

## Studienbericht (Clinical Study Report)

# My7steps App Pilotstudie

**Datum des Berichts:** 23.09.2022

**Untersuchtes Medizinprodukt:** My7steps App (Hersteller: *Ipsos Healthcare GmbH*)

**Untersuchte Indikationen (ICD-10):** F32.0, F32.1, F33.0, F33.1, F43.2

**Studiendesign:** Einarmige Post Market Clinical Follow-up mit Prä-Post-Vergleich einer webbasierten, selbstadministrierten Intervention für 80 Patienten aus dem Spektrum der depressiven Erkrankungen.

**Studienphase:** IV

**Datenerhebung:** 11.11.2021 – 06.04.2022

**Leiter der klinischen Prüfung (LKP):** Prof. Dr. med. Helmut Peter (Verhaltenstherapie Falkenried MVZ GmbH; Peter@vt-falkenried.de)

**Wissenschaftlicher Leiter:** Dipl.-Psych. Pelle Bernhold (Ipsos GmbH; p.bernhold@ipsocontext.org)

**Studienplan:** Nach ISO 14155 unter Einhaltung der Good Clinical Practices (ICH-GCP) einschließlich der Archivierung wesentlicher Dokumente.

**DRKS-ID der Studie:** DRKS00027925

**Ethikvotum:** Positives Votum/Zustimmende Bewertung durch die Ethikkommission der Ärztekammer Hamburg (Vorgangs-Nr.: 2021-10526\_1-BO-ff)

**Verfasser des Berichts:** Pelle Bernhold, Ingeborg Mißmahl-Grusche, Maryam Gardisi, Georgia Koutsouradi, Angelika Lichtwer, Dr. Ralph Grobecker, Prof. Dr. med. Helmut Peter

## Inhaltsverzeichnis

Inhaltsverzeichnis	I
Tabellenverzeichnis	IV
Abbildungsverzeichnis	V
Abkürzungsverzeichnis	VI
<b>1 Synopsis</b>	<b>1</b>
<b>2 Ethische und regulatorische Aspekte</b>	<b>9</b>
2.1 Unabhängige Ethikkommissionen	9
2.2 Ethische Durchführung der klinischen Prüfung	9
2.3 Berücksichtigte gesetzliche Bestimmungen, Leitlinien und Normen	9
2.4 Behördenmeldungen, Genehmigungen und Registrierungen	10
2.5 Aufklärung und Einwilligung der Studienteilnehmer:innen sowie der Prüfer	10
2.6 Probandenversicherung	11
2.7 Datenschutz	11
<b>3 Organisationsstruktur</b>	<b>14</b>
3.1 Sponsor	14
3.2 Weitere beteiligte Einrichtungen und Komitees	15
3.2.1 Advisory Expert	15
3.2.2 Zentrale Organisationseinheiten	15
3.2.3 Prüfer und Prüfzentren	15
<b>4 Finanzierung</b>	<b>16</b>
<b>5 Wissenschaftlicher Hintergrund</b>	<b>17</b>
5.1 Wissenschaftlicher Hintergrund psychischer Störungen: Depressionen und Anpassungsstörungen	17
5.2 Wissenschaftlicher Hintergrund der technologie- bzw. internet-gestützten Interventionen zur Behandlung von psychischen Störungen	18
5.3 Rationale der klinischen Prüfung	20
<b>6 Ziele der klinischen Prüfung</b>	<b>21</b>
6.1 Primäres Ziel	21
6.2 Sekundäre Ziele	21
<b>7 Studienplan</b>	<b>22</b>
7.1 Allgemeines Studiendesign	22
7.2 Zeitplan	22
7.3 Diskussion des Studiendesigns	23

<b>7.4 Auswahl der Studienpopulation</b>	<b>23</b>
7.4.1 Ein- und Ausschlusskriterien	24
7.4.2 Nachträglicher Ausschluss von Prüfungsteilnehmern	25
7.4.3 Rekrutierungsmethoden	26
<b>7.5 Behandlung</b>	<b>26</b>
7.5.1 Angewendete Behandlung	26
7.5.2 Beschreibung der Prüfintervention	27
7.5.3 Vorgesehene Verwendung der Prüfintervention	29
7.5.4 Änderungen der Prüfintervention während der klinischen Studie	30
7.5.5 Ausgabe und Rücknahme der Prüfintervention	30
7.5.6 Verblindung	30
7.5.7 Vorangegangene Therapie und Begleittherapie	30
7.5.8 Weiterbehandlung nach Ende der klinischen Prüfung	31
<b>7.6 Wirksamkeits- und Sicherheitsparameter</b>	<b>31</b>
7.6.1 Messung der Wirksamkeits- und Sicherheitsparameter	31
7.6.2 Angemessenheit des Messverfahrens	37
<b>7.7 Sicherstellung der Datenqualität</b>	<b>37</b>
7.7.1 Monitoring	37
7.7.2 Audits / Inspektionen	37
<b>7.8 Dokumentation</b>	<b>37</b>
7.8.1 Datenmanagement	37
<b>7.9 Statistische Methoden</b>	<b>39</b>
7.9.1 Hypothesen und Endpunkte	39
7.9.2 Subgruppen und Einflussfaktoren (Kovariaten)	39
7.9.3 Explorative Fragestellungen	40
7.9.4 Analyseverfahren	40
7.9.5 Analyse und fehlende Werte	45
<b>7.10 Festlegung der Fallzahl</b>	<b>45</b>
7.10.1 Ausgangssituation und Fallzahl	45
<b>7.11 Änderungen des Prüfplans</b>	<b>45</b>
<b>8 Stichprobe</b>	<b>47</b>
<b>8.1 Stichprobe der Studie</b>	<b>47</b>
<b>8.2 Protokollverletzungen im Rahmen der Patientenrekrutierung und -erhebung</b>	<b>48</b>
<b>9 Ergebnisse: Wirksamkeitsevaluation</b>	<b>49</b>

<b>9.1</b>	<b>Demographische Angaben zur Stichprobe</b>	<b>49</b>
<b>9.2</b>	<b>Dokumentation der weiteren Kontakte zwischen Studienteilnehmer:innen und Prüfer</b>	<b>50</b>
<b>9.3</b>	<b>Analysen zur Wirksamkeit</b>	<b>51</b>
9.3.1	Wirksamkeitsanalyse primärer Endpunkt	52
9.3.2	Wirksamkeitsanalysen sekundäre Endpunkte	53
9.3.3	Response-Analysen	54
9.3.4	Sensitivitätsanalysen primärer Endpunkt	55
9.3.5	Subgruppenanalysen	59
9.3.6	Explorative Analysen	61
<b>10</b>	<b>Sicherheitsevaluation</b>	<b>63</b>
<b>10.1</b>	<b>Auswertung der Sicherheitsparameter</b>	<b>63</b>
10.1.1	Vorkommnisse und unerwünschte Ereignisse	63
10.1.2	Laufende Risikoevaluation	63
10.1.3	Studienteilnehmer:innen mit reliabler Verschlechterung	63
10.1.4	Bewertung der Intervention durch die Studienteilnehmer:innen	63
<b>10.2</b>	<b>Sicherheitsschlussfolgerung</b>	<b>64</b>
<b>11</b>	<b>Diskussion und Schlussfolgerungen</b>	<b>65</b>
<b>11.1</b>	<b>Wirksamkeits- und Sicherheitsergebnisse</b>	<b>65</b>
<b>11.2</b>	<b>Verhältnis von Risiken und Nutzen</b>	<b>66</b>
<b>11.3</b>	<b>Einordnung der Ergebnisse in den wissenschaftlichen Kontext</b>	<b>66</b>
<b>11.4</b>	<b>Klinische Relevanz der Ergebnisse</b>	<b>67</b>
<b>11.5</b>	<b>Implikationen für die Durchführung zukünftiger Studien</b>	<b>68</b>
<b>11.6</b>	<b>Einschränkungen der Studien</b>	<b>68</b>
<b>12</b>	<b>Analysen mit externen Kontrollgruppen (literaturbasierter Kontrollarm)</b>	<b>70</b>
<b>12.1</b>	<b>Methodisches Vorgehen zur Erstellung des literaturbasierten Kontrollarms</b>	<b>70</b>
<b>12.2</b>	<b>Ergebnisse des literaturbasierten Kontrollarms</b>	<b>72</b>
<b>12.3</b>	<b>Diskussion der Ergebnisse des literaturbasierten Kontrollarms</b>	<b>78</b>
<b>13</b>	<b>Literaturverzeichnis</b>	<b>82</b>
<b>14</b>	<b>Anhang</b>	<b>89</b>
<b>14.1</b>	<b>Mittelwerte, Standardabweichungen, Zellbesetzungen und Streudiagramme für die Faktoren der Sensitivitätsanalyse</b>	<b>89</b>
<b>14.2</b>	<b>Leitfaden Check-Up Calls</b>	<b>91</b>

## Tabellenverzeichnis

Tabelle 1: Zeitlicher Rahmen der Studie	23
Tabelle 2: Zusammenfassung der erhobenen Parameter	34
Tabelle 3: Demographische Angaben zur Stichprobe	49
Tabelle 4: Mittelwerte und Effektstärken aller Endpunkte über die Messzeitpunkte (primäre Wirksamkeitsanalysen)	52
Tabelle 5: Response-Analysen für den primären Endpunkt	55
Tabelle 6: Kennwerte der mehrfaktoriellen ANCOVA mit Messwiederholungen zur Überprüfung möglicher Einflussfaktoren	56
Tabelle 7: Mittelwerte und Effektstärken des primären Endpunkts über die Messzeitpunkte (PP vs. ITT)	59
Tabelle 8: Depressions-Mittelwerte und -Standardabweichungen sowie Gruppengrößen der Diagnosegruppen	60
Tabelle 9: Kennwerte der einfaktorielle ANOVA mit Messwiederholungen zur Überprüfung der Indikationsgruppen	60
Tabelle 10: Kennwerte der Regressions-ANOVA zum quadratischen Zusammenhang von Behandlungsdauer und Wirksamkeit	61
Tabelle 11: Übersicht der Studien des literaturbasierten Kontrollarms	75
Tabelle 12: Stichprobenmerkmale der vorliegenden Untersuchung und der literaturbasierten Kontrollgruppen	76
Tabelle 13: Mittelwerte und Effektstärken des primären Endpunkts über die Messzeitpunkte und Studien	77
Tabelle 14: Baseline-Depressions-Mittelwerte und -Standardabweichungen sowie Zellbesetzungen für die Faktoren der Sensitivitätsanalyse	89
Tabelle 15: Post-Depressions-Mittelwerte und -Standardabweichungen sowie Zellbesetzungen für die Faktoren der Sensitivitätsanalyse	90

## Abbildungsverzeichnis

Abbildung 1: Übersicht der Prozedur und Datenerhebung.	36
Abbildung 2: Flussdiagramm der Registrierung, Zuordnung und Analyse	47
Abbildung 3: Baseline- und Post-Depressionswerte aufgeteilt nach Geschlecht ( $n = 87$ )	57
Abbildung 4: Baseline- und Post-Depressionswerte aufgeteilt nach Bildung ( $n = 89$ )	57
Abbildung 5: Baseline- und Post-Depressionswerte aufgeteilt nach Zusatzbehandlung ( $n = 89$ )	58
Abbildung 6: Streudiagramm Behandlungsdauer und Change-Scores des primären Endpunkts ( $n = 80$ )	62
Abbildung 7: Methodik der systematischen Identifikation von Studien, die für den literaturbasierten Kontrollarm geeignet sind	71

## Abkürzungsverzeichnis

AE	Adverse Events
AIS	Arztinformationssystem
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
CD-RISC-10	Connor-Davidson Resilience Scale
CRF	Case Report Form
DFI	Daily Functionality Impairment Scale
DiGA	Digitale Gesundheitsanwendung
DIN	Deutsches Institut für Normung
DRKS	Deutsches Register klinischer Studien
DSGVO	Datenschutz-Grundverordnung
DSM-V	Diagnostic and Statistical Manual of Mental Disorders, fünfte Edition
EN	Europäische Norm
FPFV	First Patient First Visit (Studienbeginn)
ICH-GCP	International anerkannte Leitlinien der Good Clinical Practice
ICD-10	International Statistical Classification of Diseases and Related Health Problems
IG	Interventionsgruppe
ITT	Intention-to-Treat (Prinzip der Auswertung, inklusive ausgeschiedener Studienteilnehmer:innen)
IMI	Internet and mobile-based interventions
ISO	Internationale Organisation für Normung
KG	Kontrollgruppe
KIS	Krankenhausinformationssystem
KVT	Kognitive Verhaltenstherapie
LPFV	Last Patient First Visit (Rekrutierungsende)
LPLV	Last Patient Last Visit (Studienende)
MCID	Minimal clinically important difference (minimaler klinisch bedeutsamer Unterschied)
MDD	Medical Device Directive
MPG	Medizinproduktegesetz
MPDG	Medizinprodukterecht-Durchführungsgesetz
MPAMIV	Medizinprodukte-Anwendermelde- und Informationsverordnung
MPKPV	Verordnung über klinische Prüfungen von Medizinprodukten
MPSV	Medizinprodukte-Sicherheitsplanverordnung
MVZ	Medizinisches Versorgungszentrum
PHQ-9	Patient Health Questionnaire
RCT	Randomized Controlled Trial
SMD	Standardisierte mittlere Differenz
SGB	Sozialgesetzbuch

SRM	Standardized response mean (standardisierte Mittelwertsdifferenz des Behandlungsergebnisses)
VBC	Value Based Counseling

# 1 Synopsis

<b>Sponsor</b>	<p>Ipsog Healthcare GmbH  Münsterplatz 13  78462 Konstanz</p>
----------------	-----------------------------------------------------------------------

<b>Leiter der klinischen Prüfung</b>	<p>Prof. Dr. med. Helmut Peter  Verhaltenstherapie Falkenried MVZ GmbH  Falkenried 7, 20251 Hamburg  E-Mail: Peter@vt-falkenried.de  Tel.: 0049 40 688 930 100</p>
--------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>Wissenschaftlicher Leiter</b>	<p>Dipl.-Psych. Pelle Bernhold  Ipsog GmbH  Münsterplatz 13, 78462 Konstanz  E-Mail: p.bernhold@ipsocontext.org  Telefon: 0049 176 43388782</p>
----------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>Titel der klinischen Prüfung und Name des Medizinprodukts</b>	My7steps App Pilotstudie
	My7steps App

<b>Art der Prüfung</b>	Monozentrisch (online)
	Nationale klinische Prüfung
	PMCF - Prüfung als sonstige klinische Prüfung
	Studienphase: IV

Studiendesign	
Studiendesign	Intra-individueller Prä-Post-Vergleich
Anzahl Studienarme	1
Randomisierung	Nicht randomisiert/einarmig
Verblindung	Offen/einarmig
Placebo-kontrolliert	Unkontrolliert/einarmig

Studienplan	
ISO-Konformität	Die vorliegende Studie wurde gemäß der aktuellen Version des Prüfplans unter Einhaltung der Good Clinical Practices (ICH-GCP) einschließlich der Archivierung essenzieller Dokumente durchgeführt. Der Studienplan wurde nach ISO 14155 erstellt.
ID im Deutschen Register Klinischer Studien (DRKS)	DRKS00027925
Ethikvotum	Positives Votum/Zustimmende Bewertung durch die Ethikkommission der Ärztekammer Hamburg (Vorgangs-Nr.: 2021-10526_1-BO-ff).

