

Evaluationsbericht

(gemäß Nr. 14.1 ANBest-IF)



Konsortialführung:	BKK·VBU
Förderkennzeichen:	01NVF17050
Akronym:	STEP.De
Projekttitel:	Sporttherapie bei Depression
Autoren	Christiane Stielow, Theresa Bergau, Darlene Heinen, Alba Sanchez, Andreas Heißel, Michael Rapp Prozessevaluation: Gregor Wilbertz, Stephan Heinzel Datenanalyse und statistische Auswertung: Alba Sanchez, Christiane Stielow, Gregor Wilbertz, Stephan Heinzel, Michael Rapp
Förderzeitraum:	1. April 2018 – 31. März 2022



Inhaltsverzeichnis

I.	Abkürzungsverzeichnis	3
II.	Abbildungsverzeichnis	4
III.	Tabellenverzeichnis.....	4
1.	Ziele der Evaluation:	57
2.	Darstellung des Evaluationsdesigns.....	7
2.1	Überblick	8
2.2	Fragestellung und Forschungshypothesen	10
2.3	Effektevaluation	11
2.4	Gesundheitsökonomische Evaluation	13
2.5	Prozessevaluation	14
2.6	Fallzahlenberechnung	16
2.7	Statistische Analysen.....	16
3.	Ergebnisse der Evaluation:	17
3.1	Effektevaluation	17
3.1.1	Probandencharakteristika	17
3.1.2	Analyse des primären Outcome-Parameters.....	18
3.1.3	Analyse der sekundären Outcome-Parameter	19
3.2	Gesundheitsökonomische Evaluation	20
3.2.1	Probandencharakteristika	20
3.2.2	Gesundheitsökonomische Analysen	20
3.3	Prozessevaluation	22
4.	Schlussfolgerungen und Empfehlungen des Evaluators.....	23
4.1	Schlussfolgerungen des Evaluators	24
4.2	Empfehlungen des Evaluators.....	26
5.	Literaturverzeichnis	29
6.	Anhang	34

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher, weiblicher und diverser Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

I. Abkürzungsverzeichnis

Abs.	Absatz
AU	Arbeitsunfähigkeit
BDI	Beck-Depressions-Inventar
BKK	Betriebskrankenkasse
BPNSFS	Basic Psychological Need Satisfaction and Frustration Scale
BSI	Symptom Inventory
bspw.	beispielsweise
bzgl.	bezüglich
bzw.	beziehungsweise
ca.	circa
CTM	Continuous time modell
DiD	Difference in difference
Dr.	Doktor
DPtV	Deutsche Psychotherapeuten Vereinigung
DSGVO	Datenschutzgrundverordnung
e. V.	eingetragener Verein
EXP	Experimental
et al.	et alia
etc.	et cetera
FU	Freie Universität
GEE	general estimating equation
ggfs.	gegebenenfalls
HAM-D	Hamilton Depressions-Skala
i. d. R.	in der Regel
insg.	insgesamt
IQR	Inter Quartile Range
i. R.	im Rahmen
mind.	mindestens
miTT	modifizierten Intention-to-treat-Verfahren
phil.	philosophiae
p. rd.	Post random
Prof.	Professor
RCT	raandomized controlled trial
SD	Standard Deviation
s. o.	siehe oben
s. u.	siehe unten
sog.	so genannt
SGB	Sozialgesetzbuch
StGB	Strafgesetzbuch
Tab.	Tabelle
TAU	Treatment as usual
URL	Uniform Resource Locator
VBU	Verkehrsbau Union
VDS	Verhaltensdiagnostiksystem
vgl.	vergleiche
WAI	Work Ability Index
WHO	World Health Organization
WHODAS	WHO Disability Assessment Schedule
WSAS	Work and Social Adjustment Scale
z. B.	zum Beispiel

II. Abbildungsverzeichnis

Abbildung 1: Studiendesign 10

III. Tabellenverzeichnis

Tabelle 1: Überblick über eingesetzte Messinstrumente und -zeitpunkte 13

1. Ziele der Evaluation^{1,2}

Nach Angaben der WHO sind Depressionen weltweit die häufigste Ursache für Beeinträchtigungen (World Health Organization, 2017). Sie betreffen mehr als 320 Millionen Menschen und machen 7,5 % aller Lebensjahre mit Beeinträchtigung aus (Vos et al., 2016). In Deutschland ist jeder Fünfte bis Sechste mindestens einmal im Laufe seines Lebens betroffen (Jacobi et al., 2004). Neben dem persönlichen Leid wird die Depression mit hoher Arbeitslosigkeit und verminderter Leistungsfähigkeit am Arbeitsplatz in Verbindung gebracht (Birnbaum et al., 2010).

Die noch immer wichtigsten Säulen der Behandlung von Depression sind die Pharmakotherapie und die Psychotherapie. Die Versorgung mit einer Psychotherapie ist jedoch häufig mit langen Wartezeiten verbunden (ca. 3 - 6 Monate) (Bundespsychotherapeutenkammer, 2018). Zudem erzielt sie Remissionsraten von unter 50 Prozent (De Maat et al., 2006; Shinohara et al., 2013) und verursacht hohe Kosten (Wittchen et al., 2011) z. B. durch Langzeittherapien (Maljanen et al., 2016). Neben der Therapieplatzsuche stellen die wahrgenommene öffentliche Stigmatisierung und Vorbehalte gegenüber einer Psychotherapie Hauptbarrieren für die Inanspruchnahme professioneller Hilfe dar (Kammerer et al., 2019; Kendra et al., 2014). Mehr als Zweidrittel der Menschen in Deutschland mit einer Depression erhalten keine adäquate Behandlung (Thorncroft et al., 2017). Eine unbehandelte Depression führt häufig zu einer Verschlimmerung der Symptome, einschließlich der Entwicklung von Komorbiditäten, was zu noch höheren Kosten für die Gesundheitssysteme führt (Deutsche Psychotherapeuten Vereinigung (DPTV), 2021; Kessler, 2012).

Die großen Herausforderungen in der Regelversorgung implizieren den Bedarf für alternative Behandlungsstrategien, welche schnell und einfach verfügbar sein müssen. In diesem Zusammenhang wird die Sporttherapie seit einigen Jahren als nützliche Interventionsmöglichkeit vorgeschlagen, da sie leicht zugänglich ist (Schuch et al., 2017) die körperliche Gesundheit (in einer Bevölkerungsgruppe mit einem höheren Risiko für Komorbiditäten) verbessert (Lawlor & Hopker, 2001) und nicht die Nebenwirkungen hat, die mit Antidepressiva einhergehen, insbesondere bei der Behandlung von leichten bis mittelschweren depressiven Symptomen (Schuch et al., 2017). Nachweise zur Evidenz von Sport bei Menschen mit Depression konnten bereits mehrere Meta-Analysen erbringen (Cooney et al., 2013; Rethorst et al., 2009; Schuch et al., 2016, 2018). So ermittelte beispielsweise die aktuellste veröffentlichte Cochrane Meta-Analyse aus dem Jahr 2013 unter Einschluss von 35 Studien mit mehr als 1.300 Probanden eine moderate Effektstärke von Sport bei Depression. Sie umfasste ausschließlich randomisierte kontrollierte Studien, die Gehen, Laufen (auch Laufband), Fahrradfahren (auch Ergometer), Tanzen, Rudern und Krafttraining als Intervention untersuchten (Cooney et al., 2013). Darüber hinaus konnte hinsichtlich der Wirksamkeit von Sport und Psychotherapie sowie Sport und Pharmakotherapie kein Unterschied festgestellt werden (Cooney et al., 2013; Rethorst et al., 2009). Allerdings waren hier jeweils nur 7 bzw. 3 Studien vorhanden. Eine andere Meta-Analyse zeigte sogar große signifikante Effekte von Sport bei Depression, selbst wenn die Analysen auf qualitativ hochwertige Studien begrenzt wurden (Schuch et al., 2016). Zudem wurde herausgefunden, dass erhöhte körperliche Aktivität im Zusammenhang mit einer geringeren Wahrscheinlichkeit steht, in Zukunft eine Depression zu entwickeln (Schuch et al., 2018). Die Wissenschaftler schlossen 49 prospektive Studien ein mit insgesamt 266939 Probanden und konnten zeigen, dass Sport einen Schutzfaktor für Depressionserkrankungen darstellt, unabhängig von Alter, Geschlecht und Region (Schuch et al., 2018).

Sowohl in der deutschen S3-Leitlinie/Nationale VersorgungsLeitlinie für Unipolare Depression (Schneider et al., 2017) als auch bei der European Psychiatric Association (Stubbs et al., 2018) sind Sport- und Bewegungstherapie als ergänzende Optionen bei der Behandlung von leichter bis mittelgradiger Depression bereits aufgeführt. Eine flächendeckende Umsetzung in die Versorgung ist bisher, auch aufgrund des mangelnden Nachweises aus der Versorgungsforschung, nicht gelungen. So stammt ein Großteil der Belege für die Wirksamkeit von Sporttherapie aus streng kontrollierten Wirksamkeitsstudien unter idealen Bedingungen (Phase-III-Studien) und es fehlen Informationen über die Wirksamkeit von Sporttherapien unter tatsächlichen Bedingungen (Phase-

¹ In Teilen übernommen und übersetzt aus Heißel et al., in prep.

² In Teilen übernommen aus dem Ergebnisbericht

IV-Studien) im Vergleich zur Psychotherapie in der klinischen Routinepraxis (Schuch et al., 2017). Auch ist bisher nur wenig über die praktische Frage der Kosteneffizienz der Sporttherapie bekannt. Die Kosteneffizienz ist ein wichtiges Maß für die Nutzung von Programmressourcen, das den Nachweis der Wirksamkeit ergänzt und wertvolle wirtschaftliche Erkenntnisse für Entscheidungen über die Zuweisung von Ressourcen und die Festlegung von Prioritäten liefert (Bower et al., 2000). Das Projekt STEP.De baute auf den aktuellen wissenschaftlichen Erkenntnissen auf.

Zielsetzung der STEP.De Sporttherapie ist es, Patienten mit leichter bis mittelschwerer Depression einen kurzfristigen, niedrigschwelligen Zugang zur Versorgung zu ermöglichen, um das Risiko von verschlechterten oder chronifizierten Krankheitszuständen bei den Betroffenen zu vermindern bzw. zu verhindern. Dabei kann Sporttherapie Psychotherapie teilweise substituieren oder ergänzen. Im Rahmen der STEP.De Studie wurde ein theoriegeleitetes, innovatives Versorgungskonzept entwickelt und Sporttherapie als Behandlungsoption für Depression in der Gesundheitsversorgung im Vergleich zur Standard-Psychotherapie evaluiert.

In Einklang mit den S.M.A.R.T.-Kriterien wurden als Indikatoren zur Zielerreichung zunächst der Wirksamkeitsnachweis in der komplexen Versorgungsintervention anhand des Schweregrades der Depression, gemessen mittels BDI-II und weitere sekundäre Zielvariablen zugrunde gelegt.

Um ein detailliertes Bild über die individuellen Therapie-Prozesse der Depressionspatienten während der viermonatigen Interventionsphase (Sporttherapie / TAU) zu bekommen, wurde neben der Ergebnis-Evaluation auch eine Prozessevaluation (zwei Messzeitpunkte pro Woche) durchgeführt. Da sich die Fragebogen-Items auf Symptome und andere Therapie-Prozess-Variablen jeweils eines Tages beziehen, bot diese Form der Evaluation ein akkurates, kaum verzerrtes Abbild des psychischen Ist-Zustands eines Teilnehmers. Insbesondere für eine valide Evaluation von Veränderungsprozessen der Emotionsverarbeitung und -regulation ist dies essentiell. Wie in zahlreichen Studien gezeigt werden konnte, bietet die engmaschige Prozessevaluation die Möglichkeit, ungünstige Therapieverläufe zu identifizieren und Gründe für eine Abweichung vom Behandlungsziel der Intervention exakt zu verzeichnen und deren Einfluss auf die Ergebnisse klar nachzuvollziehen.

Die ökonomische Effizienz der Intervention soll anhand von spezifischen Gesundheitskosten pro Fall nachgewiesen werden. Im Follow-up wurden zudem Nachhaltigkeitseffekte und der potenzielle Nutzen der STEP.De Sporttherapie als mögliche Alternative zur psychotherapeutischen Kurzzeittherapie evaluiert.

Das Versorgungskonzept STEP.De wurde im Rahmen der Förderung durch den Innovationsfonds an einer Stichprobe von 393 Patienten umgesetzt, die die Zugangsvoraussetzungen erfüllten. Die Ein- und Ausschlusskriterien sind in Kapitel 2.1 dargestellt.

Für diese Zielgruppe haben die Universität Potsdam, die Freie Universität Berlin und die CONVEMA in einer gemeinsamen Arbeitsgruppe anhand der aktuellen empirischen Befundlage ein viermonatiges, bisher noch nicht umgesetztes Versorgungsprogramm entworfen. Damit soll insbesondere eine Alternative zur psychotherapeutischen Kurzzeittherapie (bis zu 25 Einzelsitzungen) und geeigneten Versicherten eine zeitnahe Therapieoption angeboten werden, um den weiterhin bestehenden (und seit Corona-Krisenbeginn deutlich verschärften) Engpässen bei der kurzfristigen psychotherapeutischen Versorgung entgegenzuwirken. Das STEP.De Programm ist im ambulanten Versorgungsbereich angesiedelt und stellt eine innovative Therapie dar, die genau wie die psychotherapeutische Behandlung oder die medikamentöse Therapie helfen kann, eine depressive Symptomatik zu behandeln.

Vor Beginn der STEP.De Sporttherapie erfolgt bei einem vertraglich eingebundenen Psychotherapeuten zunächst ein diagnostisches Interventionsgespräch, bei dem die Indikation für eine Behandlung zusätzlich zur Einschlussdiagnose abgesichert wird. Anschließend (bei Eignung) erfolgt eine Sporteingangstestung in einer Sporttherapieeinrichtung durch einen geschulten und qualifizierten Sporttherapeuten. Das Ziel der Sporteingangstestung ist es:

1. Die körperliche Eignung im Umfang der durchgeführten Anamnese und Diagnostik (insbesondere orthopädische Indikationen) weitergehend zu überprüfen und damit zusätzlich zur Prüfung durch den modifizierten PAR-Q (insbesondere Herz-Kreislauf-

Erkrankungen) weitere Gründe, die zu einer Anpassung der Trainingsintervention führen können, aufzudecken.

2. Anhand der Ergebnisse und Angaben einen individuellen Sporttherapieplan zu erstellen.
3. Auf Ängste und Bedenken der Teilnehmer einzugehen.

Anschließend (bei Eignung) erhält der Teilnehmer über 16 Wochen zwei Mal wöchentlich die STEP.De Sporttherapie in einer Gruppe mit 4 -12 Teilnehmern. Die STEP.De Sporttherapie orientiert sich hinsichtlich Art, Dauer, Intensität, Häufigkeit und Inhalt an den Leitlinien der American Heart Association (American Heart Association, 2018) und weiteren Empfehlungen aus Metaanalysen zu Bewegung und Depression (Cooney et al., 2013; Schuch et al., 2016). Jede STEP.De Sporttherapieeinheit setzt sich aus einer individuell auf den Teilnehmer abgestimmten Kombination aus Ausdauer- und Krafttraining sowie Koordinations-, Entspannungs- oder Wahrnehmungsübungen zusammen und geht zudem auf mögliche bestehende Komorbiditäten wie z. B. Rückenschmerzen ein. Die STEP.De Sporttherapie wird durch ein Abschlussgespräch mit dem Sporttherapeuten beendet. In diesem wird die Aufrechterhaltung und Integration von körperlicher Aktivität in den Alltag besprochen. Während der STEP.De Sporttherapie wird der Teilnehmer zudem im Rahmen von drei telefonischen Assessments von dem betreuenden Psychotherapeuten begleitet. Ein Nachsorgegespräch mit dem Psychotherapeuten nach Abschluss der STEP.De Sporttherapie dient zudem dazu, gemeinsam mit dem Teilnehmer zu besprechen, ob und ggfs. in welchem Umfang ein weiterer psychotherapeutischer Behandlungsbedarf besteht. Psychotherapeut und Patient entscheiden gemeinsam über den Beginn (bei EXP-Gruppenzugehörigkeit) bzw. die Fortsetzung (bei TAU-Gruppenzugehörigkeit) einer Psychotherapie im Rahmen der Regelversorgung, die sich nahtlos anschließen kann bzw. über eine anschließende psychiatrische oder weitere ärztliche Behandlung.

Um der aktuellen Versorgungsrealität auch im Hinblick auf die Pandemie gerecht zu werden, wurde in Anlehnung an die Regelversorgung mit telemedizinisch gestützten Angeboten, STEP.De als Online-Option ergänzt. Hierbei wird die STEP.De Sporttherapie inklusive der Sporteingangstestung und des Abschlussgesprächs über einen zertifizierten Videodienst umgesetzt. Durch diese Flexibilität im Setting kann STEP.De auch von Versicherten aus Risikogruppen, von immobilen Versicherten, aber auch von Versicherten aus ländlichen Regionen genutzt werden und leistet damit einen Beitrag zur aktuellen Versorgungssituation.

Der Qualifikation und psychologischen Schulung (32 Lerneinheiten) der Sporttherapeuten kommt dabei eine Schlüsselrolle zu, die zusätzlich zur fachlichen sporttherapeutischen Ausbildung eine Aufbauqualifizierung in den Bereichen Psychische Erkrankungen (Schwerpunkt Depression, unter anderem um Frühwarnzeichen für eine Verschlechterung des Gesundheitszustandes zu erkennen und solche, die weitere Maßnahmen erfordern, an den Psychotherapeuten zurückzumelden sowie bei Bedarf Notfallmanagement einzuleiten) und Psychologische Qualifikation (Motivationsförderung, Autonomieförderung, gesundheitsfördernde Kommunikation etc.) beinhaltet. Voraussetzung für die Teilnahme an der Studie STEP.De und damit an der Schulung sowie der Durchführung der STEP.De Sporttherapie, ist ein Fachabschluss in Sportwissenschaften, Physiotherapie oder einem äquivalenten sport- oder bewegungsbezogenen Studien- und Ausbildungsgang im Umfang von mindestens drei Jahren bzw. mind. 4. Semester des fachspezifischen Studiengangs Sporttherapie. Für approbierte Psychotherapeuten steht eine Schulung im Umfang von acht Lerneinheiten zur Verfügung, die insbesondere auf die Ein- und Ausschlusskriterien sowie Information rund um das STEP.De Sporttherapieprogramm eingeht. Die Erfolgsfaktoren dieser Therapiestrukturen liegen vor allem in der Trainingsanleitung durch qualifizierte und geschulte Fachkräfte innerhalb eines strukturierten Zeitplans und den antidepressiven Effekten des Gruppensettings, wodurch deutlich höhere Effektstärken und geringere Abbruchraten erzielt werden (Heinzel et al., 2015; Schuch et al., 2016; Stanton, 2014).

2. Darstellung des Evaluationsdesigns

2.1 Überblick³

Um Effektivität und Effizienz von STEP.De Sporttherapie als störungsspezifische Behandlungsform unter Versorgungsbedingungen zu untersuchen, eignete sich eine prospektive, Cluster-randomisierte Interventionsstudie. Die Randomisierung erfolgte auf der Ebene von 28 approbierten Psychotherapeuten im Land Berlin, stratifiziert nach Berufserfahrung und Praxisgröße, also jeweils 14 Psychotherapeuten in die STEP.De Sporttherapiegruppe (Experimentalgruppe, EXP) und 14 Psychotherapeuten in die Psychotherapiegruppe (Treatment As Usual, TAU). Die Cluster-Randomisierung berücksichtigte insbesondere Kontaminationsbiases. Durch geeignete Adjustierung wurde eine Reduktion konfundierender Ähnlichkeiten in einem Cluster und so, das Vermeiden von Selektions- und Allokationsbiases angestrebt (Rapp et al., 2013; Schulz & Grimes, 2005). Um eine möglichst realistische Vergleichsgruppe als Bezugspunkt zu erhalten, wurde die TAU-Gruppe der üblichen „Standardbehandlung“ unterzogen (Psychotherapie). Dieses Vorgehen erweist sich als ethisch unkompliziert und methodisch streng, da mögliche Effekte der Sportintervention unter naturalistischen Bedingungen mit dem üblichen ambulanten Versorgungsangebot verglichen wurden. In der statistischen Analyse folgt das Design so folgerichtig dem Prinzip der Nichtunterlegenheit (non-inferiority).

Abbildung 1 illustriert das Studiendesign. An Depression erkrankte Versicherte wurden von einem Fallmanager ihrer Krankenkasse kontaktiert und gefragt, ob sie an Sport als Behandlungsform interessiert sind. Wenn ja, wurden sie zur Eingangsdiagnostik an einen Psychotherapeuten weitervermittelt, der zu Beginn entweder der EXP- oder der TAU-Gruppe zugeordnet worden ist. Die Fallmanager waren für die Gruppenzugehörigkeit der Psychotherapeuten verblindet. Über einen Zeitraum von 32 Monaten wurden insgesamt n = 393 Patienten, davon n = 251 in die EXP-Gruppe und n = 142 in die TAU-Gruppe, gemäß Ein- und Ausschlusskriterien aufgenommen.

Eingeschlossen wurden Patienten mit einem Lebensalter von mindestens 18 Jahren und einem ärztlichen Konsil, die die Versorgungsdiagnose vom von einem vertraglich eingebundenem Psychotherapeuten, mittels strukturierter Diagnostik, abgesicherte Diagnose einer leichten bis mittelgradigen depressiven Episode (F32.0, F32.1), rez. depressiver Episode (F33.0, F33.1), Dysthymia (F34.1), bzw. Anpassungsstörung (F43.2), Neurasthenie (F48.0), Reaktionen auf schwere Belastung (F43.8, F43.9) oder Angst und depressive Störung, gemischt (F41.2) aufwiesen. Weitere Einschlusskriterien waren gesundheitliche Teilnahmevoraussetzungen für die STEP.De Sporttherapie. Ausgeschlossen wurden Patienten, die sich bereits in einer laufenden, ambulanten psychotherapeutischen Behandlung befanden, die eine Pflegestufe und /oder ein gesetzliches Betreuungsverhältnis hatten. Außerdem wurden Patienten mit einer aktuellen Abhängigkeitserkrankung sowie bestimmten Psychopharmaka, die für die Sportintervention unerwünschte Wirkungen mit sich brächten (Trizyklika, Antipsychotika), ausgeschlossen.

Wie bereits unter Kapitel 1 kurz dargelegt, erfolgte bei einem der vertraglich eingebundenen Psychotherapeuten vor Interventionsbeginn in beiden Gruppen (TAU und EXP) ein Interventionsgespräch mit dem Ziel, die Indikation für eine Behandlung zusätzlich zur Einschlussdiagnose abzusichern. Hierfür wurden alle Psychotherapeuten darin geschult, das Strukturelle Klinische Interview I für das Diagnostische und Statistische Manual Psychischer Störungen 4 (SKID-I, Structured Clinical Interview for DSM IV, SCID-I), Achse 1, Abschnitt A, E und I anzuwenden. Das SKID-1 ist ein verbreitetes Verfahren zur Diagnostik psychischer Störungen nach dem Klassifikationssystem DSM-IV in der klinischen Praxis und Forschung. Angaben zu den Gütekriterien sind dem Anhang zu entnehmen (Anhang A, S1). Den Psychotherapeuten wurde das SKID-I (Sektion A, E und I) zur Verfügung gestellt. Zudem wurden sie zu Einsatz und Anwendung des SKID-I (Sektion A, E und I) im Rahmen der Studie umfangreich geschult.

In die TAU-Gruppe randomisierte Psychotherapeuten leisteten zunächst eine reguläre Behandlung entsprechend dem zugrundeliegenden Selektivvertrag (Basistherapie oder Kurzzeittherapie; 15 bzw. 27 Einzelsitzungen über etwa vier Monate). Im Anschluss waren im Rahmen des Selektivvertrages

³ In Teilen übernommen aus dem Evaluationskonzept (Rapp et al., 2019), mit Anpassungen in der Zeitform
Evaluationsbericht

bei Bedarf weitere Therapieverfahren möglich.

