Aus dem Institut für Systemische Neurowissenschaften

der Heinrich-Heine-Universität Düsseldorf

Direktor: Univ.-Prof. Dr. med. Simon Eickhoff

Passive smartphone-basierte Aktivitätserfassung und aktive Symptomerhebung mittels Ecological Momentary Assessment bei Erwachsenen mit depressiver Störung

–

Ergebnisse einer klinischen Kohortenstudie mit Verwendung der Smartphone-Apps „JTrack Social“ und „JTrack EMA“

Dissertation

zur Erlangung des Grades eines Doktors der Medizin

der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Theokritos Veskoukis

2026

Als Inauguraldissertation gedruckt mit Genehmigung der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Prof. Dr. med. Nikolaj Klöcker

Erstgutachter: Prof. Dr. Jürgen Dukart

Zweitgutachter: Prof. Dr. med. Leonhard Schilbach

Zusammenfassung

Veränderungen der körperlichen Aktivität gehören zu den zentralen Merkmalen depressiver Störungen und werden mit Antrieb, Affekt und psychosozialer Funktionsfähigkeit in Verbindung gebracht. Die kontinuierliche Aufzeichnung von Bewegungsdaten über Smartphones ist potenziell geeignet, sogenannte digitale Verhaltensmarker zu erfassen, die im Rahmen der Diagnostik und Verlaufsbeurteilung depressiver Störungen unterstützend eingesetzt werden könnten. Bislang liegen jedoch nur wenige Studien vor, die passiv erhobene Aktivitätsdaten und engmaschige Selbstauskünfte über mehrere Monate in einer klinisch charakterisierten Depressionskohorte kombinieren.

Die Smartphone-App „JTrack Social" wurde in der vorliegenden Studie eingesetzt, um über einen Zeitraum von 90 Tagen Aktivitätsdaten von erwachsenen Patientinnen und Patienten mit einer depressiven Störung zu erheben. Ergänzend wurden mit der App „JTrack EMA" wöchentliche Selbstauskünfte zum subjektiven Befinden erhoben, unter anderem zu Stimmung, Energie, Stress, Angst und Schlafqualität. Die klinische Fremdbeurteilung der Psychopathologie erfolgte zu Studienbeginn anhand der Brief Psychiatric Rating Scale (BPRS). Aufgezeichnet wurde, wie lange pro Tag die Aktivitätskategorien Gehen, Laufen, Fahrradfahren und Inaktivität registriert wurden [Min.]. Die Statistik wurde mit Hilfe von linearen Regressionsmodellen und linearen gemischten Modellen mit FDR-Korrektur für multiples Testen berechnet. Es wurde postuliert, dass eine höhere passiv erfasste körperliche Aktivität mit günstigeren subjektiven Selbstberichten und mit einer geringeren klinisch beurteilten Symptomschwere einhergeht..

Im Zeitraum von November 2021 bis März 2023 wurden 50 Patientinnen und Patienten mit einer klinisch diagnostizierten depressiven Störung in die Studie aufgenommen (31 weiblich, 19 männlich; mittleres Alter 39,0 Jahre). 21 Teilnehmende (42 %) absolvierten die gesamte Studiendauer von 90 Tagen. Die mittlere tägliche Gesamtaktivität betrug 53,7 Minuten (SD = 58,8) bei einer mittleren Inaktivitätszeit von 934 Minuten pro Tag; die aktive Zeit verteilte sich nahezu vollständig auf Gehen. Entgegen der Annahme zeigten sich nach FDR-Korrektur weder signifikante Zusammenhänge zwischen den Aktivitätsparametern und der klinischen Symptomschwere noch konsistente Assoziationen mit den wöchentlichen Selbstauskünften. Auffällig war eine ausgeprägte interindividuelle Heterogenität des Bewegungsverhaltens (Gesamtaktivität: 0–411 Min./Tag).

Diese Studie beschreibt somit das alltagsnahe Bewegungsverhalten von erwachsenen Patientinnen und Patienten mit einer depressiven Störung über einen Zeitraum von 90 Tagen anhand automatisch erfasster Smartphone-Daten. Die Befunde verdeutlichen den Bedarf an größeren Stichproben, längeren Beobachtungszeiträumen und einer engeren zeitlichen Kopplung klinischer Erhebungen an digitale Messungen. Die Arbeit liefert konkrete methodische Hinweise für die Konzeption zukünftiger digitaler Studien zur Entwicklung eines digitalen Verhaltensmarkers für Erwachsene mit depressiven Störungen.

Abstract

Changes in physical activity are a core feature of depressive disorders and are closely linked to drive, affect, and psychosocial functioning. The continuous recording of movement data via smartphones is potentially suitable to capture so-called digital behavioural markers, which could be used in a supportive manner for the diagnosis and longitudinal assessment of depressive disorders. To date, however, only a limited number of studies have combined passively recorded activity data with frequent self-reports over several months in a clinically characterised cohort of patients with depression.

The smartphone application 'JTrack Social' was used in the present study to collect activity data over a period of 90 days from adult patients with a depressive disorder. In addition, the application 'JTrack EMA' was used to record weekly self-reports on subjective well-being, including mood, energy, stress, anxiety, and sleep quality. Clinician-rated psychopathology was assessed at study onset by means of the Brief Psychiatric Rating Scale (BPRS). It was recorded how long per day the activity categories walking, running, cycling, and inactivity were registered [min]. The statistic was calculated using linear regression models and linear mixed models with FDR correction for multiple testing. It was postulated that a higher level of passively recorded physical activity is associated with more favourable self-reported states and with lower clinician-rated symptom severity.

In the period between November 2021 and March 2023, 50 patients with a clinically diagnosed depressive disorder were included in the study (31 female, 19 male; mean age 39.0 years). Twenty-one participants (42 %) completed the full study duration of 90 days. The mean daily total activity was 53.7 minutes (SD = 58.8) with a mean inactivity time of 934 minutes per day; active time was almost entirely attributable to walking. Contrary to the hypothesis, no statistically significant associations were found after FDR correction, neither between the activity parameters and clinical symptom severity nor as consistent associations with the weekly self-reports. A pronounced inter-individual heterogeneity in movement behaviour was apparent (total activity: 0–411 min/day).

This study therefore describes everyday movement behaviour in adult patients with a depressive disorder over a period of 90 days using automatically recorded smartphone data. The findings underscore the need for larger samples, longer observation periods, and a closer temporal coupling of clinical assessments to digital measurements. This thesis provides concrete methodological guidance for the design of future digital studies on the development of a digital behavioural marker for adults with depressive disorders.

Abkürzungsverzeichnis

ACTV	Activation (BPRS-Subskala Aktivierung)
AI Act	Verordnung über künstliche Intelligenz (Verordnung (EU) 2024/1689)
ANDP	Anxiety-Depression (BPRS-Subskala Angst-Depression)
ANER	Anergia (BPRS-Subskala Antriebslosigkeit)
API	Programmierschnittstelle (application programming interface)
BEST	„Biomarkers, EndpointS, and other Tools“ (FDA-NIH Biomarker Working Group, 2025)
BPRS	„Brief Psychiatric Rating Scale“
BPRS-E	„Brief Psychiatric Rating Scale – Expanded“ (Zanello et al., 2013)
BQP	„Biomarker Qualification Program“ (FDA)
CE	Conformité Européenne (CE-Kennzeichnung)
CGI	„Clinical Global Impression“
CGI-S-MV	„Clinical Global Impression – Severity“, modifizierte Version
COU	Anwendungskontext (context of use)
DDT	„Drug Development Tool“ (FDA)
DSGVO	Datenschutz-Grundverordnung
EMA	„Ecological Momentary Assessment“ / European Medicines Agency
FDA	U.S. Food and Drug Administration
FDR	Falscherkennungsrate (false discovery rate)
FZ	Forschungszentrum
GAF	„Global Assessment of Functioning“
GBD	„Global Burden of Disease“-Studie
GDPR	„General Data Protection Regulation“ (englische Bezeichnung der DSGVO)
GPS	„Global Positioning System“
HAMD	„Hamilton Depression Rating Scale“

HOST	Hostility (BPRS-Subskala Feindseligkeit)
HR	Hazard Ratio
ICD-10	International Statistical Classification of Diseases and Related Health Problems – 10th Version
IQR	Interquartilsabstand (interquartile range)
KI	Konfidenzintervall / Künstliche Intelligenz
LMM	Lineares gemischtes Modell (linear mixed model)
M	Mittelwert (mean)
MADRS	„Montgomery-Åsberg Depression Rating Scale“
MDCG	„Medical Device Coordination Group“
MDR	Medizinprodukteverordnung (Verordnung (EU) 2017/745)
MDSW	„Medical Device Software“
mMET-h	marginal metabolic equivalent of task – hours
NNT	„Number Needed to Treat“
OPA	berufliche körperliche Aktivität (occupational physical activity)
OR	„Odds Ratio“
PHQ-9	„Patient Health Questionnaire – 9 Items“
QALE	qualitätsadjustierte Lebenserwartung (quality-adjusted life expectancy)
QALY	qualitätsadjustiertes Lebensjahr (quality-adjusted life year)
RCT	randomisierte kontrollierte Studie (randomized controlled trial)
RDoC	„Research Domain Criteria“-Framework
RR	Relatives Risiko (risk ratio)
SD	Standardabweichung (standard deviation)
SE	Standardfehler (standard error)
SMD	standardisierte Mittelwertdifferenz (standardized mean difference)
THOT	Thought Disorder (BPRS-Subskala Denkstörung)
YLD	„Years Lived with Disability“

Inhaltsverzeichnis

1 Einleitung 1
1.1 Einführung in die Arbeit 1
1.2 Depression – klinischer und psychosozialer Kontext 1
1.3 Krankheitslast und gesellschaftliche Folgen depressiver Störungen 4
1.3.1 Krankheitslast im globalen Vergleich 4
1.3.2 Arbeitsfähigkeit und Erwerbsleben 5
1.3.3 Gesundheitsökonomische Folgen 5
1.3.4 Lebensqualität und gesunde Lebensjahre (QALYs) 6
1.3.5 Public-Health-Perspektive 6
1.4 Prävention und Therapie durch körperliche Aktivität 6
1.4.1 Prävention: körperliche Aktivität senkt das Depressionsrisiko 6
1.4.2 Wirksamkeit von Bewegungstherapie bei Depression 7
1.4.3 Ambulante Nachsorge, Verlaufskontrolle und Exazerbationen bei depressiven Störungen 8
1.4.4 BPRS 10
1.5 Präzisionspsychiatrie 11
1.5.1 Digitale Biomarker in der Psychiatrie 12
1.5.2 EMA 14
1.5.3 Passive Datenerfassung 16
1.5.4 JTrack 17
1.5.5 Forschungslücke und Beitrag der vorliegenden Arbeit 18
1.6 Ziele und Hypothesen 18
2 Material und Methoden 21
2.1 Studiendesign 21
2.2 Studiendurchführung 22
2.2.1 Ein- und Ausschlusskriterien 22
2.2.2 Rekrutierung 23
2.2.3 Ethik und Einwilligungsprozess 25
2.2.4 Datenschutz und Datenverarbeitung 25
2.3 Datenerhebung 26
2.3.1 Demografische und klinische Basisdaten 26
2.3.2 Smartphone-App: „JTrack Social" 26
2.3.3 Smartphone-App: „JTrack EMA" 28
2.3.4 Pseudonymisierung 30

2.4 Datenanalyse 30
2.4.1 Datenaufbereitung 30
2.4.2 Auswertung der demographischen Daten 37
2.4.3 Analyse der Smartphonedaten 37
3 Ergebnisse 39
3.1 Rekrutierung, Studienzeitraum und Dropouts 39
3.2 Ergebnisse der demographischen und psychometrischen Daten 41
3.2.1 Verteilung von Alter und Geschlecht 41
3.2.2 EMA 43
3.2.3 BPRS 45
3.3 Ergebnisse aus den Smartphonedaten 46
3.3.1 Vollständigkeit der Daten 46
3.3.2 Vergleich der physischen Aktivität mit der selbstberichteten Psychopathologie 48
3.3.3 Fremdbeurteilung und physische Aktivität 53
4 Diskussion 56
4.1 Analyse der körperlichen Aktivität und psychischer Symptomatik 56
4.2 Analyse der Studiendurchführung 60
4.3 Limitationen 63
4.4 Praktische und ethische Implikationen 65
4.5 Perspektiven zur Qualifizierung digitaler Verhaltensmarker als Biomarker nach FDA-Standards 66
4.6 Verifikation und analytische Validierung der JTrack-Social-Daten 66
4.7 Klinische Validierung und terminologische Einordnung 67
4.8 Regulatorischer Rahmen in Europa: MDR, AI Act, DSGVO 68
5 Schlussfolgerung 69
Literaturverzeichnis 71
6 Anhang 81
6.1 Demographischer Fragebogen 81
6.2 Rekrutierungsflyer 84
6.3 Vollständige Ergebnisse der induktiven Statistik 85

1 Einleitung

1.1 Einführung in die Arbeit

In der klinischen Praxis zeigt sich, dass die körperliche Aktivität von Erwachsenen mit depressiven Störungen im Vergleich zur gesunden Bevölkerung deutlich verändert ist. Typisch sind eine Einschränkung der Bewegungsbereitschaft und eine insgesamt reduzierte Alltagsaktivität, die sich sowohl in der Häufigkeit als auch in der Intensität körperlicher Betätigung niederschlägt (Wüthrich et al., 2022). In der vorliegenden Arbeit wird zunächst das Krankheitsbild der Depression mit seinen klinischen und psychosozialen Herausforderungen dargestellt. Anschließend erfolgt eine Analyse der bisherigen Forschung zur körperlichen Aktivität bei depressiven Störungen, mit besonderem Fokus auf deren objektive Erfassung durch digitale Verfahren. In diesem Zusammenhang wird die in der eigenen Studie eingesetzte smartphone-basierte Aktivitätsmessung erläutert, die eine kontinuierliche Erfassung von Bewegungsmustern im Alltag der Patient*innen ermöglicht. Ziel der Arbeit ist es, die Zusammenhänge zwischen smartphone-basierten Markern körperlicher Aktivität, der klinischen Fremdbeurteilung mittels BPRS und den selbstberichteten Symptomen aus EMA-Items systematisch zu untersuchen.

1.2 Depression – klinischer und psychosozialer Kontext

Als affektive Störung gehört die Depression weltweit zu den häufigsten psychischen Erkrankungen (GBD 2019 Mental Disorders Collaborators, 2022). Sie ist gekennzeichnet durch affektive, kognitive, somatische und verhaltensbezogene Symptome – wie gedrückte Stimmung, Interessen- oder Freudverlust, Antriebs- und Konzentrationsstörungen, Schuld- und Wertlosigkeitsgefühle, Schlaf- und Appetitveränderungen sowie Suizidgedanken –, die über einen Zeitraum von mindestens zwei Wochen anhalten und zu einer erheblichen Beeinträchtigung des psychosozialen Funktionsniveaus führen (*Nationale Versorgungsleitlinie Unipolare Depression 2023*). Die hohe interindividuelle Heterogenität in Ausprägung, Verlauf und Symptomkonstellation erfordert ein differenziertes und strukturiertes Vorgehen in Diagnostik und Therapie (*Nationale Versorgungsleitlinie Unipolare Depression 2023*). Im Erwachsenenalter ist die unipolare Depression besonders häufig; in Deutschland liegt die Lebenszeitprävalenz bei etwa 16–17 %, wobei Frauen signifikant häufiger betroffen sind als Männer (Jacobi et al., 2014). In der Regel beginnt die Erkrankung im frühen Erwachsenenalter. Der Verlauf ist häufig rezidivierend mit kumulativem Rezidivrisiko von 42 % innerhalb von 20 Jahren (Hardeveld et al., 2013), während etwa 15 % der Betroffenen einen chronischen, unremittierenden Verlauf ent-

wickeln (Eaton et al., 2008). Psychiatrische Komorbiditäten verschlechtern die Prognose deutlich. Bevölkerungsbasierte Registerdaten zeigen, dass nach einer depressiven Episode das Risiko für Persönlichkeitsstörungen (HR = 4,3) und Schizophrenie-Spektrum-Störungen (HR = 5,8) signifikant erhöht ist (Plana-Ripoll et al., 2019). Eine Meta-Analyse belegt zudem, dass komorbide Persönlichkeitsstörungen mit einem mehr als doppelt erhöhten Risiko für ein ungünstiges Depressions-Outcome assoziiert sind (OR = 2,16) (Newton-Howes et al., 2006).

Darüber hinaus weisen Personen mit Autismus-Spektrum-Störungen ein deutlich erhöhtes Risiko für depressive Störungen auf (aOR = 3,64) (Rai et al., 2018).

Auch kardiovaskuläre Erkrankungen stehen in enger Wechselwirkung mit Depression; depressive Störungen gehen mit einem etwa 1,8-fach erhöhten Risiko für kardiovaskuläre Ereignisse einher (RR ≈ 1,8) (Li et al., 2023).

Die Diagnose einer Depression wird anhand internationaler Klassifikationssysteme, vor allem des ICD-10 (World Health Organization, 1992) gestellt und wird über die so genannten Kernsymptome bzw. Hauptkriterien wie depressive Stimmung, Interessenverlust und Antriebsminderung sowie über Zusatzsymptome wie Schuldgefühle, Suizidgedanken oder Schlaf- und Appetitstörungen definiert. Ein wichtiges Kriterium ist die Zeitspanne des Vorliegens der Symptome: Die Symptome müssen an den meisten Tagen über mindestens zwei Wochen bestehen und eine klinisch relevante Beeinträchtigung verursachen (*Nationale Versorgungsleitlinie Unipolare Depression 2023*).

Zur Diagnose einer depressiven Episode müssen gemäß ICD-10 die oben beschriebenen operationalisierten Kriterien erfüllt sein. Die Symptome müssen mindestens zwei Wochen bestehen und mindestens zwei der drei typischen Symptome (depressive Stimmung, Interessenverlust, verminderter Antrieb) umfassen. Die Einstufung des Schweregrades erfolgt anhand der Anzahl typischer und weiterer Symptome: Eine leichte Episode (F32.0) liegt bei mindestens vier Symptomen insgesamt vor, eine mittelgradige Episode (F32.1) bei mindestens sechs Symptomen und eine schwere Episode (F32.2) bei mindestens acht Symptomen, wobei bei schweren Episoden alle drei typischen Symptome vorliegen müssen (*Nationale Versorgungsleitlinie Unipolare Depression 2023*).

Die systematische Erfassung des Schweregrades ist für die Therapieplanung von zentraler Bedeutung. Neben dem klinischen Urteil sind standardisierte Fremdbeurteilungsskalen wie die Hamilton Depression Rating Scale (HAMD), die Montgomery-Asberg Depression Rating Scale

(MADRS) oder der Patient Health Questionnaire (PHQ-9) verfügbar. Die Brief Psychiatric Rating Scale (BPRS) ermöglicht darüber hinaus eine breitere Erfassung allgemeiner psychopathologischer Symptome, einschließlich affektiver und psychotischer Dimensionen, und ist in ihrer erweiterten 24-Item-Version (BPRS-E) reliabel und sensitiv für Veränderungen beschrieben (Zanello et al., 2013).

Selbst- und Fremdbeurteilungen depressiver Symptome stimmen gut, aber nicht vollständig überein. In einer großen klinischen Stichprobe lag die Übereinstimmung für Therapieansprechen bei $\kappa = 0{,}64$ und für Remission bei $\kappa = 0{,}57$, mit relevanten Diskrepanzen auf Individualebene (Dunlop et al., 2014). Eine Vergleichsstudie mit therapieresistenten und therapienaiven Patient*innen ergab moderate bis starke Korrelationen zwischen Selbst- und Fremdratings ($r = 0{,}51–0{,}68$) (Hershenberg et al., 2020). Darüber hinaus zeigten zwei große Antidepressivastudien, dass Selbst- und Fremdbeurteilungen jeweils zusätzliche Varianz im Therapieoutcome erklärten (ca. 1–4 %), selbst nach Adjustierung für die jeweils andere Skala, was darauf hinweist, dass beide Verfahren unterschiedliche, klinisch relevante Aspekte der Depression erfassen (Uher et al., 2012). Ergänzend ermöglicht Ecological Momentary Assessment (EMA) eine wiederholte, zeitgestempelte Erfassung der aktuellen Stimmung im natürlichen Lebensumfeld und erlaubt damit die quantitative Analyse von Durchschnittsniveau, Variabilität (Standardabweichung) und emotionaler Inertia (Autokorrelation) über die Zeit. Dadurch können intraindividuelle Dynamiken depressiver Symptome differenziert abgebildet werden, die in retrospektiven Globalurteilen nicht erfasst werden (Van Genugten et al., 2021). Eine ausführlichere Darstellung des EMA-Verfahrens findet sich in Abschnitt 1.5.2.

Eine präzise Diagnostik erfordert die Abgrenzung von anderen psychischen Erkrankungen (z. B. bipolare Störungen mit manischen oder hypomanen Episoden, Angst- und Zwangsstörungen, depressive Symptomatik bei psychotischen Störungen), von somatischen Erkrankungen (z. B. endokrine, neurologische, onkologische oder systemische Erkrankungen) sowie von substanz- oder medikamenteninduzierten Syndromen. Auch die Anpassungsstörung mit depressiver Reaktion ist abzugrenzen; sie wird durch einen klar identifizierbaren Stressor ausgelöst und hat in der Regel eine kürzere Dauer. Eine systematische diagnostische Einordnung sollte den zeitlichen Verlauf der Symptomatik, mögliche auslösende Faktoren, komorbide Erkrankungen und funktionelle Beeinträchtigungen berücksichtigen. (*Nationale Versorgungsleitlinie Unipolare Depression 2023*) (Malhi & Mann, 2018).

Suizidalität stellt eines der gravierendsten Risiken depressiver Erkrankungen dar und sollte daher systematisch in der Basisdiagnostik erfasst werden (*Nationale Versorgungsleitlinie Unipolare Depression 2023*). Epidemiologische Studien zeigen, dass etwa 30–50 % der Patient*innen mit Major Depression im Verlauf einer Episode suizidale Gedanken berichten. Das Lebenszeitrisiko für einen Suizidversuch bei Major Depression liegt in bevölkerungsbasierten Studien im Bereich von 10–20 %. Das Risiko ist deutlich mit dem Schweregrad der depressiven Episode assoziiert und steigt insbesondere bei ausgeprägter Hoffnungslosigkeit, psychotischen Symptomen sowie komorbiden Substanz- oder Angststörungen. (Malhi & Mann, 2018).

Aktuelle Meta-Analysen prospektiver Kohortenstudien bestätigen diese Befunde und zeigen für den Suizidtod bei Major Depression signifikante Prädiktoren wie männliches Geschlecht (HR = 3,27), vorangegangene Suizidversuche (HR = 2,10), komorbide Angststörungen (HR = 1,77), schwere depressive Symptomatik (HR = 1,52) sowie Alkohol- oder Substanzmissbrauch (HR = 1,44) (Li et al., 2022). Neuere Meta-Analysen heben insbesondere frühere Suizidversuche, Schweregrad der Depression und Hoffnungslosigkeit als zentrale Risikofaktoren hervor (Riera-Serra et al., 2024). Darüber hinaus zeigen vergleichende Analysen, dass das Risiko suizidalen Verhaltens bei affektiven Störungen, einschließlich Major Depression, im Vergleich zur Allgemeinbevölkerung signifikant erhöht ist (Arnone et al., 2024).

1.3 Krankheitslast und gesellschaftliche Folgen depressiver Störungen

1.3.1 Krankheitslast im globalen Vergleich

Internationale Krankheitslaststudien verdeutlichen die hohe Relevanz depressiver Störungen. Im Global Burden of Disease (GBD) 2010 rangierte die Major Depression weltweit als zweithäufigste Ursache für verlorene gesunde Lebensjahre (YLDs) und trug 8,2 % zur globalen YLD-Belastung bei (Ferrari et al., 2013). Auch in der aktuellen GBD-Analyse 2019 blieb die Belastung durch depressive Störungen unverändert hoch und gehört weiterhin zu den zehn führenden Ursachen von Behinderung (GBD 2019 Mental Disorders Collaborators, 2022).

1.3.2 Arbeitsfähigkeit und Erwerbsleben

Die psychosozialen Folgen der Depression treten im Erwerbsleben besonders deutlich hervor. Bereits leichte depressive Symptome erhöhen signifikant das Risiko für langanhaltende Arbeitsunfähigkeit, bei mittelgradigen bis schweren Symptomen verdreifacht sich das Risiko für Frühpensionierung und dauerhafte Erwerbslosigkeit (Rose et al., 2023).

Studien belegen zudem eine reduzierte Erwerbsperspektive. In Dänemark verlieren Frauen im Durchschnitt 3,3 und Männer 2,3 Erwerbsjahre. Die von Depression betroffenen Patient*innen verbringen mehr Zeit in Arbeitslosigkeit und krankheitsbedingter Abwesenheit (Pedersen et al., 2019). Eine US-amerikanische Längsschnittstudie zeigt, dass persistierende Depressionen die erwartete Dauer des Erwerbslebens um bis zu 17 Jahre verkürzen können (Dobson et al., 2024). Registerdaten aus Schweden verdeutlichen, dass viele Betroffene nach längerer depressionsbedingter Arbeitsunfähigkeit dauerhaft in Mustern von Langzeiterkrankung, Arbeitslosigkeit oder Frühpensionierung verbleiben (Stutaite et al., 2025). Besonders ausgeprägt ist dieses Risiko bei therapieresistenter Depression, die mit häufigeren Krankheitsausfällen, geringerer Rückkehr ins Erwerbsleben und höheren Invaliditätsraten verbunden ist (Gronemann et al., 2022).

Auch zur Dauer krankheitsbedingter Fehlzeiten liegen belastbare Daten vor: In einer niederländischen Studie betrugen die durchschnittlichen Fehlzeiten pro depressive Episode mehr als 200 Tage, wobei Frauen länger ausfielen als Männer; rund ein Viertel der Betroffenen war ein Jahr oder länger krankgeschrieben (Koopmans et al., 2008).

1.3.3 Gesundheitsökonomische Folgen

Die volkswirtschaftliche Last ist erheblich: Für 2010 schätzten Gustavsson et al. die Gesamtkosten psychischer und neurologischer Erkrankungen in Europa auf 798 Mrd. €, davon 113 Mrd. € für affektive Störungen wie Depression. Einen großen Anteil tragen indirekte Kosten durch Produktivitätsverluste, Fehlzeiten und Frühberentung. Soziale Faktoren verstärken die Belastung: Beschäftigte mit geringerem Bildungsniveau haben ein signifikant höheres Risiko für wiederholte Arbeitsunfähigkeiten und geringere Chancen auf eine nachhaltige Rückkehr in den Beruf (Gustavsson et al., 2011; Meling et al., 2023).

1.3.4 Lebensqualität und gesunde Lebensjahre (QALYs)

Depressive Störungen gehen weit über Erwerbsfähigkeit und Kosten hinaus und führen zu erheblichen Verlusten an Lebensqualität und gesunden Lebensjahren. In US-Daten beträgt der Verlust an qualitätsadjustierter Lebenserwartung (QALE) bei Erwachsenen mit aktueller Depression etwa 28,9 Jahre, mehr als bei vielen anderen chronischen Erkrankungen (Jia et al., 2015). Bereits milde Symptome im höheren Alter sind mit deutlichen QALY-Einbußen verbunden (≈ −6,2 Jahre), bei wahrscheinlicher Major Depression mit rund −8,3 Jahren; je nach Kohorte/Modellierung werden teils höhere Werte berichtet (Jia & Lubetkin, 2017).

1.3.5 Public-Health-Perspektive

Angesichts der hohen Prävalenz, der chronischen und rezidivierenden Verläufe, der ökonomischen Folgekosten und der Einschränkung von Lebensqualität wird Depression zunehmend als zentrales Public-Health-Problem betrachtet. (McLaughlin, 2011) fordert eine stärkere interdisziplinäre Ausrichtung, die neben Behandlung auch bevölkerungsweite Präventionsmaßnahmen adressiert, um die Krankheitslast nachhaltig zu reduzieren.

1.4 Prävention und Therapie durch körperliche Aktivität

1.4.1 Prävention: körperliche Aktivität senkt das Depressionsrisiko

Prospektive Kohortenstudien und Metaanalysen stützen eine inverse Assoziation zwischen körperlicher Aktivität und dem Risiko für das erstmalige Auftreten depressiver Episoden. In einer Metaanalyse über 36 Kohorten lag die gepoolte adjustierte Odds Ratio für neu auftretende Depression bei 0,837 (95% KI 0,794–0,883) für höhere vs. niedrigere körperliche Aktivität (Schuch et al., 2018).

Eine Dosis-Response-Metaanalyse zeigte darüber hinaus eine inverse, nichtlineare Beziehung: Bereits ein Aktivitätsumfang von 4,4 mMET-h/Woche (marginal metabolic equivalent of task hours per week) war mit einer 18% geringeren Depressionsinzidenz assoziiert, während bei 8,8 mMET-h/Woche (entspricht der empfohlenen Mindestmenge) eine 25% Risikoreduktion berichtet wurde; der zusätzliche Nutzen nahm bei höheren Dosen ab. Eine weitere große Evidenzsynthese über 111 prospektive Kohortenstudien (über 3 Millionen Erwachsene) bestätigte diesen Zusammenhang: Die adjustierten Odds für neu auftretende Depression bzw. Zunahme

subklinischer Symptome waren nach Exposition gegenüber körperlicher Aktivität um 22% reduziert (adjusted OR 0,79, 95% KI 0,75–0,82). In derselben Arbeit waren die Odds niedriger, wenn Aktivität die Leitlinien erfüllte (z. B. OR 0,77, 95% KI 0,73–0,81) und wenn Aktivität im Verlauf anstieg (z. B. OR 0,69, 95% KI 0,61–0,79) im Vergleich zu einer einmaligen Baseline-Erfassung (Dishman et al., 2021). Für die Interpretation präventiver Befunde ist zudem die Aktivitätsdomäne bedeutsam, da berufliche körperliche Aktivität (occupational physical activity-OPA) nicht ohne Weiteres mit Freizeitaktivität gleichgesetzt werden kann (Park et al., 2025). In einer systematischen Übersichtsarbeit mit Dosis-Response-Metaanalyse war hohe OPA im Vergleich zu niedriger OPA mit einem erhöhten Depressionsrisiko assoziiert (gepoolter RR 1,25, 95% KI 1,12–1,40). In der linearen Dosis-Response-Analyse war jede zusätzliche Stunde OPA pro Woche mit einem erhöhten Depressionsrisiko verbunden (RR 1,05, 95% KI 1,01–1,09). (Park et al., 2025)

Ergänzend zur Evidenz zu strukturierten Bewegungsinterventionen liefert eine systematische Übersichtsarbeit Hinweise darauf, dass Gehen in natürlichen Umgebungen mit Verbesserungen verschiedener mentaler Outcomes assoziiert sein kann. In die Review wurden 17 Studien mit insgesamt 1.209 erwachsenen Teilnehmenden eingeschlossen. Aufgrund der ausgeprägten Heterogenität der Studien erfolgte eine narrative Synthese der Ergebnisse. Insgesamt weisen die Befunde darauf hin, dass Gehen in natürlichen Umgebungen (nature-based walking) mit Verbesserungen der Stimmung, des Optimismus, des mentalen Wohlbefindens sowie der Naturverbundenheit verbunden sein kann und zugleich mit einer Reduktion von Stress, Angst und Rumination einhergeht. Darüber hinaus wurden im Vergleich zu Gehen in urbanen Umgebungen teilweise stärkere Effekte beobachtet, insbesondere hinsichtlich Angst und Rumination (Ma et al., 2024).