Finanzierung	Eigenmittel der Verhaltenstherapie Falkenried MVZ GmbH Eigenmittel und Spenden der Ipsos gGmbH
--------------	---------------------------------------------------------------------------------------------------

Zusammenfassung der Studie – 1 – Hintergrund
<p>Depressionen gehören weltweit zu den häufigsten psychischen Erkrankungen, die mit körperlichen Beschwerden und einer Beeinträchtigung der Lebensqualität und der täglichen Funktionsfähigkeit einhergehen. Eine aktuelle Studie in Deutschland ergab eine Gesamtprävalenz von 10 % depressiver Symptome in der Allgemeinbevölkerung. Verbreitung, Schwere und Folgen depressiver Symptomatiken werden aber oftmals noch immer unterschätzt. Während Epidemiologie, Diagnostik und Behandlungsansätze bei depressiven Symptomatiken seit längerem systematisch erforscht werden, sind die gesicherten Befunde zur Diagnose Anpassungsstörung überschaubar. Anpassungsstörungen werden durch eine emotionale Reaktion auf ein kritisches Lebensereignis oder eine Reihe von belastenden Ereignissen hervorgerufen, wobei sich die emotionale Reaktion in der Mitte auf einem Kontinuum zwischen</p>

normalen Stressreaktionen und pathologischen Reaktionen befindet. Die Inzidenz von Anpassungsstörungen wird in Deutschland in Patientengruppen auf 5 - 21% geschätzt, in der Allgemeinbevölkerung scheint sie aber eher selten vorzuliegen.

Value Based Counseling (VBC) ist ein kultursensibler Ansatz, der das Ziel hat, durch die Entwicklung von Bewältigungskompetenzen Betroffene zu befähigen, mit Stressoren umzugehen und belastende Situationen besser zu bewältigen. Wirksamkeitsstudien von Face-to-Face und Online stattfindendem Value Based Counseling konnten zeigen, dass diese Interventionen Symptome von Depression und Angst signifikant reduzieren können und sich der Ansatz auch für die Behandlung von Anpassungsstörungen anbietet, da er psychosoziale Stressoren als direkte und indirekte Ursachen für psychische Probleme behandelt. Auch für internet and mobile-based interventions (IMIs; technologie- bzw. internet-gestützte Interventionen) konnte in den letzten Jahren gezeigt werden, dass sie bei der Behandlung von Depressionen, Angstzuständen und vielen anderen psychischen Erkrankungen ähnlich wirksam wie klassische Therapieansätze sind. IMIs konnten sich so zu einer evidenzbasierten Alternative zu persönlichen psychotherapeutischen Behandlungen entwickeln und stellen dank der hohen Popularität und Verbreitung von internetfähigen Devices weltweit eine Behandlungschance in ressourcenbeschränkten Umgebungen, bei überfüllten psychotherapeutischen Wartelisten und für schwer zu erreichende Gruppen dar. IMIs bieten dabei weitere Vorteile wie Erschwinglichkeit, Benutzerfreundlichkeit und Anonymität. Darüber hinaus kann die Bereitstellung professioneller psychiatrischer Versorgung für Menschen mit milderem psychiatrischen Symptomen die Belastung traditioneller psychiatrischer Dienste reduzieren und die psychische Gesundheit der Bevölkerung verbessern.

Die Entwicklung und Erforschung technologie-gestützter Selbsthilfe-Intervention für die Behandlung depressiver Symptomatik ist also sowohl medizinisch als auch politisch und wirtschaftlich relevant. Die von Ipsog Healthcare GmbH angebotene webbasierte Selbsthilfe-Intervention My7steps App, die auf dem VBC-Ansatz basiert, hat als IMI das Potential die depressiven Symptomatik insbesondere von Betroffenen auf Wartelisten unmittelbar nach der Diagnose zu reduzieren und damit die erste Stufe eines "stepped care"-Ansatzes darzustellen.

### Zusammenfassung der Studie – 2 – Studienziele

Primäres Ziel der vorliegenden Studie war es, zu überprüfen, ob die Versorgung von Betroffenen aus dem Spektrum der depressiven Erkrankungen durch die Webanwendung *My7steps App* der *Ipsog Healthcare GmbH* verbessert wird. Dies sollte sich in einer Reduktion der depressiven Symptomatik zeigen. Sekundäre Ziele bezogen sich auf die Stärkung der persönlichen Belastbarkeit und die Verringerung krankheitsbedingter Probleme im Alltag durch *My7steps App*.

### Zusammenfassung der Studie – 3 – Rekrutierungsmaßnahmen

Die Rekrutierung erfolgt über Wartelisten und mit Hilfe von mehrsprachigen Flyern und Beiträgen in Arztpraxen, Vereinen, Sozialen Netzwerken und Newslettern. Es wurde eine

Aufwandentschädigung von 50€ am Ende der Teilnahme ausgezahlt. Es wurde auf Einheitlichkeit beim Aufnahmeverfahren in die Studie geachtet; alle rekrutierten Studienteilnehmer:innen wurden vom Studienteam im standardisierten Einschlussverfahren aufgenommen.

#### Zusammenfassung der Studie – 4 – Indikationen (ICD-10)

Präspezifizierte Subgruppen entsprechend Studienprotokoll:

##### Indikationsgruppe 1

- F32.0 Leichte depressive Episode
- F32.1 Mittelgradige depressive Episode

##### Indikationsgruppe 2

- F33.0 Rezidivierende depressive Störung, gegenwärtig leichte Episode
- F33.1 Rezidivierende depressive Störung, gegenwärtig mittelgradig Episode

##### Indikationsgruppe 3

- F43.2 Anpassungsstörungen

#### Zusammenfassung der Studie – 5 – Ein- und Ausschlusskriterien

##### Einschlusskriterien

- Rechtskräftige Einwilligung
- Die Fähigkeit Deutsch, Englisch, Arabisch, oder Farsi/Dari zu sprechen und zu verstehen
- Alter  $\geq$  18 Jahre
- Betroffene, die in Deutschland leben
- Psychotherapeutische Diagnose einer für die Behandlung angezeigten Indikationen (F32.0/1, F33.0/1, F43.2)

##### Ausschlusskriterien

- Betroffene, die an anderen interventionellen Studien teilnehmen
- Personen, die in einem Abhängigkeits- / Arbeitsverhältnis zum Sponsor oder Prüfer stehen
- Personen in Unterbringung in einer Anstalt aufgrund gerichtlicher oder behördlicher Anordnung.

- Betroffene mit Suizidalität, schwerer Depression, Suchterkrankung, schizophrener Symptomatik, psychotischer Symptomatik, wahnhafter Störung, hirnorganischer Störung (z.B. Delir, Demenz)

### Zusammenfassung der Studie – 6 – Behandlung

*My7steps App* ist eine webbasierte Selbsthilfe-Intervention, die aus sieben Modulen zur Psychoedukation, Verhaltensaktivierung, Problemlösung, kognitiven Verhaltenstherapie, Achtsamkeit und Ressourcenorientierung besteht. Weitere Module befassen sich mit der Symptomeinschätzung, der Identifizierung psychosozialer Stressoren, Achtsamkeitsübungen, positiver Psychologie und reflektierendem Journaling. Durch das responsive Webdesign kann der Benutzer das Programm im Internet, auf einem Tablet oder Mobiltelefon verfolgen. Die tägliche Nutzung wird während der Phase der geführten Problemlösung empfohlen. Die App ist so gestaltet, dass der Benutzer die Module durchlaufen kann, während er seinem normalen Alltag und seiner Arbeit nachgeht. Für die Bearbeitung eines Modulschritts, die für einen Schritt der Anwendung etwa bis zu einer Stunde dauert, benötigen Nutzer ununterbrochenen Zugang zum Internet.

Empfohlen wird eine Anwendungsdauer von drei Monaten für *My7steps App*, um ausreichend Zeit für die Integration persönlicher Verhaltensänderungen in das eigene Leben zu gewähren. Abhängig von der selbstgewählten Teilnahmehäufigkeit und -intensität kann diese Dauer aber stark variieren. Im Rahmen der Pilotstudie wurden die Studienteilnehmer:innen gebeten sich für eine regelmäßige und intensivere Nutzung zu entscheiden, um einer übermäßig starken Varianz vorzubeugen.

### Zusammenfassung der Studie – 7 – Messinstrumente

Patient Health Questionnaire (PHQ-9)

Connor-Davidson Resilience Scale (CD-Risc-10)

Daily Functionality Impairment Scale (DFI)

### Zusammenfassung der Studie – 8 – Zeitplan

Einschluss des ersten Patienten (FPFV): 11.11.2021

Prüfungsende des letzten Patienten (LPLV): 06.04.2022

### Zusammenfassung der Studie – 9 – Studienablauf

Im Rahmen einer diagnostischen Videovisite wurden die Ein- und Ausschlusskriterien überprüft und Diagnosen gestellt. Die erste telefonische Studienvisite erfolgte umgehend und umfasste die Baseline-Messung und Ausgabe der Intervention. Nach Abschluss der selbstadministrierten Intervention wurde im Rahmen der zweiten telefonischen Studienvisite die Post-Messung und Rücknahme der Intervention vorgenommen.

### Zusammenfassung der Studie – 10 – Studienpopulation

89 Studienteilnehmer:innen wurden in die Studie aufgenommen. Die Dropout-Rate betrug 10.1% (9 von 89). Es gab keinen Antrag auf Datenlöschung, entsprechend gingen die Daten von 89 Studienteilnehmer:innen, davon 75.9% weiblich, mit einem Durchschnittsalter von 30.48 Jahren ( $SD = 7.77$ ) in die Analysen ein. 16.9% der Stichprobe wies die ICD-10 Diagnose leichte Depression (F32.0) auf, 19.1% eine mittelschwere Depression (F32.1), 20.2% eine rezidivierende leichte Depression (F33.0), weitere 20.2% eine rezidivierende mittelschwere Depression (F33.1) und 23.6% eine Anpassungsstörung (F43.2). Nach den PHQ-9-Kategorien wiesen 4.5% der Stichprobe zur Baseline subklinische Depressionswerte auf dem primären Endpunkt auf, 38.2% eine leichte depressive Symptomatik, 34.8% eine moderate, 18.0% eine mittelschwere und 4.5% eine schwere depressive Symptomatik. In der Stichprobe wiesen 4.5% zur Baseline subklinische Depressionswerte auf dem primären Endpunkt auf, 38.2% eine leichte depressive Symptomatik, 34.8% eine moderate, 18.0% eine mittelschwere und 4.5% eine schwere depressive Symptomatik nach den PHQ-9-Kategorien.

### Zusammenfassung der Studie – 11 – Statistische Methoden

Auswertung im Rahmen von *Intention-to-Treat*-Analysen, fehlende Werte wurden in einem *Baseline-Observation-Carried-Forward*-Verfahren ersetzt.

Wirksamkeitsanalysen: t-Tests für abhängige Stichproben (bei Abweichungen von der Normalverteilung zusätzlich Wilcoxon-Vorzeichen-Rang-Test)

Sensitivitätsanalysen: mehrfaktorielle ANCOVA mit Messwiederholungen; t-Tests für abhängige Stichproben (ggf. ergänzt durch Wilcoxon-Vorzeichen-Rang-Test); *Per-Protokoll*-Analyse

Subgruppenanalysen: einfaktorielle ANOVA mit Messwiederholungen; t-Test für abhängige Stichproben

Literaturbasierter Kontrollarm: deskriptive Vergleiche mit externen Kontrollgruppen (nicht inferenzstatistisch)

### Zusammenfassung der Studie – 12 – Ergebnisse zur Wirksamkeit

Die Ergebnisse der Wirksamkeitsanalysen lassen den Schluss zu, dass die Intervention in Form einer Reduktion depressiver Symptomatiken und einer Verbesserung der Resilienz und Alltagsfunktionalität wirksam ist: Die depressiven Symptomatiken waren zum Post-Messzeitpunkt signifikant gegenüber der Baseline-Messung reduziert ( $d = -0.93$ ). Die Wirksamkeit zeigte sich ebenfalls in den sekundären Endpunkten durch eine signifikant erhöhte Resilienz ( $d = 0.73$ ), eine geringere Beeinträchtigung der Alltagsfunktionalität ( $d = -0.95$ ) sowie eine deutliche Reduktion der subjektiven Symptomstärken ( $d = -1.89$ ).

In den Sensitivitätsanalysen konnten keine generellen Wirksamkeitsunterschiede hinsichtlich Geschlecht, Bildung, Alter und eventueller Zusatzbehandlungen während der Teilnahme festgestellt werden, und die Wirksamkeitsanalyse bzgl. des primären Endpunkts wurde bestätigt ( $d = 0.81$ ). Der Befund bestätigte sich des Weiteren auch nach Ausschluss aller nicht adhärenen Studienteilnehmer:innen in einer *Per-Protokoll*-Analyse ( $d = -1.09$ ) sowie nach Ausschluss aller Studienteilnehmer:innen, die von einer Zusatzbehandlung während des Studienzeitraums profitiert hatten ( $d = -0.98$ ).

Die ermittelte klinische Response-Rate von 60.8%, partielle Remissions-Rate von 54.9% und vollständige Remissions-Rate von 21.6% war mindestens vergleichbar und teils höher als die Raten einschlägiger Referenzpublikationen als auch vergleichbarer Medizinprodukte.

Eine Überprüfung der verschiedenen Indikationsgruppen hinsichtlich des Behandlungseffekts auf den primären Endpunkt ergab des Weiteren keine signifikanten Unterschiede, sodass angenommen werden kann, dass die Behandlung für alle fünf untersuchten Indikation (F32.0, F32.1, F33.0, F33.1, F43.2) gleichermaßen wirksam ist. Eine gesonderte Betrachtung der Studienteilnehmer:innen mit subklinischer bis leichter Depression ergab eine große Effektstärke ( $d = -1.34$ ), sodass auch für diese weniger schweren Fälle von einer Wirksamkeit der Intervention ausgegangen werden kann.

Deskriptive Analysen mit sechs vergleichbaren, externen Kontrollgruppen zeigten in allen Fällen einen hypothetischen, positiven between-group Interventionseffekt der *My7steps App* ( $d = -0.23$  bis  $d = -1.19$ ). Die Höhe der Differenz der within-Effekte reicht von  $\Delta$  Cohen's  $d = 0.19$  bis  $\Delta$  Cohen's  $d = 0.72$  zu Gunsten der vorliegenden Intervention. Da unter den Kontrollgruppendesigns auch solche waren, die explizit die Verwendung von Medikation und Therapie zuließen und teilweise alternative Interventionen anboten, scheint der gefundene Interventionseffekt der *My7steps App* deutlich über die verlaufs- oder begleittherapiebedingten Veränderungen der Kontrollgruppen hinauszugehen. Dies ist ein vielversprechender Hinweis auf eine Wirksamkeit der *My7steps App* unter kontrollierten Bedingungen.