Psychotherapeuten in der EXP-Gruppe schlossen Patienten in die EXP-Gruppe ein und boten eine Begleitung (telefonische Beratungsgespräche) über vier Monate an. Im Anschluss an die Sportintervention wurde ein Nachsorgegespräch geführt. Bei Bedarf waren im Rahmen des Selektivvertrages weitere Therapieverfahren (z. B. Kurzzeittherapie, Gruppentherapie) nach der STEP.De Sporttherapieintervention möglich, insofern weiterer Behandlungsbedarf bestand. Ergebnisse des nationalen niederländischen Surveys für mentale Gesundheit zeigen, dass der Median der Dauer einer depressiven Episode (MDD) bei 6 Monaten liegt, einer milden depressiven Episode bei 3 Monaten (ten Have et al., 2017). In der Literatur werden zwei konsistente Prädiktoren von länger andauernden Verläufen genannt: 1) Depressionsschwere und 2) längeres Bestehen der Symptomatik vor der Vermittlung in ein Behandlungsangebot (Holma et al., 2008; Ramana et al., 1995; Vuorilehto et al., 2009). Aufgrund der Einschlusskriterien in STEP.De (zeitnahe Vermittlung in eine Behandlung, Ausschluss schwerer depressiver Episoden) wurde ein Veränderungseffekt (response) innerhalb des Interventionszeitraums von 4 Monaten erwartet. Das Ziel einer möglichen fortführenden Behandlungsphase bestand darin, Restsymptomatik aufzulösen und das Rückfallrisiko zu mindern (Ballenger, 1999; Birmaher & Brent, 1998). Die American Psychological Association und das UK National Institute for Health and Care Excellence (NICE) gehen von einer kritischen Phase von 6 Monaten nach Symptomreduktion aus und empfehlen eine Fortführung der Medikamenteneinnahme für diesen Zeitraum, um das Rückfallrisiko zu mindern (American Psychiatric Association, 2013).

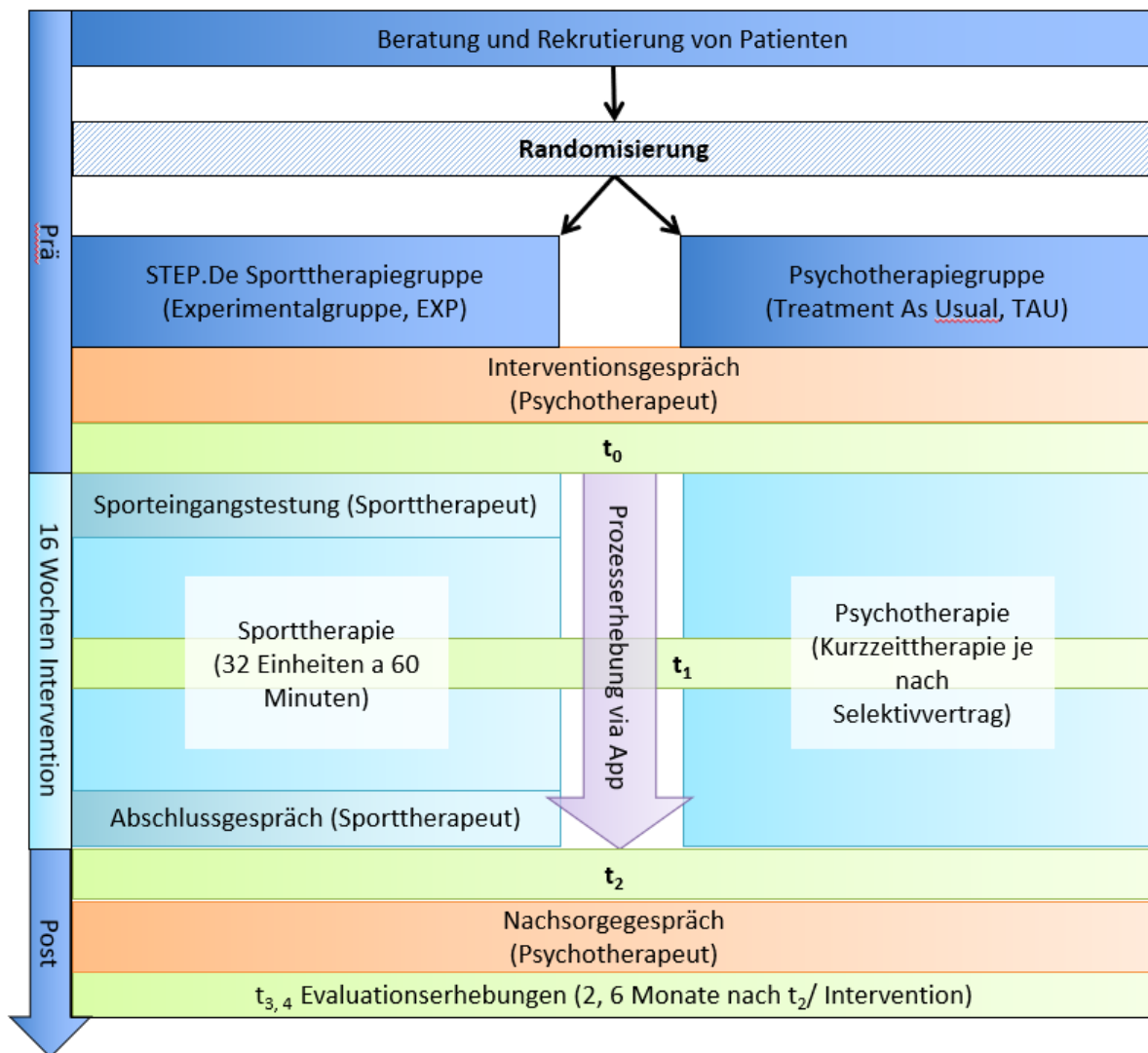


Abbildung 1: Studiendesign

2.2 Fragestellung und Forschungshypothesen⁴

Die zentrale Fragestellung der Evaluation ist, ob die neue Versorgungsform STEP.De eine der Regelversorgung (Psychotherapie, leitliniengerechte Standardbehandlung) nicht unterlegene (gleichwertige) Behandlungsoption bei der ambulanten Behandlung von Patienten mit der Diagnose einer leichten bis mittelgradigen Depression ist. Als Indikatoren werden neben dem Schweregrad der Depression, gemessen mittels BDI-II (Self-Rating) als primäre Zielvariable und dem HAM-D (Fremdrating), verschiedene weitere sekundäre Zielvariablen herangezogen.

Weiterhin wird die ökonomische Effizienz der STEP.De Sporttherapie anhand von spezifischen Gesundheitskosten pro Fall überprüft. Im Follow-up werden Nachhaltigkeitseffekte und der potentielle Nutzen der der STEP.De Sporttherapie als mögliche Behandlungsoption oder Alternative zur psychotherapeutischen Kurzzeittherapie evaluiert. Darüber hinaus wird mit einer Prozessevaluation der spezifische Wirkmechanismus der Sport-Intervention untersucht.

Folgende Forschungshypothesen werden überprüft:

(1) Nach 16 Wochen Intervention und bei der Nachbeobachtung nach 6 Monaten zeigt die

⁴ In Teilen übernommen aus Förderantrag Innovationsfonds; Heissel et al., 2020; Heißel et al., in prep.
 Evaluationsbericht

Gesamtstichprobe eine signifikante Verbesserung

- a. der depressiven Symptomatik,
- b. der Arbeitsfähigkeit,
- c. der gesundheitsbezogenen Lebensqualität.

(2) Nach 16 Wochen Intervention und bei der Nachbeobachtung nach 6 Monaten ist die STEP.De Sporttherapie der Psychotherapie bei der Verringerung depressiver Symptome nicht unterlegen.

(3) In Bezug auf die wirtschaftliche Kosteneffizienz

- a. Im Beobachtungszeitraum von 10 Monaten vor im Vergleich zu 10 Monaten nach Interventionsbeginn fallen die Gesamtkosten (ambulante Kosten, stationäre Kosten, Medikamente, Heil- und Hilfsmittel) inklusive Interventionskosten für die STEP.De Sporttherapiegruppe in einem difference-in-differences Modell geringer aus als für die Vergleichsgruppe Psychotherapie.
- b. Im Beobachtungszeitraum von 10 Monaten vor im Vergleich zu 10 Monaten nach Interventionsbeginn fallen die Gesamtkosten (ambulante Kosten, stationäre Kosten, Medikamente, Heil- und Hilfsmittel) ohne Interventionskosten für die STEP.De Sporttherapie nicht signifikant höher aus.
- c. Im Beobachtungszeitraum von 10 Monaten nach Interventionsbeginn fallen die Gesamtkosten (ambulante Kosten, stationäre Kosten, Medikamente, Heil- und Hilfsmittel) inklusive Interventionskosten für die STEP.De Sporttherapie signifikant geringer aus als für die Psychotherapie Gruppe.
- d. Die STEP.De Sporttherapie weist signifikant geringere Interventionskosten (Sport- und Psychotherapiekosten) im 10-Monatszeitraum ab Beginn der Intervention auf als die Vergleichsgruppe Psychotherapie.
- e. Die STEP.De Sporttherapie weist signifikant geringere Interventionskosten im Interventionszeitraum auf als die Vergleichsgruppe Psychotherapie.

Im Rahmen der Prozess-Evaluation von STEP.De Sporttherapie wurden folgende 3 Ziele bzw. Fragestellungen verfolgt:

1. *Ungünstige Verläufe*: Unterscheiden sich die individuellen Verläufe von Probanden der STEP.De Sporttherapie zu denen der TAU -Gruppe?
2. *Veränderungsmuster*: Gibt es Unterschiede in der Geschwindigkeit, mit der Veränderungen in verschiedenen Symptombereichen stattfinden?
3. *Wirkmechanismus*: Lässt sich der (erwartete) positive Effekt von körperlicher Aktivität auf die depressiven Symptome dadurch erklären, dass die körperliche Aktivität zunächst zu einer Verbesserung kognitiver Funktionen führt, durch eine verbesserte kognitive Kontrolle dann Emotionen besser reguliert werden können und sich hierdurch schließlich die depressiven Symptome reduzieren?

2.3 Effektevaluation⁵

Die Effektevaluation der STEP.De Sporttherapieintervention erfolgte in drei konsekutiven Messzeitpunkten (vor der Intervention, zur Mitte der Intervention, nach der Intervention) in EXP- und TAU-Gruppe.

Die Erfolgsmessung der störungsspezifischen Intervention wurde auf verschiedenen Messebenen

⁵ In Teilen übernommen aus dem Evaluationskonzept (Rapp et al., 2019), mit Anpassungen in der Zeitform
Evaluationsbericht

evaluiert (siehe Tabelle 1). Als primärer Endpunkt wurde der Schweregrad der Depression mit dem in diesem Bereich am häufigsten eingesetzten Messinstrument, dem BDI-II (Hautzinger, Keller, Kühner, et al., 2009) erhoben. Der BDI-II ist ein störungsspezifisches Messinventar, welches kognitive, affektive und somatische Symptome einer Depression abfragt und sich besonders zur Erhebung des Ausmaßes depressiver Symptomatik in klinischen Stichproben eignet. Die revidierte, deutsche Version des Fragebogens weist gute psychometrische Kennwerte auf (Kühner et al., 2007). Sekundäre Endpunkte sind das Vorliegen einer Depression gemessen mit Hilfe der klinischen Fremdbeurteilungsskala Hamilton-Depressions-Rating-Skala (HAM-D) (Hamilton, 1960). Als störungsübergreifendes Inventar wurde das Verhaltensdiagnostik-System (VDS-90-R) (Sulz & Grethe, 2005) eingesetzt, welches neben Depressivität auch Somatisierung und Ängstlichkeit erfasst und somit ein breiteres Spektrum psychologischer Pathologie einschließt. Das VDS-90-R gilt als Standardverfahren zur Erhebung psychologischer Symptome und findet eine breite Anwendung im Rahmen psychotherapeutischer Qualitäts-Managements. Weiterhin wurden Instrumente eingesetzt, die mögliche Störungsfolgen erfassen: Die Work and Social Adjustment Scale (WSAS) (Heissel et al., 2021; Mundt et al., 2002a) erfasst verschiedene Dimensionen von Arbeitsfähigkeit (u. a. derzeitige Arbeitsfähigkeit im Vergleich zu der besten je erreichten Arbeitsfähigkeit). Dieses Maß ist von besonderem Interesse, da vor allem Einschränkungen in der Arbeitsfähigkeit aufgrund psychischer Erkrankungen zu hohen Kosten für das Gesundheitssystem führen. Tage der Arbeitsunfähigkeit wurden daher für geplante Analysen in die Evaluation aufgenommen. Zur Abbildung des Gesundheitszustandes kam zudem der WHO Disability Assessment Schedule 2.0 (WHODAS 2.0) (Üstün et al., 2010) zum Einsatz. Als weiterer Indikator psychologischen Wohlbefindens wurde die gesundheitsbezogene Lebensqualität (SF-12, EQ-5D) (Gandek et al., 1998; Herdman et al., 2011) erhoben. Weiterhin wurden vermeintliche Wirkfaktoren, die in diesem Zusammenhang eine Rolle spielen, berücksichtigt. So wurde die psychologische Bedürfnisfrustration und -befriedigung über den BPNSFS (Heissel et al., 2019), sowie Selbstwirksamkeit über den GSE-6 (Romppel et al., 2013) erfragt.

Die Erhebung des BDI-II, GSE-6, EQ-5D, BPNSFS, WSAS und SF-12 erfolgte online durch secuTrial®. Die Teilnehmer erhielten eine URL sowie einen Benutzernamen und vergaben ihr Passwort bei der ersten Eingabe selbständig.

Die Erhebung des HAM-D, WHODAS 2.0 und VDS-90-R erfolgte per Telefon-Assessment durch teilverblindete Assessoren. Diese wurden nach einem spezifischen Ratertraining eingesetzt und waren für den Interventionscharakter der Studie verblindet. Teilnehmer und Psychotherapeuten wurden auf den Verblindungsgrad der Assessoren hingewiesen.

Zwei katamnestische follow-up Erhebungen (2 und 6 Monate nach t_2) ermöglichen es, die Stabilität der Veränderung zu erfassen. Studien im Kontext von Sport und Depression, die eine Auswertung von follow-up Erhebungen einschließen, sind bisher äußerst rar. So liegt die berichtete Quote in der Cochrane Metaanalyse von Cooney et al. (2013) bei 8 von 39 eingeschlossenen Studien. In einer weiteren Metaanalyse wird ein Medianwert von 6 Monaten Beobachtungszeitraum nach Interventionsende für Studien mit follow-up berichtet (Krogh et al., 2017). Besonders vor dem Hintergrund eines zunehmenden Selektionsbias, aufgrund von abnehmender Teilnahmebereitschaft zu zunehmend späteren Erhebungszeitpunkten und dem damit einhergehenden Risiko eines Ergebnisbias (Fewtrell et al., 2008) ist die Aussagekraft des 6-Monats-Follow-Up verglichen zum 12-Monats-Follow-Up als präziser einzuschätzen. Auch bleibt unklar, inwiefern langfristige Folgeeffekte auf die Intervention zurückzuführen sind, was die Aussagekraft zu einem sehr späten Erhebungszeitpunkt zusätzlich einschränkt. Um explorative Subgruppenanalysen zur Nachhaltigkeit der verschiedenen Interventionen durchzuführen, wurden ausgewählte Messinstrumente eingesetzt. Für die Subgruppenanalysen war es weiterhin erforderlich den Fortgang der weiteren Psychotherapie (Abschluss, Kurzzeittherapie, Langzeittherapie, Gruppentherapie) kategorial zu erheben.

Tabelle 1: Überblick über eingesetzte Messinstrumente und -zeitpunkte

Messinstrumente	Abk.	Messzeitpunkte					Items	Art der Erfassung	Erhoben durch
		t ₀	t ₁	t ₂	t ₃	t ₄			
Outcome-Evaluation									
Beck Depression Inventory	BDI-II(Hautzinger, Keller, Kühner, et al., 2009)	X	X	X	X	X	21	Fragebogen	Online/ secuTrial
Hamilton Rating Scale for Depression	HAM-D(Hamilton, 1960)	X		X		X		Interview	teilverblindete Assessoren
Work and Social Adjustment Scale	WSAS(Heissel et al., 2021; Mundt et al., 2002a)	X	X	X	X	X	5	Fragebogen	Online/ secuTrial
WHO Disability Assessment Schedule 2.0	WHODAS2.0(Üstün et al., 2010)	X		X		X	12	Interview	teilverblindete Assessoren
General Self-Efficacy Scale	GSE-6(Romppel et al., 2013)	X		X		X	6	Fragebogen	Online/ secuTrial
EuroQol Five-Dimensional Questionnaire	EQ-5D(Herdman et al., 2011)	X	X	X	X	X	5	Fragebogen	Online/ secuTrial
Basic Psychological Need Satisfaction and Frustration Scale	BPNSFS(Heissel et al., 2019)	X		X		X	12	Fragebogen	Online/ secuTrial
Verhaltensdiagnostik-System-90 Revision	VDS-90-R(Sulz & Grethe, 2005)	X		X		X	22	Fragebogen	teilverblindete Assessoren
12-Item Short Form Survey	SF-12(Gandek et al., 1998)	X		X		X	12	Fragebogen	Online/ secuTrial
Prozessevaluation		t₀ – t₂ (2-mal pro Woche)							
körperliche Aktivität							3	Fragen	App
Depressionssymptomatik							3	Fragen	App
Emotionen							4	Fragen	App
Emotionsregulation							6	Fragen	App
Kognitive Kontrolle							2	Fragen	App
Selbstwirksamkeitserwartung							1	Fragen	App
Therapiemotivation							1	Fragen	App
Therapiefortschritt							1	Fragen	App

2.4 Gesundheitsökonomische Evaluation

Zusätzlich zur Effektevaluation in Form einer Effektivitäts-Analyse, bei der die primären und sekundären Outcome-Parameter untersucht wurden, wurde eine gesundheitsökonomische Evaluation vorgenommen, in der die Kosten analysiert wurden. Hier wurden die Kosten für stationäre und ambulante Versorgung, Psychopharmakakosten sowie Kosten für Heil- und Hilfsmittel erfasst. Datengrundlage hierfür bildeten Abrechnungsdaten der 3 beteiligten Krankenkassen.

Dabei konzentrierten sich die Datenanalysen auf die Unterschiede zwischen EXP- und TAU-Gruppe zu den definierten Beobachtungszeitpunkten. Die für die vorliegende Evaluation herangezogenen

Endpunkte/ Kostenindikatoren sind im Anhang (Tabelle S1) detailliert dargestellt.

Es wurden die individuellen Abrechnungsdaten pro Fall für den Beobachtungszeitraum von 10 Monaten vor und nach Einschlussdatum übermittelt.

2.5 Prozessevaluation

Um ein detailliertes Bild über die individuellen Therapie-Prozesse der Depressionspatienten während der viermonatigen Interventionsphase (STEP.De Sporttherapie / TAU) zu bekommen, war neben der Ergebnis-Evaluation auch eine Prozessevaluation (zwei Messzeitpunkte pro Woche) geplant. Da sich die Fragebogen-Items auf Symptome und andere Therapie-Prozess-Variablen jeweils eines Tages beziehen, bot diese Form der Evaluation ein akkurates, kaum verzerrtes Abbild des psychischen Ist-Zustands der Teilnehmer (Hardy & Llewelyn, 2015).

Zu diesem Zwecke wurden alle Teilnehmer der STEP-De Sport- sowie Psychotherapie im Rahmen des Projekts eingeladen, an der simultanen Prozess-Evaluation teilzunehmen. Von den N=345 angefragten Personen gaben N=289 (84%) ihr Einverständnis, zogen es jedoch in 5 Fällen wieder zurück oder begannen in 3 weiteren Fällen aus unbekanntem Gründen letztendlich nie die aktive Studienteilnahme. Die aufgenommenen Probanden der Prozess-Evaluation (N=281) wurden zu Beginn der Therapie durch einen Studienmitarbeiter aufgesucht, um die Instruktionen und insbesondere die Smartphone-App zu erhalten. Nach der Installation auf dem privaten Handy des Probanden oder der Aushändigung eines vorinstallierten Studienhandys bestand die Möglichkeit, die Studien-App auszuprobieren und Test-Fragebögen durchzuführen. Nach Abschluss dieser Einführung konnten die Probanden selbstständig und in der weit überwiegenden Zahl problemlos die Kurz-Befragungen auf dem Smartphone bearbeiten, sobald die Studien-App ein entsprechendes Signal gab. Die Befragungen erstreckten sich in Form von Abendbefragungen im Abstand von 2-4 Tagen über den gesamten Therapiezeitraum (16 Wochen, insg. durchschnittlich 37 Befragungen) und umfassten jeweils 22 Fragen, insbesondere zu aktuellen depressiven Symptomen (Niedergeschlagenheit, Antriebsstörung, Anhedonie), kognitiver Kontrolle (Konzentrationsfähigkeit, Planung und Organisation der Tagesaktivitäten), Strategien der Emotionsregulation/ Umgang mit negativen Gefühlen (z. B. Rumination) sowie Einschätzungen der körperlichen Aktivität (in Minuten und getrennt für leichte, moderate und anstrengende Tätigkeiten). Es wurden insgesamt 93% aller Fragebögen beantwortet; die mittlere Bearbeitungsdauer betrug dabei 3,5 Minuten. Eine Nachbefragung ergab, dass 78% der Probanden mit der Prozess-Evaluation zufrieden waren und/oder nichts zu beanstanden hatten, 8% erwähnten sowohl positive wie negative Aspekte der Teilnahme (z. B. „Die häufigen Befragungen im Alltag waren manchmal nervig, aber halfen mir auch, mein Befinden genauer zu beobachten.“), 5% machten sich Sorgen, wenn sie einmal eine Kurz-Befragung verpasst hatten, 8% bewerteten die Smartphone-Befragungen als teilweise negativ (z. B. weil zu häufig oder wegen zu kurzer Antwortfenster), 1% der Befragten störten sich an technischen Problemen und Design-Aspekten der Studien-App. Von den Probanden der Prozess-Evaluation brachen n=36 die Studienteilnahme insgesamt und N=12 nur die Prozess-Evaluation ab. Letztere führten die Therapie aber fort und nahmen auch weiterhin an der Outcome-Evaluation teil (dieses Vorgehen eines „Teil-Dropouts“ sollte die Abbruchraten für die Studie möglichst geringhalten).

Für die Prozessevaluation wurden in einem ersten Schritt die individuellen Verläufe der Studien-Probanden über den gesamten Therapiezeitraum (16 Wochen) analysiert. Hierzu wurden die Antworten jeder Versuchsperson über die 3 Depressionsitems für jeden verfügbaren Zeitpunkt zu einem aktuellen Gesamtdepressionswert gemittelt. Basierend auf diesen Depressionswerten wurden individuelle lineare Regressionen berechnet und der jeweilige Regressionskoeffizient im Sinne eines individuellen tendenziellen Trends interpretiert. Für jede Gruppe (Therapiebedingung: STEP.De Sporttherapie oder TAU) wurde der Anteil an Studienpatient*innen berechnet, der auf diese Weise eine tendenzielle Symptomverschlechterung erlebt haben. Des Weiteren wurden dokumentierte Studienabbrüche hinsichtlich ihrer Häufigkeit für jede Gruppe summiert und

mittels Chi-Quadrat-Test zwischen den beiden Gruppen auf statistische Signifikanz getestet.

Um in einem zweiten Schritt die Veränderungsprozesse detaillierter zu analysieren, wurden die individuellen Verläufe (an dieser Stelle begrenzt auf Versuchspersonen, die einerseits bis zum Ende an der Studie teilnahmen und andererseits tendenziell positiv auf die Therapie reagierten, d.h. einen abfallenden Trend aufwiesen, s. o.) mithilfe individueller Verlaufskurven modelliert (Anhang B, Abbildung S10). Hierzu wurde die Logistische Funktion mit zwei freien Parametern a und b verwendet:

$$f(x) = \frac{1}{1 + e^{-a(x-b)}}$$

Dies sollte unterschiedliche Geschwindigkeiten in den Veränderungsprozessen sichtbar machen. Konkret ging es hierbei um die Frage, ob kognitive Symptome der Depression (Ruminationshäufigkeit) eine andere zeitliche Dynamik zeigen als die Kardinalsymptome Niedergeschlagenheit, Antriebsstörung und Anhedonie (gemittelt als Depressionswert, s.o.). Getrennt für jede Versuchsperson wurden jeweils zwei Modelle gefittet (einmal für Rumination, einmal für depressive Kardinalsymptome); hierbei kam die R Funktion nlsLM zum Einsatz. Jeweils getrennt für die beiden Gruppen (STEP.De Sporttherapie und TAU) wurden dann die beiden Modellparameter a und b mittels t-Test für abhängige Stichproben auf statistisch signifikante Unterschiede zwischen kognitiven und Kardinalsymptomen getestet. Zur Interpretation der Unterschiede wurden Kurven auf Basis der mittleren Modellparameter je Gruppe und Symptombereich erstellt.

Der postulierte Wirkmechanismus von körperlicher Aktivität auf die Depression (s. Hypothesen) wurde mittels einer Mediationsanalyse basierend auf dynamischen (Lag-) Mehrebenen-Modellen überprüft (Anhang B, Abbildung S11). Um die statistischen Effekte im Sinne kausaler Wirkmechanismen interpretieren zu können, wurden die Effekte von einem Zeitpunkt (t) auf den nächsten (t+1) analysiert. Als Voraussetzung wurde zunächst die Signifikanz aller paarweise Regressionseffekte überprüft (z. B. körperliche Aktivität->Kognitive Kontrolle(t+1), Rumination(t)->Depression(t+1), usw.). Die Effekt-Koeffizienten wurden mittels Restricted-Maximum-Likelihood geschätzt und in t-Tests auf Signifikanz geprüft. In einer Mediationsanalyse wurden schließlich der sog. indirekte Effekt (von körperlicher Aktivität über Kognitive Kontrolle und/oder Rumination auf Depression) mit dem sog. direkten Effekt (ohne Umweg von der körperlichen Aktivität auf Depression) verglichen und mithilfe einer quasi-Bayesianischen Monte-Carlo-Simulation (n=1000) auf Signifikanz geprüft.