1.4.2 Wirksamkeit von Bewegungstherapie bei Depression

Randomisierte Studien und Metaanalysen zeigen, dass strukturierte Bewegungsinterventionen depressive Symptome reduzieren können, wobei Effekte je nach Modalität und Vergleichsbedingung variieren. In einer Netzwerk-Metaanalyse gegenüber aktiven Kontrollbedingungen wurden symptomreduzierende Effekte für Gehen/Joggen (Hedges' g −0,62), Yoga (g −0,55) und Krafttraining (g −0,49) berichtet (Noetel et al., 2024). Zusätzlich wurde beschrieben, dass die Effekte proportional zur verordneten Intensität ausfielen und dass professionell angeleitete Programme mit größeren Effekten und höherer Adhärenz assoziiert waren. Ergänzend zeigte

eine Meta-Analyse von 41 RCTs gegenüber nicht-aktiven Kontrollbedingungen einen großen Gesamteffekt (SMD −0,946, 95% KI −1,18 bis −0,71; NNT = 2) sowie bei Beschränkung auf Studien mit niedrigem Verzerrungsrisiko einen moderaten Effekt (SMD −0,666, 95% KI −0,99 bis −0,34; NNT = 2,8). In einer randomisierten Studie wurden nach 16 Wochen Remissionsraten von 45–47% in den aktiven Behandlungsarmen gegenüber 31% unter Placebo berichtet (Cooney et al., 2013).

Über kurzfristige Effekte hinaus liefern Follow-up-Daten quantitative Hinweise auf Zusammenhänge zwischen fortgesetzter Aktivität und klinischem Verlauf. Im 1-Jahres-Follow-up war der erwartete Unterschied im HAM-D zwischen 180 Minuten/Woche versus 0 Minuten/Woche körperlicher Aktivität ein Rückgang um 3,1 Punkte (95% KI −5,1 bis −1,2). Zudem war die Odds, sich um eine diagnostische Kategorie zu verbessern, bei 180 Minuten/Woche im Vergleich zu 0 Minuten/Woche etwa 2,9-fach erhöht (95% KI 1,1–7,3), wobei der Zusammenhang als kurvilinear beschrieben wurde (Hoffman et al., 2011).

Ergänzend hierzu fasst eine Umbrella Review die Evidenz aus 97 systematischen Reviews (insgesamt 1.039 randomisierte kontrollierte Studien mit 128.119 Teilnehmenden) zu körperlicher Aktivität zusammen. Über verschiedene Populationen hinweg zeigte sich für depressive Symptome ein medianer Effekt von SMD = −0,43 (IQR −0,66 bis −0,27) im Vergleich zu *usual care* bzw. Kontrollbedingungen. Subgruppenanalysen nach Interventionsintensität weisen auf stärkere Effekte bei moderater (SMD = −0,56; IQR −1,03 bis −0,33) und hoher Intensität (SMD = −0,70; IQR −1,25 bis −0,24) im Vergleich zu niedriger Intensität (SMD = −0,22; IQR −0,50 bis −0,12) hin. Analysen nach Interventionsdauer legen zudem nahe, dass die Effektstärken mit zunehmender Programmdauer tendenziell abnehmen (Singh et al., 2023).

1.4.3 Ambulante Nachsorge, Verlaufskontrolle und Exazerbationen bei depressiven Störungen

Die ambulante Behandlung depressiver Störungen ist in besonderem Maße auf eine zuverlässige Verlaufskontrolle angewiesen. Im Gegensatz zu akuten somatischen Erkrankungen zeichnet sich die Depression jedoch durch eine ausgeprägte zeitliche Dynamik aus. Längsschnittanalysen zeigen, dass depressive Symptomatik innerhalb von Individuen erheblich schwanken kann und dass diese Symptomvariabilität selbst prognostische Relevanz besitzt (Price et al., 2023). Variabilität ist nicht lediglich ein methodisches Nebenprodukt, sondern eine notwendige Voraussetzung für Rückfälle, Remissionen und Exazerbationen. Darüber hinaus wurde hohe

Symptomvariabilität mit erhöhtem Suizidrisiko und ungünstigeren Langzeitverläufen assoziiert (Price et al., 2023). Eine adäquate Verlaufsbeobachtung muss diese intraindividuelle Dynamik daher systematisch erfassen.

Die ambulante Routineversorgung basiert jedoch überwiegend auf punktuellen Konsultationen und retrospektiver Selbstberichterstattung. Instrumente wie der PHQ-9 erfassen Symptome typischerweise über einen zweiwöchigen Rückblickzeitraum und werden in vielen Settings in mehrmonatigen Abständen erhoben. Dadurch entsteht eine strukturelle Diskrepanz zwischen der tatsächlichen Dynamik depressiver Symptome und der zeitlichen Auflösung ihrer Erfassung. Price et al. (2023) weisen explizit darauf hin, dass eine dreimonatige Erhebungsfrequenz bei gleichzeitigem zweiwöchigem Recall-Fenster nur begrenzt geeignet ist, natürliche Schwankungen zuverlässig abzubilden. Kurzfristige Verschlechterungen, subsyndromale Phasen oder rasche Stimmungswechsel können dadurch unentdeckt bleiben.

Diese Problematik ist nicht auf einzelne Fragebögen beschränkt, sondern betrifft das traditionelle psychiatrische Assessment insgesamt. Psychopathologische Phänomene sind zeitlich variabel, kontextabhängig und situativ moduliert. Gleichzeitig erfolgen klinische Bewertungen meist in separaten Untersuchungssituationen. Torous et al. (2017) argumentieren, dass viele psychiatrische Symptome temporär und fluktuierend auftreten und daher bei infrequenten klinischen Untersuchungen systematisch unterschätzt werden können. Hinzu kommt die bekannte Limitation retrospektiver Selbstauskunft, insbesondere bei Erkrankungen, die mit kognitiven Beeinträchtigungen, Konzentrationsstörungen oder Gedächtnisverzerrungen einhergehen. Auch Daten aus dem US-Veteranensystem verdeutlichen, wie entscheidend eine adäquate Nachsorge ist. Gerade in den zwölf Wochen nach einer stationären Behandlung oder nach Beginn einer antidepressiven Therapie, Zeiträume, die als Hochrisikophasen für suizidale Krisen gelten, erhielten ältere Patient*innen signifikant weniger ambulante Kontakte und Verlaufskontrollen als jüngere, obwohl ihr Suizidrisiko nachweislich erhöht war. (Kales et al., 2010).

Das zentrale Problem besteht hier nicht lediglich in einer allgemeinen Unterversorgung, sondern in einer fehlenden risikoadaptierten Verlaufskontrolle: Personen mit objektiv erhöhtem Gefährdungspotenzial wurden nicht engmaschiger überwacht. Damit wird deutlich, dass die Frequenz und Intensität des Monitorings im ambulanten Setting nicht systematisch an individuelle Risikoprofile angepasst sind. Gerade in Phasen erhöhter Vulnerabilität kann diese Diskrepanz dazu führen, dass klinisch relevante Verschlechterungen oder suizidale Entwicklungen erst verspätet erkannt werden.

Für die ambulante Versorgung ergibt sich daraus eine zeitliche Inkongruenz zwischen Symptomdynamik und Erhebungsfrequenz. Exazerbationen entstehen nicht abrupt im Rahmen einzelner Konsultationen, sondern entwickeln sich häufig graduell über Tage oder Wochen im Kontext natürlicher Symptomfluktuationen (Price et al., 2023). Wenn Symptomveränderungen erst beim nächsten regulären Kontakt erfasst werden, erfolgt die therapeutische Anpassung zwangsläufig verzögert. Gerade in Hochrisikophasen, etwa in den ersten Wochen nach stationärer Entlassung oder nach Beginn bzw. Änderung einer antidepressiven Medikation, ist das Risiko für Rückfälle und suizidale Krisen erhöht (Kales et al., 2010; Wiegand et al., 2020), sodass eine Verzögerung der Verlaufserfassung klinisch bedeutsam sein kann.

1.4.4 BPRS

Die Brief Psychiatric Rating Scale (BPRS) ist ein etabliertes Instrument zur standardisierten Erfassung psychopathologischer Symptome, das ursprünglich für die Schweregradmessung bei psychotischen Störungen entwickelt wurde. Aufgrund ihres breiten Itemspektrums wird sie inzwischen auch in der Depressionsforschung eingesetzt, um neben den klassischen affektiven Symptomen auch Begleitphänomene wie Antriebslosigkeit, Ängstlichkeit oder psychotische Symptome systematisch zu erfassen (Park et al., 2015).

Mehrere empirische Studien konnten zeigen, dass die BPRS auch bei unipolaren Depressionen eine differenzierte Faktorstruktur abbildet. Biancosino et al. (2010) fanden in einer stationären Stichprobe vier Dimensionen – Apathie, Dysphorie, Depression und Psychotizismus – und zeigten damit, dass depressive Störungen weit über die Kernsymptomatik hinaus durch vielfältige psychopathologische Facetten geprägt sind. Auch in einer großen multizentrischen Studie in Südkorea wurde eine Vier-Faktor-Struktur nachgewiesen, bestehend aus Stimmungsstörung, positive Symptome/Apathie, Bipolarität und Denkstörung/Manierismen, die signifikant mit klinischen Variablen korrelierten (Park et al., 2015). Diese Ergebnisse belegen, dass die BPRS eine differenzierte Charakterisierung depressiver Symptomatik ermöglicht und damit ein wertvolles Instrument über rein depressionsspezifische Skalen hinaus darstellt.

Erweiterte Versionen der Skala haben die Aussagekraft weiter gesteigert. Zanello et al., (2013b) identifizierten bei ambulanten Patient*innen mit unipolarer Depression sechs Dimensionen (Stimmungsstörung, Realitätsverzerrung, Aktivierung, Apathie, Desorganisation und Somatisierung) und konnten zeigen, dass die BPRS sensibel für Symptomveränderungen im Behandlungsverlauf ist. Internationale Studien zur BPRS-Expanded (BPRS-E) bestätigen stabile vier-

bis fünfdimensionale Strukturen und untermauern die BPRS als valides und zuverlässiges Maß psychopathologischer Schweregrade über verschiedene Diagnosegruppen hinweg (Bajraktarov et al., 2023).

Für die vorliegende Untersuchung, die den Zusammenhang zwischen körperlicher Aktivität und depressiver Symptomatik analysiert, bietet die BPRS mehrere Vorteile. Die Skala erfasst ein breites Spektrum psychopathologischer Dimensionen über die reine depressive Stimmung hinaus, darunter beispielsweise Apathie, Aktivierung oder Desorganisation, die in Faktoranalysen bei unipolaren Depressionen wiederholt identifiziert wurden (Biancosino et al., 2010; Park et al., 2015; Zanello et al., 2013).

Vor diesem Hintergrund eignet sich die BPRS, um unterschiedliche Facetten depressiver Symptomatik im Zusammenhang mit Verhaltensfaktoren wie körperlicher Aktivität zu untersuchen.

1.5 Präzisionspsychiatrie

Ziel der Präzisionspsychiatrie ist es, psychiatrische Diagnostik und Behandlung stärker an individuellen biologischen, klinischen und verhaltensbezogenen Merkmalen auszurichten. Ausgangspunkt ist die ausgeprägte Heterogenität psychischer Erkrankungen wie der Depression, bei der Patient*innen sehr unterschiedliche Symptomprofile, Krankheitsverläufe und Therapieansprechen zeigen. Vor diesem Hintergrund gewinnen datengetriebene Ansätze zunehmend an Bedeutung, die darauf abzielen, große Mengen multimodaler Daten zu integrieren, um individuelle Krankheitsmuster besser zu charakterisieren und personalisierte Behandlungsstrategien zu ermöglichen (Oudin et al., 2023).

Neben klassischen klinischen und neurobiologischen Daten rücken zunehmend auch digitale Verhaltensdaten in den Fokus der psychiatrischen Forschung. Digitale Geräte wie Smartphones oder Wearables ermöglichen eine kontinuierliche und alltagsnahe Erfassung von Aktivitäts-, Mobilitäts- und physiologischen Mustern. Smartphone-Sensoren können dabei Informationen über Schlafverhalten, Bewegungsaktivität oder Umweltbedingungen liefern und ermöglichen zudem die Durchführung von Ecological Momentary Assessments (EMA) zur Echtzeiterfassung subjektiver Zustände (Torous et al., 2025). Diese Form der Datenerhebung wird häufig im Kontext der digitalen Phänotypisierung beschrieben. Darunter versteht man die kontinuierliche

Erfassung und Quantifizierung individueller Verhaltens- und Zustandsmerkmale im Alltag mithilfe persönlicher digitaler Geräte (Torous et al., 2017). Durch die kontinuierliche Erfassung solcher Verhaltensdaten entstehen umfangreiche Datensätze, die neue Einblicke in alltägliche Verhaltensmuster und deren Zusammenhang mit psychischen Symptomen ermöglichen. Erste Studien zeigen, dass bestimmte digitale Verhaltensindikatoren, wie Veränderungen in Aktivitätsniveaus, Mobilitätsmustern oder Smartphone-Nutzung, mit depressiver Symptomatik assoziiert sein können (Bufano et al., 2023; Zhan et al., 2026).

Die aus solchen Daten abgeleiteten quantifizierbaren Parameter werden zunehmend als digitale Biomarker diskutiert, die potenziell zur Charakterisierung individueller Krankheitsverläufe und zur Entwicklung prädiktiver Modelle psychischer Erkrankungen beitragen können.

1.5.1 Digitale Biomarker in der Psychiatrie

Digitale Biomarker bezeichnen quantifizierbare physiologische oder verhaltensbezogene Parameter, die mithilfe digitaler Technologien wie Smartphones, Wearables oder anderer sensorgestützter Geräte erfasst werden können. Solche Daten entstehen beispielsweise durch Bewegungssensoren, GPS-Signale, Smartphone-Nutzungsdaten oder physiologische Sensoren und liefern Informationen über Aktivitätsmuster, Mobilität, Schlaf, soziale Interaktionen oder physiologische Zustände. Durch die hochfrequente und kontinuierliche Datenerhebung ermöglichen digitale Biomarker eine objektive Abbildung von Verhaltensmustern im Alltag und stellen damit eine wichtige Ergänzung zu traditionellen klinischen Assessments dar (Sahandi Far et al., 2021).

Die Entwicklung digitaler Biomarker steht in engem Zusammenhang mit dem Konzept der digitalen Phänotypisierung, das als kontinuierliche Quantifizierung individueller Verhaltens- und Zustandsmerkmale mithilfe persönlicher digitaler Geräte beschrieben wird (Torous et al., 2017). Moderne Smartphones verfügen über eine Vielzahl integrierter Sensoren, wie Beschleunigungsmesser, Gyroskope und GPS, und können sowohl subjektive Selbstberichte, beispielsweise über Ecological Momentary Assessments (EMA), als auch objektive Sensordaten erfassen. Auf diese Weise lassen sich psychiatrisch relevante Verhaltensaspekte wie Aktivität, Mobilität oder soziale Interaktion in natürlichen Lebenskontexten untersuchen (Onnela & Rauch, 2016).

Ein zentraler Vorteil dieses Ansatzes liegt darin, dass viele Symptome psychischer Erkrankungen vor allem im Alltag auftreten und in klinischen Untersuchungssituationen nur eingeschränkt beobachtbar sind. Digitale Technologien ermöglichen eine objektivere Erfassung alltäglicher Verhaltensmuster, etwa Veränderungen der körperlichen Aktivität, sozialer Interaktionen oder psychomotorischer Dynamik, und können dazu beitragen, frühe Veränderungen des psychischen Zustands zu identifizieren sowie individualisierte diagnostische oder therapeutische Strategien zu unterstützen (Lahnakoski et al., 2022).

Empirische Studien zeigen zunehmend, dass bestimmte digitale Verhaltensindikatoren mit depressiven Symptomen assoziiert sind. Dazu zählen unter anderem eine reduzierte Standortvielfalt, eingeschränkte soziale Interaktionen oder Veränderungen sprachlicher Merkmale, die teilweise auch Rückfälle vorhersagen können (Bufano et al., 2023).

Gleichzeitig wird betont, dass Begriffe und Messansätze im Bereich digitaler Biomarker stärker standardisiert werden sollten. Der Begriff „digitaler Biomarker“ wird derzeit breit verwendet und umfasst sowohl biologische Signale wie z.B. Herzfrequenzvariabilität als auch verhaltensbasierte Daten wie App-Nutzung oder Kommunikationsmuster. Eine stärkere Orientierung am Research Domain Criteria (RDoC)-Framework könnte dazu beitragen, die biologische Plausibilität und klinische Relevanz solcher Marker zu erhöhen (Mulinari, 2024).

Im Kontext der Präzisionspsychiatrie wird zudem angenommen, dass digitale Biomarker ihr größtes Potenzial in der Kombination mit anderen neurobiologischen und klinischen Daten entfalten. Die Integration digitaler Verhaltensdaten mit Verfahren wie Neuroimaging, genetischen Analysen oder Neurotransmitterprofilen könnte dazu beitragen, stabilere Endophenotypen zu identifizieren und Subtypen depressiver Störungen differenzierter zu charakterisieren (Dukart et al., 2024).

Die körperliche Aktivität stellt einen besonders relevanten digitalen Marker in der Depressionsforschung dar, da Veränderungen im Aktivitätsniveau eng mit zentralen Symptomen der Erkrankung wie Antriebslosigkeit, psychomotorischer Verlangsamung und sozialem Rückzug verbunden sind. Mehrere Studien haben gezeigt, dass Personen mit Depression im Vergleich zu gesunden Kontrollpersonen geringere Aktivitätsniveaus und kleinere Bewegungsradien aufweisen. Diese Befunde wurden sowohl mithilfe von Wearables und Aktigraphie als auch durch smartphonebasierte Sensoren wie Beschleunigungsmesser und GPS erhoben. So berichten Price et al. (2023), dass reduzierte tägliche Aktivitätsniveaus signifikant mit der Schwere depressiver Symptome assoziiert sind. Ähnliche Ergebnisse zeigen Studien zur digitalen Phänotypisierung, in denen GPS-basierte Mobilitätsdaten genutzt wurden, um Bewegungsradien,

Aufenthaltsorte oder Standortvielfalt zu analysieren (Moshe et al., 2021; Rykov et al., 2021). Dabei zeigte sich konsistent, dass depressive Patient*innen weniger Ortswechsel aufweisen und mehr Zeit im häuslichen Umfeld verbringen. Systematische Übersichten bestätigen, dass Mobilitätsindikatoren wie Bewegungsradius, Standortvielfalt oder „homestay“ robuste digitale Marker depressiver Symptomatik darstellen können (Zierer et al., 2024).

Trotz dieser vielversprechenden Ergebnisse bestehen weiterhin methodische Herausforderungen. Viele bisherige Studien basieren auf kurzen Beobachtungszeiträumen oder Nutzen einzelne Datenquellen isoliert, wodurch intraindividuelle Veränderungen im Alltag nur eingeschränkt erfasst werden können (Moshe et al., 2021; Zierer et al., 2024). Zudem unterscheiden sich die verwendeten Messmethoden und Auswertungsverfahren teilweise erheblich, was die Vergleichbarkeit der Ergebnisse erschwert. Vor diesem Hintergrund untersucht die vorliegende Studie körperliche Aktivität und Mobilitätsmuster mithilfe kontinuierlicher smartphonebasierter Datenerhebung über einen längeren Zeitraum. Ziel ist es, die Dynamik alltäglicher Aktivitätsmuster bei Personen mit depressiver Symptomatik genauer zu charakterisieren und damit einen Beitrag zur Weiterentwicklung digitaler Verhaltensmarker in der Depressionsforschung zu leisten.

Insgesamt stellen digitale Biomarker eine vielversprechende Ergänzung zu etablierten klinischen und biologischen Markern dar. Durch ihre alltagsnahe und skalierbare Datenerhebung eröffnen sie neue Möglichkeiten für die Früherkennung, Verlaufsbeobachtung und individualisierte Behandlung psychischer Erkrankungen und könnten damit einen wichtigen Beitrag zur Entwicklung einer datenbasierten, personalisierten Psychiatrie leisten (McGinnis et al., 2024).

1.5.2 EMA

Ecological Momentary Assessment (EMA) ist eine wiederholte, alltagsnahe Erhebung von Symptomen, Affekten und Verhaltensweisen im natürlichen Umfeld und hat sich in den letzten Jahren als wichtige Methode etabliert. Dadurch werden Einschränkungen klassischer retrospektiver Verfahren reduziert, die häufig durch Recall Bias bzw. Gedächtnisverzerrungen gekennzeichnet sind. Die Erfassung der Entwicklung depressiver Symptomatik, ihrer situativen Abhängigkeit sowie intraindividueller Schwankungen ist mithilfe von EMA besonders präzise möglich. (Armey et al., 2015).

In den EMA-Erhebungen berichten depressive Personen im Vergleich zu gesunden Kontrollgruppen konsistent von weniger positiven und mehr negativen Affekten; gleichzeitig zeigen die

hochfrequenten Messungen im Alltag, dass diese Zustände über den Tagesverlauf deutlich variieren und von situativen Faktoren beeinflusst werden können (Colombo et al., 2019; Wichers et al., 2010).

Darüber hinaus ermöglicht EMA die Untersuchung kurzfristiger Veränderungen emotionaler Zustände, die mit suizidalen Gedanken oder selbstverletzendem Verhalten in Zusammenhang stehen können. Studien zeigen, dass insbesondere negative Affekte, Stressreaktionen und affektive Instabilität mit suizidalen Gedanken assoziiert sind und sich durch wiederholte EMA-Messungen im Alltag detailliert untersuchen lassen (Armey et al., 2015). Durch die wiederholte Erfassung dieser Zustände im Alltag können solche kurzfristigen Dynamiken sichtbar werden, die in klassischen retrospektiven Erhebungen häufig nicht erfasst werden.

Außerdem zeigen die EMA-basierten Selbstberichte eine enge Korrelation mit etablierten klinischen Skalen wie der Hamilton Depression Rating Scale (HDRS), was die externe Validität des Verfahrens unterstützt (Maatoug et al., 2021; Targum et al., 2021).

EMA ermöglicht zudem die detaillierte Abbildung individueller Symptomverläufe im Alltag und macht zeitliche Muster depressiver Symptome sichtbar. Erste Studien zeigen, dass smartphonebasierte EMA-Verfahren von Patient*innen mit Depression gut akzeptiert werden und valide Informationen über den Symptomverlauf liefern (Maatoug et al., 2021). Die Visualisierung solcher Daten, beispielsweise über Dashboards, wird als potenziell hilfreicher Ansatz diskutiert, um Selbstbeobachtung, Mustererkennung und das Verständnis individueller Symptomfluktuationen zu unterstützen. Die klinische Evidenz zur Wirksamkeit solcher Feedbackmechanismen ist bislang jedoch noch begrenzt (Polhemus et al., 2022).

Es ist besonders bemerkenswert, dass EMA möglicherweise einen Vorteil in der Früherkennung hat. Eine umfassende Studie hat ergeben, dass man postpartale Depressionen schon in den ersten Wochen nach der Geburt mit hoher Genauigkeit vorhersagen kann, wenn man grundlegende klinische Daten, die EMA-basierte Symptomdynamik und die Ergebnisse von Fragebögen kombiniert (Hahn et al., 2021).

Durch die weit verbreitete Nutzung von Smartphones hat die EMA einen erheblichen Innovationsschub erlebt. Die smartphone-basierte EMA vereint aktive Selbstauskünfte mit einer passiven Datenerhebung über die integrierte Sensorik (z. B. Bewegungsprofile, Kommunikationsmuster, Schlaf-Wach-Rhythmus). Mit dieser Kombination wird die Datendichte gesteigert und gleichzeitig die ökologische Validität verbessert (Porras-Segovia et al., 2020). Die Ergebnisse von Machbarkeitsstudien belegen eine hohe Teilnahmebereitschaft und gute Compliance über

mehrere Wochen bei depressiven Patient*innen, vor allem, wenn die Anwendungen benutzerfreundlich sind (Maatoug et al., 2021; Porras-Segovia et al., 2020).

Ein weiterer vielversprechender Ansatz kombiniert EMA-Daten mit Verfahren des maschinellen Lernens, um individuelle Muster depressiver Symptomverläufe zu modellieren. Dabei werden wiederholte EMA-Selbstberichte gemeinsam mit passiv erfassten Sensordaten (z. B. zu körperlicher Aktivität, Mobilität oder Smartphone-Nutzung) analysiert. Modelle, die mehrere Datenquellen integrieren, können depressive Symptome oder kurzfristige Befindensveränderungen mit moderater Genauigkeit vorhersagen. Während Modelle auf rein passiven Sensordaten Vorhersagewerte von etwa AUC ≈ 0,60–0,70 erreichen, erzielen multimodale Modelle, die aktive EMA-Einträge mit passiven Daten kombinieren, häufig AUC ≈ 0,70–0,80 (Kathan et al., 2022).

1.5.3 Passive Datenerfassung

Als ergänzender Ansatz hat sich in den letzten Jahren das passive Monitoring etabliert. Dabei werden Verhaltens- und physiologische Daten automatisch über Sensoren digitaler Geräte wie Smartphones, Wearables oder Umweltsensoren erfasst, ohne dass aktive Eingaben der Nutzenden erforderlich sind. Dieser Ansatz wird häufig im Kontext der digitalen Phänotypisierung beschrieben, bei der digitale Technologien genutzt werden, um Verhaltensmuster im Alltag kontinuierlich zu quantifizieren und mit Aspekten der psychischen Gesundheit in Beziehung zu setzen. Hierbei werden Rohdaten aus verschiedenen Sensoren, etwa Beschleunigungssensoren, GPS, Mikrofonen oder physiologischen Sensoren, erfasst und anschließend zu Verhaltensindikatoren wie Aktivitätsniveau, Bewegungsmustern oder sozialen Interaktionen verarbeitet. Smartphones und Wearables ermöglichen damit die kontinuierliche Erhebung einer Vielzahl von Messgrößen, die potenziell mit psychischen Symptomen in Zusammenhang stehen, darunter körperliche Aktivität, Schlafverhalten, Mobilität, Kommunikationsmuster sowie physiologische Signale wie Herzfrequenz oder Hautleitfähigkeit. Diese sensorbasierten Daten können anschließend analysiert werden, um Verhaltensmarker zu identifizieren, die mit Veränderungen der psychischen Gesundheit assoziiert sind (Sheikh et al., 2021).

Empirische Studien zeigen zunehmend, dass passiv erfasste Verhaltensindikatoren mit depressiver Symptomatik assoziiert sein können. Beispielsweise können über GPS-Sensoren Mobilitätsparameter wie zurückgelegte Distanz, Aufenthaltsorte oder die Zeit, die zu Hause verbracht

wird, erfasst werden. Veränderungen solcher Mobilitätsmuster wurden wiederholt mit depressiven Symptomen in Verbindung gebracht und werden daher als potenzielle digitale Marker für Veränderungen des psychischen Zustands diskutiert (Terhorst et al., 2024).

Ein weiterer wichtiger Verhaltensparameter ist die körperliche Aktivität, da Veränderungen im Bewegungsverhalten zu den zentralen Merkmalen affektiver Störungen zählen und objektiv beispielsweise durch Aktigraphie erfasst werden können. Eine systematische Übersichtsarbeit und Meta-Analyse mit 34 Studien und insgesamt 1804 Patient*innen mit affektiven Störungen zeigte, dass Betroffene signifikant geringere Aktivitätsniveaus aufwiesen als gesunde Kontrollpersonen (SMD = −0,78). (Wüthrich et al., 2022).

1.5.4 JTrack

Eine Plattform, die passives Monitoring und aktive Selbstberichte integriert, ist das Open-Source-System JTrack. Es wurde als modulares Framework für die mobile Datenerhebung entwickelt und ermöglicht die kontinuierliche Erfassung verschiedener smartphonebasierter Verhaltensdaten, darunter Bewegungsaktivität, Standortveränderungen, Kommunikationsverhalten oder Smartphone-Nutzung.

Die Datenerhebung erfolgt automatisch im Hintergrund, ohne dass eine aktive Eingabe durch die Teilnehmenden erforderlich ist. Dadurch können alltagsnahe Verhaltensmuster über längere Zeiträume hinweg erfasst werden. Ergänzend lassen sich über die Plattform Ecological Momentary Assessments (EMA) implementieren, um subjektive Zustände wie Stimmung oder Stress in Echtzeit zu erfassen.

Durch die Kombination passiver Sensordaten und aktiver Selbstberichte ermöglicht JTrack eine multimodale Erfassung von Verhalten und Erleben im Alltag. Die modulare Architektur der Plattform erlaubt zudem die Integration verschiedener Sensor- und Datenerfassungsmodule und kann flexibel an unterschiedliche Studiendesigns angepasst werden. Ein weiterer zentraler Aspekt der Plattform ist die Berücksichtigung von Datenschutz und Datensicherheit, etwa durch kontrollierte Datenspeicherung, transparente Einwilligungsprozesse und sichere Datenübertragung. Dadurch bietet JTrack eine technische Infrastruktur zur Untersuchung digitaler Biomarker psychischer Erkrankungen und ermöglicht es, Veränderungen in Aktivität, Mobilität oder sozialen Interaktionen kontinuierlich zu erfassen und mit subjektiven Symptomangaben zu verknüpfen (Sahandi Far et al., 2021)

1.5.5 Forschungslücke und Beitrag der vorliegenden Arbeit

Die smartphone-basierte Erfassung körperlicher Aktivität bei depressiven Störungen ist in den letzten Jahren zunehmend untersucht worden (Bufano et al., 2023; Moshe et al., 2021; Zierer et al., 2024). Trotz dieser wachsenden Evidenzbasis bleiben zentrale Limitationen bestehen: Viele Studien rekrutieren über Online-Plattformen mit selbstberichteten Diagnosen und überwiegend nicht-klinischen oder mild symptomatischen Stichproben, die Beobachtungszeiträume sind häufig kurz, und die berichteten Aktivitäts-Symptom-Zusammenhänge sind unter strengeren statistischen Kriterien nur begrenzt replizierbar (Leaning et al., 2024; Moshe et al., 2021; Zierer et al., 2024).