### Zusammenfassung der Studie – 13 – Ergebnisse zur Sicherheit

Im Rahmen der Sicherheitsevaluation konnten keine besonderen Risiken oder Belastungen festgestellt werden.

### Zusammenfassung der Studie – 14 – Schlussfolgerung

Die webbasierten Selbsthilfe-Intervention *My7steps App* scheint nach den vorliegenden Daten eine wirksame Intervention bei der Behandlung von leichten bis mittelgradigen depressiven Symptomatiken und Anpassungsstörungen zu sein. Die Intervention wies eine große Effektstärke auf ( $d = 0.81$  bis  $d = 0.93$ ), welche auch im deskriptiven Vergleich mit sechs externen Kontrollgruppen bestand hatte ( $d = -0.23$  bis  $d = -1.19$ ); die Effekte konnten als klinisch relevant eingestuft werden. Die Intervention hat damit das Potential depressive Symptomatiken nach der Diagnose erfolgreich zu behandeln und die erste Stufe eines "stepped care"-Ansatzes zu bilden.

## 2 Ethische und regulatorische Aspekte

### 2.1 Unabhängige Ethikkommissionen

Die Studie und Amendments wurden von einer unabhängigen Ethikkommission begutachtet:

- Federführende Ethikkommission: Ethikkommission der Ärztekammer Hamburg
- Vorsitzender der federführenden Ethikkommission: Prof. Dr. med. Rolf Stahl (Innere Medizin, Universitätsklinikum Hamburg-Eppendorf)
  - Antragsdatum: 16.03.2021
  - Vorlagen-Nr.: 2021-10526-BO-ff
  - Votum: Positives Votum/Zustimmende Bewertung
  - Datum des positiven Votums / der zustimmenden Bewertung: 14.06.2021

Die zustimmende Bewertung der Ethikkommission der Ärztekammer Hamburg war Voraussetzung für die Initialisierung der Studie. Etwaige Studienaktivitäten im Sinne der klinischen Prüfung wurden vorher nicht begonnen.

### 2.2 Ethische Durchführung der klinischen Prüfung

In der Erstellung des Prüfplans sowie Beschreibung etwaiger Änderungen des Prüfplans kam/kommt die „Declaration of Helsinki - Medical Research Involving Human Subjects“ in ihrer Aktualisierung von 2013 zur Anwendung.

Kritische ethische Aspekte des Prüfplans lagen zum Zeitpunkt der Erstellung des Prüfplans und im Folgenden nicht vor.

### 2.3 Berücksichtigte gesetzliche Bestimmungen, Leitlinien und Normen

Zur Erstellung des verwendeten Prüfplans und des Abschlussberichts wurden die international anerkannten Leitlinien der Good Clinical Practice (ICH-GCP E3-Guideline) zugrunde gelegt. Weiterhin wurden die geltenden gesetzlichen Bestimmungen und Normen, insbesondere das Medizinprodukte-Durchführungsgesetz (MPDG), Verordnung über klinische Prüfungen von Medizinprodukten (MPKPV), Medizinprodukte-Sicherheitsplanverordnung (MPSV) sowie die DIN EN ISO-14155, berücksichtigt.

## 2.4 Behördenmeldungen, Genehmigungen und Registrierungen

Die Genehmigung der zuständigen Bundesoberbehörde (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM) war nach Anwendung des § 47 Absatz 3 MPDG (vormals § 23b MPG) nicht notwendig.

Aufgrund der Einbindung von ausschließlich nationalen Prüfzentren war eine Meldung und Genehmigung von anderen zuständigen Behörden im Ausland nicht notwendig.

Die Studie wurde beim Deutschen Register Klinischer Studien (DRKS) retrospektiv registriert:

- DRKS-ID: DRKS00027925
- Registrierungsdatum: 13.04.2022

## 2.5 Aufklärung und Einwilligung der Studienteilnehmer:innen sowie der Prüfer

Die Einwilligung der potenziellen Studienteilnehmer:innen war zwingende Voraussetzung für den Einschluss in die hier dargelegte klinische Prüfung. Patient:innen der Wartelisten der Klinik *Verhaltenstherapie Falkenried MVZ GmbH* und andere Betroffene, die sich für eine Studienteilnahme interessierten, wurden von den sie betreuenden Prüfärzten / Psychotherapeuten bei der Visite oder telefonisch bzw. videotelefonisch in verständlicher Sprache über die Studie aufgeklärt. Die Aufklärung, die schriftlich und mündlich erfolgte, umfasste neben Erklärungen zum allgemeinen Studienablauf insbesondere eine Erläuterung möglicher unerwünschter Wirkungen der Intervention. Weiterhin wurden Aspekte des Datenschutzes und der pseudonymisierten Auswertung mit dem Studienteilnehmer:innen eingehend besprochen. Die Studienteilnehmer:innen sollten die Bedeutung und Tragweite der klinischen Prüfung mit allen für sie möglichen Vor- aber auch Nachteilen überblicken und erfassen können, um eine fundierte Entscheidung zu treffen. Die Studienteilnehmer:innen wurden darüber aufgeklärt, dass er oder sie jederzeit seine/ihre Einwilligung ohne Angabe von Gründen oder Entstehen von Nachteilen widerrufen kann.

Um der Einwilligungserklärung, Datenverarbeitung und Studienteilnahme auf digitalem Weg zuzustimmen, wurde als erstes ein Termin vereinbart, zu dem die Studieninformationen und Einwilligungserklärung als PDF bereitgestellt wurde. Der betreuende Prüfer klärte bei dem Termin telefonisch über die Studie auf, beantwortete Fragen zur ausgegebenen Studieninformationen und Einwilligungserklärung und räumte nach dem Gespräch Bedenkzeit ein. Lag eine Teilnahmebereitschaft vor, wurde der/die Studienteilnehmer:in gebeten zuallererst das PDF als persönliche Kopie der Studieninformationen und Einwilligungserklärung zu speichern. Dann schickte der Prüfer eine offizielle E-Mail, die den kompletten Text der Studieninformationen und Einwilligungserklärung noch einmal enthielt, an den/die potenzielle(n) Studienteilnehmer:in mit der Bitte, für eine Einwilligung ggf. direkt auf ebendiese E-Mail zu antworten, ohne den mitgeschickten Text zu entfernen. Außerdem musste für eine Zustimmung folgender unterstrichener Text als Antwort samt Datum und vollständigem Namen unter dem Text der Antwort-E-Mail zurückgeschickt werden:

*[Vollständiger Text der Studieninformationen und Einwilligungserklärung]*

*Ich willige in die Teilnahme an dem oben beschriebenen Forschungsvorhaben ein.*



Datenübermittlung für Projektverwaltung und Fragebögen über SSL (HTTPS) Zertifikate mit Organisations-Validierung (OV) zur Verfügung; Server-Standort ist Deutschland (München). Weitere Informationen zum Datenschutz, zur Verschlüsselung, zum Löschkonzept sowie zu technischen und organisatorischen Maßnahmen (TOM) finden sich auf der Webseite des Anbieters: <https://www.soscisurvey.de/help/doku.php/de:general:privacy>.

4. Die im Rahmen des Forschungsvorhabens erhobenen persönlichen und klinischen Daten unterliegen der Schweigepflicht und den datenschutzrechtlichen Bestimmungen. Sie werden in Papierform und auf Datenträgern von der *Verhaltenstherapie Falkenried MVZ GmbH* und der *Ipsos GmbH* aufgezeichnet und nur in pseudonymisierter Form (d.h. verschlüsselt) gespeichert. Die Daten wurden entsprechend der gesetzlichen Regelungen unter Verschluss für eine Dauer von 10 Jahren archiviert. Bei der Pseudonymisierung (Verschlüsselung) wurden Namen und andere Identifikationsmerkmale durch den Zugangscode zur *My7steps App*-Anwendung ersetzt, um die Identifizierung wesentlich zu erschweren. Zugang zu diesem „Schlüssel“, der eine persönliche Zuordnung der Daten ermöglicht, hatten neben dem Leiter der klinischen Prüfung nur von diesen ausdrücklich dazu autorisierten Personen der *Verhaltenstherapie Falkenried MVZ GmbH* und der *Ipsos GmbH*. Sobald der Forschungszweck es zuließ, wurde der Schlüssel gelöscht und die erhobenen Daten damit anonymisiert.
5. Die von den Prüfern digital ausgefüllten Case Report Forms als auch die Einwilligungserklärungen der Studienteilnehmer:innen wurden bis zum Abschluss der Studie in einer eigens für die Studie eingerichteten, DSGVO-konformen, datensicheren, internen Cloud gespeichert, zu der nur das in die Erhebungskoordination eingebunden Studienpersonal Zugang hatte.
6. Der wissenschaftliche Leiter erhielt zur statistischen Auswertung pseudonymisierte Datensätze. Der Gesamtdatensatz wurde nach der Zusammenführung des Excel-Datensatzes (pseudonymisierte Zusammenfassung der CRFs), der Onlinebefragung und der App-Nutzungsdaten vom wissenschaftlichen Leiter anonymisiert. Die Veröffentlichung der Studienergebnisse erfolgt ausschließlich in anonymisierter Form.

#### Verantwortlich für die Datenverarbeitung sind:

1. Klinische Daten: Prof. Dr. med. Helmut Peter (Verhaltenstherapie Falkenried MVZ GmbH)
2. Technische Daten der *My7steps App*: Dr. Ralph Grobecker (Ipsos Healthcare GmbH)
3. Datensicherung interne Studien-Cloud: Maryam Gardisi (Ipsos GmbH)
4. Finaler Datensatz der Studie: Dipl.-Psych. Pelle Bernhold (Ipsos GmbH).

#### Rechte der Studienteilnehmer:innen:

- Die Studienteilnehmer:innen haben das Recht, über die von ihnen erhobenen personenbezogenen Daten Auskunft zu verlangen (nebst Aushändigung einer kostenfreien Kopie), sie betreffende unrichtige personenbezogene Daten berichtigen zu lassen und die Löschung von personenbezogenen Daten zu verlangen, wenn diese für den Zweck, für den sie erhoben wurden, nicht mehr notwendig sind.

- Die Verarbeitung der personenbezogenen Daten ist nur mit der Einwilligung der Studienteilnehmer:innen rechtmäßig. Im Falle des Widerrufs der Einwilligungserklärung werden die bereits erhobenen Daten gelöscht. Eine Löschung bereits anonymisierter Daten ist nicht möglich. In Ausnahmefällen dürfen die bis zum Zeitpunkt des Widerrufs erhobenen Daten weiterverarbeitet werden, wenn dies zur Erfüllung einer rechtlichen Verpflichtung erforderlich ist.
- Zur Wahrnehmung der Datenschutzrechte können sich die Studienteilnehmer:innen an den Studienleiter wenden. Sie haben daneben das Recht, sich mit einer Beschwerde an die Datenschutz-Aufsichtsbehörde zu wenden, wenn sie die erfolgte Datenverarbeitung als rechtswidrig ansehen:

Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit  
Thomas Fuchs  
Adresse Ludwig-Erhard-Str 22, 7. OG, 20459 Hamburg  
Telefonnummer 040 / 42854 - 4040  
E-Mail: [mailbox@datenschutz.hamburg.de](mailto:mailbox@datenschutz.hamburg.de)

## 3 Organisationsstruktur

### 3.1 Sponsor

<b>Sponsor</b>	
Name des Sponsors	Ipsog Healthcare GmbH
Anschrift des Sponsors	Münsterplatz 13 78462 Konstanz
Vertreter des Sponsors	Dr. Ralph Grobecker

<b>Leiter der klinischen Prüfung (LKP)</b>	
Name des LKP	Prof. Dr. med. Helmut Peter Geschäftsführer
Institution des LKP	Verhaltenstherapie Falkenried MVZ GmbH
Anschrift der Institution	Falkenried 7 20251 Hamburg

<b>Wissenschaftlicher Leiter</b>	
Name des wiss. Leiters	Dipl.-Psych. Pelle Bernhold
Institution des wiss. Leiters	Ipsog GmbH
Anschrift der Institution	Münsterplatz 13 78462 Konstanz

## 3.2 Weitere beteiligte Einrichtungen und Komitees

### 3.2.1 Advisory Expert

Im Rahmen der Planung, Durchführung sowie Bewertung dieser klinischen Prüfung sind Experten der Verhaltenstherapie Falkenried MVZ GmbH involviert worden, die mit wissenschaftlicher Expertise und klinischer Erfahrung unterstützend tätig waren. Hier ist insbesondere zu nennen:

- Dipl.-Psych. Markus Volmer (Psychologischer Bereichsleiter, Verhaltenstherapie Falkenried MVZ GmbH)

Im Rahmen der Planung, Konzeption und Berichtlegung dieser klinischen Prüfung sind Experten der fbeta GmbH (Akazienstraße 31, 10823 Berlin) involviert worden, die mit wissenschaftlicher Expertise und klinischer Erfahrung unterstützend tätig waren.

### 3.2.2 Zentrale Organisationseinheiten

Die *Ipsog GmbH* als studienbetreuende Organisationseinheit erfüllte während der Studie die Aufgabe der Datenerhebung und des Datamanagements:

- Datenmanagement: Pelle Bernhold (Wissenschaftliche Leitung, Ipsog GmbH)
- Koordination Online-Datenerhebung: Maryam Gardisi (Geschäftsführung, Ipsog GmbH)

Die gesetzlichen Meldepflichten im Rahmen der Rolle des Sponsors sowie Verantwortlichem nach § 66 des Medizinprodukte-Durchführungsgesetz (MPDG) lagen bei der *Ipsog Healthcare GmbH*.

### 3.2.3 Prüfer und Prüfzentren

Die Durchführung der Studie wurde unter Einbezug der Therapiezentren und Tageskliniken der *Verhaltenstherapie Falkenried MVZ GmbH* in Hamburg mit Unterstützung der online arbeitenden Psychologinnen und Psychologen der *Ipsog GmbH* umgesetzt.

Eine Aufstellung aller Prüfer und Prüfzentren sowie des beteiligten Studienpersonals wurde in gesonderten Listen geführt und laufend aktualisiert.

## 4 Finanzierung

Die vorliegende Studie wurde durch die folgenden Mittel finanziert:

- Eigenmittel der *Verhaltenstherapie Falkenried MVZ GmbH*
- Eigenmittel und Spenden der *Ipsog GmbH*

## 5 Wissenschaftlicher Hintergrund

### 5.1 Wissenschaftlicher Hintergrund psychischer Störungen: Depressionen und Anpassungsstörungen

Depressionen gehören weltweit zu den häufigsten psychischen Erkrankungen, die mit körperlichen Beschwerden und einer Beeinträchtigung der Lebensqualität und der täglichen Funktionsfähigkeit einhergehen [2]. Eine aktuelle Studie in Deutschland ergab eine Gesamtprävalenz von 10 % depressiver Symptome in der Allgemeinbevölkerung [81]. Das *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-V) klassifiziert psychische Erkrankungen in drei verschiedene Stufen: leicht, mittelschwer und schwer. Die Klassifizierung basiert auf der Intensität der depressiven Symptome, der Anzahl der Symptome und dem Grad der funktionellen Beeinträchtigung [1]. Hoffnungslosigkeit, Hilflosigkeit, Schuldgefühle und Verzweiflung, mangelnde Motivation, Konzentrationsschwierigkeiten, Schlaflosigkeit, Probleme mit dem Selbstwertgefühl, verminderte Produktivität und Gefühle der Wertlosigkeit sind einige der häufigsten Symptome einer leichten und mittelschweren Depression.

