Personalizing Transcranial Direct Current Stimulation for Treating Major Depressive Disorder

Reevaluation of Clinical Trials in the Context of Precision Psychiatry

Dissertation von Stephan Alexander Goerigk

München 2021

Personalizing Transcranial Direct Current Stimulation for Treating Major Depressive Disorder

Reevaluation of Clinical Trials in the Context of Precision Psychiatry

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Fakultät für Psychologie und Pädagogik
der Ludwig-Maximilians-Universität München

vorgelegt von
Stephan Alexander Goerigk

aus Kiel

München 2021

Erstgutachter/in:	Prof. Dr. Markus Bühner
Zweitgutachter/in:	Prof. Dr. Frank Padberg
Drittgutachter/in:	Prof. Dr. Nina Sarubin
Tag der mündlichen Prüfung:	15.07.2021

Contents

List of Figures

List of Tables

List of Abbreviations

ACC	anterior cingulate cortex
AE	adverse event
AI	artificial intelligence
ALE	accumulated local effect
APPA	average posterior probability of assignments
AutoML	automated machine learning
BAC	balanced accuracy
BDI	Beck Depression Inventory
BIC	Bayesian Information Criterion
BMI	body mass index
CI	confidence interval
CO-MED	Combining Medications to Enhance Depression Outcomes
CV	cross-validation
DLPFC	dorsolateral prefrontal cortex
DMPFC	dorsomedial prefrontal cortex
DSM	Diagnostic and Statistical Manual of Mental Disorders
ECT	electroconvulsive therapy
EEG	electroencephalography
ELECT-TDCS	The Escitalopram versus Electric Current Therapy for Treating Depression Clinical Study

ENIGMA	Enhancing NeuroImaging Genetics through Meta-Analysis
ES	effect size
ESC	escitalopram
FA	factor analysis
FDR	false discovery rate
FEM	finite element model
fMRI	functional magnetic resonance imaging
GABA	gamma-aminobutyric acid
GAN	generative adversarial neural networks
GDPR	General Data Protection Regulation
HAM-D	Hamilton Depression Rating Scale
HD	high definition
HIPAA	Health Insurance Portability and Accountability Act
ICD	International Classification of Diseases
IML	interpretable machine learning
ITT	intention to treat
GAD	generalized anxiety disorder
GMM	growth mixture modelling
LCLMM	latent class linear mixed models
LMM	linear mixed regression model
LTD	long-term depression
LTP	long-term potentiation
LOESS	locally estimated scatterplot smoothing
mA	milliampere
MADRS	Montgomery-Asberg Depression Rating Scale
MDD	major depressive disorder

MINI	major depressive disorder
ML	machine learning
ML	maximum likelihood
MOCA	Montreal Cognitive Assessment
MORL	multi-objective reinforcement learning models
MRI	magnetic resonance imaging
NHT	null-hypothesis testing
NIBS	non-invasive brain stimulation
NIMH	National Institute of Mental Health
OCC	odds of correct classification
OR	odds ratio
PM	precision medicine
PANAS	Positive and Negative Affect Schedule
QIDS-SR	self rated 16-Item Quick Inventory of Depressive Symptomatology
RCT	randomized controlled trial
rTMS	repetitive transcranial magnetic stimulation
sMRI	structural Magnetic Resonance Imaging
SSRI	selective serotonin reuptake inhibitor
STAI	State Trait Anxiety Inventory
STAR*D	Sequenced Treatment Alternatives to Relieve Depression
TCI	Temperament and Character Inventory
tDCS	transcranial direct current stimulation
THREE-D	Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression
TMS	transcranial magnetic stimulation
TRD	treatment resistant depression

tES	transcranial electric stimulation
VMPFC	ventromedial prefrontal cortex
XGBoost	extreme gradient tree boosting algorithms

Abstract

Transcranial direct current stimulation (tDCS) is a safe and efficient intervention for treating major depressive disorder (MDD). However, research has suggested heterogeneity of response between patients. The emerging field of precision psychiatry aims to use statistical modeling of multi-modal data to tailor treatment to the single patient. To this end, more in-depth analysis of randomized controlled trials (RCTs) will be relevant (1) due to limited availability of other large datasets with high phenotypic detail and (2) to develop tools for personalization within counterfactually controlled environments (i.e. experimental designs with sham intervention and/or active treatment comparison) to distinguish specific vs. non-specific patterns in treatment data. Previous research has aimed at identifying patient-related factors associated with better response. However, most analyses have operated on the group-level, ignoring natural clusters within the patients' constituting factors, their individual trajectories of symptom improvement, and their presented symptoms. Furthermore, group-based modeling strategies were limited to explanatory approaches using in-sample hypothesis-testing, that are ill-suited to prognosticate outcomes of single patients.

This dissertation provides a methodological framework for reevaluation of existing clinical trial data (1) to provide future investigations with more differentiated units of analysis and (2) to complement explanatory approaches with predictive modeling strategies enabling prediction of single-patient outcomes. Using data from a landmark 3-arm clinical trial paradigmatic for a rigorously controlled experimental design (10-week treatment of tDCS vs. escitalopram vs. placebo) the dissertation provides three blueprint studies for modeling heterogeneity of tDCS response:

Study 1 characterized response to tDCS by considering patient-individual dynamics of symptom change over the course of treatment. Distinct trajectories of tDCS response could be identified (rapid-, slow-, and no/minimal improvement), representing patient subgroups with varying strength and speed of improvement. These results suggest development of individualized treatment protocols and exploration of prolonged treatment courses.

Study 2 reevaluated the efficacy of tDCS, in distinct, naturally occurring clusters of depressive symptoms. Using unsupervised machine learning (ML), a global depression measure (HAM-D) was parsed into 4 distinct symptom clusters. Analysis of cluster-scores showed superiority of tDCS and escitalopram over placebo in core depressive symptoms, but only tDCS was superior in improving sleep and only escitalopram was superior in improving guilt/anxiety symptoms, suggesting treatment selection based on patients' symptom profiles.

In Study 3 supervised ML algorithms were employed to predict response to tDCS. In

this proof-of-concept approach, response could be predicted above chance on the single-patient level, but overall accuracy was modest. Features employed for model training were explored using interpretable ML methods. Trained algorithms were provided to the field for expansion as well as tests of generalizability and incremental utility.

The presented studies illustrate how in-depth secondary analyses of clinical trial data can aid personalization of treatment. The provided methodological framework can be expanded (options are discussed) and generalized to other contexts and interventions that show heterogeneity of treatment effects. Yet, the empirical studies also epitomize challenges precision psychiatry is faced with, including low data availability, low outcome granularity, and limited external validation opportunities. The dissertation concludes with a discussion of challenges and future directions resulting from infrastructural demands in data acquisition, data management, data sharing, and interdisciplinary collaboration.

Zusammenfassung

Depressive Störungen stellen eine prävalente, stark beeinträchtigende Erkrankung dar, die häufig rezidivierend auftritt und mit einer signifikanten Sterblichkeit einhergeht. Etwa 20-40% der Patient*innen profitieren nicht ausreichend von den empfohlenen Leitlinienverfahren (Pharmakotherapie und Psychotherapie), deren Einsatz zusätzlich durch Nebenwirkungen bzw. hohe Kosten und niedrige Zugänglichkeit erschwert wird. Die Entwicklung alternativer Behandlungstechniken hat demnach eine hohe klinische Relevanz.

Die transkranielle Gleichstromstimulation (tDCS) ist eine wirksame und verträgliche Methode zur Behandlung von Depressionen, die sich durch ein gutes Sicherheitsprofil und eine hohe Kosteneffizienz auszeichnet. TDCS ist ein nicht-invasives Verfahren zur Elektrostimulation des Gehirns, bei der durch das Aufsetzen von Elektroden ein schwacher Gleichstrom induziert wird. Auch wenn die genauen Wirkmechanismen der Behandlung noch nicht abschließend geklärt sind, kann vermutlich durch Veränderungen des Ruhemembranpotentials Einfluss auf die Erregbarkeit der Neuronen genommen werden, was zu Veränderungen der neuronalen Feuerrate und des Timings neuronaler Pulse führen kann.

Die Wirksamkeit der Methode konnte in einer Reihe klinischer Studien und Metaanalysen demonstriert werden, jedoch gibt es über Studien hinweg auch gemischte Befunde und nur etwa ein Drittel der Patient*innen zeigt ein ausreichendes Therapieansprechen. Eine Erklärung für die begrenzte Wirksamkeit antidepressiver Behandlungsansätze ist die Tendenz bei Auswahl und Art der Behandlung, einem standardisierten Vorgehen zu folgen ("one-size-fits-all" Paradigma), ohne sich an der spezifischen Disposition des Patienten/ der Patientin und seines/ihres Symptomprofils auszurichten. Da Behandlungen in der Regel auf Basis von Diagnosen ausgewählt werden, führt dies häufig zu einem Prozess gemäß Versuch und Irrtum, der längere Behandlungsdauern und unbefriedigende Behandlungsergebnisse zur Folge haben kann.

Im Rahmen der Präzisionspsychiatrie (Precision Psychiatry) wird versucht, Patient*innen und Behandlung besser aufeinander abzustimmen. Ein Ansatz, der zu diesem Zweck verfolgt wird, ist die Suche nach Merkmalen, die mit einem besseren Therapieansprechen assoziiert sind und eine genauere Vorhersage des Behandlungserfolgs eines Patienten/ bzw. einer Patientin ermöglichen können. Die Detektion solch komplexer, klinischer Muster erfordert die statistische Analyse großer Mengen multimodaler Behandlungsdaten. Im Bereich der Hirnstimulation macht dies die Sekundäranalyse von Daten aus randomisiert-kontrollierten Studien (RCT) notwendig.

RCTs stellen den Goldstandard für den Nachweis der klinischen Wirksamkeit einer

Behandlung dar (in der Regel durch den Vergleich mit einer Placebo-Bedingung oder einer aktiven Kontrollintervention), könnten jedoch auch für die Personalisierung von Gehirnstimulationsverfahren einen wichtigen Stellenwert einnehmen: Zum einen, weil im Bereich der Stimulationsverfahren über die RCTs hinaus noch keine weiteren großen Datensätze mit ausreichend detaillierter Beschreibung von Patient*innencharakteristika zur Verfügung stehen. Zum anderen, weil die experimentelle Ausrichtung der RCTs eine rigide Kontrolle der spezifischen und unspezifischen Behandlungseffekte auf die Variabilität im Therapieansprechen ermöglicht. In Sekundäranalysen von RCTs konnten bereits einige mit Behandlungserfolg assoziierte Merkmale identifiziert werden, jedoch wurden größtenteils gruppenbasierte Ansätze verwendet, welche natürliche Cluster in Patient*innenmerkmalen, in individuellen Trajektorien der Symptomveränderung und in den von Patient*innen präsentierten Symptomprofilen außer Acht lassen. Außerdem wurden gruppenbezogene Analysen in der Regel mittels explanativer Modellierungsstrategien und inferenzstatistischer Hypothesentests durchgeführt, welche sich für anwendungsorientierte Prognosen von Behandlungsergebnissen einzelner Patient*innen nicht eignen.

Diese Dissertation stellt am Beispiel von tDCS einen methodisch-statistischen Rahmen zur Analyse von RCTs im Rahmen der Präzisionspsychiatrie vor. Die enthaltenen statistischen Verfahren sollen eine differenzierte Modellierung der Variabilität im Therapieansprechen und somit besseres Verständnis für die Voraussetzungen von Behandlungserfolgen ermöglichen. Dieses Analyseparadigma soll (1) das Ableiten differenzierterer Konstrukte zur Beschreibung unterschiedlichen Therapieansprechens ermöglichen und (2) explanative Ansätze zur Datenmodellierung um prädiktive Ansätze ergänzen, welche Vorhersagen über das Respondieren einzelner Patient*innen treffen können.

Innerhalb dieses Rahmens wurden auf Basis der größten derzeit verfügbaren klinischen Studie zur Wirksamkeit von tDCS (ELECT-TDCS) drei empirische Studien durchgeführt. In ELECT-TDCS wurde in einem Non-Inferiority-Design über einen Behandlungszeitraum von 10 Wochen die Wirksamkeit von tDCS mit der eines Selektiven Serotonin-Wiederaufnahme-Inhibitors (Escitalopram) und einer Placebobehandlung (Sham tDCS) verglichen. Für eine verlässliche experimentelle Kontrolle der Behandlungseffekte erhielten die Patient*innen jeder Gruppe zusätzlich eine Placeboversion der jeweils anderen Behandlung (tDCS + Placebopille; Escitalopram + sham tDCS; Doppelplacebo). ELECT-TDCS kann somit als beispielhafte Studie für ein streng kontrolliertes RCT zur Überprüfung spezifischer Behandlungseffekte angesehen werden und eignet sich demnach als Blaupause zur empirischen Demonstration des hier vorgestellten Analyseparadigmas zur Modellierung von Therapieeffekten.

Studie 1

In Studie 1 wurde das Ziel verfolgt, das antidepressive Ansprechen auf tDCS unter Beachtung individueller Unterschiede in der Dynamik der symptomatischen Verbesserung zu evaluieren. Latente Growth Mixture Modelle (LGMM) wurden berechnet, um Patient*innen auf Grundlage ihrer spezifischen longitudinalen Symptomverläufe zu charakterisieren und somit Rückschlüsse auf unterschiedliche Geschwindigkeiten in der therapeutischen Ansprechrate

zuzulassen. In diesem Zusammenhang konnten drei distinkte Wachstumskurvencluster (schnell, langsam, keine/minimale Verbesserung), identifiziert werden, welche Gruppen von Patient*innen repräsentieren, die mit ähnlicher Intensität und Latenz auf die Behandlung reagierten. Die Kategorien zeigten eine hohe klinische Relevanz, da sie bereits nach einer Behandlungswoche differenziell das Theapieansprechen (Response und Remission) prädizieren konnten. Zudem wurden nur Kategorien für eine weitergehende Analyse zugelassen, die zumindest 5% der Gesamtpopulation abdeckten. Insgesamt wurden 43.6% der Patient*innen, die mit tDCS behandelt wurden, in die schnelle Verbesserungskategorie eingeordnet. Dieser Befund ist von hoher Relevanz, da gruppenbasierte Auswertungen ein Therapieansprechen erst nach Ende der akuten Behandlungsphase vermuten ließen. Ein schnelleres Respondieren wäre im Kontext der Entwicklung heimbasierter Anwendungen von tDCS vielversprechend, um die Belastung durch tägliche Besuche klinischer Behandlungszentren zu reduzieren.

Die Kategorien waren mit unterschiedlichen a-priori selektierten Charakteristika assoziiert: So zeigte sich ein schlechteres Therapieansprechen bei Patient*innen jüngeren Alters und bei Patient*innen mit stärkerer Symptomlast. Diese Patient*innengruppen scheinen somit weniger wahrscheinlich schnelle Behandlungserfolge zu zeigen und könnten von längeren Behandlungen profitieren. Zudem war ein langsameres Ansprechen auf die Behandlung mit der Einnahme von Benzodiazepinen assoziiert. Dieser Befund ist besonders relevant, da Benzodiazepine ein modifizierbares Behandlungselement darstellen und vor der Therapie mit tDCS in ihrer Dosis reduziert oder abgesetzt werden könnten. Weitere mit dem Therapieansprechen assoziierte Merkmale wurden durch ein exploratives, kreuzvalidiertes Ranking-Verfahren mit Hilfe eines elastic-net Algorithmus identifiziert. Diese beinhalteten unter anderem globale Maße der Symptomschwere (Montgomery-Åsberg Depressionsskala [MADRS] und Beck Depressionsinventar [BDI], das Ersterkrankungsalter, Einnahme von Schlafmitteln (Nicht-Benzodiazepin-Agonisten, auch als z-Drugs bezeichnet) und mit präfrontaler Aktivität assoziierte Konstrukte (negativer Affekt, Eigenschaftsangst). Die Ergebnisse sind bedeutsam für die Personalisierung von Behandlungen mit tDCS, da sie die Entwicklung individualisierter Stimulationsprotokolle nahelegen und die Überprüfung des Nutzens verlängerter Behandlungszeiträume suggerieren.

Studie 2

In Studie 2 wurden die Effekte von tDCS auf unterschiedliche depressive Symptomcluster untersucht. Wirksamkeitsstudien antidepressiver Behandlungsansätze berichten fast ausschließlich globale Depressionsmaße (z.B. Summenwerte), für deren Berechnung über die Ausprägung der einzelnen Symptome hinweg aggregiert wird. Allerdings besteht insbesondere beim Krankheitsbild der Depression eine erhebliche Heterogenität auf der Symptomebene. Das pauschale Verrechnen einzelner Symptomwerte erschwert somit eine differenzierte Betrachtung des Wirkprofils der Interventionen. In Folge der Aggregation können beispielsweise ausbleibende Effekte in einer Symptomgruppe Verbesserungen in einer anderen Symptomgruppe überdecken.

Um ein besseres Verständnis für die symptomspezifischen antidepressiven Effekte von tDCS zu erlangen, wurde ein gängiges Maß zur Symptomerfassung, die Hamilton

Depression Rating Skala (HAM-D), mit Hilfe von unsupervised Machine Learning (ML) Algorithmen hinsichtlich in den Symptomen vorhandener Untergruppen untersucht. Durch den Einsatz von hierarchischen, agglomerativen Clusterverfahren konnten vier getrennte Cluster identifiziert werden (depressive Kernsymptomatik, Schlafprobleme, Angst- und Schuldsymptome sowie atypische Symptome). Diese Clusterstruktur fand sich ebenfalls in einer analog durchgeführten Analyse einer großen psychopharmakologischen Vergleichsstudie (STAR*D) und fand sich ebenfalls in einer Sensitivitätsanalyse mit k-means Clustering Algorithmen, was die Robustheit der Ergebnisse stützt. Vergleiche der Wirksamkeit der tDCS-Behandlung innerhalb der Cluster zeigten eine ausgeprägtere Verbesserung in der depressiven Kernsymptomatik und in den mit Angst und Schuld assoziierten Symptomen als in den schlafbezogenen sowie den atypischen Symptomen. Dieses Ergebnis zeigte sich in allen drei Behandlungsgruppen. Zudem verringerte sich die depressive Kernsymptomatik in stärkerem Ausmaß als die Angst- und Schuldsymptomatik.

Beim Vergleich zwischen den Gruppen zeigte sich hinsichtlich der Verringerung der depressiven Kernsymptomatik eine Überlegenheit gegenüber der Placebobehandlung nach Behandlung mit Escitalopram und auch nach Behandlung mit tDCS. Hinsichtlich der schlafbezogenen Symptomatik führte jedoch nur die tDCS-Behandlung zu einer signifikanten (stärkeren) Reduktion, während in Bezug auf die Angst- und Schuldsymptomatik nur die Behandlung mit Escitalopram eine signifikante Reduktion zur Folge hatte. Es zeigte sich, dass Escitalopram zudem die depressive Kernsymptomatik in stärkerem Ausmaß verringerte als tDCS, was die Hauptbefunde des ELECT-TDCS Trials auf Basis des globalen Depressionswertes widerspiegelt. Dies sind die ersten Ergebnisse, die eine Überlegenheit von tDCS gegenüber Placebo in einem Symptomcluster suggerieren, in welchem eine psychopharmakologische Intervention nicht überlegen war. Die Ergebnisse können zur Personalisierung der antidepressiven Behandlung beitragen, da sie Entscheidungen über die Selektion von tDCS gegenüber pharmakologischen Interventionen auf Basis des individuellen depressiven Symptomprofils ermöglichen.

Studie 3

In Studie 3 wurde die Vorhersagbarkeit (Proof-of-Concept) des Therapieansprechens einzelner Patient*innen unter Einsatz prädiktiver Modellierungsansätze geprüft. In der prädiktiven Modellierung liegt der Fokus explizit auf einer möglichst genauen Vorhersage einzelner Datenpunkte und eignet sich daher besonders gut für die Prognose von Behandlungverläufen im klinischen Setting. Im Gegensatz dazu nutzen die in der psychiatrischen Forschung häufig eingesetzten explanativen Verfahren formale Hypothesentests zur Prüfung angenommener Gruppeneffekte oder assoziativer Zusammenhänge, deren Verallgemeinerbarkeit über die beobachtete Stichprobe hinaus anhand inferenzstatistischer Kriterien bewertet wird (p-Werte, Effektstärken). Diese Verfahren sind eignen sich daher weniger gut für die Vorhersage externer oder zukünftiger Datenpunkte. Im prädiktiven Ansatz wird am Beispiel von Patient*innen, deren Behandlungsergebnis bereits bekannt ist, ein statistisches Modell trainiert, welches Muster in den Daten erkennen kann, die mit einem wahrscheinlichen Behandlungserfolg assoziiert sind (Traning). Zur Evaluation der Vorhersagegenauigkeit des Modells (Test)

ist es wichtig, eine explizite Trennung zwischen Fällen herzustellen, die zum Training des Algorithmus herangezogen wurden und denen, die zur Evaluierung des Modells dienen. Ist diese Trennung nicht gegeben, besteht die Gefahr einer Überanpassung des Modells an die Daten (Overfitting), was zu einer schlechteren Vorhersagegenauigkeit des Modells und zu überoptimistischen Einschätzungen der Vorhersagegenauigkeit führen kann. Eine derartige Trennung von Trainings- und Testdaten wird beispielsweise im Paradigma der nested-Cross-Validation (CV) angestrebt, bei welcher die verfügbaren Daten nach dem Zufallsprinzip in ähnlich große Bestandteile (Folds) geteilt werden. Dabei wird jeweils ein Fold aus der Modellierung herausgehalten (hold-out oder Test-Set), während auf den übrigen Daten der Algorithmus trainiert werden kann (Trainings-Set). Der nicht zum Training verwendete Teil der Daten wird dann zur Vorhersageprüfung des Modells genutzt. Das iterative Wiederholen dieses Vorgangs erlaubt es, den gesamten Datensatz zur Untersuchung der Vorhersagbarkeit des Behandlungsausgangs zu nutzen.