Vor diesem Hintergrund leistet die vorliegende Arbeit in mehreren Aspekten einen Beitrag zur Forschungslage. Grundlage ist eine klinisch diagnostizierte Stichprobe aus dem stationären, teilstationären und ambulanten Versorgungssetting des LVR-Klinikums Düsseldorf, deren Diagnosen ärztlich nach ICD-10 im Rahmen der klinischen Routine gestellt wurden. Die Beobachtung erstreckt sich über 90 Tage in einem naturalistischen, nicht-interventionellen Design. Passive Aktivitätsdaten werden nicht auf eine einzelne Aktivitätskennziffer reduziert, sondern in unterschiedliche Bewegungsformen wie Gehen, Laufen, Fahrradfahren oder Inaktivitätszeiten aufgeschlüsselt. Diese Aktivitätsmaße werden mit zwei komplementären Symptomerhebungen verknüpft: einer ärztlichen Fremdbeurteilung zu Studienbeginn (BPRS) sowie wiederholten Selbstberichten im Wochenrhythmus (EMA-Items). Darüber hinaus werden die methodischen Schritte zur Weiterentwicklung der Aktivitätsmaße in den regulatorischen Rahmen zukünftiger digitaler Biomarker eingeordnet (FDA, EMA, MDR, EU-AI-Act).

1.6 Ziele und Hypothesen

Die vorliegende Arbeit untersucht die Zusammenhänge zwischen passiv über Smartphones erfasster Alltagsaktivität und psychopathologischer Symptomatik bei erwachsenem Patient*innen mit depressiver Störung. Analysiert werden zum einen die Querschnittszusammenhänge zwischen Aktivitätsmaßen und einer ärztlichen Fremdbeurteilung zu Studienbeginn (BPRS), zum anderen die längsschnittlichen Zusammenhänge zwischen Aktivitätsmaßen und wöchentlich erhobenen Selbstberichten (EMA-Items) über einen Zeitraum von 90 Tagen.

Auf Basis der bisherigen Evidenz zu reduzierter körperlicher Aktivität bei depressiven Störungen (Wüthrich et al., 2022) sowie der berichteten Assoziationen zwischen Aktivitätsmaßen und Symptomschwere (Price et al., 2023; Rykov et al., 2021) werden im Rahmen einer explorativen Datenanalyse zwei Hypothesen formuliert:

Hypothese 1: Höhere passiv erfasste Aktivitätsmaße (Gesamtaktivität, Fahrradfahren, Laufen, Gehen) sind mit geringerer selbstberichteter Symptomschwere (EMA) assoziiert; höhere Inaktivitätszeit (STILL_min) ist mit höherer Symptomschwere assoziiert. Adjustiert wird für Alter, Geschlecht und Studientag (zweiseitig, $\alpha = 0{,}05$).

Hypothese 2: Höhere passiv erfasste Aktivitätsmaße (Gesamtaktivität, Fahrradfahren, Laufen, Gehen) sind mit niedrigeren BPRS-Werten (Gesamtscore und Subskalen) assoziiert; höhere Inaktivitätszeit (STILL_min) ist mit höheren BPRS-Werten assoziiert. Adjustiert wird für Alter und Geschlecht (zweiseitig, $\alpha = 0{,}05$).

Kategorie	Zielvariable	Einheit / Aggregation
Gesamtaktivität	Total Activity (Summe: WALKING_min + RUNNING_min + On_BICYCLE)	Minuten/Tag
Fahrradfahren	On_BICYCLE	Minuten/Tag
Gehen	WALKING_min	Minuten/Tag
Laufen	RUNNING_min	Minuten/Tag
Inaktivitätszeit	STILL_min	Minuten/Tag
Psychopathologie	EMA-Items (12 Fragen): Stimmung, Freude, Energie, Schlaf, Konzentration, Stress, Angst, somat. Beschwerden etc.	Likert-Skalenpunkte (wöchentlich)
Psychopathologie	BPRS: Gesamtscore + Subskalen (THOT, ACTV, ANER, HOST) – Baseline	Skalenpunkte (Baseline)
Kovariaten	Geschlecht (kategorial), Alter, Studientag	—

Tabelle 1: Zielvariablen der Studie.

2 Material und Methoden

2.1 Studiendesign

Die Studie verwendete ein nicht-therapeutisches, nicht-invasives, longitudinales Beobachtungsdesign, um den Zusammenhang zwischen passiv erfasster körperlicher Alltagsaktivität, selbstberichteten Symptomen und klinischer Fremdbeurteilung bei Erwachsenen mit depressiver Störung zu untersuchen.

Das ursprüngliche Studiendesign sah zwei Kohorten vor: erwachsene Patient*innen mit einer Diagnose einer depressiven Störung (F32/F33 nach ICD-10) sowie erwachsene Patient*innen mit einer Diagnose einer Schizophrenie (F20). Geplant war die Rekrutierung von jeweils bis zu 100 Teilnehmenden pro Kohorte. Aufgrund einer deutlich erschwerten Rekrutierung in der Schizophrenie-Kohorte (vgl. Abschnitt 2.2.2) konnten dort nur 13 Personen eingeschlossen werden; die Hauptanalysen dieser Arbeit beschränken sich daher auf die Depressionskohorte (n = 50).

Die körperliche Aktivität der Teilnehmenden wurde über einen Zeitraum von 90 Tagen passiv über ein auf dem persönlichen Smartphone installiertes Erhebungssystem (JTrack Social) dokumentiert. Die Klassifikation der Bewegungsformen (Gehen, Laufen, Fahrradfahren, Stillstand) erfolgte über die integrierten Beschleunigungs- und Lagesensoren des Smartphones unter Verwendung der Google Activity Recognition API. Eine inhaltliche Auswertung von GPS- oder Standortdaten war nicht Gegenstand der vorliegenden Arbeit (vgl. Abschnitt 2.3.2). Parallel zur passiven Aktivitätserfassung erfolgte eine wöchentliche, aktive Selbstauskunft mittels Ecological Momentary Assessment (JTrack EMA) sowie zu Studienbeginn eine einmalige ärztliche Fremdbeurteilung mit der Brief Psychiatric Rating Scale (BPRS).

Der Beobachtungszeitraum von 90 Tagen wurde gewählt, um sowohl kurzfristige als auch längerfristige Veränderungen im Aktivitätsverhalten abzubilden und alltagsbedingte Variationen (z. B. Wochenstruktur, individuelle Tagespläne) statistisch berücksichtigen zu können.

Diese Studie ist in ein umfassenderes Forschungsprojekt am LVR-Klinikum Düsseldorf eingebettet, das sich der digitalen Erhebung und Auswertung von Alltagsverhalten bei Menschen mit psychiatrischen Erkrankungen widmet. Für die Datensammlung wurden die Anwendungen JTrack Social und JTrack EMA eingesetzt (https://jutrack.inm7.de/login/?next=/) (https://play.google.com/store/apps/details?id=inm7.JTrack.JTrack_EMA_Plus)

(https://play.google.com/store/apps/details?id=inm7.JTrack.JTrack_Social). Diese Apps wurden von der Arbeitsgruppe von Prof. Dukart am FZ Jülich dafür entwickelt, um Sensordaten der Teilnehmenden kontinuierlich und unter Einhaltung hoher Datenschutzstandards zu erfassen. Zusätzlich ermöglichen sie eine regelmäßige, wöchentliche Erhebung des psychischen Befindens.

2.2 Studiendurchführung

2.2.1 Ein- und Ausschlusskriterien

Im Rahmen der Studiendurchführung wurden klare Ein- und Ausschlusskriterien definiert, um eine homogene und wissenschaftlich aussagekräftige Stichprobe zu gewährleisten.

Für die Teilnahme an der Studie mussten die Interessierten volljährig sein (≥ 18 Jahre), in Deutschland leben und in der Lage sein, eine informierte Einwilligung zu geben. Da die für die Datenerhebung eingesetzten Anwendungen JTrack Social und JTrack EMA ausschließlich für das Betriebssystem Android verfügbar waren, stellte der Besitz und die Nutzung eines entsprechenden Smartphones eine weitere Voraussetzung für die Studienteilnahme dar. Vor Beginn der Datenerhebung wurde von allen Teilnehmenden eine schriftliche Einwilligung eingeholt. Der Studienablauf war so gestaltet, dass die Teilnahme für die Patient*innen möglichst unkompliziert und alltagsnah erfolgen konnte. Nach einem einmaligen persönlichen Termin wurden die Teilnehmenden zunächst ausführlich über Ziele, Ablauf und datenschutzrechtliche Aspekte der Studie informiert. Im Anschluss erfolgte die schriftliche Einwilligung zur Studienteilnahme.

Nach Erteilung der Einwilligung wurden die Ein- und Ausschlusskriterien überprüft und demografische sowie klinische Basisdaten erhoben. Anschließend unterstützte das Studienteam die Teilnehmenden bei der Installation und Einrichtung der Anwendungen JTrack Social und JTrack EMA auf ihrem Android-Smartphone.

Im Rahmen dieses Termins wurden zudem die wichtigsten Funktionen der Apps erläutert, insbesondere welche Daten erhoben werden und wie die Datenerfassung jederzeit gestoppt werden kann. Für die korrekte Funktion der Anwendungen mussten bestimmte Berechtigungen (z. B. Standort, Aktivitätserkennung und Benachrichtigungen) erteilt werden. Die Teilnehmenden wurden gebeten, ihr Smartphone möglichst regelmäßig bei sich zu tragen, um eine kontinuierliche Datenerhebung zu gewährleisten.

Die Rekrutierung der Patient*innen erfolgte im stationären und teilstationären Bereich des LVR-Klinikums Düsseldorf, insbesondere auf den allgemeinpsychiatrischen Stationen und in der Tagesklinik unseres Zentrums für Schema-Therapie. Eingeschlossen wurden ausschließlich Personen mit einer ärztlich diagnostizierten depressiven Störung oder Schizophrenie nach ICD-10 (F32–F33 bzw. F20). Die Diagnosestellung erfolgte durch behandelnde Fachärztinnen und Fachärzte im Rahmen der klinischen Routineversorgung und wurde anhand der klinischen Dokumentation überprüft.

Demografische Angaben sowie Informationen zu möglichen psychiatrischen Komorbiditäten wurden im Rahmen der Basisbefragung erhoben. Komorbiditäten führten grundsätzlich nicht zum Ausschluss von der Studie, sofern sie die Studienteilnahme nicht wesentlich beeinträchtigten.

Um sicherzustellen, dass die Teilnehmenden während des gesamten Studienzeitraums zur Teilnahme fähig waren, wurden Personen ausgeschlossen, bei denen eine fehlende Einwilligungsfähigkeit oder eine schwere kognitive Beeinträchtigung vorlag. Eine solche Beeinträchtigung wurde angenommen, wenn nach klinischer Einschätzung der behandelnden Ärztinnen und Ärzte eine ausreichende Verständlichkeit der Studieninformationen oder eine zuverlässige Teilnahme an den Befragungen nicht gewährleistet war. Dies betraf insbesondere Personen mit ausgeprägten kognitiven Defiziten, etwa im Rahmen einer demenziellen Erkrankung oder einer mittel- bis schwergradigen Intelligenzminderung.

2.2.2 Rekrutierung

Die Rekrutierung der Teilnehmenden mit depressiven Störungen oder Schizophrenie erfolgte über die geschützten und offenen Stationen sowie über die Tageskliniken und Ambulanzen des LVR-Klinikums Düsseldorf. Ärztinnen und Ärzte sowie therapeutisches Personal der jeweiligen Einrichtungen wurden über die Studie informiert und gebeten, potenziell geeignete Patient*innen auf die Teilnahme hinzuweisen. Zusätzlich wurde die Studie über interne Informationskanäle des Klinikums bekannt gemacht. Ziel war es, eine möglichst breite Stichprobe von Patient*innen mit einer Diagnose einer depressiven Störung zu gewinnen. Potenzielle Teilnehmende wurden entweder durch behandelnde Ärztinnen und Ärzte oder direkt durch das Studienteam angesprochen und über Ziele, Ablauf und datenschutzrechtliche Aspekte der Studie aufgeklärt.

Die Rekrutierung für die Studie begann im November 2021 und endete im März 2023. Das initiale Rekrutierungsziel war 100 Probanden mit der Diagnose einer Schizophrenie und 100 Probanden mit der Diagnose einer Depression. Es wurden insgesamt 63 Probanden, 50 mit Depression und 13 mit Schizophrenie rekrutiert. Das Rekrutierungsziel von 200 Zusagen für die beiden Kohorte wurde leider nicht erreicht. Rund ein Viertel der Kontaktaufnahmen wurden mit einer Absage beantwortet. Gründe für die Ablehnung waren, wenn angegeben, dass das Smartphone nicht genutzt werde, eine generelle technische Überforderung vorliege oder Datenschutzbedenken bestehen würden. Insbesondere die Schizophrenie Patient*innen, die gefragt wurden, hatten in mehreren Fällen kein Smartphone und litten an wahnhaften Symptomen im Sinne von Beobachtungswahn und Ängste, die die Teilnahme an unserer Studie deutlich erschwerten. Der erste Studieneinschluss erfolgte am 19.11.2021 und der letzte am 28.03.2023, sodass die gesamte Rekrutierung nach 17 Monate abgeschlossen war. Aufgrund einer individuellen Studiendauer von 90 Tage war der letzte Tag, an dem Studiendaten aufgezeichnet wurden, der 12.06.2023. Der gesamte Zeitraum der Datenaufzeichnung betrug demnach 571 Tage (ca. 19 Monate).

Die ursprüngliche Studienplanung sah eine Stichprobengröße von bis zu 200 Teilnehmenden vor. Eine a-priori-Poweranalyse mit G*Power ergab unter Annahme einer mittleren Effektgröße, eines Signifikanzniveaus von $\alpha = 0{,}05$ und einer Teststärke von 0,95 eine erforderliche Mindeststichprobe von 147 Teilnehmenden für eine zweiseitige ANCOVA. Diese Berechnung bezog sich auf den ursprünglich geplanten Zwei-Gruppen-Vergleich zwischen der Depressions- und der Schizophrenie-Kohorte. Da die Schizophrenie-Kohorte aufgrund unzureichender Rekrutierung nicht in die Hauptanalyse einging, wurde das Auswertungsverfahren an die innerhalb der Depressionsgruppe vorliegende longitudinale Datenstruktur angepasst: lineare gemischte Modelle für Hypothese 1, lineare Regression für Hypothese 2. Eine eigene Poweranalyse für diese veränderten Modelle wurde nicht durchgeführt; die Konsequenzen für die statistische Aussagekraft werden in Abschnitt 4.3 diskutiert.

Aufgrund der geringen Fallzahl von Teilnehmenden mit Schizophrenie konzentrieren sich die Hauptanalysen auf die Gruppe der Teilnehmenden mit depressiver Störung. Eine separate statistische Auswertung der Schizophrenie-Gruppe erfolgte daher nicht. Um die wiederholten Messungen innerhalb der Personen sowie potenzielle Störfaktoren angemessen zu berücksichtigen, wurden gemischte lineare Modelle (mixed-effects models) eingesetzt. Diese Modelle erlauben die Analyse von longitudinalen Daten mit mehrfachen Messzeitpunkten pro Person und sind robust gegenüber ungleichen Gruppengrößen und fehlenden Messwerten.

Bereits zu Studienbeginn wurden demografische und klinische Basisdaten erhoben, um die Stichprobe zu charakterisieren und potenzielle Einflussfaktoren in den statistischen Analysen berücksichtigen zu können. Dazu gehörten Alter, Geschlecht, klinische Diagnose, Medikation sowie klinische Schweregradeinschätzungen, die mithilfe validierter Instrumente wie der Brief Psychiatric Rating Scale (BPRS) erfasst wurden.

Die Konzeption und Datenerhebung des Forschungsprojekts erfolgten in enger Zusammenarbeit zwischen dem Doktoranden und den betreuenden Wissenschaftlerinnen und Wissenschaftlern am LVR-Klinikum Düsseldorf. Die Rekrutierung der Teilnehmenden sowie die Durchführung der Datenerhebung wurden überwiegend durch den Doktoranden koordiniert und umgesetzt.

Für die Verwaltung und Aufbereitung der erhobenen Daten wurde Microsoft Excel verwendet. Die statistischen Analysen wurden mit den Open-Source-Programmen Jamovi und R Studio durchgeführt. Zur Analyse der longitudinalen Daten kamen lineare Regressionsmodelle und gemischte lineare Modelle zum Einsatz. Die grafische Darstellung der Ergebnisse erfolgte unter Verwendung wissenschaftlicher Visualisierungsstandards.

2.2.3 Ethik und Einwilligungsprozess

Für die Studie lag ein positives Ethikvotum der zuständigen Ethikkommission der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf vor (Ethikvotum Nr. Studien-Nr.: 2021-1326). Zudem wurde eine Genehmigung zur Durchführung der Studie am 27.09.2021 durch das LVR-Klinikum Düsseldorf erteilt.

Die Teilnahme an der Studie erfolgte freiwillig. Alle Teilnehmenden wurden vor Studienbeginn umfassend über Ziele, Ablauf und Datenschutzaspekte der Studie informiert und gaben eine schriftliche informierte Einwilligung zur Teilnahme sowie zur Verarbeitung der erhobenen Daten. Die Teilnehmenden wurden darüber aufgeklärt, dass sie ihre Teilnahme jederzeit ohne Angabe von Gründen beenden können.

2.2.4 Datenschutz und Datenverarbeitung

Die erhobenen Daten wurden ausschließlich in pseudonymisierter Form gespeichert. Persönliche Identifikationsmerkmale wie Name, Kontaktdaten oder Adresse wurden nach Studienende

gelöscht, um das Risiko eines Missbrauchs zu minimieren. Die Speicherung erfolgte ausschließlich in zugangsbeschränkten, gesicherten Datenbanken. Eine Weitergabe der Daten an andere Forschende ist nur in vollständig anonymisierter Form und ausschließlich zu wissenschaftlichen Zwecken vorgesehen, sodass eine Rückverfolgung auf Einzelpersonen ausgeschlossen ist.

Die Datenerhebung umfasste keine sensiblen Inhalte wie Telefonnummern, Namen von Absendern oder Empfängern oder Kommunikationsinhalte. Erfasst wurden lediglich pseudonymisierte Informationen zu Nutzungsdauer, Zeitpunkten der App-Nutzung und Aktivitätskategorien, jedoch keine inhaltlichen oder identifizierenden Daten.

Die Verarbeitung und Speicherung der Daten erfolgte in Übereinstimmung mit der Europäischen Datenschutz-Grundverordnung (DSGVO). Die Teilnehmenden wurden über ihr Recht informiert, jederzeit Auskunft über ihre gespeicherten Daten zu erhalten, diese zu berichtigen oder löschen zu lassen und ihre Einwilligung zur Datenverarbeitung ohne Nachteile zu widerrufen. Die maximale Aufbewahrungsdauer der Daten beträgt zehn Jahre nach Studienende.

2.3 Datenerhebung

2.3.1 Demografische und klinische Basisdaten

Am Tag der Rekrutierung wurde ein strukturierter Prozess implementiert, um für jede teilnehmende Person systematisch zentrale demografische und klinische Informationen zu erfassen. Im Rahmen dieser Doktorarbeit wurde die BPRS gezielt eingesetzt, um den Schweregrad der Psychopathologie der Studienteilnehmer zum Zeitpunkt der Rekrutierung objektiv zu erfassen. Die Erhebung des BPRS-Scores erfolgte einmalig bei Studienbeginn und diente ausschließlich dazu, den Ausgangszustand der Probanden hinsichtlich ihrer psychopathologischen Belastung zu bestimmen. Eine Verlaufsbeurteilung oder die Analyse von Therapieeffekten anhand der BPRS war nicht Bestandteil des Studiendesigns.

2.3.2 Smartphone-App: „JTrack Social"

Im Rahmen der vorliegenden Studie kam die Smartphone-Anwendung „JTrack Social“ zum Einsatz, die speziell für die wissenschaftliche Erfassung von Alltagsverhalten entwickelt wurde. JTrack Social sammelt passiv eine breite Palette von Datenströmen, einschließlich GPS-

Standort, App-Nutzung und Bewegungsmuster - und unterstützt gleichzeitig die Integration aktiver ökologischer Momentaufnahmen (EMAs). Dies ermöglicht ein umfassenderes und kontinuierliches Bild des täglichen Lebens und des psychischen Gesundheitszustands eines Nutzers. JTrack Social ist so konzipiert, dass es an verschiedene Forschungskontexte angepasst werden kann, was es besonders wertvoll für Längsschnittstudien und für die Erfassung der vielschichtigen Natur depressiver Symptome in realen Umgebungen macht (Vignapiano et al., 2023).

Die App wurde am Forschungszentrum Jülich von der Arbeitsgruppe „Digitale Biomarker“ unter der Leitung von Prof. Dr. Jürgen Dukart entwickelt und steht als Open-Source-Software kostenfrei zur Verfügung. Forschende erhalten über das „JTrack Dashboard“ Zugriff auf zentrale Verwaltungsfunktionen. Während des Aufnahmegesprächs wurde für jede teilnehmende Person ein individueller QR-Code generiert, der mit der Smartphone-Kamera eingescannt werden musste, um die Datenerhebung zu starten. Das Dashboard ermöglichte es dem Forschungsteam, den Startzeitpunkt, die letzte Datenübertragung sowie das geplante Ende der Aufzeichnung jederzeit online einzusehen. Zudem konnte die Datenerhebung bei Bedarf, etwa im Falle eines Widerrufs der Einwilligung, über das Dashboard aus der Ferne beendet werden.

Im Studienprotokoll des Projekts „JuTrack Social“ wurde präzise festgelegt, welche Datenarten durch die App im Rahmen der vorliegenden Studie erfasst werden sollen. Die Datenerhebung umfasste folgende Bereiche der Smartphone-Nutzung:

App-Nutzung

- Erfasst: Name und tägliche Nutzungsdauer (Minuten) jeder im Vordergrund aktiven App
- Keine Speicherung von Inhalten, Spracheingaben oder Kameraaufnahmen
- Verwendung: Für die Forschungsfrage nicht relevant

Aktivitätserkennung

- Automatische Erkennung und Zeitmessung (Minuten) für: Stehen, Gehen, Laufen, Fahrradfahren, Telefon in der Hand halten, Autofahren
- Nutzung integrierter Sensoren
- Verwendung: Überprüfung der ersten und zweiten Hypothese

Bewegungserkennung

- Kontinuierliche Aufzeichnung relativer Positionsveränderungen (Meter) via GPS
- Kein exakter Standort gespeichert
- Verwendung: für die Forschungsfrage nicht relevant

Die erhobenen Daten wurden zunächst lokal auf dem Smartphone der Teilnehmenden zwischengespeichert. Sobald eine WLAN-Verbindung bestand, erfolgte eine verschlüsselte Übertragung der Daten an einen Server des Forschungszentrums Jülich. Nach erfolgreichem Transfer wurden die Forschungsdaten automatisch vom Gerät gelöscht, um Speicherplatz freizugeben. Eine Übertragung über das mobile Datennetz war nur dann möglich, wenn dies von den Teilnehmenden explizit in der App aktiviert wurde.

Zur Qualitätssicherung überprüfte der Doktorand mindestens einmal pro Monat im so genannten „JTrack Dashboard“, wann die letzte Datenübertragung eines Teilnehmenden erfolgte. Bei ausbleibender Übertragung über einen Zeitraum von vier Wochen wurde versucht, Kontakt mit den Teilnehmenden aufzunehmen, um mögliche Ursachen wie fehlerhafte Batterieeinstellungen oder eine fehlende WLAN-Verbindung zu klären. Falls ein vorzeitiger Studienabbruch gewünscht wurde, konnte die Datenerhebung über das Dashboard remote beendet werden.

2.3.3 Smartphone-App: „JTrack EMA"

Im Rahmen der vorliegenden Studie wurde die JTrack EMA App (Ecological Momentary Assessment) als zentrales Instrument zur Erhebung von Verhaltens- und Selbstauskunftsdaten eingesetzt. Die App wurde speziell für die kontinuierliche und alltagsnahe Erfassung psychischer und verhaltensbezogener Parameter bei Patient*innen mit affektiven und psychotischen Störungen entwickelt. Ziel war es, sowohl passive als auch aktive Daten zu erfassen, um Zusammenhänge zwischen Alltagsverhalten und Symptomen aus dem affektiven und psychotischen Spektrum im Alltag detailliert zu untersuchen. Die EMA-Abfragen fanden wöchentlich statt. Die Teilnehmer erhielten jede Woche am gleichen Wochentag eine Benachrichtigung auf ihrem Smartphone, die sie an das Ausfüllen des EMA-Fragebogens erinnerte. Diese regelmäßige Er-

innerung stellte sicher, dass die Erhebung der Selbstauskunftsdaten strukturiert und vergleichbar erfolgte. Die JTrack-EMA-App generierte die Benachrichtigungen automatisch und unterstützte die Einhaltung der Erhebungsintervalle sowie die Vollständigkeit der Datensätze. Die Teilnehmenden beantworteten dabei standardisierte Fragen zu Stimmung, Angst, Konzentration, Stress, Schlafqualität, mentaler Energie und weiteren alltagsrelevanten Aspekten. Insgesamt umfasste der EMA-Fragebogen 13 Items zur Erfassung aktueller Befindlichkeit und psychischer Symptome im Alltag.

1. Wie würden Sie Ihr gestriges allgemeines Befinden beschreiben?
2. Haben Sie sich gestern gestresst gefühlt?
3. Wie war Ihre Stimmung gestern?
4. Wie haben Sie heute Nacht geschlafen?
5. Wie würden Sie Ihre gestrige körperliche und geistige Energie beschreiben?
6. Hatten Sie gestern Angst?
7. Wie war gestern Ihre Konzentration?
8. Konnten Sie gestern Freude empfinden?
9. Hatten Sie gestern das Gefühl, dass Ihre Gedanken oder Gefühle von außen beeinflusst oder von außen wahrgenommen werden können?
10. Hatten Sie gestern das Gefühl, bedroht oder verfolgt zu werden?
11. Haben Sie gestern Dinge gehört oder gesehen, die Andere nicht hören oder sehen konnten?
12. Hatten Sie gestern körperliche Beschwerden (z. B. Kopfschmerzen, Bauchschmerzen, Brustkorbschmerzen, Kribbeln in den Armen oder Beinen, Husten, etc.)?
13. Haben Sie gestern Ihre Medikamente eingenommen?

2.3.4 Pseudonymisierung

Um eine eindeutige Zuordnung der erhobenen Daten zu gewährleisten, erhielt jede teilnehmende Person beim Eintritt in die Studie eine individuelle Studien-ID. Diese zufällig generierte vierstellige Kennziffer wurde in einem passwortgeschützten Dokument auf einem sicheren Server des LVR-Klinikums Düsseldorf mit dem jeweiligen Klarnamen verknüpft. Während der gesamten Studiendauer wurden sämtliche erhobenen Daten ausschließlich unter dieser Studien-ID gespeichert, sodass sie pseudonymisiert vorlagen.

Die Verknüpfung mit dem Klarnamen war erforderlich, um nach Abschluss der Datenerhebung die Ergebnisse der Fragebögen mit den demografischen Angaben und den über das Smartphone erfassten Daten zusammenführen zu können. Zudem ermöglichte diese Vorgehensweise eine gezielte Kontaktaufnahme mit den Teilnehmenden, beispielsweise bei technischen Problemen während der viermonatigen Datenerhebung.

Die ausgefüllten Papierfragebögen sowie die unterschriebenen Einwilligungserklärungen wurden im LVR-Klinikum Düsseldorf unter Angabe des Klarnamens sicher archiviert. Der Zuordnungsschlüssel zwischen Studien-ID und Klarnamen wurde spätestens einen Monat nach Abschluss der Datenerhebung gelöscht. Bis zu diesem Zeitpunkt hatten die Teilnehmenden die Möglichkeit, die vollständige Löschung ihrer Studiendaten zu beantragen. Nach Löschung des Schlüssels waren sämtliche Studiendaten vollständig anonymisiert, sodass eine Rückverfolgung zu einzelnen Personen nicht mehr möglich ist.

2.4 Datenanalyse

2.4.1 Datenaufbereitung

Die von den Smartphone-App „JTrack Social“ und „JTrack EMA“ erfassten Rohdaten wurden automatisch an das Forschungszentrum Jülich übermittelt. Aus diesen Daten mussten zunächst die für die Studie relevanten Zielvariablen und insbesondere die Aktivitätsdaten berechnet werden. Im Rahmen der Datenaufbereitung erfolgte die Aktivitätserkennung durch die JTrack Social App mittels eines komplexen Algorithmus, der auf den integrierten Sensoren der Smartphones der Teilnehmer basierte. Die App nutzte eine Kombination aus Beschleunigungssensor, Gyroskop und GPS-Daten, um verschiedene Bewegungsmuster zu identifizieren und zu klassifizieren. Der Beschleunigungsmesser und das Gyroskop sind entscheidend für die Erfassung der Bewegungsdynamik, während das GPS-Geolokalisierungsinformationen liefert. Die JTrack

Social App verwendete die Google Activity Recognition API (Application Programming Interface), die speziell für die präzise Erkennung verschiedener Bewegungsarten entwickelt wurde und bis zu sechs Arten von Aktivitäten, wie Gehen, Laufen und Radfahren, mit einer bestimmten Häufigkeit der Datenerfassung erkennen kann. Die Aktivitätserkennung erfolgte kontinuierlich im Hintergrund und wurde in Echtzeit verarbeitet, wobei die Daten in vordefinierten Intervallen aggregiert wurden (Sahandi Far et al., 2021).

Diese API nutzt maschinelle Lernalgorithmen, die auf umfangreichen Trainingsdaten basieren, um Bewegungsmuster zu erkennen und zu kategorisieren. Die Aktivitätserkennung erfolgte energieeffizient, um die Akkulaufzeit der Geräte zu schonen und gleichzeitig eine kontinuierliche Datenerfassung zu gewährleisten. Neben der Aktivitätsdaten wurde auch tägliche App-Nutzung ebenso berechnet. Es erfolgte eine strukturierte Aufbereitung der Rohdaten, das heißt, die thematische Einteilung der verwendeten Apps in Kategorien, um die individuellen Nutzungsprofile der Teilnehmenden vergleichbar zu machen und eine gruppenspezifische Auswertung für zukünftige Forschungsfrage zu ermöglichen.