Ein systematischer Review über die Behandlung von mittelschweren Depressionen zeigt, dass Psychotherapie oder eine Kombination aus Psycho- und Pharmakotherapie positive Effekte erzielen. Unter den psychotherapeutischen Methoden erwies sich die kognitive Verhaltenstherapie (KVT) als Einzelbehandlung oder integriert in andere Formen der Psychotherapie als wirksam. Achtsamkeitsintervention, psychodynamische Langzeittherapie sowie Hilfe zur Selbsthilfe sind weitere evidenzbasierte Interventionen bei mittelschweren Depressionen [1].

Frühere Wirksamkeitsstudien von Face-to-Face und Online Value Based Counseling (VBC) zeigten, dass diese Interventionen nach durchschnittlich drei bis vier Beratungssitzungen Symptome von Depression und Angst signifikant reduzieren können [76]. Darüber hinaus zeigte eine andere Studie, die VBC mit einer medizinischen Routinebehandlung verglich, eine signifikante Reduktion der psychologischen Symptome (Depression und Angst) und psychosozialen Stressoren der Teilnehmer:innen [5].

Die Anpassungsstörung ist eine Diagnosekategorie, die durch eine emotionale Reaktion auf ein kritisches Lebensereignis oder eine Reihe von belastenden Ereignissen hervorgerufen wird. Es handelt sich um eine subjektive und emotionale Notlage, die als Folge einer bedeutsamen Lebensveränderung ausgelöst wird, z. B. nach bedeutenden Lebensveränderungen, belastenden Lebensereignissen, schwerer körperlicher Krankheit oder der Ungewissheit einer schweren Erkrankung [15, 77]. Es wird angenommen, dass die Anpassungsstörung eine Zwischenkategorie zwischen keiner psychischen Störung und affektiven Störungen (Depression und Angststörungen) ist, da sie sich in der Mitte auf einem Kontinuum zwischen normalen Stressreaktionen und pathologischen Reaktionen befindet [24, 31].

Obwohl die Anpassungsstörung eine sehr häufige Diagnose in der klinischen Praxis ist, wurde wenig belastbare Forschung zur Klärung der Epidemiologie, der diagnostischen Relevanz, der genauen Diagnoseinstrumente, der Validität, der Reliabilität und der evidenzbasierten Behandlungen durchgeführt [15, 16, 77, 86, 101]. In der Konsiliarpsychiatrie wurde eine geschätzte Inzidenz von 5-21% festgestellt [15], während die geschätzte Prävalenz in primären Gesundheitszentren 2,94% betrug [31].

Es fehlt an randomisierten kontrollierten Studien (RCTs), die die effektivsten Psychotherapien zur Behandlung von Anpassungsstörungen untersuchen. Daher existiert keine Goldstandard-Empfehlung für die Behandlung der Anpassungsstörung, weder für die Pharmakologie noch für die Psychotherapie.

Nichtsdestotrotz besteht ein allgemeiner Konsens, dass kurze psychologische Interventionen und Psychotherapien der Pharmakotherapie überlegen sind [15, 16, 24, 75].

In Ermangelung strikter RCTs für Anpassungsstörungen schlagen Experten vor, dass Interventionen drei gemeinsame Komponenten therapeutischer Interventionen für Anpassungsstörungen berücksichtigen sollten, einschließlich der Reduktion oder des Ausschlusses des Stressors, der Bereitstellung sozialer Unterstützung für den Fall, dass der Stressor anhält, der Verbesserung der Bewältigungsfähigkeiten, der psychologischen Belastbarkeit, der Maximierung der Anpassung, der Symptomreduktion und der Verhaltensänderung [24].

Der Value Based Counseling-Ansatz kann als eine passende Alternative in der Behandlung von Anpassungsstörungen angesehen werden, da er psychosoziale Stressoren als direkte und indirekte Ursachen für psychische Probleme betrachtet, die sich in pathologischer Symptomatik ausdrücken. VBC zielt darauf ab, Beschwerden wie psychologische Symptome, intrapsychische oder interpersonelle Konflikte, traumatische Erfahrungen, störende soziale Umgebungen oder schwierige Lebensumbrüche, wie Migration oder Verlust der Existenzgrundlage, zu erfassen und ihre Bedeutung für die Betroffenen unter Berücksichtigung der individuellen Vulnerabilität in Bezug auf den ungelösten psychosozialen Stress im jeweils aktuellen soziokulturellen Lebenszusammenhang zu untersuchen. Das Ziel von VBC ist es, durch die Entwicklung von Bewältigungskompetenzen Betroffene zu befähigen, mit Stressoren umzugehen und belastende Situationen besser zu bewältigen [76].

## 5.2 Wissenschaftlicher Hintergrund der technologie- bzw. internet-gestützten Interventionen zur Behandlung von psychischen Störungen

Technologie- bzw. internet-gestützte Interventionen (*internet and mobile-based interventions (IMI)*) sind hauptsächlich selbstgesteuerte psychosoziale Interventionen in Form von webbasierten Programmen oder mobilen Apps zur Unterstützung von Personen, die gesundheitsbezogene Unterstützung suchen [9, 25]. Das Interventionsprogramm selbst versucht durch die Bereitstellung von fundiertem gesundheitsbezogenem Material und die Nutzung interaktiver webbasierter Komponenten eine positive Veränderung zu bewirken bzw. Wissen, Bewusstsein und Verständnis zu verbessern und zu erweitern [9].

In den letzten Jahren sind IMIs zu einer evidenzbasierten Alternative zu persönlichen psychotherapeutischen Behandlungen geworden [20, 25, 94]. Ihre Wirksamkeit bei der Behandlung von Depressionen [20, 66, 83, 88], Angstzuständen [25, 90, 94] und vielen anderen psychischen Erkrankungen konnte gezeigt werden [25, 30].

IMIs werden zumeist unter Verwendung von Inhalten etablierter Face-to-Face-Behandlungs- und Präventionsprogramme entwickelt, allerdings über das Internet oder mobile Anwendungen angewendet. Eine umfangreiche Meta-Analyse über den Vergleich von Face-to-Face- und web-basierten Interventionen ergab keine Unterschiede in der Wirksamkeit. Die durchschnittliche gewichtete Effektgröße für web-basierte psychologische Interventionen liegt bei 0,53 (mittlerer Effekt) was der durchschnittlichen Effektgröße traditioneller Face-to-Face-Therapie ähnlich ist [8].

IMIs können auf Basis von vier Hauptfaktoren kategorisiert werden; a) welche Technologie sie verwenden, b) wie viel menschliche Unterstützung sie beinhalten, c) auf welcher theoretischen Grundlage sie beruhen und d) welche Indikationen sie ansprechen [29].

- a) Technologien und technische Umsetzung: IMIs integrieren eine breite Palette von grundlegenden und fortgeschrittenen Technologien, wie web-basierter Selbsthilfe, interaktiver web-basierter Komponenten, wie Audio- und Videoclips sowie E-Mail-Erinnerungen, die zum Erfolg der therapeutischen Interventionen beitragen [8, 13, 37]. Eine weitere fortschrittliche Technologie in Online-Programmen ist das "Content Tailoring", welches den Nutzern ermöglicht, in Echtzeit eine Auswahl zu treffen, die je nach Präferenz oder Bedarf zu spezifischen Inhalten führt [95].
- b) Menschliche Unterstützung: Eine frühere meta-analytische Überprüfung von IMIs kategorisierte die bestehenden Interventionen in drei Gruppen:
  - (1) Eigenständige Interventionen, sogenannte *unguided interventions* (d.h. Selbsthilfe mit minimalem Therapeutenkontakt)
  - (2) Weitgehend ungeleitete Interventionen, sogenannte *mostly unguided interventions* (d. h. überwiegend Selbsthilfe mit fachlicher Anleitung durch einen Therapeuten bei Bedarf) und
  - (3) Angeleitete Interventionen, sogenannte *guided interventions* (d. h. Interventionen mit kontinuierlicher therapeutischer Unterstützung) [25].

Aktuelle systematische Reviews und Metaanalysen zu IMIs ergaben größere Effektstärken und eine höhere Adhärenz für guided IMIs im Vergleich zu den unguided IMIs (72 % vs. 26 %) [47, 99]. Obwohl angeleitete IMIs sowohl zu einer höheren Anzahl von abgeschlossenen Modulen als auch zu höheren Abschlussraten führten [25], berichteten einige neuere Meta-Analysen keinen signifikanten Unterschied zwischen guided und unguided IMIs in Bezug auf die Symptomschwere nach der Intervention und den Nachuntersuchungen [25, 44], während andere signifikant höhere Effektstärken bei Interventionen, die eine menschliche therapeutische Anleitung beinhalteten, berichten und diese auf eine verbesserte Behandlungs-Compliance zurückführen [73]. Es wird argumentiert, dass die wichtigste Funktion der menschlichen Unterstützung bei IMIs die Bereitstellung von Ermutigung und das Vermitteln von Struktur ist [82].

Unguided IMIs eignen sich am ehesten als erste Stufe des "watchful waiting" in einem "stepped care"-Ansatz und als "massive open online interventions", die der Öffentlichkeit evidenzbasierte psychologische Interventionen zur Verfügung stellen und die Reichweite, Skalierbarkeit und Bezahlbarkeit solcher Interventionen erhöhen [20]. Potential und Einsatzmöglichkeiten gezielter, menschlicher therapeutischer Anleitung sollten dafür noch ausführlicher untersucht werden [73, 82].

- c) Theoretische Grundlage: IMIs sind primär auf der Basis von face-to-face psychotherapeutischen Techniken entstanden. Ursprünglich setzten sie Prinzipien der kognitiven Verhaltenstherapie (KVT) ein, doch in den letzten Jahren wurden sie auch von anderen theoretischen Ansätzen beeinflusst, wie z.B. der emotional-fokussierten Therapie, psychodynamischen, meditations- und achtsamkeitsbasierten Methoden sowie der Akzeptanz- und Commitment-Therapie [3, 26, 42, 57, 64]. IMIs sind besonders vorteilhaft bei der Anwendung von Techniken, die auf Veränderungen von Gedanken und Verhaltensweisen abzielen, einschließlich KVT und interpersoneller Interventionen [49]. Sie zielen auf eine positive kognitive, verhaltensbezogene und emotionale Veränderung ab und verallgemeinern ihre Anwendung auf das tägliche Leben der Anwender unter Verwendung etablierter psychotherapeutischer Techniken [9, 28].

- d) Indikationen und Anwendungsbereiche: Die etabliertesten und am häufigsten untersuchten IMIs sind Programme für psychische Gesundheitsprobleme wie Angst und Depression. Eine frühere Meta-Analyse, die 22 RCTs untersuchte, die für Depressionen und Angststörungen eingesetzt wurden, zeigte eine standardisierte durchschnittliche Effektgröße von 0,38 [28]. Eine weitere Meta-Analyse, die auf 14 RCTs basiert, zeigte signifikante mittlere bis große Effektstärken für IMIs in der Behandlung von Kindern und Jugendlichen, für Interventionen, die auf Angst und Depression abzielen, sowie für transdiagnostische Interventionen im Vergleich zur Nichtbehandlung [30].

### 5.3 Rationale der klinischen Prüfung

Die hohe Popularität und Verbreitung von internetfähigen Devices weltweit bietet eine beträchtliche Chance für IMIs, zur psychischen Gesundheit beizutragen und ein breites Spektrum von Menschen zu erreichen, die eine psychische Gesundheitsversorgung benötigen, insbesondere Bewohner in ressourcenbeschränkten Umgebungen, Menschen auf Wartelisten für eine Psychotherapie oder schwer erreichbare Gruppen, wie Teenager [20]. Die Übertragbarkeit, Unmittelbarkeit und Zugänglichkeit von IMIs kann also Menschen in ländlichen und abgelegenen Gebieten zugutekommen oder den Zugang zu psychosozialer Versorgung in ressourcenarmen Gebieten verbessern. IMIs bieten weitere Vorteile, wie Erreichbarkeit, Benutzerfreundlichkeit und Anonymität, die Menschen aus niedrigeren sozioökonomischen Gruppen, Minderheiten oder Menschen, die Angst vor der mit der Inanspruchnahme psychosozialer Hilfe verbundenen Stigmatisierung haben, zugutekommen. Darüber hinaus kann die Bereitstellung von IMIs zur professionellen psychiatrischen Versorgung von Menschen mit milderen psychiatrischen Symptomen die Belastung traditioneller psychiatrischer Dienste reduzieren und die psychische Gesundheit der Bevölkerung verbessern [20].

Die von *Ipsos Healthcare GmbH* angebotene webbasierte Selbsthilfe-Intervention *My7steps App* hat als IMI das Potential die depressiven Symptomatiken insbesondere von Betroffenen auf Wartelisten unmittelbar nach der Diagnose zu reduzieren und damit die erste Stufe eines "stepped care"-Ansatzes darzustellen.

Vor diesem Hintergrund sollte in der nachfolgend beschriebenen Studie evaluiert werden, ob *My7steps App* der *Ipsos Healthcare GmbH* eine geeignete Anwendung ist, um depressive Symptomatiken, basierend auf der Depressivitätsskala des PHQ-9-Fragebogens, bei Betroffenen aus dem Spektrum der depressiven Erkrankungen zu reduzieren.

## 6 Ziele der klinischen Prüfung

Im Rahmen der Zulassung als Digitale Gesundheitsanwendung (DiGA) muss *My7steps App* das Verfahren zur Aufnahme in das DiGA-Verzeichnis in Erprobung durchlaufen. Hierfür sind die Ergebnisse einer systematischen Datenauswertung zur Nutzung der DiGA vorzulegen, die in der Anwendung gewonnen wurden. Aus der vorliegenden Pilotstudie sollten sich Anhaltspunkte für die weitere confirmatorische Erprobungsstudie ergeben; neben den Interventionseffekten sollten auch Hinweise für die Wirksamkeit der Intervention für einzelne Indikations-Subgruppen, Fallzahlen, Messinstrumente, Rekrutierungsmethoden und andere relevante Informationen zur weiteren Planung ermittelt werden.

### 6.1 Primäres Ziel

Primäres Ziel der Studie war es, zu überprüfen, ob die Versorgung von Betroffenen aus dem Spektrum der depressiven Erkrankungen durch die Webanwendung *My7steps App* der *Ipsos Healthcare GmbH* verbessert wird. Dies sollte sich in einer Reduktion der depressiven Symptomatik basierend auf dem PHQ-9 zeigen.

### 6.2 Sekundäre Ziele

Sekundäre Ziele bezogen sich auf die Stärkung der persönlichen Belastbarkeit und die Verringerung krankheitsbedingter Probleme im Alltag durch *My7steps App*. Die persönliche Belastbarkeit (Resilienz) wurde mit der Connor-Davidson Resilience Scale (CD-RISC-10) gemessen, die Beeinträchtigung der Alltagsfunktionalität durch die Erkrankung mit Hilfe der Daily Functionality Impairment Scale (DFI). Außerdem wurde die subjektive Symptomstärke in Modul 1 als auch Modul 7 der *My7steps App* Anwendung erfasst und die Differenz als sekundäre Zielgröße herangezogen.

Einer Empfehlung des BfArMs folgend, wurden auf Grundlage der Indikationen der Studienteilnehmer:innen des Weiteren drei präspezifizierte Subgruppen gebildet und auf etwaige Effektmodifikation überprüft: leichte/mittelschwere Depression (Indikationsgruppe 1) vs. rezidivierende leichte/mittelschwere Depression (Indikationsgruppe 2) vs. Anpassungsstörung (Indikationsgruppe 3).

## 7 Studienplan

### 7.1 Allgemeines Studiendesign

Zum Nachweis der Wirksamkeit der Anwendung von *My7steps App* wurde eine Ein-Gruppen-Prä-Post-Studie zum Zweck intra-individueller Vergleiche durchgeführt.

Studienzentrum war die Klinik

Verhaltenstherapie Falkenried MVZ GmbH  
Jarrestraße 8, D-22303 Hamburg  
Homepage: [www.vt-falkenried.de](http://www.vt-falkenried.de)

Die Koordination der Online-Rekrutierung, Diagnostik und Telefoninterviews übernahm die

Ipsog gemeinnützige Gesellschaft mbH  
Münsterplatz 13, 78462 Konstanz  
Homepage: [www.ipsogcontext.org](http://www.ipsogcontext.org)

Die Studienteilnehmer:innen erhielten die webbasierte Selbsthilfe-Intervention *My7steps App*, die den Nutzer selbstverwaltend durch eine Intervention aus Selbsteinschätzung, Psychoedukation, Übungen aus der kognitiven Verhaltenstherapie sowie Achtsamkeitsübungen führt. Die Erhebung der relevanten Zielgrößen zum Zeitpunkt T0 vor Beginn der Intervention und T1 nach Abschluss der Intervention, je nach der individuellen Programmabsolvierungsdauer, erfolgte durch die Prüfer.

### 7.2 Zeitplan

Die Studienteilnehmer:innen erhielten die webbasierte Selbsthilfe-Intervention *My7steps App*, mit der sie über einen Webbrowser, Tablet oder Smartphone selbstbestimmt sieben Module aus Selbsteinschätzung, Psychoedukation, Übungen aus der kognitiven Verhaltenstherapie sowie Achtsamkeitsübungen durchlaufen. Empfohlen wird eine Anwendungsdauer von drei Monaten für *My7steps App*, um ausreichend Zeit für die Integration persönlicher Verhaltensänderungen in das eigene Leben zu gewähren. Abhängig vom Verhalten der Studienteilnehmer:innen und der selbstgewählten Teilnahmehäufigkeit und -intensität kann diese Dauer aber stark variieren. Im Rahmen der Pilotstudie wurden die Studienteilnehmer:innen gebeten, sich für eine regelmäßige und intensivere Nutzung zu entscheiden. Dies sollte verhindern, dass die Nutzungsdauer in der kleinen Stichprobe übermäßig stark variiert und eventuell zur Störvariable wird. Die Pilotstudie zielte auf die Überprüfung der generellen Wirksamkeit ab; die Auswirkung einer längerer Nutzungsdauer soll in einer folgenden, größer angelegten, kontrollierten Studie überprüft werden.

#### **Prüfungsende**

Die Prüfung wurde abgeschlossen, als der/die letzte eingeschlossene Studienteilnehmer:in die Intervention durchlaufen hatte.

Tabelle 1: Zeitlicher Rahmen der Studie

Zeitplan	
Einschluss des ersten Patienten (FPFV)	11.11.2021
Einschluss letzter Patient (LPFV)	04.02.2022
Prüfungsende des letzten Patienten (LPLV)	06.04.2022

### 7.3 Diskussion des Studiendesigns

Die vorliegende Untersuchung findet im Rahmen der Zulassung als Digitale Gesundheitsanwendung im Verfahren zur Aufnahme in das DiGA-Verzeichnis in Erprobung statt. Hierfür sind die Ergebnisse einer systematischen Datenauswertung zur Nutzung der DiGA vorzulegen (Vorstudie/Pilotstudie), die in der Anwendung gewonnen wurden. Aus ihr sollen sich erste Anhaltspunkte ergeben, die die weitere konfirmatorische Erprobungsstudie vorbereiten und neben den Interventionseffekten auch Hinweise für die Fallzahlen, Messinstrumente, Rekrutierungsmethoden und andere relevante Informationen zur weiteren Planung erheben sollen.

Im Rahmen des Zulassungsprozesses als DiGA gibt es keine Vorgaben an das Studiendesign der Vorstudie. Ein intra-individueller Prä-Post-Vergleich eignet sich daher für das beschriebene Vorhaben, da die Veränderungen der depressiven Symptomatik der Studienteilnehmer:innen vor und nach der *My7steps App*-Intervention gemessen und wichtige Rückschlüsse auf die genannten Parameter der konfirmatorischen Erprobungsstudie gezogen werden können. Die Fallzahl konnte daher so klein wie möglich und so groß wie nötig gewählt werden, um valide Rückschlüsse auf erste Effekte sowie Rückschlüsse auf die genannten Parameter der konfirmatorischen Zulassungsstudie ziehen zu können und den organisatorischen Ressourcenaufwand möglichst klein zu halten.

Das vorliegende Design wurde im Rahmen eines Kick-Off-Meetings am 19.05.2021 mit dem *Bundesinstitut für Arzneimittel und Medizinprodukte* (BfArM) diskutiert und für grundsätzlich angemessen befunden.

### 7.4 Auswahl der Studienpopulation

Die Studienpopulation wurde aus der Gruppe der Betroffenen aus dem Spektrum der depressiven Erkrankungen rekrutiert. Diese Studienteilnehmer:innen konnten Patienten der Regelversorgung sein (z.B. aus Wartelisten) oder Betroffene, die bisher keine professionelle Hilfe in Anspruch genommen hatten und die Web-Intervention im Rahmen ihres Alltags als Möglichkeit der Selbsthilfe ausprobieren wollten. Im Rahmen der Studie sollte ein großes Spektrum der Versorgungsrealität abgebildet werden, und nur wenige Ausschlusskriterien sollten die Rekrutierung einschränken. Es wurde erwartet, dass gemäß der Inzidenz depressiver Erkrankungen ein signifikant höherer Anteil an Frauen rekrutiert wird, der repräsentativ für die Geschlechterverteilung der Erkrankung ist.

Folgende drei diagnostische Subgruppen wurden für die weitere Auswertung als relevant definiert:

1. leichte/mittelschwere Depression (F32.0, F32.1)
2. rezidivierende leichte/mittelschwere Depression (F33.0, F33.1)
3. Anpassungsstörung (F43.2)

In einem Interaktionstest sollte nach Abschluss der Studie eine etwaige Effektmodifikation der Behandlung in Abhängigkeit von der Indikationsgruppe überprüft werden; dieser Anforderung wurde im Rahmen einer einfaktoriellen ANOVA unter Berücksichtigung der Diagnosegruppe nachgekommen.

Ferner wurden weitere Einflussfaktoren für die spätere Betrachtung im Studienplan vermerkt. Diese sind:

- Bildung
- Alter.

Die im folgenden Kapitel beschriebenen Ein- und Ausschlusskriterien beziehen daher nur unmittelbar notwendige Kriterien zur Behandlung der Studienteilnehmer:innen mit der spezifischen Intervention ein oder basieren auf spezifischen gesetzlichen Vorgaben.

## 7.4.1 Ein- und Ausschlusskriterien

### Studienindikation

Die Studienpopulation wurde aus der Gruppe der Betroffenen aus dem Spektrum der depressiven Erkrankungen rekrutiert:

- F32.0 Leichte depressive Episode
- F32.1 Mittelgradige depressive Episode
- F33.0 Rezidivierende depressive Störung, gegenwärtig leichte Episode
- F33.1 Rezidivierende depressive Störung, gegenwärtig mittelgradige Episode
- F43.2 Anpassungsstörungen

Diagnosen, Suizidalität und Ausschlusskriterien wurden im Erstgespräch durch approbierte Psychotherapeuten studienbedingt gestellt und geprüft. Hierzu wurde die ICD-10-Symptom-Rating (ISR) Skala in der Version 2.0 auf Deutsch und Englisch eingesetzt. ICD-10-Symptom-Rating (ISR) Skala evaluiert Status und Schweregrad psychischer Störungen und erfasst das Ausmaß von Depression, Angst, Zwang, Somatisierung und Essstörung [11, 92]. Das Gespräch wurde ggf. durch eine Übersetzerin oder einen Übersetzer für Arabisch bzw. Dari/Farsi unterstützt.

### Einschlusskriterien

Die Kriterien, die für einen Einschluss von Studienteilnehmer:innen in die Studie erfüllt sein mussten, umfassten:

- Rechtskräftige Einwilligung des / der Betroffenen liegt vor
- Die Fähigkeit Deutsch, Englisch, Arabisch, oder Farsi/Dari zu sprechen und zu verstehen
- Alter  $\geq$  18 Jahre
- Betroffene, die in Deutschland leben
- Psychotherapeutische Diagnose einer der folgenden Indikationen:
  - F32.0 Leichte depressive Episode
  - F32.1 Mittelgradige depressive Episode
  - F33.0 Rezidivierende depressive Störung, gegenwärtig leichte Episode
  - F33.1 Rezidivierende depressive Störung, gegenwärtig mittelgradige Episode
  - F43.2 Anpassungsstörungen

### **Ausschlusskriterien**

- Betroffene, die die oben beschriebenen Einschlusskriterien nicht erfüllten, waren von der Studie auszuschließen.

Zusätzlich auszuschließen waren:

- Betroffene, deren Einverständniserklärung nicht vorliegt
- Betroffene, die an anderen interventionellen Studien teilnehmen
- Personen, die in einem Abhängigkeits- / Arbeitsverhältnis zum Sponsor oder Prüfer stehen
- Personen in Unterbringung in einer Anstalt aufgrund gerichtlicher oder behördlicher Anordnung.
- Betroffene mit Suizidalität, diagnostizierter schwerer Depression, Suchterkrankung, schizophrener Symptomatik, psychotischer Symptomatik, wahnhafter Störung, hirnorganischer Störung (z.B. Delir, Demenz)

### **7.4.2 Nachträglicher Ausschluss von Prüfungsteilnehmern**

Im Rahmen von Studien kann es erforderlich werden, Studienteilnehmer:innen nachträglich auszuschließen. Die Kriterien hierfür und das nachfolgende Verfahren sind für die vorliegende Untersuchung im Folgenden beschrieben:

#### **Ausschluss von der Intervention und Untersuchungen**

- Widerruf der Einwilligungserklärung seitens des/der Studienteilnehmer:in
- Abbruch der Intervention seitens des/der Studienteilnehmer:in durch beispielsweise Arztwechsel oder Verzug
- Nichtadhärenz von Studienteilnehmer:innen an die vorgesehene Intervention

Die „Nichtadhärenz des/der Studienteilnehmer:in an die vorgesehene Intervention“ wurde definiert als Nichtabschluss des vorletzten (sechsten) Moduls von *My7steps App*. Der alleinige Nichtabschluss des letzten (siebten) Moduls führte dagegen nicht zum Ausschluss.

#### **Ausschluss von der Dokumentation**

Studienteilnehmer:innen waren von der Dokumentation auszuschließen und damit im Rahmen der Intention-to-Treat- Analyse dementsprechend nicht zu berücksichtigen, wenn:

- Studienteilnehmer:innen die Zuordnung abbrechen, indem sie sich nicht für die Intervention *My7steps App* registrieren und ihre Teilnahme beenden.

### **Verfahren bei vorzeitigem Ende der Behandlung im Rahmen der klinischen Prüfung**

Schied eine Studienteilnehmer:in aus der Studie aus, waren keine Nachbeobachtungen im Rahmen der Studie zwingend notwendig, um die Sicherheit der Studienteilnehmer:innen zu gewährleisten. Im Rahmen der Abschlussdokumentation/-untersuchung war jedoch die Erhebung aller in der Studie relevanten Parameter vorgesehen, sofern der/die Studienteilnehmer:in zustimmt. Hierbei sollte zusätzlich der Grund für das Ausscheiden dokumentiert werden.

Die Schätzung der benötigten Fallzahl schloss eine erwartete Drop-out Rate von 20% ein. Eine gezielte Nachrekrutierung von Studienteilnehmer:innen war nicht vorgesehen.

## 7.4.3 Rekrutierungsmethoden

Die Rekrutierung erfolgt unter anderem über die Wartelisten der Klinik Verhaltenstherapie Falkenried MVZ GmbH, Jarrestraße 8, D-22303 Hamburg. Studienteilnehmer:innen, die den Einschlusskriterien entsprachen, wurden von den betreuenden Ärzten / Psychotherapeuten über die Möglichkeit, an der Studie teilzunehmen, informiert. Um die Rekrutierungsrate darüber hinaus zu erhöhen, wurde über die Teilnahmemöglichkeit auch mit Hilfe von mehrsprachigen Flyern und Beiträgen in Arztpraxen, Vereinen, Sozialen Netzwerken und Newslettern informiert. Es wurde eine Aufwandsentschädigung von 50€ am Ende der Teilnahme ausgezahlt. Es wurde auf Einheitlichkeit beim Aufnahmeverfahren in die Studie geachtet; alle rekrutierten Studienteilnehmer:innen wurden vom Studienteam im standardisierten Einschlussverfahren aufgenommen.

## 7.5 Behandlung

### 7.5.1 Angewendete Behandlung

#### **Interventionsgruppe *My7steps App***

Im Rahmen der Erhebung bekamen interessierte Betroffene nach Erhalt der Studieninformationen und Vorliegen der Einwilligungserklärung im Rahmen des Aufklärungsgesprächs einen Termin für ein diagnostisches Aufnahmegespräch (Videovisite) mit einer Psychotherapeutin oder einem Psychotherapeuten angeboten. Das diagnostische Gespräch fand als Videokonferenz über einen datenschutzkonformen Dienst des deutschen Anbieters BigBlueButton statt. Diagnosen, Suizidalität und Ausschlusskriterien wurden im Erstgespräch durch den Diagnostiker bzw. die Diagnostikerin gestellt und geprüft. Der/die Psychotherapeut:in füllte das CRF digital aus und vereinbarte einen weiteren Termin für die erste Studienvisite mit einem nach Sprache und Geschlecht zugeordneten Prüfer.

Die erste und zweite klinischen Studienvisite (T0 und T1) fanden in der Onlineerhebung telefonisch statt und wurden über einen Onlinefragebogen vom Prüfer dokumentiert. Der hierfür benötigte Dienst wurde von dem deutschen Onlinefragebogenanbieter SoSci Survey GmbH bereitgestellt (eine Vereinbarung über die Verarbeitung personenbezogener Daten im Auftrag gemäß Art. 28, 29 DSGVO lag vor). Am Ende der ersten Studienvisite erfolgte die Interventionsausgabe.

Psychotherapeutische Sitzungen und psychologische Beratungen in Form von Videokonferenzen und Telefongespräche gehören zur klinischen Routine. Das online stattfindende diagnostische Aufnahmegespräch sowie die beiden telefonisch stattfindenden Studienvisiten wurden in der vorliegenden Untersuchung studienbedingt durchgeführt, wobei auf eine Validierung der Messinstrumente für telefonische Erhebungen geachtet wurde.

Im Rahmen der ersten klinischen Studienvisite erhielten die Studienteilnehmer:innen Zugang zur Webanwendung *My7steps App* sowie eine kurze Einführung in die Anwendung durch die Prüfer. Zudem dokumentierten die Prüfer unbeeinflusst und ergebnisoffen den Ausgangszustand mit den in Tabelle 2 vorgegebenen standardisierten Fragebögen und Parametern.

Nach Abschluss der Erhebung der Studien- und behandlungsrelevanten Parameter erfolgte im Rahmen der Studienvisite die Besprechung des weiteren Behandlungsverlaufs. Die Studienteilnehmer:innen konnten nach Abschluss der ersten Studienvisite die Selbsthilfe-Intervention beginnen.

Im Studienplan waren zweiwöchige Checkup-Calls vorgesehen, die der Aufrechterhaltung der organisatorischen Kommunikation im Rahmen der Teilnahme dienten. Die Studienteilnehmer:innen konnten die Telefongespräche nutzen, um ggf. technische Schwierigkeiten zu berichten und organisatorische Fragen zum Studienablauf zu klären. Des Weiteren dienten die Checkup-Calls der Terminfindung für die Studienvisiten. Entsprechend dem Gesprächsleitfaden (vgl. Abschnitt 14.2 im Anhang) durften keine therapeutischen Inhalte oder anderen Themen behandelt werden. Diese Regelung wurde allen Studienteilnehmer:innen im Rahmen des Onboardings explizit erläutert. Auf die Checkup-Calls konnte ggf. verzichtet werden, falls Studienteilnehmer:innen per E-Mail oder SMS mitteilten, keine technischen Probleme oder Fragen zum Studienablauf zu haben. Eine Übersicht der stattgefundenen Kontakte zwischen Prüfern und Studienteilnehmer:innen im Rahmen der Checkup-Calls findet sich in Abschnitt **Fehler! Verweisquelle konnte nicht gefunden werden.**

Die nächsten planmäßige Studienvisite erfolgte zum Abschluss der Behandlung. Auch diese Studienvisite fand telefonisch statt.

## 7.5.2 Beschreibung der Prüfindervention

Handelsname: My7steps App

Freiname: My7steps App

Hersteller: Ipsos Healthcare GmbH

*My7steps App* ist eine webbasierte Selbsthilfe-Intervention, die aus sieben Modulen zur Psychoedukation, Verhaltensaktivierung, Problemlösung, kognitiven Verhaltenstherapie, Achtsamkeit und Ressourcenorientierung besteht. Weitere Module befassen sich mit der Symptomeinschätzung, der

Identifizierung psychosozialer Stressoren, Achtsamkeitsübungen, positiver Psychologie und reflektierendem Journaling.

Die Module bestehen aus Texten, Übungen und Praxisbeispielen und beinhalten interaktive Elemente, wie Audio- und Videoclips. Ein starker Fokus der Intervention liegt auf der Anwendung von Aufgaben und zugewiesenen Schritten, um neu erworbene Strategien und Techniken in den Alltag zu integrieren.

Jedes Modul enthält eine Aufforderung, über die Erfahrungen und den persönlichen Fortschritt bei der Verbesserung mit den vorgeschlagenen Lösungen zu reflektieren. Zu Beginn jedes Moduls wird der Kontext oder das Grundprinzip des therapeutischen Elements durch kurze Videoclips erklärt.

Die Module sind adaptiv, da der Inhalt auf die spezifischen Auswahlen des einzelnen Benutzers zugeschnitten ist, indem der Benutzer kontinuierlich aufgefordert wird, zwischen verschiedenen Antwortmöglichkeiten zu wählen. Darüber hinaus werden die Antworten vom Benutzer in Bezug auf den Zeitpunkt des Auftretens und die Dimensionen der Intensität gesetzt. Nachfolgende Inhalte werden dann auf die Antwort des Benutzers zugeschnitten.

Durch das responsive Webdesign kann der Benutzer das Programm im Internet, auf einem Tablet oder Mobiltelefon verfolgen. Die tägliche Nutzung wird während der Phase der geführten Problemlösung empfohlen. Die App ist so gestaltet, dass der Benutzer die Module durchlaufen kann, während er seinem normalen Alltag und seiner Arbeit nachgeht.

Während der Intervention bewertet der Benutzer regelmäßig seine Zuversicht, dass es ihm bald besser gehen wird, und den Fortschritt seines Status. Am Ende des Prozesses erhält der Nutzer eine Zusammenfassung der absolvierten Module und eine Übersicht über die Veränderung der Symptome.

Während des gesamten Prozesses wird der Benutzer durch das „User-Journey-Monitoring-System“ im Backend der Applikation überwacht. Die Eingaben des Anwenders können Aktionen eines zugeordneten, gleichsprachigen Counselors auslösen. Alle Counselor sind speziell für die Anwendung geschult und verfügen über ein Studium und Erfahrung im Bereich der Psychologie sowie eine Weiterbildung in Value Based Counseling, der Methode, auf der diese Anwendung basiert.

Ein VBC-geschulter Counselor wird durch Alarme benachrichtigt, die bestimmte Ereignisse erkennen, u.a.:

- Verwendung von Worten, die Schaden und Gewalt beabsichtigen
- Senkung der Vertrauensskala (Wertebereich 1-100) bei zwei aufeinanderfolgenden Assessments um 5 Punkte auf der dahinterliegenden, siebenstufigen Likert-Skala
- Bewertung der Vertrauensskala (Wertebereich 1-100) entspricht dem niedrigsten Wert (1)
- Keine Interaktion mit der Anwendung für mehr als 14 Tage

Erkennt das Nutzerreise Monitoring (*oder User Journey Monitoring*) ein relevantes Ereignis, generiert das System eine Warnung, die im Counselor Dashboard angezeigt wird. Erhält der Counselor eine Mitteilung zu einer Warnung erfordert dies eine Sicherheitsaktion von ihm. Im Counselor Dashboard hat der Counselor dann einen Hintergrundcheck zur Nutzung von *My7steps App* durch den Nutzer durchzuführen, um festzustellen, was er zu tun hat. Um eine solche Warnung aufzulösen, muss er ein Statement im Counselor Dashboard hinterlegen. Ist als Sicherheitsaktion eine Kontaktaufnahme gefordert, muss das Statement die Art, Dauer und das Ergebnis des Kontakts enthalten. Um einen Lockdown des Nutzers für *My7steps App* aufzulösen, ist ein Kontakt erforderlich. In diesem muss der Counselor prüfen, ob der

Nutzer tatsächlich kontraindiziert ist und Hilfe außerhalb von *My7steps App* in Anspruch nehmen sollte. Für den Fall, dass dies der Fall ist, informiert der Counselor den Nutzer darüber, dass und wo er Hilfe bekommen kann, und der Nutzer bleibt für die Intervention gesperrt. Andernfalls entsperrt der Counselor den Klienten.

Wenn das User-Monitoring ein relevantes Ereignis erkennt, das eine Aktion des Counselors erfordert, erhält der Counselor eine Benachrichtigung. Im Monitoring-Kontrollraum kann der Counselor einen Hintergrundcheck über den Benutzer, deren Eingaben und Nutzungsmetriken durchführen und entscheiden, wie er reagieren soll. Eine Entscheidung zur manuellen Entsperrung des Benutzers ist nur nach dem Kontakt mit dem Benutzer und abhängig vom Gespräch mit dem Benutzer bzw. ggf. von der schriftlichen Reaktion des Benutzers zu fällen. Um die Warnung aufzulösen, muss eine Begründung in der *My7steps App* hinterlegt werden. In dieser Begründung ist auf die Kontaktaufnahme und mindestens das Resultat des Nachrichtenaustauschs oder Gesprächs einzugehen.

Eine Übersicht der stattgefundenen Kontakte zwischen Prüfern und Studienteilnehmer:innen im Rahmen der Nachverfolgung von Warnungen findet sich in Abschnitt **Fehler! Verweisquelle konnte nicht gefunden werden.**

Die *My7steps App* Intervention verfolgt einen transdiagnostischen Ansatz. Das bedeutet, dass der Ansatz zur Behandlung verschiedener Störungen, wie leichter und mittelschwerer Depression sowie Anpassungsstörung, eingesetzt wird. *My7steps App* zielt nicht nur auf die Linderung spezifischer Symptome einer Depression oder Anpassungsstörung ab, sondern hat einen ganzheitlichen Ansatz für die psychische Gesundheit und den Menschen.

Die *My7steps App* ist in folgenden Sprachen erhältlich: Deutsch, Englisch, Französisch, Arabisch, Dari/Farsi und Türkisch. Im Rahmen der Studie wurde die Teilnahme auf die Sprachen Deutsch, Englisch, Arabisch, sowie Dari/Farsi beschränkt. Diese Auswahl ergab sich aus dem Vorliegen standardisierter Übersetzungen der eingesetzten Messinstrumente sowie den Kapazitäten und Sprachfähigkeiten des Studienteams.

### 7.5.3 Vorgesehene Verwendung der Prüfontervention

Nutzer von *My7steps App* müssen in der Lage sein, die Benutzeroberfläche eines Internetbrowsers zu bedienen und die Texte dort lesen zu können. Sie benötigen keine Schulung, um *My7steps App* nutzen zu können. Es sind lediglich Grundkenntnisse in der Nutzung von PC und Smartphone erforderlich.

Für die Bearbeitung eines Modulschritts, welche etwa bis zu einer Stunde dauert, benötigen Nutzer ununterbrochenen Zugang zum Internet. Zwischen einzelnen Schritten und auch Unterabschnitten wird empfohlen zu pausieren, um die Inhalte zu verarbeiten und zu reflektieren. Die *My7steps App* kann zu jeder Tages- und Nachtzeit über das Internet genutzt werden.

Empfohlen wird eine Anwendungsdauer von bis zu drei Monaten für *My7steps App*, um ausreichend Zeit für die Integration persönlicher Verhaltensänderungen in das eigene Leben zu gewähren. Abhängig vom Verhalten der Nutzer und der selbstgewählten Teilnahmehäufigkeit und -intensität kann diese Dauer aber stark variieren. Im Rahmen der Pilotstudie wurden die Studienteilnehmer:innen gebeten, sich für eine regelmäßige und intensivere Nutzung zu entscheiden. Dies sollte verhindern, dass die Nutzungsdauer in der kleinen Stichprobe übermäßig stark variiert und eventuell zur Störvariable wird.

#### 7.5.4 Änderungen der Prüfindervention während der klinischen Studie

Änderungen des Studienprodukts, die unmittelbar gegenüber dem Nutzer oder dem Behandlungsablauf wirksam waren, liegen nicht vor.

Im Studienzeitraum wurden Änderungen im Counselor Dashboard wie folgt vorgenommen:

- Das Design der Darstellung wurde von einer Einzelfallansicht zu einer Übersichtsansicht geändert. Dabei wurden die dargestellten Details des Falles nicht verändert. Da sich die Änderungen ausschließlich auf das Design des Counselor Dashboards auswirkten, waren diese für die Studienteilnehmer:innen weder sichtbar noch erfahrbar.
- Im Studienzeitraum wurden im Rahmen der regelmäßigen Qualitätssicherung einzelne im Programm eingesetzte Vokabeln und grammatikalische Strukturen von Sätzen für die Versionen Englisch, Arabisch und Dari/Farsi aktualisiert, falls entsprechende Verbesserungsvorschläge vorlagen und diese nach einer Prüfung als Angleichung an die deutschen Original-Textversion eingestuft wurden. Dabei wurde die Konsistenz zwischen der jeweiligen Sprachversionen und der deutschen Ausgangssprache verbessert und zum Beispiel durch Autokorrekturen der Textverarbeitungsprogramme verursachte Import-Fehler in der arabischen bzw. persischen Schrift behoben.

#### 7.5.5 Ausgabe und Rücknahme der Prüfindervention

Die Ausgabe der Prüfindervention erfolgte bei Studieneinschluss, indem der Prüfer dem/der Studienteilnehmer:in einen Freischaltcode der Webanwendung aushändigte. Die Codes wurden von der *Ipsos Healthcare GmbH* an die Prüfer vermittelt und dienten bis zur Anonymisierung der Daten als Studienteilnehmer:innen-Nummer. Die Rücknahme der Prüfindervention erfolgt in dem Sinne, dass der Freischaltcode nach Ende des Verschreibungszeitraums bzw. Studienzeitraums ablief oder der/die Studienteilnehmer:in alle sieben Behandlungsschritte durchlaufen hatte.

#### 7.5.6 Verblindung

Eine Verblindung war in dieser Untersuchung nicht vorgesehen.

#### 7.5.7 Vorangegangene Therapie und Begleittherapie

Im Rahmen der Studienrekrutierung und -durchführung waren keine Einschränkungen hinsichtlich vorangegangener Therapie oder Begleittherapien im Studienplan vorgesehen. Diese wurden im Rahmen der Studienvisiten erfasst.

### 7.5.8 Weiterbehandlung nach Ende der klinischen Prüfung

Eine spezifische Intervention im Sinne einer Washout- bzw. Ausschleichphase war nicht notwendig und nicht im Studienplan vorgesehen. Die Studienteilnehmer:innen blieben Patient:innen der Regelversorgung, weitere Maßnahmen konnten im Bedarfsfall durch den Prüfer oder Psychotherapeuten erfolgen.

## 7.6 Wirksamkeits- und Sicherheitsparameter

Die Wirksamkeits- und Sicherheitsparameter leiten sich direkt aus den diskutierten Haupt- und Nebenzielen der Studie ab. Sie dienen der Operationalisierung in messbare Parameter, um etwaige Effekte der Intervention statistisch nachweisen und stützen zu können.

Die im folgenden beschriebenen Parameter werden zum jeweiligen Zeitpunkt der Studienvisite durch den Prüfer erhoben. Es handelt sich hierbei um die Erhebung standardisierter und validierter Fragebögen.

Eine Übersicht aller erhobenen Parameter ist in Tabelle 2 im folgenden Kapitel dargestellt.

### 7.6.1 Messung der Wirksamkeits- und Sicherheitsparameter

#### Primäre Zielgröße

Die primäre Zielgröße der klinischen Studie stellt der Schweregrad der Depressivität dar. Dieser wurde mit dem weltweit eingesetzten und gut etablierten *Patient Health Questionnaire* (PHQ) erhoben. Der PHQ basiert auf einem Klassifikationssystem der Psychiatrie, dem *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Der PHQ-9 ist das für den Einsatz in der Primärversorgung validierte Depressionsmodul, für alle Altersgruppen ab 16 Jahren empfohlen und kann sowohl für Längsschnitt- als auch für Querschnittsstudien verwendet werden [60]. Darüber hinaus hat sich der PHQ-9 ebenso als Telefoninterview bewährt [79]. Das eingesetzte PHQ-9 Modul weist neun Items auf und ist geeignet die Schwere einer Depression zu erfassen und diesbezügliche Behandlungsverläufe zu überwachen [87]. Abgefragt werden im PHQ-9 die neun diagnostisch relevanten Symptome der Major Depression (nach DSM), wobei jeweils die Auftretenshäufigkeit in Tagen bezogen auf die letzten zwei Wochen auf einer vierstufigen Likert-Skala ermittelt wird (0 = *Überhaupt nicht* bis 3 = *Beinahe jeden Tag*). Des Weiteren wird mit einem Zusatz-Item die Funktionsbeeinträchtigung in wichtigen Lebensbereichen abgefragt (Diagnosekriterium B der Major Depression nach DSM), was eine kategoriale Auswertung ermöglicht. Bei der dimensionalen Auswertung stehen höhere Werte für eine stärkere depressive Stimmung; der Wertebereich erstreckt sich von 0 bis 27, wobei eine Summe von 10 oder höher in der Regel als Schwellenwert für moderate Depressivität und klinische Signifikanz fungiert [78]. Eine weitere Unterteilung ergeben die Grenzwerte von 5, 10, 15 und 20 für leichte, moderate, mittelschwere und schwere Grade der depressiven Symptomatik [54]. Eine klinisch signifikante Verbesserung wird für den PHQ-9 als 5-Punkte-Reduktion vom Baseline- zum Post-Messzeitpunkt definiert [54, 61, 69].

Der PHQ-9 ist ein valides und reliables Messinstrument mit hervorragenden psychometrischen Eigenschaften: Die interne Konsistenz von  $\alpha = 0.87$  bis  $0.89$  und Test-Retest-Reliabilität von  $r_{tt} = 0.81$  bis  $0.96$  sind sehr gut [35, 51, 53, 61]. Die Kriteriumsvalidität ist bei dimensionaler Auswertung und

Verwendung des 10-Punkte-Cut-off-Werts mit einer Sensitivität von 80 bis 90 % und einer Spezifität von 80 bis 88 % hoch [53, 54, 80].

### Sekundäre Zielgrößen

Die Stärkung der persönlichen Belastbarkeit (Resilienz) durch die *My7steps App* wurde mit der etablierten, aus 10 Items bestehenden *Connor-Davidson Resilience Scale* (CD-Risc-10) gemessen. Die Skala erfasst Eigenschaften wie Widerstandsfähigkeit (z. B. Engagement, Veränderung als Herausforderung ansehen, Kontrolle), Selbstwirksamkeit, Ziel- und Handlungsorientierung, Toleranz gegenüber negativen Affekten, Geduld, Sinn für Humor angesichts von Stress und die Tendenz, sich von Stress zu erholen [18]. Der Wertebereich der CD-Risc-10 Skala erstreckt sich von 0 bis 40, wobei höhere Werte für eine höhere Resilienz stehen. Die Items werden auf einer 5-Punkte-Likert-Skala beantwortet (0 = *Überhaupt nicht wahr* bis 4 = *Fast immer wahr*). Für die 10-Item-Version findet sich eine jüngere Studie an Veteranen mit PTBS und Depressionen aus den USA, die für eine klinisch-depressive Stichprobe ein Skalenmittelwert von  $M = 23.6$  angibt [96], während die Mittelwerte bei Untersuchungen an der Allgemeinbevölkerung ohne klinische Vorgaben oftmals einen Summenwert von 30 erreichen oder überschreiten können, wie unter anderem berichtet für Spanien ( $M = 29.0$ ; Volljährige, 2015) [4], Dänemark ( $M = 30.3$ ; Krankenhauspersonal, 2017) [56], Deutschland ( $M = 31.3$ ; männliche Soldaten, 2015) [85] und die USA ( $M = 32.1$ ; Stichprobe aus nationaler Zufallsziehung, 2009) [14], sodass ein Vergleich zu klinischen und nicht-klinischen Stichproben möglich ist. Untersuchungen der Skala konnten eine gute Zuverlässigkeit, Konvergenz, Divergenz und Kriteriumsvalidität des CD-RISC sowie eine Empfindlichkeit gegenüber Veränderungen als Reaktion auf die Behandlung feststellen [23]. Die Test-Retest-Reliabilität sowie die konvergente Validität und Kriteriumsvalidität waren in einer deutschen Stichprobe zufriedenstellend [84], die gute Reliabilität konnte auch für die deutsche Übersetzung bestätigt werden [97].

Die Verringerung krankheitsbedingter Probleme im Alltag durch die *My7steps App* wurde mit der für den kulturübergreifenden Einsatz neu entwickelten *Daily Functionality Impairment Scale* (DFI) erfasst [45]. Die acht Items erfassen die Beeinträchtigung der Alltagsfunktionalität auf verschiedenen Ebenen und werden auf einer 5-Punkte-Likert-Skala bewertet (0 = *gar nicht beeinträchtigt* bis 4 = *stark beeinträchtigt*). Zu den abgefragten Ebenen gehören die krankheitsbedingte Beeinträchtigung von Beziehungen, der Haushaltsführung, von Freizeit- und sozialen Aktivitäten, der Arbeit, Ausbildung und der allgemeinen Lebenszufriedenheit. Der Wertebereich erstreckt sich von 0 bis 32, wobei höhere Werte der Skala auf eine stärkere Beeinträchtigung des täglichen Lebens durch psychische Probleme innerhalb der letzten 4 Wochen hinweisen. Aufgrund der Neuheit der Skalenentwicklung liegen bisher nur wenige Informationen zur Güte vor; die Entwickler geben ein Cronbachs  $\alpha$  von 0,79 an, was auf eine akzeptable bis gute Reliabilität des Messinstruments hindeutet [45].

Veränderungen subjektiver Symptomstärken von Modul 1 zu Modul 7: Im Rahmen des Modul 1 der *My7steps App* werden die Studienteilnehmer:innen gebeten, aus einer Anzahl von Symptomen diejenigen auszuwählen, die auf sie zutreffen. Ausgewählt werden können hierbei unter anderem *Traurigkeit, Sorgen, Gefühle der Wertlosigkeit, Ermüdung, Stress, Angst, Reizbarkeit, mangelndes Interesse oder Vergnügen, Konzentrationsschwierigkeiten, Anspannung, Essstörungen oder Schlafstörungen*. Für jedes ausgewählte Symptom wird die Stärke der Empfindung angegeben (1 = *wenig* bis 10 = *stark*). Die Likert-Skala wird hierbei als semantisches Differential bereitgestellt und mit einem Slider visualisiert, der einen Wertebereich von 1 bis 100 aufweist, sodass sich auch die Symptomstärken in diesem Bereich bewegen. Nach Auswahl aller zutreffenden Symptome werden die Studienteilnehmer:innen gebeten, die zwei

stärksten Symptome zu bestimmen (es muss mindestens ein Symptom gewählt werden), welche auch zum Abschluss der Behandlung in Modul 7 noch einmal von den Studienteilnehmer:innen hinsichtlich der Symptomstärke bewertet werden. Die Veränderungen der subjektiven Symptomstärken wurden auch für die vorliegende Studie je Studienteilnehmer:in von Modul 1 zu 7 erfasst, die Werte ggf. über beide Symptome gemittelt und diese gemittelte Differenz als sekundäre Zielgröße herangezogen. Eine negative Differenz steht hierbei für eine Verbesserung der ausgewählten Symptome, eine positive Differenz für eine Verschlechterung.

## Sicherheitsparameter

Als studien- und interventionsbezogene Sicherheitsparameter wurden das Vorkommen unerwünschter Ereignisse und unerwünschter Wirkungen sowie die laufend fortgeführte Risikoevaluation im Rahmen des User-Journey-Monitorings der Applikation als auch der händischen Sicherheitsüberprüfungen herangezogen und analysiert. Ergänzend wurde die subjektive Bewertung der Intervention durch die Studienteilnehmer:innen zum Studienabschluss ausgewertet. Im Folgenden sind die Sicherheitsparameter im Einzelnen beschrieben:

1. Zur Abschätzung unerwünschter Ereignisse und unerwünschter Wirkungen wurden Vorkommnisse für die Studie nach Artikel 2 Nr. 64 (Vorkommnis) und Nr. 65 (schwerwiegendes Vorkommnis) RL 2017/745 (MDR) sowie nach § 2 MPAMIV (mutmaßliches schwerwiegendes Vorkommnis) als Funktionsstörung, Mangel an Gebrauchstauglichkeit mit konsekutiver Fehlanwendung, Ausfall, Änderung von Merkmalen/Leistungen oder unsachgemäße Kennzeichnung oder Gebrauchsanweisung von *My7steps App*, die mittel- oder unmittelbar zum Tod oder schwerwiegender Verschlechterung der Gesundheit eines Patienten, Anwenders oder einer anderen Person geführt hat, geführt haben könnte oder führen können, definiert. Vorkommnisse waren zu dokumentieren und innerhalb der gesetzlichen Fristen meldepflichtig. Als unerwünschte Ereignisse bzw. Adverse Events (AE) waren prinzipiell alle medizinischen Ereignisse definiert, die sich zum Nachteil des/der Studienteilnehmer:in, der/die *My7steps App* nutzt, ausgewirkt haben, unabhängig davon, ob sie auf die Wirkung/Nebenwirkung der Prüfintervention zunächst offensichtlich ursächlich zurückzuführen waren. Diese Wirkung/Nebenwirkung der Prüfintervention unterlag ebenfalls der Dokumentationspflicht und wurde im CRF vom Zeitpunkt des Beginns der Prüfintervention bis vier Wochen nach der letzten Studienvisite vermerkt. Da keine primären Nebenwirkungen außer des Ausbleibens einer Verbesserung durch das Medizinprodukt bekannt waren und somit erwartet werden konnten, war jede Nebenwirkung zunächst als unerwartet einzustufen.
2. Im Rahmen der technischen Überwachung der Webanwendung und des User-Journey-Monitoring-Systems sowie der händischen, dreitägigen Sicherheitsüberprüfung durch die Counselor wurden Häufigkeiten, Inhalte und Risikoklassen sowohl von individuellen Rückmeldungen als auch der vom System generierten Warnungen laufend analysiert. Die Bewertungsparameter ergaben sich hierbei jeweils aus dem qualitativen Schweregrad des Risikos und der quantitativen Eintrittswahrscheinlichkeit des Risikos. Festgestellte Vorkommnisse und Art sowie Häufigkeit von Alarmen/Warnungen konnten zu einer Überprüfung und Aktualisierung der Bewertung des Restrisikos führen, die im Falle einer geringfügigen Änderung auch zu einer Aktualisierung der Studieninformationen sowie der Einwilligungserklärung und ggf. zur

Aktivierung des vorbereiteten Recall-Verfahrens und zu entsprechenden Meldungen im Rahmen der gesetzlichen Meldepflichten geführt hätte.

3. Der Anteil von Studienteilnehmer:innen mit reliabler Verschlechterung der depressiven Symptomatik wurde für die gesamte Stichprobe ausgewertet. Hierbei kamen sowohl das Kriterium der *minimal clinically important difference* (RCI > 1.96) als auch das Response-Kriterium des PHQ-9 zum Einsatz.
4. Als ergänzender Sicherheitsparameter wurden die Ergebnisse der abschließenden Befragung der Studienteilnehmer:innen zur subjektiven Bewertung der *My7steps App* deskriptiv ausgewertet. Die Studienteilnehmer:innen sollten nach Beendigung der Intervention die App generell bewerten, einschätzen, ob die Anwendung hilfreich war, ob sie ggf. bei sich festgestellte Veränderungen auf die App zurückführen und ob sie die Behandlung weiterempfehlen würden. Hierbei waren auch stark negative Bewertungen möglich, sodass entsprechende Sicherheitsindikatoren abgeleitet werden konnten.

## Erfassung der Zielgrößen

Das diagnostische Aufnahmegespräch zu Beginn der Studie erfolgte als Videovisite, um den verantwortlichen Psychotherapeuten einen möglichst umfassenden Eindruck von den Studienteilnehmer:innen zu ermöglichen. Die Messung der Studienendpunkte wurde von geschultem Personal ausgeführt und erfolgte in standardisierten Telefoninterviews vor und nach der Intervention. Der Messzeitpunkt T1 ergab sich aus der selbstgewählten Behandlungsdauer: die zweite Studienvisite erfolgte jeweils nach Behandlungsabschluss. Diese Regelung für die Pilotstudie diente der direkten Effektmessung, der Reduktion des zeitlichen Aufwands für die Studienteilnehmer:innen und der offenen Exploration individueller Behandlungsgeschwindigkeiten, die bei subjektivem Bedarf auch drei Monate überschreiten konnten. Alle Prüfer wurden für die Untersuchung geschult und folgten einem einheitlichen Interviewleitfaden. Die Sicherheitsparameter bzgl. unerwünschter Ereignisse und unerwünschter Wirkungen sowie der Risikoevaluation wurden laufend erhoben; der ergänzende Sicherheitsparameter der Bewertung der Behandlung durch die Studienteilnehmer:innen wurde zum Studienabschluss (T1) durch den jeweiligen Prüfer dokumentiert. Nutzungsdaten aus der Webanwendung stammen aus dem Anwendungsprotokoll der *My7steps App*. Weitere Details zum Ablauf der Erhebung finden sich in Abbildung 1, Details zu allen erhobenen Parametern finden sich in Tabelle 2.

Tabelle 2: Zusammenfassung der erhobenen Parameter

	Gruppe	Parameter	Frequenz	Erheber
	Diagnostisches Aufnahmegespräch (Videovisite)	Behandlungssprache	Einmalig vor der ersten Studienvisite	Psychotherapeut
		Ein- und Ausschlusskriterien		
		Medizinische Vorgeschichte und		

	Gruppe	Parameter	Frequenz	Erheber
		aktuelle Versorgungssituation		
		ICD-10 Diagnose(n)		
	Demographische Daten	Alter	Einmalig zur ersten Studienvisite T0	Prüfer
		Geschlecht		
		Bildungsniveau		
		Beruf		
	Fragebögen	Depression (PHQ-9)	Zur Studienvisite zu den Zeitpunkten T0 und T1	Prüfer
		Resilienz (CD-RISC-10)		
		Beeinträchtigung der Alltagsfunktionalität (DFI)		
	Feedback zur Webanwendung	Subjektive Nutzungsdauer und Nutzungshäufigkeit Bewertung der My7steps App	Einmalig zur zweiten Studienvisite T1	Prüfer
	Organisatorische Dokumentation	Prüfer-Studienteilnehmer:in-Kontakte insgesamt	Fortlaufend	Prüfer
		Zahl der Prüfer-Studienteilnehmer:in-Kontakte im Rahmen auftretender Symptome/ Beschwerden		
		Krankenhausaufenthalte		
		Außerplanmäßige Vorstellungen beim Arzt		
		Notfällige Vorstellungen beim Arzt		

	Gruppe	Parameter	Frequenz	Erheber
		Rekrutierungsquelle	Einmalig zu T1	
	Abschlussdokumentation bei Ausscheiden	Dokumentation des Grunds für Studienabbruch/-ausscheiden	Fortlaufend	Prüfer
	Nutzungsdaten der Webanwendung <i>My7steps App</i>	Modulabschlüsse Nutzungsdauer Nutzungshäufigkeit Zwei Symptomstärken in Modul 1 und Modul 7	Fortlaufend bis T1 Einmalig zu T1 Einmalig zu T1 Einmalig zu T1	Anwendungsprotokoll der <i>Ipsog Healthcare GmbH</i>

### Übersicht der Prozedur und Datenerhebung

1. Onboarding (telefonisch sowie per E-Mail): Patienteninformation und Einwilligungserklärung
2. Diagnostisches Aufnahmegespräch (Videovisite): Überprüfung der Ein- und Ausschlusskriterien
3. Studienvisite T0 (Telefoninterview): Baseline-Messung und Ausgabe der Intervention
4. Intervention (selbstadministriert)
5. Studienvisite T1 (Telefoninterview): Post-Messung und Rücknahme der Intervention

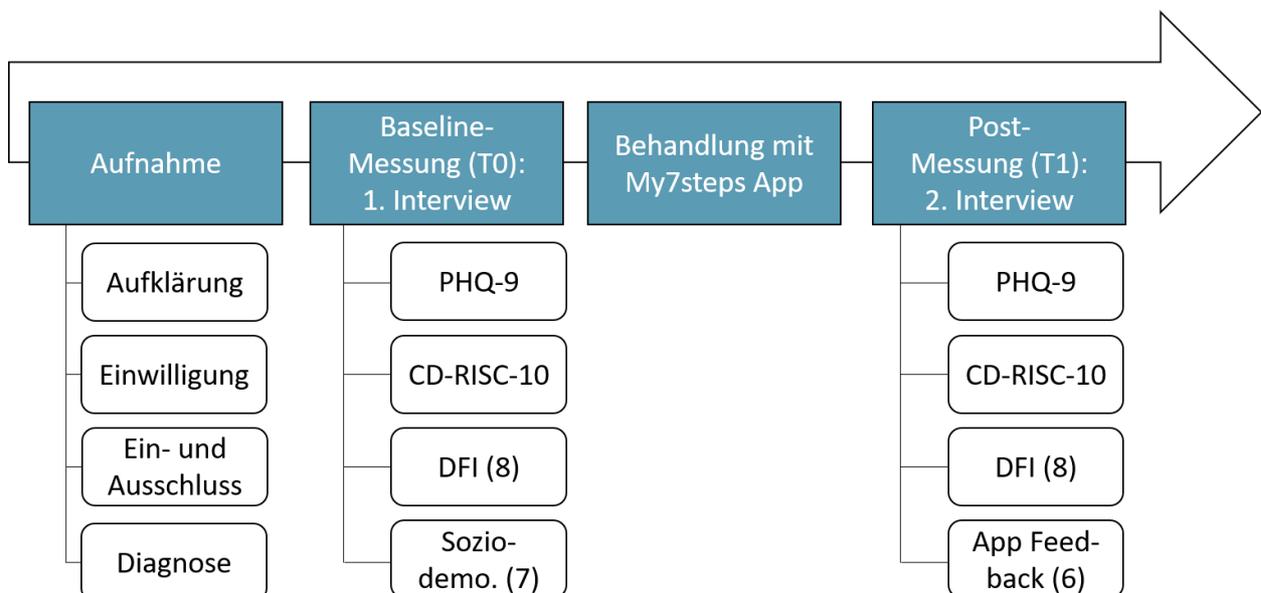


Abbildung 1: Übersicht der Prozedur und Datenerhebung.

## 7.6.2 Angemessenheit des Messverfahrens

Die gewählten und zuvor dargestellten Messverfahren und -instrumente, die in der Studie zur Generierung der Ergebnisse benötigt wurden, stellen sich in der Gesamtschau als zweckmäßig, verhältnismäßig sowie angemessen dar. Sie repräsentieren keine nennenswerte zusätzliche Belastung der Studienteilnehmer:innen, da keine invasiven Maßnahmen eingesetzt wurden und die Erhebungen in Form von validierten und in Klinik und Forschung etablierten Fragebögen erfolgten. Der Aufwand der drei im Rahmen der Studie online bzw. telefonisch stattfindenden Konsultationen (diagnostisches Gespräch, erste und zweite Studienvisite) war vergleichsweise gering: Die Gespräche haben je nach Fall 20 bis 45 Minuten gedauert, es war keine Anreise nötig und es gab keine Wartezeiten.

## 7.7 Sicherstellung der Datenqualität

Alle an der vorliegenden Studie beteiligten Prüfer verfügen über eine medizinische oder psychologische Ausbildung sowie mehrjährige Erfahrungen im Umgang mit Personen, die von depressiven Symptomatiken oder Anpassungsstörungen betroffen sind. Das gesamte Erhebungsteam wurde für die Studiendurchführung und für den Einsatz der standardisierten Messinstrumente trainiert, die Teilnahme an den diesbezüglichen Trainings war verpflichtend und wurde dokumentiert. Der Studienverlauf wurde in wöchentlichen Sitzungen gemonitort und die Einhaltung des Studienprotokolls durch das Leitungsteam und die Erhebungskoordinatoren fortlaufend überwacht.

### 7.7.1 Monitoring

Im Rahmen der Pilotstudie war keine Implementierung eines Monitorings vorgesehen.

### 7.7.2 Audits / Inspektionen

Gemäß den gesetzlichen Bestimmungen konnten jederzeit Inspektionen der Prüfzentren erfolgen. Feste Audits seitens des Sponsors waren nicht vorgesehen.

## 7.8 Dokumentation

### 7.8.1 Datenmanagement

#### **Allgemeine Dokumentation**

Alle Daten, die im Rahmen der Studie erfasst und ausgewertet wurden, wurden entweder auf Dokumentationsbögen in Form des Case Report Forms (CRF) oder den standardisierten und digital verarbeiteten Auswertungsbögen der eingesetzten Messinstrumente vom jeweiligen Prüfer dokumentiert. Das CRF stand allen Prüfern im Rahmen der Basis-Dokumentation zur Verfügung, alle Messinstrumente wurden digital bereitgestellt. Die Dokumentation organisatorischer Daten erfolgte ebenfalls durch die Prüfer: Dies betraf die Prüfer-Studienteilnehmer:innen-Kontakte insgesamt, ggf. Zahl

der Prüfer-Studienteilnehmer:innen-Kontakte im Rahmen auftretender Symptome/ Beschwerden, Krankenhausaufenthalte, außerplanmäßige Vorstellungen, notfällige Vorstellungen sowie die Dokumentation des Grunds für Studienabbruch/-ausscheiden.

Die Daten wurden im Rahmen der klinischen Datensicherung separat aufbewahrt. Die zuvor von der *Ipsog Healthcare GmbH* gelieferten App-Zugangscodes dienten als Studienteilnehmer:innen-Nummern; die Patientenidentifikationsliste mit den Studienteilnehmer:innen-Nummern wurde getrennt von den Studiendaten gelagert, der Zugang zu diesem Dokument war streng auf Prüfer mit koordinierender Funktion beschränkt. Die von den Prüfern digital ausgefüllten Case Report Forms als auch die Einwilligungserklärungen der Studienteilnehmer:innen wurden in einer eigens für die Studie eingerichteten, DSGVO-konformen, datensicheren, internen Cloud gespeichert, zu der nur das in die Erhebungsorganisation eingebundene Studienpersonal Zugang hatte. Auch bei den jeweils freigegebenen Dokumenten wurden Zugriffe, Einträge, Veränderungen und Löschungen ggf. verzeichnet und sind somit nachvollziehbar. Zum Studienabschluss wurden auch die in dieser internen Cloud gespeicherten Dokumente lokal archiviert und aus der Cloud entfernt.

### **Anonymisierung**

Das aus der Gruppe der involvierten Psychotherapeuten und psychologischen Counselor rekrutierte Studienteam überführte die Daten (inklusive der demographischen Variablen aus der Basis-Dokumentation) in einen Excel-Datensatz, der nur die Studienteilnehmer:innen-Nummern (App-Zugangscodes) enthielt; die Daten lagen hier pseudonymisiert vor. Die im Rahmen der Studienvisiten durchgeführten Befragungen mit standardisierten Messinstrumenten (PHQ-9, CD-RISC, DFI) wurden in dem Onlinefragebogen ebenso bereits pseudonymisiert gespeichert (Fallzuordnung über App-Zugangscodes). Die *Ipsog Healthcare GmbH* lieferte am Ende der Erhebung die technischen Nutzungsdaten (Nutzungsdauer und -häufigkeit, Modulabschlüsse, Symptomstärken Modul 1 und Modul 7) der Webanwendung *My7steps App* für die Studienteilnehmer:innen. Diese Daten wurden gleichfalls in pseudonymisierter Form bereitgestellt. Nach Beendigung der Intervention wurden die Daten dieser drei Quellen (CRFs, Onlinefragebögen, App-Nutzungsdaten) durch den wissenschaftlichen Leiter in pseudonymisierter Form zusammengeführt und der Gesamtdatensatz dann anonymisiert, d.h. die Löschung aller Studienteilnehmer:innen-Nummern (App-Zugangscodes) wurde im Datensatz vorgenommen.

### **Archivierung**

Alle wichtigen Unterlagen, die mit der Studie direkt oder indirekt in Verbindung stehen, insbesondere Einwilligungserklärungen, Dokumentationsbögen der Psychotherapeuten und Prüfer, aber auch der Studienteilnehmer:innen sowie relevante Dokumente der klinischen Prüfung, werden gemäß den Regeln der ICH-GCP mindestens für 10 Jahre aufbewahrt. Die Dokumentationen wurden zur Wahrung der Identität der Studienteilnehmer:innen getrennt von den Patientenidentifikationslisten aufbewahrt. Die sonstigen Regelungen zur Aufbewahrung medizinischer Unterlagen bleiben davon unberührt. Dementsprechend verbleiben die Originalbögen unter Verschluss und werden für 10 Jahre archiviert; Zugang hatte nur das im Studienprotokoll genannte Studienpersonal, das in die Erhebungsorganisation eingebunden war.

## 7.9 Statistische Methoden

Da es sich bei der vorliegenden Studie um eine Vorstudie handelt, spezifizierte der statistische Analyseplan nur die Grundlagen des Vorgehens zur Wirksamkeitsevaluation. Im folgenden Kapitel werden diese aufgeführt, weitere Maßnahmen zur Überprüfung und Differenzierung der Evaluation ergänzt und begründet sowie ggf. als Ergänzungen zum Studienplan ausgewiesen.

### 7.9.1 Hypothesen und Endpunkte

**Studienhypothese bzgl. des primären Endpunkts:** Die Intervention *My7steps App* führt bei Studienteilnehmer:innen aus dem Spektrum der depressiven Erkrankungen (ICD-10-Indikationen: F32.0 Leichte depressive Episode, F32.1 Mittelgradige depressive Episode, F33.0 Rezidivierende depressive Störung, gegenwärtig leichte Episode, F33.1 Rezidivierende depressive Störung, gegenwärtig mittelgradige Episode, F43.2 Anpassungsstörungen) zu einer Reduktion der depressiven Symptomatik, operationalisiert durch die Differenz der PHQ-9-Werte vom Baseline- zum Post-Messzeitpunkt.

**Studienhypothesen bzgl. der sekundären Endpunkte:** Die sekundären Fragestellungen werden in gleicher Weise wie der primäre Endpunkt überprüft. Diese Verfahren analysieren die persönliche Belastbarkeit (Resilienz), gemessen mit der Connor-Davidson Resilience Scale (CD-RISC-10), sowie die Beeinträchtigung der Alltagsfunktionalität durch die Erkrankung, die mit Hilfe der Daily Functionality Impairment Scale (DFI) untersucht wurde. Außerdem wird die gemittelte Differenz zweier subjektiver Symptomstärken, gemessen in Modul 1 als auch Modul 7 der *My7steps App* Anwendung, als weitere sekundäre Zielgröße herangezogen.

Es wurden keine vom Studienprotokoll abweichenden Hypothesen oder Endpunkte analysiert.

### 7.9.2 Subgruppen und Einflussfaktoren (Kovariaten)

#### Subgruppen

Folgende drei diagnostische Subgruppen wurden im Studienplan für die weitere Auswertung als relevant definiert, und eine etwaige Effektmodifikation in Anhängigkeit von der Indikationsgruppe wurde im Rahmen von Interaktionstests überprüft:

1. leichte/mittelschwere Depression (F32.0, F32.1)
2. rezidivierende leichte/mittelschwere Depression (F33.0, F33.1)
3. Anpassungsstörung (F43.2)

Ergänzend zur vorab spezifizierten Subgruppenanalyse der ICD-10-Diagnosegruppen wurde die Wirksamkeit für die Gruppe der subklinisch bis leicht depressiven Studienteilnehmer:innen auf Basis der PHQ-9-Ergebnisse zur Baseline untersucht. Diese Analyse wurde nachträglich beschlossen, um aktualisierten Anforderungen des DiGA-Zulassungsverfahrens zu entsprechen.

#### Einflussfaktoren

Ferner wurden weitere Einflussfaktoren für die spätere Betrachtung im Studienplan vermerkt. Diese sind:

- Bildung
- Alter.

Im Sinne einer kritischen Überprüfung der Wirksamkeitsergebnisse wurden zusätzlich zu den im Studienprotokoll explizierten potenziellen Einflussfaktoren folgende Variablen in den Analysen berücksichtigt:

- Geschlecht
- Zusatzbehandlung (= psychotherapeutische Behandlung zwischen Baseline- und Postmessung).

Der potenzielle Einflussfaktor „Zusatzbehandlung“ ergab sich aus der mit dem BfArM vereinbarten Aufhebung des entsprechenden Ausschlusskriteriums und berücksichtigt sowohl ambulante, teilstationäre als auch stationäre Behandlungen. Der Einflussfaktor „Geschlecht“ wurde hinzugezogen, da sich eine depressive Erkrankung bei Frauen und Männer auf unterschiedliche Weise ausprägen kann [7]. Bisher liegen aber keine Hinweise dafür vor, dass die Behandlung mit *My7steps App* für die Geschlechter unterschiedlich wirksam sein könnte.

Die vier hier aufgeführten potenziellen Einflussfaktoren können die Wirksamkeit der Behandlung von Depressionen nachweislich beeinträchtigen [93] und wurden dementsprechend im Rahmen von Sensitivitätsanalysen berücksichtigt.

### 7.9.3 Explorative Fragestellungen

Explorative Fragestellungen waren im Studienprotokoll nicht weiter spezifiziert. Da die selbstgewählte Behandlungsdauer in der Stichprobe relativ stark variierte, wurde eine Exploration dieser Variable nachträglich beschlossen.

### 7.9.4 Analyseverfahren

Alle statistischen Analyseverfahren wurden mit der Software *IBM SPSS Statistics* (Version 25) durchgeführt. