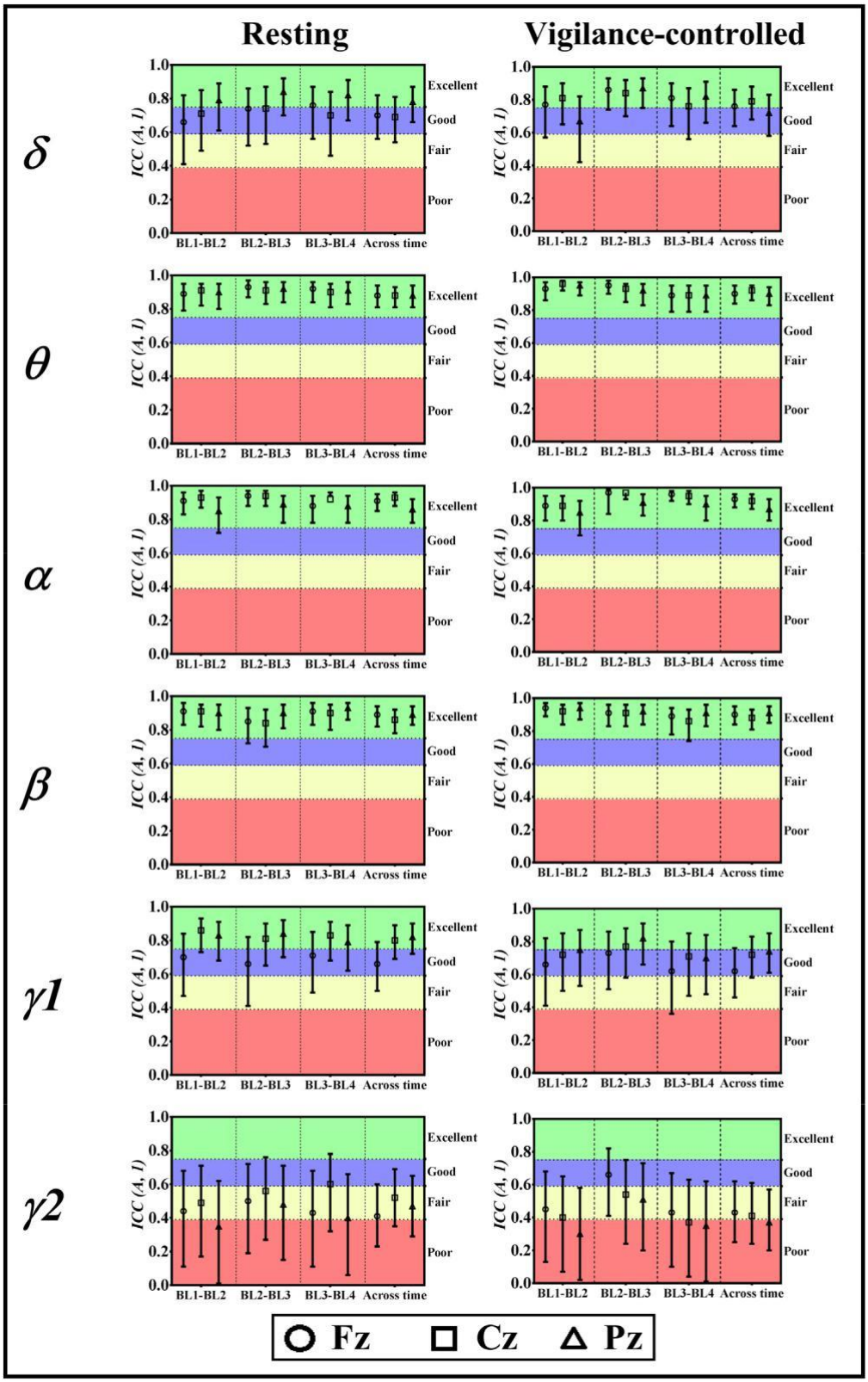


Fig. 5. Intra-class correlation coefficient (ICC) for continuous EEG across four sessions. ICCs of adjacent time and across four-time points are reported. Test-retest reliability is estimated by the single measure ICC (A, 1). The mean and confident intervals for ICCs are shown in the figure.

the range of fair to excellent (0.43–0.89), except for the ICC of first two sessions, which showed only poor reliability (0.28). The ICCs of the P200 latency were in the range of poor and fair (−0.48–0.49). Compared to standard tones, the fronto-central N100 amplitude of deviant tones exhibited lower reliability, with the ICC range of poor to fair (0.18–0.45). The parietal N100 amplitude showed poor reliability (−0.05–0.19). The midline P300 amplitude yielded fair to excellent reliability for adjacent sessions and across time (0.55–0.80). Compared

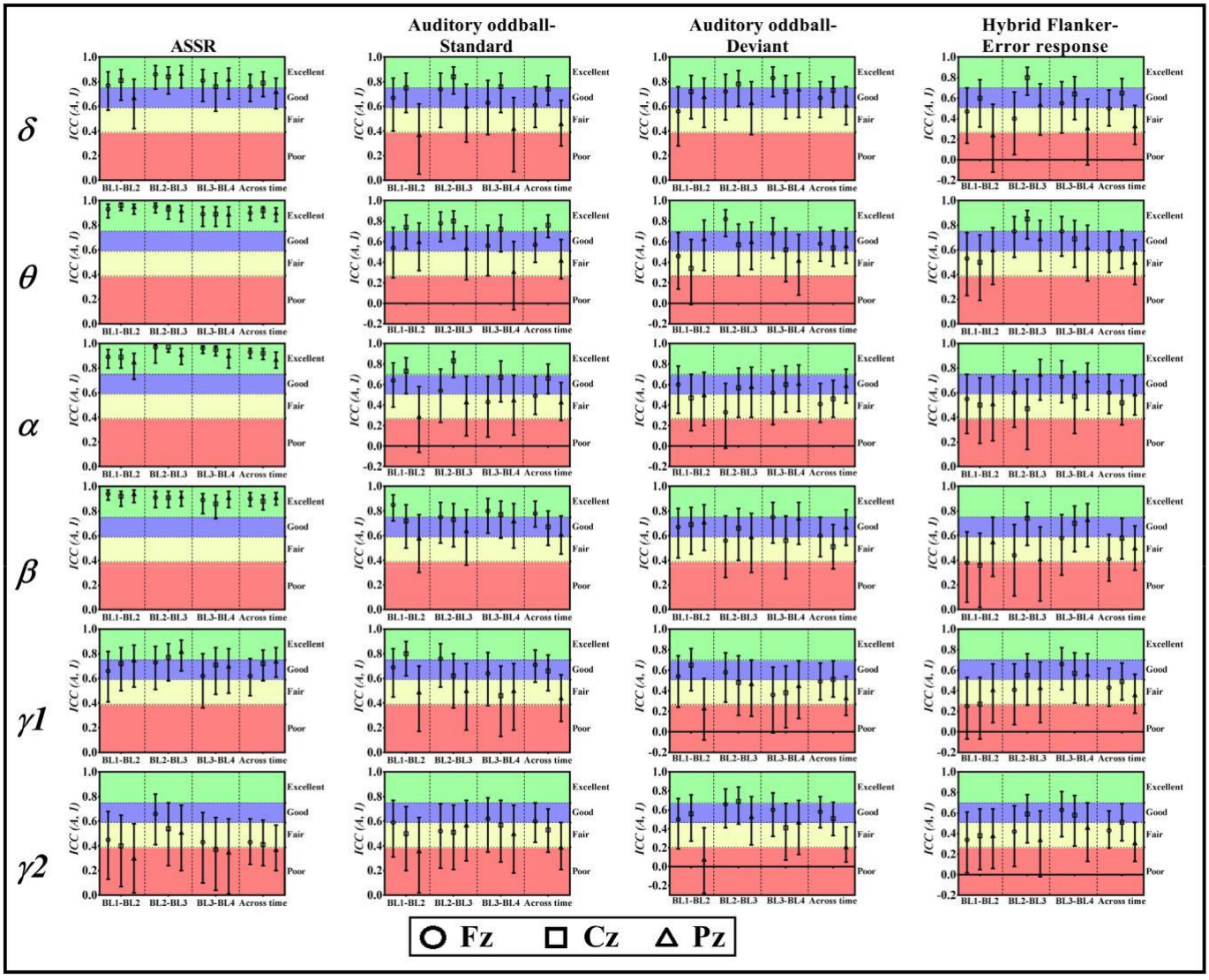


Fig. 6. Intra-class correlation coefficient (ICC) for ASSR, auditory oddball and hybrid flanker tasks across four sessions. ICCs of adjacent time and across four-time points are reported. Test-retest reliability is estimated by the single measure ICC (A, 1). The mean and confident intervals for ICCs are shown in the figure.

to the standard ERPs, the latency measures for deviant ERPs demon-strated less variations between sessions but were still indicative of poor to good levels of reliability (N100: −0.10–0.57; P300: 0.19–0.63).

For the error ERPs in the hybrid flanker task, the fronto-central ERN amplitude demonstrated poor to good levels of reliability for adjacent sessions (0.12–0.61) and poor reliability across time (0.32–0.38). The centro-parietal Pe tented to exhibited higher reliability compared to ERN, with the ICC ranging of good to excellent for both adjacent time points and across time (0.60–0.82). Latency measures showed less re-liability compared to amplitude measures. The ICCs of ERN latency were poor (0.12–0.35) while Pe latencies were poor to good (0.19–0.71).

3.5.4. Exploratory analysis: test-retest reliability with increasing percentage of accepted trials

Across time ICCs for the auditory oddball and hybrid flanker tasks for an increasing percentage of accepted trials are presented in Fig. 8 (see Supplementary materials for adjacent time points). Four percen-tages were assessed with 25% as an increment: 25%, 50%, 75% and 100%. Percentages were calculated relative to the total amount of ac-cepted trials individually. Then the corresponding number of trials would be successively selected from the total amount of accepted trials, i.e. the first 25% (or 50% and 75%) of the total accepted trials. The number of accepted epochs for different percentages is shown in Table 2. Mean amplitude, which was calculated using the same window as peak amplitude, was included here for comparison to peak ampli-tude.

As expected, reliability increased with increasing percentage of ac-cepted trials. Peak amplitude demonstrated comparable results with mean amplitude for all components. Latency measures were more susceptible to changes of percentage compared to peak amplitude and mean amplitude measures.

For the ERPs in the auditory oddball task, the ICCs increased with increasing percentage of accepted trials. Hence, it could be possible that increasing the number of accepted trials could increase the test-retest reliability. The grand average (100%) exhibited the highest reliability for almost all components, except for the N1 latency evoked by deviant tones, where the ICCs for grand average were lower than that of the first 75% of accepted trials.

For the error ERPs in the hybrid flanker task, the results for the ERN and Pe measures are similar. They were less affected by the increasing percentage of accepted trials. The ICCs increased slightly with in-creasing percentage of accepted trials up to the first 50%, but then remained at the same level of reliability as the grand average.

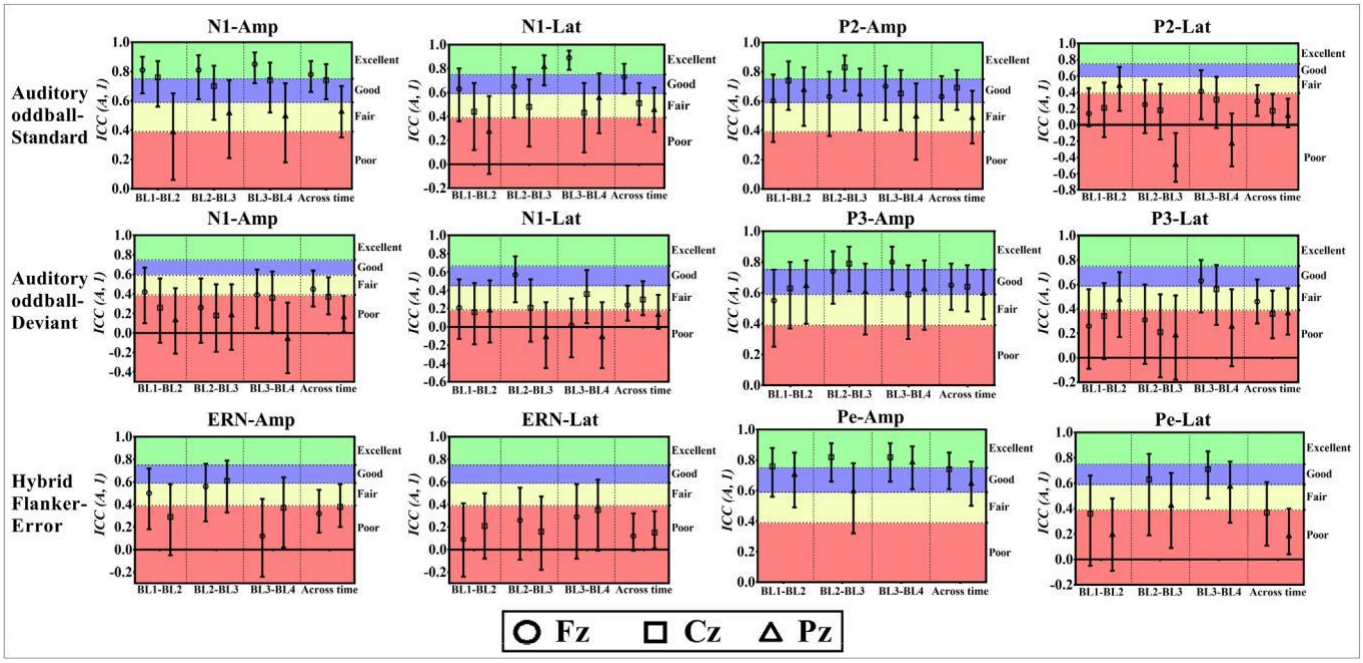


Fig. 7. Intra-class correlation coefficient (ICC) for peak amplitude and latency measures of auditory oddball and hybrid flanker tasks across four sessions. ICCs of adjacent time and across four-time points are reported. Test-retest reliability is estimated by the single measure ICC (A, 1). The mean and confident intervals for ICCs are shown in the figure.

4. Discussion

In this study, we examined the test-retest reliability of an EEG battery over four recording intervals. The EEG battery was comprised of continuous EEG (resting state and vigilance-controlled), ASSR as well as an auditory oddball paradigm and a hybrid flanker task. A linear mixed model with unstructured covariance matrix was used to identify any significant effect of recording session or the assigned intervention se-quence. The test-retest reliability was quantified by an absolute agreement type of ICC. For healthy participants, the results demon-strated that the EEG battery was found to be reliable over four sessions. The absolute power of continuous EEG showed excellent reliability in θ,
α and β (ICC > 0.84). Evoked power for ERP tasks demonstrated itself to be less stable compared to the absolute power of continuous EEG. The absolute evoked power of ASSR showed fair reliability in δ, β, γ1 and γ2 bands. For the auditory oddball task, the β band exhibited fair reliability (ICC > 0.51) in both standard and deviant conditions. The ICCs of θ in the hybrid flanker task were the most stable among all the frequency bands. While the ERP components showed lower reliability than the power spectral analysis, they still showed good test-retest re-liability at their maximal sites. The P300 amplitude obtained from the auditory oddball paradigm had consistently fair to excellent reliability at the central sites (ICC = 0.55–0.80) as well as the amplitude of the midline P2 (ICC = 0.49–0.83). The centro-parietal Pe amplitude ob-tained from the hybrid flanker task also exhibited good to excellent reliability (ICC = 0.60–0.82). Compared to amplitude measures, peak latency measures showed poor to good reliability with greater varia-bility, thus they are less reliable compared to other measures. A washout period and pharmacokinetic assessment were included to avoid a carry-over drug effect from the previous intervention.

4.1. The absolute power analysis of continuous EEG is highly reliable

The observed excellent test-retest reliability of continuous EEG in the power analysis is consistent with previous studies (Corsi-Cabrera et al., 2007; Gudmundsson et al., 2007; McEvoy et al., 2000; Williams et al., 2005). In the study of Gudmundsson et al. (2007), researchers compared different qEEG features such as power spectral parameters, entropy, complexity and coherence measures and suggested that power spectral analysis exhibits higher reliability than others types of analysis. This was confirmed by our results, and indicates that power spectral analysis of continuous EEG is reliable over time and is sufficient for clinical use.

4.2. The reliability of ERP measures was affected by various factors

Our results showed that ERP measures exhibited more variation and are less stable compared to continuous EEG. According to previous results, there are many factors that could cause the variability of ERP measures. They include the number of averaged trials (Larson et al., 2010) and the scoring methods (Brunner et al., 2013). Larson et al. (2010) investigated the influence of the number of averaged trials on error-related ERP components, and showed that adding trials increases the test-retest reliability for both the amplitude and latency measures. This was confirmed by our exploratory analysis. The results highlighted that increasing the percentage of accepted trials improved the test-retest reliability. For the ERPs in the auditory oddball task, the first 75% of accepted trials produced a comparable reliability to the grand average. Particularly for the latency measures, the increasing trend indicated that an increasing number of accepted trials improved the reliability. Except for the N1 latency of deviant tones, the ICCs of the grand average had lower reliabilities than the corresponding ICCs based on the first 75% of accepted trials. This result was mainly due to the high reliability of the first 75% of accepted trials at BL1-BL2 (Fig. S1). We surmised that fatigue/impatience due to the unfamiliarity of the task might be a possible reason. Participants could be tired or lose motivation during the last 25% of the accepted trials. Since this phe-nomenon didn't extend to the subsequent sessions, the result reiterated the importance of guiding the participants in a proper way so that they can be more comfortable with the experiments during the first re-cording session. For the error ERPs in the hybrid flanker task, the re-liability of different percentages was similar, especially for percentages above the first 25%. This finding supports previous findings by Olvet and Hajcak (2009b), in which they reported that stable ERN and Pe can

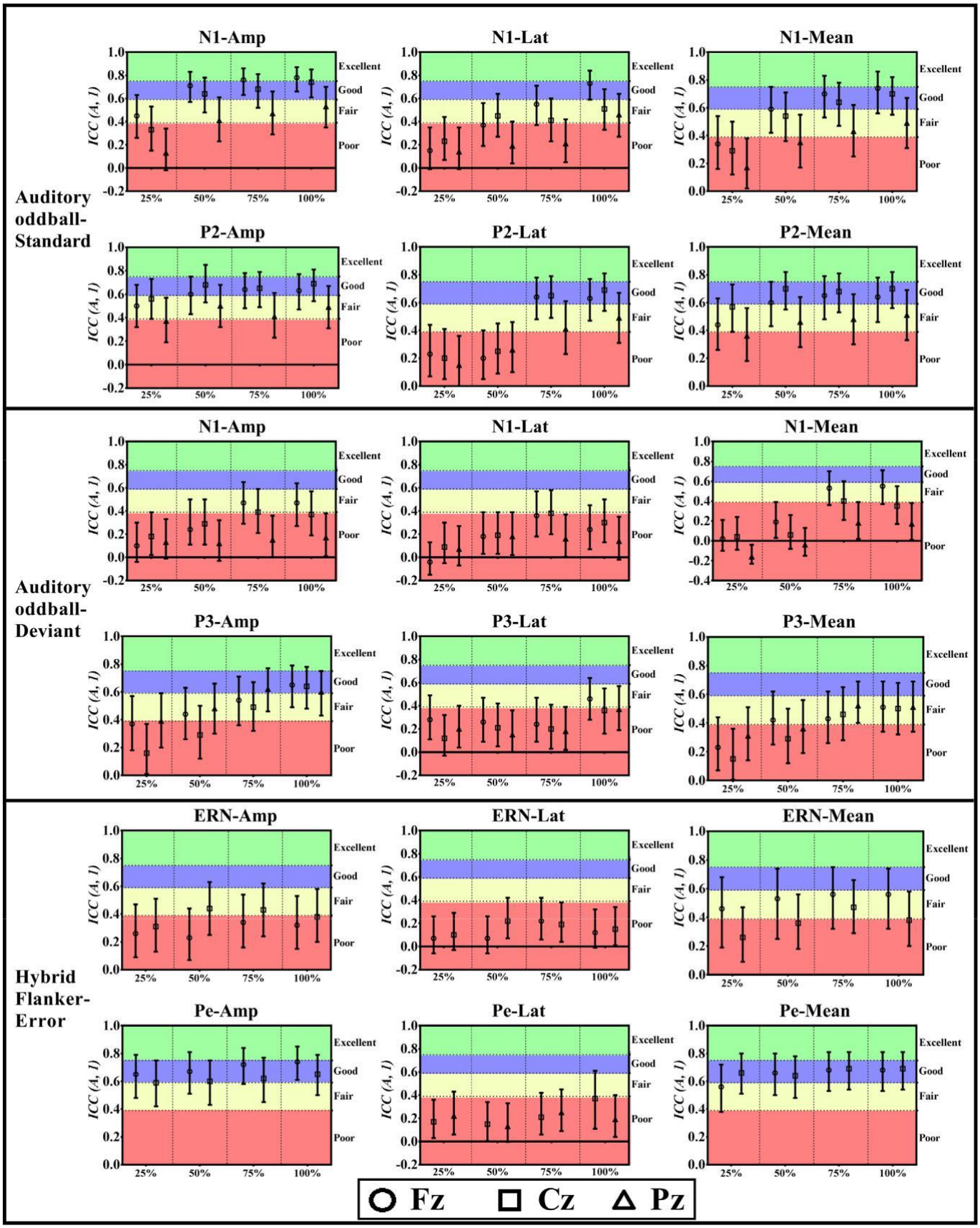


Fig. 8. Intra-class correlation coefficient (ICC) with increasing percentage of accepted trials of across time for ERP tasks (please refer to the Supplementary materials for adjacent time points). Four percentages of accepted epochs are assessed: 25%, 50%, 75% and 100%. Test-retest reliability is estimated by the single measure ICC (A, 1). The mean and confident intervals for ICCs are shown in the figure.