Aufgrund der besonderen Bedeutung innerhalb des Wirkmodells (s. u.), wurden zwei Variablen der Prozessevaluation neben der Selbsteinschätzung mittels zusätzlicher Methoden erfasst. Zur Quantifizierung der Bewegung im Alltag erhielten die Probanden einen Bewegungssensor (GENEActiv Accelerometer) zu drei Zeitpunkten im Laufe der Studie (Beginn, Mitte, Ende), um ihn für jeweils 7 Tage am Handgelenk zu tragen. Die Daten wurden hierbei kontinuierlich mit 50 Hz aufgezeichnet (Range +/- 8 g) und anhand des etablierter ENMO-Maßes ausgewertet (Euclidean Norm Minus One). Auf diese Weise ließen sich Episoden identifizieren, in denen die Probanden eine mindestens leichte körperliche Aktivität ausführten (definiert als ENMO > 30 mG, entspricht ca. 1.6 MET [Metabolisches Äquivalent]) und mit ihrer momentanen Befindlichkeit in Verbindung setzen.

Zur objektiven Erfassung der momentanen kognitiven Leistungsfähigkeit führten die Probanden außerdem regelmäßig (ca. 1 mal pro Woche) ein kognitiv-experimentelles Paradigma (N-Back) in der App durch. Es handelt sich hierbei um einen etablierten Test, der unter 3 verschiedenen Schwierigkeitslevels (kognitive Beanspruchung) durchgeführt wird. Aus Reaktionszeiten und Fehlerraten lässt sich die momentane kognitive Performanz errechnen.

2.6 Fallzahlenberechnung⁶

Die Stichprobengröße wurde für eine Nichtunterlegenheitsstudie mit folgenden Parametern geschätzt: Die Nichtunterlegenheitsmarge wurde auf 0.3 SD für die klinischen Skalen festgelegt, was drei Punkten des BDI-II Gesamtscores zum Zeitpunkt der Nachbehandlung entspricht (mit SD=10, gemäß den Ausgangsdaten der aktuellen Studie).

Die Nicht-Unterlegenheitsschwelle wurde in Anlehnung an Norman et al. (Norman et al., 2003) (empfehlen eine halbe SD) mit $0.3 * SD$ konservativer gesetzt als empfohlen. Diese Schwelle ist zwar geringfügig höher als die minimal important difference (MID) von 0.24 in Depressionsstudien (Cuijpers et al., 2021), andere Studien zur Depressionsbehandlung stützen hingegen die Lockerung der Toleranz und nehmen ähnliche Werte an ($0.35*SD$ (Rhodes et al., 2014); $0.39*SD$ (Richards et al., 2016)).

Die Verwendung des R-Pakets "SampleSize4ClinicalTrials" mit einem wahren Effekt von Null, einer α -Fehlerrate von 5 % und einer Power von 80 % sowie einer Nichtunterlegenheitsgrenze von 3 und einer SD von 10 ergibt eine Gesamtzahl von 312 Teilnehmern, um die Nichtunterlegenheit zwischen den Gruppen zum Zeitpunkt der Nachbehandlung zu testen. Unter Berücksichtigung zusätzlicher 25 % Ausfälle führt dies zu einer geschätzten Stichprobengröße von 390 Teilnehmern.

2.7 Statistische Analysen⁷

Die Auswertung des primären Evaluationsoutcomes erfolgte durch die FU Berlin/ Prof. Stephan Heinzel. Diese erhielten die anonymisierten und den Gruppen nicht zuzuordnenden Datensatz. Damit sollte die Unabhängigkeit der primären Evaluation gewährleistet werden.

Für den primären Endpunkt (Schweregrad der Depression, gemessen mittels BDI-II) wurden lineare Regressionsanalysen mit Clusteradjustierung berechnet. Um die Abhängigkeit der Messungen innerhalb eines Clusters zu berücksichtigen, wurden für die Überprüfung der Effekte verallgemeinerte lineare Schätzungsgleichungen (GEE) gerechnet. Die GEE-Analysen der MITT Population wurden unter Verwendung der LOCF-Imputation (Last Observation Carried Forward) durchgeführt. Für die LOCF-Imputation des BDI-II-Scores wurden auch die Daten der Zeitpunkte T1 und T3 einbezogen. Weiterhin wurden für die sekundären Endpunkte HAM-D und WSAS gemischte lineare und logistische Regressionsmodelle mit festen und zufälligen Faktoren zur Abschätzung der Heterogenität über Cluster hinweg durchgeführt (vgl. Rapp et al., 2013). Für die weiteren sekundären Endpunkte (WHODAS 2.0, GSE-6, EQ-5D-5L, BPNSFS, SF-12) wurden Varianzanalysen (ANOVA) mit Messwiederholung durchgeführt. Werte zu Studienbeginn und Unterschiede in den Baselineerhebungen wurden in den Modellschätzungen als Confounder berücksichtigt. Für die Berechnung der Effektstärke der Zwischengruppeneffekte ($d_{\text{intervention}}$) wurden der Parameterschätzer des Gruppeneffektes herangezogen und an der gepoolten Standardabweichung (SD) der beobachteten Messwerte standardisiert. Der besseren Vergleichbarkeit wegen wurde bei der dichotomen Ergebnisgröße Vorliegen einer Depression die geschätzte Odds Ratio mittels logit-Transformation ebenfalls in standardisierte Mittelwertdifferenzen überführt. Die Höhe der berechneten Effekte wurde entsprechend der von Cohen vorgeschlagenen Konvention interpretiert: kleine Effekte ab $d=0,2$, mittlere Effekte ab $d=0,5$ und hohe Effekte ab $d=0,8$.

Für gesundheitsökonomische Analysen wurden Kosten je Fall aus Perspektive der GKV untersucht. Um die Veränderungen der Gesamtkosten zwischen den Gruppen (EXP vs. TAU) 10 Monate vor und 10 Monate nach Interventionsbeginn zu untersuchen, wurde ein Difference-in-Difference-Ansatz (DiD) unter Verwendung gemischter Modelle (GLMM) durchgeführt. Es wurde eine Normalverteilung und ein Identitätslink verwendet.

⁶ In Teilen übernommen und übersetzt aus Heißel et al., in prep.; siehe auch Heißel et al., 2020 für weitere Einzelheiten

⁷ In Teilen übernommen aus Rapp et al., 2019, mit Anpassungen in der Zeitform

Die vorgestellten gemischten Regressionsmodelle wurden mit SPSS (Version 28.0) geschätzt.

Eine Zusammenfassung bisheriger Forschung zur Kosteneffektivität im Kontext der Depressionsbehandlung zeigt, dass die überwiegende Anzahl an Studien einen Zeithorizont von 6 Monaten umfassen, der die Akutphase und die fortführende Behandlungsphase erfasst (Zimovetz et al., 2012). Das primäre Ziel der Kostenevaluation in STEP.De war es, die Gesamtkosten vor und nach der Intervention in einem difference-in-differences Ansatzes zu vergleichen und sekundär die Kostendaten während der Interventionsphase von 16 Wochen zwischen den Gruppen zu vergleichen. Die Erhebung von Kostendaten für den Zeitraum der Interventionsphase ist üblich. Bei einem Vergleich der Kosteneffizienz einer medikamentösen Depressionsbehandlung mit den Kosten einer kognitiven Verhaltenstherapie wurde dafür der jeweilige Interventionszeitraum angesetzt (Haby et al., 2004).

Um die Kosten für mögliche anschließende Behandlungen auch bei länger andauernden depressiven Episoden berücksichtigen zu können, wurde ein Erhebungszeitraum von 10 Monaten nach Einschluss geplant, welcher zeitgleich mit der t4-follow-up-Erhebung endet (6 Monate nach Interventionsende). Aufgrund der zeitverzögerten Übermittlung ambulanter Datenpunkte (um 9 Monate) konnte hier nur eine geringere, jedoch aussagefähige Stichprobengröße (Verlust von ca. 16,6%) für die post-10-Monats-Variablen berücksichtigt werden.

3. Ergebnisse der Evaluation⁸

3.1 Effektevaluation

3.1.1 Probandencharakteristika

Es wurden 393 Patienten mit leichter bis mittelschwerer Depression (Mittelwert des BDI-II = 22.7, SD 9.9) in die STEP.De-Studie aufgenommen. Um die Konsistenz zu bewerten, wurde die Diagnose zusätzlich in einer Teilstichprobe von n = 57 mit einem zusätzlichen SKID-I Interview, das von einer im SKID-1 geschulten wissenschaftlichen Mitarbeiterin (M.Sc. Psychologie) der Universität Potsdam durchgeführt. Von den 393 Patienten waren 251 Patienten der EXP-Gruppe und 142 Patienten der TAU-Gruppe zugewiesen (Anhang B, Abbildung S1). 32 Patienten in der EXP-Gruppe und 17 Patienten in der TAU-Gruppe erhielten nicht die zugewiesene Therapie und wurden daher nicht in die mITT-Population (modifizierte Intention-to-Treat Analyse) aufgenommen. Demnach konnten 344 Patienten in die mITT-Analyse einbezogen werden (219 EXP und 125 TAU) (Anhang B, Abbildung S1 für weitere Informationen).

Die Probandencharakteristika wurden zu Studienbeginn erfasst und sind in Tabelle S1 aufgeführt. Die Probanden der Gesamtstichprobe waren zu 71.5% weiblich (EXP: 69.9, TAU: 74.4). Das durchschnittliche Alter der Teilnehmer betrug 49.0 Jahre (IQR 39.0-57.0) und unterschied sich signifikant zwischen den Gruppen. Die meisten Teilnehmer (ca. 56%) hatten als höchsten allgemeinbildenden Schulabschluss einen Realschulabschluss. Zudem lag der größte Anteil der Probanden (ca. 57%) im Bereich eines mittleren Einkommens. Im Mittel wies die Gesamtstichprobe einen BMI von ca. 26 kg/m² auf.

Die demografischen und klinischen Ausgangscharakteristika waren zwischen den Behandlungsgruppen ausgewogen, mit Ausnahme des Alters (p=0.005). Aus diesem Grund wurde das Alter als Kovariate in die Hauptanalyse einbezogen.

Von den 393 Patienten, die in die Studie eingeschlossen wurden, wurde bis zum 6 Monats-Nachbeobachtungszeitraum für insgesamt 106 Personen ein Dropout verzeichnet. Dies entspricht einer Dropout-Rate von 27 % (EXP 29 %, TAU 24 %). Im Zeitraum zwischen Interventionsbeginn (t0) und Nachbehandlung (t2) waren 54 Dropouts zu verzeichnen, was einer Dropout-Rate von 14 % (EXP 15 %, TAU 12 %) entspricht. Sobald bekannt wurde, dass sich eine Person gegen eine

⁸ In Teilen übernommen und übersetzt aus Heißel et al., in prep.
Evaluationsbericht

Weiterführung der Studienteilnahme entschieden, wurde sie telefonisch kontaktiert und im Falle einer Zustimmung zu den Dropout Gründen befragt. Als Gründe für Dropouts bei STEP.De vor Interventionsbeginn sind beispielsweise nicht erfüllte Einschlusskriterien oder ein zu hoher Aufwand für die Studie zu nennen. Personen, die im Laufe der Intervention entschieden haben, die Studie nicht zu beenden, gaben als Gründe beispielsweise die Unvereinbarkeit mit der Arbeit, zusätzlichen Aufwand oder familiäre Probleme an.

Die mittlere Anwesenheitsrate für die STEP.De Sporttherapie betrug in der mITT-Population 45.3 % (IQR 15.6-71.9) (Median 14.5 Einheiten, IQR 5.0-23.0), die der Psychotherapie 56.3 % (IQR 40.6-75.0) (Median 9.0 Sitzungen, IQR 6.5-12.0). Vier Teilnehmer aus der EXP-Gruppe überschritten aufgrund der parallelen Teilnahme am Online-Training zusätzlich zur lokalen STEP.De Sporttherapie durch das Covid-19-Verfahren das Maximum von 32 Einheiten. Von den 219 Teilnehmenden, die in die mITT Analyse einbezogen wurden, haben 3 Personen (1.4%) die STEP.De Sporttherapie ausschließlich online, 174 (79.5%) ausschließlich vor Ort und 42 Personen (19.2%) sowohl online als auch vor Ort in Anspruch genommen. 21.9% der Teilnehmenden aus der EXP-Gruppe nahmen nach der Intervention eine Psychotherapie in Anspruch, während 76.8% der TAU-Gruppe nach dem Interventionszeitraum eine Psychotherapie fortsetzten ($\chi^2=4.564$, $p=.033$).

Im Rahmen der Psychotherapiegruppe kamen zwei Psychotherapieverfahren zum Einsatz: Die Verhaltenstherapie und die Tiefenpsychologisch fundierte Therapie. In ersterem Therapieverfahren waren 26 Psychotherapeutencluster ausgebildet. In zweiterem waren 2 Psychotherapeutencluster ausgebildet. Die Psychotherapieverfahren entsprachen der in der Versorgung vorgefundenen Verteilung in den eingebundenen Psychotherapiezentren. Durch die geringe Anzahl an Psychotherapeuten ausgebildet in Tiefenpsychologisch fundierter Therapie wurden keine Subgruppenanalysen in Bezug auf die Psychotherapieverfahren durchgeführt.

3.1.2 Analyse des primären Outcome-Parameters

Der Rückgang des mittleren BDI-II Scores zwischen t0 und t2 betrug in der EXP-Gruppe 33.0% und in der TAU-Gruppe 32.3%. Auch 6 Monate nach der STEP.De Sporttherapie bzw. nach der Psychotherapie (t4) betrug die Verringerung des mittleren BDI-II Scores 38.4 % bzw. 40.5 % im Vergleich zu den Ausgangswerten. Der Anteil der Patienten mit mittelschweren oder schweren depressiven Symptomen verringerte sich in der EXP-Gruppe von 60.4 % bei t0 auf 29.3 % bei t2 und in der TAU-Gruppe von 59.0 % auf 32.6 % (Anhang C, Tabelle S3).

Tabelle S2 im Anhang zeigt die Mittelwerte, SDs und Stichprobengrößen des BDI-II für beide Gruppen zu allen Messzeitpunkten auf der Grundlage der mITT-Population.

Für das primäre Ergebnis, den mit dem BDI-II gemessenen Schweregrad der Depression, wurden GEE-Analysen durchgeführt. Das erste Modell (Anhang C, Tabelle S5, Modell 1) verglich den Effekt der STEP.De Sporttherapiegruppe auf den Schweregrad der Depression, gemessen mit dem BDI-II, nach 16 Wochen Intervention (t2). Für die mITT-Population zeigte sich ein signifikanter Haupteffekt der Zeit auf den BDI-II Gesamtscore ($B = 5.81$ [95% CI 4.37 bis 7.25], $p<0.001$), was auf eine Abnahme des Depressionsschweregrads nach 16 Wochen Intervention sowohl für die EXP- als auch für die TAU-Gruppe hinweist. Es wurde weder ein signifikanter Effekt der Intervention noch eine Interaktion zwischen Gruppe und Zeit gefunden. Der Unterschied im BDI-II-Gesamtscore zwischen beiden Gruppen zum Zeitpunkt des Interventionsabschlusses (t2) war geringer als die vorgegebene Nichtunterlegenheitsgrenze von 3 Punkten: 16.57 (Standardfehler [SE] 0.76) in der STEP.De Sporttherapiegruppe vs. 17.42 (SE 0.99) in der Psychotherapiegruppe (mittlere Differenz [MD]= -0,85 [95% CI -3,34 bis 1,63], $p=0501$).

In einem zweiten Modell (Anhang C, Tabelle S6, Modell 2) wurde die 6-monatige Nachbeobachtung (t4) miteinbezogen, um zu untersuchen, ob die Auswirkungen der Intervention 6 Monate nach Ende der Behandlung erhalten blieben. Es zeigte sich ein signifikanter Haupteffekt der Zeit (Tabelle S6), der auf eine Verbesserung des BDI-II Scores in beiden Gruppen von t0 zu t4 (7.97 [6.17 bis 9.77]; $p<0.001$) und von t2 zu t4 (2.16 [0.90 bis 3.42]; $p<0.001$) hinweist. Es wurde weder ein Haupteffekt

der Gruppe noch eine Interaktion zwischen Gruppe und Zeit festgestellt.

3.1.3 Analyse der sekundären Outcome-Parameter

In der sekundären Analyse wurde die Nichtunterlegenheit für depressive Symptome, die mit dem HAM-D bewertet wurden, durch eine GEE Analyse untersucht. Die Nicht-Unterlegenheitsschwelle von $0.3 * SD_{t0}$ entspricht 2.16 Punkte auf dem HAM-D Summenscore. Es wird also angenommen, dass die STEP.De Sporttherapie der Psychotherapie bei bis zu 2.16 Punkten Unterschied zu Messzeitpunkt t2 (Interventionsabschluss) nicht unterlegen ist. Es wurde bewusst eine geringe Schwelle festgelegt, um sicher sein zu können, dass wenn die STEP.De Sporttherapie innerhalb dieser Schwelle liegt, eine nicht-unterlegene und damit gleichwertige Behandlungsoption zur Psychotherapie bei leichter bis mittlerer Depression darstellt. Damit bewegen wir uns noch unter den Kriterien der NICE-Guidelines (National Collaborating Centre for Mental Health, 2010). In dem GEE Modell wurde der Effekt der Sporttherapie auf depressive Symptome nach 16 Wochen Intervention (t2) analysiert. Für die mITT-Population wurde ein signifikanter Haupteffekt der Zeit auf den HAM-D-Gesamtscore gefunden (3.10 [95% CI 1-96 bis 4-25], $p < 0.001$). Dies zeigt eine Reduktion der depressiven Symptomatik nach 16 Wochen Intervention sowohl für die Sporttherapie- als auch für die Psychotherapiegruppe. Es gab weder einen signifikanten Effekt der Intervention noch eine Interaktion zwischen Gruppe und Zeit. Der Unterschied im HAM-D-Summenscore zwischen beiden Gruppen zum Zeitpunkt des Interventionsabschlusses (t2) war geringer als die vordefinierte Nicht-Unterlegenheitsschwelle von 2.16 Punkten: 11.44 (Standardfehler [SE] 0.52) in der Sporttherapiegruppe vs. 12.48 (SE 0.65) in der Psychotherapiegruppe (mittlere Differenz [MD] = -1.04 [95% CI -2.70 bis 0.61], $p = 0.218$). Der HAM-D-Score verringerte sich um 4.42 (SE 0.50) ($p < 0.001$) Punkte in der Sporttherapiegruppe gegenüber 3.10 (0.58) ($p < 0.001$) Punkten in der Psychotherapiegruppe.

Weiterhin wurden für die sekundären Outcome-Parameter depressive Symptome sowie berufliche und soziale Anpassung GLMM-Analysen durchgeführt, welche weder einen signifikanten Gruppeneffekt der Intervention noch eine Interaktion zwischen Gruppe und Zeit zeigten. Für depressive Symptome, die mit dem HAM-D gemessen wurden, zeigten die mITT-Analysen einen signifikanten Effekt der Zeit vor vs. nach der Behandlung (3.95 [1.99 bis 5.91]; $p < 0.001$) (Anhang C, Tabelle S7, Modell 3), was auf einen Rückgang der HAM-D Gesamtscores nach 16 Wochen Intervention für beide Gruppen hinweist. Bei Einbeziehung der 6-monatigen Nachbeobachtung in das Modell (t4) (Anhang C, Tabelle S8, Modell 4) wurde ebenfalls ein signifikanter Haupteffekt der Zeit beobachtet (5.96 [3.94 bis 7.99]; $p < 0.001$), der eine Abnahme der depressiven Symptome vom Interventionsbeginn bis zur 6-monatigen Nachbeobachtung anzeigt.

Hinsichtlich der beruflichen und sozialen Anpassung (Anhang C, Tabelle S9, Modell 5) wurde ein signifikanter Effekt der Zeit vor vs. nach der Behandlung festgestellt (5.25 [2.50 bis 8.01]; $p < 0.001$), was auf eine Verbesserung des WSAS Scores sowohl für die STEP.De Sporttherapie als auch für die Psychotherapie hinweist. Bei Einbeziehung der 6-monatigen Nachbeobachtung in das Modell (t4) (Anhang C, Tabelle S10, Modell 6) wurde ebenfalls ein signifikanter Haupteffekt der Zeit beobachtet (7.12 [4.18 bis 10.06]; $p < 0.001$), der auf eine Verbesserung der beruflichen und sozialen Anpassung vom Ausgangswert bis zur 6-monatigen Nachbeobachtung hinweist.

Tabelle S2 im Anhang zeigt die Mittelwerte, SDs und Stichprobengrößen der sekundären Ergebnisse für beide Gruppen zu allen Messzeitpunkten auf der Grundlage der mITT-Population. In Bezug auf Angst lagen die VDS-90 Scores bei t0 in der EXP-Gruppe bei 0.96 und in der TAU-Gruppe bei 0.88, was auf keine pathologisch erhöhten Werte hinweist.

Für weitere sekundäre Endpunkte wurde eine zweifache Varianzanalyse mit wiederholten Messungen (ANOVA) durchgeführt, wobei der Zeitpunkt (t0 und t2) als Faktor innerhalb und die Gruppe als Variable zwischen den Probanden diente.

Die Ergebnisse der ANOVA mit Messwiederholung zeigten einen signifikanten Haupteffekt der Zeit (t0 bis t2) für den WHODAS 2.0 Score ($F[1,2152] = 72.684$, $p < 0.001$), den GSE-6 Gesamtscore

($F[1,247]= 24.583$, $p<0.001$), den EQ-5D-5L-Indexwert ($F[1,240]= 36.197$, $p<0.001$), die mentale (MCS-12; $F[1,220]=94.083$, $p<0,001$) und körperliche Komponente des SF-12 (PCS-12) ($F[1,220]= 13,698$, $p<0001$) sowie der Befriedigung der psychologischen Grundbedürfnisse ($F[1,245]= 11,569$, $p<0,001$) und deren Frustration ($F[1,247]= 26,893$, $p<0,001$). Für keinen der sekundären Endpunkte wurde ein signifikanter Effekt der Gruppe oder eine Interaktion zwischen Gruppe und Zeit gefunden. In den Abbildungen S2 bis S8 in Anhang B sind diese Ergebnisse für jeden sekundären Endpunkt grafisch dargestellt. Nach Hinzufügen des Alters als Kovariate (ANCOVAs mit Messwiederholung), blieb der signifikante Zeiteffekt für alle Variablen erhalten, außer für die Frustration der psychologischen Grundbedürfnisse und den SF-12 PCS-12. Es wurde auch bei den ANCOVAs weder ein signifikanter Effekt der Gruppe noch eine Interaktion zwischen Gruppe und Zeit gefunden.

Bei Einbeziehung der 6-monatigen Nachbeobachtung (t4) wurden ebenfalls eine zweifaktorielle Varianzanalyse mit wiederholten Messungen (ANOVA) durchgeführt, wobei der Zeitpunkt (t0 und t4) als Faktor innerhalb und die Gruppe als Variable zwischen den Probanden diente. Die Ergebnisse der ANOVA mit Messwiederholung zeigten einen signifikanten Haupteffekt der Zeit (t0 bis t4) für den WHODAS 2.0 Score ($F[2, 108]=49.220$, $p<0.001$), den GSE-6 Gesamtscore ($F[2, 196]=22.485$, $p<0.001$), den EQ-5D-5L-Indexwert ($F[2, 192]=23.275$, $p<0.001$), die mentale ($F[2, 164]= 56.749$, $p<0.001$) und die körperliche Komponente des SF-12 ($F[2, 164]= 7.964$, $p<0.001$) sowie die Befriedigung der psychologischen Grundbedürfnisse ($F[2, 194]= 6.290$, $p = 0.002$) und deren Frustration ($F[2, 196]= 17.460$, $p < 0.001$). Auch wenn das Alter als Kovariate berücksichtigt wurde (ANCOVAs mit Messwiederholung), blieb der Zeiteffekt überwiegend signifikant, so für den WHODAS 2.0, den EQ-5D-5L-Indexwert und den GSE-6. Für den BPNSFS blieb der signifikante Zeiteffekt nach Hinzufügen des Alters als Kovariate nur hinsichtlich der Frustration der psychologischen Grundbedürfnisse erhalten. Für den SF-12 zeigte sich unter Berücksichtigung des Alters als Kovariate kein signifikanter Zeiteffekt.