Basierend auf der zuvor beschriebenen Kategorisierung wurden aus den Rohdaten der JTrack Social App für jede Teilnehmerin und jeden Teilnehmer die Minuten pro Tag in den jeweiligen Aktivitätskategorien Inaktivitätszeit (STILL_min), Gehen (WALKING_min), Laufen (RUNNING_min), Fahrradfahren (ON_BICYCLE_min) berechnet. Der verwendete Programmiercode verarbeitete die vom Google Play Activity Recognition Service erkannten Aktivitäten und aggregierte sie zu täglichen Summen. Die resultierenden Daten bildeten die Grundlage für die weiteren statistischen Analysen, wobei jede Zeile im Datensatz einem Teilnehmer-Tag mit den entsprechenden Aktivitätszeiten in Minuten entsprach.

Auch bei den Aktivitätsdaten gab es Qualitätskriterien: Nach Angaben der App-Entwickler sollten pro Tag mindestens zehn Aktivitätsdatenpunkte aufgezeichnet werden – selbst wenn das Smartphone kaum genutzt wurde, da dann beispielsweise für „Gehen“ eine Dauer von 0 Minuten dokumentiert wird. Wurden an einem Tag weniger als zehn Aktivitätsdatenpunkte erfasst, wurden für alle Aktivitätskategorien sowie bei „Datenpunkte_Aktivität“ fehlende Werte eingetragen. Die Zeitspannen, in denen das Smartphone die Aktivitäten „Gehen“, „Laufen“ und „Fahrradfahren“ erkannt hat, wurden für jeden Tag und jede teilnehmende Person aufsummiert und als Zielvariable „Gesamtaktivität (Total_Activity)“ in Minuten festgehalten.

Die wöchentlichen Ecological Momentary Assessment (EMA) Daten wurden einer strukturierten Aufbereitungsprozedur unterzogen. Die EMA-Erhebung umfasste 13 standardisierte Fragen, die von den Teilnehmerinnen und Teilnehmern einmal wöchentlich über die Studiendauer

von drei Monaten beantwortet wurden. Die Rohdaten wurden zunächst auf Vollständigkeit überprüft, wobei fehlende Werte entsprechend gekennzeichnet wurden. Die Antworten wurden je nach Fragetyp in geeignete numerische Formate transformiert, um statistische Analysen zu ermöglichen. Bei Likert-skalierten Fragen wurden die Antworten in entsprechende numerische Werte konvertiert, während die Textantwort bei der dreizehnten Frage mittels qualitativer Inhaltsanalyse kategorisiert wurde. Die aufbereiteten EMA-Daten wurden mit den passiven Aktivitätsdaten der JTrack Social App zeitlich synchronisiert, um Zusammenhänge zwischen subjektiv berichteten Erfahrungen und objektiv gemessenen Verhaltensmustern untersuchen zu können. Dieser Prozess ermöglichte eine integrierte Analyse der psychischen Befindlichkeit und des Aktivitätsverhaltens im Wochenverlauf.

Um die Qualität und Vollständigkeit der Datenerhebung durch die App „JTrack Social“ zu dokumentieren, wurde analysiert, wie viele Datenpunkte täglich aufgezeichnet wurden. Dabei erfolgte eine getrennte Erfassung der Datenpunkte für die verschiedene Aktivitäten. Ein Datenpunkt könnte beispielsweise sein, dass eine Teilnehmerin am fünften Studientag von 14:32 bis 14:43 Uhr die Aktivität „Walking“ nachgegangen ist.

Nachdem sämtliche Studiendaten in einer einheitlichen Tabelle zusammengeführt worden waren, wurden abschließende Anpassungen vorgenommen, um die Daten optimal für die statistische Analyse vorzubereiten. Dazu gehörte die präzise Festlegung des Skalenniveaus für jede Variable (vgl. Tabelle 1), was für die Auswahl geeigneter statistischer Verfahren essenziell war. Anschließend erfolgte eine systematische Prüfung der Datenvollständigkeit sowie der Plausibilität aller erfassten Werte.

Bei der detaillierten Sichtung der Forschungsdaten wurden vereinzelt Werte identifiziert, die sich nicht durch normale tägliche Aktivitätsmuster erklären ließen. So wurden beispielsweise an einzelnen Tagen Inaktivitätszeit-Werte von über 1200 Minuten (mehr als 20 Stunden) registriert – ein Wert, der physiologisch unplausibel erscheint und auf eine fehlerhafte Messung, etwa durch Systemupdates, Hintergrundprozesse oder temporäre Sensorausfälle, hindeutet. Für solche Extremwerte wurden spezifische Filter angewendet, die die Variablen auf realistische Zeitspannen begrenzten. Diese Filterkriterien wurden auf Basis physiologischer Plausibilitätsannahmen definiert.

Konkret wurden folgende Plausibilitätsbereiche definiert und angewendet:

Inaktivitätszeit: 240-1230 Minuten pro Tag

Gehen: 0-300 Minuten pro Tag

Fahrradfahren: 0-150 Minuten pro Tag

Laufen: 0-100 Minuten pro Tag

Die Werte außerhalb dieser Bereiche wurden als fehlende Werte gekennzeichnet und in den nachfolgenden Analysen entsprechend behandelt. Zusätzlich wurde sichergestellt, dass die Summe aller Aktivitätszeiten eines Tages den logischen Maximalwert von 1440 Minuten (24 Stunden) nicht überschritt. Diese Qualitätssicherungsmaßnahmen gewährleisteten die Validität der Daten für die nachfolgenden statistischen Analysen.

Vollständige Daten für einen Studientag lagen somit nur dann vor, wenn nach diesen Korrekturen weder bei den App-Nutzungsdaten noch bei den Aktivitätsdaten fehlende Werte vermerkt waren.

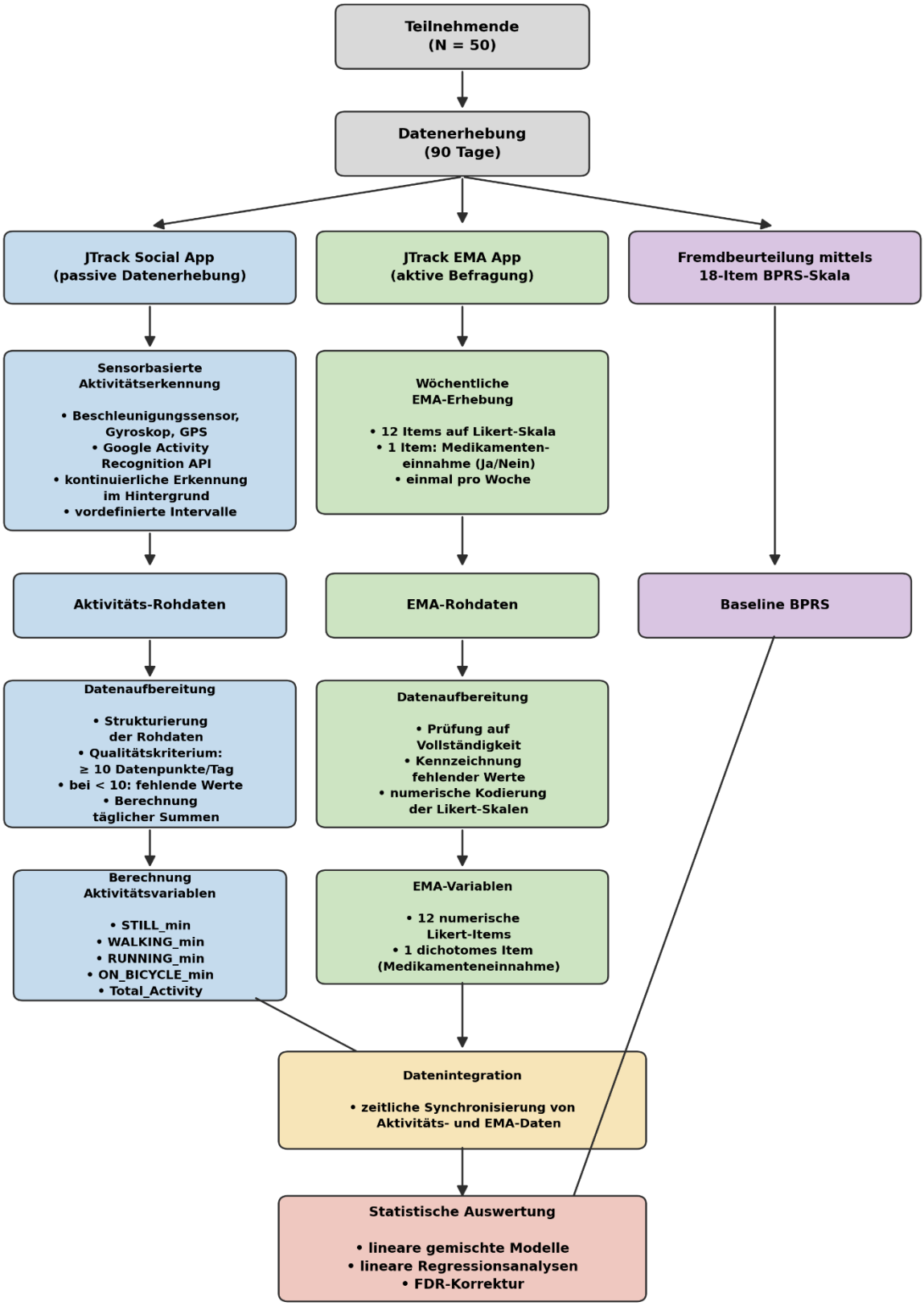


Abbildung1: Schematische Darstellung des Studienablaufs mit den drei parallelen Datenquellen (passive Aktivitätserfassung über JTrack Social, wöchentliche EMA-Selbstauskünfte über JuTrack EMA, klinische Fremdbeurteilung mittels 18-Item BPRS), deren Datenaufbereitung und -integration sowie der anschließenden statistischen Auswertung.

Variable	Berechnung / Beschreibung	Einheit	Skalenniveau
ID	Zufällige, fünfstellige Studien-ID	ID	Nominal
Kohorte	Diagnosegruppe: Depression / Schizophrenie	–	Nominal
Hauptdiagnose	Primäre psychiatrische Diagnose (ICD-Code)	ICD-Code	Nominal
Psychiatrische_ Komorbidität	Nebendiagnosen (ICD-Codes)	ICD-Codes	Nominal
Bildungsniveau	Höchster Abschluss (z. B. Hauptschule, Realschule, Abitur, Studium)	Kategorie	Ordinal
Geschlecht	Selbstangabe: w/m/d	–	Nominal
Alter	Alter bei Rekrutierung	Jahre	Kontinuierlich
Indexjahr	Jahr des Erkrankungsbeginns	Jahr (JJJJ)	Nominal
Tag	Studientag: 1-90		Kontinuierlich
Psych_Aufenthalte	Stationäre Behandlungen in der Vorgeschichte	Anzahl	Kontinuierlich
Erwerbstätigkeit	Vollzeit/Teilzeit/arbeitslos/in Ausbildung/Studium/berentet/arbeitsunfähig	Kategorie	Nominal
Familienstand	Ledig/verheiratet/getrennt/in Partnerschaft	Kategorie	Nominal
Tage_seit_Aufnahme	Tage seit stationärer/teilstationärer Aufnahme	Tage	Kontinuierlich
CGI-S-MV	Schweregrad (Modifizierte Version)	Punkte (1–7)	Ordinal
GAF	Globales Funktionsniveau	Punkte (0–100)	Kontinuierlich
Rekrutierungssetting	Geschützt / offen / Tagesklinik / Ambulant	Kategorie	Nominal

Variable	Berechnung / Beschreibung	Einheit	Skalenniveau
BPRS_Score	Summenwert (z. B. BPRS-18)	Punkte	Kontinuierlich
ANDP	Subskala (5-Faktoren-Modell)	Punkte	Kontinuierlich
THOT	Subskala (5-Faktoren-Modell)	Punkte	Kontinuierlich
ANER	Subskala (5-Faktoren-Modell)	Punkte	Kontinuierlich
HOST	Subskala (5-Faktoren-Modell)	Punkte	Kontinuierlich
ACTV	Subskala (5-Faktoren-Modell)	Punkte	Kontinuierlich
EMA_Q1-Q12	Wöchentliche EMA-Frage (Likert)	Punkte (1-10)	Ordinal
EMA_Q13_Kategorie	Kodierte Textantwort (qualitative Inhaltsanalyse)	Kategorie	Nominal
Datenpunkte_Aktivität	Anzahl Aktivitäts-Events pro Tag	Datenpunkte/ Tag	Kontinuierlich
Inaktivitätszeit	Aktivitätsminuten pro Tag	Min/Tag	Kontinuierlich
Stehen	Aktivitätsminuten pro Tag	Min/Tag	Kontinuierlich
Gehen	Aktivitätsminuten pro Tag	Min/Tag	Kontinuierlich
Laufen	Aktivitätsminuten pro Tag	Min/Tag	Kontinuierlich
Fahrradfahren	Aktivitätsminuten pro Tag	Min/Tag	Kontinuierlich
Tilting	Aktivitätsminuten pro Tag	Min/Tag	Kontinuierlich
Gesamtaktivität	Summe aus Gehen, Laufen, Fahrradfahren	Min/Tag	Kontinuierlich

Tabelle 2: Eine Erläuterung aller erhobenen Variablen. Alle fett geschriebenen Variablen entsprechen Zielvariablen. In der zweiten Spalte ist die Beschreibung und/oder die Berechnung der entsprechenden Variable angegeben.

2.4.2 Auswertung der demographischen Daten

Für die nähere Charakterisierung der Kohorte wurden die Häufigkeitsverteilungen der relevanten demografischen Variablen Geschlecht (Sex) und Alter (Age) erfasst und ausgewertet.

Zur klinischen Charakterisierung der Stichprobe wurden die Skalenwerte der BPRS (Brief Psychiatric Rating Scale), CGI (Clinical Global Impression) und GAF (Global Assessment of Functioning) am Tag der Rekrutierung erhoben und deskriptiv ausgewertet.

2.4.3 Analyse der Smartphonedaten

Nach Abschluss der vorbereitenden Schritte wurden die erhobenen Smartphonedaten ausgewertet.

Die deskriptive Analyse der Aktivität umfasste die grafische Darstellung sowohl der gesamten Aktivität als auch der verschiedenen Aktivitätsformen. Mithilfe von Boxplots wurde veranschaulicht, wie sich die durchschnittlichen Aktivitäten verteilen. Für die Zielvariablen der Aktivität wurden Mittelwert, Median, Standardabweichung, Minimum und Maximum berechnet (vgl. Abb 6).

Zur Prüfung der ersten Hypothese untersuchten wir den Zusammenhang zwischen Aktivitätsmustern und dem selbstberichteten psychischen Zustand mithilfe linearer gemischter Modelle (LMM). Dieses Verfahren ist für das vorliegende Design geeignet, da pro Person wiederholte Messungen vorliegen, die Anzahl der Beobachtungen je Person variiert und einzelne Werte fehlen; LMMs erlauben die gemeinsame Schätzung fester und zufälliger Effekte (Fahrmeir et al., 2013). Als Zielvariablen dienten die Aktivitätsmaße (z. B. Bewegungsdauer, Inaktivitätszeit, Gesamtaktivität), die EMA-Items gingen als Hauptprädiktoren in die Modelle ein. Geschlecht wurde als kategoriale Variable berücksichtigt, Alter und Tag als kontinuierliche Kovariaten; zusätzlich erfassten wir Items-Unterschiede über eine Indikatorvariable für Question 1–12. Interindividuelle Unterschiede modellierten wir über einen zufälligen Interzept auf Personenebene (ID).

Die allgemeine Modellformel lautete:

$$\text{Zielvariable} \sim 1 + \text{Geschlecht} + \text{Tag} + \text{Alter} + \text{Question} + (1 \mid \text{ID})$$

Die gemischten linearen Modelle wurden mittels Restricted Maximum Likelihood (REML) geschätzt; zur Kontrolle des multiplen Testens wurden die p-Werte mittels False-Discovery-Rate-Korrektur nach Benjamini–Hochberg (FDR) angepasst. Für jeden Prädiktor berichten wir die Schätzung (β), den Standardfehler (SE), das 95 %-Konfidenzintervall, den t-Wert, den p-Wert sowie den FDR-korrigierten p-Wert.

Die statistischen Analysen erfolgten mit Jamovi (Version 2.3.28) (https://www.jamovi.org/)

Die zweite Hypothese war die Beurteilung, ob höhere klinische Symptomschwere mit geringerer körperlicher Aktivität (Gesamtaktivität, Gehen, Laufen, Fahrradfahren) und mit höherer Inaktivitätszeit assoziiert ist. In der vorliegenden Studie wurde untersucht, wie sich psychiatrische Symptome auf das durchschnittliche Aktivitätsniveau (Activity_mean_time) von Patient*innen auswirken. Hierzu wurden die Gesamt-Scores der Brief Psychiatric Rating Scale (BPRS) sowie deren Subskalen als Prädiktoren in die Analyse aufgenommen. Die abhängige Variable war die durchschnittliche tägliche Aktivitätszeit, die mithilfe von Smartphone-Sensordaten objektiv erfasst wurde. Für die statistische Auswertung wurde eine lineare Regressionsanalyse durchgeführt

Zur Untersuchung des Zusammenhangs zwischen der täglichen Aktivitätszeit und der psychopathologischen Symptomschwere (gemessen mit der Brief Psychiatric Rating Scale, BPRS) wurden zwei Modelle getrennt geschätzt. Die BPRS wurde einmalig zu Studienbeginn erhoben; für die Analyse wurde die über den 90-tägigen Beobachtungszeitraum erfasste Aktivität pro Teilnehmer*in zu einem Mittelwert aggregiert. Die lineare Regression prüft somit, ob die zu Studieneinschluss beobachtete Symptomschwere mit dem mittleren Aktivitätsniveau im nachfolgenden Beobachtungszeitraum assoziiert ist. In Modell A wurde der BPRS-Gesamtscore als Prädiktor berücksichtigt. In Modell B wurden stattdessen die BPRS-Subskalen gemäß der verwendeten Faktorstruktur einbezogen (THOT = Thought Disturbance, ANDP = Anxiety/Depression, ACTV = Activation, ANER = Anergia, HOST = Hostility). Der Gesamtscore und die Subskalen wurden nicht gemeinsam in einem Modell berücksichtigt, um Multikollinearität zu vermeiden. Alter und Geschlecht wurden als potenzielle Störfaktoren kontrolliert.

Für jeden Prädiktor wurden der β-Koeffizient, der Standardfehler, das 95-%-Konfidenzintervall, der p-Wert sowie der FDR-korrigierte p-Wert nach Benjamini–Hochberg berichtet.

Die allgemeine Modellformel lautete:

Zielvariable (Activity_mean_time) ~1+Geschlecht + Age+ BPRS

3 Ergebnisse

3.1 Rekrutierung, Studienzeitraum und Dropouts

Von den 50 Teilnehmenden mit Depressionsdiagnose erreichten 21 Personen (42 %) die gesamte Studiendauer von 90 Tagen und schlossen damit formal die Studie ab. Weitere 5 Personen (10 %) verblieben bis mindestens Tag 75, schieden jedoch kurz vor dem Studienende aus. 12 Teilnehmende (24 %) beendeten ihre Teilnahme zwischen Tag 50 und 74, 6 Teilnehmende (12 %) nahmen zwischen 25 und 49 Tagen teil, und weitere 6 Teilnehmende (12 %) brachen bereits innerhalb der ersten 24 Tage ab. (vgl. Tabelle 3) Damit verblieb insgesamt gut die Hälfte der Stichprobe mindestens 75 Tage in der Studie.

Kategorie	Anzahl	Prozent
90 Tage	21	42 %
75–89 Tage	5	10 %
50–74 Tage	12	24 %
25–49 Tage	6	12 %
0–24 Tage	6	12 %

Tabelle 3: Der Tabelle ist zu entnehmen, wie viele Teilnehmende die angestrebte Studiendauer von 90 Tagen erreicht bzw. nicht erreicht haben. n=50.

Betrachtet man nicht die Verweildauer, sondern die tatsächliche Zahl aktiver Tage mit Dateneingabe, zeigt sich ein abweichendes Bild: 15 Teilnehmende (30 %) übermittelten an allen 90 Studientagen Daten. Weitere 5 Personen (10 %) waren an 75 bis 89 Tagen aktiv, 12 Teilnehmende (24 %) lieferten Daten an 50 bis 74 Tagen, 6 Personen (12 %) an 25 bis 49 Tagen. Ein weiteres Viertel der Gruppe (12 Personen, 24 %) übermittelte Daten an weniger als 25 Tagen. (vgl. Tabelle 4) Eine Mindestteilnahmedauer war im Studiendesign nicht vorgesehen.

Kategorie	Anzahl	Prozent
90 Tage	15	30 %
75-89 Tage	5	10 %
50-74 Tage	12	24 %
25-49 Tage	6	12 %
0-24 Tage	12	24 %

Tabelle 4: Die Tabelle gibt die Datencompliance wieder. 30 % der Teilnehmenden übermittelten an allen 90 Tagen Daten, während etwa die Hälfte der Stichprobe weniger als 50 aktive Tage aufwies.n=50

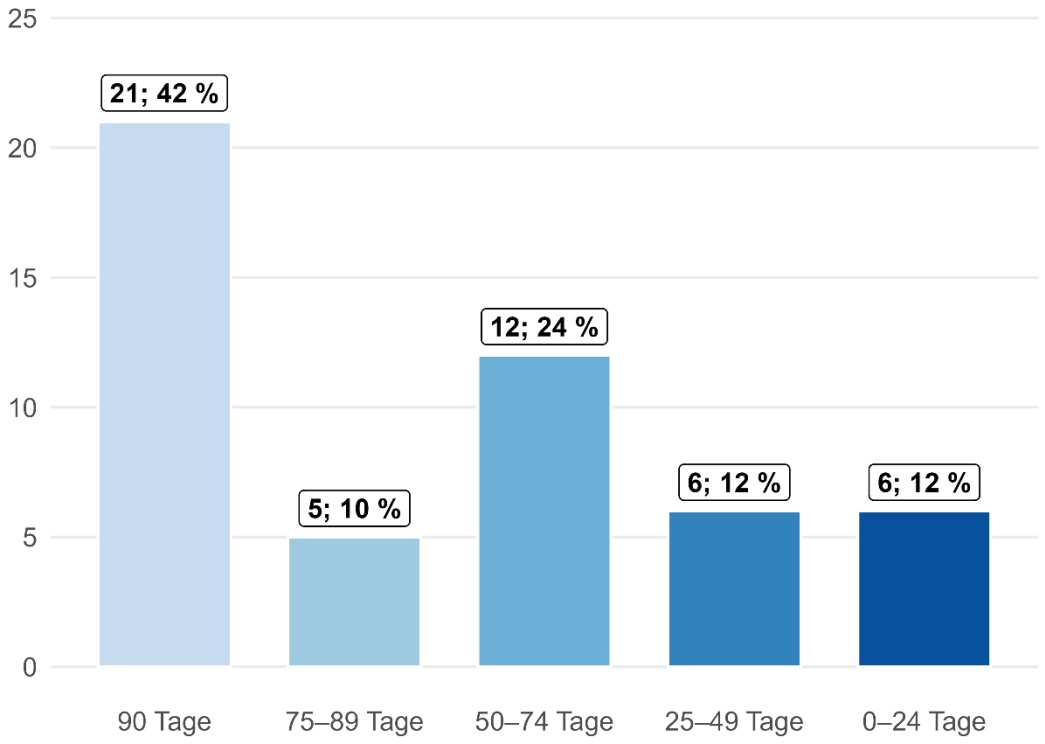


Abbildung 2: Verteilung der Verweildauer der Studienteilnehmenden (N=50)

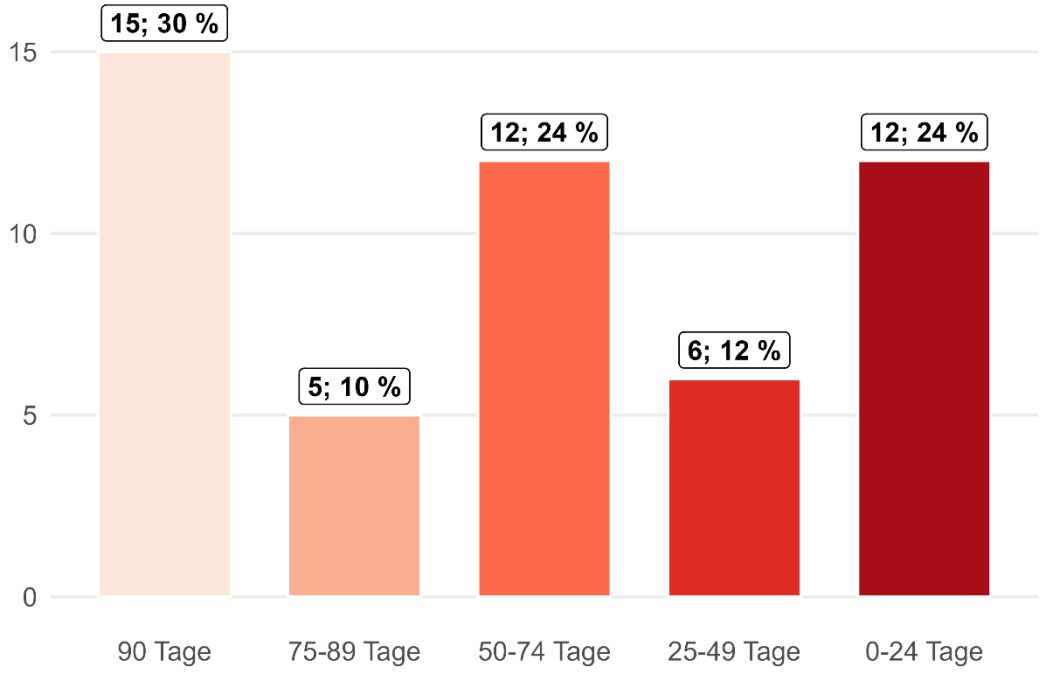


Abbildung 3: Verteilung der aktiven Studientage (Datendichte).

3.2 Ergebnisse der demographischen und psychometrischen Daten

3.2.1 Verteilung von Alter und Geschlecht

Am Rekrutierungstag wurden die demographischen und psychometrischen Daten allen Patient*innen erhoben. Die demographischen Daten umfassten die Hauptdiagnose, die Nebendiagnosen, den Bildungsgrad, die familiäre Situation, die berufliche Situation, das Alter, das Geschlecht, das Index Jahr der Erkrankung sowie die Anzahl der stationären Aufenthalte.

Die Stichprobe umfasste 50 Teilnehmende (31 weiblich, 19 männlich). Das Geschlecht „Divers“ wurde von niemandem angegeben. Das mittlere Alter war in beiden Gruppen nahezu gleich (Männer: 39,3 Jahre, Frauen: 38,8 Jahre; SD ≈ 14). Der Median lag bei 40 Jahren (Männer) bzw. 34 Jahren (Frauen); die Spannweite betrug 19–61 bzw. 18–63 Jahre. (vgl. Tabelle 5)

	Männlich	Weiblich
N	19	31
Fehlende Werte	0	0
Mittelwert	39.3	38.8
Median	40.0	34.0
Standardabweichung	14.2	14.3
Minimum	19.0	18.0
Maximum	61.0	63.0

Tabelle 5: Alter der Teilnehmenden nach Geschlecht (in Jahren); deskriptive Kennwerte.

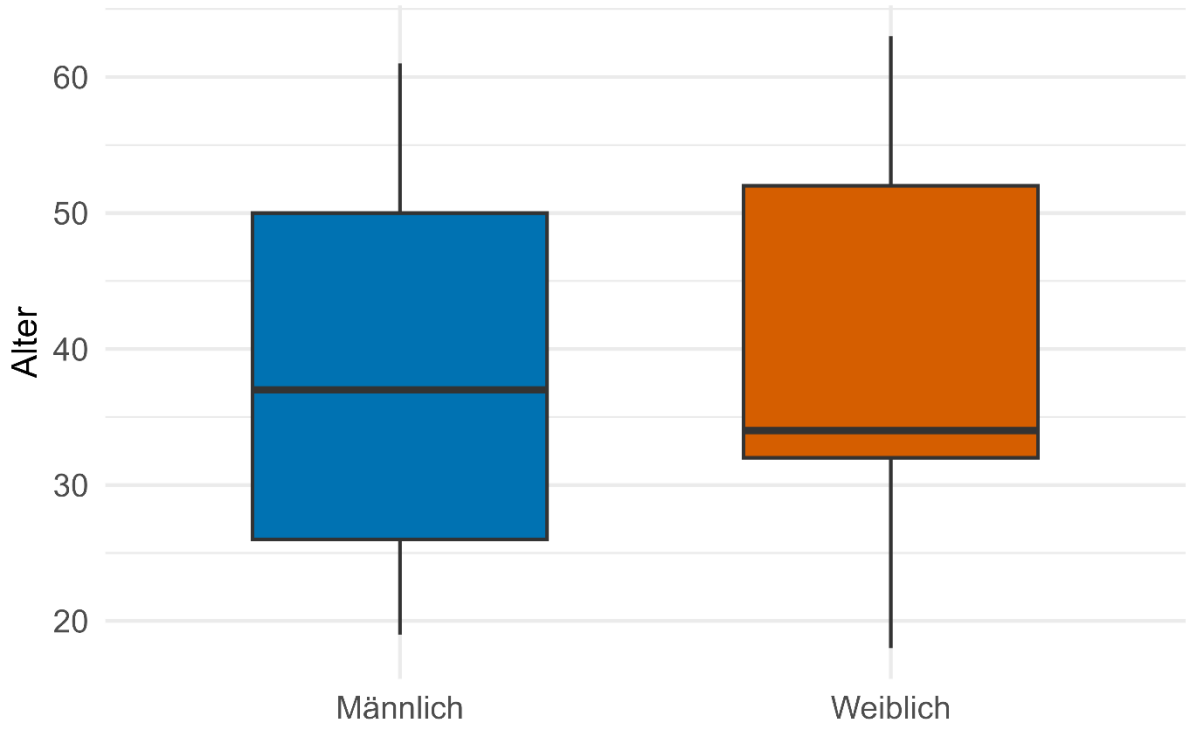


Abbildung 4: Alter nach Geschlecht in Jahren. (Median = Linie, Box = IQR, Whisker = 1,5×IQR)

3.2.2 EMA

Die folgende Tabelle gibt einen Überblick über die deskriptiven Statistiken der zwölf wöchentlich erhobenen EMA-Items. Für jede Frage ist die Anzahl der Antworten (N), die Anzahl der fehlenden Werte, der Mittelwert (Mean), der Median, die Standardabweichung (SD), das Minimum und das Maximum aufgeführt. (vgl. Tabelle 6)

Alle Items hatten einen Wertebereich von 1 bis 10.