Zur Vorhersage des Therapieansprechens in ELECT-TDCS wurden XG-Boost Algorithmen mit klinischen und soziodemographischen Daten trainiert, die ausschließlich vor Therapiebeginn erhoben wurden. Diese Auswahl von Prädiktoren (Features) hat den Vorteil besonders kostengünstig und mit verhältnismäßig geringem Aufwand erhebbar zu sein. Das Therapieansprechen (Response) wurde definiert als $\geq$50% Symptomreduktion von Baseline bis Behandlungsende in Woche 10. Die Algorithmen wurden im Rahmen einer 5-fach wiederholten, 10-fold nested CV trainiert und hinsichtlich ihrer Vorhersagegenauigkeit evaluiert. Dieser Prozess wurde analog innerhalb jeder Behandlungsgruppe wiederholt (tDCS, Escitalopram, Placebo). Die kreuzvalidierte Evaluation der Modelle ergab eine signifikant überzufällige Vorhersagegenauigkeit für das tDCS-Modell ($BAC = 64.75\%$) und das Escitalopram-Modell ($BAC = 59.80\%$). Eine Placebo Response konnte nicht systematisch vorhergesagt werden ($BAC = 42.90\%$). Diese Erkennungsraten fallen im Vergleich zur Genauigkeit ($BAC = 85\%$) von Algorithmen, die zur Prädikton von Behandlunganspreche auf repititive transkranielle Magnetstimulation (rTMS) verwendet, wurden gering aus. Die rTMS Modelle wurden jedoch mit Bildgebungsdaten aus der strukturellen Magnetresonanztomographie trainiert. Die in ELECT-TDCS erreichte Genauigkeit ist jedoch vergleichbar mit der von Algorithmen aus großen psychopharmakologischen Studien (STAR*D), die ebenfalls nur mit klinischen (Fragebogen-)daten trainiert wurden ($BAC = 65\%$).

Zur Prüfung behandlungsspezifischer Vorhersagemuster, wurden der Algorithmus, der zur Erkennung einer tDCS Response trainiert wurde, auf Daten der Patient*innen angewendet die Escitalopram erhielten und umgekehrt. Response konnte jeweils nicht durch den für die andere Intervention trainierten Algorithmus vorhergesagt werden, was den Schluss von unterschiedlichen Wirkmechanismen der Behandlungen nahelegt. Patient*innen, die nach 10-wöchiger Behandlung mit Escitalopram und Placebo noch keine Verbesserung der Symptome zeigten, konnten im Rahmen einer, an ELECT-TDCS angeschlossenen, Open-Label Phase eine zusätzliche tDCS Behandlung erhalten. Zur weiteren Validierung der tDCS Modelle wurde das Therapieansprechen innerhalb dieser Open-Label Phase vorhergesagt. Als Responder prognostizierte Patient*innen zeigten eine stärkere Verbesserung in ihrer Symptomatik. Dieser Effekt war jedoch nicht signifikant (möglicherweise aufgrund der

kleinen Stichprobengröße in der Open-Label Phase).

Eine weitere Exploration der Modelle wurde mittels interpretierbarer Machine Learning (IML) Methoden durchgeführt. Diese Methoden erlauben durch Permutation der Werte einzelner Prädiktorvariablen den Informationsgehalt dieser Variablen für das Modell unbrauchbar zu machen. Die Reduktion der Vorhersagegenauigkeit durch die systematische Eliminierung der jeweiligen Variable kann somit Schlüsse über ihren Beitrag zum Klassifikationsprozess zulassen. Prospektiv wird eine höhere Vorhersagegenauigkeit notwendig sein, um prädiktive Modelle zur Stützung klinischer Therapieentscheidungen nutzen zu können.

Die in Studie 3 trainierten Algorithmen stellen einen Proof-of-Concept Ansatz dar, wurden aber dennoch der tDCS Forschungsgemeinschaft zur Erweiterung und zur Überprüfung ihrer Generalisierbarkeit (anhand externer Datensätze) und ihrer klinischen Anwendbarkeit zur Verfügung gestellt.

Fazit

Traditionell wurde eine inter-individuelle Variabilität im Therapieansprechen als ein Hindernis für die Entwicklung standardisierter Behandlungprotokolle angesehen. Unter Berücksichtigung der phänotypischen Variabilität depressiver Patient*innen, wird der "One-size-fits-all" Ansatz psychiatrischer Behandlungen mehr und mehr in Frage gestellt. Stattdessen werden vermehrt präzisionsorientierte Ansätze zur Verbesserung der Passung zwischen Behandlung und Patient*in verfolgt.

Im Rahmen dieser Dissertation wurde der Beitrag von RCTs zum präzisionspsychiatrischen Paradigma (1) als Datenquelle für Modellentwicklung und Beschreibung der Heterogenität von Effekten psychiatrischer Behandlung und (2) als experimenteller Rahmen für die divergente Validierung mit Behandlungseffekten assoziierter klinischer Muster im Vergleich zu Mustern innerhalb aktiver und Placebokontrollbedingungen herausgestellt. Analysen undifferenzierter RCT-Ergebnismaße können jedoch ein Übersehen relevanter Informationen in den Patient*innenencharakteristika, den individuellen Symptomverläufen und den Symptomen der Patient*innen zur Folge haben und häufig zur RCT-Auswertung eingesetzte explanative Verfahren lassen keine direkten Vorhersagen über Behandlungsverläufe einzelner Patient*innen zu.

Die Dissertation stellt einen methodisch-statistischen Rahmen für Sekundäranalysen klinischer Studiendaten am Beispiel der ELECT Studie vor, welcher Lösungsansätze mit direktem Bezug zu den Problemen traditioneller Analysen klinischer Studiendaten beinhaltet. Die Implementation dieser Lösungsansätze wurde beispielhaft in drei empirischen Studien demonstriert. Die Ergebnisse dieser Studien weisen auf Möglichkeiten hin, die Personalisierung antidepressiver Behandlung durch individuelle Behandlungsprotokolle und Adaptation modifizierbarer Behandlungsfaktoren, durch Interventionsselektion auf Basis des individuellen Symptomprofils und durch prädiktiv-analytisch ermittelte Wahrscheinlichkeiten des Behandlungserfolgs einzelner Patient*innen zu verbessern. Jedoch beschränkt sich die Aussagekraft dieser Ergebnisse auf den speziellen Behandlungskontext von ELECT-TDCS, was eine weitergehende Überprüfung der Generalisierbarkeit der Ergebnisse erforderlich macht. Innerhalb des Analyseparadigmas wurde in der Dissertation die

Kombinierbarkeit explanativer und prädiktiver Modellierungsansätze demonstriert und die Nutzung von ML-Methoden in die präzisionsorientierte Erforschung von tDCS eingeführt.

Zusammenfassend zeigt sich am Beispiel von tDCS, dass die Individualisierung von psychiatrischer Versorgung noch am Anfang ihrer Entwicklung steht. Jedoch kann das präsentierte Paradigma für tiefergehende Analysen klinischer Studiendaten auf andere Interventionen und Kontexte übertragen werden, in welchen die bestehende Variabilität von Behandlungsergebnissen eine Personalisierung nahelegt. Die Dissertation kann zu einem verbesserten Verständnis der Effekte von tDCS verhelfen und somit zur Verbesserung der Intervention beitragen. Während es unklar ist, ob die Behandlung mit tDCS vollständig individualisierbar ist, stellen die Methoden des präzisionspsychiatrischen Paradigmas einen vielversprechenden Ansatz zur Optimierung psychiatrischer Forschung und Behandlung dar. Jedoch wird es analog zu jedem aufkommenden Forschungsparadigma notwendig sein, verfrühtem Optimismus mit Vorsicht zu begegnen, damit das Wohl der Patient*innen an erster Stelle steht.

Chapter 1

Introduction

1.1 Treatment of Depression with TDCS

Major depressive disorder (MDD) is a prevalent (Kupfer et al., 2016), seriously impairing mental disorder, which is often recurrent and causes significant mortality (Kessler et al., 2003; Liu et al., 2020). The World Health Organization ranks MDD as one of the leading causes of disability worldwide as measured by years lived with the disability (Murray et al., 1996; Lopez et al., 2006). Chronic and treatment-resistant courses of illness are experienced by an estimated 15-25% of MDD patients and 20-40% of patients do not benefit sufficiently from available treatments that are recommended by current psychiatric guidelines (Härter et al., 2010; Parikh et al., 2016; Kennedy et al., 2016), including pharmacotherapy (Wong et al., 2010; Rush et al., 2006b) and psychotherapy (Hofmann et al., 2012). Furthermore, pharmacological interventions are limited by adverse effects (Carvalho et al., 2016) and while psychotherapeutic interventions are moderately effective for treating depression (Cuijpers et al., 2011) they are curbed by delayed effects (Keller et al., 2000), costs, and limited availability (Cuijpers, 2017). Consequently, there is an urgent demand for alternative methods of treatment, which makes the development and optimization of novel techniques a priority.

One such class of new interventions are non-invasive brain stimulation techniques (NIBS). Their therapeutic application is based on the modulation of brain activity through the delivery of a stimulus without the introduction of instruments inside the body or breaking the skin (Dayan et al., 2013). In MDD, the application of NIBS techniques targeting prefrontal cortical areas has emerged as a novel treatment option with promising effects (Palm et al., 2016; Dunlop et al., 2017).

In 2009, transcranial magnetic stimulation (TMS), a method that applies a changing magnetic field using a magnetic coil to cause electric currents at specific areas of the brain through electromagnetic induction (Barker et al., 1985; Lefaucheur et al., 2014), was approved as the first NIBS technique for the treatment of MDD by the Food and Drug Administration (FDA) (Connolly et al., 2012). While TMS has shown promising results in various clinical trials (Brunoni et al., 2017a), it is associated with small risks of seizure

(Rosa and Lisanby, 2012) and relatively expensive.

1.1.1 Rationale for TDCS

Within NIBS, transcranial electric stimulation (tES) constitutes a class of techniques, where a weak direct or alternating current is applied via electrodes, that are placed directly on the scalp, with the largest body of evidence for transcranial direct current stimulation (tDCS). TDCS is a promising treatment alternative due to its excellent safety profile in short-term application and has not been associated with seizures (Bikson et al., 2016). Further advantages of tDCS are its ease of use and superior cost effectiveness, as compared to other NIBS methods (Gandiga et al., 2006). Mechanisms of tDCS action are not fully elucidated, however, tDCS has been shown to influence neuronal firing rates and spike timing probably by inducing shifts in resting membrane potentials towards de- or hyperpolarization, i.e. eliciting an increase or decrease of the neuron's excitability, respectively (Fritsch et al., 2010; Nitsche and Paulus, 2000; Purpura and McMurtry, 1965; Woods et al., 2016).

One strategy for effective targeting of network structures involved in MDD is a wide spread stimulation, simultaneously reaching relevant brain regions as well as their connecting fiber tracts. In MDD, tDCS is usually applied over the dorsolateral prefrontal cortex (DLPFC), which is considered a particularly relevant target for mood regulation (Mayberg et al., 2000; McTeague et al., 2017). By acting on neuronal activity in target regions and functional connectivity between electrodes, tDCS can potentially modulate multiple neural networks involved in MDD, such as distinct resting state networks (Keeser et al., 2011). This has been shown for prefrontal tDCS (Wörsching et al., 2016; Chan et al., 2021), though findings were not consistent across studies. Very recently, engagement of stimulation targets and modulation of depression-relevant circuits have been shown in a pilot study in MDD, supporting prefrontal tDCS as a mechanism-based antidepressant intervention (Jog et al., 2021).

1.1.2 Results of Clinical Trials for Treatment of MDD

The efficacy and safety of tDCS has been demonstrated in several clinical trials (Fregni et al., 2006; Boggio et al., 2008; Brunoni et al., 2013c; Martin et al., 2011) and meta-analyses (Nitsche et al., 2009; Brunoni et al., 2016; Moffa et al., 2020; Mutz et al., 2019). However, results are heterogeneous across studies and across patients with only about a third showing a clinical response (Brunoni et al., 2016; Moffa et al., 2020)). In a recent randomized controlled trial (RCT), tDCS did not show non-inferiority to escitalopram (Brunoni et al., 2017b), and in another large multicenter trial no superiority of tDCS over sham (placebo) could be demonstrated (Loo et al., 2018).

The variability of responses to tDCS challenges its successful application in psychiatric care. Thus, while there is need for further confirmatory RCTs in order to develop tDCS towards clinical implementation, it seems pivotal to further improve its efficacy and to gain a deeper understanding of its antidepressant mechanisms.

1.1.3 Improving Efficacy of TDCS for Treating MDD

Small between-group effect sizes and substantial between-subject heterogeneity of treatment response emphasize the relevance of improving efficacy of tDCS (Razza et al., 2020; Fregni et al., 2020). As in other NIBS methods, this heterogeneity may result from a variety of sociodemographic, connectivity-based, neurofunctional, neuroanatomical, and genetic factors that mediate the variability of stimulation effectiveness (Brunoni et al., 2012). One explanation for the limited efficacy of antidepressant treatments in general, is that they are typically selected within a one-size-fits-all paradigm with little guidance on patient-specific symptomatic burden, the neurobiological underpinnings of the symptoms, or related lifestyle factors. However, each patient is unique and particularly for treatment of MDD, their uniqueness may be of relevance to care (Chekroud et al., 2017b). Discovery and validation of response predictors that reliably operate in the individual patient setting could increase effectiveness and applicability of tDCS (Borrione and Brunoni, 2019). To address this lack of personalization in psychiatric treatment, the framework of precision psychiatry has been proposed, aiming to tailor psychiatric treatment to the individual characteristics of each patient (Fernandes et al., 2017).

1.2 Personalizing TDCS in the Context of Precision Medicine

1.2.1 Precision Psychiatry and the Status Quo in NIBS and TDCS Research

The underlying concept of precision medicine (PM), as previously defined by Hodson Hodson (2016), is constituted in individually tailoring health care according to the patient's genes, lifestyle and environment. In recent years, this framework has gained increased traction, as highlighted by the founding of the Precision Medicine Initiative in 2015 (Collins and Varmus, 2015). Since then, it has been conceptualized in several medical disciplines including psychiatry (precision psychiatry) (Williams, 2016; Fernandes et al., 2017; Friston, 2017).

In precision psychiatry, there is a strong focus on exploiting statistical analyses of large bodies of data to uncover complex clinical patterns. These data-derived tools can in turn be used to optimize patient stratification and clinical management, selection between treatment modalities, treatment adjustment, and improved prognosis tailored to each patient (Drysdale et al., 2017; Bzdok and Meyer-Lindenberg, 2018). Since a clinician's choice of the optimal treatment often does not exclusively depend on knowledge about causes or maintaining factors of the symptoms of a given patient, systematic benchmarking of treatment response probabilities for intermediate phenotypes has a translational potential in matching patients with the right treatment options and in reducing delays between bench and bedside (Perna and Nemeroff, 2017).

Personalization of NIBS

Parallel to research on other therapies such as psychotherapy (Cuijpers et al., 2016) and pharmacotherapy (Menke, 2018), NIBS techniques are a highly promising field for the development of personalized interventions which may replace current "one-size-fits-all" approaches. As stimulation parameters can be directed to specifically affected brain areas, NIBS techniques are well suited for a precision-oriented approach that is guided by information about the circuits underlying emotional, cognitive, and self-reflective functions relevant to MDD, in order to individualize patient-oriented treatments (Williams, 2016).

Within this framework, directions for precision-oriented personalization of the intervention itself include (1) optimization of stimulation timing using closed-loop approaches that may allow synchronization with neural oscillatory network activation through live functional magnetic resonance imaging (fMRI) or electroencephalography (EEG) read-outs (Schestatsky et al., 2013) (2) optimization of location and dosage by systematic variation of treatment parameters (e.g. electrode positioning and size, intensity and duration of stimulation, number of sessions per day, and interval between sessions) and establishment of dose-effect relationships using computational finite element models (FEM) of electric fields (efields) as a proxy of dosage (Thielscher et al., 2011; Mezger et al., 2021) (3) combined and synergistic approaches with other interventions such as pharmacological treatments (Brunoni et al., 2013c), (computerized) cognitive training (Wolkenstein and Plewnia, 2013; Brunoni et al., 2014), and psychotherapy (Bajbouj and Padberg, 2014; Bajbouj et al., 2018). For general summaries on the opportunities and challenges of precision psychiatry in NIBS, see the review articles by Borrione and colleagues Borrione et al. (2020) and Padberg and colleagues Padberg et al. (2021). A fourth approach of precision-oriented personalization is the search for patient-specific factors that allow prediction of a patient's response to treatment. The objectives of this dissertation are primarily focused on this strategy.

Identification of Responders and Non-Responders

The search for responders and non-responders is motivated by the assumption that particular patients, subgroups of patients, or subgroups of symptoms may benefit from the intervention while others do not (Guerra et al., 2020). Instead of a trial-and-error approach, the objective is to identify these distinctly responding entities and to match the right treatment with the right patient or the right symptom, respectively.

In brain stimulation, most attempts to characterize preconditions for a patient's response to treatment have been focused on identification of associated factors on the group-level. Early predictor analyses in data from open and controlled rTMS and tDCS trials proposed disease-related factors such as treatment resistant depression (TRD) and episode duration that were later replicated (Brakemeier et al., 2007, 2008; Fregni et al., 2006; Holtzheimer III et al., 2004).

While there are many studies on associations of TMS treatment response with neurophysiological markers including neuroimaging data, it is more difficult to focally target cortex regions using tES methods. Still, there is an increasing number of findings, relating

biomarkers to antidepressant tDCS response, including individual factors of cortex morphology such as grey matter volumes of PFC sub-regions (Bulubas et al., 2019), individual efield strength in the left anterior cingulate cortex (ACC) (Suen et al., 2021), and regional functional activation of the DLPFC (Nord et al., 2019), plasma levels of neurotrophic factors, interleukins and their receptors (Brunoni et al., 2018; Goerigk et al., 2021), heart rate variability (Brunoni et al., 2013b), and genetics (Brunoni et al., 2013a).

Despite the rapidly growing number of associative findings, there does not yet seem to be a consistent functional or structural pattern to explain heterogeneity of response to tDCS. Consequently, the identified predictors have not yet been of clinical relevance for stratified treatment. The limited understanding of MDD and its neurobiological underpinnings might itself be a large factor for the insufficient explainability of heterogeneous treatment effects. However, some of the uncertainty may be attributable to undifferentiated concepts of response (Uher et al., 2010), the measures used to define it (Fried and Nesse, 2015; Olbert et al., 2014), and the framework used to identify and validate predictors. While some of these shortcomings may be explained by axioms within the analysis of RCT data, these trials may assume (or retain) an integral role in precision psychiatry and should thus be analysed using more in-depth concepts of analysis. The following paragraphs will outline some of the upsides and pitfalls of using RCT data for treatment personalisation.

Precision Medicine and Randomised Controlled Trials

The gold-standard research paradigm for the evaluation of tDCS treatment effects in MDD is the demonstration of group-level treatment response, usually compared to a placebo group within RCTs. While these trials are indispensable for demonstration of efficacy on the population level (evidence based practice), they have also been the main source of data to be used in the attempt to personalize brain stimulation techniques, i.e. due to limited availability of data, most secondary analyses attempting to find predictors for response have focused on data from RCTs.

However, the role of placebo and RCTs in the context of PM is still unclear. To identify responders and non-responders to treatment, one would tend to focus on variability in treatment outcomes of those patients that have received an active stimulation and discard data from those who received a sham treatment. However, while the variability in treatment effects is the core assumption motivating the search for responders and predictive biomarkers, it has recently been questioned whether there is indeed a larger induction of variability in response to active brain stimulation (i.e. variability as a consequence of treatment) as compared to the observed variability in response to sham (placebo stimulation). In a large meta analysis (N = 5748), Homan et al. (2021) found only little evidence for different amounts of treatment response variability after active compared to sham treatment. These results suggest, that the need for and possibility of personalization through explanation of treatment-induced variability remains an open question.