3.2 Gesundheitsökonomische Evaluation

3.2.1 Probandencharakteristika

Teilnehmer, die der Weitergabe ihrer Kostendaten zustimmten, wurden in die gesundheitsökonomische Analyse aus Sicht der gesetzlichen Krankenversicherung einbezogen. Somit umfasste die Stichprobe für die gesundheitsökonomischen Datenanalysen zum Zeitpunkt der aktuellen Datenauswertung (September 2022) 207 Teilnehmer in der EXP-Gruppe (94.5% der mITT Population der EXP-Gruppe) und weitere 124 Teilnehmer in der TAU-Gruppe (99.2% der mITT Population der TAU-Gruppe).

3.2.2 Gesundheitsökonomische Analysen

Die Analyse des primären Endpunktes (Schweregrad der Depression gemessen mittels BDI-II) wurde mit der Stichprobe der gesundheitsökonomischen Daten wiederholt.

Es wurden verallgemeinerte Schätzungsgleichungen (GEE) mit Last Observation Carried Forward (LOCF) verwendet, die die Abhängigkeit der Messungen innerhalb eines Clusters berücksichtigen. Das erste Modell (Anhang C, Tabelle S5, Modell 1) untersuchte den Effekt der STEP.De-Intervention (16 Wochen) auf den Schweregrad der Depression, gemessen mit dem BDI-II. Für die mITT-Population wurde ein signifikanter Haupteffekt der Zeit auf den BDI-II-Gesamtscore (5.93 [95% CI 4.53 bis 7.33], $p<0.001$) gefunden, was auf eine Abnahme des Depressionsschweregrads nach 16 Wochen Intervention sowohl für die STEP.De- als auch für die Psychotherapiegruppe hinweist. Es wurde weder ein signifikanter Effekt der Gruppe noch eine Interaktion zwischen Gruppe und Zeit festgestellt. Der Unterschied im BDI-II-Gesamtscore zwischen beiden Gruppen zum Zeitpunkt des Interventionsendes (t2) war geringer als die vorgegebene Nichtunterlegenheitsgrenze von 3

Punkten: 17.23 in der STEP.De Sporttherapiegruppe vs. 17.27 in der Psychotherapiegruppe.

In einem zweiten Modell (Anhang C, Tabelle S6, Modell 2) wurde die 6-monatige Nachbeobachtung (t4) eingeschlossen, um zu untersuchen, ob die Effekte der Intervention 6 Monate nach Ende der Intervention erhalten blieben. Die Analyse ergab einen signifikanten Haupteffekt der Zeit, der auf eine Verbesserung des BDI-II-Scores in beiden Gruppen von t0 bis t4 (7.97 [6.17 bis 9.77]; $p < 0.001$) und von t2 bis t4 (2.16 [0.90 bis 3.42]; $p < 0.001$) hinweist.

Es wurde weder ein Haupteffekt der Gruppe noch eine Interaktion zwischen Gruppe und Zeit gefunden.

Sport- und Psychotherapiekosten

Die Analyse der Sport- und Psychotherapiekosten pro Fall innerhalb eines Zeitraums von 10 Monaten nach Beginn der Intervention zeigte einen signifikanten Unterschied zwischen den Gruppen ($p < 0.001$) mit durchschnittlichen Ausgaben von 1356.40 € (SD 1550.76 €) in der EXP-Gruppe und 2411.26 € (SD 1603.62 €) in der TAU-Gruppe. Diese Kosten umfassen sowohl die Kosten für die jeweilige Intervention als auch die Kosten für Psychotherapie, die innerhalb des 10-monatigen Zeitraums fortgesetzt bzw. begonnen wurde. Die Analyse der reinen Interventionskosten ergab für 4 Monate STEP.De Sporttherapie (M 681.20 €, SD 269.87 €) signifikant niedrigere Kosten als für 4 Monate Psychotherapie (M 993.52 €, SD 411.08 €; $p < 0.001$). Um zudem die Heterogenität der Cluster zu berücksichtigen, wurden verallgemeinerte lineare gemischte Modelle (GLMM) durchgeführt. Die Cluster (Psychotherapeutenpaare/triplets) wurden als zufällige Effekte einbezogen. Die GLMM Analyse der Kosten pro Fall innerhalb des Zeitraums von 10 Monaten nach Interventionsbeginn zeigte einen signifikanten Gruppeneffekt (-1054,86 [-1405,75 bis -703,97]; $p < 0,001$), der zeigt, dass die Zugehörigkeit zur STEP.De Sporttherapiegruppe mit signifikant niedrigeren Kosten verbunden ist. Auch das GLMM für die Intervention von 4 Monaten zeigte einen signifikanten Gruppeneffekt (-313,33 [-386,89 bis -238,77]; $p < 0,001$), was darauf hinweist, dass die Zugehörigkeit zur STEP.De Sporttherapiegruppe mit signifikant niedrigeren Interventionskosten verbunden ist.

Weitere gesundheitsökonomische Analysen

Weiterhin wurden Arbeitsunfähigkeitstage (AU-Tage), ambulante Kosten, stationäre Kosten, Medikamentenkosten sowie Kosten für Heil- und Hilfsmittel innerhalb der 10 Monate vor Interventionsbeginn sowie innerhalb der 10 Monate nach Interventionsbeginn analysiert. Diese Kosten sind Tabelle S12 im Anhang zu entnehmen.

Die AU-Tage geben die Anzahl an Tagen an, an denen die Versicherten bei der Krankenkasse als arbeitsunfähig gemeldet waren. Die durchschnittlichen AU-Tage vor Interventionsbeginn (t0) betragen 66.5 (SD 60.9) Tage in der EXP- und 69.3 (SD 65.2) Tage in der TAU-Gruppe. Seit Interventionsbeginn ist die Anzahl der AU-Tage in beiden Gruppen gestiegen: EXP 125.8 (122.7) Tage und TAU M 125.5 (SD118.5) Tage. Für die AU-Tage ergab sich in der Varianzanalyse für beide Gruppen ein signifikanter Zeiteffekt ($p < 0.001$), jedoch kein signifikanter Interaktions- (Gruppe x Zeit, $p = 0.809$) oder Gruppeneffekt ($p = 0.893$).

Hinsichtlich der Inanspruchnahme von ambulanten Leistungen konnte im Mittel in beiden Gruppen eine Reduktion von ca. 12 Facharztkontakten vor Interventionsbeginn auf ca. 11 nach Beginn der Intervention festgestellt werden. Die dominierende Fachgruppe ist aufgrund vieler fehlender Angaben nicht eindeutig zu benennen. Die Kosten für ambulante Leistungen waren sowohl vor als auch nach Interventionsbeginn etwas höher in der EXP-, als in der TAU- Gruppe (nicht signifikant), zeigten jedoch keine Veränderung ab t0.

Insgesamt wurden bei 10.1% (EXP) bzw. bei 14.5% (TAU) der Probanden vor Interventionsbeginn stationäre Aufenthalte verzeichnet. Diese Probanden, welche stationäre Aufenthalte aufwiesen, verbrachten durchschnittlich 14.8 (SD 23.8) Tage (EXP) bzw. 15.7 (SD 20.2) Tage (TAU) im Krankenhaus. Seit Interventionsbeginn wurden 10.6% (EXP) bzw. 16.9% (TAU) der Probanden stationär aufgenommen. Die durchschnittliche Anzahl der stationären Tage betrug 10.8 (SD 15.1) Tage in der EXP- bzw. 24.2 (SD 30.1) Tage in der TAU-Gruppe. In beiden Gruppen stiegen die Kosten

für stationäre Aufenthalte nach Interventionsbeginn leicht an, wobei die Varianzanalyse weder einen signifikanten Zeiteffekt ($p=0.076$), noch einen signifikanten Interaktions- ($p=0.249$) oder Gruppeneffekt ($p=0.114$) ergab.

Bei der Auswertung der Arzneimittel zeigt sich, dass Probanden in beiden Gruppen vor Interventionsbeginn durchschnittlich etwa 6 und in den 10 Monaten nach t_0 etwa 7-mal Medikamente verordnet wurden. Die Medikamentenkosten zeigten keine Veränderung über die Zeit und keinen Unterschied zwischen den Gruppen.

Hinsichtlich Heil- und Hilfsmittel konnte festgestellt werden, dass die Anzahl der Verordnungen in beiden Gruppen sowohl vor (Min. 0; Max. 14 [EXP] bzw. 10 [TAU]) als auch nach Interventionsbeginn (Min. 0; Max. 12 [EXP] bzw. 8 [TAU]) durchschnittlich etwa 1 betrug. Die Kosten für Heil- und Hilfsmittel zeigten keine signifikante Veränderung über die Zeit und keinen signifikanten Unterschied zwischen den Gruppen.

Difference in difference (DiD) GLMM

Mithilfe eines Difference in difference (DiD)-Ansatzes wurde die Wirkung der Intervention auf Gesamtkosten (ambulante Kosten, stationäre Kosten, Medikamentenkosten sowie Kosten für Heil- und Hilfsmittel) zwischen der EXP- und der TAU-Gruppe untersucht. Der Effekt wurde mit gemischten Modellen (GLMM) in SPSS modelliert. Es wurde "Treatment" als Dummy-Variable erstellt, welche den Effekt der STEP.De Sporttherapiegruppe kodiert (EXP-Gruppe (=1) und TAU-Gruppe (=0)); "Post period" als Dummy-Variable, die den Effekt für den Zeitraum nach der Intervention angibt (10 Monate vor der Intervention (=0) und 10 Monate nach der Intervention (=1)). Der Interaktionsterm (Behandlung*Post) spiegelt somit den DiD-Parameter wider, der die Wirkung der STEP.De Sporttherapie auf die Veränderung der Gesundheitskosten angibt. Weiterhin wurde für Clusterzugehörigkeit (Psychotherapeutenpaare/triplets) kontrolliert und die Interaktion "Behandlung*Cluster" in das Modell aufgenommen. Die IDs der Teilnehmenden wurden in den Modellen als Zufallseffekt behandelt.

Das DiD-Modell zeigte, dass die Zugehörigkeit zur EXP-Gruppe statistisch signifikante Auswirkungen auf die Höhe der Gesamtkosten einschließlich der Interventionskosten (10 Monate Sport- und Psychotherapiekosten) hatte (1232.58 € [95% CI 214.39 € bis 2250.77 €]; $p < 0,018$; Tabelle S14). Der in diesem gemischten Modell dargestellte Effekt entsteht durch einen Anstieg der Gesamtkosten einschließlich der Interventionskosten (10 Monate Sport- und Psychotherapiekosten) von M 2246.80 € (SD 5301.59 €) auf M 5119.95 € (SD 5762.09 €) in der TAU Gruppe bzw. von M 2065.77 € (SD 5252.81 €) auf M 3674.07 € (SD 7304.30 €) in der EXP Gruppe. Somit zeigen sich im Vergleich zur Psychotherapiegruppe in der STEP.De Sporttherapiegruppe um M 1232.58 € (SE = 518.51, $p < .05$) geringere Gesamtkosten einschließlich der Interventionskosten (10 Monate Sport- und Psychotherapiekosten).

In einer Sensitivitätsanalyse untersuchten wir darüber hinaus den Effekt auf die Gesamtkosten ohne Interventionskosten und anschließende Regelpsychotherapie. Das DiD-Modell zeigte hier, dass die Zugehörigkeit zur EXP-Gruppe keine statistisch signifikanten Auswirkungen auf die Höhe der Gesamtkosten (ohne Interventionskosten und anschließende Regelpsychotherapie) hatte ($p = 0,726$; Anhang C, Tabelle S15).

3.3 Prozessevaluation

Für die Prozessevaluation wurden in einem ersten Schritt die individuellen Verläufe der Studien- Probanden über den gesamten Therapiezeitraum (16 Wochen) analysiert. Es zeigte sich, dass ein Großteil der Verläufe grundsätzlich positiv (d. h. durch tendenziell abfallende depressive Symptomatik) gekennzeichnet war, wohingegen ca. ein Drittel der Probanden in beiden Gruppen eine insgesamt Symptomverschlechterung berichtete (tendenzieller Anstieg der depressiven Symptome: 33% in der STEP.De Sporttherapie, 29% in der Psychotherapie (siehe Abbildung S9). Hinsichtlich Studienabbruch ist der Anteil an Probanden innerhalb der Prozess-Evaluation zwar

höher in der STEP.De Sporttherapie (16%) als in der Psychotherapie (9%), aber nicht signifikant (X^2 -Test $p=.25$).

Um in einem zweiten Schritt die Veränderungsprozesse detaillierter zu analysieren, wurden die individuellen Verläufe (an dieser Stelle begrenzt auf Probanden, die einerseits bis zum Ende an der Studie teilnahmen und andererseits tendenziell positiv auf die Therapie reagierten, d.h. einen abfallenden Trend aufwiesen, s. o.) mithilfe individueller Verlaufskurven modelliert (Anhang B, Abbildung S10). Tatsächlich zeigte sich ein signifikanter Unterschied in den Modellparametern der STEP.De Sporttherapie, der auf eine insgesamt schnellere Veränderung der kognitiven Symptome (Rumination) im Vergleich zu den depressiven Symptomen hinweist ($p=.003$ und $p<.001$ für die Modell-Parameter a und b); die Veränderungen innerhalb der Psychotherapiegruppe sahen insgesamt ähnlich aus, unterschieden sich aber nicht signifikant (alle $ps>.100$). Dies könnte einerseits auf einen Sport-spezifischen Mechanismus hinweisen, jedoch auch mit der geringeren Stichprobengröße und damit geringeren statischen Power in der TAU-Gruppen zusammenhängen (vgl. Gruppenzuweisung 2:1).

Der postulierte Wirkmechanismus von körperlicher Aktivität auf die Depression (s. Hypothesen) wurde mittels einer Mediationsanalyse basierend auf dynamischen (Lag-) Mehrebenen-Modellen überprüft (Anhang B, Abbildung S11). Die Ergebnisse sind in Abbildung S11 im Anhang B dargestellt: alle Pfade weisen einen signifikanten ($p<.05$) Koeffizienten im Bereich von .04 bis .13 auf; entsprechend der Kodierung der Variablen sind die Effekte entweder positiv (z. B. zwischen körperlicher Aktivität und Kognitiver Kontrolle oder zwischen Rumination und Depression) oder negativ (z. B. zwischen Körperlicher Aktivität und Depression oder zwischen Kognitiver Kontrolle und Rumination). Das Ergebnis der Mediationsanalyse, bei welcher der sog. indirekte Effekt (von körperlicher Aktivität über Kognitive Kontrolle und/oder Rumination auf Depression) mit dem sog. direkten Effekt (ohne Umweg von der körperlichen Aktivität auf Depression) verglichen wurde, spricht für eine partielle Mediation, da beide Effekte signifikant sind (alle $ps<.05$).

Für eine weitere Ausdifferenzierung der zentralen Variablen „Bewegung“ und „kognitive Kontrolle“ wurden diese jeweils zusätzlich mit einer gesonderten Methode erhoben. Mithilfe von Bewegungssensoren (Akzelerometrie) konnten zeitlich begrenzte Phasen von erhöhter körperlicher Aktivität im Alltag identifiziert werden (z. B. auf dem Weg zum Einkaufen, Radfahren, sowie eigenständiger Sport) und mit den fortlaufend durchgeführten Ratings der Probanden in Verbindung gesetzt werden. Nach Standardisierung (am individuellen Mittelwert und der individuellen Standardabweichung), ergab sich durch Übereinanderlegen dieser Zeitfenster von allen Probanden ein mittlerer Anstieg der positiven Emotionsratings (z. B. Freude), sowie ein Rückgang negativer Emotionen (z. B. Traurigkeit) über mehrere Stunden nach der Aktivität (siehe Anhang B, Abbildung S12). In ähnlicher Weise ging die Phase körperlicher Aktivität mit einer nachfolgenden Phase subjektiv verbesserter kognitiver Kontrolle und weniger Grübeln einher.

Da insbesondere die eigene momentane kognitive Kontrolle nicht ohne Schwierigkeiten selbst einzuschätzen ist, wurde hier ebenfalls ein ergänzendes Verfahren zur Erfassung eingesetzt. In wöchentlich stattfindenden Tests, bei denen die Probanden möglichst schnell und akkurat auf visuelle Reize auf dem Smartphone reagieren sollten, wurde die Veränderung der kognitiven Kontrolle im Alltag und über die Wochen der Studienteilnahme erhoben. Die Ergebnisse zeigten einen generellen Trend zu besserer kognitiver Kontrolle bei den Probanden (z. B. weniger Fehler, bessere Genauigkeit, schnellere Reaktionen) über die Wochen der Studienteilnahme (Anhang B, Abbildung S1).

4. Schlussfolgerungen und Empfehlungen des Evaluators

Die neue Versorgungsform „STEP.De Sporttherapie bei Depression“ beinhaltet die folgenden Qualitätskriterien als theoriegeleitetes und innovatives Versorgungskonzept:

- Leitliniengerechte Behandlung bei leichter bis mittlerer Depression
- Evidenzbasierte Sporttherapie

- Unter Anleitung von geschulten und qualifizierten Sporttherapeuten
- Begleitung und Unterstützung durch Psychotherapeuten

Das Ziel der Evaluation war der Nachweis der STEP.De Sporttherapie bei Depression als nicht unterlegene bzw. gleichwertige Behandlungsoption gegenüber der Regelversorgung (Psychotherapie) in Bezug auf Effektivität, Nachhaltigkeit und Kosten.

4.1 Schlussfolgerungen des Evaluators

Die STEP.De Sporttherapie bei Depression Studie als ergänzende therapeutische Leistung zeigte sich als wichtige und gleichwertige Ergänzung zur bestehenden Psychotherapie in der Regelversorgung.

Innerhalb der vierjährigen Studienphase konnte der wissenschaftliche Nachweis zu Umfang und Nachhaltigkeit der Wirksamkeit auf die Depressionssymptomatik, Arbeits- und Soziale Anpassungsfähigkeit sowie Kosteneffizienz erbracht werden.

Im Einklang mit vorherigen Forschungsergebnissen, meta-analytischen Studien sowie Leitlinien und Richtlinien konnte ein positiver Effekt auf die Reduktion der Depressionssymptomatik durch die spezifische STEP.De Sporttherapieintervention nachgewiesen werden.

Insgesamt wurden 393 Teilnehmer der neuen Versorgungsform oder der Regelversorgung zugewiesen. Die Drop-out Rate (s. 4.1.1) befand sich im erwarteten Bereich und wies keine Unterschiede in Bezug auf die Gruppen auf. Als einzige Variable wies das Alter zum ersten Erhebungszeitpunkt einen Gruppenunterschied auf und wurde daher in den Analysen berücksichtigt.

Die Höhe der gemessenen signifikanten Reduktion des primären Endpunktes (Schweregrad der Depression) von 7.4 Punkten in der STEP.De Sporttherapiegruppe und von 7.5 Punkten in der Psychotherapiegruppe) auf dem BDI-II ist in Orientierung an internationalen Referenzwerten (National Institute for Health and Care Excellence, 2018) als mittlere Veränderung mit deutlicher klinischer Relevanz zu klassifizieren (eine Reduktion um >3 Punkte wird als klinisch relevant angegeben) und stellt einen besonders zu würdigenden Interventionseffekt dar. Der Schweregrad der depressiven Symptomatik reduziert sich von einer durchschnittlich mittelgradigen Depression zu Beginn der Interventionen auf eine leichte bis minimale Depression nach Beendigung der Interventionen für beide Gruppen.

Darüber hinaus konnte mit dem Nicht-Unterlegenheits-Design gezeigt werden, dass die STEP.De Sporttherapie bei Depression der leitliniengerechten Erstlinientherapie (Psychotherapie) sowohl nach dem 16-wöchigen Interventionszeitraum als auch im 6-monatigen Nachbeobachtungszeitraum nicht unterlegen war und damit eine gleichwertige Alternative in der Behandlung leichter bis mittlerer Depression zur Psychotherapie darstellen kann.

Auch in Bezug auf die wahrgenommene Arbeits- und soziale Anpassungsfähigkeit und die (gesundheitsbezogene) Lebensqualität weist die EXP-Gruppe signifikant positive Veränderungen über die Zeit sowohl nach der Intervention als auch im 6-monatigen Nachbeobachtungszeitraum auf, wobei keine Unterschiede zur TAU-Gruppe zu verzeichnen sind.

Für die gesundheitsökonomische Analyse zeigen sich für die Kosten der Intervention selbst um etwa 300€ niedrigere Kosten bei der STEP.De Sporttherapiegruppe im Interventionszeitraum. Zusammen mit den Kosten für weitere Psychotherapie in den sechs Monaten nach Abschluss der STEP.De Sporttherapie, die auch Kosten für nach der STEP.De Sporttherapie begonnene Regelpsychotherapie beinhalten, eine Reduktion der Kosten in der STEP.De Sporttherapiegruppe um etwa 1200€. Dieser Unterschied zeigt sich auch als signifikanter Unterschied in den Gesamtkosten inklusive der Sporttherapie- und Psychotherapiekosten. Werden die Interventions- und Psychotherapiekosten aber aus der Analyse herausgenommen, unterscheiden sich die übrigen Gesamtkosten für alle Kostenarten darüber hinaus nicht.

Folgende Limitationen müssen angemerkt werden: Die Randomisierung erfolgte aufgrund des

Studiendesigns im Versorgungsforschungskontext bereits zu einem recht frühen Zeitpunkt, dadurch wurden im weiteren Verlauf z. B. bei dem Psychotherapeuten bei Diagnoseabsicherung weitere Patienten aus der Studie ausgeschlossen von denen aufgrund der fehlenden Einwilligungserklärung zur Datenerhebung und -verarbeitung keine Daten vorliegen, wodurch sich durch einen hohen prä-Interventions-Dropout eine stärker selektierte finale Stichprobe ergab, bei der sich Drop-outs auch durch die Präferenz der Teilnehmer ergeben haben könnten (siehe Flow Chart). Ein weiterer prä-Interventions-Drop-out ergab sich durch den Einschluss von Patienten aufgrund der Einschlussdiagnosen durch den behandelnden Arzt, die nicht immer durch den Psychotherapeuten beim Erstgespräch bestätigt wurden bzw. weitere Ausschlussdiagnosen zur Anwendung kamen (z. B. Abhängigkeitserkrankungen). Insgesamt konnten so trotz des Erreichens der Einschlusszahlen zu Beginn der Studie zum Zeitpunkt nach der Intervention für die mITT Analyse nur 259 Probanden erhoben werden. Dies entspricht 83% der ursprünglich angestrebten Fallzahl von 312. Die für die Überprüfung der Interventionseffekte auf das primäre Zielkriterium verwendeten verallgemeinerten lineare Schätzungsgleichungen (GEE) gelten aber für längsschnittlichen Dropout als robust (Lin & Rodriguez, 2015).

Des Weiteren sicherte der Psychotherapeut die Diagnose zwar entsprechend der Einschlussdiagnosen ab, war instruiert diese anhand des affektiven Moduls des SKID-II abzusichern und bestätigte therapeutischen Behandlungsbedarf, trotzdem ergibt sich zu Beginn der Interventionen, dass ca. 20% der Teilnehmer einen BDI Wert von unter 14 Punkten aufweisen, was auf eine minimale bis keine Depression hinweist und damit auch eine weitere Reduktion der Symptomatik limitiert.

Für die Interpretierbarkeit und Anwendbarkeit auf die Gesamtbevölkerung sind diese Einschränkungen zu berücksichtigen, jedoch auch aus der Perspektive Versorgungsforschung und Versorgungsrealität z. B. in Bezug auf die Präferenz der Behandlung zu sehen.

Als besonders hilfreich und qualitativ hochwertig muss die psychologische Schulung der Sporttherapeuten und Schulung der Psychotherapeuten sowie die individuelle Betreuung durch die Sporttherapeuten in der Gruppe und die Begleitung durch die Psychotherapeuten angesehen werden.

Die psychologische Schulung der Sporttherapeuten befähigte diese, die Teilnehmenden in Bezug auf u.a. Krisen- und Notfallsituationen, gesundheitsfördernde Kommunikation und Unterstützung der psychologischen Grundbedürfnisse kompetent zu begleiten.

Die Psychotherapeuten wurden durch die Schulung befähigt, die STEP.De Sporttherapie als adäquate Behandlungsmöglichkeit in Absicherung der Diagnose durch Anwendung der Ein- und Ausschlusskriterien im Interventionsgespräch einzuschätzen. Des Weiteren begleiteten sie durch telefonische Termine sowie der Erörterung von möglichem weiteren Behandlungsbedarf im Nachsorgegespräch.

Durch die Sporteingangstestung konnten Ängste und Bedenken der Teilnehmer aufgenommen und adressiert werden sowie physiologische Indikationen anhand von u. a. Anamnese und sportphysiologischen Testungen mit in die Trainingsplanung einbezogen und durch das Zielsetzungsgespräch durch den Sporttherapeut individuell auf die Bedürfnisse des Teilnehmers abgestimmt werden.

Die individuelle Betreuung in der Gruppe durch die Sporttherapeuten in den STEP.De Sporttherapieeinheiten ermöglichte sowohl zielgerichtete Sporttherapie als auch den Austausch mit anderen Teilnehmenden und dem Sporttherapeuten.