Table 2
The number of accepted epochs for different percentages.

Task	Percentage	Condition	BL1	BL2	BL3	BL4
Auditory oddball	25%	Standard	44 ± 7(30–53)[a]	39 ± 9(25–57)	40 ± 8(24–54)	41 ± 8(20–57)
		Deviant	8 ± 1(6–10)	7 ± 2(4–10)	7 ± 2(3−10)	7 ± 2(4–10)
	50%	Standard	87 ± 13(59–106)	78 ± 19(51–114)	79 ± 16(49–108)	81 ± 17(40–114)
		Deviant	15 ± 2(11–19)	14 ± 4(8–20)	14 ± 3(6–19)	14 ± 3(8–20)
	75%	Standard	130 ± 20(88–159)	116 ± 28(76–170)	119 ± 24(73–161)	122 ± 25(60–170)
		Deviant	23 ± 3(17–29)	21 ± 5(12−30)	21 ± 5(9–29)	21 ± 5(11−30)
	100%	Standard	180 ± 23(117–212)	169 ± 37(101–227)	167 ± 27(97–215)	174 ± 34(80–227)
		Deviant	31 ± 5(22–38)	30 ± 7(16–40)	29 ± 6(12–38)	30 ± 7(15–40)
Hybrid Flanker	25%	Error	22 ± 9(7–45)	18 ± 7(2–34)	17 ± 9(2–46)	18 ± 8(1–36)
	50%		43 ± 18(14–90)	35 ± 14(4–67)	34 ± 18(4–92)	36 ± 16(2–71)
	75%		64 ± 26(21–135)	53 ± 21(6–100)	51 ± 28(6–138)	54 ± 23(3–106)
	100%		85 ± 35(28–180)	70 ± 28(8–133)	67 ± 36(8–183)	72 ± 31(3–141)

Notes. [a] The minimum and maximum of epochs are provided in the brackets. The mean and standard deviation are reported.

be obtained with 6 and 2 error trials, correspondingly. Our data ex-tended the results by presenting similar reliability when the number of accepted trials was increased. Moreover, we found that the reliability of ERP measures is affected by the size of the components. Smaller-sized components such as N100 and ERN exhibit lower reliability relative to larger-sized components, such as P300 and Pe (Fig. 7). This discrepancy could be caused by the difference in SNRs existing in different sizes of ERP components (Luck, 2005). Increasing the number of averaged trials and a better control of artifacts could increase the SNR for ERP com-ponents, thereby leading to a higher reliability.

Furthermore, we found that amplitude measures are more stable than latency measures, which is consistent with previous findings (Cassidy et al., 2012). The scoring method could play an important role in causing the discrepancies between ERP parameters, in which am-plitude seems to be less susceptible to different scoring methods (Olvet and Hajcak, 2009a; Weinberg and Hajcak, 2011) than latency (Brunner et al., 2013). Brunner et al. (2013) compared the reliability of con-ventional peak measures and of the fractional area approach (FA) for the measures of independent component analysis (ICA). Their results suggested that the FA approach leads to an increase in the reliability of latency measures between two recording sessions, especially for the late components. On the other hand, Olvet and Hajcak (2009a) found si-milar reliabilities using both the area and peak measures, which in-dicate that the reliability of amplitude measurements is affected by different scoring methods to a lesser degree. In the present study, only peak-picking analysis was implemented, and thus it was difficult to capture the best method for the reliability of latency measures. Further investigation is needed to improve the reliability of latency measures in general.

4.3. Statistical methodology

To date, EEG/ERP reliability studies have mainly been conducted over two recording sessions, while more than two sessions are involved in most pharmacological studies. It is important to evaluate how EEG changes across longer periods of time and across multiple sessions. The reason for choosing a linear mixed model for the present analysis is that it can evaluate different recording sessions through an unstructured covariance matrix, an approach which is assumption-free on the cov-ariance matrix, given the fact that we cannot be sure whether EEG/ERP parameters decay, increase or remain stable between the different re-cordings. This approach is adequate for assessing multiple recordings. We used the ICC coefficients to quantify the reliability between sub-sequent recording sessions instead of the Pearson correlation coefficient since the latter is not a proper measure of reliability (see Chapter 1, (Lin et al., 2012)).

4.4. Limitations

There are some considerations that need to be taken into account before an interpretation and further generalization of our results can be made. First, to reduce the variability of our data, we excluded women in our recruited population. Even though the effect of menstrual cycle was not the main interest in the current study, it could result in lower re-liabilities based on previous findings (O'Reilly et al., 2004; Walpurger et al., 2004). Moreover, Bazanova et al. (2017) demonstrated how the α amplitude suppression could change in different phases of the men-strual cycle, but the effects of different phases and the relation between phases and the reliability of EEG (or ERP) remain unclear. Hence, the presented results should be carefully interpreted since menstrual cycle could influence the test-retest reliability. Although there was previous evidence showing that the test-retest reliability is highly comparable for both genders (Tenke et al., 2018), future studies must address whether test-retest reliability changes across genders, e.g. with the menstrual phases.

In addition, the mixed model showed a significant effect of assigned sequence. We believe that this might be caused by the spurious age differences within the different sequence groups ($F(3, 31) = 4.057$, $p < .05$), since unfortunately age was not taken into account when the participants were randomized. This could cause low EEG/ERP relia-bility since age does have an impact on EEG/ERP (Hämmerer et al., 2013). However, we cannot be certain due to the relatively small sample size of the present study (eight participants per intervention sequence).

Another related issue is that of the low number of accepted trials of error ERPs at BL4 which was observed for one participant (Table 1, the minimum is 3). This participant contributed the lowest number of ac-cepted trials (the second lowest is 12) in all sessions. This was the case due to low committed errors instead of a noisy signal (i.e. non-phy-siological signal). We didn't exclude the participant for two main rea-sons: first, this study is a clinical trial and thus it is important to report the actual data. Second, in the study of Olvet and Hajcak (2009b), they demonstrated that stable error ERPs could be measured with a minimum of six error trials, even two trials for the component Pe. Olvet and Hajcak (2009a) extended the results by comparing the reliability of high versus low number of error trials, and demonstrated similar test-retest reliability between groups. Therefore, it could be possible that the reported reliability for EPRs is underestimated but we don't believe the exclusion of this single participant's data would improve the reliability significantly.

Finally, carry-over drug effects might have existed in our dataset even though blood tests confirmed that there was a complete washout. This is because while during later baseline measurements, blood tests can eliminate the presence of previous interventions in the bloodstream (BL2, BL3, BL4), it cannot completely rule out the indirect influence a previous treatment had on a participant's subsequent test performance.

5. Conclusions

In conclusion, we find excellent test-retest reliability for power spectra measures of continuous EEG in the frequency bands of θ, α and β and an acceptable reliability for the evoked power of ERP tasks. Amplitude measures are more consistent across sessions relative to la-tency measures, and the amplitudes of P300 and Pe are more reliable than the amplitudes of ERN and Pe. Our results support that these EEG/ ERP parameters are reliable across three-week intervals and thus are sufficiently reliable for future investigations examining pharmacolo-gical effects.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpsycho.2018.09.007.

Declaration of interest

All authors declare no conflict of interest.

Acknowledgement

Collection of the data included in the study was supported by H. Lundbeck A/S. C.I. was supported by the Innovationsfonden, Denmark (5189-00087A). M.G. was supported by the Lundbeck Foundation (R181-2014-3586). B.O. was supported by the Lundbeck Foundation (R231-2016-3236) and Marie-Curie-NEUROMODEL (746850).

References

Akin, M., 2002. Comparison of wavelet transform and FFT methods in the analysis of EEG signals. J. Med. Syst. 26, 241–247. https://doi.org/10.1023/A:1015075101937. Alperin, B.R., Mott, K.K., Rentz, D.M., Holcomb, P.J., Daffner, K.R., 2014. Investigating the age-related "anterior shift" in the scalp distribution of the P3b component using principal component analysis. 51, 620–633. https://doi.org/10.1111/psyp.12206.

Bazanova, O.M., Nikolenko, E.D., Barry, R.J., 2017. Reactivity of alpha rhythms to eyes opening (the Berger effect) during menstrual cycle phases. Int. J. Psychophysiol. 122, 56–64. https://doi.org/10.1016/j.ijpsycho.2017.05.001.

Brassen, S., Adler, G., 2003. Short-term effects of acetylcholinesterase inhibitor treatment on EEG and memory performance in Alzheimer patients: an open, controlled trial. Pharmacopsychiatry 36, 304–308. https://doi.org/10.1055/s-2003-45118.

Brunner, J.F., Hansen, T.I., Olsen, A., Skandsen, T., Håberg, A., Kropotov, J., 2013. Long-term test-retest reliability of the P3 NoGo wave and two independent components decomposed from the P3 NoGo wave in a visual Go/NoGo task. Int. J. Psychophysiol. 89, 106–114. https://doi.org/10.1016/j.ijpsycho.2013.06.005.

Cassidy, S.M., Robertson, I.H., O'Connell, R.G., 2012. Retest reliability of event-related potentials: evidence from a variety of paradigms. Psychophysiology 49, 659–664. https://doi.org/10.1111/j.1469-8986.2011.01349.x.

Corsi-Cabrera, M., Calindo-Vilchis, L., Del-Río-Portilla, Y., Arce, C., Ramos-Loyo, J., 2007. Within-subject reliability and inter-session stability of EEG power and coherent ac-tivity in women evaluated monthly over nine months. Clin. Neurophysiol. 118, 9–21. https://doi.org/10.1016/j.clinph.2006.08.013.

Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J., 2000. ERP components on re-action errors and their functional significance: a tutorial. Biol. Psychol. 51, 87–107. https://doi.org/10.1016/S0301-0511(99)00031-9.

Fernandez, T., Harmony, T., Rodriguez, M., Reyes, A., Marosi, E., Bernal, J., 1993. Test-retest reliability of EEG spectral parameters during cognitive tasks: I. Absolute and relative power. Int. J. Neurosci. 68, 255–261. https://doi.org/10.3109/ 00207459308994280.

Gudmundsson, S., Runarsson, T.P., Sigurdsson, S., Eiriksdottir, G., Johnsen, K., 2007. Reliability of quantitative EEG features. Clin. Neurophysiol. 118, 2162–2171. https://doi.org/10.1016/j.clinph.2007.06.018.

Hämmerer, D., Li, S.-C., Völkle, M., Müller, V., Lindenberger, U., 2013. A lifespan com-parison of the reliability, test-retest stability, and signal-to-noise ratio of event-re-lated potentials assessed during performance monitoring. Psychophysiology 50, 111–123. https://doi.org/10.1111/j.1469-8986.2012.01476.x.

Huffmeijer, R., Bakermans-Kranenburg, M.J., Alink, L.R.A., Van IJzendoorn, M.H., 2014. Reliability of event-related potentials: the influence of number of trials and elec-trodes. Physiol. Behav. 130, 13–22. https://doi.org/10.1016/j.physbeh.2014.03.008.

Kemp, A.H., Pe Benito, L., Quintana, D.S., Clark, C.R., McFarlane, A., Mayur, P., Harris, A., Boyce, P., Williams, L.M., 2010. Impact of depression heterogeneity on attention: an auditory oddball event related potential study. J. Affect. Disord. 123, 202–207. https://doi.org/10.1016/j.jad.2009.08.010.

Kinoshita, S., Inoue, M., Maeda, H., Nakamura, J., Morita, K., 1996. Long-term patterns of change in ERPs across repeated measurements. Physiol. Behav. 60, 1087–1092. https://doi.org/10.1016/0031-9384(96)00130-8.

Larson, M.J., Baldwin, S.A., Good, D.A., Fair, J.E., 2010. Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): the role of number of trials. Psychophysiology 47, 1167–1171. https://doi.org/10.1111/j.1469-8986.2010.01022.x.

Leiser, S.C., Pehrson, A.L., Robichaud, P.J., Sanchez, C., 2014. Multimodal antidepressant vortioxetine increases frontal cortical oscillations unlike escitalopram and duloxetine - a quantitative EEG study in rats. Br. J. Pharmacol. 171, 4255–4272. https://doi.org/10.1111/bph.12782.

Lin, L., Hedayat, A.S., Wu, W., 2012. Statistical Tools for Measuring Agreement. Springer, New York. https://doi.org/10.1007/978-1-4614-0562-7.

Loo, S.K., Specter, E., Smolen, A., Hopfer, C., Teale, P.D., Reite, M.L., 2000. Functional effects of the DAT1 polymorphism on EEG measures in ADHD. J. Am. Acad. Child Adolesc. Psychiatry 42, 986–993. https://doi.org/10.1097/01.CHI.0000046890.27264.88.

Luck, S.J., 2005. Ten Simple Rules for Designing ERP Experiments. In: Handy, T.C. (Ed.), Event-Related Potentials: A Methods Handbook. MIT Press, Cambridge, MA. https://doi.org/10.1371/journal.pcbi.0020012.

McEvoy, L.K., Smith, M.K., Gevins, A., 2000. Test–retest reliability of cognitive EEG. Clin. Neurophysiol. 111, 457–463. https://doi.org/10.1016/S1388-2457(99)00258-8.

McFadden, K.L., Steinmetz, S.E., Carroll, A.M., Simon, S.T., Wallace, A., Rojas, D.C., 2014. Test-retest reliability of the 40 Hz EEG auditory steady-state response. PLoS One 9, 59–61. https://doi.org/10.1371/journal.pone.0085748.

Mcgraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1, 30–46. https://doi.org/10.1037/1082-989X.1.1.30.

Mulert, C., Juckel, G., Brunnmeier, M., Karch, S., Leicht, G., Mergl, R., Möller, H., Hegerl, U., Pogarell, O., 2007. Prediction of treatment response in major depression: in-tegration of concepts. J. Affect. Disord. 98, 215–225. https://doi.org/10.1016/j.jad. 2006.07.021.

Näpflin, M., Wildi, M., Sarnthein, J., 2007. Test-retest reliability of resting EEG spectra validates a statistical signature of persons. Clin. Neurophysiol. 118, 2519–2524. https://doi.org/10.1016/j.clinph.2007.07.022.

Olbrich, S., Arns, M., 2013. EEG biomarkers in major depressive disorder: discriminative power and prediction of treatment response. Int. Rev. Psychiatry 25, 604–618. https://doi.org/10.3109/09540261.2013.816269.

Olvet, D.M., Hajcak, G., 2009a. Reliability of error-related brain activity. Brain Res. 1284, 89–99. https://doi.org/10.1016/j.brainres.2009.05.079.

Olvet, D.M., Hajcak, G., 2009b. The stability of error-related brain activity with in-creasing trials. Psychophysiology 46, 957–961. https://doi.org/10.1111/j.1469-8986.2009.00848.x.

O'Reilly, M.A., Cunningham, C.J., Lawlor, B.A., Walsh, C.D., Rowan, M.J., 2004. The effect of the menstrual cycle on electrophysiological and behavioral measures of memory and mood. Psychophysiology 41, 592–603. https://doi.org/10.1111/j.1469-8986.2004.00194.x.

Poyraz, B.Ç., Sakallı Kani, A., Aksoy Poyraz, C., Öcek Baş, T., Arıkan, M.K., 2017. Cognitive psychophysiological substrates of affective temperaments: a P300 study. Clin. EEG Neurosci. 48, 96–102. https://doi.org/10.1177/1550059416650112.

Rentzsch, J., Jockers-Scherübl, M.C., Boutros, N.N., Gallinat, J., 2008. Test-retest relia-bility of P50, N100 and P200 auditory sensory gating in healthy subjects. Int. J. Psychophysiol. 67, 81–90. https://doi.org/10.1016/j.ijpsycho.2007.10.006.

Ruchsow, M., Spitzer, M., Grön, G., Grothe, J., Kiefer, M., 2005. Error processing and impulsiveness in normals: evidence from event-related potentials. Cogn. Brain Res. 24, 317–325. https://doi.org/10.1016/j.cogbrainres.2005.02.003.

Ruchsow, M., Walter, H., Buchheim, A., Martius, P., Spitzer, M., Kächele, H., Grön, G., Kiefer, M., 2006. Electrophysiological correlates of error processing in borderline personality disorder. Biol. Psychol. 72, 133–140. https://doi.org/10.1016/j. biopsycho.2005.08.006.

Sandman, C.A., Patterson, J.V., 2000. The auditory event-related potential is a stable and reliable measure in elderly subjects over a 3 year period. Clin. Neurophysiol. 111, 1427–1437. https://doi.org/10.1016/S1388-2457(00)00320-5.

Segalowitz, S.J., Santesso, D.L., Murphy, T.I., Homan, D., Chantziantoniou, D.K., Khan, S., 2010. Retest reliability of medial frontal negativities during performance monitoring. Psychophysiology 47, 260–270. https://doi.org/10.1111/j.1469-8986.2009.00942.x.

Semlitsch, H.V., Anderer, P., Schuster, P., Presslich, O., 1986. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. Psychophysiology 23, 695–703. https://doi.org/10.1111/j.1469-8986.1986.tb00696.x.

Tenke, C.E., Kayser, J., Manna, C.G., Fekri, S., Kroppmann, C.J., Schaller, J.D., Alschuler, D.M., Stewart, J.W., Mcgrath, P.J., Bruder, G.E., 2011. Current source density mea-sures of electroencephalographic alpha predict antidepressant treatment response. Biol. Psychiatry 70, 388–394. https://doi.org/10.1016/j.biopsych.2011.02.016. Tenke, C.E., Kayser, J., Alvarenga, J.E., Abraham, K.S., Warner, V., Talati, A., Weissman, M.M., Bruder, G.E., 2018. Temporal stability of posterior EEG alpha over twelve years. Clin. Neurophysiol. 129, 1410–1417. https://doi.org/10.1016/j.clinph.2018.03.037.

Towers, D.N., Allen, J.J.B., 2009. A better estimate of the internal consistency reliability of frontal EEG asymmetry scores. Psychophysiology 46, 132–142. https://doi.org/10. 1111/j.1469-8986.2008.00759.x.

Van Deursen, J.A., Vuurman, E.F.P.M., Van Kranen-Mastenbroek, V.H.J.M., Verhey, F.R.J., Riedel, W.J., 2011. 40-Hz steady state response in Alzheimer's disease and mild cognitive impairment. Neurobiol. Aging 32, 24–30. https://doi.org/10.1016/j.neurobiolaging.2009.01.002.

Walhovd, K.B., Fjell, A.M., 2002. One-year test-retest reliability of auditory ERPs in young and old adults. Int. J. Psychophysiol. 46, 29–40. https://doi.org/10.1016/S0167-8760(02)00039-9.

Walpurger, V., Pietrowsky, R., Kirschbaum, C., Wolf, O.T., 2004. Effects of the menstrual cycle on auditory event-related potentials. Horm. Behav. 46, 600–606. https://doi.org/10.1016/j.yhbeh.2004.07.002.

Weinberg, A., Hajcak, G., 2011. Longer term test-retest reliability of error-related brain activity. Psychophysiology 48, 1420–1425. https://doi.org/10.1111/j.1469-8986.

2011.01206.x.

Williams, L., Simms, E., Clark, C.R., Paul, R.H., Rowe, D., Gordon, E., 2005. The test-retest reliability of a standardized neurocognitive and neurophysiological test battery: "Neuromarker.". Int. J. Neurosci. 115, 1605–1630. https://doi.org/10.1080/00207450590958475.

Yener, G.G., Güntekin, B., Öniz, A., Başar, E., 2007. Increased frontal phase-locking of event-related theta oscillations in Alzheimer patients treated with cholinesterase inhibitors. Int. J. Psychophysiol. 64, 46–52. https://doi.org/10.1016/j.ijpsycho. 2006.07.006.