The incorporation of RCTs as rigorously planned experiments to provide counterfactual control (i.e. what would effects have looked like without the treatment or with another treatment) not only of the efficacy of treatments but also of the models to predict variability

in their effects might therefore remain of relevance to PM. Conversely, a precision-oriented approach may contribute to tackling and distinguishing the specific vs. non-specific (placebo) effects of treatment. Thus, the reconciliation of the experimentally (placebo-)controlled approach of the RCT with individual-level personalization seems necessary, both from a practical (data availability) and a conceptual point of view.

However, in understanding the variability that can be observed after active treatment, there are challenges that require new statistical frameworks for dissecting patient-individual effects. One of the challenges for the prediction of treatment response in data from clinical trials, is that they are often evaluated on arbitrarily chosen or undifferentiated primary and secondary outcomes, while other variance components relevant for unraveling treatment effects may be overlooked. Thus, to profit from clinical trial data in an attempt to personalize treatment, some caveats in analyzing RCT data should be considered.

1.2.2 Caveats in Analyzing Classical Primary and Secondary Outcomes of RCTs

The primary hypotheses of RCTs are often evaluated on units of analysis that make it difficult to comprehend variability on the individual patient level (i.e. demonstration of group average improvement in global depression scores). In fact, RCTs are designed to minimize the impact of individual patient nuance, but these nuances could be crucial in the explanation and definition of treatment response. Secondary outcomes often include the use of statistical cut-offs for categorization of treatment response for better interpretability, which can lead to arbitrary dichotomies between patients (e.g. 50% improvement cutoff conceptually assigns 1% and 49% improvement to same and 49% and 51% to a different response category)(Senn, 2018). In the decision-making process about who to treat with tDCS, this often leaves three relevant questions unanswered: (1) Is a specific patient likely to show response to tDCS? (2) Which specific aspects of the condition can be expected to improve? (3) To what extent and when is the treatment going to help? The following paragraphs summarize the caveats in traditional definitions and analyses of treatment response.

Problem 1: Assumption of Homogeneous Change over the Course of Treatment

One aspect limiting the understanding of heterogeneous tDCS response and its predictors is constituted in the dynamics of symptomatic improvement. The main interest of the efficacy evaluation in RCTs is the treatment-dependent difference in symptomatic change over time. In statistical methods used to test the hypotheses of clinical trials on continuous outcomes (e.g. linear mixed regression modeling (LMM) (Cnaan et al., 1997)) this difference in symptom change, acting as the primary unit of analysis, is represented in the time x group interaction effect. However, the axioms of these hypothesis tests assume an average trajectory within each treatment group (single-class model) (Uher et al., 2010). In other words, from baseline until endpoint (i.e. when clinical response rates are defined cross-sectionally), all patients who received the treatment are assumed to improve in the same

way. While sub-grouping approaches defined by cut-offs on the last observed measurement of depression severity take into account information about the degree of improvement (e.g. response vs. non-response), but ignore the time course of symptomatic change and can consequently lead to inefficient analyses (Quitkin et al., 1984; Royston et al., 2006; Muthén et al., 2008). However, if continuous change is merely evaluated on the treatment group level, information about the more fine-grained dynamics of response, such as subgroups of fast responding vs. slow responding patients, is lost (Beunckens et al., 2008).

Problem 2: Measurement of Treatment Effects Using Aggregate Outcomes

Clinical trials including patients with MDD, who are treated with tDCS, almost exclusively report aggregate symptom severity scores as primary outcomes (e.g. Hamilton Depression Rating Scale (HAM-D) (Hamilton, 1960); Beck Depression Inventory (BDI) (Beck et al., 1961); Montgomery-Åsberg Depression Rating Scale (MADRS) (Montgomery and Åsberg, 1979)). However, there is also heterogeneity on the level of depressive symptoms (psychometrically speaking, on the item level) presented by the patients (Fava et al., 1997; Musil et al., 2018), which may impede the evaluation of clinical interventions for depression (Fried and Nesse, 2015; Olbert et al., 2014) and the identification of predictors for response. Factor analytic studies (Li et al., 2014; Romera et al., 2008) and meta-analyses (Shafer, 2006) of large MDD patient populations indicate the organisation of MDD symptoms into up to 5 clusters depending on the employed symptom checklist. Treatment effects on one cluster of symptoms may be masked in the final sum scores by a lack of treatment effects on another group of symptoms. This has been suggested as one explanation for the mixed results from larger comparative meta-analyses for antidepressant treatments (Chekroud et al., 2017a; Cipriani et al., 2009; Gartlehner et al., 2011). While symptoms may be organized based on clinical experience (Lin and Stevens, 2014), without clear recommendations on matching symptoms (rather than diagnostic categories) to treatments this approach may contribute to the above-mentioned trial-and-error selection process, leading to longer treatment duration (Rush et al., 2006a; Chekroud et al., 2017a).

Problem 3: Losing Sight of the Individual Outcome

The primary research goal in RCTs is to introduce novel treatment options that benefit a majority of one particular clinical group. Consequently, statistical methods used in the evaluation of treatment response have a long-standing focus on formally testing group effects (null-hypothesis testing, NHT), which is achieved by fitting a project-specific probability model to explain treatment-induced variance in symptom outcomes (Breiman et al., 2001b). As an important practical consequence these methods based on the concept of statistical significance have less obvious potential for judgments on single individuals within a treatment group (Bzdok and Meyer-Lindenberg, 2018).

Applications of NHT methods usually aim at finding statistical effects in the data at hand (in-sample estimates), without evaluating fitted models on unseen or future data points (Bzdok and Yeo, 2017). This lack of ubiquitous out-of-sample model validation in

NHT may explain some of the mixed results across different RCTs. Hence, NHT takes the form of a one-step procedure by producing an effect size or p-value that can itself not be extrapolated for prognosis of a patient's improvement in a later step (Friedman et al., 2001). In fact, observed statistically significant effects by a p-value do not measure reproducibility or replication (Goodman, 1992). In psychiatry, this has led to a large number of statistically significant, but not reproducible results (Ioannidis, 2005; Nuzzo, 2014) and even if a result is reproducible, significant findings may not be clinically meaningful (Abi-Dargham and Horga, 2016; Ioannidis, 2016) and may not reach a high prediction accuracy in unseen data (Arbabshirani et al., 2017; Breiman et al., 2001b; Shmueli et al., 2010). By adopting this retrospective point of view, the prognostic potential of NHT is reduced and its relevance for clinical practice is impeded.

Summarily, while the approach of investigating predictors for better treatment outcomes and characteristics of responders is generally not new, data are often under-analysed and associations are masked by undifferentiated units of analysis. These shortcomings limit the understanding of response heterogeneity which has traditionally been regarded as a limitation for developing standardized tDCS treatment procedures.

1.2.3 Using TDCS Studies for Developing a Paradigm of In-Depth Analyses of Clinical Trials

With more frequent application of the precision-oriented approach and its tools for in-depth analyses, there may be new opportunities to better exploit the available clinical trial data, both on the level of more differentiated treatment outcomes and on the level of predictive analytics to model them (Bzdok and Meyer-Lindenberg, 2018). Driven by a general increase in data availability (UK Biobank (Sudlow et al., 2015); Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) (Thompson et al., 2014)) more computational methods such as latent variable modeling for meaningful reduction of data dimensionality (Galbraith et al., 2002) and predictive approaches from the domain of artificial intelligence (AI) (Jameson and Longo, 2015) are being applied in psychiatry (Chekroud et al., 2017b; Bzdok and Meyer-Lindenberg, 2018).

These methods (including the subset of machine learning [ML] algorithms) (1) can detect general principles within a series of observations and (2) can handle large numbers of predictors where researchers were historically forced to limit the number of variables in their analyses (Zou and Hastie, 2005), while (3) making few formal assumptions about associations between variables and the distribution of the data (Goodfellow et al., 2016; Bzdok, 2017). Thus, they follow data-driven decision rules, rather than arbitrarily defining outcomes (e.g. traditional dichotomous cutoff values) or selecting variables as predictors. Available modeling tools comprise unsupervised algorithms for dimensionality reduction and clustering that can effectively discover unknown constellations in patient data and supervised methods specialized for accurate outcome prediction (Mohri et al., 2018). Hence, complementing the benefits of classical NHT in psychiatry, these new methods provide promising alternatives in addressing current challenges in application of brain stimulation

techniques.

While these computational tools have not been used much in the field of brain stimulation and, to our knowledge, not at all in tDCS research, their application has been called for in a number of recent commentaries and review articles on NIBS within precision psychiatry (Passos and Mwangi, 2018; Borrione and Brunoni, 2019; Borrione et al., 2020; Padberg et al., 2021).

In their review on personalizing non-invasive transcranial brain stimulation in psychiatry, Padberg and colleagues Padberg et al. (2021), suggested more in-depth analysis of available trial data aligned with comparable approaches in pharmacotherapy (Chekroud et al., 2017a) and rTMS research (Drysdale et al., 2017; Kaster et al., 2019; Siddiqi et al., 2020). Employing computational methods, these approaches look beyond average treatment-group-level analyses and aggregate symptom measures (see 1.2.2 Problems 1-2) by deriving more meaningful, data-driven units of analysis. Aside from these refinements on the outcome level, a shift of focus towards prediction as a two-step procedure, incorporating cross-validation (CV) frameworks as a gold standard to evaluate fitted models capacity to extrapolate predictions to unseen instances, may alleviate insufficient model validation and limited applicability in single-subject settings (see 1.2.2 Problem 3) (Friedman et al., 2001; Passos and Mwangi, 2018; Borrione et al., 2020; Padberg et al., 2021)

Based on the proposals by Padberg et al. (2021), a framework for in-depth clinical trial analysis was constructed. The next paragraphs give a brief summary on the methodological approaches within that framework, with direct reference to the above-described problems associated with traditional statistical methods (see 1.2.2). Thereafter, blueprint empirical studies for the implementation of each approach are presented.

Remedy Problem 1: Parsing Symptomatic Improvement

One approach to evaluate response to antidepressant treatment, but without ignoring differences in the time course of symptomatic improvement is to categorize patients based on their specific temporal patterns of change using latent variable methods (Beunckens et al., 2008; Muthén et al., 2008). Contrasting conventional methods of categorization (Montgomery, 1994; Frank et al., 1991), latent variable methods such as growth mixture modelling (GMM) capture heterogeneity as it naturally occurs in the observations by allowing individual variation of straight and curved trajectories (i.e. linear, quadratic, cubic) to identify different rates of change at different stages of the treatment process. In NIBS, this procedure has previously been adopted from pharmacological (Smagula et al., 2015) and psychotherapy studies (Uher et al., 2010) to identify distinct trajectories of change during standard and accelerated rTMS treatment (Kaster et al., 2019, 2020). The concept of parsing symptomatic improvement using latent variable modeling, as suggested by Padberg et al. (2021), is schematically summarized in Figure 1.1.

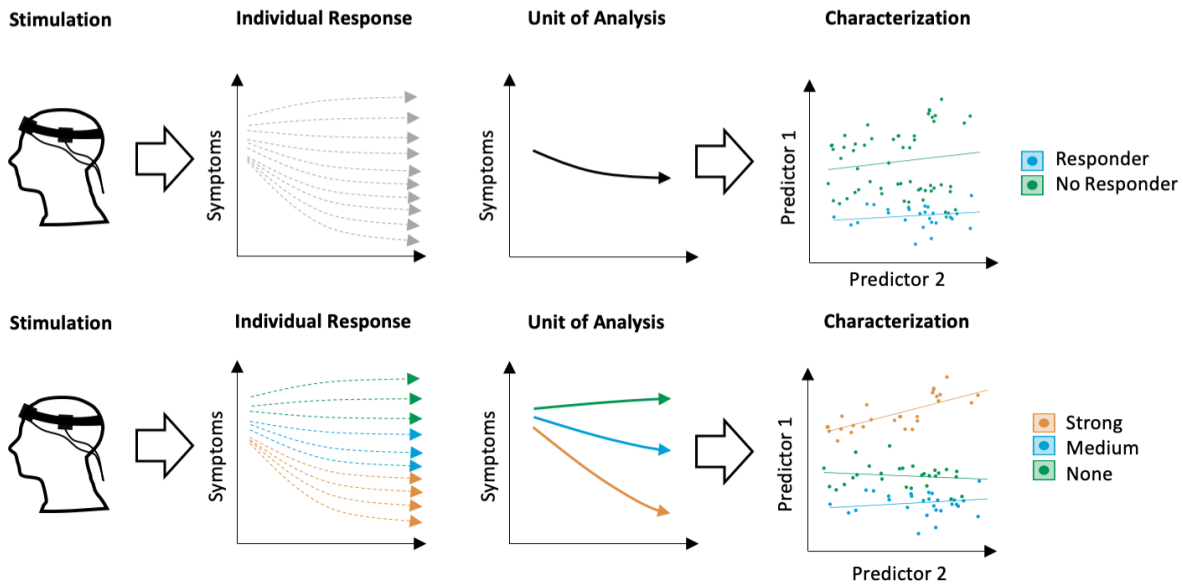

Figure 1.1: Upper row: Traditional RCT analysis: patient-specific change is condensed into one group-average trajectory resulting in imprecise representations of response, as patients' symptoms may change with different celerity and intensity; Consequently, responders can only be characterized based on associations with undifferentiated units of analysis. Lower row: Deriving of new units of analysis for precision medicine: Patients are grouped based on their longitudinal patterns of change (latent-class solution). Associated characteristics and predictors masked in the one-class solution can be explored more effectively. All displayed graphs are hypothetical, i.e. not based on real data. Adopted and modified from (Padberg et al., 2021).

Remedy Problem 2: Parsing Study Outcomes

Treatment effects on specific depressive symptoms may be undetectable, if they are analyzed after aggregation (e.g. sum scores), as is intended by most scale manuals and reporting guidelines (Kearns et al., 1982). However, evaluating effects on the individual symptom level may be prone to family-wise errors and impractical, since clinicians would have to remember procedures specific to each symptom (Chekroud et al., 2017a). Computational methods can be applied to establish natural clusters within the symptoms to restructure measures into data-driven sub-components. Traditional statistical approaches on dimensionality reduction have some shortcomings, including difficult-to-interpret constellations of symptoms groups with cross-loadings (e.g. in factor analysis, FA) (Uher et al., 2009) and susceptibility to experimenter bias when choosing the number of groups to retain (Williams et al., 2010). Unsupervised ML provides a subset of methods that supply deterministic, easy-to-visualize solutions, that assign each symptom to a fixed cluster while avoiding assumptions about the pre-specified number of clusters to keep (Murtagh and Contreras, 2012). In NIBS, this

procedure has previously been adopted from pharmacological and psychotherapy studies (Chekroud et al., 2017a; Bondar et al., 2020) to identify distinct symptom-specific treatment targets for circuit-based rTMS treatment (Siddiqi et al., 2019). The concept of parsing study outcomes using unsupervised ML techniques, as suggested by Padberg et al. (2021), is schematically summarized in Figure 1.2.

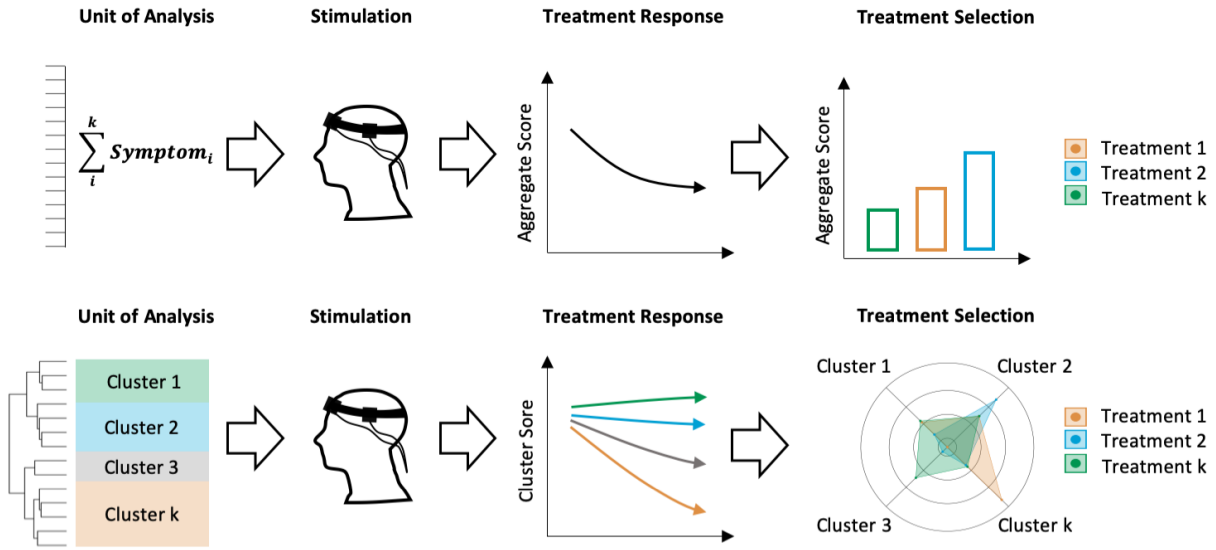

Figure 1.2: Upper row: Traditional RCT analysis: Symptom change is measured and reported in aggregated scale units (e.g. sum-scores), resulting in imprecise representations of phenotypes as information about response in natural symptom subgroups is masked; Consequently, treatments alternatives can only be selected based on overall severity ratings. Lower row: Derivation of new units of analysis for precision medicine: Symptoms are clustered based on the similarity of their responses (e.g. all sleep related components are grouped but remain independent of other domains). Treatment-induced change can be analyzed on the level of the cluster score. Treatment alternatives can be selected based on their response profiles in treating specific groups of symptoms. All displayed graphs are hypothetical, i.e. not based on real data. Adopted and modified from (Padberg et al., 2021).

Remedy Problem 3: Single-patient vs. Group-level Prediction of Treatment Response

For decades, the primary working unit of mental health research and evidence-based practice has been "the group". The primary data-modeling strategy was mostly association-based (explanatory), aiming to describe observational data by formally testing hypotheses about group-differences and factors related to better treatment response (Breiman et al., 2001a; Shmueli et al., 2010).

An alternative modeling strategy is algorithmic modeling or predictive modeling. It has a long-established focus on prediction as the primary criterion of statistical quality (Friedman et al., 2001). While the quality of explanatory models may be judged by using statistical concepts such as explained variance, significance, and effect sizes within the sample (in-sample-estimates), success in predictive modeling is quantified using estimates of accuracy, i.e. in successfully predicting category or value of previously unseen data points (out-of-sample estimates)(Breiman et al., 2001a), as is the gold-standard in translational science (Cannon et al., 2016; Carrion et al., 2016). Because predictive models can be applied to and obtain answers from single observations, they have the potential to bring the single patient as a new working unit into reach (Dwyer et al., 2018; Bzdok and Meyer-Lindenberg, 2018; Topol, 2019). Taking the form of a two-step procedure, predictive models are initially fit on an amount of available data (training set), then the generalizability of the trained model is evaluated by testing its capacity to extrapolate to new data (test set) (Friedman et al., 2001). This forward-oriented workflow yields a particular notion of clinical relevance and makes techniques for out-of-sample generalization of predictive models candidates for catalyzing progress in personalized tDCS treatments (Bzdok and Meyer-Lindenberg, 2018; Chekroud et al., 2017b).

Supervised ML algorithms, such as support vector machines (Cortes and Vapnik, 1995), decision trees (Breiman, 2001; Buntine, 1992), and neural-networks for deep learning (LeCun et al., 2015) are a promising set of tools for predictive modeling (Bzdok and Meyer-Lindenberg, 2018; Dwyer et al., 2018). While their approximation to the data often requires optimization of additional, learner-specific parameters (hyperparameter-tuning) (Feurer and Hutter, 2019), which can be computationally intensive, they have advantages that lend themselves particularly well to modeling of observational mental health data: They can handle high-dimensional input data (i.e. large numbers of predictors), are capable of identifying highly non-linear patterns within the observations (James et al., 2015), and can even handle multiple outcomes at once (multi-class prediction) (Aly, 2005). Importantly, within ML frameworks generalizability can be optimized and evaluated efficiently using resampling techniques via computer simulation (e.g. leave-one-out resampling, bootstrapping, k-fold cross validation). In these nested frameworks data can iteratively be split into training and test data. The average accuracy within all hold-out test sets is returned as an estimate for the out-of-sample generalizability (Stone, 1974; Filzmoser et al., 2009). The degree of required model generalizibility can be increased by selecting left-out folds that are more detached from the training data systematically (e.g. leave-site out, leave-continent out) (Koutsouleris et al., 2016) or temporally (e.g. prospective validation) (Dwyer et al., 2018). A prototypical procedure of training a predictive model on clinical trial data to predict patient-level response is displayed in Figure 1.3.