Im Abschlussgespräch berichteten die Teilnehmer über Schwierigkeiten und Erfolge und entwickelten mit Hilfe des Sporttherapeuten konkrete Pläne, wie körperliche Aktivität und Bewegung sowie weitere gesundheitsfördernde Maßnahmen langfristig in ihren Alltag integriert werden können.

Zusammen genommen können diese Kernbestandteile des STEP.De Programms als ausschlaggebende Faktoren für die erfolgreiche Umsetzung der innovativen Versorgungsform angenommen werden.

Bei der Terminvereinbarung mit den begleitenden Psychotherapeuten der EXP-Gruppe wurden immer wieder Engpässe gemeldet, da diese mit regulären Patienten außerhalb der Studie, die eine Psychotherapie beanspruchten, für längere Zeiträume blockiert waren.

Grundsätzlich ist der zeitliche Aufwand der Begleitung durch einen Psychotherapeuten als gering einzuschätzen (2 Termine plus 3 Telefontermine). Die Begleitung im Online-Verfahren bzw. durch psychotherapeutische Zentren, die einen Schwerpunkt auf das STEP.De Programm legen, könnte möglicherweise Abhilfe schaffen.

Damit besteht der Vorteil des STEP.De Programms in der schnellen Versorgung innerhalb von i. d. R. 7-14 Tagen der Patienten. Einen Nachweis für diesen Vorteil konnte die Studie aufgrund des Designs nicht erbringen (die Einsteuerung in STEP.De Sporttherapie- und Psychotherapiegruppe erfolgte im gleichen zeitlichen Rahmen), wird aber in der Regelversorgung angesichts der deutlich längeren Wartezeiten auf einen Psychotherapieplatz als gegeben angenommen.

Hinsichtlich der Fragestellungen der Prozessevaluation konnten folgende Schlussfolgerungen gezogen werden: In Ergänzung zu den Prä-Post Effekten mittels Fragebögen (z. B. BDI, s. o.) wiesen auch die zeitlich hoch-aufgelösten Symptoms-Ratings (mittels Smartphone-App) einen überwiegend positiven Trend im Therapieverlauf auf. Der Anteil von Probanden mit negativem Verlauf oder Studien-Abbruch unterschied sich nicht signifikant zwischen den Gruppen. Jedoch zeigte sich ein differentieller Effekt in der Geschwindigkeit einzelner Variablen: nur in der Gruppe der STEP.De Sporttherapie-Probanden verbesserten sich die kognitiven Symptome (Grübeln) signifikant schneller als die depressiven Hauptsymptome (Stimmung, Antrieb, Anhedonie). In einem Mediationsmodell konnte schließlich der medierende Effekt von Kognitiver Kontrolle und Grübeln im Zusammenhang des positiven Effekts von körperlicher Aktivität auf die depressiven Symptome als partielle Mediation bestätigt werden. Entsprechend der Hypothese bedeutet dies, dass der Wirkmechanismus von Sport bei Depression zumindest teilweise auf eine initiale Verbesserung allgemeiner kognitiver Ressourcen baut. Diese erlaubt es ihrerseits den Betroffenen, das dysfunktionale Grübeln zu reduzieren und somit Verbesserungen der depressiven Symptomatik zu erreichen. Zukünftige Studien können überprüfen, ob die besondere Rolle der Kognitiven Kontrolle im Rahmen von Sporttherapie auch als Ansatzpunkt für Indikationsstellungen genutzt werden kann oder muss: Profitieren Menschen mit ausgeprägten Defiziten der Kognitiven Kontrolle in besonderem Maße von Sporttherapie? Oder ist ein anfängliches Mindestmaß an kognitiver Kontrolle erforderlich, um positive Entwicklungen im Rahmen von Sporttherapie zu initiieren?

4.2 Empfehlungen des Evaluators

Ein Großteil der unterstützenden Evidenz für Sporttherapie bei Depression stammt aus streng kontrollierten Wirksamkeitsstudien unter idealen, artifiziellen Bedingungen (Phase III-Studien) (Schuch, Morres, Ekkekakis, Rosenbaum & Stubbs, 2017). Dies ist nicht ausreichend, um Ärzte und politische Entscheidungsträger bei der Wahl der optimalen Behandlung für Patienten hinreichend unterstützen zu können. Daher ist es wichtig, methodisch hochwertige, pragmatische RCTs durchzuführen, um die Wirksamkeit und Kosteneffizienz von Sporttherapie in der Regelversorgung zu untersuchen. Sie füllt die Umsetzungslücke zwischen Evidenz und Praxis und liefert entscheidende Informationen über die Generalisierbarkeit der Studienbefunde und damit eine adäquate Grundlage für praktische Entscheidungen in natürlichen Anwendungskontexten (Heissel et al., 2020).

Aufgrund dessen zielte die STEP.De Studie darauf ab, die Wirksamkeit und den Nutzen sowie die Kosteneffizienz von STEP.De Sporttherapie im Vergleich zu Psychotherapie unter realen Bedingungen (Phase IV-Studien) und somit in der klinischen Routine zu untersuchen. Damit geht die Notwendigkeit einher, auf sich ändernde Umgebungsbedingungen wie bspw. das aktuelle

pandemische Geschehen flexibel zu reagieren und entsprechende Anpassungen vorzunehmen.

Insgesamt konnte in der Evaluation gezeigt werden, dass die neue Versorgungsform STEP.De zur Verbesserung der Behandlung von Patienten mit leichter bis mittlerer Depression in einer Cluster-randomisierten kontrollierten Nicht-Unterlegenheits-Studie beiträgt, vergleichbar mit der Standardbehandlung.

Vor dem Hintergrund von u. a. bestehenden Versorgungsengpässen, langen Wartezeiten auf bestehende Versorgungsleistungen, der notwendigen Senkung von Chronifizierungsraten, den krisenbedingten Anstiegen der Prävalenz von Depression sowie der Entstigmatisierung von Psychotherapie stellt die evaluierte neue Versorgungsform STEP.De eine notwendige Behandlungsoption in der Versorgung dar.

Die für die Implementierung gewählte Angliederung der neuen Versorgungsform an im Rahmen des Projektes genutzten und weiterhin bestehenden selektivvertraglichen Strukturen nach § 140a SGB V erscheint im ersten Schritt grundsätzlich geeignet, da hierdurch die Fortführung des Programms sichergestellt werden kann, bis die Implementierung in die Regelversorgung vollumfänglich abgeschlossen ist.

Nicht nur erbrachte die neue Versorgungsform ihren Wirksamkeitsnachweis in Bezug auf die Primär- und Sekundärdatenanalyse (Depressionssymptomatik, Arbeitsfähigkeit und Lebensqualität), gleichzeitig zeigte sich bei gesundheitsökonomischer Betrachtung eine Verringerung der Gesamtkosten verglichen mit der Regelversorgung, die insbesondere aus kürzeren psychotherapeutischen Behandlungszeiträumen und niedrigeren Kosten im Bereich Sport- und Psychotherapie resultieren. Überdies konnte eine geringere Rate an verschriebenen Arzneimitteln beobachtet werden. Aus Perspektive der Patienten und Kostenträger sind die beobachteten Effekte als wünschenswert zu bezeichnen. Daneben wurden auch Limitationen der neuen Versorgungsform identifiziert. Eine Gesamtkostenreduktion über die psychotherapeutischen Behandlungskosten hinaus konnten nicht beobachtet werden. Lediglich die Interventionskosten selbst, sowie die Kosten für eine nach Interventionsabschluss weitergeführte bzw. begonnene Psychotherapie treiben bei den Gesundheitsökonomischen Analysen im gemischten Modell den Kostenunterschied. Dieser liegt bei etwa 1000€ pro Patient.

Die nachhaltige Umsetzung der STEP.De Intervention in die Regelversorgung stellt in der Regel eine besondere Herausforderung dar. Sie erfordert ein flächendeckendes Netz an geschulten und zertifizierten Sport- und Psychotherapeuten und damit einhergehend die Infrastruktur für (Nach-) Schulungen und (Re-) Zertifizierungen der Sport- und Psychotherapeuten möglichst im Onlineformat, um Engpässe zu vermeiden. Darüber hinaus sind (Online-) Kurzschulungen zum STEP.De Programm für Ärzte, Psychotherapeuten, Sportwissenschaftler, Fallmanager und weitere Krankenkassenmitarbeitende sowie sonstige Berater für die Erstberatung von Interessierten über online verfügbare Informationen hinaus wünschenswert. Die Prüfung der Zugangsvoraussetzungen der Sport- und Psychotherapeuten sowie Anbietern und Prüfungen zu den Schulungsinhalten stellen einen wichtigen Aspekt der Qualitätssicherung des STEP.De Programms dar. Darüber hinaus wird eine zentrale koordinierende Stelle für die Außendarstellung (z. B. Webpräsenz) und Vernetzungsaktivitäten zwischen den Leistungserbringern und Betroffenen benötigt.

Mit Verweis auf die große Zahl unbehandelter Menschen, die an Depression leiden, einhergehend mit Engpässen in der Versorgung durch u. a. lange Wartezeiten auf Psychotherapieplätze sowie mögliche Chronifizierung, Entwicklung von Komorbiditäten und Verschlechterung der Symptomatik und den damit verbundenen Kosten für das Gesundheitssystem, erhalten die Ergebnisse eine besonders hohe Relevanz und Dringlichkeit in Bezug auf die praktische Umsetzung.

Zusammenfassend zeigt STEP.De sowohl nach der Intervention als auch im Nachbeobachtungszeitraum 6 Monate später eine Verbesserung der wahrgenommenen Arbeits- und sozialen Anpassungsfähigkeit sowie der (gesundheitsbezogenen) Lebensqualität. Somit stellt sie im Hinblick auf die klinisch relevante Verbesserung der Depressionssymptomatik eine gleichwertige und im Versorgungskontext erprobte, evidenzbasierte Behandlungsoption von Patienten mit leichter bis mittlerer Depression dar. In Bezug auf die Reduzierung der Kosten der

Akronym: STEP.De

Förderkennzeichen: 01NVF17050

Intervention (Interventionskosten selbst, sowie die Kosten für eine nach Interventionsabschluss weitergeführte bzw. begonnene Psychotherapie) im Vergleich zur Standardversorgung ist das STEP.De Programm überlegen.

Wir empfehlen das STEP.De Programm als Versorgungsform, zuerst weiterführend im Rahmen von Selektivverträgen und folgend vollumfänglich in die Versorgung zu überführen.

Aufgrund der vorhandenen Evidenz von Sport in der Behandlung von weiteren psychischen Erkrankungen (u. a. Angsterkrankungen, PTSB, Burnout, Insomnia) ist zu prüfen, ob eine Erweiterung der Diagnosen für das STEP.De Programm adäquat erscheint.

5. Literaturverzeichnis

- American Heart Association. (2018). *American Heart Association Recommendations for Physical Activity in Adults and Kids*. <https://www.heart.org/en/healthy-living/fitness/fitness-basics/aha-recs-for-physical-activity-in-adults>
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: Fifth edition (DSM-5)*.
- Ballenger, J. C. (1999). Clinical guidelines for establishing remission in patients with depression and anxiety. *Journal of Clinical Psychiatry*, *60*, 29–34.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II: Beck depression inventory*. Pearson.
- Birmaher, B., & Brent, D. (1998). Practice parameters for the assessment and treatment of children and adolescents with depressive disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*(10), 635–835.
- Birnbaum, H. G., Kessler, R. C., Kelley, D., Ben-Hamadi, R., Joish, V. N., & Greenberg, P. E. (2010). Employer burden of mild, moderate, and severe major depressive disorder: Mental health services utilization and costs, and work performance. *Depression and Anxiety*, *27*(1), 78–89. <https://doi.org/10.1002/da.20580>
- Bower, P., Byford, S., Sibbald, B., Ward, E., King, M., Lloyd, M., & Gabbay, M. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression. II: Cost effectiveness. *BMJ*, *321*(7273), 1389–1392. <https://doi.org/10.1136/bmj.321.7273.1389>
- Bundespsychotherapeutenkammer. (2018). *Wartezeiten für psychisch kranke Menschen weiterhin zu lang: BPTK zum G-BA-Gutachten zur Bedarfsplanung*. <https://www.bptk.de/wartezeiten-fuer-psychisch-krank-menschen-weiterhin-zu-lang/>
- 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*. <https://doi.org/10.1002/14651858.CD004366.pub6>
- Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica*, *144*(3), 288–299. <https://doi.org/10.1111/acps.13335>
- De Maat, S., Dekker, J., Schoevers, R., & De Jonghe, F. (2006). Relative efficacy of psychotherapy and combined therapy in the treatment of depression: A meta-analysis. *Psychotherapy Research*, *16*(5), 566–578. <https://doi.org/10.1080/10503300600756402>
- Deutsche Psychotherapeuten Vereinigung (DPTV). (2021). *Report Psychotherapie 2021*. https://www.dptv.de/fileadmin/Redaktion/Bilder_und_Dokumente/Wissensdatenbank_oeffentlich/Report_Psychotherapie/DPTV_Report_Psychotherapie_2021.pdf
- Fewtrell, M. S., Kennedy, K., Singhal, A., Martin, R. M., Ness, A., Hadders-Algra, M., Koletzko, B., & Lucas, A. (2008). *How much loss to follow-up is acceptable in long-term randomised trials and prospective studies?* *93*(6), 4.
- Gandek, B., Ware, J. E., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, J. E., Bullinger, M., Kaasa, S., Leplege, A., Prieto, L., & Sullivan, M. (1998). Cross-Validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries: Results from the IQOLA Project. *Journal of Clinical Epidemiology*, *51*(11), 1171–1178. [https://doi.org/10.1016/S0895-4356\(98\)00109-7](https://doi.org/10.1016/S0895-4356(98)00109-7)
- Haby, M. M., Tonge, B., Littlefield, L., Carter, R., & Vos, T. (2004). *Cost-Effectiveness of Cognitive Behavioural Therapy and Selective Serotonin Reuptake Inhibitors for Major Depression in Children and Adolescents*. *38*(8), 579–591.
- Hamilton, M. (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry*, *23*, 56–62.
- Hardy, G. E., & Llewelyn, S. (2015). Introduction to Psychotherapy Process Research. In O. C. G. Gelo, A. Pritz, & B. Rieken (Hrsg.), *Psychotherapy Research* (S. 183–194). Springer Vienna. https://doi.org/10.1007/978-3-7091-1382-0_9
- Hautzinger, M., Keller, F., & Kühner, C. (2009). *Das Beck Depressions Inventar II: Deutsche Bearbeitung und Handbuch zum BDI II*. Harcourt Test Services. <https://www.testzentrale.de/shop/beck-depressions-inventar-revision.html>

- Hautzinger, M., Keller, F., Kühner, C., & Beck, A. T. (2009). *Beck depressions-Inventar: BDI II; manual*. Pearson Assessment.
- Heinzel, S., Lawrence, J. B., Kallies, G., Rapp, M., & Heissel, A. (2015). Using exercise to fight depression in older adults. *GeroPsych*.
- Heissel, A., Bollmann, J., Kangas, M., Abdulla, K., Rapp, M., & Sanchez, A. (2021). Validation of the German version of the work and social adjustment scale in a sample of depressed patients. *BMC Health Services Research*, *21*(1), 1–11.
- Heissel, A., Pietrek, A., Flunger, B., Fydrich, T., Rapp, M., Heinzel, S., & Vansteenkiste, M. (2019). The validation of the German basic psychological need satisfaction and frustration scale in the context of mental health. *European Journal of Health Psychology*, *25*(4), 119–132. <https://doi.org/10.1027/2512-8442/a000017>
- Heissel, A., Pietrek, A., Schwefel, M., Abula, K., Wilbertz, G., Heinzel, S., & Rapp, M. (2020). STEP.De study—A multicentre cluster-- randomised effectiveness trial of exercise therapy for patients with depressive symptoms in healthcare services: Study protocol. *BMJ Open*, *10*, e036287. <https://doi.org/10.1136/bmjopen-2019-036287>
- Heißel, A., Stielow, S., Heinen, D., Bergau, T., Sanchez, A., Pietrek, A., Schwefel, M., Wilbertz, G., Rapp, M., & Heinzel, S. (in prep.). *The STEP.De study – Exercise therapy by qualified exercise professionals’ proofs to be as effective as psychotherapy in depressed patients in health care: A randomized controlled trial. *equal contribution*.
- Herdman, M., Gudex, C., Lloyd, A., Janssen, MF., Kind, P., Parkin, D., Bonnel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736. <https://doi.org/10.1007/s11136-011-9903-x>
- Holma, K. M., Holma, I. A., Melartin, T. K., Rytälä, H. J., & Isometsä, E. T. (2008). Long-term outcome of major depressive disorder in psychiatric patients is variable. *The Journal of clinical psychiatry*, *69*(2), 196–205.
- Jacobi, F., Wittchen, H.-U., Höltling, C., Höfler, M., Pfister, H., Müller, N., & Lieb, R. (2004). Prevalence, co-morbidity and correlates of mental disorders in the general population: Results from the German Health Interview and Examination Survey (GHS). *Psychological Medicine*, *34*(4), 597–611. <https://doi.org/10.1017/S0033291703001399>
- Kammerer, K., Falk, K., Heintze, C., Döpfner, S., & Heusinger, J. (2019). Hindernisse und Voraussetzungen für die Vermittlung älterer, depressiv erkrankter Menschen in Psychotherapie aus Sicht von HausärztInnen. *Das Gesundheitswesen*, *81*(01), 58–62. <http://dx.doi.org/10.1055/s-0042-116589>
- Kendra, M. S., Mohr, J. J., & Pollard, J. W. (2014). The stigma of having psychological problems: Relations with engagement, working alliance, and depression in psychotherapy. *Psychotherapy*, *51*(4), 563–573. <https://doi.org/10.1037/a0036586>
- Kessler, R. C. (2012). The Costs of Depression. *Psychiatric Clinics of North America*, *35*(1), 1–14. <https://doi.org/10.1016/j.psc.2011.11.005>
- Kirchberger, I., Braitmayer, K., Coenen, M., Oberhauser, C., & Meisinger, C. (2014). Feasibility and psychometric properties of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0) in a population-based sample of patients with myocardial infarction from the MONICA/KORA myocardial infarction registry. *Population Health Metrics*, *12*(1), 27. <https://doi.org/10.1186/s12963-014-0027-8>
- Konerding, U. (2013). What does Cronbach’s alpha tell us about the EQ-5D? A methodological commentary to “Psychometric properties of the EuroQol Five-Dimensional Questionnaire (EQ-5D-3L) in caregivers of autistic children”. *Quality of Life Research*, *22*(10), 2939–2940. <https://doi.org/10.1007/s11136-013-0430-9>
- Krogh, J., Hjorthøj, C., Speyer, H., Gluud, C., & Nordentoft, M. (2017). Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential analysis. *BMJ Open*, *7*(9), e014820.
- Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliabilität und Validität des revidierten Beck Depressionsinventars (BDI-II). *Nervenarzt*, *78*, 651–656. <https://doi.org/10.1007/s00115-006-2098-7>
- Lawlor, D. A., & Hopker, S. W. (2001). The effectiveness of exercise as an intervention in the

- management of depression: Systematic review and meta-regression analysis of randomised controlled trials. *BMJ*, 322(7289), 763–767.
<https://doi.org/10.1136/bmj.322.7289.763>
- Lin, G., & Rodriguez, R. N. (2015). Weighted Methods for Analyzing Missing Data with the GEE Procedure. *Paper SAS 166-2015*, 1–8.
- Maier, W., & Philipp, M. (2013). *Reliabilität und Validität der Subtypisierung und Schweregradmessung depressiver Syndrome*. Springer-Verlag.
- Maljanen, T., Knekt, P., Lindfors, O., Virtala, E., Tillmann, P., Härkänen, T., & The Helsinki Psychotherapy Study Group. (2016). The cost-effectiveness of short-term and long-term psychotherapy in the treatment of depressive and anxiety disorders during a 5-year follow-up. *Journal of Affective Disorders*, 190, 254–263.
<https://doi.org/10.1016/j.jad.2015.09.065>
- Marks, I. M. (1986). *Behavioural Psychotherapy: Maudsley Pocket Book of Clinical Management*. Wright.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002a). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, 180(5), 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002b). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, 180(5), 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- National Collaborating Centre for Mental Health. (2010). *Depression: The treatment and management of depression in adults (updated edition)*.
- National Institute for Health and Care Excellence. (2018). *Overview | Depression in adults: Recognition and management | Guidance | NICE*.
<https://www.nice.org.uk/guidance/cg90/>
- Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical care*, 582–592.
- Ramana, R., Paykel, E., Cooper, Z., Hayhurst, H., Saxty, M., & Surtees, P. (1995). Remission and relapse in major depression: A two-year prospective follow-up study. *Psychological Medicine*, 25(6), 1161–1170. <https://doi.org/doi:10.1017/S0033291700033134>
- Rapp, M. A., Mell, T., Majic, T., Treusch, Y., Nordheim, J., Niemann-Mirmehdi, M., Gutzmann, H., & Heinz, A. (2013). Agitation in Nursing Home Residents With Dementia (VIDEANT Trial): Effects of a Cluster-Randomized, Controlled, Guideline Implementation Trial. *Journal of the American Medical Directors Association*, 14(9), 690–695.
<https://doi.org/10.1016/j.jamda.2013.05.017>
- Rapp, M., Heißel, A., Pietrek, A., Bergau, T., Heinen, D., Stielow, S., Sanchez, A., Schwefel, M., Wilbertz, G., & Heinz, S. (2019). *Evaluationskonzept Projekt STEP.De Sporttherapie bei Depression. Unveröffentlichtes Manuskript eingereicht beim Gemeinsamen Bundesausschuss. Zuletzt aktualisierte und bestätigte Version vom 26.05.2021*.
- Rethorst, C. D., Wipfli, B. M., & Landers, D. M. (2009). The Antidepressive Effects of Exercise: A Meta-Analysis of Randomized Trials. *Sports Medicine*, 39(6), 491–511.
<https://doi.org/10.2165/00007256-200939060-00004>
- Rhodes, S., Richards, D. A., Ekers, D., McMillan, D., Byford, S., Farrand, P. A., Gilbody, S., Hollon, S. D., Kuyken, W., Martell, C., O'Mahen, H. A., O'Neill, E., Reed, N., Taylor, R. S., Watkins, E. R., & Wright, K. A. (2014). Cost and outcome of behavioural activation versus cognitive behaviour therapy for depression (COBRA): Study protocol for a randomised controlled trial. *Trials*, 15, 29. <https://doi.org/10.1186/1745-6215-15-29>
- Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., Barrett, B., Farrand, P. A., Gilbody, S., Kuyken, W., O'Mahen, H., Watkins, E. R., Wright, K. A., Hollon, S. D., Reed, N., Rhodes, S., Fletcher, E., & Finning, K. (2016). Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): A randomised, controlled, non-inferiority trial. *Lancet (London, England)*, 388(10047), 871–880.
[https://doi.org/10.1016/S0140-6736\(16\)31140-0](https://doi.org/10.1016/S0140-6736(16)31140-0)

- Romppel, M., Herrmann-Lingen, C., Wachter, R., Edelmann, F., Düngen, H.-D., Pieske, B., & Grande, G. (2013). A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an intercultural non-clinical sample and a sample of patients at risk for heart failure. *GMS Psycho-Social-Medicine*, *10*, 1–7. <https://doi.org/10.3205/psm000091>
- Schneider, F., Härter, M., & Schorr, S. (Hrsg.). (2017). *S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression*. Springer-Verlag. <https://doi.org/10.1007/978-3-662-52906-5>
- Schuch, F. B., Morres, I. D., Ekkekakis, P., Rosenbaum, S., & Stubbs, B. (2017). A critical review of exercise as a treatment for clinically depressed adults: Time to get pragmatic. *Acta neuropsychiatrica*, *29*(2), 65–71.
- Schuch, F. B., Silva, E. S., Hallgren, M., & Fleck, M. P. (2018). Physical Activity and Incident Depression: A Meta-Analysis of Prospective Cohort Studies. *Am J Psychiatry*, *175*(7), 631–648. <https://doi.org/10.1176/appi.ajp.2018.17111194>
- Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., & Stubbs, B. (2016). Exercise as a treatment for depression: A meta-analysis adjusting for publication bias. *Journal of Psychiatric Research*, *77*, 42–51. <https://doi.org/10.1016/j.jpsychires.2016.02.023>
- Schulz, K. F., & Grimes, D. A. (2005). Sample size calculations in randomised trials: Mandatory and mystical. *The Lancet*, *365*(9467), 1348–1353. [https://doi.org/10.1016/S0140-6736\(05\)61034-3](https://doi.org/10.1016/S0140-6736(05)61034-3)
- Shinohara, K., Honyashiki, M., Imai, H., Hunot, V., Malta, D. M., Davies, P., Moore, T. H. M., Furukawa, T. A., & Churchill, R. (2013). Behavioural therapies versus other psychological therapies for depression. *Cochrane Database Systematic Reviews*, *10*. <https://doi.org/10.1002/14651858.CD008696.pub2>
- Stanton, R. (2014). Exercise and the treatment of depression: A review of the exercise program variables. *Journal of Science and Medicine in Sport*, *6*.
- Stubbs, B., Vancampfort, D., Hallgren, M., Firth, J., Veronese, N., Solmi, M., Brand, S., Cordes, J., Malchow, B., Gerber, M., Schmitt, A., Correll, C. U., De Hert, M., Gaughran, F., Schneider, F., Kinnafick, F., Falkai, P., Möller, H.-J., & Kahl, K. G. (2018). Epa guidance on physical activity as a treatment for severe mental illness: A meta-review of the evidence and position statement from the European psychiatric association (EPA), supported by the International organization of physical therapists in mental health (IOPTMH). *Eur Psychiatry*, *54*, 214–244.
- Sulz, S. K. D., & Grethe, C. (2005). *DIE VDS90-SYMPTOMLISTE – EINE ALTERNATIVE ZUR SCL90-R FÜR DIE AMBULANTE PSYCHOTHERAPIE-PRAXIS UND DAS INTERNE QUALITÄTSMANAGEMENT?* 11.
- ten Have, M., Penninx, B. W. J. H., Tuithof, M., van Dorsselaer, S., Kleinjan, M., Spijker, J., & de Graaf, R. (2017). Duration of major and minor depressive episodes and associated risk indicators in a psychiatric epidemiological cohort study of the general population. *Acta Psychiatrica Scandinavica*, *136*(3), 300–312. <https://doi.org/10.1111/acps.12753>
- Thornicroft, G., Chatterji, S., Evans-Lacko, S., Gruber, M., Sampson, N., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Andrade, L., Borges, G., Bruffaerts, R., Bunting, B., Almeida, J. M. C. de, Florescu, S., Girolamo, G. de, Gureje, O., Haro, J. M., He, Y., Hinkov, H., ... Kessler, R. C. (2017). Undertreatment of people with major depressive disorder in 21 countries. *The British Journal of Psychiatry*, *210*(2), 119–124. <https://doi.org/10.1192/bjp.bp.116.188078>
- Üstün, T. B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., Saxena, S., von Korf, M., & Pull, C. (2010). Developing the World Health Organization Disability Assessment Schedule 2.0. *Bulletin of the World Health Organization*, *88*(11), 815–823. <https://doi.org/10.2471/BLT.09.067231>
- Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., Carter, A., Casey, D. C., Charlson, F. J., Chen, A. Z., Coggeshall, M., Cornaby, L., Dandona, L., Dicker, D. J., Dillegge, T., Erskine, H. E., Ferrari, A. J., Fitzmaurice, C., Fleming, T., ... Murray, C. J. L. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease

- Study 2015. *The Lancet*, 388(10053), 1545–1602. [https://doi.org/10.1016/S0140-6736\(16\)31678-6](https://doi.org/10.1016/S0140-6736(16)31678-6)
- Vuorilehto, M. S., Melartin, T. K., & Isometsä, E. T. (2009). Course and outcome of depressive disorders in primary care: A prospective 18-month study. *Psychological Medicine*, 39(10), 1697–1707.
- Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., & Steinhausen, H.-C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, 21(9), 655–679. <https://doi.org/10.1016/j.euroneuro.2011.07.018>
- World Health Organization. (2004). *International statistical classification of diseases and related health problems: Instruction manual* (Bd. 2). World Health Organization.
- World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates*. WorldHealthOrganization.
- Zimovetz, E. A., Wolowacz, S. E., Classi, P. M., & Birt, J. (2012). *Methodologies used in cost-effectiveness models for evaluating treatments in major depressive disorder: A systematic review*. 19.