Die Anzahl gültiger Antworten pro Frage variierte zwischen 386 (Frage 10) und 392 (Frage 3); pro Frage fehlten 149–153 Angaben (ca. 27–28 % der Gesamtfälle).

Die Mittelwerte lagen für die Items 1–8 (Stimmung, Stress, Befinden, Schlaf, Energie, soziale Interaktion, Konzentration, Freude) im mittleren Skalenbereich zwischen M = 5,03 und M = 5,57. Die größte Streuung wies Item 12 (körperliche Beschwerden; SD = 3,06) auf, was auf eine ausgeprägte interindividuelle Variation in diesem Bereich hinweist.

Die niedrigsten Mittelwerte zeigten sich bei den Items 9 (M = 2,13), 10 (M = 1,50) und 11 (M = 1,42). Diese Items erfassen Beeinflussungserleben, Bedrohungserleben und halluzinatorisches Erleben — Phänomene, die für die ursprünglich vorgesehene Schizophrenie-Kohorte konzipiert waren und in der vorliegenden Depressionsstichprobe entsprechend selten berichtet wurden.

Die Mediane folgen demselben Muster: meist 5,00–6,00; für die Items 9–11 konstant 1,00.

Die Standardabweichungen reichten von SD = 1,39 (Item 11) bis SD = 3,06 (Item 12), wobei die Items 9–11 auch die geringste Streuung aufwiesen.

Frage	Item	N	Fehlende Werte	Mittelwert	Median	SD	Minimum	Maximum
1	Allgemeines Befinden	391	150	5.30	5.0	2.29	1.0	9.0
2	Stress	390	149	5.57	6.0	2.62	1.0	10.0
3	Stimmung	392	149	5.21	5.0	2.36	1.0	9.0
4	Schlafqualität	389	151	5.34	5.0	2.31	1.0	10.0
5	Energie (körperlich/geistig)	390	150	5.03	5.0	2.15	1.0	9.0
6	Angst	389	150	3.78	3.0	2.69	1.0	10.0
7	Konzentration	387	153	5.30	5.0	2.18	1.0	9.0
8	Freude	388	152	5.35	5.0	2.41	1.0	9.0
9	Beeinflussungserleben	390	151	2.13	1.0	2.12	1.0	10.0
10	Bedrohungs-/Verfolgungserleben	386	151	1.50	1.0	1.42	1.0	10.0
11	Halluzinationen	391	149	1.42	1.0	1.39	1.0	10.0
12	Körperliche Beschwerden	391	149	5.10	5.0	3.06	1.0	10.0

Tabelle 6: Deskriptive Kennwerte der 12 EMAs (Skala 1–10): Anzahl beobachteter Antworten (N), fehlende Werte, Mittelwert, Median, Standardabweichung (SD), Minimum und Maximum. N variiert je Item aufgrund fehlender Messungen (≈ 386–392).

3.2.3 BPRS

Tabelle 7 zeigt die deskriptiven Statistiken des BPRS-Gesamtscores und der fünf Subskalen für die Stichprobe von 50 Teilnehmenden. Fehlende Werte traten nicht auf.

Der BPRS-Gesamtscore lag im Mittel bei M = 32,7 (SD = 6,49; Median = 32; Range 21–47). Die individuellen Werte verteilten sich über einen Bereich von 26 Punkten, was die ausgeprägte interindividuelle Variabilität der psychopathologischen Belastung in der Stichprobe widerspiegelt.

Auf Subskalenebene wies Anxiety/Depression (ANDP) den höchsten Mittelwert und zugleich die größte Streuung auf (M = 12,7; SD = 4,27; Median = 12; Range 6–22). Es folgten Anergia (ANER; M = 6,82; SD = 2,61; Median = 6; Range 4–13) und Activation (ACTV; M = 5,86; SD = 1,69; Median = 5; Range 3–11).

Für die Subskalen Thought Disturbance (THOT; M = 4,10; SD = 0,46; Median = 4; Range 4–7) und Hostility (HOST; M = 3,22; SD = 0,62; Median = 3; Range 3–6) lag die Mehrheit der Teilnehmenden am Skalenminimum. Die Standardabweichungen dieser beiden Subskalen waren entsprechend deutlich geringer als die der übrigen Subskalen.

Eine grafische Darstellung der Verteilungen findet sich in Abbildung 5.

Variable	N	Missing	Mittelwert	Median	Standardabweichung	Minimum	Maximum
BPRSscore	50	0	32.7	32	6.49	21	47
ANDP	50	0	12.7	12	4.27	6	22
ANER	50	0	6.82	6	2.61	4	13
THOT	50	0	4.10	4	0.463	4	7
ACTV	50	0	5.86	5	1.69	3	11
HOST	50	0	3.22	3	0.616	3	6

Tabelle 7: BPRS-Gesamtscore und Subskalen (ANDP, ANER, THOT, ACTV, HOST): deskriptive Kennwerte (n = 50). Angegeben sind N, fehlende Werte, Mittelwert, Median, Standardabweichung, Minimum und Maximum; es traten keine fehlende Werte auf.

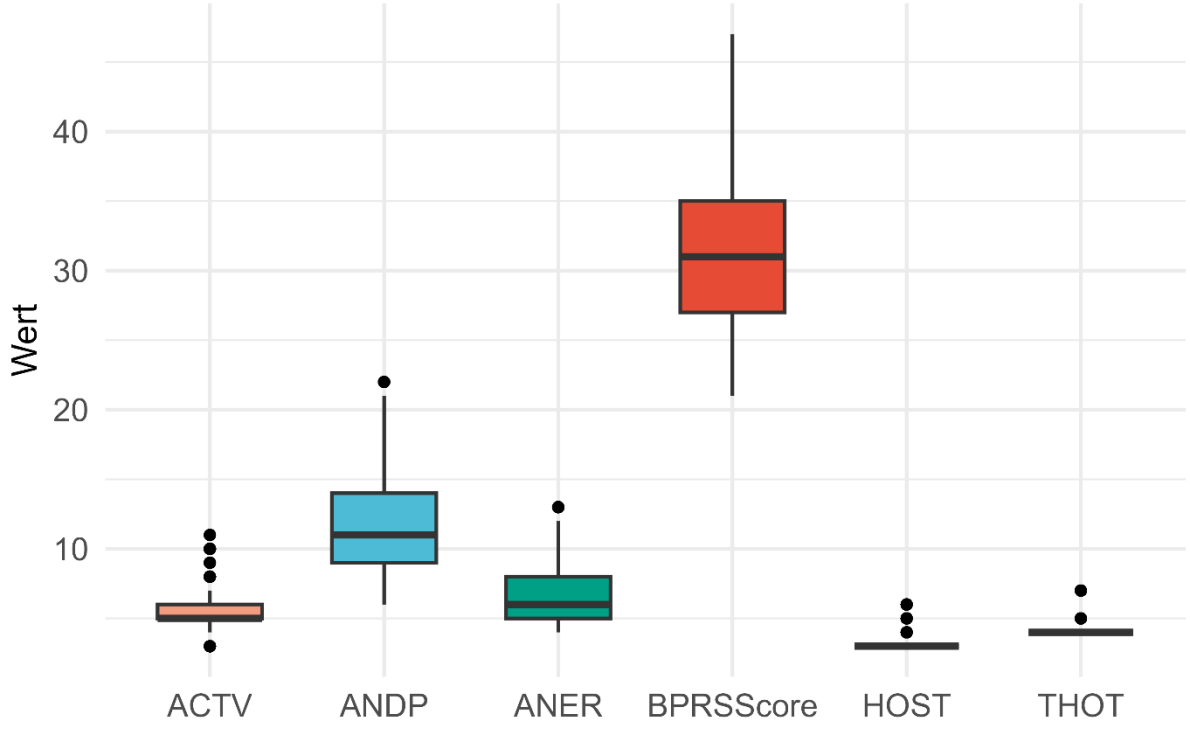


Abbildung 5: Boxplots von BPRS-Gesamtscore und Subskalen: (Box = IQR, Linie = Median, Whisker = 1,5×IQR) mit überlagertem Mittelwert (roter Punkt) und ±1 SD (rote Balken); n = 50.

3.3 Ergebnisse aus den Smartphonedaten

3.3.1 Vollständigkeit der Daten

Um die Forschungsfrage dieser Arbeit beantworten zu können, wurde das Smartphone Aktivitätsverhalten von Patient*innen mit Depression bestimmt. Begonnen wurde mit der Analyse der Vollständigkeit der Datenaufzeichnung über die App „JTrack Social“.

Tabelle 8 zeigt die deskriptiven Statistiken der fünf Aktivitätsvariablen (ON_BICYCLE_min = Fahrradfahren, STILL_min = Inaktivitätszeit, WALKING_min = Gehen, RUNNING_min = Laufen, Total_Activity = Gesamtaktivität) über insgesamt 2.908 Beobachtungstage. Die Anzahl gültiger Werte (N) variierte je nach Aktivitätskategorie zwischen 1.715 und 2.237, mit 671 bis 1.193 fehlenden Werten pro Variable.

Statistik	Fahrradfahren	Inaktivitätszeit	Gehen	Laufen	Gesamtaktivität

N	2228	1715	2234	2237	2225
fehlende Werte	680	1193	674	671	683
Mittelwert	3.69	934	49.4	0.690	53.7
Median	0.00	970	32.0	0.00	36.0
SD	13.4	210	54.6	4.77	58.8
Minimum	0	422	0	0.00	0.00
Maximum	145	1230	296	61.4	411

Tabelle 8: Vollständigkeit der Datenaufzeichnung mithilfe der App „JTrack Social“. Vollständige. Daten liegen bei „Aktivität“ vor, wenn an dem Tag mindestens 10 Datenpunkte aufgezeichnet sind.

Hinsichtlich der Vollständigkeit der Aktivitätsvariablen wiesen Gehen (WALKING_min; N = 2.234), Laufen (RUNNING_min; N = 2.237), Gesamtaktivität (Total_Activity; N = 2.225) und Fahrradfahren (ON_BICYCLE_min; N = 2.228) eine vergleichbare Datenabdeckung mit jeweils etwa 23 % fehlenden Werten auf. Die Variable Inaktivitätszeit (STILL_min) war demgegenüber unvollständiger erfasst (N = 1.715; ca. 41 % fehlende Werte). Dies ist auf die Festlegung der Wertspanne von 240–1230 Minuten pro Tag zurückzuführen, durch die Werte außerhalb dieses Bereichs ausgeschlossen wurden.

Die Inaktivitätszeit (STILL_min) nahm mit M = 934 Min/Tag den größten Anteil der erfassten Zeit ein. Für Gehen lag der Mittelwert bei M = 49,4 Min/Tag, für Fahrradfahren bei M = 3,69 Min/Tag und für Laufen bei M = 0,69 Min/Tag. Die Gesamtaktivität (Total_Activity) belief sich im Mittel auf 53,7 Min/Tag bei einer Standardabweichung von SD = 58,8. Die individuellen Werte der Gesamtaktivität reichten von 0 bis 411 Min/Tag.

Alle Aktivitätsvariablen waren rechtsschief verteilt; Median und Mittelwert wichen voneinander ab, und das Minimum lag jeweils bei 0 Minuten. Bei Laufen und Fahrradfahren betrug der Median 0 Min/Tag. Für die Inaktivitätszeit lag kein Wert bei 0 Min/Tag.

Das Boxplot-Diagramm in Abbildung 6 zeigt die Verteilung der Aktivitätsvariablen. Der größte Zeitanteil entfiel auf die Inaktivitätszeit (Median = 970 Min/Tag). Für Gehen betrug der Median 32 Min/Tag, für die Gesamtaktivität 36 Min/Tag. Fahrradfahren und Laufen wiesen jeweils einen Median von 0 Min/Tag auf; einzelne Werte lagen deutlich oberhalb des Medians.

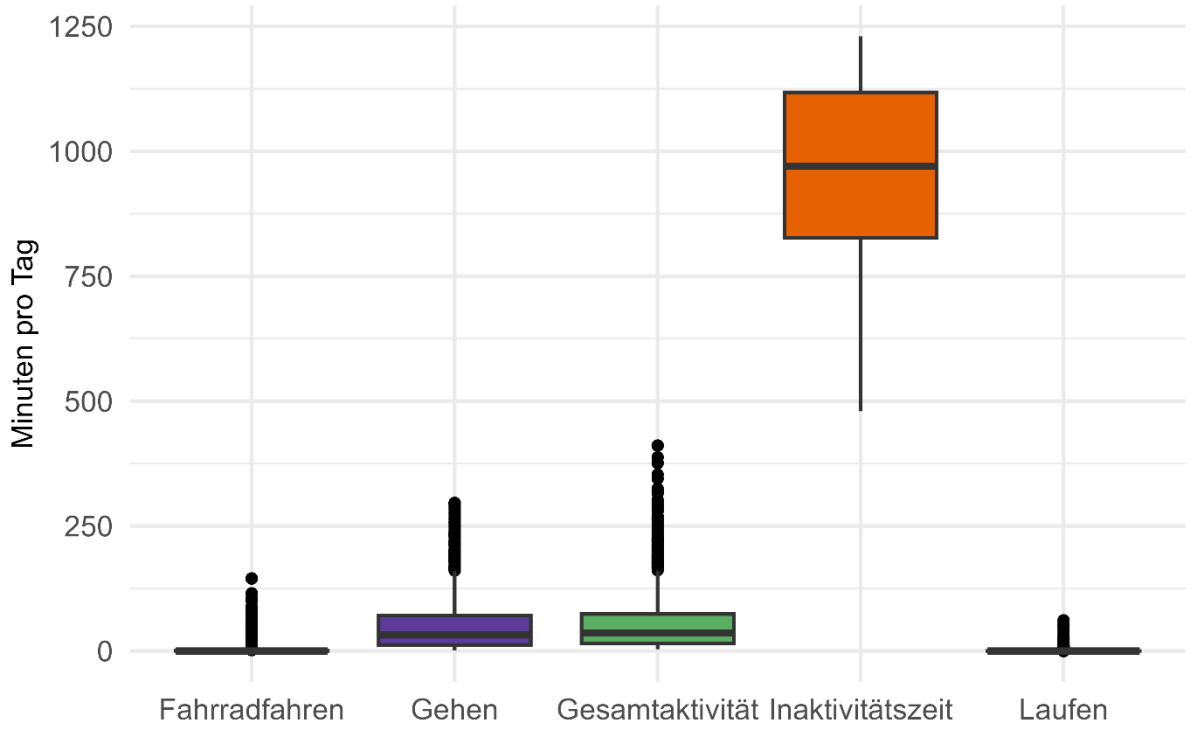


Abbildung 6: Tägliche Aktivitätsdauer nach Kategorie (Minuten/Tag); Boxplots mit Median und Ausreißern.

3.3.2 Vergleich der physischen Aktivität mit der selbstberichteten Psychopathologie

Zur Überprüfung der ersten Hypothese wurden lineare gemischte Modelle berechnet, um den Zusammenhang zwischen digital erfasster körperlicher Aktivität und den wöchentlich erhobenen EMA-Angaben zum psychischen Befinden zu untersuchen. Berücksichtigt wurden die fünf Aktivitätsmaße Gesamtaktivität, Fahrradzeit, Laufzeit, Gehzeit und Inaktivitätszeit sowie die zwölf EMA-Items. Die durchschnittliche tägliche Dauer dieser Aktivitätsformen über den Studienverlauf ist in den Abbildungen 8 und 9 dargestellt.

Für jede der fünf Bewegungsvariablen wurden zwölf Modelle geschätzt, in denen die EMA-Items (Fragen 1–12) als Prädiktoren der jeweiligen Aktivitätsdauer dienten. Alter, Geschlecht und Studientag wurden in allen Modellen als feste Effekte berücksichtigt, ergänzt durch einen zufälligen Intercept auf Subjektebene. Die inferenzielle Bewertung basierte auf FDR-korrigierten p-Werten. Die Regressionskoeffizienten (β) geben die erwartete Veränderung der Aktivitätsdauer in Minuten pro Anstieg des jeweiligen Prädiktors um eine Einheit an (vgl. Tabelle 9).

Nach Korrektur für multiples Testen ergaben sich keine statistisch signifikanten Zusammenhänge zwischen den EMA-Items und den Aktivitätsmaßen.

Für die Gesamtaktivität variierten die β-Schätzungen zwischen −1,65 und +5,35; alle 95 %-Konfidenzintervalle umfassten die Null. Die höchsten unadjustierten Schätzungen zeigten sich für Bedrohungs-/Verfolgungserleben (Frage 10; β = 5,35 [−0,18; 10,97]) und halluzinatorisches Erleben (Frage 11; β = 3,36 [−2,59; 9,37]); nach FDR-Korrektur waren beide Effekte nicht signifikant.

Für Fahrradfahren lagen die β-Werte zwischen −0,41 und +0,25 mit durchweg breiten Konfidenzintervallen, die die Null einschlossen. Systematische Zusammenhänge mit den EMA-Items ließen sich weder in positiver noch in negativer Richtung erkennen.

Bei der Laufzeit bewegten sich die β-Schätzungen zwischen −0,23 und +0,29. Für die EMA-Items allgemeines Befinden (Frage 1), Stimmung (Frage 3), körperliche/geistige Energie (Frage 5) und Freude (Frage 8) waren die β-Schätzungen positiv. Alle Konfidenzintervalle umfassten jedoch die Null, und nach FDR-Korrektur bestand keine Signifikanz (vgl. Abb. 7).

Für Gehen ergaben sich β-Schätzungen zwischen −1,73 und +6,57. Der stärkste unadjustierte Zusammenhang bestand zwischen Bedrohungs-/Verfolgungserleben (Frage 10) und Gehzeit (β = 6,57 [1,93; 12,07]); das Konfidenzintervall umfasste die Null nicht. Nach FDR-Korrektur war auch dieser Zusammenhang nicht mehr signifikant (vgl. Abb. 6).

Bei der Inaktivitätszeit reichten die β-Schätzungen von −11,68 bis +12,92. Ein positiver Trend zeigte sich für körperliche Beschwerden (Frage 12; β = +12,92 [1,37; 24,94]), der nach FDR-Korrektur ebenfalls nicht signifikant blieb.

Die in Hypothese 1 postulierten Zusammenhänge zwischen körperlicher Aktivität und selbstberichteter Symptomschwere konnten in dieser Stichprobe statistisch nicht bestätigt werden. Unkorrigierte Effektrichtungen werden in Abschnitt 4.1 explorativ eingeordnet.

EMA-Item	Aktivität	β	95 %-KI	p (unkorrigiert)	p (FDR-korrigiert)
Bedrohungs-/Verfolgungserleben	Gehen	6.57	[1.93, 12.07]	.012	.144
Allgemeines Befinden	Laufen	0.29	[0.02, 0.53]	.021	.106
Stress	Laufen	−0.23	[−0.46, −0.02]	.032	.106
Stimmung	Laufen	0.25	[0.03, 0.54]	.039	.106
Körperliche/geistige Energie	Laufen	0.27	[0.01, 0.53]	.044	.106
Freude	Laufen	0.25	[0.02, 0.49]	.038	.106
Körperliche Beschwerden	Laufen	−0.19	[−0.38, 0.02]	.061	.122
Körperliche Beschwerden	Inaktivitätszeit	+12.92	[1.37, 24.94]	.036	.432

Tabelle 9: Zusammenhang zwischen EMA-Items und passiv erfasster körperlicher Aktivität. Ergebnisse linearer gemischter Modelle mit β-Koeffizienten, 95 %-Konfidenzintervallen sowie unkorrigierten und BH-FDR-korrigierten p-Werte

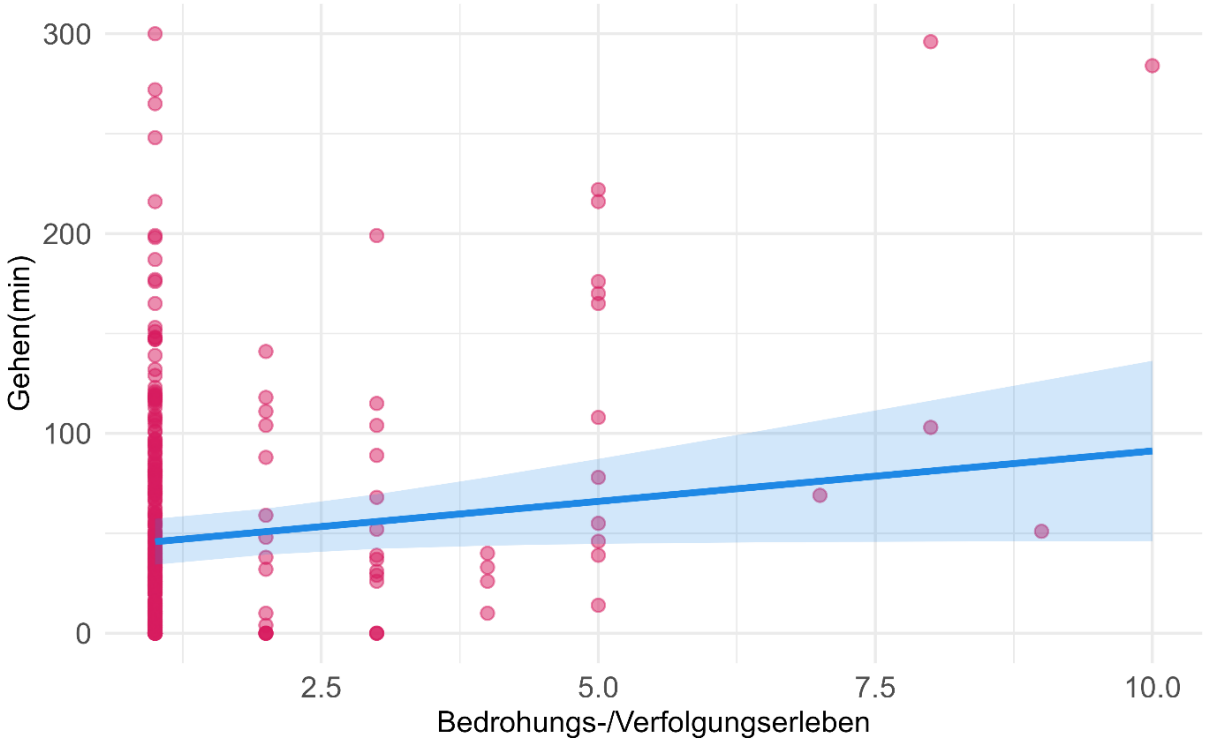


Abbildung 7: Modellbasierter Effektplot des Zusammenhangs zwischen Bedrohungs-/Verfolgungserleben (EMA) und Gehaktivität (Minuten/Tag), adjustiert für Alter, Geschlecht und Studientag.

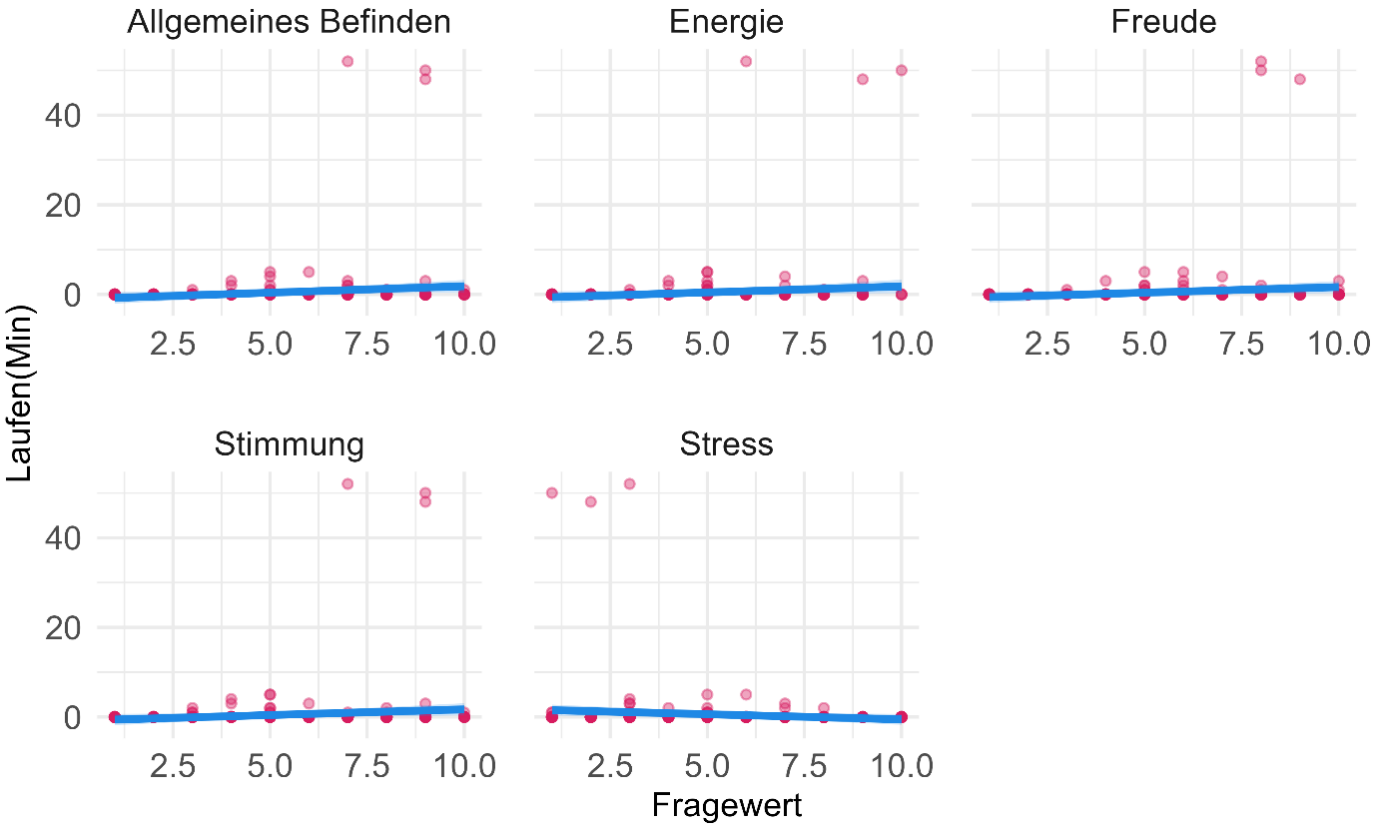


Abbildung 8: Modellbasierte Zusammenhänge zwischen der Laufaktivität (in Minuten pro Tag) und sechs EMA-Items, welche verschiedene Aspekte des subjektiven Befindens abbilden.

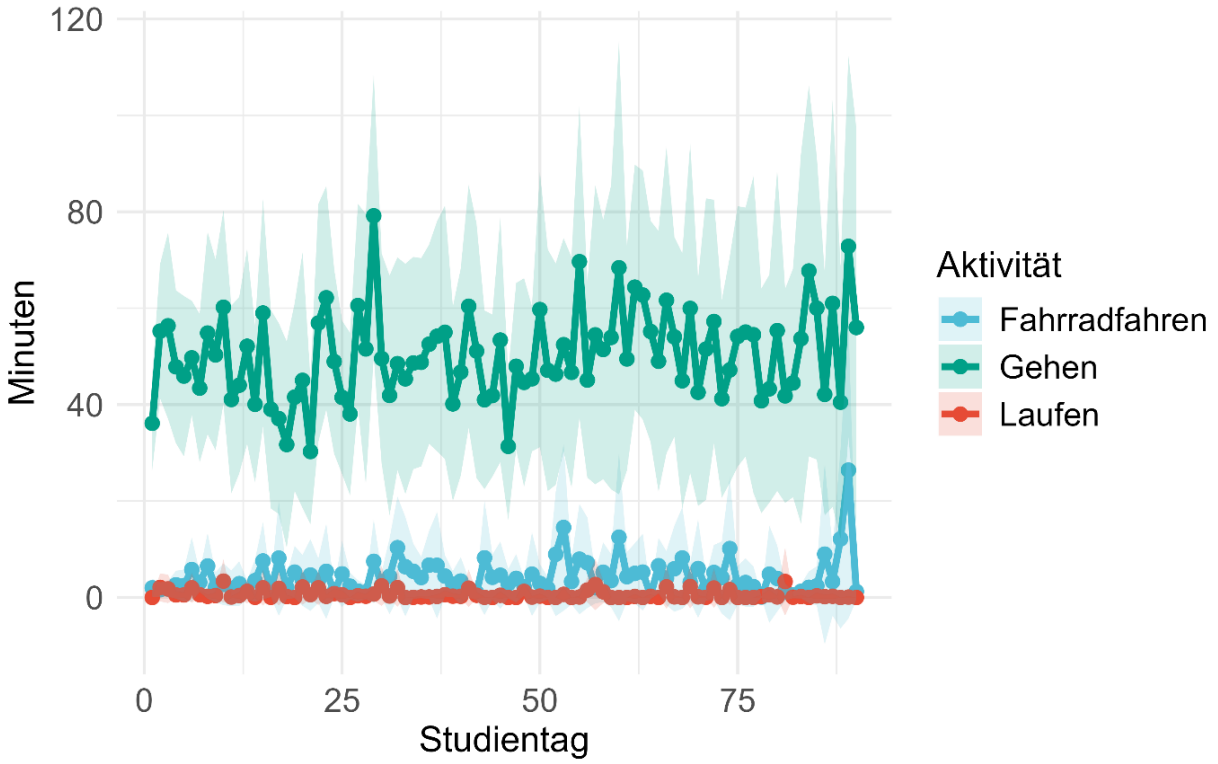


Abbildung 9: Durchschnittliche körperliche Aktivität im Studienverlauf. Gezeigt sind die über alle Teilnehmenden aggregierten mittleren täglichen Minuten für Gehen, Laufen und Fahrradfahren pro Studientag.

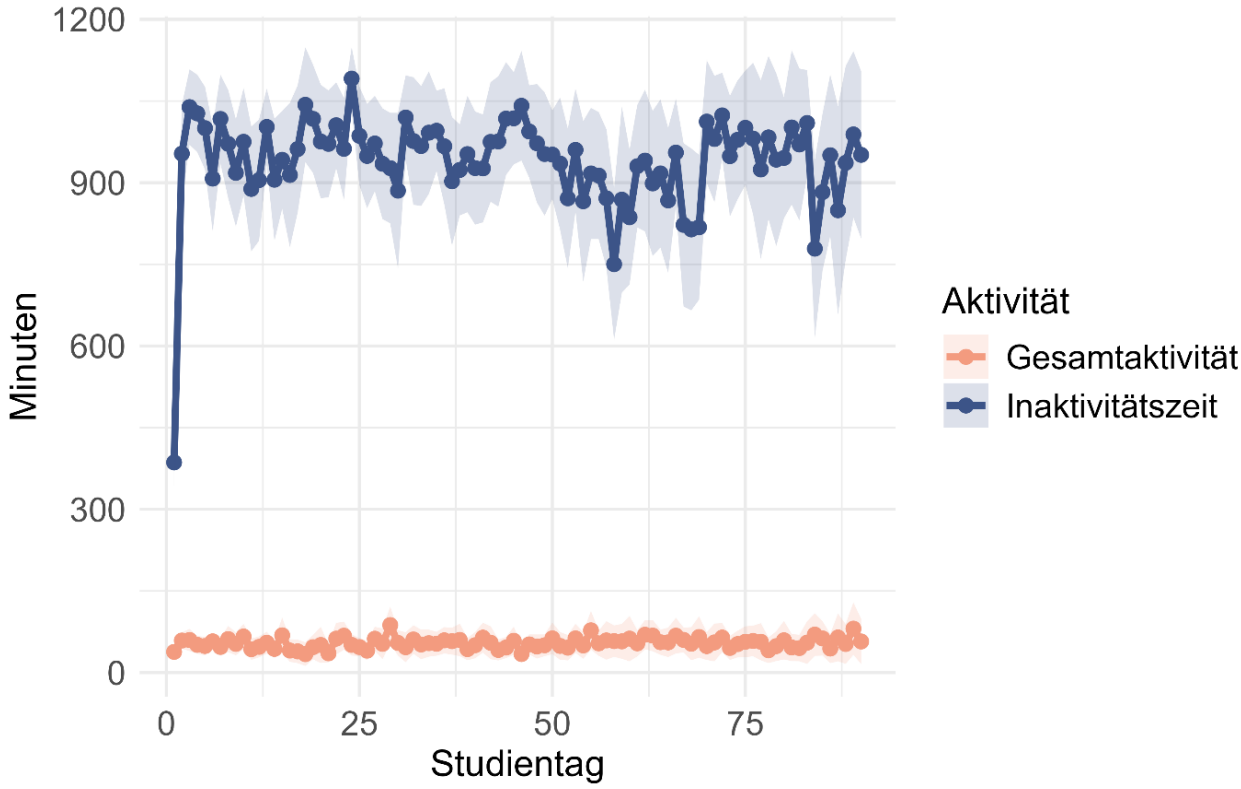


Abbildung 10: Durchschnittliche körperliche Aktivität im Studienverlauf. Gezeigt sind die über alle Teilnehmenden aggregierten mittleren täglichen Minuten für die Gesamtaktivität und die Inaktivitätszeit pro Studientag.

3.3.3 Fremdbeurteilung und physische Aktivität

Zur Überprüfung der zweiten Hypothese wurde für jedes der fünf Aktivitätsmaße (Minuten pro Tag) eine lineare Regression geschätzt. Als Prädiktor diente die einmalig zu Studienbeginn erhobene BPRS, wobei der Gesamtscore in Modell A und die fünf Subskalen in Modell B getrennt analysiert wurden. Berichtet werden Regressionskoeffizient (β), Standardfehler (SE), 95 %-Konfidenzintervall, t-Wert und FDR-korrigierter p-Wert (vgl. Tabelle 10).

Nach Korrektur für multiples Testen ergaben sich keine statistisch signifikanten Zusammenhänge zwischen den BPRS-Prädiktoren und den Aktivitätsmaßen.

Für die Gesamtaktivität variierten die β-Schätzungen zwischen −189,61 und +37,68. Weder der BPRS-Gesamtscore (β = −5,19; p_FDR = 0,685) noch die Subskalen zeigten signifikante Zusammenhänge mit der durchschnittlichen Aktivität.

Bei der Fahrradzeit reichten die β-Werte von −1,03 bis +4,16. Vor FDR-Korrektur ergab sich ein positiver Zusammenhang zwischen der Subskala Hostility (HOST) und der Fahrradzeit (β = 4,16 [0,41; 7,92]); das 95 %-Konfidenzintervall umfasste die Null nicht. Nach FDR-Korrektur war dieser Effekt nicht mehr signifikant (p_FDR = 0,186; vgl. Abb. 11).

Für die Laufzeit bewegten sich die β-Schätzungen zwischen −0,14 und +0,10. Die numerisch größte Schätzung zeigte sich für Anxiety/Depression (ANDP) und die Laufzeit (β = −0,10 [−0,20; 0,00]; p_FDR = 0,336); auch dieser Effekt war nach FDR-Korrektur nicht signifikant.

Bei der Gehzeit variierten die β-Werte zwischen −1,80 und +5,12, bei der Inaktivitätszeit zwischen −44,49 und +44,90. In keinem dieser Modelle ergaben sich nach FDR-Korrektur signifikante Effekte.

Weder Alter noch Geschlecht zeigten nach FDR-Korrektur signifikante Zusammenhänge mit den Aktivitätsmaßen. Vor FDR-Korrektur wiesen weibliche Teilnehmende für einzelne Aktivitätsmaße (z. B. Fahrradzeit) niedrigere Werte auf.

Hypothese 2 ließ sich in dieser Stichprobe nach FDR-Korrektur nicht nachweisen. Unkorrigierte Effektrichtungen werden in Abschnitt 4.1 eingeordnet.

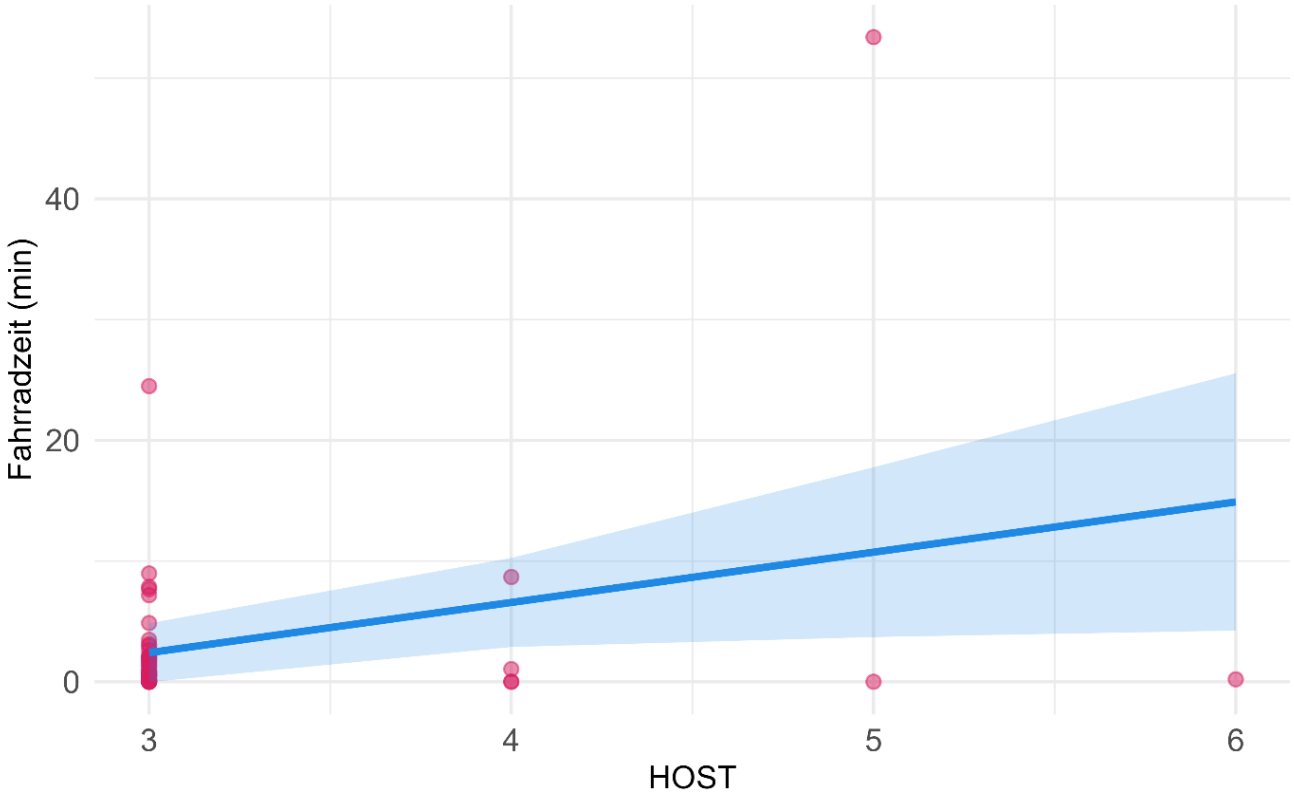


Abbildung 11: Effektplot zur linearen Regression zwischen der BPRS-Subskala Hostilität/Feindseligkeit (HOST) und der durchschnittlichen Fahrradzeit (Minuten pro Tag). Der unadjustierte positive Zusammenhang (β = 4.16; 95 %-KI [0.41; 7.92]) verlor nach FDR-Korrektur seine Signifikanz.

BPRS (Subskala)	Aktivität	β	95 %-KI	p (unkorrigiert)	p (FDR-korrigiert)
HOST	Fahrradfahren	4.16	[0.41; 7.92]	.031	0.186
ANDP	Laufen	−0.10	[−0.20; 0.00]	.056	0.336
ANER	Gesamtaktivität	31.22	[−17.40; 79.82]	.197	0.494
ANDP	Gesamtaktivität	−14.62	[−42.20; 12.96]	.285	0.494
ACTV	Gesamtaktivität	−35.90	[−110.30; 38.51]	.329	0.494
HOST	Gesamtaktivität	−189.61	[−441.60; 62.36]	.133	0.494

Tabelle 10: Zusammenhang zwischen BPRS-Subskalen und passiv erfasster körperlicher Aktivität. Ergebnisse linearer Regressionsmodelle mit β-Koeffizienten, 95 %-Konfidenzintervallen sowie unkorrigierten und FDR-korrigierten p-Werten.

4 Diskussion

4.1 Analyse der körperlichen Aktivität und psychischer Symptomatik

Die vorliegende Studie untersuchte über einen Zeitraum von 90 Tagen den Zusammenhang zwischen passiv erfasster körperlicher Aktivität und psychopathologischer Symptomatik bei 50 erwachsenen Patient*innen mit depressiver Störung. Im Vergleich zu bisherigen Arbeiten zur digitalen Phänotypisierung bei Depression zeichnet sich die Studie durch einen langen Beobachtungszeitraum und eine engmaschige passive Datenerfassung aus. Mithilfe der über das persönliche Smartphone erhobenen Daten ließ sich das alltägliche Bewegungsverhalten der Teilnehmenden quantitativ charakterisieren und mit selbstberichteter sowie klinisch fremdbeurteilter Symptomatik in Beziehung setzen. Im Folgenden werden zunächst die deskriptiven Befunde zur körperlichen Aktivität in den Forschungskontext eingeordnet; anschließend werden die Ergebnisse zu den beiden präregistrierten Hypothesen diskutiert und die eigenständigen Beiträge der Arbeit herausgearbeitet.

Die durchschnittliche tägliche Gesamtaktivität betrug 53,7 Minuten (SD = 58,8), bei einer mittleren Inaktivitätszeit von 934 Minuten pro Tag (SD = 210). Die aktive Zeit verteilte sich überwiegend auf Gehen (M = 49,4 Min/Tag), während Laufen (M = 0,69 Min/Tag) und Fahrradfahren (M = 3,69 Min/Tag) eine untergeordnete Rolle spielten. Dieses Muster aus niedriger Gesamtaktivität, hoher Inaktivitätszeit und einer Verschiebung hin zu energieschonenden Bewegungsformen steht im Einklang mit Befunden aus Wearable- und Aktigraphiestudien zu depressionsassoziierten Aktivitätsmustern (Rykov et al., 2021; Wüthrich et al., 2022). Ein möglicher Erklärungsansatz liegt in depressionsbedingten Antriebs- und Aktivierungseinschränkungen, die die Aufnahme intensiver körperlicher Aktivität erschweren könnten (Noetel et al., 2024).

Eine Einordnung der gemessenen 53,7 Minuten täglicher Aktivität gegenüber der Allgemeinbevölkerung ist nur mit Vorbehalt möglich. Troiano et al., (2008) beobachteten in der US-Bevölkerung etwa 144 Minuten, Willetts et al., (2018) in der UK Biobank etwa 150 Minuten tägliche Aktivität. Beide Studien nutzten tragbare Accelerometer (hüft- bzw. handgelenksgetragen), während in der vorliegenden Arbeit die smartphonebasierte Klassifikation über die Google Activity Recognition API zum Einsatz kam. Da sich diese Verfahren methodisch unterscheiden und nicht direkt ineinander umrechenbar sind, sind die hier berichteten Werte als

smartphonebasierte Aktivitätsschätzung in einer klinischen Stichprobe zu lesen und nicht als quantifizierbares Defizit gegenüber populationsbasierten Referenzwerten.

Hypothese 1 postulierte einen Zusammenhang zwischen höheren passiv erfassten Aktivitätsmaßen und geringerer selbstberichteter Symptomschwere sowie zwischen höherer Inaktivitätszeit und stärkerer Symptomschwere. Nach FDR-Korrektur ließ sich diese Hypothese nicht bestätigen: Sämtliche der 60 geschätzten Zusammenhänge zwischen den fünf Aktivitätsmaßen und den zwölf EMA-Items waren nach Adjustierung statistisch nicht signifikant (alle $p_{FDR} > 0{,}10$). Hypothese 1 wird daher zurückgewiesen.

Die unkorrigierten Schätzungen zeigen ein kohärentes Bild: Laufaktivität geht mit positiven β-Werten für allgemeines Befinden, Stimmung, Energie und Freude einher und mit negativen β-Werten für Stress und körperliche Beschwerden. Dieses Muster reproduziert die Richtung der etablierten Befunde zum antidepressiven Effekt aerober Bewegung (Noetel et al., 2024; Singh et al., 2023). Bei der hier vorliegenden Stichprobengröße liegt die FDR-Schranke erwartungsgemäß über den beobachteten Effekten; der konsistente Richtungsbefund ist daher als Replikationshinweis und als Begründung für hinreichend gepowerte Folgestudien zu werten.

Der unadjustierte Zusammenhang zwischen Bedrohungs-/Verfolgungserleben (Frage 10) und Gehzeit ($\beta = 6{,}57$ [1,93; 12,07]; $p_{FDR} = 0{,}14$) verdient eine methodische Einordnung. Das Item zeigt in der vorliegenden Stichprobe einen ausgeprägten Bodeneffekt ($M = 1{,}50$; Median = 1,00). Bei einer derart schiefen Verteilung wird die Steigungsschätzung durch wenige Antworten am oberen Skalenende stark beeinflusst, sodass der β-Wert keine belastbare Aussage über einen inhaltlichen Zusammenhang erlaubt.

Auch die Literatur stützt diese Einordnung. In einer randomisiert-kontrollierten Studie zu einem dreimonatigen Walking-Programm bei chronischer Schizophrenie berichten Loh et al. (2015) zwar Verbesserungen der Lebensqualität und des Funktionsniveaus, jedoch keine konsistenten Effekte auf die Positivsymptomatik. Die Metaanalyse von (Oliva et al., 2023) belegte für Bewegungsinterventionen einen kleinen, aber signifikanten Effekt auf positive Symptome ($MD = -0{,}75$), wobei die Effekte auf negative und allgemeine Symptome deutlich stärker ausfielen. Gehen erscheint somit auch in der Literatur als wenig spezifische Intervention für paranoides Erleben oder andere Positivsymptome. Dieser Befund passt zu der hier vorgenommenen methodischen Einordnung des Bedrohungserleben-Effekts.

Mehrere methodische Faktoren erschweren die Detektion robuster Effekte. Erstens waren Laufen (Median = 0 Min/Tag) und Fahrradfahren (Median = 0 Min/Tag) so selten, dass eine zuverlässige Modellierung dieser Variablen kaum möglich war. Zweitens war die wöchentliche EMA-Frequenz vermutlich zu grob aufgelöst, um tagesnahe Schwankungen abzubilden, die in höher aufgelösten EMA-Designs (mehrmals täglich) als zentrale Stärke des Verfahrens gelten (Maatoug et al., 2021; Van Genugten et al., 2021). Drittens fehlten aufgrund der reduzierten Stichprobengröße (n = 50 statt der a-priori berechneten 147) die statistischen Voraussetzungen zur Detektion kleiner bis mittlerer Effekte. Vergleichbare Schwierigkeiten in der Replikation robuster Aktivität-EMA-Assoziationen unter konservativen statistischen Kriterien werden in anderen digitalen Phänotypisierungsstudien berichtet (Moshe et al., 2021; Rykov et al., 2021; Zierer et al., 2024)

Auch Hypothese 2, die einen Zusammenhang zwischen passiv erfassten Aktivitätsmaßen und der ärztlich erhobenen Symptomschwere (BPRS-Gesamtscore und Subskalen) postulierte, wird zurückgewiesen: Weder der BPRS-Gesamtscore noch die Subskalen zeigten nach FDR-Korrektur signifikante Zusammenhänge mit einem der Aktivitätsmaße (alle $p_{FDR} > 0{,}18$).

Der unadjustierte Zusammenhang zwischen der HOST-Subskala und Fahrradzeit ($\beta = 4{,}16$ [0,41; 7,92]) ist analog zum Befund bei Frage 10 methodisch einzuordnen: Die HOST-Werte lagen in der Stichprobe zwischen 3 und 6 mit $M = 3{,}22$ und damit bei einem deutlichen Bodeneffekt. Eine sinnvolle Modellierung dieser Subskala als kontinuierlicher Prädiktor war damit nur eingeschränkt möglich, und der unadjustierte Effekt ist überwiegend durch wenige Personen mit Werten oberhalb des Skalenminimums bedingt. Eine inhaltliche Interpretation des HOST-Effekts ist daher nicht angezeigt.

Theoretisch wäre ein Zusammenhang zwischen körperlicher Aktivität und Aspekten von Hostilität bzw. emotionaler Regulation grundsätzlich plausibel. Van Der Sluys et al. (2024) berichten Hinweise darauf, dass körperliche Aktivität mit einer Reduktion von Ärger-, Hostilitäts- und antisozialen Verhaltensweisen assoziiert sein kann. Die vorliegende Stichprobe bietet jedoch aufgrund der Bodeneffekte der HOST-Subskala keine geeignete Datengrundlage, um diese Hypothesen empirisch zu prüfen; sie können allenfalls als Anregung für künftige Studien mit klinisch heterogeneren Stichproben dienen.

Über die geringe Stichprobengröße hinaus ist eine weitere methodische Besonderheit hervorzuheben. Die BPRS wurde ausschließlich zu Studienbeginn erhoben und mit der über 90 Tage gemittelten Aktivität in Beziehung gesetzt. Diese Erhebungsstruktur prüft den Zusammenhang zwischen Symptomschwere bei Studieneinschluss und mittlerem Aktivitätsniveau im

nachfolgenden Beobachtungszeitraum. Eine Modellierung dynamischer Veränderungen, also ob Symptomverbesserungen mit Aktivitätsveränderungen einhergehen, war mit diesem Design nicht möglich. Eine wiederholte BPRS-Erhebung im Studienverlauf wäre für eine solche dynamische Analyse erforderlich und stellt eine zentrale Empfehlung für künftige Studien dar.

Unabhängig von der formalen Hypothesenprüfung liefert die vorliegende Studie mehrere Befunde, die für die digitale Phänotypisierung depressiver Störungen relevant sind. Zum einen wurde gezeigt, dass eine kontinuierliche, smartphonebasierte Aktivitätserfassung über 90 Tage in einer klinischen Depressionsstichprobe technisch und organisatorisch realisierbar ist. Mit einer Datendichte von rund 77 % über 2.908 Beobachtungstage und einer Halterate von 42 % über die gesamten 90 Tage erreicht die Studie eine Datenbasis, die im Vergleich mit kürzeren Beobachtungsdauern in der bisherigen Literatur als tragfähig zu bewerten ist. Maatoug et al. (2021) berichten in einer EMA-Machbarkeitsstudie über 28 Tage bei Patient*innen mit Major Depression eine durchschnittliche Antwortrate von 62,5 %, wobei rund zwei Drittel der eingeschlossenen Personen den geplanten Erhebungszeitraum vollständig durchliefen. Die vorliegende Studie erreichte trotz dreifach längerer Beobachtungsdauer eine vergleichbar hohe Datendichte. Auffällig war zugleich ein erhöhtes Abbruchrisiko in den ersten Studienwochen: 24 % aller Teilnehmenden brachen innerhalb der ersten 24 Tage ab, was auf eine kritische Frühphase hinweist und für künftige Studien eine niedrigschwellige Begleitung der Initialphase nahelegt.

Darüber hinaus liefert die Arbeit eine quantitative Charakterisierung des Bewegungsprofils erwachsener Patient*innen mit unipolarer Depression unter Alltagsbedingungen. Die mittlere Inaktivitätszeit von 934 Minuten pro Tag entspricht nach Abzug einer durchschnittlichen Schlafdauer etwa sieben bis acht Stunden sedentärer Wachzeit. Die Verteilung der aktiven Zeit auf überwiegend Gehen bei nahezu vollständiger Abwesenheit von Laufen und Fahrradfahren bestätigt unter den Vorbehalten der Methodenvergleichbarkeit das in der Literatur beschriebene Muster energieschonender Alltagsmobilität bei Depression (Rykov et al., 2021; Wüthrich et al., 2022). Schließlich verdeutlichen die Befunde die ausgeprägte interindividuelle Heterogenität des Bewegungsverhaltens innerhalb der Diagnose Depression. Die Gesamtaktivität rangierte zwischen 0 und 411 Minuten pro Tag, die Inaktivitätszeit zwischen 422 und 1.230 Minuten, eine Spannweite, die die Standardabweichung des Mittelwerts überschreitet. Dieser Befund stützt die in der Literatur diskutierte Annahme, dass die diagnostische Katego-

rie „Depression“ sehr heterogene Verhaltensphänotypen umfasst (Brady et al., 2023). Der Befund unterstreicht den in der Präzisionspsychiatrie diskutierten Bedarf an individualisierten, verhaltensbezogenen Phänotypisierungsansätzen (vgl. Abschnitt 1.5).

Die vorliegende Arbeit reiht sich in ein methodisch noch junges Forschungsfeld ein. Die systematische Übersichtsarbeit von Bufano et al. (2023) analysiert 29 Studien zur digitalen Phänotypisierung psychischer Erkrankungen und kommt zu dem Schluss, dass smartphonebasiert erfasste Mobilitäts- und Aktivitätsdaten in zahlreichen Studien konsistente Assoziationen mit Symptomverläufen zeigen, die Übertragung in robuste, replizierbare klinische Vorhersagen jedoch noch erheblicher methodischer Weiterentwicklung bedarf. Als zentrale Herausforderungen werden insbesondere heterogene Stichproben, begrenzte klinische Vergleichsmaße und die große Variabilität der eingesetzten passiven Parameter beschrieben. Die Befunde der vorliegenden Studie passen gut zu diesem Bild. Dazu zählen die stabile Datenerhebung über 90 Tage, konsistente Effektrichtungen trotz ausbleibender statistischer Signifikanz nach FDR-Korrektur sowie methodische Bodeneffekte einzelner Skalenitems. Gleichzeitig verdeutlichen sie, welche methodischen Anforderungen zukünftige Studien zur digitalen Phänotypisierung bei Depression berücksichtigen sollten.

Eine zusätzliche methodische Anmerkung betrifft den Stellenwert der EMA-Selbstauskünfte gegenüber der klinischen Fremdbeurteilung. Während die BPRS punktuell und zu einem festen Zeitpunkt erhoben wurde, liefern die EMA-Daten einen dichten, ökologisch validen Zugang zum subjektiven Erleben der Patient*innen über den gesamten Beobachtungszeitraum. Selbstberichte sind in der Literatur als methodisch valide und für die Erfassung subjektiver Zustände indirekten Verfahren häufig überlegen anerkannt (Corneille & Gawronski, 2024). EMA-Verfahren besitzen darüber hinaus eine hohe ökologische Validität und lassen sich gut mit passiv erfassten Bewegungsdaten kombinieren (Browning et al., 2024).

4.2 Analyse der Studiendurchführung

Eine kritische Bewertung der Studiendurchführung erlaubt Rückschlüsse auf die Qualität der erhobenen Daten und die darauf basierenden Schlussfolgerungen. Im folgenden Abschnitt soll daher die Rekrutierung, die Vollständigkeit der Daten und die demografischen sowie psychometrischen Charakteristika der Teilnehmenden analysiert werden.

Der Anteil von Zusagewilligen an allen Kontaktaufnahmen deutet mit rund einem Viertel (≈ 25 %) auf ein moderates Interesse an der Studienteilnahme hin. Häufig genannte Ablehnungsgründe waren eine generelle technische Überforderung, die Nicht-Nutzung von Smartphones oder explizite Datenschutzbedenken. Die während der Entwicklung der Applikation angewandten hohen Datenschutzstandards dürften das Vertrauen der Teilnehmenden positiv beeinflusst haben (Sahandi Far et al., 2021). Besonders auffällig war, dass die Rekrutierung von Personen mit Schizophrenie durch wahnhafte Symptomatik (z. B. Beobachtungswahn) deutlich erschwert war. Dies erklärt, warum das ursprüngliche Rekrutierungsziel von 200 Teilnehmenden (100 mit Depression, 100 mit Schizophrenie) nicht erreicht werden konnte und die Stichprobe schließlich 50 Patient*innen mit Depression und 13 mit Schizophrenie umfasste. Damit musste sich die Auswertung der Daten auf die Depressionskohorte konzentrieren.

Die Dauer der Rekrutierung (November 2021 bis März 2023) sowie der Beobachtungszeitraum von 90 Tagen pro Person führten insgesamt zu einer Datenaufzeichnungsperiode von rund 19 Monaten. Es ist positiv hervorzuheben, dass bei allen Teilnehmenden die vollständigen demografischen Angaben vorlagen.

Die Datenerhebung über die App funktionierte überwiegend zuverlässig. An etwa 77 % der Beobachtungstage lagen gültige Aktivitätsdaten vor. Dennoch zeigte sich eine erhebliche Anzahl fehlender Werte mit Anteilen zwischen 23 % (Gehen, Laufen, Fahrradfahren) und 41 % (Inaktivitätszeit). Die Ursachen lagen einerseits in technischen Ausfällen (z. B. durch Batterieoptimierung oder fehlende Internetverbindung), andererseits in der Anwendung von Plausibilitätsfiltern, durch die unplausible Werte (z. B. Inaktivitätszeit über 20 Stunden pro Tag) ausgeschlossen wurden. Diese Filter erhöhten die Validität der Daten, führten jedoch gleichzeitig zu einer reduzierten Vollständigkeit.

Die Verweildauer in der Studie zeigte eine deutliche Heterogenität: 42 % der Teilnehmenden (n = 21) erreichten die vollen 90 Tage, weitere 10 % (n = 5) verblieben zwischen 75 und 89 Tagen. Etwa 58 % der Stichprobe beendeten die Teilnahme vorzeitig, einige bereits in den ersten 30 Tagen. Die Analyse der Datenkonformität ergab, dass nur 30 % der Teilnehmenden (n = 15) an allen 90 Tagen Daten übermittelten. Es bestand also ein Unterschied zwischen der formalen Verweildauer und der tatsächlichen Datenvollständigkeit. Eine hohe Abbruchrate gefährdet die Langzeitvalidität und kann zu Selektionsverzerrungen führen; die beobachteten Muster liegen jedoch im Rahmen dessen, was auch in anderen naturalistischen Digitalstudien festgestellt wurde.

Ein weiteres zentrales Merkmal der Studie ist die große Menge an erfassten Aktivitätsdaten über den gesamten Beobachtungszeitraum. Für jeden Teilnehmenden wurden kontinuierlich Bewegungsdaten in den Kategorien Gehen, Laufen, Fahrradfahren, Inaktivitätszeit und Gesamtaktivität aufgezeichnet. Diese mehrdimensionale Erfassung ermöglichte eine differenzierte Analyse des Alltagsverhaltens und der individuellen Bewegungsmuster.

Während die passive Bewegungsaufzeichnung über JTrack Social erfolgte, wurde die Datenerhebung aktiv durch wöchentliche EMA-Abfragen (Ecological Momentary Assessment) mit der App „JTrack EMA" ergänzt. Diese Komponente erfasste selbstberichtete Stimmungs-, Aktivitäts- und Befindlichkeitsverläufe und ist deshalb ein wesentlicher Bestandteil der Studiendurchführung.

Insgesamt lagen zwischen 386 und 392 gültige Datensätze pro Item vor, was einer Antwortquote von etwa 72 % entspricht. Etwa ein Viertel der geplanten EMA-Antworten fehlte, was einer moderaten, aber für naturalistische digitale Längsschnittstudien typischen Ausfallrate entspricht.

Bei der demografischen Zusammensetzung fällt eine leichte Überrepräsentation von Frauen (n = 31) gegenüber Männern (n = 19) auf. Das mittlere Alter lag bei 39 Jahren (Spannweite 18–63), sodass eine überwiegend jüngere bis mittlere erwachsene Kohorte abgebildet wurde. Der Anteil an komorbiden psychiatrischen Diagnosen und stationären Aufenthalten wurde bei Studieneinschluss dokumentiert.

Die App „JTrack Social" erwies sich insgesamt als benutzerfreundlich und praktikabel. Technische Einschränkungen, insbesondere durch Batterieoptimierung und Betriebssystemabhängigkeiten, führten jedoch vereinzelt zu Unterbrechungen der Datenerhebung.

Insgesamt wurde die Studiendurchführung erfolgreich umgesetzt; es konnte eine umfangreiche Aktivitätsdatenbasis über mehrere Monate erhoben werden. Die Dropouts und fehlenden Werte schränken die statistische Aussagekraft ein und müssen bei der Interpretation der Ergebnisse berücksichtigt werden. Trotz dieser Einschränkungen ist es gelungen, eine für naturalistische digitale Studien valide Datengrundlage zu schaffen.

4.3 Limitationen

Die Arbeit hat einige wichtige Limitationen, welche sich sowohl auf die Studienplanung als auch auf die Durchführung und die Datenauswertung beziehen. Die Erfassung dieser Einschränkungen stellt eine wesentliche Voraussetzung dar, um in einem weiteren Schritt Empfehlungen für zukünftige Studien ableiten zu können.

Eine zentrale Limitation ist die Stichprobengröße. Die ursprüngliche Studienplanung sah eine Gesamtstichprobe von bis zu 200 Teilnehmenden auf zwei Kohorten (Depression und Schizophrenie) verteilt vor. Die ausgewertete Stichprobe der vorliegenden Arbeit umfasste 50 Erwachsene mit einer depressiven Störung (vgl. Abschnitt 2.2.2). Die Power zur Detektion schwacher Effekte war damit begrenzt; das Ausbleiben signifikanter Zusammenhänge nach FDR-Korrektur ist auch im Kontext dieser reduzierten Fallzahl zu betrachten. Zukünftige Studien sollten mit einer größeren Stichprobe und gegebenenfalls einem multizentrischen Design geplant werden.

Ein weiterer Punkt betrifft die zeitliche Auflösung der klinischen Erhebung. Die Aktivität wurde über 90 Tage kontinuierlich gemessen, während die Fremdbeurteilung der Symptomatik mittels BPRS lediglich einmal zu Studienbeginn erfolgte. Dieses zeitliche Missverhältnis erschwert die Abbildung dynamischer Beziehungen zwischen psychopathologischer Schwere und Bewegungsverhalten. Auch die wöchentliche Frequenz der EMA-Abfragen war zu grob, um tagesnahe Dynamiken zuverlässig zu erfassen. Eine höhere zeitliche Auflösung würde es zudem ermöglichen, verzögerte oder nichtlineare Beziehungen („Lag-Effekte") zwischen Aktivität und Symptomverlauf zu modellieren.

Obwohl die Applikation „JTrack Social" fortlaufend verbessert wird, weist sie technische Einschränkungen auf. Je nach Hardware, Betriebssystem und Netzwerkbedingungen kann es zu Datenverlusten kommen (Sahandi Far et al., 2021). Die Aktivierung der Batterieoptimierung führte teilweise dazu, dass die Aufzeichnung unterbrochen wurde, ohne dass die Teilnehmenden es bemerkten. Im Studienzeitraum war die App ausschließlich für Android-Geräte verfügbar. Bei einem Android-Marktanteil von ca. 60,7 % in Deutschland (*Mobile Operating System Market Share Germany*, Oktober 2025) wurde damit ein Großteil der Zielpopulation erreicht; seit 2023 ist die Anwendung auch im Apple App Store verfügbar, was zukünftigen Studien eine breitere Rekrutierungsbasis ermöglicht.

Eine weitere Limitation betrifft die Qualität der Aktivitätsdaten. Der Anteil fehlender Werte lag je nach Variable zwischen 23 und 41 % und ist auf Hintergrundprozesse, Batterieoptimierung

und Internetverbindungsunterbrechungen zurückzuführen. Für die Inaktivitätszeit (STILL_min) wurden zusätzlich Werte außerhalb der definierten Wertspanne von 240–1230 Min/Tag durch einen Plausibilitätsfilter ausgeschlossen. Für die Inaktivitätszeit (STILL_min) wurden zusätzlich Werte außerhalb der definierten Wertspanne von 240–1230 Min/Tag durch einen Plausibilitätsfilter ausgeschlossen. Hierbei werden allerdings auch Tage ausgeschlossen, an denen Patientinnen und Patienten besonders inaktiv waren. Bei schweren depressiven Episoden ist genau dieses Verhalten klinisch zu erwarten. Die Inaktivitätszeiten in dieser Arbeit sind damit eher unterschätzt.

Auch die Aktivitätsvariablen selbst erwiesen sich als nur begrenzt aussagekräftig. Da Laufen und Fahrradfahren in der untersuchten Population kaum auftraten (Median = 0 Min/Tag), waren mögliche Zusammenhänge mit psychischen Variablen aufgrund von Bodeneffekten kaum detektierbar. Für zukünftige Analysen wäre es ratsam, komplexere Parameter wie Variabilitätsmaße oder Tagesmuster zu integrieren, anstatt sich auf absolute Minutenwerte zu beschränken.

Ein weiterer limitierender Aspekt der vorliegenden Untersuchung ist die Compliance der Teilnehmenden. Die Teilnahme an psychiatrischen Studien ist keineswegs selbstverständlich, da sie von zahlreichen individuellen und strukturellen Faktoren beeinflusst wird. In ihrer umfassenden Studie haben Bener et al. (2025) festgestellt, dass Non-Compliance oft durch Aspekte wie fehlende soziale Unterstützung, unregelmäßige Inanspruchnahme psychiatrischer Dienste, mangelnde Krankheitseinsicht, geringe Zufriedenheit mit der ärztlichen Behandlung, unzureichende Aufklärung über Medikamente und Nebenwirkungen sowie eine subjektiv empfundene geringe Wirksamkeit der Behandlung begünstigt wird. Auch die ökologische Momentanerhebung (EMA) hat mit ähnlichen Problemen zu kämpfen: In ihrer Untersuchung stellten Porras-Segovia et al. (2020) fest, dass die Retentionsraten über sechs Monate erheblich gesunken sind: Die Teilnahme an aktiver EMA („MEmind"-App) fiel nach sechs Monaten auf unter 25 %, während die passive EMA („eB2"-App) mit knapp 47 % zwar bessere, aber dennoch reduzierte Retentionsraten aufwies. Offensichtlich kann die variable Compliance von psychiatrischen Patient*innen Studien, die auf eine kontinuierliche Datenerfassung angewiesen sind, erheblich stören. Daher muss auch in der vorliegenden Arbeit berücksichtigt werden, dass die Generalisierbarkeit und Validität der Ergebnisse durch selektive Teilnahmebereitschaft und Ausfälle potenziell beeinträchtigt sind.

Auch die klinische Heterogenität der Stichprobe stellt eine Einschränkung dar. Erfasst wurden Erwachsene mit einer unipolaren Depression, jedoch ohne weitere Differenzierung nach Schweregrad, Krankheitsdauer, Medikation oder Komorbiditäten. Diese Variablen wurden

zwar dokumentiert, aber nicht in die statistischen Modelle integriert. Damit bleibt unklar, ob bestimmte Subgruppen, etwa Patientinnen und Patienten mit therapieresistenter Depression oder komorbider Angststörung, andere Aktivitätsmuster aufweisen.

Schließlich ist auf eine methodische Limitation in der Kovariatenauswahl hinzuweisen. In allen Modellen wurden ausschließlich Alter, Geschlecht und Studientag berücksichtigt. Wichtige Störgrößen wie aktuelle Medikation, Therapieereignisse oder psychosoziale Belastungen blieben unberücksichtigt. Eine Erweiterung der Modelle um diese klinisch relevanten Kovariaten könnte in zukünftigen Analysen helfen, konfundierende Einflüsse präziser zu kontrollieren.

4.4 Praktische und ethische Implikationen

Die vorliegende Studie zeigt, dass die passive Erfassung körperlicher Aktivität über eine Smartphone-App eine praktikable Ergänzung zu klassischen klinischen Erhebungsmethoden darstellen kann. Für die psychiatrische Versorgung ergibt sich daraus die Möglichkeit, objektive Aktivitätsdaten künftig ergänzend zu nutzen, um Krankheitsverläufe im Alltag besser abzubilden und Veränderungen frühzeitig zu erkennen. Die frei verfügbare Open-Source-App „JTrack Social“ erwies sich dabei als grundsätzlich praktikabel, technisch stabil und niedrigschwellig bedienbar. Die Implementierung orientierte sich an datenschutzrechtlichen Prinzipien und erscheint daher grundsätzlich für zukünftige Forschungsprojekte geeignet.

Ein möglicher klinischer Nutzen digitaler Phänotypisierung liegt in der frühen Erfassung von Verhaltensänderungen, die einer klinisch erkennbaren Symptomverschlechterung vorausgehen können. Voraussetzung hierfür ist allerdings eine deutlich größere und längsschnittlich stärker strukturierte Datenbasis als in der vorliegenden Studie. Die Ergebnisse dieser Arbeit verdeutlichen vor allem, welche methodischen Anforderungen zukünftige Studien berücksichtigen sollten. Dazu zählen insbesondere größere Stichproben, häufigere Erhebungen und eine sorgfältigere Auswahl der verwendeten Skalenitems.

Neben den methodischen Potentialen sind ethische und datenschutzrechtliche Aspekte digitaler Phänotypisierung sorgfältig zu beachten. Sensible Gesundheitsdaten erfordern hohe Standards in Bezug auf informierte Einwilligung, Datensicherheit und Wahrung der ärztlichen Schweigepflicht (Andreoletti et al., 2024; Martinez-Martin et al., 2021). Zudem wird in der Literatur diskutiert, ob eine zu starke Dominanz algorithmischer Verfahren die therapeutische Beziehung

beeinträchtigen könnte, wenn das subjektive Erleben der Betroffenen in den Hintergrund tritt (Martinez-Martin et al., 2021). Digitale Verhaltensmarker sollten daher das klinische Gespräch ergänzen, nicht ersetzen.

4.5 Perspektiven zur Qualifizierung digitaler Verhaltensmarker als Biomarker nach FDA-Standards

Gemäß der US-amerikanischen Food and Drug Administration (FDA) bezeichnet ein Biomarker eine definierte Charakteristik, die als Indikator für normale biologische Prozesse, pathologische Vorgänge oder Reaktionen auf eine Intervention gemessen wird (FDA-NIH Biomarker WorkingGroup, 2025). Zur regulatorischen Anerkennung solcher Marker hat die FDA das Biomarker Qualification Program (BQP) als Teil der Drug Development Tool (DDT) Qualification Programs etabliert. Voraussetzung ist ein klar definierter Context of Use (COU), der den spezifischen Einsatzbereich des Biomarkers festlegt (Biomarker Qualification Program (BQP), 2025).

Der Qualifizierungsprozess erfolgt in drei Schritten: Letter of Intent (LOI), Qualification Plan (QP) und Full Qualification Package (FQP), das die Daten zur analytischen und klinischen Validität enthält (Drug Development Tool (DDT), 2025). Ein Biomarker erhält erst nach einer positiven Bewertung durch die FDA die offizielle Genehmigung für den deklarierten Einsatzbereich.

Für digitale Verhaltensmarker und potenzielle digitale Biomarker, die aus Smartphone- oder Sensordaten abgeleitet werden, gelten dieselben Grundprinzipien. Goldsack et al. (2020) operationalisieren dies in einem dreistufigen Validierungsrahmen aus Verification, Analytical Validation und Clinical Validation. Die folgenden Abschnitte ordnen die in dieser Arbeit eingesetzten Aktivitätsmaße in diesen Rahmen ein und benennen die Schritte, die für eine zukünftige Qualifizierung als digitale Biomarker erforderlich wären.

4.6 Verifikation und analytische Validierung der JTrack-Social-Daten

Die analytische Verifikation prüft, ob ein Messinstrument das beabsichtigte Merkmal technisch korrekt, wiederholbar und unter verschiedenen Bedingungen zuverlässig erfasst (FDA-

NIH Biomarker Working Group, 2025; U.S. Food and Drug Administration (FDA), 2024). Für die JTrack-Social-Plattform sind in diesem Schritt mehrere offene Punkte zu adressieren.

Ein zentrales methodisches Ziel der Weiterentwicklung der App betrifft die plattformübergreifende Vergleichbarkeit zwischen Android und iOS. Die Systeme zur Datenerfassung unterscheiden sich strukturell: Kommunikationsparameter wie die Frequenz von Anrufen oder Textnachrichten dürfen unter iOS aus Sicherheitsgründen nicht von Drittanbieter-Apps aufgezeichnet werden (Bentley et al., 2019); die Kategorisierung installierter Apps ist im Apple App Store taxonomisch geregelt, im Google Play Store hingegen nicht. Auch die Verfügbarkeit passiver Sensordaten und die zugrunde liegende Algorithmik variieren zwischen den Betriebssystemen (Ringwald et al., 2025). Ein anerkanntes Verfahren zur systematischen Harmonisierung dieser Datenquellen existiert bislang nicht (Daniore et al., 2024; Sameh et al., 2024). Eine methodische Vergleichsstudie, die App-Daten gegen unabhängige Systemstatistiken stellt, wäre ein notwendiger Schritt zur analytischen Verifikation der JTrack-Aktivitätsmaße.

Hinzu kommt die Frage der Reproduzierbarkeit der Datenpipeline. Die in dieser Arbeit eingesetzten Schritte – Aktivitätserkennung über die Google Activity Recognition API, Aggregation zu täglichen Summen, Anwendung von Plausibilitätsfiltern – müssen für eine regulatorische Anerkennung vollständig dokumentiert und unabhängig reproduzierbar sein (Coravos et al., 2019; Izmailova et al., 2018). Sameh et al. (2024) identifizieren fehlende Dokumentation und eingeschränkten Zugang zu Quellcodes als zentrale Hindernisse für die Replizierbarkeit von Modellen, die aus Smartphone-Daten abgeleitet werden. Dass passiv erhobene Smartphone-Daten bei sauberer Pipeline klinisch interpretierbare Ergebnisse liefern können, zeigt exemplarisch die Arbeit von Langholm et al. (2023), in der smartphonebasiert erfasste Schlafparameter eng mit selbstberichteten Schlaffragebögen korrelierten ($r = 0{,}83$), sofern eine ausreichende Datendeckung gegeben war.

4.7 Klinische Validierung und terminologische Einordnung

Die klinische Validierung prüft, ob ein Biomarker das intendierte Merkmal in der Zielpopulation zuverlässig erfasst und ob seine Ergebnisse mit etablierten klinischen Referenzverfahren übereinstimmen (Goldsack et al., 2020). Nach dem Konzept der fit-for-purpose validation hängen Maß und Detailgenauigkeit der Validierung vom jeweiligen Einsatzzweck ab; zu prüfen sind insbesondere Genauigkeit, Präzision, Sensitivität, Spezifität und Reproduzierbarkeit

(Izmailova et al., 2018; U.S. Food and Drug Administration (FDA), 2024). Für die in dieser Arbeit erhobenen Aktivitätsmaße wäre die klinische Validierung in einem longitudinalen Design mit wiederholten Fremdratings (z. B. MADRS, BDI oder BPRS) zu adressieren, das den hier nicht möglichen dynamischen Vergleich zwischen Symptomverlauf und Aktivitätsverlauf erlaubt.

Bei der inhaltlichen Einordnung der Befunde ist terminologische Präzision erforderlich. Mulinari, (2024) argumentiert, dass der Begriff Biomarker biologisch fundierten Parametern vorbehalten bleiben sollte, für die ein plausibler pathophysiologischer Zusammenhang mit der Erkrankung besteht. Bei depressiven Störungen kämen hierfür physiologische Messgrößen wie circadiane Aktivitätsmuster, neurovegetative Parameter oder stressbezogene endokrine Veränderungen in Betracht. Rein verhaltensbasierte Maße, wie die in dieser Arbeit erhobenen Aktivitätsminuten, sind unter dieser Konvention präziser als digitale Verhaltensmarker oder digitale Phänotypen zu bezeichnen. Multimodale Ansätze, die Sensorik, Aktivitäts- und Schlafparameter, Sprachanalyse und physiologische Signale integrieren, gelten als besonders aussichtsreich für die Annäherung an einen Biomarker im engeren Sinn (Khoo et al., 2024; McGinnis et al., 2024).

4.8 Regulatorischer Rahmen in Europa: MDR, AI Act, DSGVO

Ein digitaler Verhaltensmarker, der zur Diagnostik oder Verlaufsbeurteilung depressiver Störungen eingesetzt wird, fällt nach der Medizinprodukteverordnung (EU) 2017/745 (MDR) unter die Definition eines Medizinprodukts, da er der Erkennung oder Überwachung einer Krankheit dient. Für den klinischen Einsatz sind eine Konformitätsbewertung und eine CE-Kennzeichnung erforderlich, in der Regel unter Einbeziehung einer benannten Stelle. Gemäß der aktualisierten MDCG-Leitlinie 2019-11 Rev. 1, 2025) ist Software, die diagnostische oder therapeutische Entscheidungen unterstützt, als Medical Device Software (MDSW) zu klassifizieren und wird abhängig vom Anwendungskontext häufig mindestens der Risikoklasse IIa zugeordnet.

Der EU-AI-Act (Regulation (EU) 2024/1689) ergänzt diesen Rahmen für Anwendungen mit KI-Komponenten um Anforderungen an Transparenz, Risikomanagement, algorithmische Nachvollziehbarkeit und menschliche Aufsicht. Während die MDR-Anforderungen an Sicherheit und klinische Leistungsbewertung definiert, adressiert der AI-Act die spezifischen Risiken algorithmischer Verfahren. Zusammen mit der Datenschutz-Grundverordnung (DSGVO)

bilden beide Regelwerke den europäischen Rahmen für die zukünftige klinische Nutzung digitaler Verhaltensmarker. Colloud et al. (2023) und Tackney et al. (2024) heben in diesem Zusammenhang die Bedeutung eines frühzeitigen Dialogs mit Regulierungsbehörden hervor: Die analytische und klinische Validierung sollte nicht erst nachträglich an einen bestehenden Marker angepasst werden, sondern bereits in das Studiendesign eingehen.

5 Schlussfolgerung

Die im Rahmen dieser Arbeit durchgeführte explorative Studie zielte darauf ab, das Aktivitätsverhalten von Patient*innen mit depressiven Störungen anhand automatisch erfasster Smartphonedaten aktiv und passiv zu erfassen. Die App JTrack Social diente dazu, über einen Zeitraum von 90 Tagen alltagsnahe Aktivitätsmuster zu erfassen und diese systematisch mit wöchentlichen Selbstberichten (EMA) sowie dem initialen klinischen Rating (BPRS) zu verknüpfen.

Es wurden tendenzielle, nach FDR-Korrektur jedoch keine signifikanten Assoziationen zwischen Aktivitätsmaßen und psychischen Zustandsmaßen beobachtet. Diese explorativen Befunde unterstreichen die Bedeutung langfristiger digitaler Datenerhebungen in der psychiatrischen Forschung und verdeutlichen, dass größere Stichproben und längere Beobachtungszeiträume erforderlich sind, um robuste Aussagen zu ermöglichen.

Neben den empirischen Ergebnissen leistet diese Arbeit einen methodischen Beitrag zur Entwicklung und Validierung digitaler Biomarker in der Psychiatrie. Sie zeigt, dass Smartphone- und Sensordaten, die passiv erhoben werden, klinisch relevante Verhaltensmuster abbilden können – vorausgesetzt, sie werden nach klar definierten Qualitäts- und Datenschutzstandards verarbeitet. Die in dieser Dissertation beschriebenen Verfahren zur Erhebung, Kategorisierung und Auswertung der Daten bilden eine Grundlage für künftige Studien, die digitale Verhaltensparameter systematisch in klinische Forschung integrieren möchten.

Darüber hinaus berücksichtigt die Arbeit regulatorische und ethische Anforderungen, die für die sichere Nutzung digitaler Biomarker in der klinischen Versorgung unverzichtbar sind. Zusammenfassend zeigt diese Dissertation, dass digitale Messverfahren das Potenzial haben, Diagnostik, Verlaufsbeurteilung und Therapieevaluation bei depressiven Störungen zu verbessern. Die Kombination aus kontinuierlicher Datenerhebung, methodischer Standardisierung

und regulatorischer Absicherung eröffnet neue Perspektiven für eine präzisere, objektivere und personenzentrierte Psychiatrie.

Literaturverzeichnis

Andreoletti, M., Haller, L., Vayena, E., & Blasimme, A. (2024). Mapping the ethical landscape of digital biomarkers: A scoping review. *PLOS Digital Health*, *3*(5), e0000519. https://doi.org/10.1371/journal.pdig.0000519

Armey, M. F., Schatten, H. T., Haradhvala, N., & Miller, I. W. (2015). Ecological momentary assessment (EMA) of depression-related phenomena. *Current Opinion in Psychology*, *4*, 21–25. https://doi.org/10.1016/j.copsyc.2015.01.002

Arnone, D., Karmegam, S. R., Östlundh, L., Alkhyeli, F., Alhammadi, L., Alhammadi, S., Alkhoori, A., & Selvaraj, S. (2024). Risk of suicidal behavior in patients with major depression and bipolar disorder – A systematic review and meta-analysis of registry-based studies. *Neuroscience & Biobehavioral Reviews*, *159*, 105594. https://doi.org/10.1016/j.neubiorev.2024.105594

Bajraktarov, S., Blazhevska Stoilkovska, B., Russo, M., Repišti, S., Maric, N. P., Dzubur Kulenovic, A., Arënliu, A., Stevovic, L. I., Novotni, L., Ribic, E., Konjufca, J., Ristic, I., Novotni, A., & Jovanovic, N. (2023). Factor structure of the brief psychiatric rating scale-expanded among outpatients with psychotic disorders in five Southeast European countries: Evidence for five factors. *Frontiers in Psychiatry*, *14*, 1207577. https://doi.org/10.3389/fpsyt.2023.1207577

Bener, A., Saldı, M. H., Bhugra, D., & Ventriglio, A. (2025). Factors contributing to treatment noncompliance in a large sample of psychiatric patients in Turkey. *Asian Journal of Psychiatry*, *109*, 104528. https://doi.org/10.1016/j.ajp.2025.104528

Bentley, K. H., Kleiman, E. M., Elliott, G., Huffman, J. C., & Nock, M. K. (2019). Real-time monitoring technology in single-case experimental design research: Opportunities and challenges. *Behaviour Research and Therapy*, *117*, 87–96. https://doi.org/10.1016/j.brat.2018.11.017

Biancosino, B., Picardi, A., Marmai, L., Biondi, M., & Grassi, L. (2010). Factor structure of the Brief Psychiatric Rating Scale in unipolar depression. *Journal of Affective Disorders*, *124*(3), 329–334. https://doi.org/10.1016/j.jad.2009.11.019

Biomarker Qualification Program (BQP), C. for D. E. and. (2025, Juni 17). *Biomarker Qualification Program*. FDA. FDA. https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/biomarker-qualification-program

Brady, R., Brown, W. J., & Mielke, G. I. (2023). Day-to-day variability in accelerometer-measured physical activity in mid-aged Australian adults. *BMC Public Health*, *23*(1), 1880. https://doi.org/10.1186/s12889-023-16734-0

Browning, C. R., Pinchak, N. P., Calder, C. A., & Boettner, B. (2024). Leveraging Experience Sampling/Ecological Momentary Assessment for Sociological Investigations of Everyday Life. *Annual Review of Sociology*, *50*(1), 41–59. https://doi.org/10.1146/annurev-soc-091523-013249

Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, *25*, e46778. https://doi.org/10.2196/46778

Colloud, S., Metcalfe, T., Askin, S., Belachew, S., Ammann, J., Bos, E., Kilchenmann, T., Strijbos, P., Eggenspieler, D., Servais, L., Garay, C., Konstantakopoulos, A., Ritzhaupt, A., Vetter, T., Vincenzi, C., & Cerreta, F. (2023). Evolving regulatory perspectives on digital health technologies

for medicinal product development. *Npj Digital Medicine*, *6*(1), 56. https://doi.org/10.1038/s41746-023-00790-2

Colombo, D., Fernández-Álvarez, J., Patané, A., Semonella, M., Kwiatkowska, M., García-Palacios, A., Cipresso, P., Riva, G., & Botella, C. (2019). Current State and Future Directions of Technology-Based Ecological Momentary Assessment and Intervention for Major Depressive Disorder: A Systematic Review. *Journal of Clinical Medicine*, *8*(4), 465. https://doi.org/10.3390/jcm8040465

Cooney, G. M., Dwan, K., Greig, C. A., Lawlor, D. A., Rimer, J., Waugh, F. R., McMurdo, M., & Mead, G. E. (2013). Exercise for depression. *Cochrane Database of Systematic Reviews*, *2013*(9). https://doi.org/10.1002/14651858.CD004366.pub6

Coravos, A., Khozin, S., & Mandl, K. D. (2019). Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *Npj Digital Medicine*, *2*(1), 14. https://doi.org/10.1038/s41746-019-0090-4

Corneille, O., & Gawronski, B. (2024). Self-reports are better measurement instruments than implicit measures. *Nature Reviews Psychology*, *3*(12), 835–846. https://doi.org/10.1038/s44159-024-00376-z

Daniore, P., Nittas, V., Haag, C., Bernard, J., Gonzenbach, R., & Von Wyl, V. (2024). From wearable sensor data to digital biomarker development: Ten lessons learned and a framework proposal. *Npj Digital Medicine*, *7*(1), 161. https://doi.org/10.1038/s41746-024-01151-3

Dishman, R. K., McDowell, C. P., & Herring, M. P. (2021). Customary physical activity and odds of depression: A systematic review and meta-analysis of 111 prospective cohort studies. *British Journal of Sports Medicine*, *55*(16), 926–934. https://doi.org/10.1136/bjsports-2020-103140

Dobson, K. G., Gignac, M. A. M., & Mustard, C. A. (2024). The working life expectancy of American adults experiencing depression. *Social Psychiatry and Psychiatric Epidemiology*, *59*(6), 1013–1027. https://doi.org/10.1007/s00127-023-02547-4

Drug Development Tool (DDT), C. for D. E. and. (2025, Juli 21). *Drug Development Tool (DDT) Qualification Programs*. FDA. FDA. https://www.fda.gov/drugs/development-approval-process-drugs/drug-development-tool-ddt-qualification-programs

Dukart, J., Lotter, L. D., & Eickhoff, S. B. (2024). Moving towards precision psychiatry: The hard nut of depression. *Signal Transduction and Targeted Therapy*, *9*(1), 310. https://doi.org/10.1038/s41392-024-02023-8

Dunlop, B. W., McCabe, B., Eudicone, J. M., Sheehan, J. J., & Baker, R. A. (2014). How well do clinicians and patients agree on depression treatment outcomes? Implications for personalized medicine. *Human Psychopharmacology: Clinical and Experimental*, *29*(6), 528–536. https://doi.org/10.1002/hup.2428

Eaton, W. W., Shao, H., Nestadt, G., Lee, B. H., Bienvenu, O. J., & Zandi, P. (2008). Population-Based Study of First Onset and Chronicity in Major Depressive Disorder. *Archives of General Psychiatry*, *65*(5), 513. https://doi.org/10.1001/archpsyc.65.5.513

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34333-9

FDA-NIH Biomarker WorkingGroup, F.-N. B. W. (2025). FDA-NIH Biomarker Working Group. In *BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]*. Food and Drug Administration (US). https://www.ncbi.nlm.nih.gov/books/NBK338449/

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J. L., Vos, T., & Whiteford, H. A. (2013). Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. *PLoS Medicine*, *10*(11), e1001547. https://doi.org/10.1371/journal.pmed.1001547

GBD 2019 Mental Disorders Collaborators, 2022. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*, *9*(2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3

Goldsack, J. C., Coravos, A., Bakker, J. P., Bent, B., Dowling, A. V., Fitzer-Attas, C., Godfrey, A., Godino, J. G., Gujar, N., Izmailova, E., Manta, C., Peterson, B., Vandendriessche, B., Wood, W. A., Wang, K. W., & Dunn, J. (2020). Verification, analytical validation, and clinical validation (V3): The foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *Npj Digital Medicine*, *3*(1), 55. https://doi.org/10.1038/s41746-020-0260-4

Gronemann, F. H., Lund, T., Lindholdt, L., Madsen, K. B., Jørgensen, M. B., Nordentoft, M., & Osler, M. (2022). Treatment-resistant depression and labor market affiliation in the Danish welfare society: A register-based study. *Social Psychiatry and Psychiatric Epidemiology*, *57*(6), 1189–1199. https://doi.org/10.1007/s00127-022-02243-9

Gustavsson, A., Svensson, M., Jacobi, F., Allgulander, C., Alonso, J., Beghi, E., Dodel, R., Ekman, M., Faravelli, C., Fratiglioni, L., Gannon, B., Jones, D. H., Jennum, P., Jordanova, A., Jönsson, L., Karampampa, K., Knapp, M., Kobelt, G., Kurth, T., … Olesen, J. (2011). Cost of disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(10), 718–779. https://doi.org/10.1016/j.euroneuro.2011.08.008

Hahn, L., Eickhoff, S. B., Habel, U., Stickeler, E., Schnakenberg, P., Goecke, T. W., Stickel, S., Franz, M., Dukart, J., & Chechko, N. (2021). Early identification of postpartum depression using demographic, clinical, and digital phenotyping. *Translational Psychiatry*, *11*(1), 121. https://doi.org/10.1038/s41398-021-01245-6

Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A., & Beekman, A. T. F. (2013). Recurrence of major depressive disorder and its predictors in the general population: Results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Psychological Medicine*, *43*(1), 39–48. https://doi.org/10.1017/S0033291712002395

Hershenberg, R., McDonald, W. M., Crowell, A., Riva-Posse, P., Craighead, W. E., Mayberg, H. S., & Dunlop, B. W. (2020). Concordance between clinician-rated and patient reported outcome measures of depressive symptoms in treatment resistant depression. *Journal of Affective Disorders*, *266*, 22–29. https://doi.org/10.1016/j.jad.2020.01.108

Hoffman, B. M., Babyak, M. A., Craighead, W. E., Sherwood, A., Murali Doraiswamy, P., Coons, M. J., & Blumenthal, J. A. (2011). Exercise and Pharmacotherapy in Patients With Major Depression: One-Year Follow-Up of the SMILE Study. *Psychosomatic Medicine*, *73*(2), 127–133. https://doi.org/10.1097/PSY.0b013e31820433a5

Izmailova, E. S., Wagner, J. A., & Perakslis, E. D. (2018). Wearable Devices in Clinical Trials: Hype and Hypothesis. *Clinical Pharmacology & Therapeutics*, *104*(1), 42–52. https://doi.org/10.1002/cpt.966

Jacobi, F., Höfler, M., Siegert, J., Mack, S., Gerschler, A., Scholl, L., Busch, M. A., Hapke, U., Maske, U., Seiffert, I., Gaebel, W., Maier, W., Wagner, M., Zielasek, J., & Wittchen, H. (2014). Twelve-month prevalence, comorbidity and correlates of mental disorders in Germany: The Mental Health Module of the German Health Interview and Examination Survey for Adults (DEGS1-MH). *International Journal of Methods in Psychiatric Research*, *23*(3), 304–319. https://doi.org/10.1002/mpr.1439

Jia, H., & Lubetkin, E. I. (2017). Incremental decreases in quality-adjusted life years (QALY) associated with higher levels of depressive symptoms for U.S. Adults aged 65 years and older. *Health and Quality of Life Outcomes*, *15*(1), 9. https://doi.org/10.1186/s12955-016-0582-8

Jia, H., Zack, M. M., Thompson, W. W., Crosby, A. E., & Gottesman, I. I. (2015). Impact of depression on quality-adjusted life expectancy (QALE) directly as well as indirectly through suicide. *Social Psychiatry and Psychiatric Epidemiology*, *50*(6), 939–949. https://doi.org/10.1007/s00127-015-1019-0

Kales, H. C., Kim, H. M., Austin, K. L., & Valenstein, M. (2010). Who Receives Outpatient Monitoring During High-Risk Depression Treatment Periods? *Journal of the American Geriatrics Society*, *58*(5), 908–913. https://doi.org/10.1111/j.1532-5415.2010.02810.x

Kathan, A., Harrer, M., Küster, L., Triantafyllopoulos, A., He, X., Milling, M., Gerczuk, M., Yan, T., Rajamani, S. T., Heber, E., Grossmann, I., Ebert, D. D., & Schuller, B. W. (2022). Personalised depression forecasting using mobile sensor data and ecological momentary assessment. *Frontiers in Digital Health*, *4*, 964582. https://doi.org/10.3389/fdgth.2022.964582

Khoo, L. S., Lim, M. K., Chong, C. Y., & McNaney, R. (2024). Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches. *Sensors*, *24*(2), 348. https://doi.org/10.3390/s24020348

Koopmans, P. C., Roelen, C. A. M., & Groothoff, J. W. (2008). Sickness absence due to depressive symptoms. *International Archives of Occupational and Environmental Health*, *81*(6), 711–719. https://doi.org/10.1007/s00420-007-0243-7

Lahnakoski, J. M., Eickhoff, S. B., Dukart, J., & Schilbach, L. (2022). Naturalizing psychopathology—Towards a quantitative real-world psychiatry. *Molecular Psychiatry*, *27*(2), 781–783. https://doi.org/10.1038/s41380-021-01322-8

Leaning, I. E., Ikani, N., Savage, H. S., Leow, A., Beckmann, C., Ruhé, H. G., & Marquand, A. F. (2024). From smartphone data to clinically relevant predictions: A systematic review of digital phenotyping methods in depression. *Neuroscience & Biobehavioral Reviews*, *158*, 105541. https://doi.org/10.1016/j.neubiorev.2024.105541

Li, X., Mu, F., Liu, D., Zhu, J., Yue, S., Liu, M., Liu, Y., & Wang, J. (2022). Predictors of suicidal ideation, suicide attempt and suicide death among people with major depressive disorder: A systematic review and meta-analysis of cohort studies. *Journal of Affective Disorders*, *302*, 332–351. https://doi.org/10.1016/j.jad.2022.01.103

Li, X., Zhou, J., Wang, M., Yang, C., & Sun, G. (2023). Cardiovascular disease and depression: A narrative review. *Frontiers in Cardiovascular Medicine*, *10*, 1274595. https://doi.org/10.3389/fcvm.2023.1274595

Loh, S. Y., Abdullah, A., Abu Bakar, A. K., Thambu, M., & Nik Jaafar, N. R. (2015). Structured Walking and Chronic Institutionalized Schizophrenia Inmates: A pilot RCT Study on Quality of Life. *Global Journal of Health Science*, *8*(1), 238. https://doi.org/10.5539/gjhs.v8n1p238

Ma, J., Lin, P., & Williams, J. (2024). Effectiveness of nature-based walking interventions in improving mental health in adults: A systematic review. *Current Psychology*, *43*(11), 9521–9539. https://doi.org/10.1007/s12144-023-05112-z

Maatoug, R., Peiffer-Smadja, N., Delval, G., Brochu, T., Pitrat, B., & Millet, B. (2021). Ecological Momentary Assessment Using Smartphones in Patients With Depression: Feasibility Study. *JMIR Formative Research*, *5*(2), e14179. https://doi.org/10.2196/14179

Malhi, G. S., & Mann, J. J. (2018). Depression. *The Lancet*, *392*(10161), 2299–2312. https://doi.org/10.1016/S0140-6736(18)31948-2

Martinez-Martin, N., Greely, H. T., & Cho, M. K. (2021). Ethical Development of Digital Phenotyping Tools for Mental Health Applications: Delphi Study. *JMIR mHealth and uHealth*, *9*(7), e27343. https://doi.org/10.2196/27343

McGinnis, E. W., Cherian, J., & McGinnis, R. S. (2024). The State of Digital Biomarkers in Mental Health. *Digital Biomarkers*, *8*(1), 210–217. https://doi.org/10.1159/000542320

McLaughlin, K. A. (2011). The Public Health Impact of Major Depression: A Call for Interdisciplinary Prevention Efforts. *Prevention Science*, *12*(4), 361–371. https://doi.org/10.1007/s11121-011-0231-8

MDCG-Leitlinie 2019-11 Rev. 1. (2025). *Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR*. https://health.ec.europa.eu/latest-updates/update-mdcg-2019-11-rev1-qualification-and-classification-software-regulation-eu-2017745-and-2025-06-17_en?utm_source=chatgpt.com

Meling, H. M., Ruths, S., Baste, V., Hensing, G., & Haukenes, I. (2023). Level of education and sustainable return to work among long-term sick-listed workers with depression: A register-based cohort study (The Norwegian GP-DEP Study). *BMJ Open*, *13*(7), e072051. https://doi.org/10.1136/bmjopen-2023-072051

Mobile Operating System Market Share Germany. (o. J.). StatCounter Global Stats. Abgerufen 8. November 2025, von https://gs.statcounter.com/os-market-share/mobile/germany

Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., Mohr, D. C., & Pulkki-Råback, L. (2021). Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. *Frontiers in Psychiatry*, *12*, 625247. https://doi.org/10.3389/fpsyt.2021.625247

Mulinari, S. (2024). Aligning digital biomarker definitions in psychiatry with the National Institute of Mental Health Research Domain Criteria framework. *NPP—Digital Psychiatry and Neuroscience*, *2*(1), 15. https://doi.org/10.1038/s44277-024-00017-6

Nationale Versorgungsleitlinie Unipolare Depression 2023. (o. J.).

Newton-Howes, G., Tyrer, P., & Johnson, T. (2006). Personality disorder and the outcome of depression: Meta-analysis of published studies. *British Journal of Psychiatry*, *188*(1), 13–20. https://doi.org/10.1192/bjp.188.1.13

Noetel, M., Sanders, T., Gallardo-Gómez, D., Taylor, P., Del Pozo Cruz, B., Van Den Hoek, D., Smith, J. J., Mahoney, J., Spathis, J., Moresi, M., Pagano, R., Pagano, L., Vasconcellos, R., Arnott, H., Varley, B., Parker, P., Biddle, S., & Lonsdale, C. (2024). Effect of exercise for depression: Systematic review and network meta-analysis of randomised controlled trials. *BMJ*, *384*, e075847. https://doi.org/10.1136/bmj-2023-075847

Oliva, H. N. P., Monteiro-Junior, R. S., Oliva, I. O., & Powers, A. R. (2023). Effects of exercise intervention on psychotic symptoms: A meta-analysis and hypothetical model of neurobiological mechanisms. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *125*, 110771. https://doi.org/10.1016/j.pnpbp.2023.110771

Onnela, J.-P., & Rauch, S. L. (2016). Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology*, *41*(7), 1691–1696. https://doi.org/10.1038/npp.2016.7

Oudin, A., Maatoug, R., Bourla, A., Ferreri, F., Bonnot, O., Millet, B., Schoeller, F., Mouchabac, S., & Adrien, V. (2023). Digital Phenotyping: Data-Driven Psychiatry to Redefine Mental Health. *Journal of Medical Internet Research*, *25*, e44502. https://doi.org/10.2196/44502

Park, S., Jang, E. Y., Kim, D., Jun, T., Lee, M., Kim, J., Kim, J., Jo, S., & Park, Y. C. (2015). Dimensional approach to symptom factors of major depressive disorder in Koreans, using the Brief Psychiatric Rating Scale: The Clinical Research Center for Depression of South Korea Study. *The Kaohsiung Journal of Medical Sciences*, *31*(1), 47–54. https://doi.org/10.1016/j.kjms.2014.11.001

Park, Y., An, S. Y., Jeon, J. Y., & Lee, D. H. (2025). Occupational physical activity and depression risk: A systematic review and dose-response meta-analysis. *Journal of Affective Disorders*, *391*, 119983. https://doi.org/10.1016/j.jad.2025.119983

Pedersen, J., Thorsen, S. V., Andersen, M. F., Hanvold, T. N., Schlünssen, V., & Bültmann, U. (2019). Impact of depressive symptoms on worklife expectancy: A longitudinal study on Danish employees. *Occupational and Environmental Medicine*, *76*(11), 838–844. https://doi.org/10.1136/oemed-2019-105961

Plana-Ripoll, O., Pedersen, C. B., Holtz, Y., Benros, M. E., Dalsgaard, S., De Jonge, P., Fan, C. C., Degenhardt, L., Ganna, A., Greve, A. N., Gunn, J., Iburg, K. M., Kessing, L. V., Lee, B. K., Lim, C. C. W., Mors, O., Nordentoft, M., Prior, A., Roest, A. M., … McGrath, J. J. (2019). Exploring Comorbidity Within Mental Disorders Among a Danish National Population. *JAMA Psychiatry*, *76*(3), 259. https://doi.org/10.1001/jamapsychiatry.2018.3658

Polhemus, A., Simblett, S., Dawe-Lane, E., Gilpin, G., Elliott, B., Jilka, S., Novak, J., Nica, R. I., Temesi, G., & Wykes, T. (2022). Health Tracking via Mobile Apps for Depression Self-management: Qualitative Content Analysis of User Reviews. *JMIR Human Factors*, *9*(4), e40133. https://doi.org/10.2196/40133

Porras-Segovia, A., Molina-Madueño, R. M., Berrouiguet, S., López-Castroman, J., Barrigón, M. L., Pérez-Rodríguez, M. S., Marco, J. H., Díaz-Oliván, I., De León, S., Courtet, P., Artés-Rodríguez, A., & Baca-García, E. (2020). Smartphone-based ecological momentary assessment (EMA) in

psychiatric patients and student controls: A real-world feasibility study. *Journal of Affective Disorders*, *274*, 733–741. https://doi.org/10.1016/j.jad.2020.05.067

Price, G. D., Heinz, M. V., Song, S. H., Nemesure, M. D., & Jacobson, N. C. (2023). Using digital phenotyping to capture depression symptom variability: Detecting naturalistic variability in depression symptoms across one year using passively collected wearable movement and sleep data. *Translational Psychiatry*, *13*(1), 381. https://doi.org/10.1038/s41398-023-02669-y

Rai, D., Culpin, I., Heuvelman, H., Magnusson, C. M. K., Carpenter, P., Jones, H. J., Emond, A. M., Zammit, S., Golding, J., & Pearson, R. M. (2018). Association of Autistic Traits With Depression From Childhood to Age 18 Years. *JAMA Psychiatry*, *75*(8), 835. https://doi.org/10.1001/jamapsychiatry.2018.1323

Riera-Serra, P., Navarra-Ventura, G., Castro, A., Gili, M., Salazar-Cedillo, A., Ricci-Cabello, I., Roldán-Espínola, L., Coronado-Simsic, V., García-Toro, M., Gómez-Juanes, R., & Roca, M. (2024). Clinical predictors of suicidal ideation, suicide attempts and suicide death in depressive disorder: A systematic review and meta-analysis. *European Archives of Psychiatry and Clinical Neuroscience*, *274*(7), 1543–1563. https://doi.org/10.1007/s00406-023-01716-5

Ringwald, W. R., King, G., Vize, C. E., & Wright, A. G. C. (2025). Passive Smartphone Sensors for Detecting Psychopathology. *JAMA Network Open*, *8*(7), e2519047. https://doi.org/10.1001/jamanetworkopen.2025.19047

Rose, U., Kersten, N., Pattloch, D., Conway, P. M., & Burr, H. (2023). Associations between depressive symptoms and 5-year subsequent work nonparticipation due to long-term sickness absence, unemployment and early retirement in a cohort of 2,413 employees in Germany. *BMC Public Health*, *23*(1), 2159. https://doi.org/10.1186/s12889-023-17090-9

Rykov, Y., Thach, T.-Q., Bojic, I., Christopoulos, G., & Car, J. (2021). Digital Biomarkers for Depression Screening With Wearable Devices: Cross-sectional Study With Machine Learning Modeling. *JMIR mHealth and uHealth*, *9*(10), e24872. https://doi.org/10.2196/24872

Sahandi Far, M., Stolz, M., Fischer, J. M., Eickhoff, S. B., & Dukart, J. (2021). JTrack: A Digital Biomarker Platform for Remote Monitoring of Daily-Life Behaviour in Health and Disease. *Frontiers in Public Health*, *9*, 763621. https://doi.org/10.3389/fpubh.2021.763621

Sameh, A., Rostami, M., Oussalah, M., Korpelainen, R., & Farrahi, V. (2024). Digital phenotypes and digital biomarkers for health and diseases: A systematic review of machine learning approaches utilizing passive non-invasive signals collected via wearable devices and smartphones. *Artificial Intelligence Review*, *58*(2), 66. https://doi.org/10.1007/s10462-024-11009-5

Schuch, F. B., Vancampfort, D., Firth, J., Rosenbaum, S., Ward, P. B., Silva, E. S., Hallgren, M., Ponce De Leon, A., Dunn, A. L., Deslandes, A. C., Fleck, M. P., Carvalho, A. F., & Stubbs, B. (2018). Physical Activity and Incident Depression: A Meta-Analysis of Prospective Cohort Studies. *American Journal of Psychiatry*, *175*(7), 631–648. https://doi.org/10.1176/appi.ajp.2018.17111194

Sheikh, M., Qassem, M., & Kyriacou, P. A. (2021). Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring. *Frontiers in Digital Health*, *3*, 662811. https://doi.org/10.3389/fdgth.2021.662811

Singh, B., Olds, T., Curtis, R., Dumuid, D., Virgara, R., Watson, A., Szeto, K., O'Connor, E., Ferguson, T., Eglitis, E., Miatke, A., Simpson, C. E., & Maher, C. (2023a). Effectiveness of physical

activity interventions for improving depression, anxiety and distress: An overview of systematic reviews. *British Journal of Sports Medicine*, *57*(18), 1203–1209. https://doi.org/10.1136/bjsports-2022-106195

Singh, B., Olds, T., Curtis, R., Dumuid, D., Virgara, R., Watson, A., Szeto, K., O'Connor, E., Ferguson, T., Eglitis, E., Miatke, A., Simpson, C. E., & Maher, C. (2023b). Effectiveness of physical activity interventions for improving depression, anxiety and distress: An overview of systematic reviews. *British Journal of Sports Medicine*, *57*(18), 1203–1209. https://doi.org/10.1136/bjsports-2022-106195

Stutaite, G., Mittendorfer-Rutz, E., Helgesson, M., Kautzky, A., Finnes, A., & Gémes, K. (2025). Working life patterns after sickness absence due to depression: A 15-year register-based prospective cohort study. *Journal of Affective Disorders*, *379*, 822–834. https://doi.org/10.1016/j.jad.2025.03.073

Tackney, M. S., Steele, A., Newman, J., Fritzsche, M.-C., Lucivero, F., Khadjesari, Z., Lynch, J., Abbott, R. A., Barber, V. S., Carpenter, J. R., Copsey, B., Davies, E. H., Dixon, W. G., Fox, L., González, J., Griffiths, J., Hinchliffe, C. H. L., Kolanko, M. A., McGagh, D., … Villar, S. S. (2024). Digital endpoints in clinical trials: Emerging themes from a multi-stakeholder Knowledge Exchange event. *Trials*, *25*(1), 521. https://doi.org/10.1186/s13063-024-08356-7

Targum, S. D., Sauder, C., Evans, M., Saber, J. N., & Harvey, P. D. (2021). Ecological momentary assessment as a measurement tool in depression trials. *Journal of Psychiatric Research*, *136*, 256–264. https://doi.org/10.1016/j.jpsychires.2021.02.012

Terhorst, Y., Knauer, J., Philippi, P., & Baumeister, H. (2024). The Relation Between Passively Collected GPS Mobility Metrics and Depressive Symptoms: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, *26*, e51875. https://doi.org/10.2196/51875

Torous, J., Linardon, J., Goldberg, S. B., Sun, S., Bell, I., Nicholas, J., Hassan, L., Hua, Y., Milton, A., & Firth, J. (2025). The evolving field of digital mental health: Current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry*, *24*(2), 156–174. https://doi.org/10.1002/wps.21299

Torous, J., Onnela, J.-P., & Keshavan, M. (2017). New dimensions and new tools to realize the potential of RDoC: Digital phenotyping via smartphones and connected devices. *Translational Psychiatry*, *7*(3), e1053–e1053. https://doi.org/10.1038/tp.2017.25

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., & Mcdowell, M. (2008). Physical Activity in the United States Measured by Accelerometer. *Medicine & Science in Sports & Exercise*, *40*(1), 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3

Uher, R., Perlis, R. H., Placentino, A., Dernovšek, M. Z., Henigsberg, N., Mors, O., Maier, W., McGuffin, P., & Farmer, A. (2012). SELF-REPORT AND CLINICIAN-RATED MEASURES OF DEPRESSION SEVERITY: CAN ONE REPLACE THE OTHER?: Research Article: Interchangeablility of Self-Report and Clinician-Rated Measures of Depression Severity. *Depression and Anxiety*, *29*(12), 1043–1049. https://doi.org/10.1002/da.21993

U.S. Food and Drug Administration (FDA), C. for D. E. and. (2024, Juli 29). *Digital Health Technologies for Remote Data Acquisition in Clinical Investigations*. FDA. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations

Van Der Sluys, M. E., Zijlmans, J., Ket, J. C. F., Marhe, R., Popma, A., Scherder, E. J. A., & Van Der Laan, P. H. (2024). The efficacy of physical activity interventions in reducing antisocial behavior: A meta-analytic review. *Journal of Experimental Criminology*, *20*(2), 347–373. https://doi.org/10.1007/s11292-022-09536-8

Van Genugten, C. R., Schuurmans, J., Van Ballegooijen, W., Hoogendoorn, A. W., Smit, J. H., & Riper, H. (2021). Discovering different profiles in the dynamics of depression based on real–time monitoring of mood: A first exploration. *Internet Interventions*, *26*, 100437. https://doi.org/10.1016/j.invent.2021.100437

Vignapiano, A., Monaco, F., Pagano, C., Piacente, M., Farina, F., Petrillo, G., Sica, R., Marenna, A., Shin, J. I., Solmi, M., & Corrivetti, G. (2023). A narrative review of digital biomarkers in the management of major depressive disorder and treatment-resistant forms. *Frontiers in Psychiatry*, *14*, 1321345. https://doi.org/10.3389/fpsyt.2023.1321345

Wichers, M., Peeters, F., Geschwind, N., Jacobs, N., Simons, C. J. P., Derom, C., Thiery, E., Delespaul, P. H., & Van Os, J. (2010). Unveiling patterns of affective responses in daily life may improve outcome prediction in depression: A momentary assessment study. *Journal of Affective Disorders*, *124*(1–2), 191–195. https://doi.org/10.1016/j.jad.2009.11.010

Wiegand, H. F., Saam, J., Marschall, U., Chmitorz, A., Kriston, L., Berger, M., Lieb, K., & Hölzel, L. P. (2020). Challenges in the Transition from In-Patient to Out-Patient Treatment in Depression. *Deutsches Ärzteblatt International*. https://doi.org/10.3238/arztebl.2020.0472

Willetts, M., Hollowell, S., Aslett, L., Holmes, C., & Doherty, A. (2018). Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. *Scientific Reports*, *8*(1), 7961. https://doi.org/10.1038/s41598-018-26174-1

World Health Organization. (1992). *International Statistical Classification of Diseases and Related Health Problems (ICD-10)*. https://icd.who.int/browse10/2019/en#/F32.2 (Ursprünglich erschienen World Health Organization)

Wüthrich, F., Nabb, C. B., Mittal, V. A., Shankman, S. A., & Walther, S. (2022). Actigraphically measured psychomotor slowing in depression: Systematic review and meta-analysis. *Psychological Medicine*, *52*(7), 1208–1221. https://doi.org/10.1017/S0033291722000903

Zanello, A., Berthoud, L., Ventura, J., & Merlo, M. C. G. (2013a). The Brief Psychiatric Rating Scale (version 4.0) factorial structure and its sensitivity in the treatment of outpatients with unipolar depression. *Psychiatry Research*, *210*(2), 626–633. https://doi.org/10.1016/j.psychres.2013.07.001

Zanello, A., Berthoud, L., Ventura, J., & Merlo, M. C. G. (2013b). The Brief Psychiatric Rating Scale (version 4.0) factorial structure and its sensitivity in the treatment of outpatients with unipolar depression. *Psychiatry Research*, *210*(2), 626–633. https://doi.org/10.1016/j.psychres.2013.07.001

Zhan, Y., Liu, H., & Wang, Y. (2026). Digital phenotyping of depression: A multi-modal passive sensing approach to identifying novel behavioral and physiological markers of treatment response. *Journal of Psychiatric Research*, *194*, 40–50. https://doi.org/10.1016/j.jpsychires.2025.12.036

Zierer, C., Behrendt, C., & Lepach-Engelhardt, A. C. (2024). Digital biomarkers in depression: A systematic review and call for standardization and harmonization of feature engineering. *Journal of Affective Disorders*, *356*, 438–449. https://doi.org/10.1016/j.jad.2024.03.163

6 Anhang

6.1 Demographischer Fragebogen

JTrack

Datum:
Fallnummer:

Vom Versuchsleiter auszufüllen

Hauptdiagnose:

ICD-Code:

Tage seit Aufnahme:

CGI-S-MV:
GAF:

Rekrutierungsstation: ◯ Geschützt ◯ Offen ◯ Tagesklinisch ◯ Ambulant

Wie alt sind Sie?	
Welches Geschlecht haben Sie?	◯ männlich ◯ weiblich ◯ divers
Welchen Familienstand haben Sie?	◯ Ledig ◯ Verheiratet

	◯ Geschieden ◯ In fester Partnerschaft
Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?	◯ Keinen Schulabschluss ◯ Volksschule/Hauptschule ◯ Mittlere Reife ◯ fachgebundene oder allgemeine Hochschulreife
Sind Sie zur Zeit erwerbstätig?	◯ Nein, ich bin nicht erwerbstätig ◯ Ja, ich bin Student/in oder Auszubildende/r ◯ Ja, ich bin Akademiker(in) in freiem Beruf (z.B. Arzt/Ärztin, Rechtsanwalt/-anwältin, Steuerberater/in) ◯ Ja, ich bin selbständig im Handel, Gewerbe, Handwerk, Industrie, Dienstleistung ◯ Ja, ich bin Beamter/Beamtin ◯ Ja, ich bin Angestellte/r ◯ Ja, ich bin Arbeiter/in ◯ Ja, ich arbeite in einer Werkstatt für Menschen mit Behinderung
Welchen zeitlichen Umfang hat Ihre Tätigkeit?	◯ Vollzeit-erwerbstätig (35h pro Woche oder mehr) ◯ Teilzeit-erwerbstätig (15h bis 34h pro Woche) ◯ Stundenweise erwerbstätig (bis 15h pro Woche)
Haben Sie weitere psychische Erkrankungen?	◯ Ja, und zwar ◯ Nein

Wie viele Klinikaufenthalt hatten Sie bereits auf Grund Ihrer Diagnose?	
Wann kam es zum ersten Mal zum Ausbruch Ihrer Erkrankung?	Im Jahr

Tabelle 11: Der nachfolgende Fragebogen wurde bei Studieneinschluss im Rahmen der Rekrutierung von allen Teilnehmenden ausgefüllt und diente der Erfassung soziodemographischer Basisdaten.

6.2 Rekrutierungsflyer

LVR Qualität für Menschen

hhu Heinrich Heine Universität Düsseldorf

StudienteilnehmerInnen gesucht!

Kann man mit dem Handy messen, wie es mir geht?



Wie stimmen Smartphone-Nutzung und Alltagsverhalten mit dem seelischen Zustand und dem eigenen Wohlbefinden überein? Wie gut spiegeln diese Informationsquellen die in der Klinik erfassten Symptome wider?

Für diese wissenschaftliche Studie suchen wir Patient*innen, die

- ✓ an Depressionen oder einer psychotischen Erkrankung (z.B. Schizophrenie) leiden
- ✓ ein Android-Smartphone besitzen
- ✓ bereit wären auf dem eigenen Smartphone 1x/Woche einen kurzen Fragebogen auszufüllen und Smartphone-Nutzungsinformationen pseudonymisiert zu teilen

Diese Informationen helfen uns dabei zu verstehen, ob Smartphone-basierte Daten zur Verbesserung von Diagnostik und Behandlung von psychischen Erkrankungen hilfreich sein könnten.

FÜR IHRE TEILNAHME ERHALTEN SIE EINE AUFWANDSENTSCHÄDIGUNG VON 15€

Wir würden uns über Ihre Teilnahme freuen und stehen für Rückfragen gerne zur Verfügung:

Theokritos.Veskoukis@lvr.de
Fiona.Radtke@lvr.de

JÜLICH Forschungszentrum

6.3 Vollständige Ergebnisse der induktiven Statistik

Frage	Gesamtaktivität (β [95%-KI], p korr.)	Fahrradfahren (β [95%-KI], p korr.)	Laufen (β [95%-KI], p korr.)	Gehen (β [95%-KI], p korr.)	Inaktivitätszeit (β [95%-KI], p korr.)
1	-0.31 [-3.15; 2.42], p=0.953	0.25 [-0.35; 0.90], p=0.999	0.29 [0.02; 0.53], p=0.106	-0.82 [-3.50; 1.98], p=0.887	-4.14 [-20.70; 12.66], p=0.867
2	0.51 [-1.68; 3.01], p=0.953	-0.26 [-0.81; 0.32], p=0.999	-0.23 [-0.45; -0.02], p=0.106	0.96 [-1.27; 3.08], p=0.887	-1.14 [-13.52; 12.15], p=0.867
3	-0.08 [-3.02; 2.78], p=0.953	0.00 [-0.61; 0.61], p=0.999	0.25 [0.03; 0.54], p=0.106	-0.28 [-2.75; 2.23], p=0.887	-4.54 [-18.59; 9.35], p=0.867
4	-1.65 [-4.22; 1.03], p=0.909	0.08 [-0.56; 0.69], p=0.999	-0.01 [-0.22; 0.26], p=0.965	-1.73 [-4.40; 0.81], p=0.744	-11.68 [-26.13; 4.36], p=0.669
5	1.57 [-1.31; 4.73], p=0.909	0.05 [-0.63; 0.75], p=0.999	0.27 [0.01; 0.53], p=0.106	1.34 [-1.40; 3.87], p=0.887	11.15 [-5.70; 27.96], p=0.669
6	-1.09 [-3.43; 1.59], p=0.943	-0.07 [-0.60; 0.48], p=0.999	-0.20 [-0.43; 0.03], p=0.125	-0.83 [-3.07; 1.71], p=0.887	1.22 [-12.34; 15.23], p=0.867
7	-0.23 [-3.12; 3.10], p=0.953	0.22 [-0.41; 0.89], p=0.999	0.17 [-0.12; 0.47], p=0.327	-0.56 [-3.76; 2.40], p=0.887	4.72 [-10.72; 21.89], p=0.867
8	-0.33 [-3.00; 2.43], p=0.953	-0.13 [-0.77; 0.48], p=0.999	0.25 [0.02; 0.49], p=0.106	-0.39 [-2.81; 1.86], p=0.887	9.49 [-6.04; 23.94], p=0.669
9	-0.64 [-4.09; 2.79], p=0.953	-0.03 [-0.78; 0.71], p=0.999	-0.13 [-0.40; 0.14], p=0.476	-0.54 [-3.42; 2.67], p=0.887	2.09 [-15.77; 20.48], p=0.867
10	5.35 [-0.18; 10.97], p=0.744	-0.35 [-1.66; 0.87], p=0.999	0.19 [-0.23; 0.68], p=0.535	6.57 [1.93; 12.07], p=0.144	5.73 [-25.61; 38.55], p=0.867

Frage	Gesamtaktivität (β [95%-KI], p korr.)	Fahrradfahren (β [95%-KI], p korr.)	Laufen (β [95%-KI], p korr.)	Gehen (β [95%-KI], p korr.)	Inaktivitätszeit (β [95%-KI], p korr.)
11	3.36 [-2.59; 9.37], p=0.909	-0.41 [-1.76; 1.09], p=0.999	-0.12 [-0.64; 0.37], p=0.703	3.76 [-2.02; 9.18], p=0.744	-9.07 [-38.52; 21.25], p=0.867
12	0.12 [-1.93; 2.66], p=0.953	0.19 [-0.32; 0.75], p=0.999	-0.19 [-0.38; 0.02], p=0.122	0.15 [-1.96; 2.20], p=0.887	12.92 [1.37; 24.94], p=0.432

Tabelle 12: In dieser Tabelle werden die Effekte der zwölf wöchentlich erhobenen EMAs auf die fünf Aktivitäts-parameter (Gesamtaktivität, Fahrradfahren, Laufen, Gehen und Inaktivitätszeit) dargestellt.

Prädiktor	Gesamtaktivität (Mittelwert) (β [95%-KI], p korr.)	Gehen (Mittelwert) (β [95%-KI], p korr.)	Laufen (Mittelwert) (β [95%-KI], p korr.)	Fahrradfahren (Mittelwert) (β [95%-KI], p korr.)	Still-Zeit (Mittelwert) (β [95%-KI], p korr.)
ANER	31.22 [-17.40; 79.82], p_FDR=0.494	-1.04 [-4.84; 2.77], p_FDR=0.727	-0.09 [-0.28; 0.09], p_FDR=0.614	-0.60 [-1.57; 0.38], p_FDR=0.338	21.24 [-9.35; 51.82], p_FDR=0.871
ANDP	-14.62 [-42.20; 12.96], p_FDR=0.494	-1.80 [-3.94; 0.34], p_FDR=0.576	-0.10 [-0.20; 0.00], p_FDR=0.336	-0.39 [-0.95; 0.18], p_FDR=0.338	-5.73 [-23.73; 12.27], p_FDR=0.871
ACTV	-35.90 [-110.30; 38.51], p_FDR=0.494	2.96 [-2.46; 8.39], p_FDR=0.656	0.10 [-0.17; 0.36], p_FDR=0.698	-0.20 [-1.62; 1.23], p_FDR=0.783	-7.33 [-52.31; 37.64], p_FDR=0.893
BPRS-Score	-5.19 [-23.80; 13.44], p_FDR=0.685	-0.72 [-2.18; 0.74], p_FDR=0.656	-0.05 [-0.12; 0.02], p_FDR=0.405	-0.24 [-0.62; 0.13], p_FDR=0.338	-0.09 [-12.19; 12.02], p_FDR=0.989
HOST	-189.61 [-441.60; 62.36], p_FDR=0.494	0.41 [-14.85; 15.67], p_FDR=0.957	-0.14 [-0.87; 0.60], p_FDR=0.847	4.16 [0.41; 7.92], p_FDR=0.186	-44.49 [-168.74; 79.76], p_FDR=0.871
THOT	37.68 [-431.50; 506.86], p_FDR=0.870	5.12 [-14.74; 24.99], p_FDR=0.727	0.06 [-0.90; 1.02], p_FDR=0.894	-1.03 [-6.19; 4.14], p_FDR=0.783	44.90 [-117.76; 207.56], p_FDR=0.871

Tabelle 13: Effekte der BPRS-(Sub)Skalen auf die fünf Aktivitätsparameter. Angegeben sind β mit 95 %-KI und FDR-korrigierten p-Werten.