Supplementary

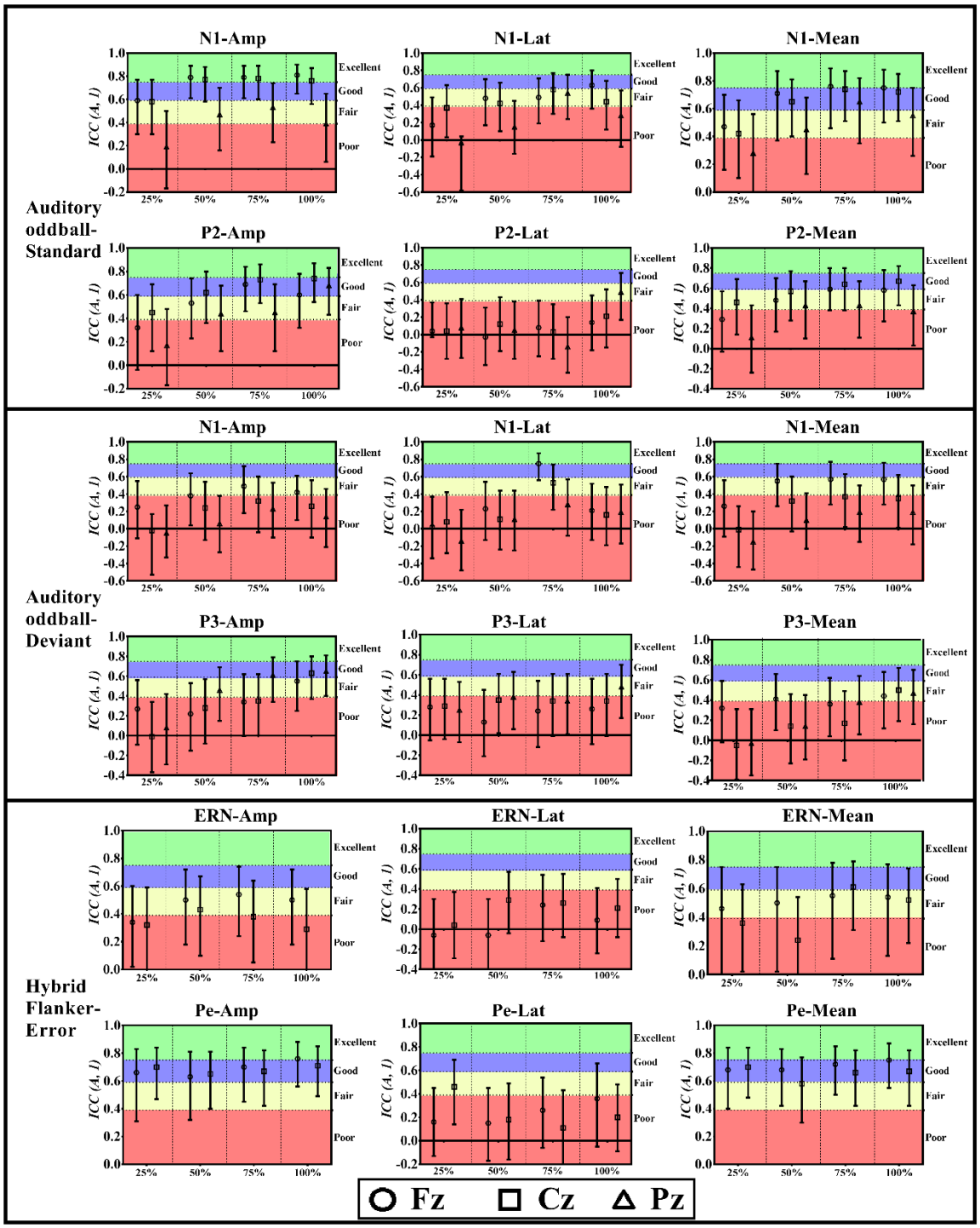


Fig. S1 Intra-class correlation coefficient (*ICC*) with increasing percentage of accepted trials of BL1-BL2 for ERP tasks. Four percentages of accepted epochs were assessed: 25%, 50%, 75% and 100 %. Test-retest reliability was estimated by single measure *ICC (A, 1)*. Mean and confident intervals for *ICC*s are shown in the figure.

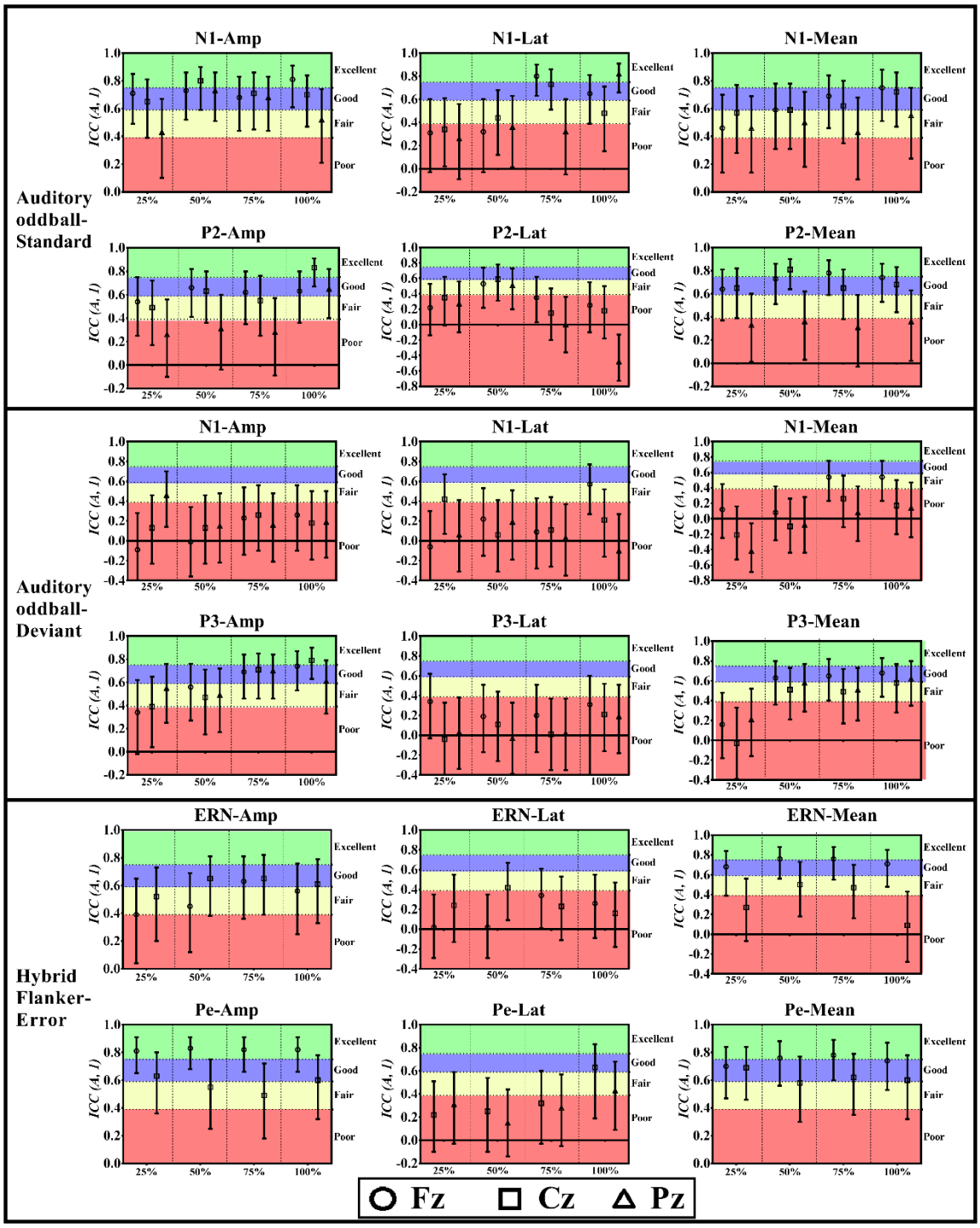


Fig. S2 Intra-class correlation coefficient (*ICC*) with increasing percentage of accepted trials of BL1-BL2 for ERP tasks. Four percentages of accepted epochs were assessed: 25%, 50%, 75% and 100 %. Test-retest reliability was estimated by single measure *ICC (A, 1)*. Mean and confident intervals for *ICC*s are shown in the figure.

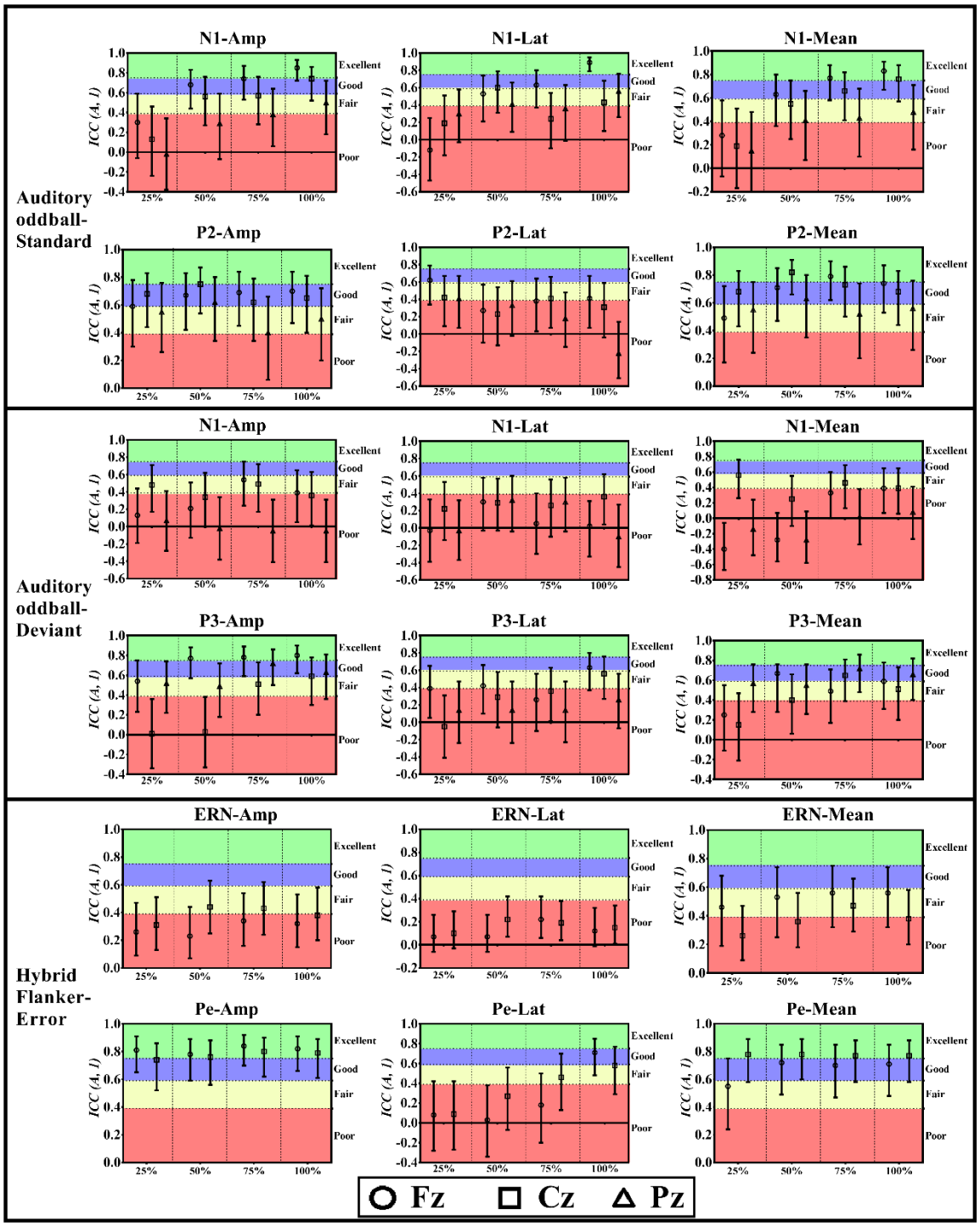


Fig. S3 Intra-class correlation coefficient (*ICC*) with increasing percentage of accepted trials of BL1-BL2 for ERP tasks. Four percentages of accepted epochs were assessed: 25%, 50%, 75% and 100 %. Test-retest reliability was estimated by single measure *ICC (A, 1)*. Mean and confident intervals for *ICC*s are shown in the figure.

Title: Pretreatment qEEG biomarkers for predicting pharmacological treatment outcome in Major Depressive Disorder: Independent validation from the NeuroPharm study

Authors: Cheng-Teng Ip[1,2,3], MSc, Sebastian Olbrich[4], MD, PhD, Melanie Ganz[2,5], PhD, Brice Ozenne[2,6], PhD, Kristin Köhler-Forsberg[2,3], MD, MSc, Vibeke H. Dam[2,3], MSc, Sándor Beniczky[7,8,9], PhD, Martin B. Jørgensen[3,10], MD, PhD, Vibe G. Frokjaer[2], MD, PhD, Birgitte Søgaard[1], PhD, Søren R. Christensen[1], PhD, Gitte M. Knudsen[2,3], MD, DMSc.

Affiliations:

[1]Department of Clinical Pharmacology, H. Lundbeck A/S, Valby, Denmark

[2]Neurobiology Research Unit, University Hospital Rigshospitalet, Copenhagen, Denmark

[3]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[4]Department of Psychiatry, Psychotherapy and Psychosomatic, University Zürich, Switzerland

[5]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

[6]Section of Biostatistics, Department of Public Health, University of Copenhagen. Denmark

[7]Department of Clinical Neurophysiology, Danish Epilepsy Center, Dianalund, Denmark

[8]Department of Clinical Medicine, University of Aarhus, Aarhus, Denmark

[9]Department of Clinical Neurophysiology, Aarhus University Hospital, Aarhus, Denmark

[10]Department of Psychiatry, Psychiatric Centre Copenhagen, Copenhagen, Denmark

Corresponding author:

Gitte M. Knudsen, MD, PhD
Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet, Building 6931, Blegdamsvej 9, DK-2100 Copenhagen, Denmark

Phone: (+45) 3545 6720, Email: gmk@nru.dk

Disclosures: All authors declare no conflict of interest.

Acknowledgement: We thank all the participants who participated in this study, the contributions of all involved referral sites and the great help from Gerda Thomsen, Svitlana Olsen and Lone Freyr for the recruitment of healthy controls and the preparation of the blood samples. We would like to express our sincere gratitude to Peter Jensen for the help of data management. Collection of the data included in the study was supported by Innovationsfonden, Denmark (4108-00004B). C.I. was supported by the Innovationsfonden, Denmark (5189-00087A). M.G. was supported by the Lundbeck Foundation (R181-2014-3586). B.O. was supported by the Lundbeck Foundation (R231-2016-3236).

Abstract:

Several electroencephalogram (EEG) biomarkers for prediction of drug response in major depressive disorder (MDD) have been proposed, but validations in larger independent datasets are missing. In the current study, we investigated the prognostic value of previously suggested EEG biomarkers. We gathered data that matched prior studies in terms of EEG methodology, clinical criteria for MDD, and statistical approach as closely as possible. The NeuroPharm study is a non-randomized and open label prospective clinical trial. One hundred antidepressant free patients with MDD were enrolled in the study and 79 (57 female) were included in the per-protocol analysis. The biomarkers candidates for cross-validation were derived from prior studies such as iSPOT-D and EMBARC and include frontal and occipital alpha power and asymmetry and delta and theta activity at anterior cingulate cortex (ACC). The alpha asymmetry, reported in two out of six prior studies, could be partially validated. We found that in female patients, larger right than left frontal alpha power prior to drug treatment was associated with better clinical outcome 8 weeks later. Moreover, female non-responder had higher central left alpha power relative to the right. In contrast to prior reports, we found that lower theta activity at ACC was present in remitters and was associated with greater improvement at week 8. We provide evidence that in women with MDD, alpha asymmetry seems to be the most promising EEG biomarker for prediction of treatment response.

Registration number: NCT02869035.

Keywords: qEEG, major depressive disorder, treatment response, pretreatment biomarker

1. Introduction

Major depressive disorder (MDD) is a heterogeneous disease with widespread biological causes, which explains why not all patients benefit from the same treatment (Spronk et al., 2011). Further, there is an often-prolonged time of treatment with trial-and-error regarding the choice of intervention (Baskaran et al., 2012). The situation could be remediated if patients were appropriately stratified for treatment selection (Olbrich and Arns, 2013), based on, e.g. their neurophysiological endophenotypes (Wu et al., 2020; Zhdanov et al., 2020).

Due to its low cost, high temporal resolution, and high accessibility, electroencephalography (EEG) shows great promise as a useful biomarker in clinical settings. As summarized in Table 1, several quantitative EEG (qEEG) biomarkers have shown potential to predict treatment outcome in MDD. Among those markers, alpha (Bruder et al., 2008) power (8–12 Hz rhythmic brain wave) and theta (Korb et al., 2009; Pizzagalli et al., 2001) power (4–8 Hz), sometimes in combination with a source localization technique (Pascual-Marqui, 1999; Pascual-Marqui et al., 1994), are the most promising candidates for clinical usage.

While these results are promising, they still suffer from low effect sizes, small samples, and importantly a lack of independent cross-validation in other cohorts. Another challenge is the use of different methodology e.g. by means of vigilance states under resting conditions (alert: Korb et al., 2009; relaxed: Pizzagalli et al., 2018) and heterogeneous treatment response criteria (Widge et al., 2018). In particular, differences in how treatment responders are defined limit comparisons between studies because some patients are considered responders in one study but not in another. Together these inconsistencies limit the validity and comparability of the reported findings.

We here aim to validate previously reported biomarkers relating to treatment outcome in MDD in an independent dataset (NeuroPharm). We meticulously applied the same criteria as previous studies regarding treatment response, biomarker definition, preprocessing/analysis methods and statistical approaches. We hypothesize that validation is achievable in our independent prospective study cohort.

2. Methods

2.1. Study design

NeuroPharm is a non-randomized, open label clinical trial (Figure 1). Details of the study protocol are described elsewhere (Köhler-Forsberg et al., 2020)

2.2. Participants and treatment

Two-hundred and fifty-nine antidepressant-free, MDD outpatients were screened from the central referral site and general practitioners, the Mental Health Services in the Capital Region of Denmark. MDD diagnosis was confirmed by a certified psychiatrist and confirmed by a Mini-International Neuropsychiatric Interview (Sheehan et al., 1998). Inclusion criteria for patients were: 18–65 years, moderate to severe, first or recurrent major depressive episode and a minimum score of 18 on Hamilton Depression Rating Scale 17 items ($HDRS_{17}$). Exclusion criteria: clinically significant psychosis, severe somatic co-morbidity, current or previous psychiatric severe co-morbidity, acute suicidal ideation and duration > 2 years of the current episode. One-hundred patients who met the criteria were allocated to treatment. All participants provided written informed consent prior to participation. The study was approval by the National Committee on Health Research Ethics (protocol: H-15017713).

Patients were treated with the selective serotonin reuptake inhibitor (SSRI) escitalopram at flexible doses of 5–20 mg/day adjusted depending on effects and side effects by trained physicians at each visit: week 1, 2, 4, 8 and 12. $HDRS_6$ was extracted from the $HDRS_{17}$ and used for assessment of antidepressant response (Bech et al., 2010, 2006; Østergaard et al., 2016). Definitions of treatment response are shown in Figure 1.

Consistent with clinical practice, patients with no response to escitalopram after 4 weeks, defined as early non-responders, or with intolerable side effect were offered second line treatment with the selective serotonin noradrenalin reuptake inhibitor (SNRI) duloxetine, with a dose ranging of 30–120 mg/day (n = 15). Medication compliance, side-effects to antidepressant treatment, and depressive symptoms were monitored at each visit.

A total of 79 patients were included in the per-protocol analyses: Two patients opted out of the study and for technical reasons, pretreatment EEG data was missing in five patients. One patient was excluded due to spontaneous remission and one patient was excluded due to suicidal ideations. Twelve patients dropped out before week 8. Three patients suspected of non-compliance because of drug levels below ≤ 5 nM at week 8 were excluded from the post-hoc analysis.

2.3. Electrophysiological recording

EEG recording was conducted 2.5 ± 2.4 days (mean ± SD) before treatment was initiated. All patients were seated on a comfortable armchair in a quiet room. Resting EEG was recorded with both eyes closed and open. Participants were instructed to remain quiet and relaxed, avoid eye-blinks and movements and to relax chin muscles during recording. Resting EEG was recorded during four 3-min periods with a counterbalanced order of OCOC (O for eyes open, C for eyes closed) or COCO between subjects. EEG data were recorded using a 256-channel HydroCel Sensor Net system (EGI, Inc., Eugene, OR) at 1000 Hz with 0.1–100 Hz analog filtering, the vertex electrode as reference. Impedances were kept below 50 kΩ. The high-density electrode array allowed the resampling of montages used by the different studies that we ought to validate.

2.4. Chosen biomarkers for validation

Besides the biomarkers showing promise in the meta-analysis, the study of Pizzaglli et al. (Pizzagalli et al., 2018) was also included (Table 1) since it is until present day the largest multicenter randomized placebo controlled clinical trial (Trivedi et al., 2016) and was published after the meta-analysis. Theta cordance and antidepressant treatment response index are treatment emergent biomarkers (for assessment one week after intervention), and thus were excluded from this analysis. Alpha activity derived from current source density was excluded (Tenke et al., 2011) because the criteria of acquiring the exact principal components were not reliably described. Since we only recorded EEG while patients were relaxed, we excluded studies where patients were kept alert (Cook et al., 2009; Iosifescu et al., 2009; Korb et al., 2009).

2.5. Electrophysiological Preprocessing and spectral analyses

Corrupted channels were interpolated using spline interpolation (Perrin et al., 1989). Resting EEG data were re-referenced offline to an average reference. Eye-movement artifacts were corrected using independent component analysis in EEGLAB (Delorme and Makeig, 2004). A zero-phase digital IIR Butterworth bandpass filter was applied, cut-off frequencies were 0.5 and 70 Hz, with orders of 2. An additional 50 Hz notch filter (order of 3) was applied. The data were cut into 1s epochs for visual artifact inspection (movement, muscle and electrical artifacts) and were further processed in Matlab 2017b (The Mathworks, Inc., Natick, MA, USA).

The data were re-epoched into 4s (adjusted when assessing different biomarkers) epochs with 50% overlapping for spectral analysis. Average power spectra were computed for O and C conditions separately using short-time Fourier transform after tapering with a Hanning window to suppress spectral leakage. For each electrode (see Table 1 for the selected electrodes), the resulting absolute power spectra were calculated over segmentations and over the following bands: overall alpha (8–13 Hz for Arns et al., 2016; 7.8–12.5 Hz for Bruder et al., 2008, 2001) and high alpha (10–12.5 Hz for Bruder et al., 2001). Logarithms of power values were applied.