For summaries on the application of ML in psychiatry see the review articles by Dywer and colleagues Dwyer et al. (2018) and Bzdok and Meyer-Lindenberg Bzdok and Meyer-Lindenberg (2018). While there are numerous studies that have applied ML algorithms for pre-treatment prediction of therapeutic outcomes in MDD (for a review see Lee et al. (2018)), results from the field of non-invasive brain stimulation are scarce (Koutsouleris et al., 2018) and no studies have used clinical data to predict response to tDCS.

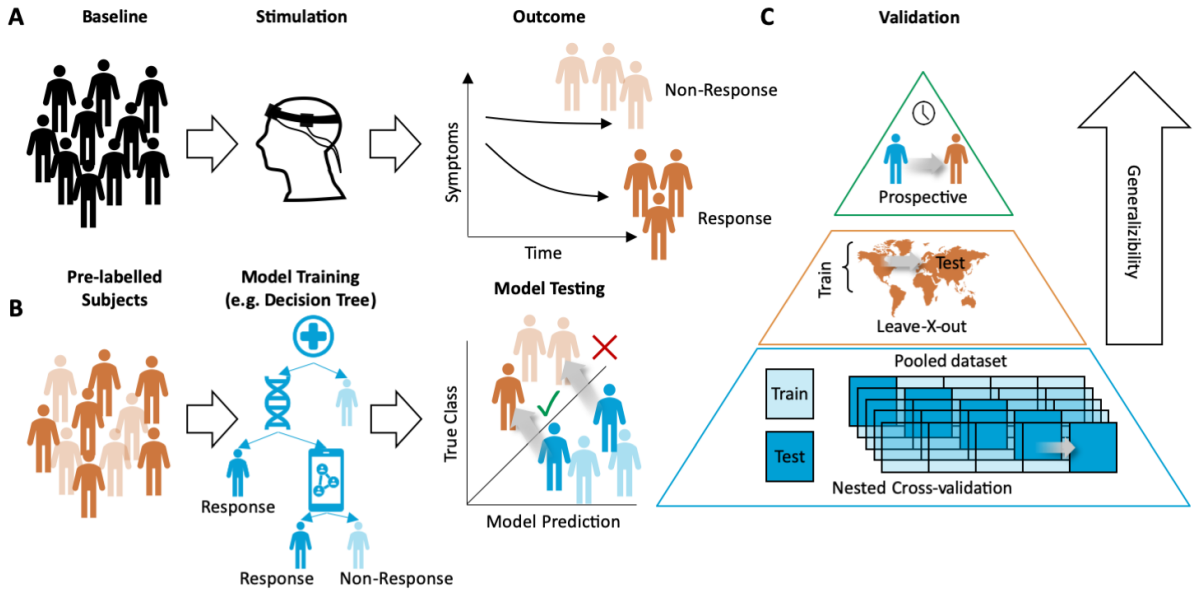

Figure 1.3: Prediction of treatment response on the single-patient level. (A) Following treatment with tDCS, patients are labeled as responders or non-responders. (B) Multimodal data can be used to train a predictive model (e.g. supervised ML model) to classify pre-labeled patients into responders and non-responders. This may include hyperparameter-tuning, weighing of observations, and sampling strategies to counter class-imbalance. Within a nested cross-validation (CV) paradigm, this step is embedded in the inner CV loop and is usually combined with preprocessing steps including imputation, scaling, and feature selection. Trained models are used to predict out-of-sample instances. Prediction takes place both in the inner CV loop validation folds and in the outer CV loop for unbiased performance evaluation. (C) Generalizability can be assessed by validating the model in increasingly diverse test sets. While in nested CV, one test fold consisting of randomly selected patients from the same dataset or trial is iteratively left out to predict unseen instances while training on the rest of the data, models can be evaluated in naturally left-out test sets (e.g. leave-site out validation). Prospective validation involves the employment of existing models to new subjects, i.e., in real life circumstances or in algorithm-informed clinical trials.

1.3 The Present Dissertation

1.3.1 Rationale

While tDCS has been shown to be a safe and effective treatment for MDD, its efficacy should be increased and heterogeneity in treatment response should be better understood. Up to this point, RCTs provide the largest available body of systematically collected data. However, they often remain under-analyzed with regard to differential clinical response on

the subgroup-level, patient-level, and symptom-level. This lack of predictability hinders prognosis in individual cases and may lead to misinformed treatment choices.

Precision psychiatry is a developing paradigm grounded in the notion of selecting the right treatment for the right patient at the right time, taking into account each person's variability in constituting factors and symptomatic burden. The present dissertation aims to provide a methodological framework for reevaluation of clinical trials by example of a paradigmatic analysis of ELECT-TDCS. By conducting more in-depth analyses of representative clinical data, the goals of the empirical studies embedded in the dissertation were three-fold:

The **first objective** was to gain a deeper understanding for the longitudinal dynamics of response within patients treated with tDCS. For this purpose, we aimed to identify distinct, naturally occurring trajectories of symptom improvement over the duration of treatment. By categorizing patients based on their longitudinal patterns of change, rather than at the least observed measurement, these trajectory classes would capture heterogeneity in development of response over the course of tDCS treatment (e.g. fast responding, slow responding, delayed responding). Following this approach, we aimed to formally test, whether longitudinal response categorization could add explanatory value to a single mean growth curve (i.e. challenge the assumption of homogeneous change, see 1.2.1) and to characterize the identified trajectory classes by exploring associated characteristics.

The **second objective** was to reevaluate the efficacy of tDCS in treating specific subgroups of depressive symptoms. For this goal, alternative outcome measures with a more precise capture of distinct symptom domains had to be derived, while avoiding statistical shortcomings of experimenter bias (e.g. number of clusters to retain), or difficult to interpret cross-loadings between domains. Unsupervised machine-learning was employed to establish a data-driven grouping of depressive symptoms, as assessed by items of a common aggregate measure (HAM-D) into depressive symptom clusters. Resulting clusters would be assessed for their robustness by comparing them to other clustering solutions on depressive symptoms following the same rationale. Based on these newly created outcomes, we aimed to test whether tDCS or an alternative antidepressant intervention could be chosen according to their effects on specific clusters of symptoms to achieve better overall response.

The **third objective** was to translate the question of who is going to benefit from tDCS to a predictive modeling paradigm. This objective was chosen, as there are currently no objective tools available to guide clinical decision making in the single-patient setting. Our goal was to use supervised ML algorithms trained on easy-to-obtain clinical data measured ahead of treatment to predict individual tDCS response at the end of treatment. To make most use of the available data while getting unbiased accuracy estimates, we chose state-of-the-art nested cross-validation frameworks for hyperparameter-optimization and model evaluation. Importantly, in this setup the steps of model development and model validation are strictly separated, thus mimicking the real clinical situation where existing models are used for prognosis of success over an individual course of treatment. Trained predictive models would then be made available as a collaborative research product. As supervised ML models have often been labeled as "black-box models", part of our objective was to incorporate general model-agnostic methods that make the creation of the ML

models and their decisions more interpretable.

To test the robustness of our findings following these three objectives and suspend the possibility that any uncovered patterns of tDCS response merely resemble general mechanisms of response to antidepressant treatment, we chose to conduct our analyses in a double-controlled design, including analysis of a second active pharmacological intervention with a distinct mode of action and analysis of a placebo group.

In detail, the three empirical studies presented in this dissertation contribute to the above-described goals as follows:

Study 1

In Study 1, distinct trajectories of tDCS response were characterized using latent trajectory modeling in the largest RCT (N = 245) for treating MDD to date (The Escitalopram versus Electric Current Therapy for Treating Depression Clinical Study, ELECT-TDCS, ClinicalTrials.gov, NCT01894815). The main objective was to gain insight into the dynamics of symptomatic change above the group level. The ELECT-TDCS trial enrolled antidepressant-free patients diagnosed with an acute unipolar, non-psychotic, depressive episode, and with a minimum baseline score of 17 points on the HAM-D Scale (17-items). Participants were randomized to a 10-week treatment with either prefrontal tDCS (22 sessions of 30 minutes, 2 mA bifrontal stimulation), an active pharmacotherapy (20 milligram/day escitalopram), or placebo. The resulting trajectory classes were representative of patient subgroups, who showed similar strength and speed of symptomatic improvement. Only groups with a clinically relevant capture of patients were accepted for further analysis. Assignment rates to the identified trajectory classes were compared both relative to patients, who received the same treatment, as well as relative to the entire sample. A selection of top-down (hypothesis-driven) predictors from previous literature were tested using multinomial logistic regression models with conservative correction for multiple testing. Additionally, bottom-up (data-driven) methods were employed to explore associated characteristics using a cross-validated stability ranking procedure combined with elastic net regularization.

Study 2

In Study 2, the issue of small antidepressant efficacy of tDCS and considerable inter-individual variability of response was reevaluated at the outcome level, assuming that aggregate measures might be insufficient to address the diverse spectrum of antidepressant treatment effects. Using data from the ELECT-TDCS clinical trial, which originally employed the HAM-D as the primary measurement of depressive symptoms, we applied unsupervised ML (agglomerative hierarchical clustering) to identify naturally occurring symptom clusters within the HAM-D rating scale. This approach followed a rationale that was previously applied in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial (N = 4039). Resulting symptom clusters were then compared to the STAR*D clustering solution which was derived in a larger sample and robust to replication in a second cohort (Combining medications to enhance depression outcomes, CO-MED, N =

640). Symptom clusters were tested for differences in their responsiveness to treatment both across and within the treatment arms, using multilevel modeling on within-cluster severity scores regularized for the number of items within each cluster, addressing the question, if any of the interventions were favourable for treating a specific group of symptoms.

Study 3

Study 3 is to our knowledge the first study to employ supervised ML for prediction of tDCS treatment response on the single-subject level. Using a comprehensive dataset of easily obtainable, clinical and neuropsychological variables from the ELECT-TDCS trial, extreme gradient tree boosting algorithms (XGBoost) were employed to predict response, defined as a minimum improvement of 50% on the HAM-D scale within each of the treatment groups (tDCS, escitalopram, placebo). Predictive models were trained and evaluated within a state-of-the-art framework of repeated nested cross-validation. To test the specificity of the ML models and to explore potential mechanistic differences between the interventions, we applied classifiers trained to predict escitalopram response on the tDCS data and vice-versa. The tDCS classifier was further validated in an external open-label phase at the end of the trial. Finally, interpretable machine learning (IML) methods were applied to explore the role of associated characteristics and to inspire future explanatory research.

1.3.2 Parts of the Dissertation and Author Contributions

Table 1.1: Publications in the dissertation and author contributions

Study	Publication	Author Contributions
1	**Goerigk, S.**, Padberg, F., Bühner, M., Sarubin, N., Kaster, TS., Daskalakis, ZJ., Blumberger, DM., Borrione, L., Razza, LB., and Brunoni, AR., (2021). Distinct Trajectories of Response to Prefrontal tDCS in Major Depression: Results from a 3-arm randomized controlled trial. Neuropsychopharmacology, 46(4), 774-782. https://doi.org/10.1038/s41386-020-00935-x	**SG**, **FP**, and **ARB** conceptualized the study, analyzed the data, and interpreted the results. **LB** and **LBR** were involved in data acquisition and interpretation of the findings. **MB**, **NS**, **TSK**, **ZJD**, and **DMB** were involved in the interpretation of the findings.
2	**Goerigk, S.**, Padberg, F., Chekroud, A., Kambeitz, J. Bühner, M., and Brunoni, AR., (2021). Parsing the Antidepressant Effects of Non-Invasive Brain Stimulation and Pharmacotherapy: a Symptom Clustering Approach on ELECT-TDCS. Brain Stimulation, 14(4), 906-912. https://doi.org/10.1016/j.brs.2021.05.008	**SG:** Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original Draft, Visualization **FP:** Conceptualization, Resources, Writing - Review and Editing, Supervision, Funding acquisition **AC:** Conceptualization, Methodology, Writing - Review and Editing **JK:** Conceptualization, Methodology, Writing - Review and Editing **MB:** Writing - Review and Editing, Supervision **ARB:** Conceptualization, Validation, Resources, Writing - Original Draft, Visualization, Supervision, Project administration, Funding acquisition
3	Kambeitz, J., **Goerigk, S.**, Gattaz W., Falkai P., Benseñor, I.M., Lotufo P.A., Bühner M., Koutsouleris N., Padberg, F., Brunoni, A.R., (2020), Clinical patterns differentially predict response to transcranial direct current stimulation (tDCS) and escitalopram in major depression: a machine learning analysis of the ELECT-TDCS study. Journal of affective disorders, 265, 460-467. https://doi.org/10.1016/j.jad.2020.01.118	**JK:** Conceptualization, Formal analysis, Writing - original draft, Writing - review and editing. **SG:** Conceptualization, Formal analysis, Writing - original draft, Writing - review and editing. **WG:** Project administration, Writing - original draft, Writing - review and editing. **PF:** Writing - original draft, Writing - review and editing. **IMB:** Funding acquisition, Writing - original draft, Writing - review and editing. **PAL:** Funding acquisition, Writing - original draft, Writing - review and editing. **MB:** Writing - original draft, Writing - review and editing. **NK:** Funding acquisition, Project administration, Writing - original draft, Writing - review and editing. **FP:** Funding acquisition, Project administration, Writing - original draft, Writing - review and editing. **ARB:** Conceptualization, Funding acquisition, Project administration, Writing - original draft, Writing - review and editing.

Note. Contributions of the author of this dissertation are in bold. Formatting of contributions in Study 1 follows the ICMJE best practice recommendations as required by Springer Nature (www.icmje.org); Contributions in Studies 2 and 3 are organized according to the Elsevier CRediT author statement guidelines (https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement).

Table 1.1 lists the empirical studies that contribute to this dissertation. All authors have contributed significantly to the research presented in the articles. The right column of Table 1.1 shows the respective individual contributions.

1.3.3 Open Science Statement

While patient-sensitive data were not acquired with consent to be uploaded to a public repository, for the purpose of replicability of results each article contains a data availability section, stating that data can obtained upon reasonable request from the corresponding authors. Codes for supervised ML algorithms are deposited on github: `https://github.com/biolpsychlab/Predict_TDCS`.

Bibliography

Abi-Dargham, A. and Horga, G. (2016). The search for imaging biomarkers in psychiatric disorders. *Nature medicine*, 22(11):1248.

Aly, M. (2005). Survey on multiclass classification methods. *Neural Netw*, 19:1–9.

Arbabshirani, M. R., Plis, S., Sui, J., and Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145:137–165.

Bajbouj, M., Aust, S., Spies, J., Herrera-Melendez, A.-L., Mayer, S. V., Peters, M., Plewnia, C., Fallgatter, A. J., Frase, L., Normann, C., et al. (2018). Psychotherapyplus: augmentation of cognitive behavioral therapy (cbt) with prefrontal transcranial direct current stimulation (tdcs) in major depressive disorder—study design and methodology of a multicenter double-blind randomized placebo-controlled trial. *European archives of psychiatry and clinical neuroscience*, 268(8):797–808.

Bajbouj, M. and Padberg, F. (2014). A perfect match: noninvasive brain stimulation and psychotherapy. *European archives of psychiatry and clinical neuroscience*, 264(1):27–33.

Barker, A. T., Jalinous, R., and Freeston, I. L. (1985). Non-invasive magnetic stimulation of human motor cortex. *The Lancet*, 325(8437):1106–1107.

Beck, A. T., Ward, C., Mendelson, M., Mock, J., and Erbaugh, J. (1961). Beck depression inventory (bdi). *Arch Gen Psychiatry*, 4(6):561–571.

Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal gaussian data. *Biometrics*, 64(1):96–105.

Bikson, M., Grossman, P., Thomas, C., Zannou, A. L., Jiang, J., Adnan, T., Mourdoukoutas, A. P., Kronberg, G., Truong, D., Boggio, P., et al. (2016). Safety of transcranial direct current stimulation: evidence based update 2016. *Brain stimulation*, 9(5):641–661.

Boggio, P. S., Rigonatti, S. P., Ribeiro, R. B., Myczkowski, M. L., Nitsche, M. A., Pascual-Leone, A., and Fregni, F. (2008). A randomized, double-blind clinical trial on the efficacy of cortical direct current stimulation for the treatment of major depression. *International Journal of Neuropsychopharmacology*, 11(2):249–254.

Bondar, J., Caye, A., Chekroud, A. M., and Kieling, C. (2020). Symptom clusters in adolescent depression and differential response to treatment: a secondary analysis of the treatment for adolescents with depression study randomised trial. *The Lancet Psychiatry*, 7(4):337–343.

Borrione, L., Bellini, H., Razza, L. B., Avila, A. G., Baeken, C., Brem, A.-K., Busatto, G., Carvalho, A. F., Chekroud, A., Daskalakis, Z. J., et al. (2020). Precision non-implantable neuromodulation therapies: a perspective for the depressed brain. *Brazilian Journal of Psychiatry*, (AHEAD).

Borrione, L. and Brunoni, A. R. (2019). Precision noninvasive brain stimulation: is it precise? is it needed? *Brazilian Journal of Psychiatry*, 41(5):376–377.

Brakemeier, E.-L., Luborzewski, A., Danker-Hopfe, H., Kathmann, N., and Bajbouj, M. (2007). Positive predictors for antidepressive response to prefrontal repetitive transcranial magnetic stimulation (rtms). *Journal of psychiatric research*, 41(5):395–403.

Brakemeier, E.-L., Wilbertz, G., Rodax, S., Danker-Hopfe, H., Zinka, B., Zwanzger, P., Grossheinrich, N., Várkuti, B., Rupprecht, R., Bajbouj, M., et al. (2008). Patterns of response to repetitive transcranial magnetic stimulation (rtms) in major depression: replication study in drug-free patients. *Journal of affective disorders*, 108(1-2):59–70.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Cox, D., and Breiman, L. (2001a). Comment-statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231.

Breiman, L. et al. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.

Brunoni, A., Boggio, P., De Raedt, R., Benseñor, I., Lotufo, P., Namur, V., Valiengo, L., and Vanderhasselt, M.-A. (2014). Cognitive control therapy and transcranial direct current stimulation for depression: a randomized, double-blinded, controlled trial. *Journal of affective disorders*, 162:43–49.

Brunoni, A., Kemp, A., Shiozawa, P., Cordeiro, Q., Valiengo, L., Goulart, A., Coprerski, B., Lotufo, P., Brunoni, D., Perez, A., et al. (2013a). Impact of 5-httlpr and bdnf polymorphisms on response to sertraline versus transcranial direct current stimulation: implications for the serotonergic system. *European Neuropsychopharmacology*, 23(11):1530–1540.

Brunoni, A. R., Chaimani, A., Moffa, A. H., Razza, L. B., Gattaz, W. F., Daskalakis, Z. J., and Carvalho, A. F. (2017a). Repetitive transcranial magnetic stimulation for the acute treatment of major depressive episodes: a systematic review with network meta-analysis. *JAMA psychiatry*, 74(2):143–152.

Brunoni, A. R., Kemp, A. H., Dantas, E. M., Goulart, A. C., Nunes, M. A., Boggio, P. S., Mill, J. G., Lotufo, P. A., Fregni, F., and Benseñor, I. M. (2013b). Heart rate variability is a trait marker of major depressive disorder: evidence from the sertraline vs. electric current therapy to treat depression clinical study. *International Journal of Neuropsychopharmacology*, 16(9):1937–1949.

Brunoni, A. R., Moffa, A. H., Fregni, F., Palm, U., Padberg, F., Blumberger, D. M., Daskalakis, Z. J., Bennabi, D., Haffen, E., Alonzo, A., et al. (2016). Transcranial direct current stimulation for acute major depressive episodes: meta-analysis of individual patient data. *The British Journal of Psychiatry*, 208(6):522–531.

Brunoni, A. R., Moffa, A. H., Sampaio-Junior, B., Borrione, L., Moreno, M. L., Fernandes, R. A., Veronezi, B. P., Nogueira, B. S., Aparicio, L. V., Razza, L. B., et al. (2017b). Trial of electrical direct-current therapy versus escitalopram for depression. *New England Journal of Medicine*, 376(26):2523–2533.

Brunoni, A. R., Nitsche, M. A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., Edwards, D. J., Valero-Cabre, A., Rotenberg, A., Pascual-Leone, A., et al. (2012). Clinical research with transcranial direct current stimulation (tdcs): challenges and future directions. *Brain stimulation*, 5(3):175–195.

Brunoni, A. R., Padberg, F., Vieira, E. L. M., Teixeira, A. L., Carvalho, A. F., Lotufo, P. A., Gattaz, W. F., and Benseñor, I. M. (2018). Plasma biomarkers in a placebo-controlled trial comparing tdcs and escitalopram efficacy in major depression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 86:211–217.

Brunoni, A. R., Valiengo, L., Baccaro, A., Zanão, T. A., de Oliveira, J. F., Goulart, A., Boggio, P. S., Lotufo, P. A., Benseñor, I. M., and Fregni, F. (2013c). The sertraline vs electrical current therapy for treating depression clinical study: results from a factorial, randomized, controlled trial. *JAMA psychiatry*, 70(4):383–391.