6. Anhang

Anhang A: Methoden

S1. Erhebungsinstrumente

S2. Kostenindikatoren der gesundheitsökonomischen Evaluation

Anhang B: Zusätzliche Abbildungen

Abbildung S1. Flow Chart

Abbildung S2. Veränderungen der Arbeitsfähigkeit (WHODAS 2.0) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

Abbildung S3. Veränderungen der allgemeinen Selbstwirksamkeit (GSE-6) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

Abbildung S4. Veränderungen der gesundheitsbedingten Lebensqualität (EQ-5D) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

Abbildung S5. Veränderungen der Befriedigung psychologischer Grundbedürfnisse (BPNSFS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

Abbildung S6. Veränderungen der Frustration psychologischer Grundbedürfnisse (BPNSFS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

Abbildung S7. Veränderungen des physischen Gesundheitsstatus (SF-12, PCS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

Abbildung S8. Veränderungen des mentalen Gesundheitsstatus (SF-12, MCS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

Abbildung S9. Individuelle Trends (basierend auf linearer Regression) der depressiven Symptomatik über 16 Wochen für alle inkludierte Studien-Probanden der Prozessevaluation

Abbildung S10. Modellierter Verlauf von depressiven Symptomen und Rumination (Grübeln) in beiden Therapiegruppen

Abbildung S11. Mediationsanalyse

Abbildung S12. Kurzfristige Dynamik von Emotionalen Zuständen (Freude, Traurigkeit) sowie Kognitionen (kognitive Kontrolle und Grübeln) im Anschluss an eine mindestens leichte körperliche Aktivität

Abbildung S13. kognitive Tests

Anhang C: Zusatztabelle

Tabelle S1. Probandencharakteristika

Tabelle S2. Mittelwerte (SD) für primäre und sekundäre Endpunkte in der mITT-Population

Tabelle S3. Schweregrad der Depression (BDI-II) in der mITT-Population

Tabelle S4. Grad der Beeinträchtigung nach WSAS in der mITT-Population

Tabelle S5. GEE-Analyse mit BDI-II als abhängige Variable (Modell 1) in der mITT-Population

Tabelle S6. GEE-Analyse mit BDI-II als abhängige Variable (Modell 2) in der mITT-Population

Tabelle S7. GLMM-Analyse mit HAM-D als abhängige Variable (Modell 3) in der mITT-Population

Tabelle S8. GLMM-Analyse mit HAMD als abhängige Variable (Modell 4) in der mITT-Population

Tabelle S9. GLMM-Analyse mit WSAS als abhängige Variable (Modell 5) in der mITT-Population

Tabelle S10. GLMM-Analyse mit WSAS als abhängige Variable (Modell 6) in der mITT-Population

Tabelle S11. Beschreibung der Interventionskosten der EXP- und TAU-Gruppe

Akronym: STEP.De

Förderkennzeichen: 01NVF17050

Tabelle S12. Darstellung der Interventionskosten inkl. (anschließende)
Psychotherapiekosten

Tabelle S13. Deskriptive Darstellung der Gesundheitskosten der MITT Population

Tabelle S14. Difference-in-Difference Modell zur Analyse der Gesamtkosten einschließlich
der Interventionskosten (10 Monate)

Tabelle S15. Difference-in-Difference Modell zur Analyse der Gesamtkosten (ohne
Interventionskosten)

7. Anlagen

Nicht zutreffend.

Anhang A: Methoden¹

S1. Erhebungsinstrumente

Strukturelles Klinische Interview I für das Diagnostische und Statistische Manual Psychischer Störungen 4 (SKID-I für DSM IV)

Um die Indikation für eine Behandlung zusätzlich zur Einschlussdiagnose im Erstgespräch noch einmal abzusichern, wurden alle Psychotherapeuten angewiesen, das Strukturelle Klinische Interview I für das Diagnostische und Statistische Manual Psychischer Störungen 4 (SKID-I, Structured Clinical Interview for DSM IV, SCID-I), Achse 1, Abschnitt A, E und I zu verwenden. Das SKID-I ist ein ökonomisches, effizientes und reliables Verfahren zur strukturierten Erfassung ausgewählter Achse-I-Störungen. Es ist ein verbreitetes Verfahren in der klinischen Praxis und Forschung. Nach eindeutig formulierten diagnostischen Kriterien werden unter Abschnitt A Affektive Symptome (wie z.B. Major Depression, Manie, Dysthymie), unter Abschnitt E Missbrauch und Abhängigkeit von psychotropen Substanzen und unter Abschnitt I Anpassungsstörungen (z. B. depressive Verstimmung, ängstliche Gehemmtheit, Verhaltensstörung) mithilfe von bestimmten Fragen erfasst (Wittchen et al., 1997). Anschließend werden die Antworten kodiert, in Abhängigkeit davon, inwieweit ein diagnostisches Kriterium erfüllt ist oder nicht, um am Ende eine Differenzialdiagnose stellen zu können (Doering et al., 2012). Einige Studien zu Reliabilität zeigen gute Kennwerte. Es werden Cohen's Kappa für SKID-I von 0.61-0.83 (MD = 0.71) berichtet (Lobbstaël et al., 2011).

Primärer Endpunkt

BDI-II

Depressive Symptome wurden mit der deutschen Version (Hautzinger, Keller, & Kühner, 2009) des Beck-Depressions-Inventars (BDI-II) (Beck et al., 1996) gemessen. Der BDI-II ist ein 21 Items umfassendes Selbstbericht-Screeningverfahren für Depression. Jedes Item wird auf einer 4-Punkte-Likert-Skala von 0 bis 3 bewertet, wobei höhere Punktzahlen ein höheres Ausmaß an depressiven Symptomen anzeigen. Nach dem Leitfaden für den BDI-II (Beck et al., 1996) bedeutet eine Punktzahl von 0-13 eine minimale Depression, 14-19 eine leichte Depression, 20-28 eine mittelschwere Depression und 29-63 eine schwere Depression.

Sekundäre Endpunkte

Demographischer Fragebogen

Von allen an der Studie teilnehmenden Personen wurden zu Beginn demographische Daten erhoben. Die Probanden füllten selbst einen online Fragebogen aus, in welchem sie Informationen über Alter, Geschlecht, Körpermaße, Familienstand, Bildungsgrad, sozialen Status und gesundheitsgefährdendes Verhalten angaben. Für den Bildungsgrad wurde eine Variable mit drei Kategorien erstellt: niedriger (Hauptschule), mittlerer (Realschulabschluss) und hoher Bildungsgrad (Hochschulreife und Hochschulabschluss). Die Einkommensvariable wurde in niedriges (<1000€), mittleres (1000-2000€) und hohes (>2000€) persönliches monatliches Nettoeinkommen eingeteilt.

Hamilton Depressions-Skala (HAM-D)

Die HAM-D (Hamilton, 1960) ist ein weit verbreitetes Screening-Instrument zur Beurteilung depressiver Symptome, welche in der vergangenen Woche aufgetreten sind. Die 17 Items bestehen aus neun Symptomen (depressive Stimmung, Selbstabwertung und Schuldgefühle, Suizidalität, Beeinträchtigung der Arbeit und Aktivitäten des täglichen Lebens, psychomotorische Verlangsamung, Unruhe, psychische Angstsymptome, somatische Angstsymptome, Hypochondrie) die mit Werten zwischen 0 (nicht vorhanden) und 4 (andauernd) bewertet werden, und aus acht Symptomen (Einschlafstörungen, Durchschlafstörungen, morgendliche Schlafstörungen, gastrointestinale Beschwerden, allgemeine somatische Beschwerden, Verlust des sexuellen Interesses, mangelnde Krankheitseinsicht, Gewichtsverlust), die mit Werten zwischen 0 (nicht vorhanden) und 2 (deutlich ausgeprägt) bewertet werden, sodass die maximale Punktzahl 52

¹ In Teilen übernommen und übersetzt aus Heißel et al., in prep.

beträgt. Ein Wert von ≤ 7 liegt im Normalbereich. Die interne Konsistenz der HAM-D beträgt Cronbach's $\alpha=0.81$ (Maier & Philipp, 2013).

Verhaltensdiagnostik-System -90 Symptomliste (VDS-90)

Die VDS-90 (Sulz & Grethe, 2005) ist ein Screening-Instrument zur Erkennung psychischer Störungen gemäß ICD-10 (World Health Organization, 2004). Die Punktwerte des Items liegen zwischen 0 (keine psychischen Probleme) und 3 (schwere psychische Probleme). Der Gesamtwert für Angst ergibt sich aus dem Durchschnitt von 5 Items.

Work and Social Adjustment Scale (WSAS)

Die WSAS (Marks, 1986; Mundt et al., 2002b) besteht aus 5 Items (Arbeit, Haushaltsführung, soziale Freizeit, private Freizeit und Beziehungen), die jeweils auf einer 9-Punkte-Likert-Skala von 0 (überhaupt nicht beeinträchtigt) bis 8 (sehr stark beeinträchtigt) von den Teilnehmenden selbst bewertet werden. Die Gesamtpunktzahl reicht von 0 bis 40, wobei höhere Punktwerte einen höheren Grad der Beeinträchtigung kennzeichnen. Die interne Konsistenz der WSAS beträgt Cronbach's $\alpha=0.89$ (Heissel et al., 2021).

WHO Disability Assessment Schedule 2.0 (WHODAS 2.0)

WHODAS 2.0 (Üstün et al., 2010) ist ein generalisiertes Bewertungsschema zur Erfassung des Grades der Funktionsfähigkeit und Behinderung einer Person in sechs wichtigen Lebensbereichen, basierend auf dem konzeptionellen Rahmen der Internationalen Klassifikation der Funktionsfähigkeit (ICF) der WHO. Die Befragung wurde telefonisch von geschulten Assessoren durchgeführt. Der WHODAS 2.0 ist ein Fragebogen, mit dem die Funktionsfähigkeit einer Person in sechs Bereichen bewertet wird: Kognition, Mobilität, Selbstfürsorge, soziale Interaktion, Alltagsaktivitäten und gesellschaftliche Teilhabe. In dieser Studie wurde die deutsche 12-Item-Screening-Version von WHODAS 2.0 (Kirchberger et al., 2014) verwendet. Für jedes Item gaben die Befragten den Grad der in den letzten 30 Tagen erlebten Schwierigkeiten auf einer fünfstufigen Likert-Skala von 1 "keine Schwierigkeiten" bis 5 "extreme Schwierigkeiten/ nicht möglich" an. Die Gesamtpunktzahl für die globale Behinderung reicht von 0 "keine Behinderung" bis 60 "vollständige Behinderung". Die interne Konsistenz des WHODAS 2.0 beträgt Cronbach's $\alpha=0.98$ (Üstün et al., 2010).

General-Self-Efficacy-Scale (GSE-6)

Die Generalisierte Selbstwirksamkeitsskala ist ein Screening-Instrument zur Erfassung des Selbstbildes hinsichtlich der eigenen Selbstwirksamkeitserwartung. Anhand von 6 Items wird das Vertrauen in die eigene Fähigkeit, mit verschiedenen anspruchsvollen Situationen durch adaptives Handeln umzugehen, ermittelt. Dazu wird jedes Item vom Probanden selbst mit Werten von 1 (stimmt nicht) bis 4 (stimmt genau) bewertet, sodass die Gesamtpunktzahl zwischen 6 und 24 liegt. Die interne Konsistenz der GSE-6 beträgt Cronbach's $\alpha=0.79-0.88$ (Romppel et al., 2013).

EuroQol Five-Dimensional-Five-Level Questionnaire (EQ-5D-5L)

Der EQ-5D-5L Fragebogen ist ein allgemeines Messinstrument zur Beschreibung und Bewertung der Gesundheit. Er basiert auf einem deskriptiven System, das Gesundheit anhand von fünf Dimensionen definiert: Mobilität, Selbstversorgung, alltägliche Aktivitäten, Schmerzen/Beschwerden, und Angst/ Depression. Jede Dimension wird vom Patienten selbst mit einer der fünf zur Verfügung stehenden Antwortkategorien bewertet (keine, leichte, mäßige, große oder extreme Probleme/Schmerzen/Angst). Im Anschluss erfolgt zusätzlich eine Bewertung des Gesundheitszustands auf einer vertikalen visuellen Analogskala (EQ-VAS) mit Werten von 0-100. Die Schätzung der Reliabilität mittels Cronbachs Alpha ist für ein Indexmessinstrument, wie der EQ-5D, der verschiedene Dimensionen erfasst, nicht geeignet (Konerding, 2013).

Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS)

Die BPNSFS ist ein Messinstrument zur Bewertung der Befriedigung und Frustration der drei psychologischen Grundbedürfnisse Autonomie, Kompetenz und Verbundenheit. Das 6-Faktoren Modell fragt insgesamt 24 Items ab. Jedes der drei Grundbedürfnisse wird durch acht Items adressiert, von denen sich jeweils vier auf die Befriedigung und vier auf die Frustration des Bedürfnisses beziehen. Die Items werden auf einer 5-Punkte-Likert-Skala von 1 (vollkommen

anderer Meinung) bis 5 (vollkommene Zustimmung) vom Patienten selbst bewertet. Sowohl für die Subskala der Bedürfnisfrustration als auch für die Subskala der Bedürfnisbefriedigung beträgt die interne Konsistenz Cronbach's $\alpha=0.84$ (Heissel et al., 2019). In der vorliegenden Studie wurde eine 12-item Version verwendet.

12-Item Short Form Survey (SF-12)

Der SF-12 Fragebogen ist ein Zwei-Komponenten-Messinstrument aus insgesamt 12 Items, die zur Erfassung des körperlichen und mentalen Gesundheitsstatus dienen. Die Skala der körperlichen Komponenten (PCS) setzt sich aus 6 Items zusammen, von denen sich zwei auf die physische Gesundheit, zwei auf die physisch bedingte Rollenfunktion sowie jeweils ein Item auf die körperlichen Schmerzen und die allgemeine Gesundheitswahrnehmung beziehen. Die Skala der mentalen Komponenten (MCS) setzt sich aus 6 Items zusammen, von denen sich zwei auf die mentale Gesundheit, zwei auf die emotional bedingte Rollenfunktion sowie jeweils ein Item auf die Vitalität und die soziale Funktionsfähigkeit beziehen. Die zwei Items zur physischen Gesundheit werden anhand von drei Stufen (stark eingeschränkt bis überhaupt nicht eingeschränkt) bewertet. Alle anderen Items werden auf einer 5-Punkte-Linkert-Skala bewertet. Mithilfe einer Software wird eine Gesamtpunktzahl mit Werten zwischen 0 bis 100 ermittelt, wobei eine höhere Punktzahl mit einem besseren körperlichen und mentalen Gesundheitsstatus assoziiert ist.

S2. Kostenindikatoren der gesundheitsökonomischen Evaluation (Sperrfrist: 30.09.2023)

STEP.De - Variablennamen für gesundheitsökonomische Daten

1. Tabelle: Arbeitsunfähigkeit

Spaltenname	Beschreibung
Pseudonym	ID des Versicherten
Einschreibedatum	Einschreibedatum des Versicherten in Step.de-Programm im Format 9999-12-31
AU_von	Beginndatum des AU-Zeitraums im Format 9999-12-31
AU_bis	Enddatum des AU-Zeitraums im Format 9999-12-31
AU_Tage	Länge des AU-Zeitraumes in Tagen (=AU_bis – AU_von + 1) als Integer
AU_Hauptdiagnose	ICD der Hauptdiagnose des AU-Falls, z.B. Z73, F48.0, M07.30*
AU_Leistungsart	Art der Leistung als Text. Kann „Krankengeld“ oder „Verletztengeld“ sein.
AU_Fallkosten	Gesamtkosten des AU-Falls
ICD1	Auflistung aller ICDs der einzelnen F-Diagnosen eines AU-Falls. Die maximale Anzahl der Diagnosen ist für die betrachteten Fälle der Kassen 9. Deshalb gibt es hier 9 Spalten. Die Werte sind alphabetisch aufsteigend sortiert pro AU-Fall. ICDs mit gleicher Bezeichnung werden zu einer ICD zusammengefasst.
ICD2	
...	
ICD9	

2. Tabelle: stationäre Kosten

Spaltenname	Beschreibung
Pseudonym	ID des Versicherten

Einschreibedatum	Einschreibedatum des Versicherten in Step.de-Programm im Format 9999-12-31
KH_von	Aufnahmetag im Format 9999-12-31
KH_bis	Entlassungstag im Format 9999-12-31
KH_Fachabteilung_Code	4-stelliger Fachabteilungscode (z.B. 1600) der Fachabteilung, die die Hauptdiagnose gestellt hat
KH_Fachabteilung_Text	Name der Fachabteilung als Übersetzung des Codes, z.B. „Unfallchirurgie“ (für Code 4000)
Leistungsart	Art der Leistung als Text, z.B. „Vollstationär“ oder „Teilstationär“
KH_DRG_Gruppe	Klassifizierungsnummer des KH-Falls nach DRG, z.B. 7010K12Z
KHdiag_ICD	ICD der Hauptdiagnose des KH-Falls, z.B. I10.00, L40.5+, K20
KH_Gesamtbruttobetrag	Summe der Zahlungen inklusive der Zuzahlungsbeträge des Versicherten
KH_Gesamtnettobetrag	Summe der Zahlungen ohne Zuzahlungsbeträge des Versicherten
KH_Gesamtzuzahlungist	Summe der geleisteten Zuzahlungen des Versicherten
ICD1	Auflistung aller ICDs der einzelnen Diagnosen eines KH-Falls. Die maximale Anzahl der Diagnosen ist für die betrachteten Fälle der Kassen 29. Deshalb gibt es hier 29 Spalten. Die Werte sind alphabetisch aufsteigend sortiert pro KH-Fall. ICDs mit gleicher Bezeichnung werden zu einer ICD zusammengefasst.
ICD2	
...	
ICD29	

3. Tabelle: ambulante Kosten

Spaltenname	Beschreibung
Pseudonym	ID des Versicherten
Einschreibedatum	Einschreibedatum des Versicherten in Step.de-Programm im Format 9999-12-31
TP1_von	Beginndatum des ambulanten Falls im Format 9999-12-31
TP1_bis	Enddatum des ambulanten Falls im Format 9999-12-31
TP1_Fallpunkte	Summe der Punktzahlen des Behandlungsfalls als Dezimalzahl mit 2 Nachkommastellen
TP1_Fallkosten	Gesamtkosten des Behandlungsfalls
TP1_Punktwert	Wert eines Punktes
TP1_Kosten	= Fallpunkte * Punktwert / 100 + Fallkosten
TP1_Dialysesachkosten	Kosten für Dialyse für den ambulanten Fall
TP1_Fachgruppe	Bezeichnung der Fachgruppe des behandelnden Arztes, z.B. „Radiologie“
ICD1	Auflistung aller ICDs der einzelnen Diagnosen eines TP1-Falls. Die maximale Anzahl der Diagnosen ist für die betrachteten Fälle der Kassen momentan 46. Deshalb gibt es hier 46 Spalten. Die Werte sind alphabetisch aufsteigend sortiert pro TP1-Fall. ICDs mit gleicher Bezeichnung werden zu einer ICD zusammengefasst.
ICD2	
...	
ICD46	

4. Tabelle: Medikamentenkosten

Spaltenname	Beschreibung
Pseudonym	ID des Versicherten
Einschreibedatum	Einschreibedatum des Versicherten in Step.de-Programm im Format 9999-12-31
TP3rez_ausgestellt_am	Ausstellungsdatum des Rezeptes im Format 9999-12-31
TP3evo_Zuzahlungsbetrag_anteilig	Anteil des Zuzahlungsbetrags des Rezeptes an der Einzelverordnung, d.h.: (Zuzahlungsbetrag des Versicherten am Rezept)* TP3evo_Betrag/(Gesamtbetrag des Rezeptes)
TP3evo_Betrag	Betrag der Einzelverordnung (des einzelnen Mittels)
TP3Evo_Rabatt_summe	Rabattbetrag der Einzelverordnung (für das einzelne Mittel)
Fachgruppe	Bezeichnung der zugehörigen Fachgruppe des verordnenden Arztes, z.B. „Augenheilkunde“
widoStamm_ATC_code	ATC-Code des Medikaments, z.B. „C07AB07“

5. Tabelle: Heil- und Hilfsmittel

Spaltenname	Beschreibung
Pseudonym	ID des Versicherten
Einschreibedatum	Einschreibedatum des Versicherten in Step.de-Programm im Format 9999-12-31
Lm_von	Beginndatum der Leistung bzw. Abgabedatum im Format 9999-12-31
Lm_bis	Enddatum der Leistung bzw. Abgabedatum im Format 9999-12-31
Lm_Maßnahmeart	Art der Leistungsmaßnahme, z.B. „Heilmittel“
Lm_Leistungsart	Art der Leistung, z.B. „Physiotherapie“
Lm_Kosten_Kasse_ist	Von der Kasse geleistete Zahlungen
Lm_Zuzahlung_ist	Vom Versicherten geleistete Zuzahlungen
Lm_Gesamtbruttobetrag	Summe aller Zahlungen inklusive der Versichertenanteile
Lm_Hauptdiagnose	, z.B. M75.5 oder X59.9!

Anhang B: Zusätzliche Abbildungen¹

Abbildung S1. Flow Chart

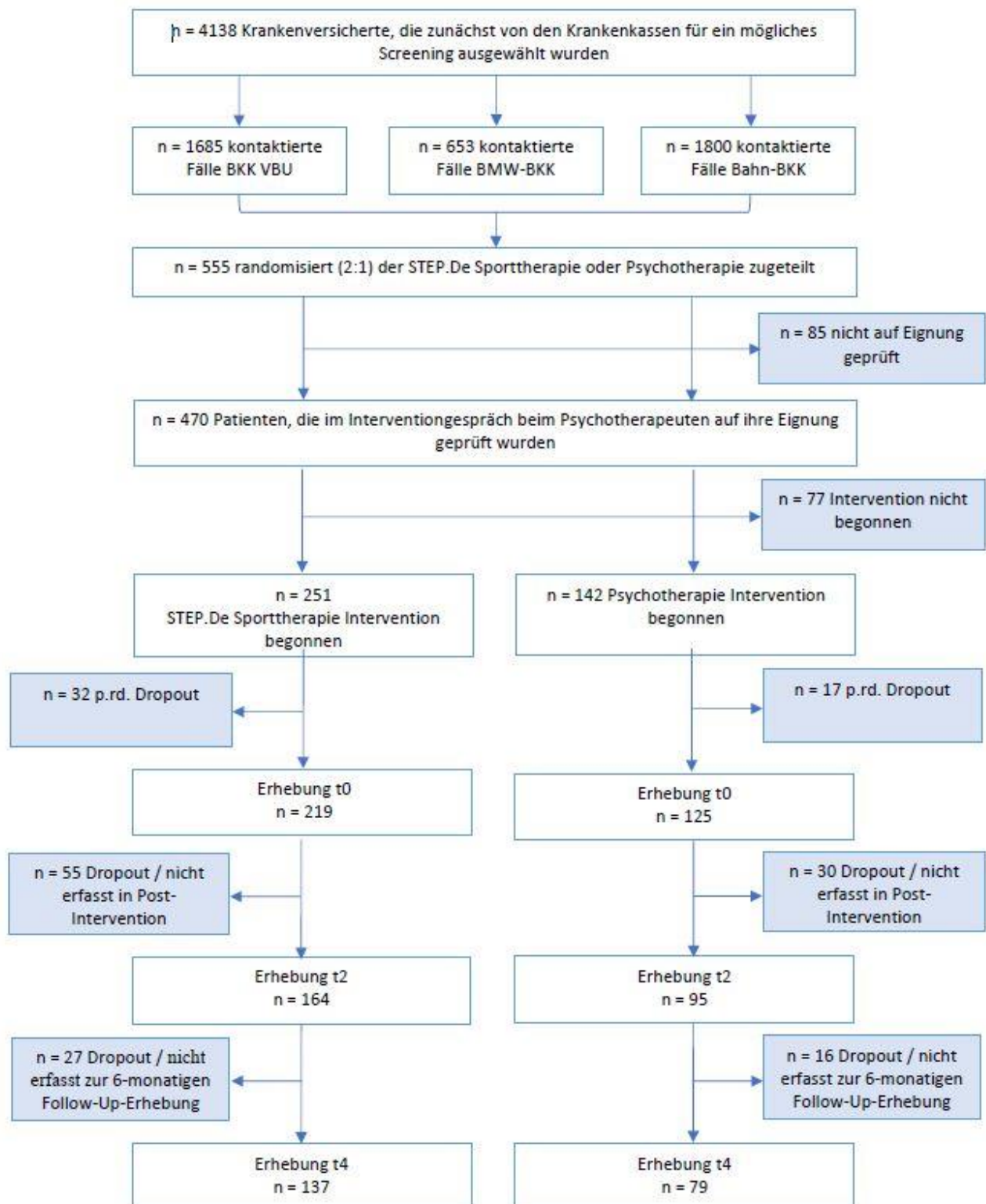


Abbildung 1: Studienprofil für die MITT Population

¹ In Teilen übernommen und übersetzt aus Heißel et al., in prep.

Abbildung S2. Veränderungen der Arbeitsfähigkeit (WHODAS 2.0) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

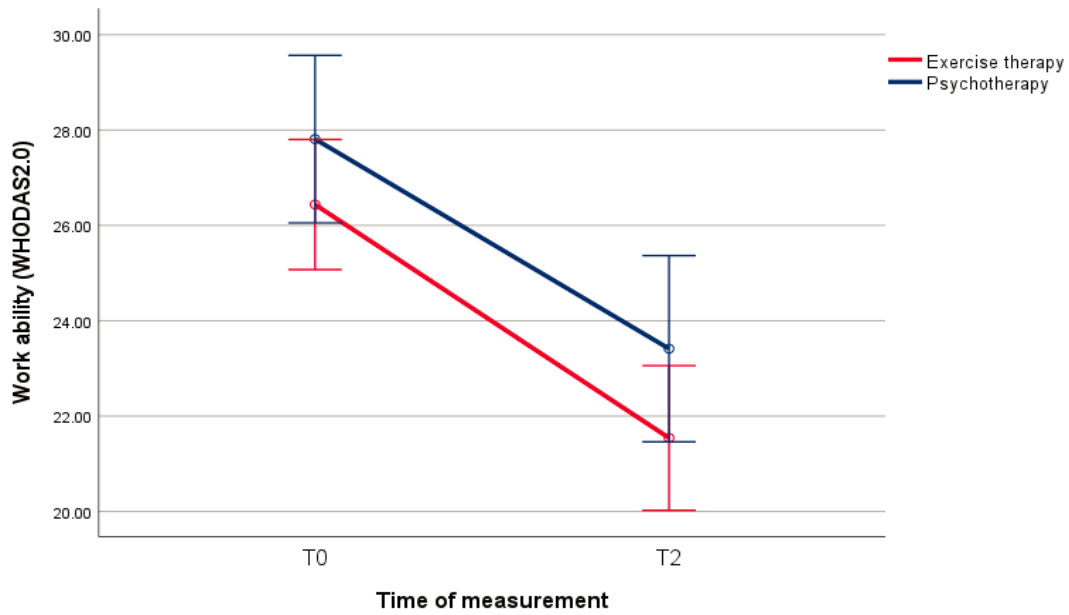


Abbildung S3. Veränderungen der allgemeinen Selbstwirksamkeit (GSE-6) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

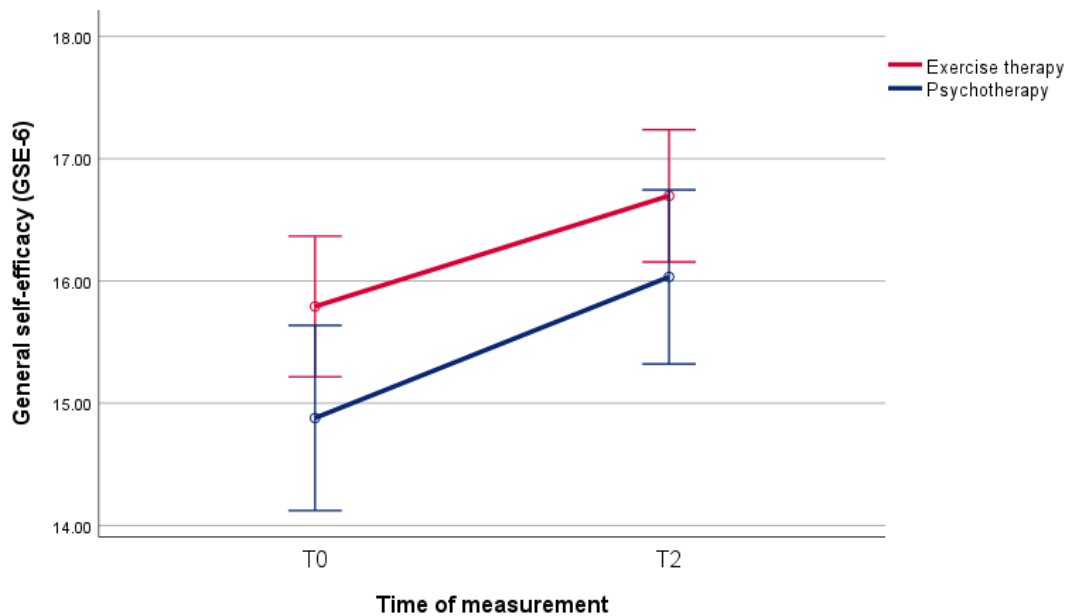


Abbildung S4. Veränderungen der gesundheitsbedingten Lebensqualität (EQ-5D) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

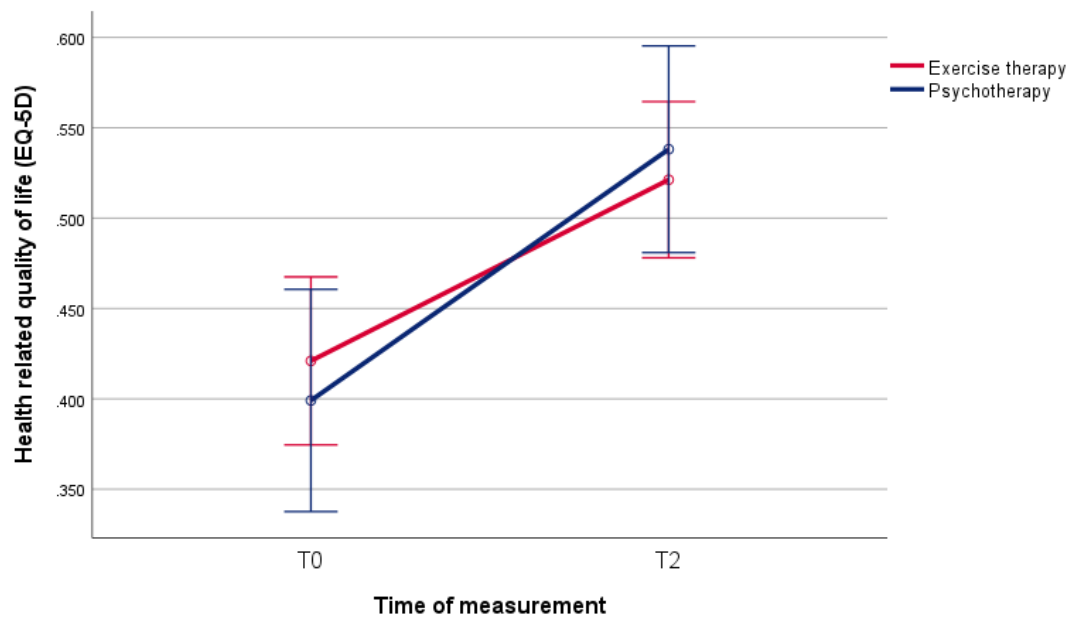


Abbildung S5. Veränderungen der Befriedigung psychologischer Grundbedürfnisse (BPNSFS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

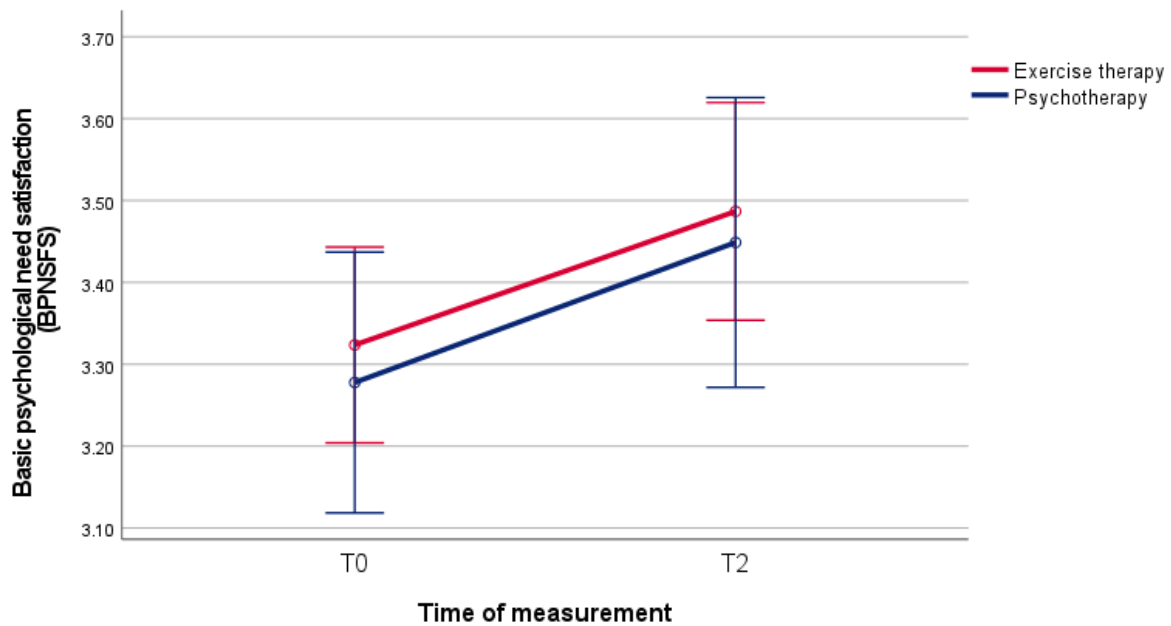


Abbildung S6. Veränderungen der Frustration psychologischer Grundbedürfnisse (BPNSFS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

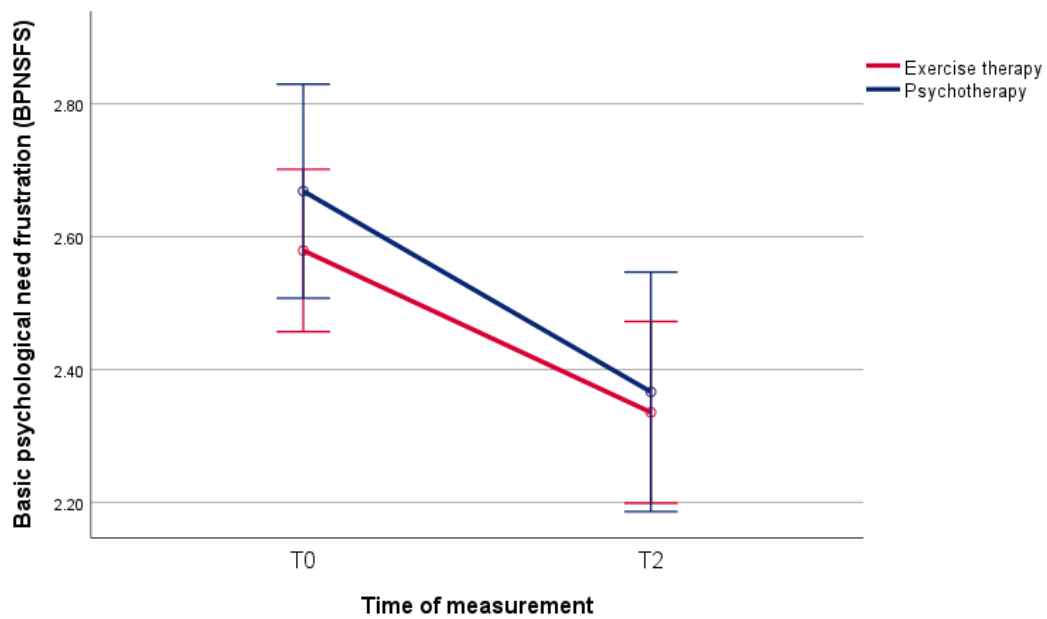


Abbildung S7. Veränderungen des physischen Gesundheitsstatus (SF-12, PCS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

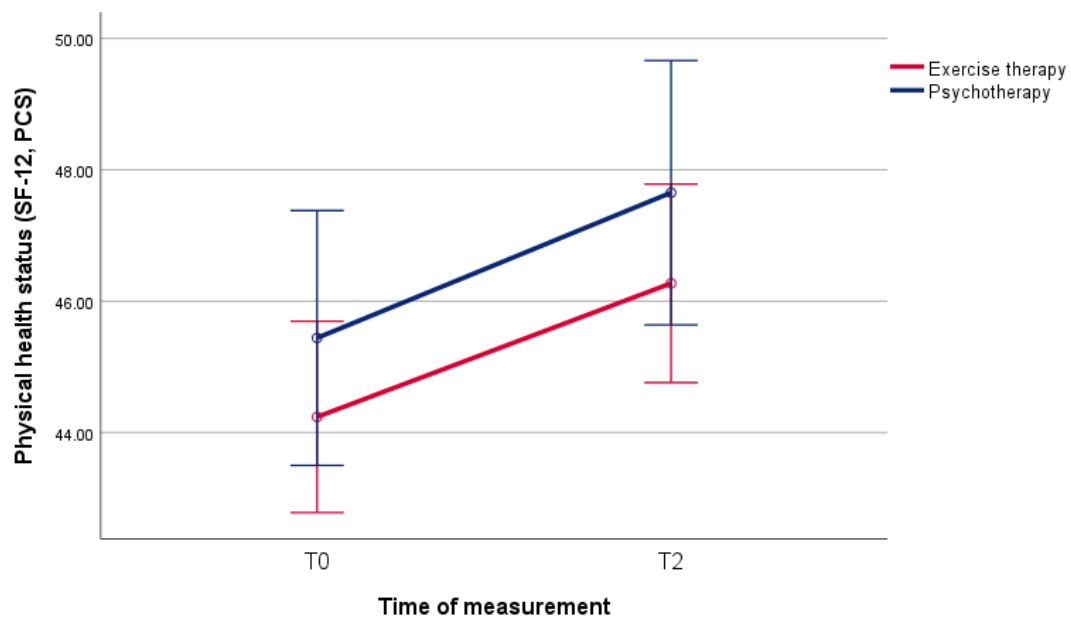


Abbildung S8. Veränderungen des mentalen Gesundheitsstatus (SF-12, MCS) vom Beginn bis zum Abschluss der Intervention in der mITT-Population

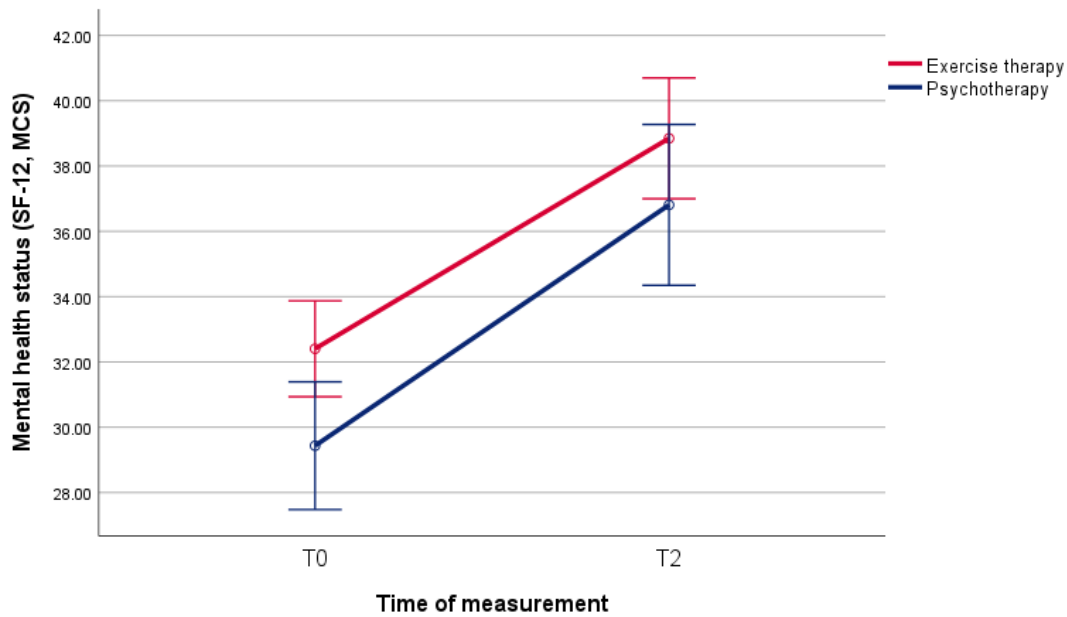


Abbildung S9. Individuelle Trends (basierend auf linearer Regression) der depressiven Symptomatik über 16 Wochen für alle inkludierte Studien-Probanden der Prozessevaluation.

Grüne Linien (abfallend) weisen auf eine tendenzielle Verbesserung der depressiven Symptomatik im Therapieverlauf hin, rote Linien auf eine tendenzielle Verschlechterung (ansteigende Symptomatik); schwarze Kreuze signalisieren Studienabbruch.

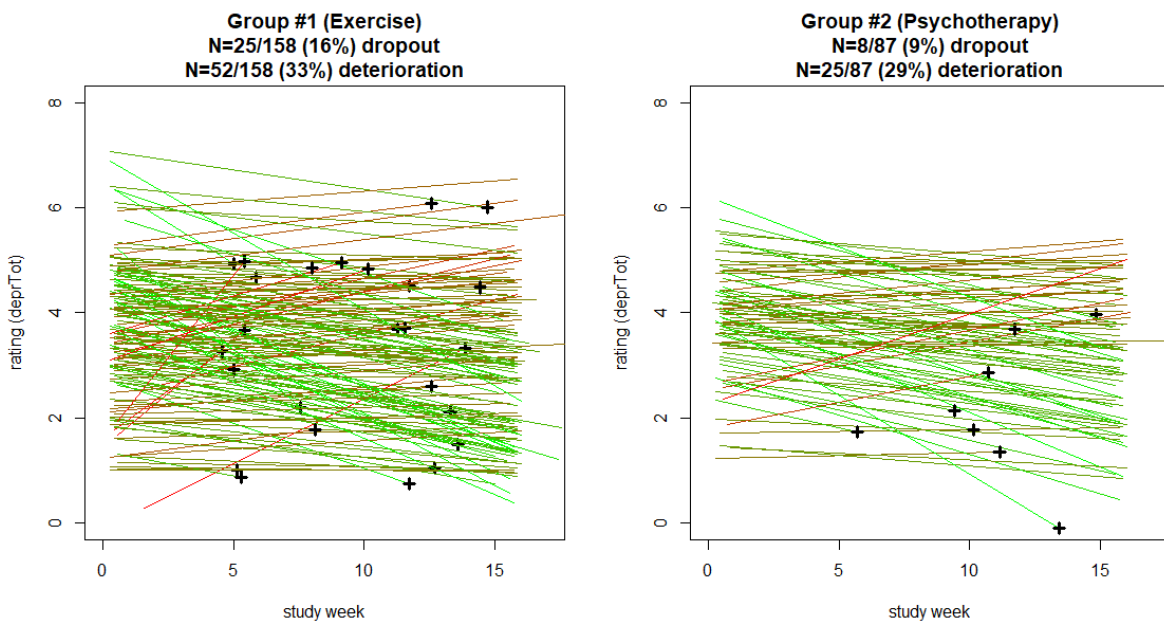


Abbildung S10. Modellierter Verlauf von depressiven Symptomen und Rumination (Grübeln) in beiden Therapiegruppen.

Tendenziell verbesserten sich kognitive Symptome (Rumination) schneller als die sog. Kardinalsymptome der Depression (Niedergeschlagenheit, Antriebsstörung, Anhedonie); innerhalb der Sporttherapie-Gruppe ist dieser Unterschied signifikant ($p < .01$), was auf einen besonderen Wirkmechanismus von Sport hinweisen kann.

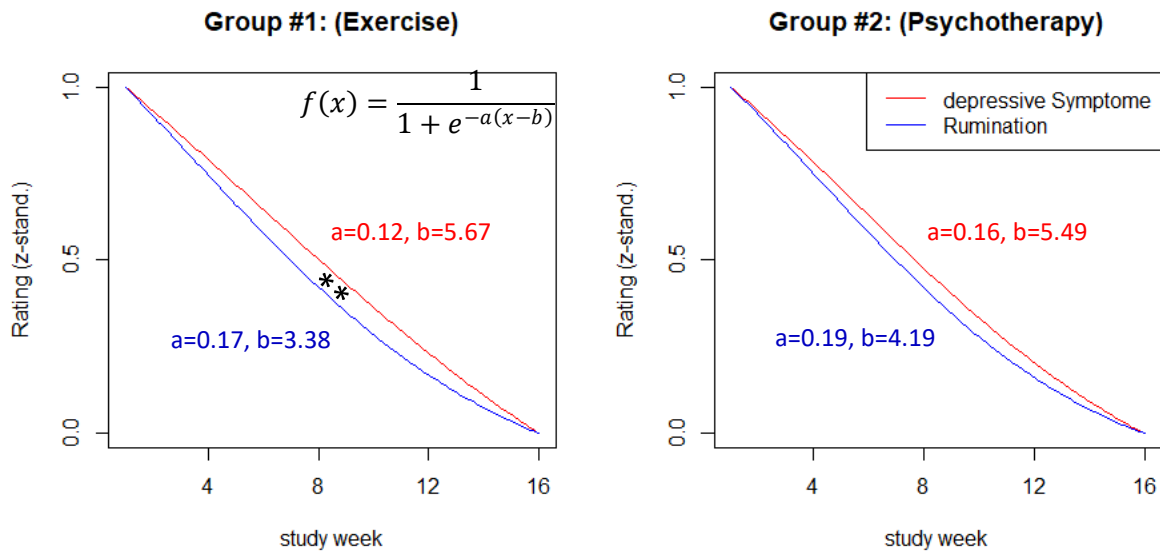


Abbildung S11. Mediationsanalyse

Die Mediationsanalyse ergab eine partielle Mediation des Effekts von körperlicher Aktivität auf depressive Symptome, der (zumindest teilweise) über die Zwischenschritte einer verbesserten kognitiven Kontrolle und weniger Grübeln (Rumination) vermittelt wird. Durch die dynamische Modellierung der Effekte (time-lagged) können diese Effekte als kausale Wirkmechanismen interpretiert werden. Dargestellt sind die standardisierten Koeffizienten sowie das Signifikanz-Niveau (* $p < .05$, ** $p < .01$, *** $p < .001$).

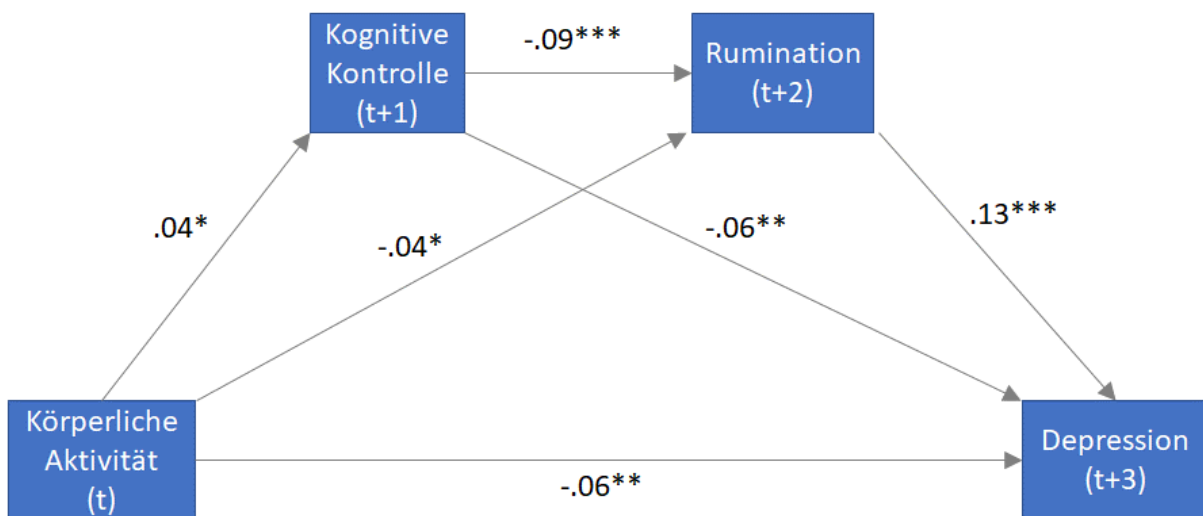


Abbildung S12. Kurzfristige Dynamik von Emotionalen Zuständen (Freude, Traurigkeit) sowie Kognitionen (kognitive Kontrolle und Grübeln) im Anschluss an eine mindestens leichte körperliche Aktivität

Kurzfristige Dynamik von Emotionalen Zuständen (Freude, Traurigkeit) sowie Kognitionen (kognitive Kontrolle und Grübeln) im Anschluss an eine mindestens leichte körperliche Aktivität (definiert über Bewegungssensoren: 10 min oder länger eine kumulierte Aktivität [ENMO] von 30 mG oder mehr). Dargestellt sind z-standardisierte Werte der Ratings als sliding average (mit window size=120 min).

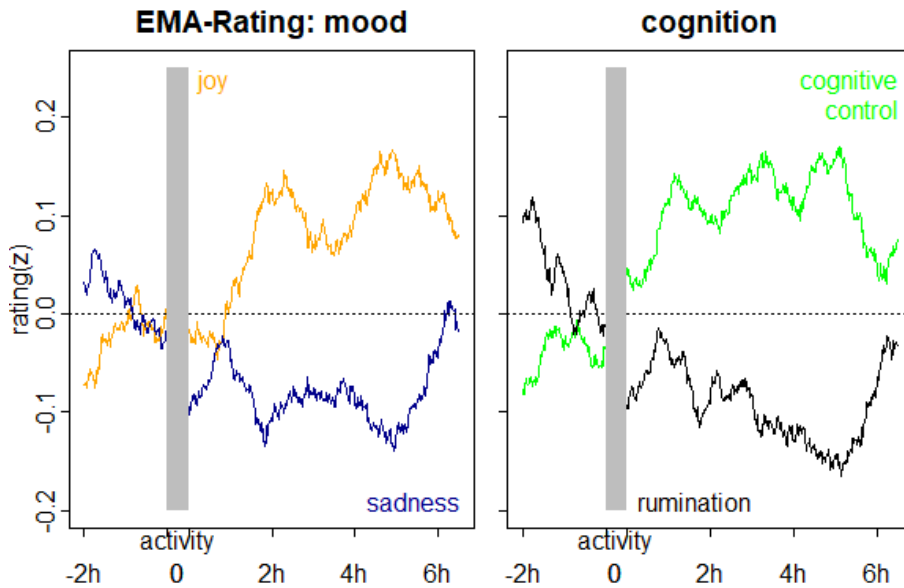
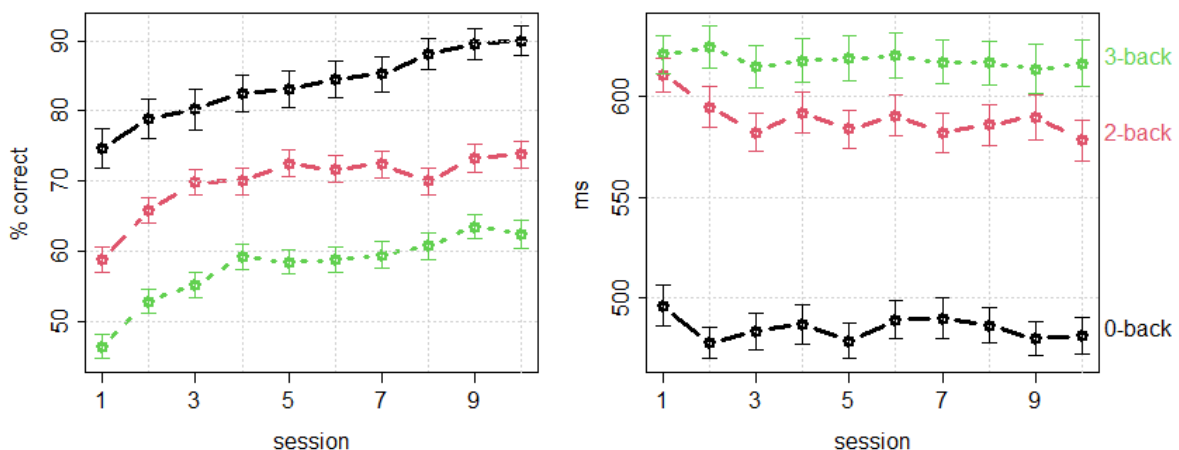


Abbildung S13. kognitive Tests

Wöchentlich wiederholte kognitive Tests (N-Back, durchgeführt am Smartphone an zufälligen Tagen im Alltag) weisen stetige Verbesserungen über die Zeit auf (links: Treffer-Genauigkeit wird höher, rechts: Reaktionsgeschwindigkeit wird schneller) sowie stabile Unterschiede zwischen den Schwierigkeitsgraden (0-back [schwarz], als leichteste Aufgabe, führt zu den schnellsten Reaktionen und der höchsten Treffer-Genauigkeit; 2-back [rot], als mittleres Niveau, führt zu langsameren Reaktionen und geringerer Treffer-Genauigkeit; 3-back [grün], als schwierigste Variante, führt zu den langsamsten Reaktionen und der geringsten Treffer-Genauigkeit). Anmerkung: Da nicht alle Probanden die max. Anzahl von 16 Sessions absolvierten, sind hier nur die ersten 10 Sessions dargestellt.



Anhang C: Zusatztabelle¹

Tabelle S1. Probandencharakteristika

	Sporttherapie (n=219)	Psychotherapie (n=125)	Gesamt (n=344)
Alter (Jahre), Median (IQR), Range	51.0 (40.0-58.0), 21-69	47.0 (36.0-55.0), 20-64	49.0 (39.0-57.0), 20-69
Geschlecht			
Weiblich	153/218 (69.9)	93/125 (74.4)	259/343 (71.5)
Männlich	65/218 (29.7)	32/125 (25.6)	97/343 (28.2)
Bildungsniveau			
Untere Sekundarschule	19/210 (9.0)	9/110 (8.2)	28/320 (8.8)
Sekundarschule	113/210 (53.8)	66/110 (60.0)	179/320 (55.9)
Höhere Bildung	78/210 (37.1)	35/110 (31.8)	113/320 (35.3)
Nettoeinkommen			
Geringes Einkommen	23/207 (11.1)	10/109 (9.2)	33/316 (10.4)
Mittleres Einkommen	110/207 (53.1)	69/109 (63.3)	179/316 (56.6)
Hohes Einkommen	74/207 (35.7)	30/109 (27.5)	104/316 (32.9)
BMI (kg/m ²), Median (IQR)	26.2 (23.4-30.3)	25.5 (22.4-30.1)	25.9 (23.0-30.3)
Anzahl der Medikamente, Median (IQR)	0.0 (0.0-1.0)	0.0 (0.0-1.0)	0.0 (0.0-1.0)
WHODAS 2.0 Punktzahl	27.0 (22.0-32.7)	28.0 (23.0-32.8)	27.3 (22.0-32.7)
Depressive Symptome (BDI-II), Mittelwert (SD)	22.4 (10.0)	23.2 (9.8)	22.7 (9.9)

Daten sind in n/N (%) sofern nicht anders angegeben. Die Stichprobengröße reichte aufgrund fehlender Antworten von n = 219 bis n = 207 in der STEP.De Sporttherapie-Gruppe und von n = 125 bis n = 109 in der Psychotherapie-Gruppe. BMI=Body Mass Index. WHODAS 2.0= World Health Organization Disability Assessment Schedule.

Tabelle S2. Mittelwerte (SD) für primäre und sekundäre Endpunkte in der mITT-Population

	T0		T1		T2		T3		T4	
	MW (SD)	N	MW (SD)	N	MW (SD)	N	MW (SD)	N	MW (SD)	N
BDI-II										
Sporttherapie	22.4 (10.0)	215	17.9 (10.2)	175	15.0 (10.1)	164	14.6 (9.9)	154	13.8 (10.9)	137
Psychotherapie	23.2 (9.8)	117	19.0 (10.2)	101	15.7 (10.0)	95	14.6 (10.8)	83	13.8 (11.1)	79
HAM-D										
Sporttherapie	15.9 (7.5)	215			10.2 (7.3)	168			9.4 (7.2)	146
Psychotherapie	15.5 (6.8)	116			11.6 (6.7)	93			9.6 (8.0)	86
WSAS										
Sporttherapie	18.6 (10.1)	214	16.1 (10.8)	175	13.7 (10.4)	163	13.1 (10.4)	153	11.8 (10.7)	136
Psychotherapie	20.2 (9.4)	116	18.1 (10.1)	101	15.0 (10.4)	94	13.2 (10.9)	83	13.2 (11.0)	79
WHODAS 2.0										
Sporttherapie	27.4 (7.8)	160			22.2 (8.0)	130			21.4 (9.2)	110
Psychotherapie	28.0 (6.1)	92			23.2 (7.3)	74			22.6 (7.1)	72
GSE6										
Sporttherapie	15.7 (3.7)	211			16.7 (3.4)	161			16.9 (3.5)	134
Psychotherapie	15.0 (3.8)	113			15.9 (3.7)	93			16.3 (4.3)	77
EQ-5D-5L										
Sporttherapie	·40 (-.33)	206	·45 (-.31)	171	·50 (-.29)	161	·52 (-.29)	151	·52 (-.32)	134
Psychotherapie	·38 (-.28)	108	·46 (-.31)	98	·53 (-.28)	93	·54 (-.30)	82	53 (-.29)	78
VDS-90										
Sporttherapie	·96 (-.51)	215			·71 (-.48)	170			·71 (-.54)	149
Psychotherapie	·94 (-.44)	116			·75 (-.47)	95			·66 (-.45)	86
VDS-90 Angst										
Sporttherapie	·96 (-.77)	215			·75 (-.72)	170			·76 (-.78)	149
Psychotherapie	·88 (-.70)	116			·79 (-.74)	95			·68 (-.65)	86
BPNSFS										
Satisfaction										
Sporttherapie	3.3 (-.78)	211			3.5 (-.85)	159			3.6 (-.82)	134
Psychotherapie	3.3 (-.77)	112			3.4 (-.82)	94			3.4 (-.92)	77
Frustration										
Sporttherapie	2.6 (-.83)	211			2.3 (-.90)	160			2.2 (-.87)	134
Psychotherapie	2.7 (-.74)	112			2.4 (-.84)	93			2.3 (-.91)	79
SF-12										
PCS										
Sporttherapie	43.3 (9.1)	196			46.3 (8.9)	154			45.6 (9.4)	125
Psychotherapie	43.8 (9.1)	107			47.6 (9.2)	86			46.9 (9.5)	74

¹ In Teilen übernommen und übersetzt aus Heißel et al., in prep.

MCS

Sporttherapie	32.0 (9.1)	196	39.2 (11.4)	154	41.2 (11.4)	125
Psychotherapie	30.3 (8.6)	107	37.0 (10.7)	86	40.0 (12.8)	74

*mITT=modified intention-to-treat. BDI-II=Beck Depression Inventory-II. HAM-D=Hamilton Rating Scale for Depression. WSAS=Work and Social Adjustment Scale. WHODAS 2.0=WHO Disability Assessment Schedule 2.0. GSE6=General Self-Efficacy Scale. EQ-5D-5L=EuroQoL-5-dimension-5-level questionnaire. VDS-90=Verhaltensdiagnostiksystem 90. BPNSFS=Basic Psychological Need Satisfaction and Frustration Scale. SF-12= 12-Item Short-Form-Health-Survey. PCS=Physical Component Score. MCS=Mental Component Score.

Tabelle S3. Schweregrad der Depression (BDI-II) in der mITT-Population

	T0		T2		T4	
	Sporttherapie (n=215)	Psychotherapie (n=117)	Sporttherapie (n=164)	Psychotherapie (n=95)	Sporttherapie (n=137)	Psychotherapie (n=79)
Minimal	41 (19.1)	21 (17.9)	78 (47.6)	43 (45.3)	75 (54.7)	46 (58.2)
Leicht	44 (20.5)	27 (23.1)	38 (23.2)	21 (22.1)	23 (16.8)	10 (12.7)
Mittelschwer	71 (33.0)	27 (23.1)	28 (17.1)	16 (16.8)	24 (17.5)	14 (17.7)
Schwer	59 (27.4)	42 (35.9)	20 (12.2)	15 (15.8)	15 (10.9)	9 (11.4)

*Daten sind dargestellt in n (%). mITT=modified intention to treat.
BDI-II=Beck Depression Inventory-II.

Tabelle S4. Grad der Beeinträchtigung nach WSAS in der mITT-Population

	T0		T2		T4	
	Sporttherapie (n=214)	Psychotherapie (n=116)	Sporttherapie (n=163)	Psychotherapie (n=94)	Sporttherapie (n=133)	Psychotherapie (n=79)
Subklinisch	44 (20.6)	18 (15.5)	64 (39.3)	33 (35.1)	71 (52.2)	39 (49.4)
Signifikante Funktionsbeeinträchtigung	74 (34.6)	39 (33.6)	55 (33.7)	32 (34.0)	35 (25.7)	15 (19.0)
Mittelschwere oder schwere Beeinträchtigung	96 (44.9)	59 (50.9)	44 (27.0)	29 (30.9)	30 (22.1)	25 (31.6)

*Daten sind dargestellt in n (%). mITT=modified intention to treat.
WSAS=Work and Social Adjustment Scale

Tabelle S5. GEE-Analyse mit BDI-II als abhängige Variable (Modell 1) in der mITT-Population

	Koeffizient (B)	SE	Wald χ^2	p-Wert
Konstante	17.49 (12.56 to 22.42)	2.51	48.363	<0.001
Zeit				
T0	5.81 (4.37 to 7.25)	0.74	62.476	<0.001
T2	Ref.			
Intervention				
Sporttherapie	-0.85 (-3.34 to 1.63)	1.27	0.453	0.501
Psychotherapie	Ref.			
Intervention*Zeit				
Sporttherapie*T0	0.04 (-1.72 to 1.79)	0.89	0.002	0.966
Psychotherapie*T0	Ref.			
Sporttherapie*T2	Ref.			
Psychotherapie*T2	Ref.			
Alter	-0.00 (-0.1 to 0.10)	0.05	0.001	0.978

BDI-II=Beck Depression Inventory II. mITT=modified intention to treat.
Quasi Likelihood Under Independence Model Criterion (QIC): 71995.931
Corrected QIC (QICC): 71993.978

Tabelle S6. GEE-Analysis mit BDI-II als abhängige Variable (Modell 2) in der mITT-Population

	Koeffizient (B)	SE	Wald χ^2	p-Wert
Konstante	15.21 (10.39 to 20.04)	2.48	38.145	<0.001
Zeit				
T0	7.97 (6.17 to 9.77)	0.92	75.276	<0.001
T2	2.16 (0.90 to 3.42)	0.64	11.322	<0.001
T4	Ref.			
Intervention				
Exercise therapy	0.32 (-2.32 to 2.96)	1.35	0.058	0.810

Psychotherapy	Ref.			
Intervention*Zeit				
Sporttherapie*T0	-1.15 (-3.2 to 1.03)	1.11	1.072	0.301
Psychotherapie*T0	-1.19 (-2.72 to 0.35)	0.78	2.283	0.131
Sporttherapie*T2	Ref.			
Psychotherapie*T2	Ref.			
Sporttherapie*T4	Ref.			
Psychotherapie*T4	Ref.			
Alter	-0.001 (-0.10 to 0.10)	0.05	0.001	0.980

BDI-II=Beck Depression Inventory II. mITT=modified intention to treat.
 Quasi Likelihood Under Independence Model Criterion (QIC): 115707.781
 Corrected QIC (QICC): 115704.300

Tabelle S7. GLMM-Analyse mit HAM-D als abhängige Variable (Modell 3) in der mITT-Population

	Koeffizient (95% CI)	SE	p-Wert
Konstante	9.34 (-1.02 bis 19.70)	5.27	0.077
Zeit			
T0	3.95 (1.99 bis 5.91)	1.00	<0.001
T2	Ref.		
Intervention			
Sporttherapie	-1.60 (-3.43 bis 0.23)	0.93	0.087
Psychotherapie	Ref.		
Intervention*Zeit			
Sporttherapie*T0	1.82 (-0.61 bis 4.26)	1.24	0.142
Psychotherapie*T0	Ref.		
Sporttherapie*T2	Ref.		
Psychotherapie*T2	Ref.		
Alter	0.05 (-0.004 bis 0.10)	0.03	0.069

Akaike information criterion (AIC)=4014.834, und Bayesian Information Criterion (BIC)=4027.918.
 HAM-D=Hamilton Rating Scale for Depression. mITT=modified intention to treat.

Tabelle S8. GLMM-Analyse mit HAMD als abhängige Variable (model 4) in der mITT-Population

	Koeffizient (95% CI)	SE	p-Wert
Konstante	7.26 (-3.17 to 17.68)	5.31	0.172
Zeit			
T0	5.96 (3.94 to 7.99)	1.03	<0.001
T2	2.01 (-0.12 to 4.14)	1.09	0.065
T4	Ref.		
Intervention			
Sporttherapie	-0.44 (-2.38 to 1.51)	0.99	0.660
Psychotherapie	Ref.		
Intervention*Zeit			
Sporttherapie *T0	0.66 (-1.88 to 3.19)	1.29	0.612
Psychotherapie*T0	Ref.		
Sporttherapie *T2	-1.17 (-3.84 to 1.50)	1.36	0.391
Psychotherapie*T2	Ref.		
Sporttherapie *T4	Ref.		
Psychotherapie*T4	Ref.		
Alter	0.05 (0.005 to 0.100)	0.02	0.030

*Akaike information criterion (AIC) 5603.582, and Bayesian Information Criterion (BIC)=5617.670.
 HAM-D=Hamilton Rating Scale for Depression. mITT=modified intention to treat.

Tabelle S9. GLMM-Analyse mit WSAS als abhängige Variable (Modell 5) in der mITT-Population

	Koeffizient (95% CI)	SE	p-Wert
Konstante	11.39 (-3.22 to 26.01)	7.45	0.127
Zeit			
T0	5.25 (2.50 to 8.01)	1.40	<0.001
T2	Ref.		
Intervention			
Sporttherapie	-1.56 (-4.15 to 1.04)	1.32	0.239
Psychotherapie	Ref.		
Intervention*Zeit			
Sporttherapie *T0	-0.34 (-3.78 to 3.10)	1.75	0.846
Psychotherapie*T0	Ref.		
Sporttherapie *T2	Ref.		
Psychotherapie*T2	Ref.		
Alter	0.07 (-0.01 to 0.15)	0.04	0.067

*Akaike information criterion (AIC)=4380.530, and Bayesian Information Criterion (BIC)=4393.561.
WSAS=Work and Social Adjustment Scale. mITT=modified intention to treat.

Tabelle S10. GLMM-Analyse mit WSAS als abhängige Variable (Modell 6) in der mITT-Population

	Koeffizient (95% CI)	SE	p-Wert
Konstante	9.55 (-5.24 bis 24.33)	7.53	0.205
Zeit			
T0	7.12 (4.18 bis 10.06)	1.50	<0.001
T2	1.88 (-1.19 bis 4.96)	1.57	0.230
T4	Ref.		
Intervention			
Sporttherapie	-1.77 (-4.64 bis 1.10)	1.46	0.225
Psychotherapie	Ref.		
Intervention*Zeit			
Sporttherapie*T0	-0.16 (-3.84 bis 3.52)	1.87	0.933
Psychotherapie*T0	Ref.		
Sporttherapie*T2	0.15 (-3.71 bis 4.02)	1.97	0.938
Psychotherapie*T2	Ref.		
Sporttherapie*T4	Ref.		
Psychotherapie*T4	Ref.		
Alter	0.08 (0.01 bis 0.14)	0.03	0.021

Akaike information criterion (AIC)=6091.883, und Bayesian Information Criterion (BIC)=6105.936.
WSAS=Work and Social Adjustment Scale. mITT=modified intention to treat

Tabelle S11: Beschreibung der Interventionskosten der EXP- und TAU-Gruppe

Interventionskosten	EXP (n = 207)	TAU (n = 124)	p-Wert
MW (SD)	681.20 (269.87)	993.52 (411.08)	<0.001
min-max	125.92 – 1201.60	149.00 -1933.00	
Median (IQR)	709.34 (469.78 - 916.82)	1025.00 (718.38 – 1270.00)	

IQR: Interquartile range; SD: Standard deviation.

Tabelle S12: Darstellung der Interventionskosten inkl. (anschließende) Psychotherapiekosten

	EXP (n = 207)	TAU (n = 124)	p-Wert
MW (SD)	1356.40 (1550.76)	2411.26 (1603.62)	<0.001
min-max	125.92 – 8909.97	160.00-7157.50	
Median (IQR)	798.50 (509.34-1062.00)	2072.50 (1250.13-3122.50)	

IQR: Interquartile range; SD: Standard deviation.

Tabelle S13. Deskriptive Darstellung der Gesundheitskosten der mITT Population

Kostenart	vor Interventionsbeginn		nach Interventionsbeginn	
	EXP (n=207)	TAU (n=124)	EXP (n=207)	TAU (n=124)
Ambulante Kosten	1061·2 (4157·5)	712·2 (517·2)	1031·8 (4326·4)	685·4 (726·2)
Stationäre Kosten	484·3 (1846·1)	589·7 (1788·3)	621·3 (3058·0)	1235·0 (4099·8)
Medikamentenkosten	307·8 (1294·6)	673·3 (4735·6)	437·4 (2237·4)	562·0 (3342·2)
Heil- und Hilfsmittel	212·5 (450·1)	271·6 (565·5)	227·2 (529·1)	226·3 (421·1)

Daten sind dargestellt in MW (SD).

Tabelle S14: Difference-in-Difference Modell zur Analyse der Gesamtkosten einschließlich der Interventionskosten (10 Monate)

	Koef.	Std. Fehler	p-Wert	95 % CI	
Konstante	11042·86	3355·27	0·001	4454·19	17631·53
Post period	-2717·80	410·04	<0·001	-3523·00	-1912·61
Interventionsgruppe (EXP)	-8293·11	1823·65	<0·001	-11874·17	-4712·06
Post period x Intervention (EXP)	1232·58	518·51	0·018	214·39	2250·77

Tabelle S15: Difference-in-Difference Modell zur Analyse der Gesamtkosten (ohne Interventionskosten)

	Koef.	Std. Fehler	p-Wert	95 % CI	
Konstante	8834·45	3285·14	0·007	2383·50	15285·40
Post period	-306·55	393·87	0·437	-1079·97	466·88
Interventionsgruppe (EXP)	-7401·46	1795·50	<0·001	-10927·24	-3875·67
Post period x Intervention (EXP)	174·81	498·06	0·726	-803·213	1152·83