2.5.1. eLORETA analysis

All resting data with eyes-closed were resampled to 250 Hz, in alignment with previous studies (Pizzagalli et al., 2001, 2018; Rentzsch et al., 2014), and further analyzed with exact low-resolution electromagnetic tomography (eLORETA) (for technical details refer to Pascual-Marqui, 2007; Pascual-Marqui et al., 2011). The eLORETA software computes three-dimensional intracerebral source distributions. Cross-spectra for delta (1.5–6 Hz) (Rentzsch et al., 2014), narrow (6.5–8 Hz) (Pizzagalli et al., 2001, 2018) and broad theta (4.5–7 Hz) bands (Pizzagalli et al., 2018) were computed. The extracted power of current source density was normalized for every subject and every band. Logarithms of power were extracted from the following regions of interest (ROI): Rostral anterior cingulate cortex (identical 14 voxels from Pizzagalli et al., 2001, 2018) and perigenual and anterior dorsal (pg/ad) ACC (identical 22 voxels from Rentzsch et al., 2014) based on previously evidence on treatment response (See Figure 2a for the visualization of the Montreal Neurological Institute brain atlas (MNI) coordinates).

2.6. Statistical analyses

The statistics were performed in SPSS version 24 (IBM Corp., Armonk, NY). The statistical models and covariates used to carry out the analyses are listed in table 2; they were kept as close as possible to the original studies. For treatment response definition, criteria of NeuroPharm (Figure 1) and of the resampled studies (Arns et al., 2016; Bruder et al., 2008, 2001; Pizzagalli et al., 2001, 2018; Rentzsch et al., 2014) were assessed. To allow for comparison across studies using different rating scales of treatment response, scales were transformed whenever possible (Riedel et al., 2010). Otherwise, responders were defined by at least 50% improvement of depressive symptoms as assessed by the $HDRS_{17}$ score at week 8. To investigate the value of biomarkers adds to the preexisting drug efficacy, number needed to treat (NNT) was reported when a significant effect of treatment responses appeared.

Bonferroni's correction was used to adjust for multiple comparisons in post-hoc analyses. Degrees of freedom were corrected by Greenhouse-Geisser correction when necessary. One-sided *p* values were chosen, for the sake of validation. Group differences in sex, age, education, pretreatment generalized anxiety disorder-10 (GAD-10) (Bech et al., 2005) and pretreatment HDRS scores were tested by simple *t* statistic or using the χ^2 statistic (sex). Two-sided *p* values were chosen when testing the demographical features.

3. Results

Groups did not differ in age, sex, education, pretreatment GAD_{10} and pretreatment HDRS scores for both criteria (all *p* values > .10). Table 2 summarizes the validation outcomes.

3.1. Frontal alpha asymmetry from Arns et al. (2016)

To align with the previous study (Arns et al., 2016), frontal alpha asymmetry (FAA) was assessed by the formula (F4-F3) of raw power values extracted from electrodes F4 and F3. A positive FAA indicates greater right than left alpha activity. A greater right FAA was reported in SSRI female responders but not in SNRI female responders previously (Arns et al., 2016), thus additional analysis without patients shifted to duloxetine was performed to allow direct comparisons between studies.

We found no group effect of FAA score (NeuroPharm: F (1, 24) = 0.14, p = .712) and no interaction was observed between conditions and group (NeuroPharm: F (1, 24) = 0.70, p = .412). Repeating the analysis for females only did not change the results (all p values > .31).

However, after excluding patients with low serum concentrations and patients shifted to duloxetine (only female patients, remaining n = 35), the results revealed significant sex-specific partial correlations between FAA score and pretreatment $HDRS_6$ scores (NeuroPharm: r (30) = -0.29, p = .048) and for the improvement at week 8 (NeuroPharm: r (30) = 0.32, p = .036) in eyes closed condition (Figure 3a), as well as in the eyes open condition (NeuroPharm: $HDRS_6$*:* r (30) = -0.28, p = .060; $HDRS_6$*:* r (30) = -0.31, p = .041). Consistent with previous study Arns et al. (2016), this association between treatment response and FAA was found in women but not in men. Since the general remission rate for female patients were 19.3%, and 24% of females with right FAA in favour for SSRI response. Therefore, the NNT for FAA is 21. When applying the response criteria of Arns et al. (2016), we did not find any discriminative value of FAA (all p values > .10).

3.2. Alpha asymmetry from Bruder et al. (2001)

Alpha asymmetry was assessed by integrating the alpha power at three regions: anterior (left, F3, F7; right, F4, F8), central (C3, T7; C4, T8), and posterior (P3, P7; P4, P8) (Bruder et al., 2001). Powers of overall alpha (7.8–12.5 Hz) and high alpha (10–12.5 Hz) were extracted independently at these three regions. Only data with eyes opened were assessed to align with previous study (Bruder et al., 2001).

We found no significant main group effects (NeuroPharm: F (1, 31) = 2.05, p = .162; criteria of the previous study: F (1, 76) = 1.44, p = .235) but we saw a significant interaction between sex and group on alpha asymmetry (NeuroPharm: F (1, 31) = 6.36, p = .017. When using the same criteria as in Bruder et al. (2001), we saw a trend only (F (1, 76) = 2.89, p = .093). The analyses of simple effects showed that male non-responders had a greater right than left alpha power (NeuroPharm: 0.14 vs. -0.15 on logarithmic scale, p = .007; criteria of the previous study: 0.06 vs. -0.05, p = .03). This interaction disappeared after the exclusion of patients with low serum drug concentrations (p values > .05). No significant difference was found for high alpha power (p values > .10).

Exploratory analysis

For validation purposes, i.e. adding regions as a within-subject factor, we also conducted the ANOVA without averaging the log alpha power at three regions. We found a significant four-way interaction between groups, sex, regions and hemisphere (NeuroPharm: F (2, 64) = 3.87, p = .041; criteria of the previous study: F (2, 150) = 4.31, p = .025). Analyses of simple effects showed not only greater right posterior alpha in male non-responders (NeuroPharm: 0.19 vs. -0.11, p = .022; NNT = 15; criteria of the previous study: 0.26 vs. -0.02, p = .008; NNT = 25), but also a less right central alpha in female non-responders (Criteria of the previous study: -0.62 vs. -0.55, p = .031; NNT = 42 (Figure 3b)). Post-hoc analysis after excluding patients with low serum concentrations did not change the results (criteria of the previous study: F (2, 144) = 3.99, p = .032). No significant result was found for high alpha (p values > .10).

3.3. Alpha power and alpha asymmetry from Bruder et al. (2008)

Alpha power (7.8–12.5 Hz) was extracted from the occipital sites (O1, O2) and entered the statistical model. There was no significant difference in occipital alpha among groups (NeuroPharm: F (1, 29) = 0.13, p = .723; criteria of the previous study: F (1, 65) = 0.83, p = .366), nor in occipital alpha asymmetry (NeuroPharm: F (1, 29) = 0.48, p = .496; criteria of the previous study: F (1, 65) = 0.19, p = .668). Repeating the analysis for right-handed only did not change the results (p values > .30).

3.4. Theta current source density at ACC from Pizzagalli et al. (2001)

A repeated ANOVA yielded a significant group effect (NeuroPharm: F (1, 33) = 5.10, p = .03), suggesting an overall higher ACC theta in non-responders compared to remitters (NeuroPharm: -2.38 vs. -2.64 eLORETA unit, p = .03, Figure 2b). Furthermore, the Pearson correlation revealed a negative trend between the whole cluster (14 voxels) and the improvement at week 8 (NeuroPharm: $HDRS_6$: r (79) = -0.18, p = .056), indicating that the lower pretreatment ACC theta was associated with the greater improvement on the depressive symptoms. The trend remained after exclusion of patients with low serum concentrations (NeuroPharm: r (76) = -0.17, p = .067). No significant results

were found when median cut of $HDRS_{17}$ (Pizzagalli et al., 2001) was applied as defining treatment response (criteria of the previous study: *p* values > .12, Figure 2b).

3.5. Theta current source density at ACC from Pizzagalli et al. (2018)

For this analysis, we focused on determining whether the narrow theta (6.5–8 Hz) and the broad theta (4.5–7 Hz) bands were associated with symptom improvement at week 8.

The partial correlation showed a trend of negative correlation between the narrow theta at the whole cluster (14 voxels) and the improvement at week 8 (NeuroPharm: $HDRS_6$, r (44) = -0.21, p = .085; criteria of the previous study: $HDRS_{17}$, r (44) = -0.23, p = .067), indicating that the lower pretreatment ACC narrow-theta was associated with greater improvement on the depressive symptoms at week 8. Excluding patients with low serum concentrations did not change the results (NeuroPharm: r (44) = -0.21, p = .087; criteria of the previous study: r (44) = -0.22, p = .075). No significant result was found for broad theta (*p* values > .29).

3.6. Delta current source density at pg/ad ACC from Rentzsch et al. (2014)

Pg/ad ACC delta was compared between groups by using the early responses at week 4 to align with the previous study (Rentzsch et al., 2014). Since $HDRS_{21}$ was not included in the current study, $HDRS_{17}$ Was used instead. Both criteria were used independently: NeuroPharm (see Figure 1 for the early status on treatment responses at week 4) and at least 50% reduction in $HDRS_{17}$.

There was no significant difference in pd/ad ACC delta among groups (NeuroPharm: F (1, 49) = 0.59, p = .448; criteria of the previous study: F (1, 81) = 0.99, p = .323). No significant correlation was found between pd/ad ACC delta and HDRS scores at week 4 (NeuroPharm: $HDRS_6$, r (83) = -0.03, p = .390; criteria of the previous study: $HDRS_{17}$, r (83) = 0.01, p = .477).

4. Discussion

This work aimed at cross-validating candidate qEEG biomarkers in MDD in a large independent prospective study. To the extent possible, the same EEG methodology, clinical definition of treatment response and statistical approach were applied, resulting in the partial validation of two studies (Table

2), both related to alpha asymmetry. In accordance with the literature (Arns et al., 2016), a sex-specific effect for alpha asymmetry was observed with higher right frontal alpha power being associated with greater improvement of symptoms at week 8 in female MDD. Furthermore, female non-responders showed higher left alpha power at the central sites. Alpha asymmetry *adds* to the already existing treatment efficacy of drug with an NNT of 21 in female MDD (and 42 for female non-responders). Previous findings of ACC theta and pg/ad ACC delta for treatment prediction were not validated. In contrast to the previously reported higher ACC theta in favor for SSRI response, we found that non-responders had *higher* ACC theta compared to remitters (Figure 2).

4.1. Alpha asymmetry

In partial validation of prior work (Arns et al., 2016), we found that female patients with greater right FAA (decreased right cortical activation) had milder depressive symptoms on $HDRS_6$ score prior to medication and show better improvement at week 8. In female non-responders, we found evidence for higher left central alpha asymmetry in the eyes-open condition compared to female responders. Although prior reports indicated an overall alpha asymmetry for non-responders (Bruder et al., 2001), with greater right cortical activity than left, and occipital alpha asymmetry (Bruder et al., 2008), our findings are only supportive of right hemisphere hyperactivation in non-responders (Arns et al., 2016; Bruder et al., 2008, 2001). Furthermore, we did not observe that FAA differ between remitters and non-responders (NeuroPharm) or between responders and non-responders (Arns et al., 2016) in neither eyes-open nor eyes-closed conditions. The reported NNT of alpha asymmetry reveals that every 21 women would benefit from the pretreatment selection, when a greater right FAA is present.

The sex-specific finding supports the presence of a sex-dependent lateralization in the serotonergic neurotransmitter system. A sex-specific cortical lateralization has been associated with the 5-HTTLPR (serotonin-transporter-linked polymorphic region) polymorphism (Volf et al., 2015). Furthermore, previous positron emission tomography (PET) studies have also found sex differences in cortical asymmetry of both the serotonin transporter (Kranz et al., 2014) and the serotonin 1A receptor (Fink et al., 2009). Future investigations on the sex-related asymmetry on cortical activation

(measured by EEG) and serotonergic system (measured by PET) could help understand this observation.

4.2. Delta/Theta current source density at ACC

Unlike prior studies (Korb et al., 2009; Mulert et al., 2007; Pizzagalli et al., 2001, 2018; Rentzsch et al., 2014), we did not find that higher slow-frequency activity at ACC was associated with better response; instead we found that remitters had lower theta activity at ACC compared to non-responders (Figure 2). Our results are aligned with the finding from one of the largest qEEG studies in MDD so far, the International Study to Predict Optimized Treatment in Depression (iSPOT-D), where they also report a lower theta activity in treatment responders (Arns et al., 2015). There seems to exist treatment specificity on theta activity in relation to treatment response, as proposed by Arns et al. (2015). Future studies are needed to investigate whether ACC theta power could inform about which type of antidepressant treatment to use.

4.3. Outcome depends on MDD scale or clinical criteria

The discriminative power of qEEG differed depending on the clinical rating scale used: We observed an association between FAA and $HDRS_6$ at week 8 but not when $HDRS_{17}$ was applied. We argure it is possible that $HDRS_6$ is more sensitive to the treatment outcome. The $HDRS_6$ excludes three negative side effects (somatic and gastrointestinal; sexual dysfunction; loss of weight). Thus the exclusion of less relevant or noisy items in the $HDRS_6$ should provide a better signal for estimating antidepressant response compared with $HDRS_{17}$ (Bech et al., 2010, 2006; Østergaard et al., 2016). We found that ACC theta was associated with treatment responses defined by $HDRS_6$ but not with $HDRS_{17}$. We propose that the $HDRS_6$ should be favored in future EEG studies of treatment responses.

4.4. Limitations

Despite the efforts to validate previous studies as closely as possible, several factors limit the interpretation of this study. First, the different recording lengths of resting EEG between studies was disregarded in the current analysis although vigilance states and EEG profiles might be tightly

associated (Hegerl and Hensch, 2014; Olbrich et al., 2016). In addition, high-dimensional montage (257-channel) was used for source analysis instead of the montages used in previous studies (Pizzagalli et al., 2001, 2018; Rentzsch et al., 2014). The high-dimensional recording aimed at reducing the possible topographical interpolation, resulting in a more accurate estimation of the current source density. However, we cannot exclude that this difference could result in inconsistencies between the current and the prior results. Further investigation of the robustness of the location of rACC derived from various montages is needed. Finally, validations were impacted by the variability in clinical measures used in previous studies although we attempted to take that into account.

5. Conclusions

We could partially validate two out of six qEEG candidate biomarkers, which were both related to alpha asymmetry. Applying FAA alongside with treatment would increase efficacy around 20%. Future, prospective studies of qEEG biomarkers should aim to standardize recording length, montages, marker calculation and clinical measures, and it is also possible that a combination of markers might increase the accuracy for treatment prediction.

6. References

Arns, M., Bruder, G., Hegerl, U., Spooner, C., Palmer, D.M., Etkin, A., Fallahpour, K., Gatt, J.M., Hirshberg, L., Gordon, E., 2016. EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized iSPOT-D study. Clin. Neurophysiol. 127, 509–519. https://doi.org/10.1016/j.clinph.2015.05.032

Arns, M., Etkin, A., Hegerl, U., Williams, L.M., DeBattista, C., Palmer, D.M., Fitzgerald, P.B., Harris, A., deBeuss, R., Gordon, E., 2015. Frontal and rostral anterior cingulate (rACC) theta EEG in depression: Implications for treatment outcome? Eur. Neuropsychopharmacol. 25, 1190–1200. https://doi.org/10.1016/j.euroneuro.2015.03.007

Baskaran, A., Milev, R., McIntyre, R.S., 2012. The neurobiology of the EEG biomarker as a predictor of treatment response in depression. Neuropharmacology 63, 507–513. https://doi.org/10.1016/j.neuropharm.2012.04.021

Bech, P., Boyer, P., Germain, J.-M., Padmanabhan, K., Haudiquet, V., Pitrosky, B., Tourian, K.A., 2010. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. Pharmacopsychiatry 43, 271–276. https://doi.org/10.1055/s-0030-1263173

Bech, P., Kajdasz, D.K., Porsdal, V., 2006. Dose-response relationship of duloxetine in placebo-controlled clinical trials in patients with major depressive disorder. Psychopharmacology (Berl). 188, 273–280. https://doi.org/10.1007/s00213-006-0505-1

Bech, P., Lunde, M., Unden, M., 2005. An inventory for the measurement of generalized anxiety distress symptoms, the GAD-10 Inventory. Acta Psychiatr Belg 105, 111–118.

Bruder, G.E., Sedoruk, J.P., Stewart, J.W., McGrath, P.J., Quitkin, F.M., Tenke, C.E., 2008. Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: Pre- and post-treatment findings. Biol. Psychiatry 63, 1171–1177. https://doi.org/10.1016/j.biopsych.2007.10.009

Bruder, G.E., Stewart, J.W., Tenke, C.E., McGrath, P.J., Leite, P., Bhattacharya, N., Quitkin, F.M., 2001. Electroencephalographic and perceptual asymmetry differences between responders and nonresponders to an SSRI antidepressant. Biol. Psychiatry 49, 416–425. https://doi.org/10.1016/S0006-3223(00)01016-7

Cook, I.A., Hunter, A.M., Abrams, M., Siegman, B., Leuchter, A.F., 2009. Midline and right frontal brain function as a physiologic biomarker of remission in major depression. Psychiatry Res. - Neuroimaging 174, 152–157. https://doi.org/10.1016/j.pscychresns.2009.04.011

Delorme, A., Makeig, S., 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. J. Neurosci. Methods 13, 9–21. https://doi.org/http://dx.doi.org/10.1016/j.jneumeth.2003.10.009

Fink, M., Wadsak, W., Savli, M., Stein, P., Moser, U., Hahn, A., Mien, L.K., Kletter, K., Mitterhauser, M., Kasper, S., Lanzenberger, R., 2009. Lateralization of the serotonin-1A receptor distribution in language areas revealed by PET. Neuroimage 45, 598–605. https://doi.org/10.1016/j.neuroimage.2008.11.033

Hegerl, U., Hensch, T., 2014. The vigilance regulation model of affective disorders and ADHD. Neurosci. Biobehav. Rev. 44, 45–57. https://doi.org/10.1016/j.neubiorev.2012.10.008

Iosifescu, D. V., Greenwald, S., Devlin, P., Mischoulon, D., Denninger, J.W., Alpert, J.E., Fava, M., 2009. Frontal EEG predictors of treatment outcome in major depressive disorder. Eur. Neuropsychopharmacol. 19, 772–777. https://doi.org/10.1016/j.euroneuro.2009.06.001

Köhler-Forsberg, K., Jørgensen, A., Dam, V., Stenbæk, D., Fisher, P., Ip, C.-T., Melanie, G., Poulsen, H., Giraldi, A., Ozenne, B., Jørgensen, M., Knudsen, G., Frokjaer, V., 2020. Predicting treatment outcome in Major Depressive Disorder using serotonin 4 receptor PET brain imaging, functional MRI, cognitive-, EEG-based and peripheral biomarkers: a NeuroPharm open label clinical trial protocol. Front. Psychiatry.

Korb, A.S., Hunter, A.M., Cook, I.A., Leuchter, A.F., 2009. Rostral anterior cingulate cortex theta

current density and response to antidepressants and placebo in major depression. Clin. Neurophysiol. 120, 1313–1319. https://doi.org/10.1016/j.clinph.2009.05.008

Kranz, G.S., Hahn, A., Baldinger, P., Haeusler, D., Philippe, C., Kaufmann, U., Wadsak, W., Savli, M., Hoeflich, A., Kraus, C., Vanicek, T., Mitterhauser, M., Kasper, S., Lanzenberger, R., 2014. Cerebral serotonin transporter asymmetry in females, males and male-to-female transsexuals measured by PET in vivo. Brain Struct. Funct. 219, 171–183. https://doi.org/10.1007/s00429-012-0492-4

Mulert, C., Juckel, G., Brunnmeier, M., Karch, S., Leicht, G., Mergl, R., Möller, H.J., Hegerl, U., Pogarell, O., 2007. Prediction of treatment response in major depression: Integration of concepts. J. Affect. Disord. 98, 215–225. https://doi.org/10.1016/j.jad.2006.07.021

Olbrich, S., Arns, M., 2013. EEG biomarkers in major depressive disorder: Discriminative power and prediction of treatment response. Int. Rev. Psychiatry 25, 604–618. https://doi.org/10.3109/09540261.2013.816269

Olbrich, S., Tränkner, A., Surova, G., Gevirtz, R., Gordon, E., Hegerl, U., Arns, M., 2016. CNS- and ANS-arousal predict response to antidepressant medication: Findings from the randomized iSPOT-D study. J. Psychiatr. Res. 73, 108–115. https://doi.org/10.1016/j.jpsychires.2015.12.001

Østergaard, S.D., Bech, P., Miskowiak, K.W., 2016. Fewer study participants needed to demonstrate superior antidepressant efficacy when using the Hamilton melancholia subscale (HAM-D6) as outcome measure. J. Affect. Disord. 190, 842–845. https://doi.org/10.1016/j.jad.2014.10.047

Pascual-Marqui, R.D., 2007. Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv:0710.3341 [math-ph], 1–16.

Pascual-Marqui, R.D., 1999. Review of Methods for Solving the EEG Inverse Problem. Int. J. Bioelectromagn. 1, 75–86.

Pascual-Marqui, R.D., Lehmann, D., Koukkou, M., Kochi, K., Anderer, P., Saletu, B., Tanaka, H., Hirata, K., John, E.R., Prichep, L., Biscay-Lirio, R., Kinoshita, T., 2011. Assessing interactions

in the brain with exact low-resolution electromagnetic tomography. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 369, 3768–3784. https://doi.org/10.1098/rsta.2011.0081

Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int. J. Psychophysiol. 18, 49–65. https://doi.org/10.1016/0167-8760(84)90014-X

Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187. https://doi.org/https://doi.org/10.1016/0013-4694(89)90180-6

Pizzagalli, D., Pascual-Marqui, R.D., Nitschke, J.B., Oakes, T.R., Larson, C.L., Abercrombie, H.C., Schaefer, S.M., Koger, J. V., Benca, R.M., Davidson, R.J., 2001. Anterior cingulate activity as a predictor of degree of treatment response in major depression: Evidence from brain electrical tomography analysis. Am. J. Psychiatry 158, 405–415. https://doi.org/10.1176/appi.ajp.158.3.405

Pizzagalli, D.A., Webb, C.A., Dillon, D.G., Tenke, C.E., Kayser, J., Goer, F., Fava, M., McGrath, P., Weissman, M., Parsey, R., Adams, P., Trombello, J., Cooper, C., Deldin, P., Oquendo, M.A., McInnis, M.G., Carmody, T., Bruder, G., Trivedi, M.H., 2018. Pretreatment rostral anterior cingulate cortex theta activity in relation to symptom improvement in depression: A randomized clinical trial. JAMA Psychiatry 75, 547–554. https://doi.org/10.1001/jamapsychiatry.2018.0252

Rentzsch, J., Adli, M., Wiethoff, K., Gómez-Carrillo De Castro, A., Gallinat, J., 2014. Pretreatment anterior cingulate activity predicts antidepressant treatment response in major depressive episodes. Eur. Arch. Psychiatry Clin. Neurosci. 264, 213–223. https://doi.org/10.1007/s00406-013-0424-1

Sheehan, D. V, Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry.

Spronk, D., Arns, M., Barnett, K.J., Cooper, N.J., Gordon, E., 2011. An investigation of EEG, genetic and cognitive markers of treatment response to antidepressant medication in patients with major depressive disorder: A pilot study. J. Affect. Disord. 128, 41–48. https://doi.org/10.1016/j.jad.2010.06.021

Tenke, C.E., Kayser, J., Manna, C.G., Fekri, S., Kroppmann, C.J., Schaller, J.D., Alschuler, D.M., Stewart, J.W., McGrath, P.J., Bruder, G.E., 2011. Current source density measures of electroencephalographic alpha predict antidepressant treatment response. Biol. Psychiatry 70, 388–394. https://doi.org/10.1016/j.biopsych.2011.02.016

Trivedi, M.H., McGrath, P.J., Fava, M., Parsey, R. V., Kurian, B.T., Phillips, M.L., Oquendo, M.A., Bruder, G., Pizzagalli, D., Toups, M., Cooper, C., Adams, P., Weyandt, S., Morris, D.W., Grannemann, B.D., Ogden, R.T., Buckner, R., McInnis, M., Kraemer, H.C., Petkova, E., Carmody, T.J., Weissman, M.M., 2016. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. J. Psychiatr. Res. 78, 11–23. https://doi.org/10.1016/j.jpsychires.2016.03.001

Volf, N. V., Belousova, L. V., Knyazev, G.G., Kulikov, A. V., 2015. Gender differences in association between serotonin transporter gene polymorphism and resting-state EEG activity. Neuroscience 284, 513–521. https://doi.org/10.1016/j.neuroscience.2014.10.030

Widge, A.S., Bilge, M.T., Montana, R., Chang, W., Rodriguez, C.I., Deckersbach, T., Carpenter, L.L., Kalin, N.H., Nemeroff, C.B., 2018. Electroencephalographic biomarkers for treatment response prediction in major depressive illness: A meta-analysis. Am. J. Psychiatry 176, 44–56. https://doi.org/10.1176/appi.ajp.2018.17121358

Wu, W., Zhang, Y., Jiang, J., Lucas, M. V, Fonzo, G.A., Rolle, C.E., Cooper, C., Chin-Fatt, C.,

Krepel, N., Cornelssen, C.A., 2020. An electroencephalographic signature predicts antidepressant response in major depression. Nat. Biotechnol. 1–9. https://doi.org/https://doi.org/10.1038/s41587-019-0397-3

Zhdanov, A., Atluri, S., Wong, W., Vaghei, Y., Daskalakis, Z.J., Blumberger, D.M., Frey, B.N., Giacobbe, P., Lam, R.W., Milev, R., Mueller, D.J., Turecki, G., Parikh, S. V., Rotzinger, S., Soares, C.N., Brenner, C.A., Vila-Rodriguez, F., McAndrews, M.P., Kleffner, K., Alonso-Prieto, E., Arnott, S.R., Foster, J.A., Strother, S.C., Uher, R., Kennedy, S.H., Farzan, F., 2020. Use of Machine Learning for Predicting Escitalopram Treatment Outcome From Electroencephalography Recordings in Adult Patients With Depression. JAMA Netw. open 3, e1918377. https://doi.org/10.1001/jamanetworkopen.2019.18377

Table 1 The listed studies for validation and the details of previous published biomarkers

Studies	qEEG biomarkers	Recording duration	R	NR	Clinical evaluation	Medication	Response criteria
Arns et al., (2016)	Greater FAA [(F4+F3)/(F4-F3)] in R and better response to SSRI in female	Two 2-min (CO[1] or OC)	427	240	**$HDRS_{17}$**; Baseline, week 8	escitalopram, sertraline, venlafaxine	R: ≥ 50% improved on $HDRS_{17}$. Remission: score of ≤ 7 on the $HRSD_{17}$
Bruder et al., (2001)	Less right alpha (F3, F7, F4, F8, C3, T7, C4, T8, P3, P7, P4, P8) than left in female NR	Four 2-min (COOC or OCCO)	34	19	**CGI-I**; Baseline, week 12	fluoxetine	R: "Much improved" or "very much improved" on CGI-I
Bruder et al., (2008)	1. Greater occipital alpha (O1, O2) in R; 2. Greater right hemispheric alpha in R.	Four 2-min (COOC or OCCO)	11	7	**CGI-I**; Baseline, week 12	fluoxetine	R: "Much improved" or "very much improved" on CGI-I
Pizzagalli et al., (2001)	Higher rACC theta (LORETA) activity, better response	Ten 3-min (COCOCOCOCO or OCOCOCOCOC)	9	9	**BDI**; Baseline, 4-6 months	nortriptyline	Median split of BDI scores
Pizzagalli et al., (2018)	Higher rACC theta (LORETA) activity, better response	Four 2-min (COOC or OCCO)	248[2]		**$HDRS_{17}$**; Baseline, weeks 1, 2, 3, 4, 6, and 8	sertraline, placebo	Absolute score on $HDRS_{17}$
Rentzsch et al., (2014)	Higher right pg/adACC (LORETA) delta in R	≥ 10 min (C)	11	20	**$HDRS_{21}$**; Baseline, weeks 2 and 4	various SSRIs	R: ≥ 50% improved on $HDRS_{21}$ at week 4.

Notes: [1] C refers to eyes closed, and O refers to eyes open. [2] The study did not provide information on the numbers for treatment responders/non-responders.

Abbreviation: R, responders; NR, non-responders; SSRI, selective serotonin reuptake inhibitor; FAA: frontal alpha asymmetry; rACC, rostral anterior cingulate cortex; pg/ad ACC, perigenual and anterior dorsal anterior cingulate cortex; HDRS, Hamilton Depression Rating Scale; CGI-I, Clinical Global Impression Improvement scale; BDI, Beck Depression Inventory.

Table 2 Summary of the statistical analysis and the validation results

Study	Statistical analysis[1]	Included covariates	Results[3]	
			Criteria of NeuroPharm	Criteria of the previous study
Arns et al. (2016)	**Primary**: ANOVA with FAA score under different conditions and responses' groups. Sex-specific effect on female patients. **Secondary**: Partial correlation between FAA and HDRS scores. **Exploratory:** ANOVA with hemisphere and groups.	Age, sex, pretreatment anxiety level and pretreatment HDRS score. Age and sex were included in the exploratory analysis.	Partial validation: Greater right FAA, better response in female	No validation
Bruder et al. (2001)	**Primary:** ANOVA with alpha power (overall and high alpha) on both hemisphere, groups and sex were tested. **Exploratory**: ANOVA with regions (anterior, central, posterior), hemisphere, groups and sex were tested.	Not applicable	No validation	Partial validation: Less right central alpha in female NR
Bruder et al. (2008)	**Primary:** ANOVA with hemisphere and condition and groups. Possible handedness effect on right-handed patients were examined.	Number of years of education	No validation	No validation
Pizzagalli et al. (2001)	**Primary:** ANOVA with narrow ACC theta (14 voxels) and groups. **Secondary:** Pearson correlation between ACC theta and HDRS scores.	Not applicable	No validation[4]	No validation
Pizzagalli et al. (2018)	**Primary:** Partial correlation between ACC theta (narrow and broad theta) and the ΔHDRS at week 8.	Age, sex, race, marital status, employment status, pretreatment anxiety level and pretreatment HDRS score[3]	No validation[4]	No validation[4]
Rentzsch et al. (2014)	**Primary:** ANOVA with ACC delta (22 voxels) and groups. **Secondary:** Pearson correlation between ACC delta and ΔHDRS at week 4.	Not applicable	No validation	No validation

Notes: [1] The chosen models followed the resampled study as close as possible. Only the biomarkers that has been found to associate with treatment response in the previous study were examined here. [2] Tailored from the final model reported in the study of Pizzagalli et al., (2018). [3] Criteria of NeuroPharm and of the previous studies were reported separately. [4]An opposite direction of the resample study was observed.

Abbreviation: ANOVA, repeated-measures analysis of variance; FAA, Frontal alpha asymmetry; ACC, anterior cingulate cortex; HDRS, Hamilton Depression Rating Scale.

Figure captions:

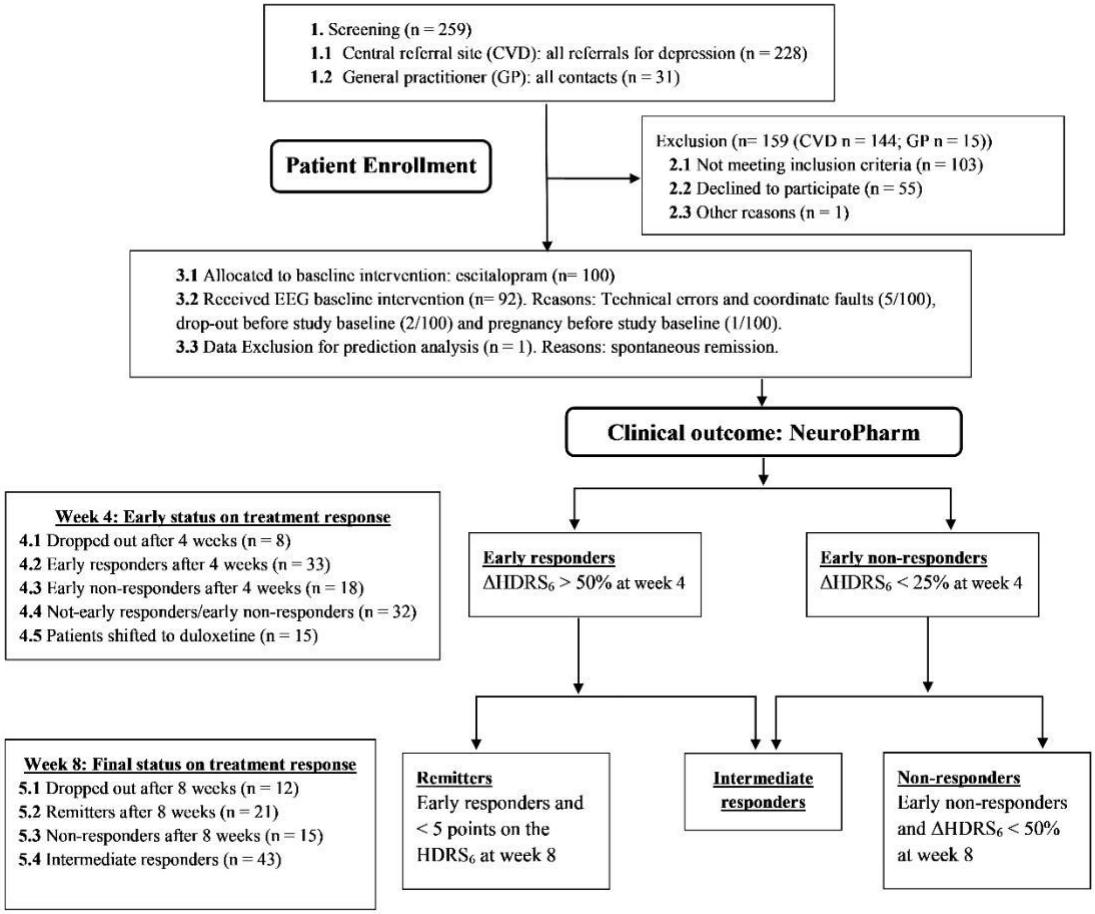


Figure 1. Study flowchart and the definition of treatment response in NeuroPharm. Antidepressant-free patients diagnosed with MDD were included and allocated to baseline intervention. Resting EEG data were collected at pretreatment and week 8 after treatment intervention. Compliance, side-effects to antidepressant treatment, and clinical evaluations of patients were conducted by a trained clinician at each visit (week 1, 2, 4, 8 and 12). The treatment responses were measured by $HDRS_6$ ((HDRS score at week 4/8 - HDRS score at pretreatment)/ HDRS score at pretreatment) and defined as follows: Patients with $HDRS_6$ > 50% at week 4 were considered early responders; Patients with $HDRS_6$ < 25% at week 4 were considered early non-responders; Patients with $HDRS_6$ > 50% at week 4, and < 5 points on the $HDRS_6$ scale at week 8 were considered remitters; Patients with $HDRS_6$ < 25% at week 4 and $HDRS_6$ < 50% at week 8 were considered non-responders. The other patients were considered intermediate responders.

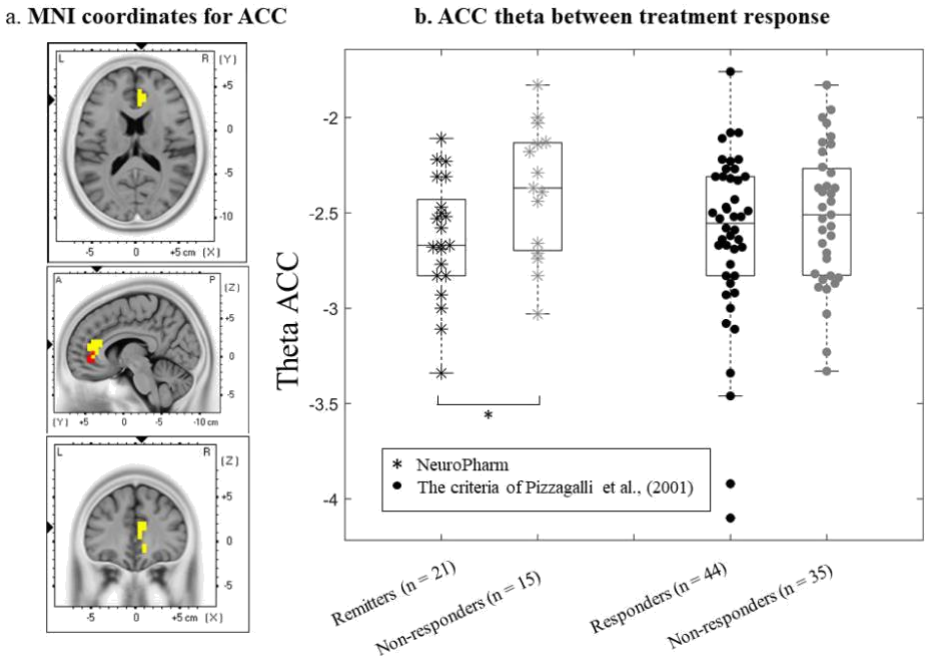


Figure 2. **a**. The visualization of the chosen MNI-coordinates (view angle: [X, Y, Z] = [7, 35, 16] mm). Theta power was extracted from a rACC cluster (identical 14 voxels from 5,9, highlighted as red) and delta power was extracted from pg/adACC (identical 22 voxels from 23, highlighted as yellow). **b**. Theta activity at rACC cluster (red blocks) for both response criteria: NeuroPharm and the previous study (Pizzagalli et al., 2001). Logarithmic theta current source density was extracted from rACC using eLORETA. When the criteria of NeuroPharm was assessed, the results showed a significant higher ACC theta in non-responders compared to remitters. No such difference was observed when the criteria of the resampled study was assessed. Data were visualized the same way as the resampled study (Pizzagalli et al., 2001).

Note: *: $p < .05$.

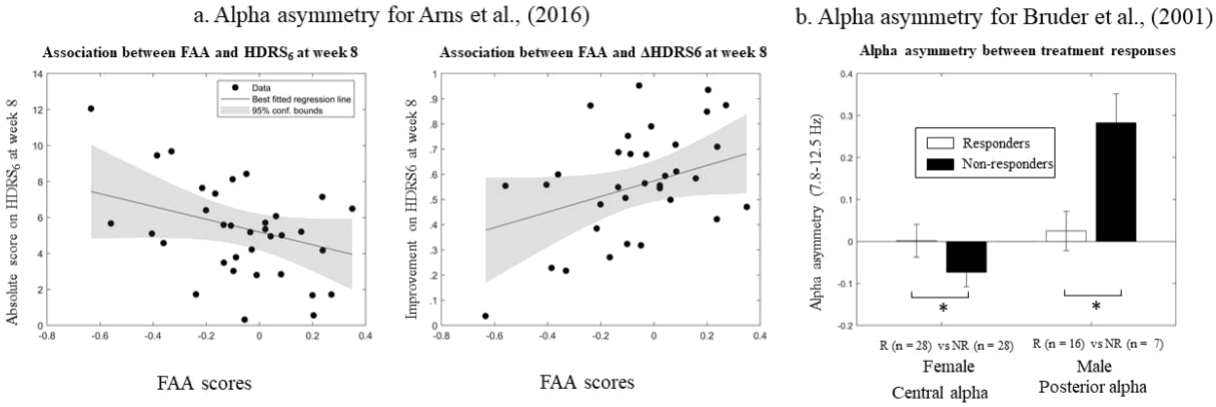


Figure 3. **a.** A negative partial correlation was found between FAA score and $HDRS_6$ scores at week 8 (Figure 3a left panel) and a positive partial correlation was found between FAA score and $\Delta HDRS_6$ at week 8 (Figure 3a right panel). Grey areas indicate 95% confident intervals for the fitted lines. **b.** Mean alpha asymmetry when the criterion of the study of Bruder et al. (2001) was used. Error bars indicate standard deviations. A positive score means that alpha power is larger in the right than in the left region. Data were visualized the same way as the resampled study (Bruder et al., 2001).

Note: *: $p < .05$.

NeuroPharm Study: EEG wakefulness regulation as a biomarker in MDD

Cheng-Teng Ip[1,2,3], Melanie Ganz[2,4], Vibeke H. Dam[2,3], Annia Rüesch[5], Kristin Köhler-Forsberg[2,3], Martin B. Jørgensen[3,6], Vibe G. Frokjaer[2,3,6], Birgitte Søgaard[1], Søren R. Christensen[1], Gitte M. Knudsen[2,3], Sebastian Olbrich[5]*

Affiliations:

[1] Department of Experimental Medicine, H. Lundbeck A/S, Valby, Denmark

[2] Neurobiology Research Unit, University Hospital Rigshospitalet, Copenhagen, Denmark

[3] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[4] Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

[5] Department for Psychiatry, Psychotherapy and Psychosomatic; University Zurich, Switzerland

[6] Department of Psychiatry, Psychiatric Centre Copenhagen, Copenhagen, Denmark

Word count abstract: 233

Word count text: app. 3997

Figures: 4

Tables: 2

* Correspondence to: Sebastian Olbrich

Clinic for Psychiatry, Psychotherapy and Psychosomatic, University Zürich

Militärstrasse 8

8021 Zürich, Switzerland

Tel:t +41 (0)44 2967400

E-mail: Sebastian.olbrich@puk.zh.ch

Running head: EEG-wakefulness regulation in MDD

ABSTRACT:

While several electroencephalogram (EEG)-based biomarkers have been proposed as diagnostic or predictive tools in major depressive disorder (MDD), there is a clear lack of replication studies in this field. Markers that link clinical features such as disturbed wakefulness regulation in MDD with neurophysiological patterns are particularly promising candidates for e.g. EEG-informed choices of antidepressive treatment. We investigate if we in an independent MDD sample can replicate abnormal findings of EEG-vigilance regulation during rest and as a predictor for antidepressive treatment response. EEG-resting state was recorded in 91 patients and 35 healthy controls from the NeuroPharm trial. EEG-vigilance was assessed using the Vigilance Algorithm Leipzig (VIGALL). We compared the vigilance regulation during rest between patients and healthy controls and between remitters/responders and non-remitters/non-responders after eight weeks of SSRI/SNRI treatment using two different sets of response criteria (NeuroPharm and iSPOT-D). We replicated previous findings showing hyperstable EEG-wakefulness regulation in patients in comparison to healthy subjects. Responders defined by the iSPOT-D criteria showed faster declines toward low vigilance stages in comparison to patients with no response at pretreatment, however this did not apply when using the NeuroPharm criteria. EEG-wakefulness regulation patterns normalized toward patterns of healthy controls especially in responders and remitters after 8 weeks of treatment. This replication study supports the diagnostic value of EEG-vigilance regulation and its usefulness as biomarker for the choice of treatment in MDD.

Clinical Trials Registration: https://clinicaltrials.gov/ct2/show/NCT02869035?draw=1

Registration number: NCT0286903.

Keywords: EEG, VIGALL, antidepressant treatment effect, SSRI, MDD, biomarker

INTRODUCTION

Despite concerted efforts, there is still a lack of validated objective biomarkers in major depressive disorder (MDD) (Kennis et al., 2020) although this would be useful not only for differential diagnostic purposes but also for predicting pharmacological treatment response (Olbrich and Conradi, 2016). Neurophysiological methods have gained attention as they provide an affordable framework that reflects the functional aspects of the nervous system at a high-resolution timescale (Olbrich and Arns, 2013; Widge et al., 2018). In particular, electroencephalogram (EEG) is a non-invasive measurement of neuronal activity with the potential to provide clinically relevant biomarkers (Grzenda and Widge, 2020; Olbrich and Arns, 2013; Rolle et al., 2020a).

Dysregulation of sleep and wakefulness is part of the diagnostic criteria for MDD (Nutt et al., 2008; Seifritz, 2001): Patients often feel tired and fatigued during the day but have difficulties falling asleep and wake up early which leads to a vicious cycle of sleep disturbances and tiredness. Polysomnographic measurements have with some success been used to differentiate patients and healthy controls (Thase, 2006), but there is only limited value for those markers to predict antidepressive drug response (Steiger and Pawlowski, 2019) and is not sensitive to important aspects of the sleep-wake dysregulation in MDD. The Vigilance Algorithm Leipzig (VIGALL) has been developed to identify and analyze different functional brain states from full wakefulness to sleep onset with closed eyes by using EEG and electrooculogram data (Hegerl et al., 2012; Olbrich et al., 2015). The outcome of the algorithm has been compared with other imaging modalities (Guenther et al., 2011; Olbrich et al., 2009) and clinical data (Jawinski et al., 2015).

According to the VIGALL framework of EEG wakefulness regulation, MDD patients tend to show a hyperstable wakefulness regulation with fewer declines toward relaxation and sleep stages in comparison to healthy subject (Hegerl and Hensch, 2014; Olbrich et al., 2012). For the full clinical utility of VIGALL, it is essential to demonstrate its prognostic power for treatment outcome. So

far, two studies have investigated this. Data from the International Study to Predict Optimised Treatment - in Depression (iSPOT-D) (Williams et al., 2011) showed that patients with a good response to selective serotonin reuptake inhibitors (SSRIs) had a faster decline of wakefulness during a two-minute resting state than non-responders (Olbrich et al., 2016a). Contradictory, a study by Schmidt et al., (2017) showed that patients with a pronounced hyperstable wakefulness regulation during 15 minutes of rest were more likely to respond to SSRI treatment.

To resolve these differences and to overcome the limitations of missing replications (Widge et al., 2018), the aim of this study was to investigate the VIGALL outcomes in an independent cohort of patients with MDD from the NeuroPharm trial (Köhler-Forsberg et al., 2020). We wanted to see if we could replicate 1) the findings of a hyperstable EEG wakefulness regulation in MDD in comparison to healthy controls and 2) the predictive properties of the VIGALL algorithm with respect to treatment outcome for SSRIs and SNRIs. It was hypothesized that 1) patients suffering from MDD will show more high vigilance stages and fewer declines toward sleep stages in comparison to healthy controls and 2) that responders will show a less stable EEG wakefulness regulation over time as assessed with the VIGALL algorithm. In an exploratory analysis, we also assess the treatment effect on VIGALL parameters.

MATERIALS AND METHODS

NeuroPharm is a non-randomized, open label clinical trial in an outpatient setting. A completed consortium diagram and the flowchart of the study protocol are available elsewhere (Köhler-Forsberg et al., 2020, also see Figure 1).

Subjects

One-hundred medication-free MDD outpatients were recruited and their diagnosis was ascertained by a Mini-International Neuropsychiatric Interview (Sheehan et al., 1998) and

confirmed by a certified psychiatrist. Details of inclusion and exclusion criteria are described elsewhere (Köhler-Forsberg et al., 2020). Thirty-five healthy subjects were included as controls (See Table 1 for demographic features). All participants provided written informed consent prior to participation. Ethical approval was obtained by the National Committee on Health Research Ethics (protocol: H-15017713).

After pretreatment visit, patients were treated first line with the SSRI escitalopram at flexible doses of 5–20 mg/day. Doses were adjusted depending on effects and side effects by trained physicians at each visit at weeks 1, 2, 4, 8 and 12. Hamilton Depression Rating Scale 6 items ($HDRS_6$) was used for assessment of antidepressant response since it has been shown to be more sensitive to treatment outcome (Bech et al., 2010, 2006; Østergaard et al., 2016). Patients with no response to escitalopram after 4 weeks, or with intolerable side effects were offered second line treatment with the selective serotonin noradrenalin reuptake inhibitor (SNRI) duloxetine, with a dose ranging of 30–120 mg/day (n = 15).

EEG recordings were obtained at pretreatment visits (unmedicated) for all included participants, and 40 of the patients were recorded again after 8 weeks of treatment. A total of 79 patients (See Figure 1 for the exclusion reason) and 35 healthy controls were included in the per-protocol analyses for the pretreatment data. In order to explore the drug effect, three patients suspected of non-compliance because of drug levels below ≤ 5 nM at week 8 were excluded from the post-hoc analysis. EEG data from 39 patients´ data was available for further analyses at 8 weeks follow-up.

Clinical measures and treatment response

The treatment responses of NeuroPharm were measured as percentage change in $HDRS_6$ score from baseline to week 4 or 8 ($\Delta HDRS_6$). Three categorical treatment outcomes were defined

(Figure 1): Remitters included patients with $\Delta HDRS_6$ > 50% at week 4 and < 5 points on the $HDRS_6$ scale at week 8; non-responders included patients with $\Delta HDRS_6$ < 25% at week 4 and $\Delta HDRS_6$ < 50% at week 8 were considered non-responders. Patients who did fulfill either remitter or non-responder criteria were classified as intermediate responders (Figure 1).

To allow comparisons with already existing studies, the criterium defined by the NeuroPharm consortium (Köhler-Forsberg et al., 2020) and the one used for the iSPOT-D data (Olbrich et al., 2016b) were assessed. When treatment response of iSPOT-D was assessed, using the NeuroPharm data, remission was defined as a score of 7 at week 8 on $HDRS_{17}$ and response as > 50% decrease in $\Delta HDRS_{17}$ (from pretreatment to week 8).

EEG recordings

Resting EEG was recorded during four 3 min periods with a counterbalanced order of OCOC (O for eyes open, C for eyes closed) or COCO between subjects (details of recording condition are detailed elsewhere: Ip et al., submitted). EEG data was acquired from a 256-channel HydroCel Sensor Net system (EGI, Inc., Eugene, OR) at 1000 Hz sampling rate and referenced to the vertex electrode. Impedances were kept below 50 kΩ.

EEG processing and classification of EEG-vigilance

Corrupted channels were interpolated using spline interpolation (Perrin et al., 1989). Twenty-five VIGALL EEG channels were selected from our high-density net according to the VIGALL manual (VIGALL 2.1 manual; https://research.uni-leipzig.de/vigall/). HEOG was recorded from two electrodes at the outer canthi of the right and left eye and VEOG was recorded from the infraorbital and supraorbital regions of the right eye, keeping it as close as possible to the VIGALL manual. The EEG with the applied VIGALL montage was down-sampled to 250 Hz and re-referenced

offline to an average reference. The data was cut into 1s epochs for visual artifact inspection such as movement and electrical artifacts and was further processed in Brain Vision Analyzer 2.0 (Brain Products GmbH, Glitching, Germany). Artefact segments were marked, but not removed to retain the full time series for each subject. A zero-phase digital IIR Butterworth bandpass filter with cut-off frequencies 0.5 and 70 Hz, and a 50 Hz notch filter was applied to the EEG-data after eye-movement artifacts correction (independent component analysis approach). Electrooculogram channels were bandpass filtered with cut-off frequencies below 0.01 and above 70 Hz to retain slow eye movements.

EEG-vigilance was assessed from the eye-closed resting EEG data by using the algorithm-based Vigilance Algorithm Leipzig (VIGALL 2.0, 11, 12). Slow eye movements (SEMs) criteria were set to 100 µV with a 6 second window length to detect any drowsiness in the recording (Jödicke et al., 2013; Santamaria and Chiappa, 1987). Each 1s epochs was automatically classified into the following arousal states, resulting in a vigilance time-course: stage 0 (highest arousal), A1, A2, A3, B1, B2/3, C (lowest arousal, sleep onset, classified visually by sleep grapho-elements from an experienced rater), according to the classification from (Bente, 1976; Roth, 1961; Santamaria and Chiappa, 1987). The arousal continuum was classified mainly based on the distribution of alpha cortical current density over four distinct regions of interests (ROIs): frontal, central, temporal and occipital. In a typical case, non-alpha activity with the absence of SEMs would first appear after closing eyes (stage 0). The alpha activity would then dominate gradually from occipital (A1) to central and frontal (A2), and to mainly centralized at frontal area (A3) along with the relaxation of the participant. Subsequently, alpha activity would disappear and replaced by low amplitude activity with SEM (B1) then dominate by delta and theta activity (B2/3) (for details refer to VIGALL 2.1 manual). Since no subject showed stage C segments and the prevalence of stages A2 and A3 have been quite low in previous studies, we followed the usual procedure of pooling two

A stages (A2/3), resulting in five different vigilance stages (0, A1, A2/3, B1, B2/3) and were assigned numerically with a range from 6 (stage 0) to 2 (stage B2/3). To align with previous studies, median vigilance of each 1 minute-block of eyes closed condition (in total of 6 blocks) was calculated for its ordinal-scale feature (Olbrich et al., 2016a). Further, the slope of the median vigilance was calculated from each 3 min eyes closed recording (vigilance slope). A positive slope indicated less decline toward sleep while a negative slope indicated more pronounced decrease of vigilance toward drowsiness. Percentages of the different vigilance stages were calculated after subtraction of the number of artefact segments.

Statistics

Percentages of each vigilance stage at each block (5 states × 6 blocks), median vigilance of each block (6 blocks) and vigilance slope (two separate eyes closed recordings from one session: 1st and 2nd recordings) were correspondingly subjected to repeated-measures analysis of variance (ANOVA) to determine whether there were temporal dynamic changes in EEG vigilance patterns > between pretreatment MDD and healthy controls, 2) in pretreatment vigilance between remitters and non-responders (NeuroPharm criteria); between responders and non-responders (iSPOT-D criteria); between remitters and non-remitters (iSPOT-D criteria), 3) between pretreatment and after 8 weeks of treatment (follow-up), and whether these differences were distinct in remitters and non-responders (NeuroPharm criteria); between responders and non-responders (iSPOT-D criteria); between remitters and non-remitters (iSPOT-D criteria). Age was included as covariate. Significant interactions were examined by analysis of simple effects. Bonferroni's correction was used for multiple comparisons and post hoc analyses. Degrees of freedom were corrected by Greenhouse-Geisser correction when necessary. In order to assess the discriminative power, the analysis of Receiver Operator Curve (ROC) was performed for any successful discriminant on the clinical outcome. Group differences in sex, age, educational scores, pretreatment generalized

anxiety disorder-10 score (GAD_{10}) (Bech et al., 2005), pretreatment HDRS scores and HDRS scores at week 8 were tested by simple *t* statistic or using the χ^2 statistic (sex). Two-sided *p* values were chosen for all tests. To investigate the vigilance effect on clinical outcome, a correlation analysis on HDRS scores (both $HDRS_6$ and $HDRS_{17}$) and vigilance slope was performed.

RESULTS

Sociodemographic characteristics

Patients with MDD and healthy controls did not differ in age and sex (all *p* values > .32, Table 1), but the healthy controls had a higher education score compared to patients (t (98) = -2.607, p = .011). Pretreatment HDRS score did not differ significantly between remitters and non-responders (NeuroPharm criteria) nor between responders and non-responders (iSPOT-D criteria) (*p* values > .23). No significant difference was found on GAD_{10} measures (*p* values > .05, Table 1)

Pretreatment MDD patients vs. Healthy Controls

A repeated ANOVA with age as covariate yielded a significant group effect of median slope (F (1, 123) = 4.59, p = .034, Table 2), indicating that prior to treatment MDD patients had less pronounced declines toward sleep stages in comparison to healthy controls in the first 3 min recording (-0.02 vs. -0.26, p = .011, Figure 2 left panel). No significant group effect was found for median EEG-vigilance or percentages of each stage at each block (*p* values > .05).

VIGALL as predictor for clinical outcome

There was no significant difference in median vigilance, percentages of each stage or vigilance slope between remitters and non-responders prior to treatment (NeuroPharm criteria, *p* values > .51). When using the iSPOT-D criteria, we observed a significant interaction between vigilance slope and group (F (1, 76) = 4.16, p = .045, Table 2). The analyses of simple effects showed that responders had more pronounced declines toward sleep stages compared to non-responders in the first 3 min recording (1st recording: -0.11 vs. 0.13, p = .042), but not the second 3 min recording (2nd recording: p = .70, Figure 2, right panel). Post-hoc analysis indicated that excluding patients with low serum concentrations slightly affected the results, but the trend remained (F (1, 73) = 3.86, p = .053; 1st recording: -0.11 vs. 0.14, p = .044). However, receiver operator characteristics yielded an area under the curve of only .60 (p = .12). No significant difference was found for median vigilance or percentages of each stage (p values > .32). The associations between $HDRS_6$ change and vigilance slope was r (80) = -.16 with p =.158 for the 1st and r (80) = .03, with p =.81 for the 2nd recording. We did not find any significant correlation when correlating $\Delta HDRS_{17}$ scores at week 8 and vigilance slope (1st recording: r (80) = -.12, p =.286; 2nd recording: r (80) = .05, p =.650).

Treatment effects on VIGALL parameters

By pooling all patients' data regardless of their treatment responses, we found a significant interaction between recording time (pretreatment vs. follow-up) and stage (F (4, 152) = 3.11, p = .026), demonstrating that MDD patients had a higher amount of low vigilance stage B1 (23% vs. 35%) at follow-up in comparison to pretreatment (Figure 3). Moreover, there was a trend of decreased median vigilance in MDD patients when comparing follow-up versus pretreatment (4.32 vs. 4.06, p = .08), suggested by the borderline significant effect of recording time (F (1, 190) = 3.24, p = .080). Post-hoc analysis after excluding patients with low serum concentrations did not

change the direction of the results (recording time × stage: F (4, 144) = 3.68, p = .012; median vigilance: F (1, 36) = 4.12, p = .049). No significant result was found for median slope (p > .79). We further investigated the treatment effects on patients with different clinical outcomes. There was a significant interaction between recording time (pretreatment and treatment) and vigilance stages on remitters (i.e. pretreatment remitters vs. follow-up remitters: F (4, 40) = 3.16, p = .049) but not non-responders (NeuroPharm criteria, p = .625, Figure 4). Analyses of simple effects showed that remitters had a higher percentage of stage B1 segments (14% vs. 35%, p = .008) after 8 weeks of treatment. A similar effect was observed when examining the iSPOT-D criteria (F (4, 96) = 4.84, p = .002), indicating that responders had a higher amount of vigilance stage B1 (23% vs. 40%, p = .002, Figure 4) after treatment but not non-responders (p = .58). Remitters (iSPOT-D criteria) also showed a trend towards a larger amount of stage B1 (21% vs. 39%, p = .026) after treatment, suggested by a borderline significant interaction between recording time and stage (F (4, 52) = 2.72, p = .061, Figure 4). Furthermore, a significant main recording time effect of median vigilance (F (1, 24) = 7.17, p = .013) was observed for responders (iSPOT-D criteria), showing that responders had lower median vigilance at follow-up (4.41 vs. 3.93, p = .013) in comparison to pretreatment. Post-hoc analysis after excluding patients with low serum concentrations did not change the results.

Exploratory analysis: Normalization effect on VIGALL parameters after 8 weeks of treatment

Since treatment effects for vigilance stage B1 were observed for all MDD patients and patients with good treatment response, we decided to further examine if there was any normalization effect towards patterns of EEG-vigilance of healthy controls in these patients. T statistics for recording time and comparing patients with healthy controls at stage B1 revealed that pretreatment MDD had a lower amount of vigilance stage B1 compared to healthy controls (t (72) = -2.09, p = .040, Figure

4). This difference had disappeared after 8 weeks of treatment (treatment MDD vs. HC: t (72) = -0.11, p = .910). When assessing normalization effects on patients with good treatment response, we found that remitters had a lower percentage of stage B1 compared to healthy controls (NeuroPharm criteria: t (44) = -2.38, p = .022, Figure 4) while this difference was not present when assessing follow-up remitters and healthy controls (p = .972). Similarly, responders (t (58) = -1.78, p = .081) and remitters (t (47) = -1.69, p = .098) showed a trend of lower stage B1 at the pretreatment visit (iSPOT-D criteria, Figure 4). The trend disappeared after 8 weeks of treatment (responders vs. healthy controls: p = .55; remitters vs. healthy controls: p = .66).

DISCUSSION

This study on EEG-vigilance based biomarkers replicated in an independent sample of MDD patients: the discriminative feature of the attenuated wakefulness regulation toward sleep onset during rest for depressed patients in comparison to healthy subjects and the distinction between responders and non-responders to SSRI/SNRI treatment. Although EEG research recently showed the possible usage of EEG biomakers in depression (Arns et al., 2016; Olbrich et al., 2016a; Pizzagalli et al., 2018; Rolle et al., 2020b; Wu et al., 2020), many limitations restrict the clinical usage up to today (Widge et al., 2018). One of the main limitations was the missing replication in independent datasets of any of the markers (Widge et al., 2018). Hence these findings and replications might help to guide the way into clinical practise.

EEG-vigilance in Major Depression

The Vigilance framework (Hegerl and Hensch, 2014) states that patients suffering from MDD show a more rigid wakefulness regulation during rest. That means that their functional brain states, as

assessed with EEG, stay at high wakefulness levels during the resting state, as shown before (Hegerl et al., 2012; Olbrich et al., 2012). This observation may reflect the difficulties of depressed patients to fall asleep and the high inner tensions. At the behavioral level, patients may try to counteract the hyperstable vigilance by avoiding any arousing activities (Hegerl and Hensch, 2014). The EEG Vigilance measurements allow to assess the timecourses of this wakefulness regulation. Previous studies showed a hyperstable vigilance regulation in unmedicated patients, having more high vigilance stages than heatlhy controls (Hegerl et al., 2012; Olbrich et al., 2012; Schmidt et al., 2017). The data from the NeuroPharm study independently replicates that there is a significant difference between patients and controls. Although this is a replication of the discriminative feature, the presented study only used a relatively short EEG resting state segment in comparison to previous studies, thus proving that the differences can be assessed already with a 3 min resting state paradigm.

EEG vigilance and treatment response

So far, two studies have investigated the predictive value of EEG-vigilance regulation in MDD for pharmacological antidepressant treatment (Olbrich et al., 2016a; Schmidt et al., 2017). In a large cohort of 599 patients, one study found an association between the slope of EEG vigilance, i.e. the change of vigilance over time, and the response to SSRI treatment using a two-minute resting state EEG (Olbrich et al., 2016a). In a smaller cohort of 65 patients, hyperstable EEG vigilance regulation was associated with a better response to different types of antidepressive drug when using a much longer EEG vigilance resting state session of 15 min. The response in the Schmidt et al., (2017) study was defined as a decline of HDRS of 50% after 28 days whereas the NeuroPharm study assessed clinical response after 56 days. In the present study, we replicate the findings from the iSPOT-D study (Olbrich et al., 2016a), showing that a faster decline towards lower vigilance

stages are linked with a significantly better outcome following SSRI treatment. Importantly, however, the replication could only be found when the iSPOT-D response criteria were applied, and not with the NeuroPharm criteria. This highlights the importance of using standardized outcome measures when comparing studies. As stated in Widge et al., (2018) replication needs to stick to the used methods from the studies to be replicated. Since this is the first independent replication of treatment prediction using EEG vigilance measures in a retrospective analysis, the next step would be a prospective study to show the efficacy of an EEG-informed treatment choice in a randomiuzed controlled design.

Treatment effect on EEG vigilance

The findings from the present study showed that remitters (NeuroPharm criteria) and responders (iSPOT-D criteria) after 8 weeks of treatment revealed lowered vigilance and increased downregulation of EEG-vigilance. Further, when comparing remitters and responders to EEG-vigilance profiles of healthy subjects, the wakefulness regulation seemed to normalize. This was not the case in non-remitters or non-responders (for neither the NeuroPharm nor iSPOT-D criterion). It is noteworthy that the normalization of EEG wakefulness parameters was mainly restricted to the stage B1, a brain state with desynchronized EEG activity and thus low amplitude in association with horizontal slow eye movement (Santamaria and Chiappa, 1987). This stage reflects the transition between the alpha-dominated period after closing the eyes and the occurrence of slow wave activity just before sleep onset. The increasing amount of B1 stages in responders after 8 weeks of treatment may reflect the normalization of the wakefulness regulation toward patterns of healthy subjects. Thus, it might be hypothesized that stage B1 is a gate-keeping brain state: Patients suffering from MDD might have difficulties to pass through these desynchronized stages to achieve recreational rest.

In summary, our findings support that wakefulness regulation in patients suffering from MDD shows hyperstable patterns with less drifts toward lowered vigilance during rest. The degree of rigidity further seems to be associated with the response to pharmacological treatment: Patients with the most hyperstable vigilance regulation seem to profit less from this kind of treatment. In addition, the wakefulness regulation profiles of responders and remitters seems to normalize after successful treatment, while this is not the case in non-responders or non-remitters. In a clinical context, these observations could be useful to 1) initiate other treatments than SSRI treatment in subjects with hyperrigid EEG vigilance regulation 2) augment treatment early in the course of intervention in patients with hyperrigid EEG-vigilance regulation and 3) provide patients with pronounced declines of EEG-vigilance regulation a treatment with SSRIs in a shared decision-making process. 4) stratify patients in future drug development programs.

Limitations

The NeuroPharm study is a naturalistic study without any placebo arm, meaning that eventual drug effects on vigilance measures cannot be assessed. This was, however, been done in preclinical studies that SSRI treatment decrease the firing rate of the locus coeruleus (LC) and thus decrease vigilance stages (Hegerl and Hensch, 2014). Another potential limitation is that the recording of the EEG-resting state was limited to 2x3 min; in Schmidt et al., (2017) recordings were longer. However, 2x3 min recording seems to be sufficient since we could replicate the differential properties of EEG wakefulness regulation for healthy controls and patients.

Conclusions

Our findings based on EEG data from the NeuroPharm trial independently confirm the previously reported usefulness of EEG-vigilance biomarkers for predicting outcome to SSRI treatment in MDD. Whereas, this is an important step towards possible clinical applications. Future research should include 1) identification of treatment approaches that work better for patients with a hyperrigid EEG-vigilance regulation and 2) a larger prospective randomized controlled trial with a treatment as usual group and an EEG-vigilance regulation informed treatment group.

Acknowledgments: We thank all the participants that included in the study. Special thanks to Gerda Thomsen, Svitlana Olsen and Lone Freyr for the recruitment of healthy controls and the preparation of the blood samples.

Funding and Disclosures: Collection of the data included in the study was supported by Innovationsfonden, Denmark (4108-00004B) and by the Lundbeck Foundation (Cimbi: R90-A7722). C.I. was supported by the Innovationsfonden, Denmark (5189-00087A). M.G. was supported by the Lundbeck Foundation (R181-2014-3586) and the Elsass foundation (18-3-0147). V.D. was supported by Augustinus Foundation (16-0058) and Rigshospitalet's Research (R149-A6325). M.J. was supported by Savværksejer Jeppe Juhl og hustru Ovita Juhls Mindelegat. V.F. was supported by Augustinus Foundation (16-0058). G.K. was supported by Innovationsfonden, Denmark (4108-00004B).

All authors declare no conflict of interest.

REFERENCES

Arns, M., Bruder, G., Hegerl, U., Spooner, C., Palmer, D.M., Etkin, A., Fallahpour, K., Gatt, J.M., Hirshberg, L., Gordon, E., 2016. EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized iSPOT-D study. Clin. Neurophysiol. 127, 509–519. https://doi.org/10.1016/j.clinph.2015.05.032

Bech, P., Boyer, P., Germain, J.M., Padmanabhan, K., Haudiquet, V., Pitrosky, B., Tourian, K.A., 2010. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. Pharmacopsychiatry. https://doi.org/10.1055/s-0030-1263173

Bech, P., Kajdasz, D.K., Porsdal, V., 2006. Dose-response relationship of duloxetine in placebo-controlled clinical trials in patients with major depressive disorder. Psychopharmacology (Berl). https://doi.org/10.1007/s00213-006-0505-1

Bech, P., M., L., M., U., 2005. An inventory for the measurement of generalized anxiety distress symptoms, the GAD-10 Inventory. Acta Psychiatr Belg 105, 111–118.

Bente, D., 1976. EEG aspects of waking sleep behaviour and the chronophysiology of endogenous depressions. Arzneimittel-Forschung/Drug Res.

Grzenda, A., Widge, A.S., 2020. Electroencephalographic Biomarkers for Predicting Antidepressant Response: New Methods, Old Questions. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2019.3749

Guenther, T., Schönknecht, P., Becker, G., Olbrich, S., Sander, C., Hesse, S., Meyer, P.M., Luthardt, J., Hegerl, U., Sabri, O., 2011. Impact of EEG-vigilance on brain glucose uptake measured with [18F]FDG and PET in patients with depressive episode or mild cognitive impairment. Neuroimage. https://doi.org/10.1016/j.neuroimage.2011.01.059

Hegerl, U., Hensch, T., 2014. The vigilance regulation model of affective disorders and ADHD. Neurosci. Biobehav. Rev. 44, 45–57. https://doi.org/10.1016/j.neubiorev.2012.10.008

Hegerl, U., Wilk, K., Olbrich, S., Schoenknecht, P., Sander, C., 2012. Hyperstable regulation of vigilance in patients with major depressive disorder. World J. Biol. Psychiatry. https://doi.org/10.3109/15622975.2011.579164

Jawinski, P., Sander, C., Mauche, N., Spada, J., Huang, J., Schmidt, A., Häntzsch, M., Burkhardt, R., Scholz, M., Hegerl, U., Hensch, T., 2015. Brain Arousal Regulation in Carriers of Bipolar Disorder Risk Alleles. Neuropsychobiology. https://doi.org/10.1159/000437438

Jödicke, J., Olbrich, S., Sander, C., Minkwitz, J., Chittka, T., Himmerich, H., Hegerl, U., 2013. Separation of low-voltage EEG-activity during mental activation from that duringtransition to drowsiness. Brain Topogr. https://doi.org/10.1007/s10548-013-0287-9

Kennis, M., Gerritsen, L., van Dalen, M., Williams, A., Cuijpers, P., Bockting, C., 2020. Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis. Mol. Psychiatry. https://doi.org/10.1038/s41380-019-0585-z

Köhler-Forsberg, K., Jørgensen, A., Dam, V., Stenbæk, D., Fisher, P., Ip, C.-T., Melanie, G., Poulsen, H., Giraldi, A., Ozenne, B., Jørgensen, M., Knudsen, G., Frokjaer, V., 2020. Predicting treatment outcome in Major Depressive Disorder using serotonin 4 receptor PET brain imaging, functional MRI, cognitive-, EEG-based and peripheral biomarkers: a NeuroPharm open label clinical trial protocol. Front. Psychiatry.

Nutt, D.J., Wilson, S., Paterson, L., 2008. Sleep disorders as core symptoms of depression. Dialogues Clin. Neurosci.

Olbrich, S., Arns, M., 2013. EEG biomarkers in major depressive disorder: Discriminative power and prediction of treatment response. Int. Rev. Psychiatry. https://doi.org/10.3109/09540261.2013.816269

Olbrich, S., Conradi, J., 2016. Future of clinical EEG in psychiatric disorders: Shifting the focus from diagnosis to the choice of optimal treatment. Clin. Neurophysiol. https://doi.org/10.1016/j.clinph.2015.06.018

Olbrich, S., Fischer, M.M., Sander, C., Hegerl, U., Wirtz, H., Bosse-Henck, A., 2015. Objective markers for sleep propensity: Comparison between the Multiple Sleep Latency Test and the

Vigilance Algorithm Leipzig. J. Sleep Res. https://doi.org/10.1111/jsr.12290
Olbrich, S., Mulert, C., Karch, S., Trenner, M., Leicht, G., Pogarell, O., Hegerl, U., 2009. EEG-vigilance and BOLD effect during simultaneous EEG/fMRI measurement. Neuroimage. https://doi.org/10.1016/j.neuroimage.2008.11.014
Olbrich, S., Sander, C., Minkwitz, J., Chittka, T., Mergl, R., Hegerl, U., Himmerich, H., 2012. EEG vigilance regulation patterns and their discriminative power to separate patients with major depression from healthy controls. Neuropsychobiology. https://doi.org/10.1159/000337000
Olbrich, S., Tränkner, A., Surova, G., Gevirtz, R., Gordon, E., Hegerl, U., Arns, M., 2016a. CNS- and ANS-arousal predict response to antidepressant medication: Findings from the randomized iSPOT-D study. J. Psychiatr. Res. https://doi.org/10.1016/j.jpsychires.2015.12.001
Olbrich, S., Tränkner, A., Surova, G., Gevirtz, R., Gordon, E., Hegerl, U., Arns, M., 2016b. CNS- and ANS-arousal predict response to antidepressant medication: Findings from the randomized iSPOT-D study. J. Psychiatr. Res. 73, 108–115. https://doi.org/10.1016/j.jpsychires.2015.12.001
Østergaard, S.D., Bech, P., Miskowiak, K.W., 2016. Fewer study participants needed to demonstrate superior antidepressant efficacy when using the Hamilton melancholia subscale (HAM-D6) as outcome measure. J. Affect. Disord. 190, 842–845. https://doi.org/10.1016/j.jad.2014.10.047
Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187. https://doi.org/https://doi.org/10.1016/0013-4694(89)90180-6
Pizzagalli, D.A., Webb, C.A., Dillon, D.G., Tenke, C.E., Kayser, J., Goer, F., Fava, M., McGrath, P., Weissman, M., Parsey, R., Adams, P., Trombello, J., Cooper, C., Deldin, P., Oquendo, M.A., McInnis, M.G., Carmody, T., Bruder, G., Trivedi, M.H., 2018. Pretreatment rostral anterior cingulate cortex theta activity in relation to symptom improvement in depression: A randomized clinical trial. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2018.0252
Rolle, C.E., Fonzo, G.A., Wu, W., Toll, R., Jha, M.K., Cooper, C., Chin-Fatt, C., Pizzagalli, D.A., Trombello, J.M., Deckersbach, T., Fava, M., Weissman, M.M., Trivedi, M.H., Etkin, A., 2020a. Cortical Connectivity Moderators of Antidepressant vs Placebo Treatment Response in Major Depressive Disorder. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2019.3867
Rolle, C.E., Fonzo, G.A., Wu, W., Toll, R., Jha, M.K., Cooper, C., Chin-Fatt, C., Pizzagalli, D.A., Trombello, J.M., Deckersbach, T., Fava, M., Weissman, M.M., Trivedi, M.H., Etkin, A., 2020b. Cortical Connectivity Moderators of Antidepressant vs Placebo Treatment Response in Major Depressive Disorder: Secondary Analysis of a Randomized Clinical Trial. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2019.3867
Roth, B., 1961. The clinical and theoretical importance of EEG rhythms corresponding to states of lowered vigilance. Electroencephalogr. Clin. Neurophysiol. https://doi.org/10.1016/0013-4694(61)90008-6
Santamaria, J., Chiappa, K.H., 1987. The EEG of drowsiness in normal adults. J. Clin. Neurophysiol. https://doi.org/10.1097/00004691-198710000-00002
Schmidt, F.M., Sander, C., Dietz, M.E., Nowak, C., Schröder, T., Mergl, R., Schönknecht, P., Himmerich, H., Hegerl, U., 2017. Brain arousal regulation as response predictor for antidepressant therapy in major depression. Sci. Rep. https://doi.org/10.1038/srep45187

Seifritz, E., 2001. Contribution of sleep physiology to depressive pathophysiology. Neuropsychopharmacology. https://doi.org/10.1016/S0893-133X(01)00319-0
Sheehan, D. V, Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry.
Steiger, A., Pawlowski, M., 2019. Depression and sleep. Int. J. Mol. Sci. https://doi.org/10.3390/ijms20030607
Thase, M.E., 2006. Depression and sleep: Pathophysiology and treatment. Dialogues Clin. Neurosci.
Widge, A.S., Bilge, M.T., Montana, R., Chang, W., Rodriguez, C.I., Deckersbach, T., Carpenter, L.L., Kalin, N.H., Nemeroff, C.B., 2018. Electroencephalographic biomarkers for treatment response prediction in major depressive illness: A meta-analysis. Am. J. Psychiatry 176, 44–56. https://doi.org/10.1176/appi.ajp.2018.17121358
Williams, L.M., Rush, A.J., Koslow, S.H., Wisniewski, S.R., Cooper, N.J., Nemeroff, C.B., Schatzberg, A.F., Gordon, E., 2011. International Study to Predict Optimized Treatment for Depression (iSPOT-D), a randomized clinical trial: Rationale and protocol. Trials. https://doi.org/10.1186/1745-6215-12-4
Wu, W., Zhang, Y., Jiang, J., Lucas, M. V., Fonzo, G.A., Rolle, C.E., Cooper, C., Chin-Fatt, C., Krepel, N., Cornelssen, C.A., Wright, R., Toll, R.T., Trivedi, H.M., Monuszko, K., Caudle, T.L., Sarhadi, K., Jha, M.K., Trombello, J.M., Deckersbach, T., Adams, P., McGrath, P.J., Weissman, M.M., Fava, M., Pizzagalli, D.A., Arns, M., Trivedi, M.H., Etkin, A., 2020. An electroencephalographic signature predicts antidepressant response in major depression. Nat. Biotechnol. https://doi.org/10.1038/s41587-019-0397-3

Table and Figure Legends

Table 1: Sociodemographic characteristics for pretreatment visit and clinical outcome at week 8 according to either NeuroPharm or iSPOT-D criteria

	Pretreatment visit		**Clinical outcome at week 8**			
			NeuroPharm[1]		iSPOT-D [2]	
	Healthy controls	MDD	Remitters	Non-responders	Responders[4]	Non-responders[4]
N	35	91	21	15	44	35
Sex (M/F)	10/25	25/66	11/10	4/11	16/28	6/29
Age (Mean±SD)	29.0±9.7	27.4±8.3	29.4±9.7	25.7±9.3	28.5±8.8	25.5±7.2
Education	16.0±1.4	14.9±2.2[3]	15.8±1.3	14.8±2.3	15.2±1.9	14.6±2.5
GAD$_{10}$		22.9±9.7	22.3±8.5	22.1±8.6	23.4±7.9	22.1±10.3
Pretreatment HDRS		12.4±1.7[1] (22.9±3.4[2])	11.8±1.6	11.4±1.8	23.0±3.5	22.1±3.0
Week 8 HDRS			2.4±1.3	10.1±2.4[5]	7.0±3.3	17.5±4.5[5]

Notes: [1] HDRS$_6$ scores are shown in both pretreatment and week 8 HDRS$_6$ scores

[2] HDRS$_{17}$ scores are shown in both pretreatment and week 8 HDRS$_{17}$ scores

[3] Healthy controls had a significant higher education score compared to MDD (t (98) = -2.607, p = .011)

[4]Responders were defined by at least 50% improvement of depressive symptoms assessed by HDRS$_{17}$ score. HDRS$_{17}$ scores are shown in both pretreatment and week 8 HDRS$_{17}$ scores

[5]Between group comparisons showed that remitters had significant lower HDRS$_6$ score at week 8 compared to non-responders (t (34) = -12.71, p < .001); responders had significant lower HDRS$_{17}$ score at week 8 compared to non-responders (t (77) = -11.90, p < .001). No significant was found for other demographic characteristics (p values >.05).

Abbreviation: GAD$_{10}$, generalized anxiety disorder-10 score; HDRS, Hamilton Depression Rating Scale.

Table 2: EEG-vigilance parameters at pretreatment visit and at week 8

	Pretreatment visit		**Clinical outcome at week 8**			
			NeuroPharm[1]		iSPOT-D[2]	
	Healthy controls	MDD	Remitters	Non-responders	Responders	Non-responders
Stage 0 (%, Mean±SD)	13.7±3.3	15.3±2.1	14.2±5.1	22.6±6.0	16.2±3.4	15.3±3.6
Stage A1	32.6±5.7	36.3±3.5	36.1±7.7	35.4±9.2	35.2±5.3	37.0±5.7
Stage A23	10.1±3.2	14.1±2.0	14.1±3.9	6.1±4.6	13.0±3.3	17.1±3.5
Stage B1	34.8±4.4	26.4±2.7	23.2±5.5	23.6±6.6	27.7±3.9	22.0±4.1
Stage B23	8.8±2.4	7.8±1.5	12.4±4.8	12.4±5.7	7.9±2.4	8.5±2.6
Median vigilance (Mean±SD)	4.07±0.15	4.27±0.93	4.24±0.25	4.31±0.30	4.28±0.15	4.30±0.16
Vigilance slope at 1st recording (Mean±SD)	-0.26±0.43	-0.02±0.50*	-0.09±0.39	0.03±0.68	-0.11±0.48	0.13±0.54*
Vigilance slope at 2nd recording (Mean±SD)	-0.01±0.51	0.05±0.43	0.10±0.24	0.16±0.56	0.10±0.38	0.06±0.42

Notes: * $p < .05$, age was included as covariate in all models

Figure 1: The flowchart of the study design. See the dashed frame for the definition of treatment responses in NeuroPharm. Medication-free patients diagnosed with MDD were included and allocated to baseline intervention. The treatment responses were measured as percentage change in HDRS6 score from baseline to week 4 or 8 (ΔHDRS6). Three categorical treatment outcomes were defined: Remitters included patients with ΔHDRS6 > 50% at week 4 and < 5 points on the HDRS6 scale at week 8; non-responders included patients with ΔHDRS6 < 25% at week 4 and ΔHDRS6 < 50% at week 8 were considered non-responders. Patients who did not fulfill either remitter or non-responder criteria were classified as intermediate responders.

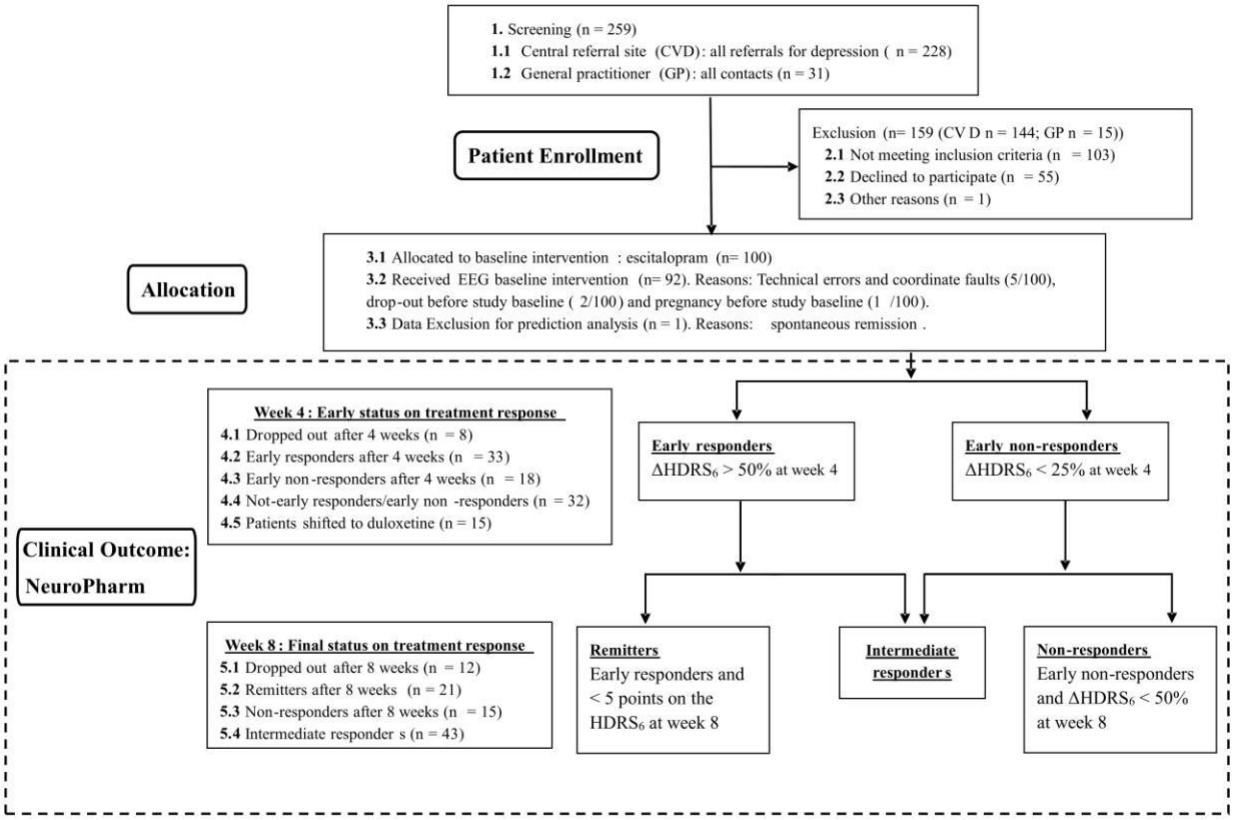


Figure 2: Mean median vigilance and the corresponding error bar were depicted in the figure. **Left panel:** Vigilance slope between pretreatment MDD patients and healthy controls. Pretreatment MDD patients had less declines on vigilance slope compared to healthy controls. **Right panel:** Responders defined by iSPOT-D criteria had faster declines toward sleep stages compared to non-responders in the first recording, but not the second recording.

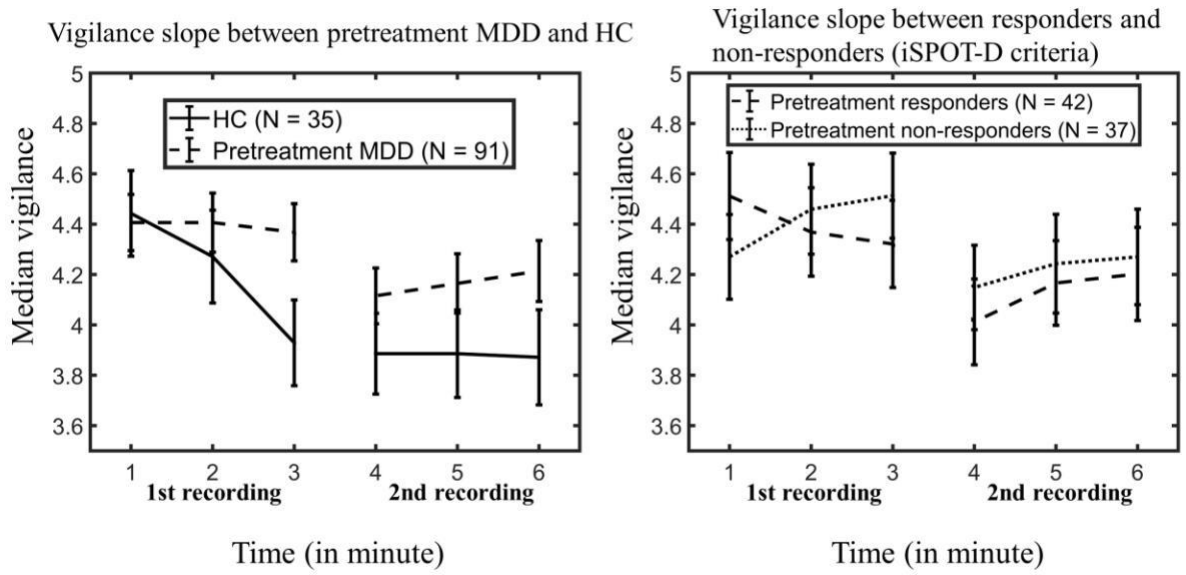


Figure 3: Treatment effects on different vigilance stages on MDD patients. Mean percentage and the corresponding error bar were depicted in the figure. After 8 weeks of treatment, MDD patients had a higher percentage of stage B1 compared to pretreatment visit. Moreover, depressed patients at follow-up showed a normalization effect towards the vigilance pattern of healthy controls (HC) at stage B1.

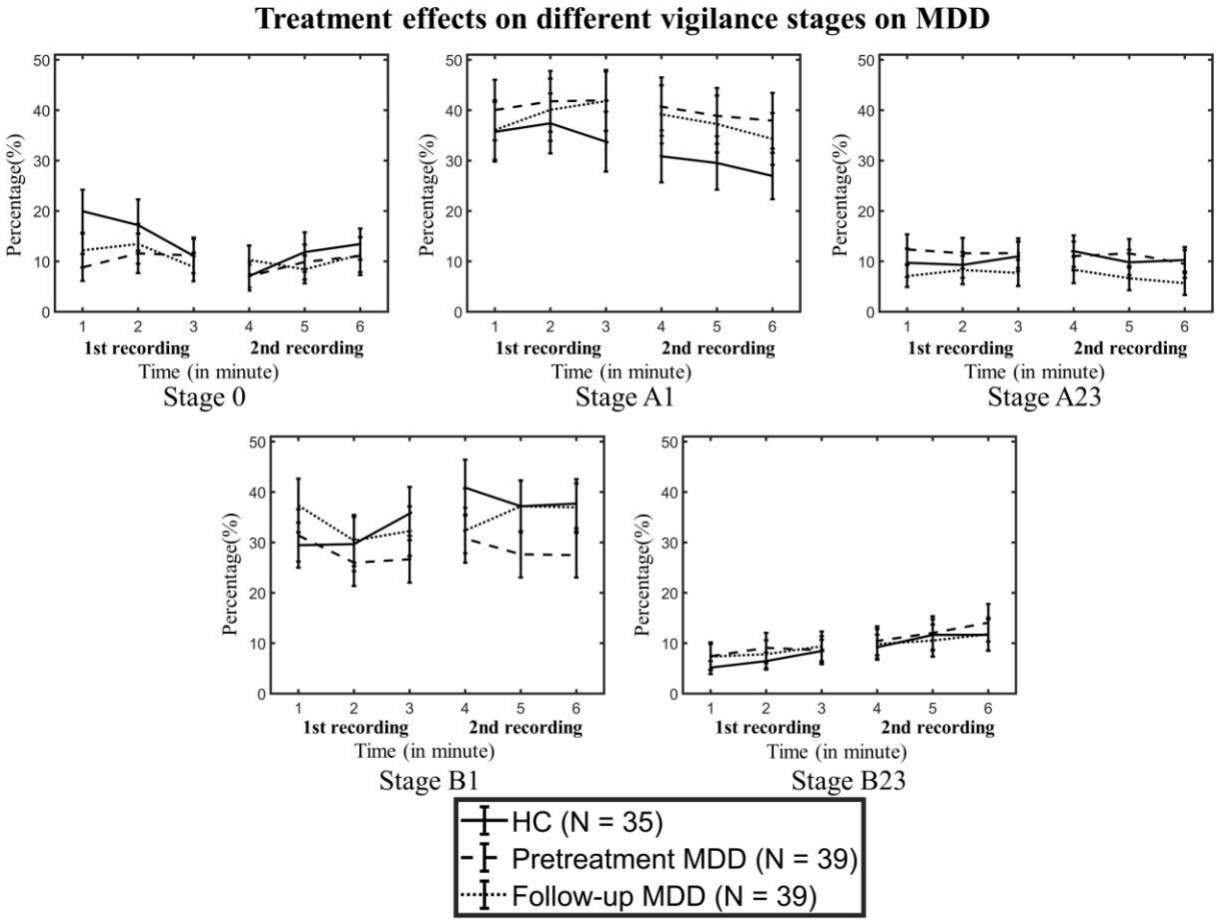
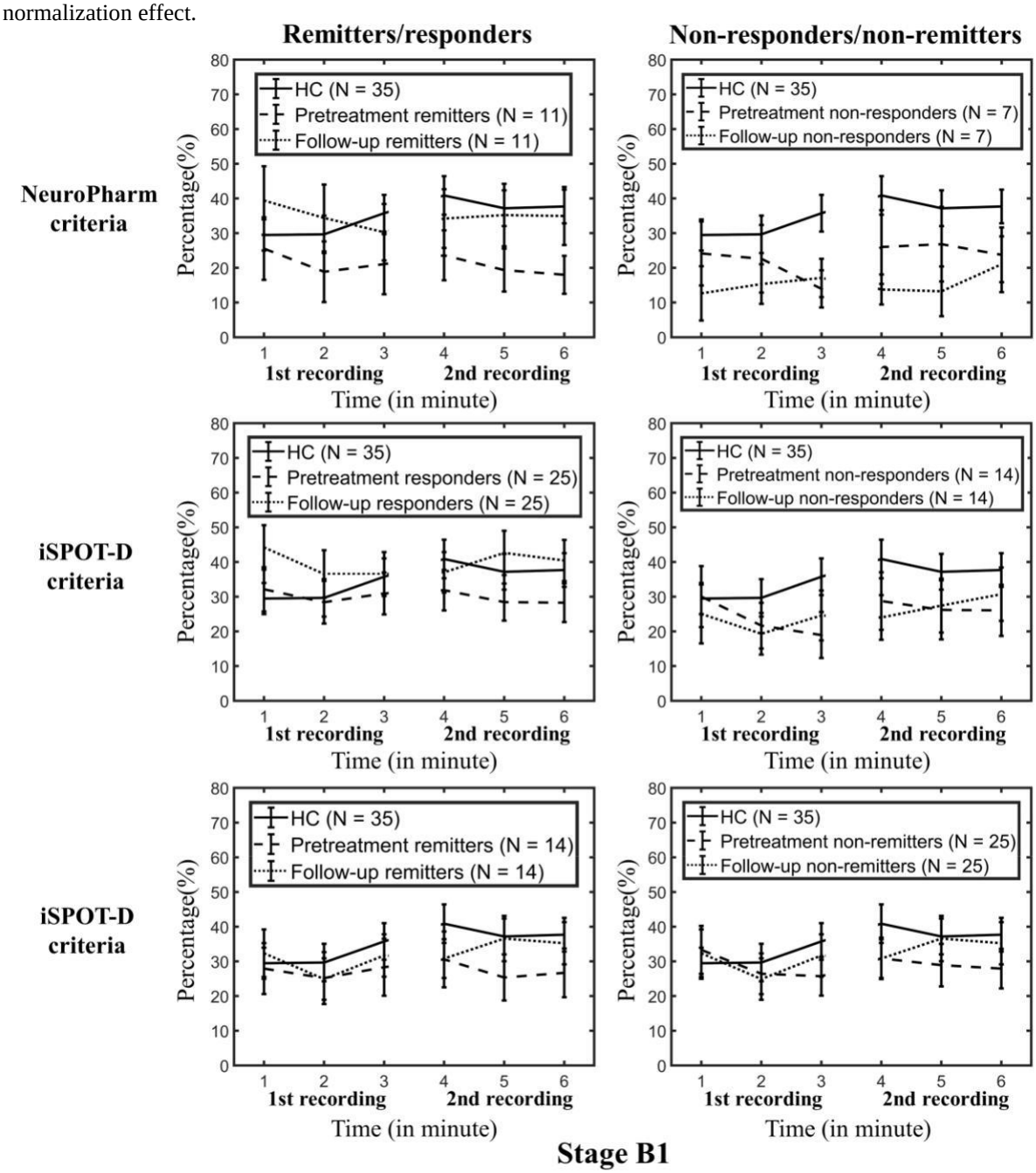


Figure 4: Treatment effects on MDD patients with different clinical outcomes at stage B1. Mean percentage and the corresponding error bar were depicted in the figure. Patients with good clinical responses (remitters/responders) showed a higher percentage of stage B1 after 8 weeks of treatment, whilst, this was not found on non-responders and non-remitters. The data of healthy controls (HC) were shown for illustrating the normalization effect.



GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

PHD-THESIS
DECLARATION OF CO-AUTHORSHIP

The declaration is for PhD students and must be completed for each conjointly authored article. Please note that if a manuscript or published paper has ten or less co-authors, all co-authors must sign the declaration of co-authorship. If it has more than ten co-authors, declarations of co-authorship from the corresponding author(s), the senior author and the principal supervisor (if relevant) are a minimum requirement.

1. Declaration by	
Name of PhD student	Cheng-Teng Ip
E-mail	chengtengip@nru.dk
Name of principal supervisor	Gitte Moos Knudsen
Title of the PhD thesis	Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depression Disorder

2. The declaration applies to the following article	
Title of article	Pre-intervention test-retest reliability of EEG and ERP over four recording intervals
Article status	
Published ☒ Date: Dec 2018	Accepted for publication ☐ Date:
Manuscript submitted ☐ Date:	Manuscript not submitted ☐
If the article is published or accepted for publication, please state the name of journal, year, volume, page and DOI (if you have the information).	doi:10.1016/j.ijpsycho.2018.09.007

3. The PhD student's contribution to the article *(please use the scale A-F as benchmark)* Benchmark scale of the PhD-student's contribution to the article **A.** Has essentially done all the work (> 90 %) **B.** Has done most of the work (60-90 %) **C.** Has contributed considerably (30-60 %) **D.** Has contributed (10-30 %) **E.** No or little contribution (<10 %) **F.** Not relevant	**A, B, C, D, E, F**
1. Formulation/identification of the scientific problem	D
2. Development of the key methods	E
3. Planning of the experiments and methodology design and development	F
4. Conducting the experimental work/clinical studies/data collection/obtaining access to data	F
5. Conducting the analysis of data	A
6. Interpretation of the results	A
7. Writing of the first draft of the manuscript	A
8. Finalisation of the manuscript and submission	A
Provide a short description of the PhD student´s specific contribution to the article.[i] The PhD student analyzed data and interpreted the results. She wrote the frist draft of the manuscript and revised it after co-authors' comments	

4. Material from another thesis / dissertation[ii]	
Does the article contain work which has also formed part of another thesis, e.g. master's thesis, PhD thesis or doctoral dissertation (the PhD student's or another person's)?	Yes: ☒ No: ☐
If yes, please state name of the author and title of thesis / dissertation.	Cheng-Teng Ip: Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depression Disorder
If the article is part of another author's academic degree, please describe the PhD student's and the author's contributions to the article so that the individual contributions are clearly distinguishable from one another.	-

5. Signatures of the co-authors[iii]

	Date	Name	Title	Signature
1.	2 / 2	Melanie Ganz	PhD	Melanie Ganz-Benjamin
2.		Brice Ozenne	PhD	
3.	10/03/20	Lasse B. Sluth	PhD	
4.	04/08-2020	Mikkel Gram	PhD	Mikkel Gram
5.	27/07/20	Geoffrey Viardot	PhD	Viardot
6.	27/07/20	Philippe l'Hostis	PhD	
7.	26/07/20	Philippe Danjou	MD, PhD	
8.	Oct 23 2020	Gitte M. Knudsen	MD, DMSc	
9.	13 OCT 2020	Søren R. Christensen	PhD	
10.				

6. Signature of the principal supervisor

I solemnly declare that the information provided in this declaration is accurate to the best of my knowledge.
Date: Oct 23. 2020
Principal supervisor:

7. Signature of the PhD student

I solemnly declare that the information provided in this declaration is accurate to the best of my knowledge.
Date: 23-10-2020
PhD student: Cheng Teng Ip

Please learn more about responsible conduct of research on the Faculty of Health and Medical Sciences' website.

[i] This can be supplemented with an additional letter if needed.
[ii] Please see Ministerial Order on the PhD Programme at the Universities and Certain Higher Artistic Educational Institutions (PhD Order) § 12 (4):
"Any articles included in the thesis may be written in cooperation with others, provided that each of the co-authors submits a written declaration stating the PhD student's or the author's contribution to the work."
[iii] If more signatures are needed please add an extra sheet.

GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

PHD-THESIS
DECLARATION OF CO-AUTHORSHIP

The declaration is for PhD students and must be completed for each conjointly authored article. Please note that if a manuscript or published paper has ten or less co-authors, all co-authors must sign the declaration of co-authorship. If it has more than ten co-authors, declarations of co-authorship from the corresponding author(s), the senior author and the principal supervisor (if relevant) are a minimum requirement.

1. Declaration by	
Name of PhD student	Cheng-Teng Ip
E-mail	chengtengip@nru.dk
Name of principal supervisor	Gitte Moos Knudsen
Title of the PhD thesis	Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depression Disorder

2. The declaration applies to the following article	
Title of article	Pretreatment qEEG biomarkers for predicting pharmacological treatment outcome in Major Depressive Disorder: Independent validation from the NeuroPharm study
Article status	
Published ☐ Date:	Accepted for publication ☐ Date:
Manuscript submitted ☒ Date: 24 June 2020	Manuscript not submitted ☐
If the article is published or accepted for publication, please state the name of journal, year, volume, page and DOI (if you have the information).	

3. The PhD student's contribution to the article *(please use the scale A-F as benchmark)* Benchmark scale of the PhD-student's contribution to the article **A.** Has essentially done all the work (> 90 %) **B.** Has done most of the work (60-90 %) **C.** Has contributed considerably (30-60 %) **D.** Has contributed (10-30 %) **E.** No or little contribution (<10 %) **F.** Not relevant	**A, B, C, D, E, F**
1. Formulation/identification of the scientific problem	C
2. Development of the key methods	A
3. Planning of the experiments and methodology design and development	B
4. Conducting the experimental work/clinical studies/data collection/obtaining access to data	A
5. Conducting the analysis of data	A
6. Interpretation of the results	B
7. Writing of the first draft of the manuscript	A
8. Finalisation of the manuscript and submission	A
Provide a short description of the PhD student´s specific contribution to the article.[1] The PhD student developed the testing EEG package, acquired neurophysiological data, analyzed data and interpreted the resutls. She wrote the frist draft of the manuscript and revised it after co-authors' comments.	

4. Material from another thesis / dissertation[ii]	
Does the article contain work which has also formed part of another thesis, e.g. master's thesis, PhD thesis or doctoral dissertation (the PhD student's or another person's)?	Yes: ☒ No: ☐
If yes, please state name of the author and title of thesis / dissertation.	Cheng-Teng Ip: Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depression Disorder
If the article is part of another author's academic degree, please describe the PhD student´s and the author´s contributions to the article so that the individual contributions are clearly distinguishable from one another.	

5. Signatures of the co-authors[iii]

	Date	Name	Title	Signature
1.	Oct 26th 2020	Sebastian Olbrich	MD, PhD	
2.	25/10/20	Melanie Ganz	PhD	
3.	27/07/20	Brice Ozenne	PhD	
4.	13 Oct 2020	Søren R. Christensen	PhD	
5.	Oct 23 2020	Gitte M. Knudsen	MD, DMSc	
6.	Oct. 26th 2020			
7.				
8.				
9.				
10.				

6. Signature of the principal supervisor

I solemnly declare that the information provided in this declaration is accurate to the best of my knowledge.
Date: Oct 23, 2020
Principal supervisor:

7. Signature of the PhD student

I solemnly declare that the information provided in this declaration is accurate to the best of my knowledge.
Date: 23-10-2020
PhD student: Cheng-Teng Ip

Please learn more about responsible conduct of research on the Faculty of Health and Medical Sciences' website.

7

[i] This can be supplemented with an additional letter if needed.
[ii] Please see Ministerial Order on the PhD Programme at the Universities and Certain Higher Artistic Educational Institutions (PhD Order) § 12 (4):
"Any articles included in the thesis may be written in cooperation with others, provided that each of the co-authors submits a written declaration stating the PhD student's or the author's contribution to the work."
[iii] If more signatures are needed please add an extra sheet.

GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

PHD-THESIS
DECLARATION OF CO-AUTHORSHIP

The declaration is for PhD students and must be completed for each conjointly authored article. Please note that if a manuscript or published paper has ten or less co-authors, all co-authors must sign the declaration of co-authorship. If it has more than ten co-authors, declarations of co-authorship from the corresponding author(s), the senior author and the principal supervisor (if relevant) are a minimum requirement.

1. Declaration by	
Name of PhD student	Cheng-Teng Ip
E-mail	chengtengip@nru.dk
Name of principal supervisor	Gitte Moos Knudsen
Title of the PhD thesis	Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depression Disorder

2. The declaration applies to the following article	
Title of article	NeuroPharm Study: EEG wakefulness regulation as biomarker in MDD
Article status	
Published ☐ Date:	Accepted for publication ☐ Date:
Manuscript submitted ☒ Date: 14 Sep 2020	Manuscript not submitted ☐
If the article is published or accepted for publication, please state the name of journal, year, volume, page and DOI (if you have the information).	

3. The PhD student's contribution to the article ***(please use the scale A-F as benchmark)*** Benchmark scale of the PhD-student's contribution to the article A. Has essentially done all the work (> 90 %) B. Has done most of the work (60-90 %) C. Has contributed considerably (30-60 %) D. Has contributed (10-30 %) E. No or little contribution (<10 %) F. Not relevant	**A, B, C, D, E, F**
1. Formulation/identification of the scientific problem	B
2. Development of the key methods	E
3. Planning of the experiments and methodology design and development	A
4. Conducting the experimental work/clinical studies/data collection/obtaining access to data	A
5. Conducting the analysis of data	A
6. Interpretation of the results	B
7. Writing of the first draft of the manuscript	B
8. Finalisation of the manuscript and submission	A
Provide a short description of the PhD student's specific contribution to the article.[1] The PhD student analyzed data and interpreted the results. She wrote the frist draft of the manuscript and revised it after co-authors' comments	

4. Material from another thesis / dissertation[ii]	
Does the article contain work which has also formed part of another thesis, e.g. master's thesis, PhD thesis or doctoral dissertation (the PhD student's or another person's)?	Yes: ☒ No: ☐
If yes, please state name of the author and title of thesis / dissertation.	Cheng-Teng Ip: Towards personalized medicine: Effectiveness of pretreatment EEG biomarker in Major Depression Disorder
If the article is part of another author's academic degree, please describe the PhD student's and the author's contributions to the article so that the individual contributions are clearly distinguishable from one another.	

5. Signatures of the co-authors[iii]

	Date	Name	Title	Signature
1.	6/10/20	Melanie Ganz	PhD	
2.	6/10-20	Vibeke Dam	PhD	
3.	27/10.20	Annia Rüesch	MSc	
4.	6/10-20	Kristin Köhler-Forsberg	PhD	
5.	6/10-2020	Martin Balslev Jørgensen	MD, DMSc	
6.	6/10-20	Vibe G. Frokjaer	MD, PhD	
7.	13/10/20	Birgitte Søgaard	PhD	p.p.
8.	13 OCT 2020	Søren R. Christensen	PhD	
9.	Oct 23 2020	Gitte M. Knudsen	MD, DMSc	
10.	26th OCT 2020	Sebastian Olbrich	MD, PhD	

6. Signature of the principal supervisor

I solemnly declare that the information provided in this declaration is accurate to the best of my knowledge.
Date: Oct 23, 2020
Principal supervisor:

7. Signature of the PhD student

I solemnly declare that the information provided in this declaration is accurate to the best of my knowledge.
Date: 23-10-2020
PhD student: Cheng-Teng Ip

Please learn more about responsible conduct of research on the Faculty of Health and Medical Sciences' website.

[i] This can be supplemented with an additional letter if needed.

[ii] Please see Ministerial Order on the PhD Programme at the Universities and Certain Higher Artistic Educational Institutions (PhD Order) § 12 (4):

"Any articles included in the thesis may be written in cooperation with others, provided that each of the co-authors submits a written declaration stating the PhD student's or the author's contribution to the work."

[iii] If more signatures are needed please add an extra sheet.