Bulubas, L., Padberg, F., Bueno, P. V., Duran, F., Busatto, G., Amaro Jr, E., Benseñor, I. M., Lotufo, P. A., Goerigk, S., Gattaz, W., et al. (2019). Antidepressant effects of tdcs are associated with prefrontal gray matter volumes at baseline: Evidence from the elect-tdcs trial. *Brain stimulation*, 12(5):1197–1204.

Buntine, W. (1992). Learning classification trees. *Statistics and computing*, 2(2):63–73.

Bzdok, D. (2017). Classical statistics and statistical learning in imaging neuroscience. *Frontiers in neuroscience*, 11:543.

Bzdok, D. and Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3):223–230.

Bzdok, D. and Yeo, B. T. (2017). Inference in the age of big data: Future perspectives on neuroscience. *Neuroimage*, 155:549–564.

Cannon, T. D., Yu, C., Addington, J., Bearden, C. E., Cadenhead, K. S., Cornblatt, B. A., Heinssen, R., Jeffries, C. D., Mathalon, D. H., McGlashan, T. H., et al. (2016). An individualized risk calculator for research in prodromal psychosis. *American Journal of Psychiatry*, 173(10):980–988.

Carrion, R. E., Cornblatt, B. A., Burton, C. Z., Tso, I. F., Auther, A. M., Adelsheim, S., Calkins, R., Carter, C. S., Niendam, T., Sale, T. G., et al. (2016). Personalized prediction of psychosis: external validation of the napls-2 psychosis risk calculator with the edippp project. *American Journal of Psychiatry*, 173(10):989–996.

Carvalho, A. F., Sharma, M. S., Brunoni, A. R., Vieta, E., and Fava, G. A. (2016). The safety, tolerability and risks associated with the use of newer generation antidepressant drugs: a critical review of the literature. *Psychotherapy and psychosomatics*, 85(5):270–288.

Chan, M. M., Yau, S. S., and Han, Y. M. (2021). The neurobiology of prefrontal transcranial direct current stimulation (tdcs) in promoting brain plasticity: a systematic review and meta-analyses of human and rodent studies. *Neuroscience & Biobehavioral Reviews*.

Chekroud, A. M., Gueorguieva, R., Krumholz, H. M., Trivedi, M. H., Krystal, J. H., and McCarthy, G. (2017a). Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA psychiatry*, 74(4):370–378.

Chekroud, A. M., Lane, C. E., and Ross, D. A. (2017b). Computational psychiatry: embracing uncertainty and focusing on individuals, not averages. *Biological psychiatry*, 82(6):e45–e47.

Cipriani, A., Furukawa, T. A., Salanti, G., Geddes, J. R., Higgins, J. P., Churchill, R., Watanabe, N., Nakagawa, A., Omori, I. M., McGuire, H., et al. (2009). Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *The lancet*, 373(9665):746–758.

Cnaan, A., Laird, N. M., and Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in medicine*, 16(20):2349–2380.

Collins, F. S. and Varmus, H. (2015). A new initiative on precision medicine. *New England journal of medicine*, 372(9):793–795.

Connolly, K. R., Helmer, A., Cristancho, M. A., Cristancho, P., and O'Reardon, J. P. (2012). Effectiveness of transcranial magnetic stimulation in clinical practice post-fda approval in the united states: results observed with the first 100 consecutive cases of depression at an academic medical center. *The Journal of clinical psychiatry*, 73(4):567–573.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Cuijpers, P. (2017). Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses. *Canadian Psychology/psychologie canadienne*, 58(1):7.

Cuijpers, P., Ebert, D. D., Acarturk, C., Andersson, G., and Cristea, I. A. (2016). Personalized psychotherapy for adult depression: A meta-analytic review. *Behavior Therapy*, 47(6):966–980.

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., and van Straten, A. (2011). Interpersonal psychotherapy for depression: a meta-analysis. *American Journal of Psychiatry*, 168(6):581–592.

Dayan, E., Censor, N., Buch, E. R., Sandrini, M., and Cohen, L. G. (2013). Noninvasive brain stimulation: from physiology to network dynamics and back. *Nature neuroscience*, 16(7):838–844.

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R. N., Zebley, B., Oathes, D. J., Etkin, A., et al. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine*, 23(1):28–38.

Dunlop, K., Hanlon, C. A., and Downar, J. (2017). Noninvasive brain stimulation treatments for addiction and major depression. *Annals of the New York Academy of Sciences*, 1394(1):31.

Dwyer, D. B., Falkai, P., and Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual review of clinical psychology*, 14:91–118.

Fava, M., Uebelacker, L. A., Alpert, J. E., Nierenberg, A. A., Pava, J. A., and Rosenbaum, J. F. (1997). Major depressive subtypes and treatment response. *Biological psychiatry*, 42(7):568–576.

Fernandes, B. S., Williams, L. M., Steiner, J., Leboyer, M., Carvalho, A. F., and Berk, M. (2017). The new field of 'precision psychiatry'. *BMC medicine*, 15(1):1–7.

Feurer, M. and Hutter, F. (2019). Hyperparameter optimization. In *Automated Machine Learning*, pages 3–33. Springer, Cham.

Filzmoser, P., Liebmann, B., and Varmuza, K. (2009). Repeated double cross validation. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 23(4):160–171.

Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., Rush, A. J., and Weissman, M. M. (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Archives of general psychiatry*, 48(9):851–855.

Fregni, F., Boggio, P. S., Nitsche, M. A., Marcolin, M. A., Rigonatti, S. P., and Pascual-Leone, A. (2006). Treatment of major depression with transcranial direct current stimulation.

Fregni, F., El-Hagrassy, M. M., Pacheco-Barrios, K., Carvalho, S., Leite, J., Simis, M., Brunelin, J., Nakamura-Palacios, E. M., Marangolo, P., Venkatasubramanian, G., et al. (2020). Evidence-based guidelines and secondary meta-analysis for the use of transcranial direct current stimulation (tdcs) in neurological and psychiatric disorders. *International Journal of Neuropsychopharmacology*.

Fried, E. I. and Nesse, R. M. (2015). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the star* d study. *Journal of affective disorders*, 172:96–102.

Friedman, J., Hastie, T., Tibshirani, R., et al. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Friston, K. J. (2017). Precision psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2(8):640–643.

Fritsch, B., Reis, J., Martinowich, K., Schambra, H. M., Ji, Y., Cohen, L. G., and Lu, B. (2010). Direct current stimulation promotes bdnf-dependent synaptic plasticity: potential implications for motor learning. *Neuron*, 66(2):198–204.

Galbraith, J., Moustaki, I., Bartholomew, D. J., and Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. Crc Press.

Gandiga, P. C., Hummel, F. C., and Cohen, L. G. (2006). Transcranial dc stimulation (tdcs): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clinical neurophysiology*, 117(4):845–850.

Gartlehner, G., Hansen, R. A., Morgan, L. C., Thaler, K., Lux, L., Van Noord, M., Mager, U., Thieda, P., Gaynes, B. N., Wilkins, T., et al. (2011). Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Annals of internal medicine*, 155(11):772–785.

Goerigk, S., Cretaz, E., Sampaio-Junior, B., Vieira, É. L. M., Gattaz, W., Klein, I., Lafer, B., Teixeira, A. L., Carvalho, A. F., Lotufo, P. A., et al. (2021). Effects of tdcs on neuroplasticity and inflammatory biomarkers in bipolar depression: Results from a sham-controlled study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 105:110119.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning cambridge.

Goodman, S. N. (1992). A comment on replication, p-values and evidence. *Statistics in medicine*, 11(7):875–879.

Guerra, A., López-Alonso, V., Cheeran, B., and Suppa, A. (2020). Solutions for managing variability in non-invasive brain stimulation studies. *Neuroscience letters*, 719:133332.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry*, 23(1):56–62.

Härter, M., Berger, M., Schneider, F., and Ollenschläger, G. (2010). *S3 Praxisleitlinien in Psychiatrie und Psychotherapie: nationale VersorgungsLeitlinie Unipolare Depression*. Springer.

Hodson, R. (2016). Precision medicine. *Nature*, 537(7619):S49–S49.

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., and Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive therapy and research*, 36(5):427–440.

Holtzheimer III, P. E., Russo, J., Claypoole, K. H., Roy-Byrne, P., and Avery, D. H. (2004). Shorter duration of depressive episode may predict response to repetitive transcranial magnetic stimulation. *Depression and anxiety*, 19(1):24–30.

Homan, S., Muscat, W., Joanlanne, A., Marousis, N., Cecere, G., Hofmann, L., Ji, E., Neumeier, M., Vetter, S., Seifritz, E., et al. (2021). Treatment effect variability in brain stimulation across psychiatric disorders: A meta-analysis of variance. *Neuroscience & Biobehavioral Reviews*.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8):e124.

Ioannidis, J. P. (2016). Why most clinical research is not useful. *PLoS medicine*, 13(6):e1002049.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2015). An introduction to statistical learning with applications in r (springer-verlag new york).

Jameson, J. L. and Longo, D. L. (2015). Precision medicine—personalized, problematic, and promising. *Obstetrical & gynecological survey*, 70(10):612–614.

Jog, M. S., Kim, E., Anderson, C., Kubicki, A., Kayathi, R., Jann, K., Yan, L., Leaver, A., Hellemann, G., Iacoboni, M., et al. (2021). In-vivo imaging of targeting and modulation of depression-relevant circuitry by transcranial direct current stimulation: a randomized clinical trial. *Translational psychiatry*, 11(1):1–13.

Kaster, T. S., Chen, L., Daskalakis, Z. J., Hoy, K. E., Blumberger, D. M., and Fitzgerald, P. B. (2020). Depressive symptom trajectories associated with standard and accelerated rtms. *Brain stimulation*, 13(3):850–857.

Kaster, T. S., Downar, J., Vila-Rodriguez, F., Thorpe, K. E., Feffer, K., Noda, Y., Giacobbe, P., Knyahnytska, Y., Kennedy, S. H., Lam, R. W., et al. (2019). Trajectories of response to dorsolateral prefrontal rtms in major depression: a three-d study. *American Journal of Psychiatry*, 176(5):367–375.

Kearns, N. P., Cruickshank, C., McGuigan, K., Riley, S., Shaw, S. P., and Snaith, R. (1982). A comparison of depression rating scales. *The British Journal of Psychiatry*, 141(1):45–49.

Keeser, D., Padberg, F., Reisinger, E., Pogarell, O., Kirsch, V., Palm, U., Karch, S., Möller, H.-J., Nitsche, M., and Mulert, C. (2011). Prefrontal direct current stimulation modulates resting eeg and event-related potentials in healthy subjects: a standardized low resolution tomography (sloreta) study. *Neuroimage*, 55(2):644–657.

Keller, M. B., McCullough, J. P., Klein, D. N., Arnow, B., Dunner, D. L., Gelenberg, A. J., Markowitz, J. C., Nemeroff, C. B., Russell, J. M., Thase, M. E., et al. (2000). A

comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine*, 342(20):1462–1470.

Kennedy, S. H., Lam, R. W., McIntyre, R. S., Tourjman, S. V., Bhat, V., Blier, P., Hasnain, M., Jollant, F., Levitt, A. J., MacQueen, G. M., et al. (2016). Canadian network for mood and anxiety treatments (canmat) 2016 clinical guidelines for the management of adults with major depressive disorder: section 3. pharmacological treatments. *The Canadian Journal of Psychiatry*, 61(9):540–560.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., and Wang, P. S. (2003). The epidemiology of major depressive disorder: results from the national comorbidity survey replication (ncs-r). *Jama*, 289(23):3095–3105.

Koutsouleris, N., Kahn, R. S., Chekroud, A. M., Leucht, S., Falkai, P., Wobrock, T., Derks, E. M., Fleischhacker, W. W., and Hasan, A. (2016). Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *The Lancet Psychiatry*, 3(10):935–946.

Koutsouleris, N., Wobrock, T., Guse, B., Langguth, B., Landgrebe, M., Eichhammer, P., Frank, E., Cordes, J., Wölwer, W., Musso, F., et al. (2018). Predicting response to repetitive transcranial magnetic stimulation in patients with schizophrenia using structural magnetic resonance imaging: a multisite machine learning analysis. *Schizophrenia bulletin*, 44(5):1021–1034.

Kupfer, D. J., Frank, E., and Phillips, M. L. (2016). Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *Focus*, 14(2):266–276.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Lee, Y., Ragguett, R.-M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., Brietzke, E., Lin, K., Pan, Z., Subramaniapillai, M., et al. (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *Journal of affective disorders*, 241:519–532.

Lefaucheur, J.-P., André-Obadia, N., Antal, A., Ayache, S. S., Baeken, C., Benninger, D. H., Cantello, R. M., Cincotta, M., de Carvalho, M., De Ridder, D., et al. (2014). Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rtms). *Clinical Neurophysiology*, 125(11):2150–2206.

Li, Y., Aggen, S., Shi, S., Gao, J., Tao, M., Zhang, K., Wang, X., Gao, C., Yang, L., Liu, Y., et al. (2014). The structure of the symptoms of major depression: exploratory and confirmatory factor analysis in depressed han chinese women. *Psychological medicine*, 44(7):1391–1401.

Lin, S. Y. and Stevens, M. B. (2014). The symptom cluster-based approach to individualize patient-centered treatment for major depression. *The Journal of the American Board of Family Medicine*, 27(1):151–159.

Liu, Q., He, H., Yang, J., Feng, X., Zhao, F., and Lyu, J. (2020). Changes in the global burden of depression from 1990 to 2017: Findings from the global burden of disease study. *Journal of psychiatric research*, 126:134–140.

Loo, C. K., Husain, M. M., McDonald, W. M., Aaronson, S., O'Reardon, J. P., Alonzo, A., Weickert, C. S., Martin, D. M., McClintock, S. M., Mohan, A., et al. (2018). International randomized-controlled trial of transcranial direct current stimulation in depression. *Brain stimulation*, 11(1):125–133.

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., and Murray, C. J. (2006). *Global burden of disease and risk factors*. The World Bank.

Martin, D. M., Alonzo, A., Mitchell, P. B., Sachdev, P., Gálvez, V., and Loo, C. K. (2011). Fronto-extracephalic transcranial direct current stimulation as a treatment for major depression: an open-label pilot study. *Journal of affective disorders*, 134(1-3):459–463.

Mayberg, H. S., Brannan, S. K., Tekell, J. L., Silva, J. A., Mahurin, R. K., McGinnis, S., and Jerabek, P. A. (2000). Regional metabolic effects of fluoxetine in major depression: serial changes and relationship to clinical response. *Biological psychiatry*, 48(8):830–843.

McTeague, L. M., Huemer, J., Carreon, D. M., Jiang, Y., Eickhoff, S. B., and Etkin, A. (2017). Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. *American Journal of Psychiatry*, 174(7):676–685.

Menke, A. (2018). Precision pharmacotherapy: psychiatry's future direction in preventing, diagnosing, and treating mental disorders. *Pharmacogenomics and personalized medicine*, 11:211.

Mezger, E., Rauchmann, B.-S., Brunoni, A. R., Bulubas, L., Thielscher, A., Werle, J., Mortazavi, M., Karali, T., Stöcklein, S., Ertl-Wagner, B., et al. (2021). Effects of bifrontal transcranial direct current stimulation on brain glutamate levels and resting state connectivity: multimodal mri data for the cathodal stimulation site. *European archives of psychiatry and clinical neuroscience*, 271(1):111–122.

Moffa, A. H., Martin, D., Alonzo, A., Bennabi, D., Blumberger, D. M., Benseñor, I. M., Daskalakis, Z., Fregni, F., Haffen, E., Lisanby, S. H., et al. (2020). Efficacy and acceptability of transcranial direct current stimulation (tdcs) for major depressive disorder: An individual patient data meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 99:109836.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Montgomery, S. A. (1994). Clinically relevant effect sizes in depression. *European Neuropsychopharmacology*, 4(3):283–284.

Montgomery, S. A. and Åsberg, M. (1979). A new depression scale designed to be sensitive to change. *The British journal of psychiatry*, 134(4):382–389.

Murray, C. J., Lopez, A. D., Organization, W. H., et al. (1996). *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary*. World Health Organization.

Murtagh, F. and Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97.

Musil, R., Seemüller, F., Meyer, S., Spellmann, I., Adli, M., Bauer, M., Kronmüller, K.-T., Brieger, P., Laux, G., Bender, W., et al. (2018). Subtypes of depression and their overlap in a naturalistic inpatient sample of major depressive disorder. *International journal of methods in psychiatric research*, 27(1):e1569.

Muthén, B., Brown, H., Leuchter, A., and Hunter, A. (2008). General approaches to analysis of course: Applying growth mixture modeling to randomized trials of depression medication. *Causality and psychopathology: Finding the determinants of disorders and their cures. Washington, DC: American Psychiatric Publishing*, pages 159–78.

Mutz, J., Vipulananthan, V., Carter, B., Hurlemann, R., Fu, C. H., and Young, A. H. (2019). Comparative efficacy and acceptability of non-surgical brain stimulation for the acute treatment of major depressive episodes in adults: systematic review and network meta-analysis. *bmj*, 364.

Nitsche, M. A., Boggio, P. S., Fregni, F., and Pascual-Leone, A. (2009). Treatment of depression with transcranial direct current stimulation (tdcs): a review. *Experimental neurology*, 219(1):14–19.

Nitsche, M. A. and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of physiology*, 527(3):633–639.

Nord, C. L., Halahakoon, D. C., Limbachya, T., Charpentier, C., Lally, N., Walsh, V., Leibowitz, J., Pilling, S., and Roiser, J. P. (2019). Neural predictors of treatment response to brain stimulation and psychological therapy in depression: a double-blind randomized controlled trial. *Neuropsychopharmacology*, 44(9):1613–1622.

Nuzzo, R. (2014). Scientific method: statistical errors. *Nature News*, 506(7487):150.

Olbert, C. M., Gala, G. J., and Tupler, L. A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. *Journal of Abnormal Psychology*, 123(2):452.

Padberg, F., Bulubas, L., Mizutani-Tiebel, Y., Burkhardt, G., Kranz, G. S., Koutsouleris, N., Kambeitz, J., Hasan, A., Takahashi, S., Keeser, D., et al. (2021). The intervention, the patient and the illness–personalizing non-invasive brain stimulation in psychiatry. *Experimental Neurology*, 341:113713.

Palm, U., Hasan, A., Strube, W., and Padberg, F. (2016). tdcs for the treatment of depression: a comprehensive review. *European archives of psychiatry and clinical neuroscience*, 266(8):681–694.

Parikh, S. V., Quilty, L. C., Ravitz, P., Rosenbluth, M., Pavlova, B., Grigoriadis, S., Velyvis, V., Kennedy, S. H., Lam, R. W., MacQueen, G. M., et al. (2016). Canadian network for mood and anxiety treatments (canmat) 2016 clinical guidelines for the management of adults with major depressive disorder: section 2. psychological treatments. *The Canadian Journal of Psychiatry*, 61(9):524–539.

Passos, I. C. and Mwangi, B. (2018). Machine learning-guided intervention trials to predict treatment response at an individual patient level: an important second step following randomized clinical trials. *Mol Psychiatry*.

Perna, G. and Nemeroff, C. B. (2017). Personalized medicine in psychiatry: back to the future. *Personalized medicine in psychiatry*, 1:1.

Purpura, D. P. and McMurtry, J. G. (1965). Intracellular activities and evoked potential changes during polarization of motor cortex. *Journal of neurophysiology*, 28(1):166–185.

Quitkin, F. M., Rabkin, J. G., Ross, D., and Stewart, J. W. (1984). Identification of true drug response to antidepressants: use of pattern analysis. *Archives of general psychiatry*, 41(8):782–786.

Razza, L. B., Palumbo, P., Moffa, A. H., Carvalho, A. F., Solmi, M., Loo, C. K., and Brunoni, A. R. (2020). A systematic review and meta-analysis on the effects of transcranial direct current stimulation in depressive episodes. *Depression and Anxiety*, 37(7):594–608.

Romera, I., Delgado-Cohen, H., Perez, T., Caballero, L., and Gilaberte, I. (2008). Factor analysis of the zung self-rating depression scale in a large sample of patients with major depressive disorder in primary care. *BMC psychiatry*, 8(1):1–8.

Rosa, M. A. and Lisanby, S. H. (2012). Somatic treatments for mood disorders. *Neuropsychopharmacology*, 37(1):102–116.

Royston, P., Altman, D. G., and Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*, 25(1):127–141.

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., Niederehe, G., Thase, M. E., Lavori, P. W., Lebowitz, B. D., et al. (2006a). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a star* d report. *American Journal of Psychiatry*, 163(11):1905–1917.

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Stewart, J. W., Nierenberg, A. A., Thase, M. E., Ritz, L., Biggs, M. M., Warden, D., Luther, J. F., et al. (2006b). Bupropion-sr, sertraline, or venlafaxine-xr after failure of ssris for depression. *New England Journal of Medicine*, 354(12):1231–1242.

Schestatsky, P., Morales-Quezada, L., and Fregni, F. (2013). Simultaneous eeg monitoring during transcranial direct current stimulation. *Journal of visualized experiments: JoVE*, (76).

Senn, S. (2018). Statistical pitfalls of personalized medicine.

Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, ces-d, hamilton, and zung. *Journal of clinical psychology*, 62(1):123–146.

Shmueli, G. et al. (2010). To explain or to predict? *Statistical science*, 25(3):289–310.

Siddiqi, S. H., Taylor, S., Cooke, D., Pascual-Leone, A., George, M. S., and Fox, M. D. (2019). Distinct symptom-specific treatment targets for circuit-based neuromodulation. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, 12(4):e138.

Siddiqi, S. H., Taylor, S. F., Cooke, D., Pascual-Leone, A., George, M. S., and Fox, M. D. (2020). Distinct symptom-specific treatment targets for circuit-based neuromodulation. *American Journal of Psychiatry*, 177(5):435–446.

Smagula, S. F., Butters, M. A., Anderson, S. J., Lenze, E. J., Dew, M. A., Mulsant, B. H., Lotrich, F. E., Aizenstein, H., and Reynolds, C. F. (2015). Antidepressant response trajectories and associated clinical prognostic factors among older adults. *JAMA psychiatry*, 72(10):1021–1028.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–133.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *Plos med*, 12(3):e1001779.

Suen, P. J., Doll, S., Batistuzzo, M. C., Busatto, G., Razza, L. B., Padberg, F., Mezger, E., Bulubas, L., Keeser, D., Deng, Z.-D., et al. (2021). Association between tdcs computational modeling and clinical outcomes in depression: data from the elect-tdcs trial. *European archives of psychiatry and clinical neuroscience*, 271(1):101–110.

Thielscher, A., Opitz, A., and Windhoff, M. (2011). Impact of the gyral geometry on the electric field induced by transcranial magnetic stimulation. *Neuroimage*, 54(1):234–243.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al. (2014). The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior*, 8(2):153–182.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1):44–56.

Uher, R., Maier, W., Hauser, J., Marušič, A., Schmael, C., Mors, O., Henigsberg, N., Souery, D., Placentino, A., Rietschel, M., et al. (2009). Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. *The British Journal of Psychiatry*, 194(3):252–259.

Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., Petrovic, A., Zobel, A., Henigsberg, N., Rietschel, M., et al. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological medicine*, 40(8):1367.

Williams, B., Onsman, A., and Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian journal of paramedicine*, 8(3).

Williams, L. M. (2016). Precision psychiatry: a neural circuit taxonomy for depression and anxiety. *The Lancet Psychiatry*, 3(5):472–480.

Wolkenstein, L. and Plewnia, C. (2013). Amelioration of cognitive control in depression by transcranial direct current stimulation. *Biological psychiatry*, 73(7):646–651.

Wong, E. H., Yocca, F., Smith, M. A., and Lee, C.-M. (2010). Challenges and opportunities for drug discovery in psychiatric disorders: the drug hunters' perspective. *International Journal of Neuropsychopharmacology*, 13(9):1269–1284.

Woods, A. J., Antal, A., Bikson, M., Boggio, P. S., Brunoni, A. R., Celnik, P., Cohen, L. G., Fregni, F., Herrmann, C. S., Kappenman, E. S., et al. (2016). A technical guide to tdcs, and related non-invasive brain stimulation tools. *Clinical neurophysiology*, 127(2):1031–1048.

Wörsching, J., Padberg, F., Ertl-Wagner, B., Kumpf, U., Kirsch, B., and Keeser, D. (2016). Imaging transcranial direct current stimulation (tdcs) of the prefrontal cortex—correlation or causality in stimulation-mediated effects? *Neuroscience & Biobehavioral Reviews*, 69:333–356.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Chapter 2

Study 1: Distinct Trajectories of Response to Prefrontal tDCS in Major Depression: Results from a 3-arm randomized controlled trial

Goerigk, S. A., Padberg, F., Bühner, M., Sarubin, N., Kaster, T. S., Daskalakis, Z. J., Blumberger, D., Borrione, L., Razza, L., and Brunoni, A. R. (2021). Distinct trajectories of response to prefrontal tDCS in major depression: results from a 3-arm randomized controlled trial. Neuropsychopharmacology, 46(4), 774-782.

Contents of this study are published elsewhere and can be accessed here:
https://doi.org/10.1038/s41386-020-00935-x

Chapter 3

Study 2: Parsing the Antidepressant Effects of Non-Invasive Brain Stimulation and Pharmacotherapy: a Symptom Clustering Approach on ELECT-TDCS

Goerigk, S., Padberg, F., Chekroud, A., Kambeitz, J. Bühner, M., and Brunoni, AR., (2021). Parsing the Antidepressant Effects of Non-Invasive Brain Stimulation and Pharmacotherapy: a Symptom Clustering Approach on ELECT-TDCS. Brain Stimulation, 14(4), 906-912.

Contents of this study are published elsewhere and can be accessed here: https://doi.org/10.1016/j.brs.2021.05.008

Chapter 4

Study 3: Clinical patterns predict treatment response to transcranial direct current stimulation (tDCS) in patients with major depression: A machine learning analysis of the ELECT-TDCS study

Kambeitz, J., Goerigk, S., Gattaz W., Falkai P., Benseñor, I.M., Lotufo P.A., Bühner M., Koutsouleris N., Padberg, F., Brunoni, A.R. (2020), Clinical patterns differentially predict response to transcranial direct current stimulation (tDCS) and escitalopram in major depression: a machine learning analysis of the ELECT-TDCS study. Journal of affective disorders, 265, 460-467.

Contents of this study are published elsewhere and can be accessed here:
https://doi.org/10.1016/j.jad.2020.01.118

Chapter 5

General Discussion

The present dissertation introduced a methodological framework to reevaluate clinical trials in the context of precision psychiatry using the example tDCS. In-depth methods of analysis within this framework allow parsing of patient-individual components and response-dynamics, as well as the prediction of treatment response on the single-patient level. Based on newly derived and traditional units of analysis, statistical models of both the explanatory and algorithmic modeling traditions were applied over the course of this exploratory research, to gain a deeper understanding for the heterogeneity of response to tDCS on the patient- and subgroup-level. The framework was empirically taken to the test, using a comprehensive dataset from ELECT-TDCS, i.e. a landmark tDCS trial. To differentiate general antidepressant and tDCS-specific effects, all clinical patterns identified within the group of patients treated with tDCS were compared to another active pharmacological intervention, and to placebo.

Study 1 investigated whether group-level response to tDCS could be better understood as a combination of naturally occurring trajectories of symptom change over the course of treatment. By using latent growth mixture modelling, classes of rapidly, slowly, and no/minimally responding patients could be identified. In tDCS, these classes showed distinct profiles of associated characteristics, suggesting, amongst others, the relevance of higher depression severity, use of benzodiazepines, and age for poorer response to tDCS.

In **Study 2** it was explored whether efficacy of tDCS, as commonly evaluated on aggregate measures for the severity depressive symptoms, could be better understood in terms of its effects on distinct, naturally occurring clusters of symptoms. Using unsupervised ML methods, the HAM-D (17-items) scale was parsed into 4 distinct symptom clusters, effectively replicating results found within a large, representative sample for antidepressant treatment (STAR*D). Results of efficacy analyses suggested that antidepressant effects of tDCS and escitalopram were stronger in core depressive and guilt/anxiety symptoms than in clusters related to sleep and atypical symptoms. However, while both tDCS and escitalopram were superior to placebo in reducing core depressive symptoms, only tDCS was superior to placebo in improving sleep symptoms and only escitalopram was superior in improving guilt/anxiety symptoms.

In **Study 3** supervised ML algorithms trained on easily-obtainable clinical and de-

mographic baseline data were employed to predict end-of-treatment response to tDCS. Cross-validated out-of-sample accuracy was modest but above chance levels for algorithms trained on pre-labelled tDCS and escitalopram data, while placebo response could not effectively be predicted. Importantly, models were sensitive only to predict response to the intervention that they were trained for, but could not be successfully be generalized to the other treatment (i.e. tDCS models could not predict escitalopram response and vice versa), even though they were fit on identical sets of features. IML methods identified highly contributing features that were distinct for both active interventions. For tDCS, clinical features related to prefrontal cortex activation, such as negative affect, were among the most highly ranked variables.

A detailed discussion of the three empirical studies has already been given in their respective chapters. Thus, the general scientific contribution of the present dissertation and its implications for the field of tDCS research and for secondary analyses of clinical trials in general are the focal topic of the general discussion. It is discussed what the present work can and cannot contribute to the existing literature on prediction of successful tDCS treatment. Finally, future directions and challenges for the emerging field of precision tDCS therapies are summarized.

5.1 Overall Contribution of the Present Dissertation

One key in improving tDCS treatment in MDD is its personalization to the patient. While in precision psychiatry the question of who is going to benefit from what kind of treatment is more strongly coming into focus, methods applied within this paradigm (such as ML) often require large samples with high phenotypic detail (Bzdok and Meyer-Lindenberg, 2018). However, in tDCS research data is mostly limited to clinical trials, thus, to provide evidence-based strategies for personalization, the data that already exists must be explouted efficiently. Another issue that may make RCTs a key component of future precision approaches, is the uncertainty about specific, treatment-induced variability after NIBS treatment, which would be a prerequisite for personalisation of the intervention Homan et al. (2021). However, this open question cannot be pursued by analysis of NIBS data alone, but only in comparison with other active control groups or in comparison with placebo. Thus, given demand for the experimental attributes of the RCT design, the framework for in-depth RCT analysis presented within this presentation may contribute to the distinction between specific vs. non-specific (placebo) effects of treatment and thus to better personalisation options.

Importantly, the methods embedded within the presented methodological framework enable better exploitation of already existing clinical trial data. While efficacy analyses of RCTs are primarily focused on demonstrating group-based improvement of the treatment, ancillary analyses using new statistical models can provide more differentiated focus in explaining varying treatment effects and enable the translational step to other trials and settings.

This framework should be regarded as an intermediary step, linking data that was

gathered under systematically controlled conditions to new investigations within precision psychiatry (e.g. stratified treatment scenarios) as well as to more naturalistic big data contexts using data from routine care (see 5.3).

Using data from a high-quality sample this dissertation provides a blueprint for this type of systematic RCT analysis combining both sub-grouping approaches for more meaningful categories of response as well as prediction of treatment response within the single subject setting.

5.1.1 Deriving New Units of Analysis

One contribution of the present dissertation was to illustrate how statistical methods could be applied to parse traditional measures used in RCTs. Two different data-driven approaches were presented, each targeting a different aspect of the treatment: the symptoms and the patients. An increasing number of studies suggest that traditional concepts of clinical categorization are not ideal sets of reference (Gabrieli et al., 2015). This implies that neither a diagnosis by itself nor a global score on a diagnosis-specific measure seem to adequately portray, if a patient is an eligible recipient for tDCS and how the treatment is going to take effect. Thus, focusing on more differentiated units of analysis provides relevant information about the treatment process and the treatment selection that would otherwise have been missed:

One aim of the present dissertation was to gain a deeper understanding for the longitudinal dynamics of response within patients treated with tDCS. While previous research investigating response to tDCS has traditionally been focused on explaining change as presented by the entire group of patients on average, studies on other antidepressant interventions have identified distinct response trajectories to treatment (Uher et al., 2010; Kaster et al., 2019). For tDCS in particular, a differentiation with respect to response latency seems relevant, as "late effects" of the treatment have previously been suggested (Brunoni et al., 2017; Li et al., 2019). Consequently, determination of non-response may be premature for the slowly-improving patients in shorter treatment procedures.

Study 1 was the first to describe depression response trajectories over a 10 week treatment with tDCS. Importantly, assignment of each patient to a change-sensitive category was clinically relevant as the three identified latent trajectory classes could differentiate in terms of response and remission rates after only one week of treatment. Without a categorization procedure capable of taking individual trajectories of symptom change into account, associations with clinically relevant characteristics, such as the modifiable treatment factor concomitant benzodiazepine use that was related to poorer treatment response, may have been masked within a cross-sectional two-class solution. Associations between higher baseline severity as well as younger age with worse response suggest that these patients are less likely to respond instantaneously and may require a longer duration of tDCS treatment. While further work will be necessary to replicate these findings, the existence of slow and fast responding sub-populations suggests possibilities for development of individualized treatment protocols and exploration of the utility of prolonged tDCS treatment courses. Furthermore, these results raise the possibility that identified trajectory

classes represent separate neurophysiological phenotypes of MDD with preferential response to tDCS (Drysdale et al., 2017).

Another aim was the reevaluation of tDCS efficacy for distinct depressive symptom clusters. While most clinical trials have previously been evaluated on the change in global measures of depression severity (aggregate measures), studies on other antidepressant interventions (Chekroud et al., 2017; Bondar et al., 2020) have identified distinct treatment effects on different groups of symptoms. This goal was pursued in Study 2 and required derivation of another set of new units of analysis, this time in the outcome space (i.e. on the symptom-level). Evaluation of HAM-D cluster scores, rather than HAM-D sum scores provided insight in the effects tDCS had on each symptom cluster. Furthermore, it allowed more high-resolution comparisons between the treatment-profiles of the two active antidepressant interventions. These are the first results suggesting superiority of tDCS over placebo in a symptom cluster, where pharmacological treatment was not found to be superior. These findings have relevant implications for the personalization of treatment choices, however, the differences would have been missed using an aggregate measure of symptom severity.

While Study 2 used unsupervised ML to find natural sub-groups within the item space, another application related to the derivation of new units of analysis which could be used to expand the here-presented mythological framework, is to group patients (instead of symptoms) according to their constituting factors. For instance, the diagnostic categories in the Diagnostic and Statistical Manual of Mental Disorders (DSM) and International Classification of Diseases (ICD) are mostly based on symptom phenomenology and are thus often not congruent with patterns in brain and behavior. Thus, by assigning patients to a single diagnostic category, the contribution of different pathophysiological mechanisms on the clinical picture is ignored. ML methods provide automatic extraction of previously unknown patterns in single patients from different levels of biological and behaviour data, cutting across the diagnostic categories (manifolds). These data-driven sub-groups can be characterized by their interaction with disease and treatment processes (Bzdok and Meyer-Lindenberg, 2018; Insel and Cuthbert, 2015).

In an intermediary step, statistical reassessments that provide new units of analysis can be hypothesis-generating for future studies. In the context of experimental studies, a cluster of symptoms (e.g., insomnia cluster) could be the dependent variable in which the efficacy of tDCS (e.g. using different sets of parameters or in combination with other treatments) would be assessed. Furthermore, these units provide options for systematically testing the intervention in different sub-groups. For instance, studies relying on testing a specific predictor would involve stratifying the sample into patients who have and those who do not have the respective characteristic (alternative stratification could be into groups of patients who predominantly present symptoms of specific clusters), and then testing the efficacy of the intervention, ideally in comparison to a placebo group.

5.1.2 Integration of Explanatory and Predictive Modeling

The methods applied in the framework for reanalyzing clinical trials in the context of precision psychiatry, as presented in this dissertation, constitute a pluralistic composition of models operating both in the notion of significance as well as in the prediction of previously unseen instances. The framework should thus be discussed in terms of its contribution to the ongoing controversy about the explanatory versus the predictive approach to modeling mental health outcomes within psychiatric research and data science in general (Yarkoni and Westfall, 2017; Breiman et al., 2001).

While explanation and prediction are the two main goals of most research endeavors (Simon, 2001), they are qualitatively distinct in terms of their approach, their principled assessment for extrapolation of an effect beyond the available data, and mathematical models, hence, they answer distinct questions (Shmueli et al., 2010). In recent years, classical explanatory modeling has been increasingly scrutinized due to its lack of reproducibility and predictive power (Ioannidis, 2005; Abi-Dargham and Horga, 2016; Dwyer et al., 2018). The predictive modeling approach is better suited to create highly accurate models than previous explanatory work and can be readily applied within the single patient setting (Bzdok, 2017). However, often these highly accurate models can neither explain the underlying (causal) mechanisms nor the statistical mechanisms that cause their predictions. For this reason, ML models have often been criticized of being black-box algorithms (Castelvecchi, 2016).

Interestingly, while both approaches are not incompatible, they are rarely used together (Mahmoodi et al., 2017). In the endeavor of personalising treatments an integration of predictive and explanatory modeling may be helpful to optimize trade-offs in both approaches as has been previously suggested for social and behavioral research (Mahmoodi et al., 2017). On the one hand, even in studies that have primarily explanatory objectives, ML methods can still offer invaluable benefits when they are used instrumentally by reducing dimensionality of data, increasing research efficiency, and minimizing p-hacking (Yarkoni and Westfall, 2017). For example, in Studies 1 and 3 of the present dissertation, machine learning methods could provide a comprehensive ranking of baseline characteristics by their contribution to classify outcomes of tDCS treatment. However, these large numbers of predictors would have inevitably caused issues of multicollinearity and model identification in traditional statistical models (Zou and Hastie, 2005). Specifically, predictive modeling might advance the understanding of tDCS response heterogeneity by identifying relevant targets that could then be targeted in process-oriented clinical studies.

On the other hand, explanation can improve prediction of future instances by specifying data-informed models. These models would potentially be more sparse, more robust to changes in the data they are trained on, and less dependent on continuous re-calibration (Von Rueden et al., 2019). Mahmoodi et al. (2017) suggest establishing a productive cycle by creating novel, theoretical insights based on prediction efforts and including explanatory insights to create better informed prediction models.

5.1.3 Ingetration of Data-driven and Theory-driven Approaches

The dichotomy of data-driven versus hypothesis testing methods also touches upon deductive (knowledge-driven) approaches (Popper and Keuth, 2005) and induction (data-driven) approaches for the identification of clinical patterns within the patient data: In the context of precision psychiatry, the accumulation of large datasets has introduced a scientific practice that creates insights exclusively from data (e.g. data-derived subgroups of patients vs. DSM/ICD diagnoses)(Bzdok and Meyer-Lindenberg, 2018; Drysdale et al., 2017). While this approach can avoid confirmation bias within the research community (Nickerson, 1998), claims that big data approaches make the scientific method obsolete (Anderson, 2008) have elicited warnings from a "big data hybris" (Lazer et al., 2014).

Arguably, the epistemological strategy to better understand heterogeneity of treatment response should be to combine both, deductive and inductive practices. For example, in Study 1 both top-down (i.e. literature-driven) and bottom-up strategies were employed for the selection of variables to test as predictors of trajectory class assignment probabilities, leading both to the confirmation of candidate predictors as well as to the identification of novel associations to be tested in future explanatory trials. Thus, data-driven analysis combined with well-formalized theories about mechanisms of action leading to the investigation of increasingly refined research questions appears to be a realistic and promising hybrid approach to investigate response to tDCS (Platt, 1998).

5.1.4 Using Machine Learning for tDCS Research

Finally, this dissertation illustrates two distinct scenarios for employment of ML in clinical data. The main promise of using ML methods in psychiatric intervention research is to harvest their potential to detect and formalize complex patterns in clinical information, behavior, brain, and genes, both to advance a more biologically grounded definition of conventional psychiatric disorders (i.e. using objectively measurable endophenotypes) as well as to improve predictability of treatment outcomes. The studies included in this dissertation are the first ones in the field to use supervised and unsupervised ML approaches for the assessment of response to tDCS.

Study 2 employed unsupervised ML methods to uncover natural depressive symptom clusters within aggregate measures of symptom severity. While person-centered unsupervised ML approaches can potentially parse the population of patients treated with brain stimulation into data-driven endophenotypes (Drysdale et al., 2017), this symptom-centered approach is an example for transposing commonly used model inputs into the outcome space (i.e. clustering symptoms instead of patients).

In contrast, supervised ML algorithms are well suited for accurate outcome prediction and have the potential to give a pre-informed prognosis about treatment success as well as predicting valid and immediately usable clinical objects, such as stimulation dosage and location. Following methodological procedures of previous work from the rTMS field (Koutsouleris et al., 2018b), Study 3 applied supervised ML to predict response to tDCS after a 10-week treatment period using only a set of easily collectable clinical and

demographic information at baseline. While the developed supervised ML models should still be considered on the proof-of-concept level, they have been provided to the field and can, once data availability increases, be expanded (e.g. by adding new features) and, with improving accuracy, be clinically evaluated in terms of generalization (independent or third party test sets), model scope (locally representative samples and samples of the target population), and incremental utility (real-life workflow, complementing state-of-the art).

Thus, both studies show that reevaluation of clinical trial data in combination with supervised and unsupervised ML approaches, provides a wide range of opportunities for optimizing tDCS treatment and treatment selection.

5.2 Limitations of the Present Dissertation and Implications for tDCS Research

The present dissertation has some limitations, out of which several epitomize the fact that personalisation of tDCS treatment is still in its beginnings. The primary limitation in the deployment of state-of-the-art statistical models to personalize tDCS is likely the size of available datasets, both in terms of the number of included subjects as well as the insufficient granularity of information (e.g. biomarkers, medical history, number of measurements per week). As clinical efficacy of tDCS remains unproven, further phase-3 controlled studies are still necessary. Consequently, systematic collection of vast amounts of data, for example from routine clinical care, has not yet been feasible. Further limitations within the present dissertation are summarised below:

5.2.1 Outcome Measures and Unclear Prognostic Labels

One of the objectives of the present dissertation was to derive more meaningful units of analysis, as compared to those measures primarily used in the evaluation of clinical trials. However, while the resulting units provide an increased detail in capturing variability of treatment response, they are still fundamentally based on a clinician-rated scale (HAM-D, 17 items). Thus, they remain susceptible to the caveats associated with this type of symptomatic assessment. These caveats include rater effects (Bühner, 2011) and situational measurement bias.

Study 3 used a traditional definition of response (>50% improvement) as the outcome to be predicted by the ML models to allow comparisons with previously developed algorithms in the field of NIBS (Koutsouleris et al., 2018b). The pitfalls of introducing this type of arbitrary dichotomies have been summarized above (see 1.2.2). Thus, it would be a logical next research step to combine the building blocks that were separately illustrated within the empirical studies of this dissertation, and predict newly derived, data-driven units in the outcome space (i.e. assignment to trajectory classes from Study 1 or symptom-cluster scores from Study 2) within the single-subject prediction paradigm.

Furthermore, the self- or clinican-rated symptom assessment may be complimented using objective behavioral correlates of MDD. A promising approach for the acquisition

of objective behavioral data may be offered by technical devices carried by patients. For instance, smartphone data could enable early detection of clinical events, such as rumination, unrest, or suicide attempts (Farhan et al., 2016). Ambulatory assessments using smartphones have previously been classified into passive sensing (recording from sensors and phone usage in the background)(Stachl et al., 2017) and active logging (in-situ reported experiences using diaries or so-called experience sampling) (Harari et al., 2017, 2020; Schödel, 2020). Furthermore, recorders for biosignals such as heart rate monitors (Wilhelm and Grossman, 2010) and actigraphs (Van De Water et al., 2011), have been used in the past, however, these are exceedingly getting integrated into smart devices (Miller, 2012). Moreover, digital sensors able to monitor movement, communication habits, and diverse behaviors are entering everyday life (Internet of things, (Ashton et al., 2009)).

Within the present dissertation, effects of tDCS were only investigated in terms of better response to treatment but not in terms of adverse effects. Since safety of treatment represents a second relevant outcome in brain stimulation interventions, adverse events (AE) should be considered as an integral outcome in future studies on personalizing tDCS. For instance, the loss-functions of certain ML methods such as multi-objective reinforcement learning models (MORL) can be defined to simultaneously optimize conflicting alternatives (e.g. dose-response vs. dose-safety trade-off) (Tzeng and Huang, 2011).

Furthermore, it might be relevant to probe both, depressive symptoms and time-varying clinical predictors over longer measurement periods, instead of restricting inputs for model training only to baseline data. This is relevant, because depression-related constructs and self-concepts (e.g. mood) are highly volatile and situation dependent. Thus, they can vary over the span of a day (Peeters et al., 2003). While it would be advantageous to give a patient a precise as possible prognosis before the onset of the treatment (baseline-informed model), even estimations of treatment success after a number of weeks (early response) would considerably shorten treatment times and spare the patients a lengthy process of trial-and-error. In this context, Senn (2018) recommends testing treatment response at least twice in the same individual, as only treatments with some consistency of response can efficiently be tailored to the patient, while inconsistent treatment outcomes are too futile to identity subsets of responders. One promising approach in gaining some insight into these (situation- and patient-dependent) inconsistencies of response are N-of-1 studies that monitor frequent treatment application over a longer period (Araujo et al., 2016).

5.2.2 Predictors

Another limitation besides the deficiency of objectivity in the outcome space is the lack of detailed information in the feature space used to predict response to tDCS. While the clinical, demographic, and neurocognitive baseline variables in Studies 1 and 3 have the advantage of being obtainable without larger efforts or costs, no neuroanatomical, functional, or genetic markers were used in the training of the statistical models. However, these data might hold relevant detail for the classification of responders vs. non-responders. In fact, using a subset of patients from the same clinical trial (ELECT-TDCS), two studies could identify associations between clinical outcomes and grey matter volumes of PFC subregions

as well as with individual efield strength in the left ACC (Bulubas et al., 2019; Suen et al., 2021). While these results suggest that therapeutic outcomes may be related to individual cortex morphology, above-mentioned factors have neither been tested for their predictability in a cross-validated framework by themselves, nor in combination with the variables used in the present dissertation yet.

Interestingly, Koutsouleris et al. (2018b) could predict response to 21 days of rTMS treatment with higher cross-validated accuracy (85%) using pre-treatment structural Magnetic Resonance Images (sMRI) of patients with schizophrenia, while comparable studies predicting remission from pharmacological treatment across large datasets (STAR*D) using only clinical data reached similar accuracy as in this dissertation (64.6%)(Chekroud et al., 2016). This raises the possibility of a multi-level classification approach for the trade-off in cost-efficiency and accuracy by applying a model trained on clinical characteristics in a first step and then applying high-effort and high-cost structural MRI models in patients with previously low classification thresholds (Koutsouleris et al., 2018a).

5.2.3 Validity and Generalizability

It is important to note that the results from the present dissertation are specific to the interventions conducted in the context of the ELECT-TDCS trial (Brunoni et al., 2015, 2017). Besides potential effects due to the double-controlled trial design (i.e. tDCS patients received an additional placebo pill rather than pure stimulation), study-specific factors include the setup of the tDCS device and the parameters employed. While effects may vary depending on the specific treatment administered (Brunoni et al., 2012), no "optimal" or "standard" tDCS protocols for MDD are currently available. Furthermore, ELECT-TDCS included only antidepressant-free, unipolar, non-psychotic patients without substance abuse or dependence, and without personality disorders.

This raises a general issue regarding generalizibility of models fit on data from RCTs: While strict criteria on the inclusion and exclusion of eligible patients are understood as a marker of quality within the experimental character of randomized controlled trials (e.g. multiple comorbidities, concomitant therapeutic regimens), this can impede the translation of findings and algorithmic tools to other settings, such as "real-world" clinical practice. Thus, external validation is warranted to generalize study results to other contexts and avoid translational issues related to idiosyncrasies of patients (e.g., cultural homogeneity), materials and staff (e.g. study personnel, stimulation devices) and methods (e.g., measurement procedure) (see Figure 1.3).(Dwyer et al., 2018)

The cross-validation frameworks applied for model evaluation in this dissertation, tested predictive performance on previously unseen data. However, no validation on treatment data from completely external contexts could be conducted, as for now, there are no other tDCS trials that use comparable feature sets while having similar treatment duration. Thus, data availability, the harmonization of existing datasets, as well as the standardization of gathered information in future trials remain a challenge (see 5.3).

To avoid false-positive findings (Type-I errors) due to multiple hypothesis-testing and pairwise comparisons between treatment groups, statistical confidence measures were

consistently adjusted for the false discovery rate (FDR) (Benjamini and Hochberg, 1995). In completely data-driven procedures, even more conservative corrections were applied (99.9% CIs). However, the findings presented in this dissertation are the result of secondary analyses the ELECT-TDCS trial was neither explicitly planned nor powered for. Thus, they should be understood as exploratory and they mandate further investigation in future studies.

5.2.4 Long-term Effects of Treatment

Within the ELECT-TDCS trial, effects of tDCS were investigated over a 10-week period. Thus, the time frame that this dissertation can provide insides for is limited to this acute treatment phase. While only few studies with relatively small sample sizes report follow-up data (e.g. over a 6-month course) relapse rates vary between 27-50% (Aparicio et al., 2019; Martin et al., 2013; Valiengo et al., 2013). However, for the personalization of tDCS treatment the durability of treatment effects would be another relevant aspect. For example, one study suggested differences in relapse rates between nontreatment- versus antidepressant treatment-resistant patients (Aparicio et al., 2019). While high rates of relapse are also problematic after treatments with electroconvulsive therapy (ECT) (Sackeim et al., 2001) and rTMS (Kedzior et al., 2015) in the absence of maintenance treatment, tDCS might provide the advantage of being applicable at home due to its portability. In this context, self-administered application of tDCS can be distinguished between remotely supervised tDCS and home-use (domiciliary) tDCS (Palm et al., 2018). Combining predictive tools for relapse prediction with home-based maintenance application may provide a promising approach for more stable and long-lasting treatment effects in the single-patient setting.

5.3 Challenges and Future Directions in Personalized Application of tDCS

There are many reasons to extend the framework of precision psychiatry to tDCS research. Foremost, this goal should be pursued with a view to an improved clinical applicability. However, some challenges have to be overcome before tDCS can be established as a personalized treatment option for MDD. While there are hurdles in each of the ingredient disciplines that act jointly in the effort of personalizing tDCS, the following section will focus on more general obstacles related to the results of this dissertation. The interested reader may be referred to specific summaries of pitfalls of tDCS (Brunoni et al., 2012), computational neuroscience (Huys et al., 2016), and ML (James et al., 2015; Domingos, 2012).

5.3.1 Routine Data Collection and Data Sharing

As previously stated, availability of large, multi-modal datasets is one of the primary limitations, when it comes to predictive analytics in mental health. Mega-cohorts such as

the UK Biobank (Sudlow et al., 2015) and international consortia like ENIGMA (Thompson et al., 2014) provide multi-modal acquisition of clinical and biological data at scale. While analyses of complex multidimensional patterns on different biological levels may reveal new pathways for conducting research on individual factors (Bzdok and Meyer-Lindenberg, 2018; Padberg et al., 2021), the collection of larger quantities of treatment data remains a challenge in brain stimulation research.

In tDCS, further phase-3 controlled clinical trials will first have to be conducted to prove its efficacy. However, to successfully exploit the methods from the precision psychiatry framework, patient documentation should be acquired as meticulously as possible in such trials. Whenever possible, investigators should collect comprehensive sets of biomarkers to be analyzed in association with the primary hypothesis of the study (Borrione et al., 2020). Once efficacy of tDCS can be proven, further sham controlled trials will no longer be ethically feasible, as equipoise to placebo could no longer be assumed (Miller and Brody, 2002). At this point clinical and academic centers performing tDCS should routinely incorporate collection of patient characteristics (clinical and demographic information, rating-scales, questionnaires, and tests on cognitive changes) as well as biomarkers (brain scans and molecular data). A comparable approach has already been implemented for treatment with ECT (Global ECT-MRI Research Collaboration) which could serve as a prototypical model for tDCS (Oltedal et al., 2017).

To reach adequate sample sizes and to appropriately consider generalizability when designing or reporting studies with translational aims, consolidation of models and data across different study centers or even continents is inevitable. This requires global efforts in research collaboration and data sharing. To this end, support by open-access and (properly anonymized) data sharing initiatives (Ayris et al., 2016; Soderberg, 2018), will be essential for overcoming difficulties in assessing literature and primary data. Thus, regulatory agencies should further endorse policies that ensure clinical trial data can be shared under request (Borrione et al., 2020).

5.3.2 Basic Research Outside of Clinical Application

Basic research outside of the clinical application of tDCS in MDD could provide further advances in its clinical efficacy. A better understanding of brain networks involved in pathophysiological processes, that do not only represent topographically linked nodes, but dynamic networks coupled by oscillatory brain activities could be driven by advances in disciplines such as genetics or multi-modal brain imaging and may help directing the stimulation to pathophysiologically relevant targets (Padberg et al., 2021). Furthermore, prospects for increasing tDCS response in MDD improve with a better understanding of the neurological effects of tDCS. A common feature in tDCS trials is its capability to elicit lasting changes in regional cortical excitability. Investigations for a deepened understanding of the extent to which synaptic plasticity can be induced by the stimulation (e.g. long-term potentiation (LTP)-type or long-term depression (LTD)-type changes) may be beneficial for potentiation of its antidepressant effects. Basic research outside of clinical populations may help to improve the understanding of the technique's effects. For instance,

pre-clinical studies (e.g. translational reseach in animals) (Brunoni et al., 2011) and computer simulation (Thielscher et al., 2011) may give insights about mechanisms of action that are still undetermined. Finally, studies in healthy populations could optimize selection within the stimulation parameter space and aid development of closed-loop approaches for efficient online control of the intervention (Lorenz et al., 2019).

5.3.3 Development of Predictive Tools

Automatized Model Development and Optimization

With higher complexity of data and research objectives, the process of generating a successful predictive model (e.g. supervised ML) relies increasingly on appropriate steps taken by the programmer. These steps include adequate preprocessing, feature selection or engineering, model selection and optimization, as well as valid assessments of model performance and generalization (as outlined in 4.2). Informed decisions within this algorithmic pipeline require both experience as well as trial-and-error processes, which can be a challenge for the developer. For this reason, many studies skip the optimization process of the model pipeline entirely. This has been epitomized in a survey of neuroimaging studies, which showed that 73% of the studies used the same ML model and did not perform any hyperparameter tuning (Arbabshirani et al., 2017). However, without knowledge of the underlying processes that generate the data, on average, most ML algorithms produce the same, usually not optimal, results ("No Free Lunch Theorem") (Wolpert et al., 1995). In most health disciplines, clinicians and researchers in the field of brain stimulation are not specifically trained in programming ML algorithms. However, the discipline of so-called automated machine learning (AutoML) dramatically simplifies the above-mentioned steps for non-experts (Hutter et al., 2019). AutoML covers the entire pipeline (end-to-end) from the raw dataset, over the selection of the model, to the deployable algorithm (e.g. Auto Keras (Jin et al., 2018) and auto-sklearn (Feurer et al., 2019)).While this trend can be expected to accelerate, automation will increase requirements for proper model evaluation.

Creating Complex Models Using Small Samples

Data availability can be assumed to increase quickly in the future. However, the sample sizes required for successful model training in ML can quickly amount to large numbers with increasing model complexity. In deep-learning algorithms, a particularly complex set of methods, the number of required free parameters may easily add up to millions (Goodfellow et al., 2016). Training these models with data from only hundreds of patients may induce overfitting within the training sets (i.e. the model perfectly learns patterns in the training data) and thus limit prediction performance in real-world data (Friedman et al., 2001). This problem is aggravated in high dimensional data sources (Arbabshirani et al., 2017), such as genetic and neuroimaging data (so-called large P small N problems) (Cearns et al., 2019). Acquiring datasets of this size, does not seem feasible in brain stimulation and psychiatric research in general.

However, new approaches from the ML domain specifically address the problem of training predictive models on small datasets (Cearns et al., 2019). For instance, random augmentation of existing data (data perturbing) is a mathematical means to artificially enlarge training sets. This approach has been successfully applied in image processing (Krizhevsky et al., 2017). The optimal degree of randomness to add can be introduced as a tunable hyperparameter in the model training process. In a second approach, transfer learning can be used for the transformation of variables into lower dimensional representations based on previous training on other datasets (Esteva et al., 2017). Similarly, generative adversarial neural networks (GAN) can be employed for intra-domain transfer learning, by combining both generative and classification processes to create arbitrary numbers of stimuli (e.g. brain scans) required for model training (Radford et al., 2015).

From Research Algorithms to Decision Support Frameworks

At current, most research on the development of predictive tools in psychiatry is still within the notion of proof-of-concept studies. However, real-world estimates of performance and prognostic stability will have to be demonstrated before algorithms for prediction of tDCS response can be clinically deployed (Cearns et al., 2019). In other medical disciplines, algorithmic tools have been lined up to compete against trained clinicians in terms of their ability to correctly diagnose diseases and predict treatment outcomes (Ardila et al., 2019). However, it seems unlikely that algorithms by themselves will run the clinical decision-making process in psychiatric treatment due to the interpersonal nature of mental health care, even if algorithmic tools were to outperform clinician's prognoses. Instead, decision support systems alongside human therapists could prove to be a promising and more socially accepted application scenario in the clinical use of predictive models Cearns et al. (2019). In tDCS, prospective RCTs could test if a collaboration of a clinician and an algorithmic tool could reach significantly higher accuracy in predictions than the clinician prognostication alone (i.e. outperforming the state of the art) (Passos and Mwangi, 2018). In this scenario, the human-AI synergy would be of particular importance, combining clinical expertise with carefully calibrated (Niculescu-Mizil and Caruana, 2005) probabilistic estimates as model outcomes (Hahn et al., 2013). For further recommendations on the implementation of clinical AI monitoring systems in psychiatry, including ongoing post deployment evaluation, security assessment, and algorithmic bias strategies see the review by Cearns and colleagues (Cearns et al., 2019).

Another aspect bringing the precision-oriented approach for tDCS into clinical practice is its economical feasibility. Cost-effectiveness analyses will have to be performed to examine if the personalization approach is advantageous. For example, while tDCS devices have the advantage of being relatively cheap, using bio-markers that have to be measured using expensive neuroimaging procedures may relativize the savings on the intervention side.

Understanding the Predictions

As previously outlined, (most) algorithmic tools are by no means black-boxes but apply deterministic, albeit rather complex rules to make predictions. IML methods (Molnar et al., 2020), such as the ones applied in Study 3, can help to understand these rules (e.g. to identify the most relevant variables). In a straightforward manner, so-called model-agnostic techniques systematically obscure certain sets of features and measure the resulting decrease in prediction performance. This mechanism works independently of the applied model type and can thus be applied in a variety of statistical scenarios. Another model-specific approach evaluates the fitted model itself. Examples include interpretation of weight-maps in support vector machines (Cortes and Vapnik, 1995), the gini impurity measure in tree-based classifiers (Breiman, 2001) or the visualization of layer-wise data transformation in neural networks (Bach et al., 2015). However, besides understanding the interplay of model inputs, it is also relevant to register steps taken in the algorithmic pipeline such as preprocessing, feature engineering and performance evaluation, to understand model behaviour and ensure replicability of results. Within this notion, the recently presented conceptual framework of AI Transparency (Hahn et al., 2018) provides a "checklist" to quantify the maturity of a project and to monitor steps of model development.

5.3.4 Interdisciplinarity and Collaboration

The development of an interdisciplinary infrastructure represents another challenge in establishing personalized tDCS research within the data-driven framework of precision psychiatry. When faced with a multi-omics approach (so-called panomics) for revelation of the underlying biological pathways involved in psychiatric disorders, interdisciplinary collaboration of researchers from many disciplines will be required (e.g. clinicians, psychologists, computer scientists, mathematicians). The technical infrastructure required for handling big data (e.g. multimodal neuroimaging) poses another challenge, as most clinicians and psychologists mandated to acquire clinical data have primary education in data analysis but not in management and organization of large databases and utilization of high-performance computing (e.g. cluster systems, cloud computing). Training programs may have to adapt to novel requirements, including cross-disciplinary communication of methods and clinical concepts to improve the collection, organization, and integration of mental health data (Yarkoni, 2012). Furthermore, synergy between healthcare delivery systems and academic institutions provides opportunities, as healthcare providers can extract structured data to be modelled and shared as a collaborative research product (Williams et al., 2015).

5.3.5 Ethical Considerations

Alongside the opportunities that personalization of tDCS provides for the improvement of patient care, ethical questions need to be considered as well. While organisation of large multi-modal data repositories seems essential for modeling complex patterns of brain and behaviour, patient records have to be handled responsibly in terms of data privacy and

data security (Annas et al., 2003). On an organizational level, data privacy requires making data processing steps transparent to patients. While recursive anonymization of data is usually a requirement in modern research and data sharing initiatives, vast amounts of personal information (e.g. brain scans) (Valizadeh et al., 2018) may render re-identification of patients possible. It is thus crucial that patients are given as much control over their data as possible, that they are properly informed about study purposes and types of collected data, and that no data is saved without previous informed consent (Appelbaum et al., 1987).

In terms of security, technical measures to handle patient data, for instance secure transfer and storage as well as the timely deletion after the purpose of their collection is fulfilled have to be provided. Thus, researchers should carefully rehearse data acquisition and handling procedures with ethical committees and data protection officers at their institutions to comply with national and international regulation (e.g. Europe: General Data Protection Regulation, GDPR, USA: Health Insurance Portability and Accountability Act, HIPAA).

Clinical application of algorithms for classification of treatment success will have to be approved by regulations from central agencies (e.g. USA: Food and Drug Administration, Europe: European Medicines Agency) before its widespread use. This particularly involves the implementation of guidelines for biomarker development (Amur et al., 2015), that should be informed by a strong evidence base provided by academia and industry, as well as governmental agencies Dwyer et al. (2018).

Finally, while prognostic tools can play a role in making better informed treatment choices, the stakes are high when asking whether to give or withhold a treatment, therefore, this kind of decision should not exclusively be made on probabilistic grounds, but also involve humanistic consideration.

5.4 Conclusion

Traditionally, inter-individual variability of treatment response has been regarded as a limitation for developing standardized tDCS treatment procedures for MDD. However, taking the considerable phenotypic heterogeneity of depressive patients into account, the one-size-fits-all approach is increasingly questioned throughout psychiatric treatment. Instead, precision-oriented approaches aim to improve the treatment by tailoring it to the patient.

The present dissertation discussed the role of clinical trials in PM. It suggested that clinical trials may play an important role both (1) as a source of data to use in model development for a better understanding of the variability of treatment effects and (2) as an experimental framework that allows divergent validation of these models within active and placebo control conditions. In this context, caveats associated with analyses of traditional RCT outcomes that limit detailed examination of response heterogeneity, have been described. Earlier research predominantly used the treatment group as the primary working unit and ignored relevant clinical patterns within the patients, within the individual improvement trajectories, and within the presented symptoms.

The dissertation provided a methodological paradigm for renalysis of clinical trial data by example of an analysis of ELECT-TDCS. Within this paradigm in-depth statistical modeling approaches were presented in direct reference to the above-mentioned problems. The implementation of these approaches was demonstrated in three blueprint empirical studies investigating individual trajectories of improvement over the course of treatment, tDCS efficacy in distinct depressive symptom clusters, and the predictability of response in the single-patient setting.

The findings of the empirical studies point to opportunities of personalizing tDCS treatment of MDD: Patients who show distinct trajectories of change may profit from individualized treatment protocols. Concomitant benzodiazepine use is a modifiable factor related to poorer response and should be carefully investigated within upcoming trials, while younger patients and those with higher depression severity who are more likely to respond slowly may require longer duration of tDCS treatment. TDCS and escitalopram were both found to be superior to placebo in treating core depressive symptoms, yet only tDCS was superior in treating sleep related symptoms, while only escitalopram was superior in treating guilt/anxiety symptoms. Single-subject prediction using supervised ML algorithms provides above-chance classification of treatment response, using an easily obtainable set of baseline data. While predictive modeling has high potential for guiding clinical decisions in the future, the algorithms must first reach higher accuracy and their generalizability and incremental utility must be demonstrated.

Besides the derivation of more differentiated units of analysis and prediction on the single-subject level, analyses within the presented methodological framework demonstrated that exploratory research could fruitfully integrate approaches from the explanatory modeling and the predictive modeling cultures. More generally, the application of ML techniques was introduced to precision-oriented tDCS research.

Through the empirical studies, some of the current limitations and challenges that the personalization of tDCS for treating MDD is faced with were demonstrated in the present dissertation, including limited availability of large datasets, imprecise measurement of depressive symptoms and and associated characteristics in absence of objective biomarkers and behavioural measures, and lack of external validation possibilities. Further challenges not exclusive to tDCS research but referring to precision psychiatry in general are the establishment of interdisciplinary infrastructure, responsible handling of personal data, and initiatives in open science.

In summary, the personalization of mental health care is still in its beginnings, as has been demonstrated using the example of tDCS. However, the presented framework for in-depth analysis of clinical trials can be generalized to other interventions and contexts where variability of exerted effects suggests the individualization of treatment. By providing blueprint empirical studies for the implementation of new statistical methods within the framework of precision psychiatry this dissertation can specifically contribute to a deeper understanding of the heterogeneity of tDCS effects to improve this low-cost, tolerable, and safe intervention.

While it is unclear, if tDCS can be transformed into a fully individualized tool, the methods used in precision psychiatry offer a promising future direction in psychiatry research

and clinical care. As with any emerging field of research, optimistic biases may have to be overcome with judicious caution to serve the best interests of the patients.

Bibliography

Abi-Dargham, A. and Horga, G. (2016). The search for imaging biomarkers in psychiatric disorders. *Nature medicine*, 22(11):1248.

Amur, S., LaVange, L., Zineh, I., Buckman-Garner, S., and Woodcock, J. (2015). Biomarker qualification: toward a multiple stakeholder framework for biomarker development, regulatory acceptance, and utilization. *Clinical Pharmacology & Therapeutics*, 98(1):34–46.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7):16–07.

Annas, G. J. et al. (2003). Hipaa regulations-a new era of medical-record privacy? *New England Journal of Medicine*, 348(15):1486–1490.

Aparicio, L. V., Rosa, V., Razza, L. M., Sampaio-Junior, B., Borrione, L., Valiengo, L., Lotufo, P. A., Benseñor, I. M., Fraguas, R., Moffa, A. H., et al. (2019). Transcranial direct current stimulation (tdcs) for preventing major depressive disorder relapse: Results of a 6-month follow-up. *Depression and anxiety*, 36(3):262–268.

Appelbaum, P. S., Lidz, C. W., and Meisel, A. (1987). Informed consent: Legal theory and clinical practice.

Araujo, A., Julious, S., and Senn, S. (2016). Understanding variation in sets of n-of-1 trials. *PloS one*, 11(12):e0167167.

Arbabshirani, M. R., Plis, S., Sui, J., and Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145:137–165.

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6):954–961.

Ashton, K. et al. (2009). That 'internet of things' thing. *RFID journal*, 22(7):97–114.

Ayris, P., Berthou, J.-Y., Bruce, R., Lindstaedt, S., Monreale, A., Mons, B., Murayama, Y., Södergård, C., Tochtermann, K., and Wilkinson, R. (2016). Realising the european open science cloud.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., and Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, 10(7):e0130140.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bondar, J., Caye, A., Chekroud, A. M., and Kieling, C. (2020). Symptom clusters in adolescent depression and differential response to treatment: a secondary analysis of the treatment for adolescents with depression study randomised trial. *The Lancet Psychiatry*, 7(4):337–343.

Borrione, L., Bellini, H., Razza, L. B., Avila, A. G., Baeken, C., Brem, A.-K., Busatto, G., Carvalho, A. F., Chekroud, A., Daskalakis, Z. J., et al. (2020). Precision non-implantable neuromodulation therapies: a perspective for the depressed brain. *Brazilian Journal of Psychiatry*, (AHEAD).

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L. et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.

Brunoni, A. R., Fregni, F., and Pagano, R. L. (2011). Translational research in transcranial direct current stimulation (tdcs): a systematic review of studies in animals. *Rev Neurosci*, 22(4):471–81.

Brunoni, A. R., Moffa, A. H., Sampaio-Junior, B., Borrione, L., Moreno, M. L., Fernandes, R. A., Veronezi, B. P., Nogueira, B. S., Aparicio, L. V., Razza, L. B., et al. (2017). Trial of electrical direct-current therapy versus escitalopram for depression. *New England Journal of Medicine*, 376(26):2523–2533.

Brunoni, A. R., Nitsche, M. A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., Edwards, D. J., Valero-Cabre, A., Rotenberg, A., Pascual-Leone, A., et al. (2012). Clinical research with transcranial direct current stimulation (tdcs): challenges and future directions. *Brain stimulation*, 5(3):175–195.

Brunoni, A. R., Sampaio-Junior, B., Moffa, A. H., Borrione, L., Nogueira, B. S., Aparício, L. V. M., Veronezi, B., Moreno, M., Fernandes, R. A., Tavares, D., et al. (2015). The escitalopram versus electric current therapy for treating depression clinical study (elect-tdcs): rationale and study design of a non-inferiority, triple-arm, placebo-controlled clinical trial. *São Paulo Medical Journal*, 133(3):252–263.

Bühner, M. (2011). *Einführung in die Test-und Fragebogenkonstruktion*. Pearson Deutschland GmbH.

Bulubas, L., Padberg, F., Bueno, P. V., Duran, F., Busatto, G., Amaro Jr, E., Benseñor, I. M., Lotufo, P. A., Goerigk, S., Gattaz, W., et al. (2019). Antidepressant effects of tdcs are associated with prefrontal gray matter volumes at baseline: Evidence from the elect-tdcs trial. *Brain stimulation*, 12(5):1197–1204.

Bzdok, D. (2017). Classical statistics and statistical learning in imaging neuroscience. *Frontiers in neuroscience*, 11:543.

Bzdok, D. and Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3):223–230.

Castelvecchi, D. (2016). Can we open the black box of ai? *Nature News*, 538(7623):20.

Cearns, M., Hahn, T., and Baune, B. T. (2019). Recommendations and future directions for supervised machine learning in psychiatry. *Translational psychiatry*, 9(1):1–12.

Chekroud, A. M., Gueorguieva, R., Krumholz, H. M., Trivedi, M. H., Krystal, J. H., and McCarthy, G. (2017). Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA psychiatry*, 74(4):370–378.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H., and Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry*, 3(3):243–250.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87.

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R. N., Zebley, B., Oathes, D. J., Etkin, A., et al. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine*, 23(1):28–38.

Dwyer, D. B., Falkai, P., and Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual review of clinical psychology*, 14:91–118.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118.

Farhan, A. A., Yue, C., Morillo, R., Ware, S., Lu, J., Bi, J., Kamath, J., Russell, A., Bamis, A., and Wang, B. (2016). Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. In *2016 IEEE Wireless Health (WH)*, pages 1–8. IEEE.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J. T., Blum, M., and Hutter, F. (2019). Auto-sklearn: efficient and robust automated machine learning. In *Automated Machine Learning*, pages 113–134. Springer, Cham.

Friedman, J., Hastie, T., Tibshirani, R., et al. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Gabrieli, J. D., Ghosh, S. S., and Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, 85(1):11–26.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning cambridge.

Hahn, T., Ebner-Priemer, U., and Meyer-Lindenberg, A. (2018). Transparent artificial intelligence–a conceptual framework for evaluating ai-based clinical decision support systems. *Available at SSRN 3303123*.

Hahn, T., Marquand, A. F., Plichta, M. M., Ehlis, A.-C., Schecklmann, M. W., Dresler, T., Jarczok, T. A., Eirich, E., Leonhard, C., Reif, A., et al. (2013). A novel approach to probabilistic biomarker-based classification using functional near-infrared spectroscopy. *Human brain mapping*, 34(5):1102–1114.

Harari, G. M., Müller, S. R., Aung, M. S., and Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current opinion in behavioral sciences*, 18:83–90.

Harari, G. M., Vaid, S. S., Müller, S. R., Stachl, C., Marrero, Z., Schoedel, R., Bühner, M., and Gosling, S. D. (2020). Personality sensing for theory development and assessment in the digital age. *European Journal of Personality*, 34(5):649–669.

Homan, S., Muscat, W., Joanlanne, A., Marousis, N., Cecere, G., Hofmann, L., Ji, E., Neumeier, M., Vetter, S., Seifritz, E., et al. (2021). Treatment effect variability in brain stimulation across psychiatric disorders: A meta-analysis of variance. *Neuroscience & Biobehavioral Reviews*.

Hutter, F., Kotthoff, L., and Vanschoren, J. (2019). *Automated machine learning: methods, systems, challenges*. Springer Nature.

Huys, Q. J., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, 19(3):404.

Insel, T. R. and Cuthbert, B. N. (2015). Brain disorders? precisely. *Science*, 348(6234):499–500.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8):e124.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2015). An introduction to statistical learning with applications in r (springer-verlag new york).

Jin, H., Song, Q., and Hu, X. (2018). Auto-keras: Efficient neural architecture search with network morphism. *arXiv preprint arXiv:1806.10282*, 5.

Kaster, T. S., Downar, J., Vila-Rodriguez, F., Thorpe, K. E., Feffer, K., Noda, Y., Giacobbe, P., Knyahnytska, Y., Kennedy, S. H., Lam, R. W., et al. (2019). Trajectories of response to dorsolateral prefrontal rtms in major depression: a three-d study. *American Journal of Psychiatry*, 176(5):367–375.

Kedzior, K. K., Reitz, S. K., Azorina, V., and Loo, C. (2015). Durability of the antidepressant effect of the high-frequency repetitive transcranial magnetic stimulation (rtms) in the absence of maintenance treatment in major depression: A systematic review and meta-analysis of 16 double-blind, randomized, sham-controlled trials. *Depression and anxiety*, 32(3):193–203.

Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., et al. (2018a). Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA psychiatry*, 75(11):1156–1172.

Koutsouleris, N., Wobrock, T., Guse, B., Langguth, B., Landgrebe, M., Eichhammer, P., Frank, E., Cordes, J., Wölwer, W., Musso, F., et al. (2018b). Predicting response to repetitive transcranial magnetic stimulation in patients with schizophrenia using structural magnetic resonance imaging: a multisite machine learning analysis. *Schizophrenia bulletin*, 44(5):1021–1034.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: traps in big data analysis. *Science*, 343(6176):1203–1205.

Li, M.-S., Du, X.-D., Chu, H.-C., Liao, Y.-Y., Pan, W., Li, Z., and Hung, G. C.-L. (2019). Delayed effect of bifrontal transcranial direct current stimulation in patients with treatment-resistant depression: a pilot study. *BMC psychiatry*, 19(1):1–9.

Lorenz, R., Simmons, L. E., Monti, R. P., Arthur, J. L., Limal, S., Laakso, I., Leech, R., and Violante, I. R. (2019). Efficiently searching through large tacs parameter spaces using closed-loop bayesian optimization. *Brain stimulation*, 12(6):1484–1489.

Mahmoodi, J., Leckelt, M., van Zalk, M. W., Geukes, K., and Back, M. D. (2017). Big data approaches in social and behavioral science: four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences*, 18:57–62.

Martin, D. M., Alonzo, A., Ho, K.-A., Player, M., Mitchell, P. B., Sachdev, P., and Loo, C. K. (2013). Continuation transcranial direct current stimulation for the prevention of relapse in major depression. *Journal of affective disorders*, 144(3):274–278.

Miller, F. G. and Brody, H. (2002). What makes placebo-controlled trials unethical? *The American Journal of Bioethics*, 2(2):3–9.

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on psychological science*, 7(3):221–237.

Molnar, C., Casalicchio, G., and Bischl, B. (2020). Interpretable machine learning–a brief history, state-of-the-art and challenges. *arXiv preprint arXiv:2010.09337*.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220.

Niculescu-Mizil, A. and Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning*, pages 625–632.

Oltedal, L., Bartsch, H., Sørhaug, O. J. E., Kessler, U., Abbott, C., Dols, A., Stek, M. L., Ersland, L., Emsell, L., van Eijndhoven, P., et al. (2017). The global ect-mri research collaboration (gemric): Establishing a multi-site investigation of the neural mechanisms underlying response to electroconvulsive therapy. *NeuroImage: Clinical*, 14:422–432.

Padberg, F., Bulubas, L., Mizutani-Tiebel, Y., Burkhardt, G., Kranz, G. S., Koutsouleris, N., Kambeitz, J., Hasan, A., Takahashi, S., Keeser, D., et al. (2021). The intervention, the patient and the illness–personalizing non-invasive brain stimulation in psychiatry. *Experimental Neurology*, 341:113713.

Palm, U., Kumpf, U., Behler, N., Wulf, L., Kirsch, B., Wörsching, J., Keeser, D., Hasan, A., and Padberg, F. (2018). Home use, remotely supervised, and remotely controlled transcranial direct current stimulation: a systematic review of the available evidence. *Neuromodulation: Technology at the Neural Interface*, 21(4):323–333.

Passos, I. C. and Mwangi, B. (2018). Machine learning-guided intervention trials to predict treatment response at an individual patient level: an important second step following randomized clinical trials. *Mol Psychiatry*.

Peeters, F., Nicolson, N. A., Berkhof, J., Delespaul, P., and deVries, M. (2003). Effects of daily events on mood states in major depressive disorder. *Journal of abnormal psychology*, 112(2):203.

Platt, J. R. (1998). Strong inference. In *Maritime Archaeology*, pages 155–166. Springer.

Popper, K. R. and Keuth, H. (2005). *Logik der forschung*, volume 11. Mohr Siebeck Tübingen.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Sackeim, H. A., Haskett, R. F., Mulsant, B. H., Thase, M. E., Mann, J. J., Pettinati, H. M., Greenberg, R. M., Crowe, R. R., Cooper, T. B., and Prudic, J. (2001). Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *Jama*, 285(10):1299–1307.

Schödel, R. (2020). *At the Interface of Personality Psychology and Computational Science: Using Smartphones to Investigate Individual Differences*. PhD thesis, München, Ludwig-Maximilians-Universität.

Senn, S. (2018). Statistical pitfalls of personalized medicine.

Shmueli, G. et al. (2010). To explain or to predict? *Statistical science*, 25(3):289–310.

Simon, H. A. (2001). Science seeks parsimony, not simplicity: Searching for pattern in phenomena. *Simplicity, inference and modelling: Keeping it sophisticatedly simple*, pages 32–72.

Soderberg, C. K. (2018). Using osf to share data: A step-by-step guide. *Advances in Methods and Practices in Psychological Science*, 1(1):115–120.

Stachl, C., Hilbert, S., Au, J.-Q., Buschek, D., De Luca, A., Bischl, B., Hussmann, H., Bühner, M., and Wrzus, C. (2017). Personality traits predict smartphone usage. *European Journal of Personality*, 31(6):701–722.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *Plos med*, 12(3):e1001779.

Suen, P. J., Doll, S., Batistuzzo, M. C., Busatto, G., Razza, L. B., Padberg, F., Mezger, E., Bulubas, L., Keeser, D., Deng, Z.-D., et al. (2021). Association between tdcs computational modeling and clinical outcomes in depression: data from the elect-tdcs trial. *European archives of psychiatry and clinical neuroscience*, 271(1):101–110.

Thielscher, A., Opitz, A., and Windhoff, M. (2011). Impact of the gyral geometry on the electric field induced by transcranial magnetic stimulation. *Neuroimage*, 54(1):234–243.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al. (2014). The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior*, 8(2):153–182.

Tzeng, G.-H. and Huang, J.-J. (2011). *Multiple attribute decision making: methods and applications*. CRC press.

Uher, R., Muthén, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., Petrovic, A., Zobel, A., Henigsberg, N., Rietschel, M., et al. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological medicine*, 40(8):1367.

Valiengo, L., Benseñor, I. M., Goulart, A. C., de Oliveira, J. F., Zanao, T. A., Boggio, P. S., Lotufo, P. A., Fregni, F., and Brunoni, A. R. (2013). The sertraline versus electrical current therapy for treating depression clinical study (select-tdcs): results of the crossover and follow-up phases. *Depression and anxiety*, 30(7):646–653.

Valizadeh, S. A., Liem, F., Mérillat, S., Hänggi, J., and Jäncke, L. (2018). Identification of individual subjects on the basis of their brain anatomical features. *Scientific reports*, 8(1):1–9.

Van De Water, A. T., Holmes, A., and Hurley, D. A. (2011). Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography–a systematic review. *Journal of sleep research*, 20(1pt2):183–200.

Von Rueden, L., Mayer, S., Garcke, J., Bauckhage, C., and Schuecker, J. (2019). Informed machine learning–towards a taxonomy of explicit integration of knowledge into machine learning. *Learning*, 18:19–20.

Wilhelm, F. H. and Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological psychology*, 84(3):552–569.

Williams, M. S., Ritchie, M. D., and Payne, P. R. (2015). Interdisciplinary training to build an informatics workforce for precision medicine. *Applied & translational genomics*, 6:28–30.

Wolpert, D. H., Macready, W. G., et al. (1995). No free lunch theorems for search. Technical report, Technical Report SFI-TR-95-02-010, Santa Fe Institute.

Yarkoni, T. (2012). Psychoinformatics: New horizons at the interface of the psychological and computing sciences. *Current Directions in Psychological Science*, 21(6):391–397.

Yarkoni, T. and Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6):1100–1122.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich über den Zeitraum meiner Promotion unterstützt und begleitet haben.

Bei meinem Doktorvater, Prof. Dr. Markus Bühner bedanke ich mich für sein Zutrauen, seine Begeisterungsfähigkeit und die Möglichkeit mich in unterschiedlichen Koorperationen frei bewegen zu können.

Eine wichtige Rolle hat für meine Promotion das interdisziplinäre Arbeiten gespielt. Ich möchte mich herzlich bei Prof. Dr. Frank Padberg bedanken, der mich mit seiner klinschen Expertise und in vielen persönlichen Gesprächen und unseren gemeinsamen Reisen zu Projektpartern sowohl fachlich als auch menschlich unterstützt hat. Selbiges gilt für Prof. Dr. Nina Sarubin und Prof. Dr. Sven Hilbert, denen ich bezüglich meiner akademischen Vernetzung und persönlichen Weiterentwicklung an der Universität sehr viel zu verdanken habe. Unserem brasilianischen Partner Prof. Dr. Brunoni gilt ein besonderer Dank. Ohne die enge und persönliche Zusammenarbeit sowie die großzügige Bereitstellung der Daten von Trials, die als Meilensteine in dem Feld der transkraniellen Hirnstimulation angesehen werden können, wäre diese Promotion nicht umsetzbar gewesen. Ein großes Dankeschön gilt auch an die Kollegen der unterschiedlichen Arbeitsgruppen, Dr. Daniel Keeser, Prof. Dr. Joseph Kambeitz, Dr. Jana Werle, Dr. Matthias Reinhardt, Dr. Lucia Bulubas, Eva Mezger und Gerrit Burkhardt. Zu guter Letzt ein großes Dankeschön an meine Familie und Freunde. Ich danke meinen Eltern, dass sie mir diesen Weg ermöglicht haben, mir Rückhalt geben und immer für mich da sind.

Schließlich geht mein Dank an meine Esther. Danke, dass Du an meiner Seite bist, mir Kraft gibst und auch ein Auge darauf hast, dass nicht immer nur alles Arbeit ist.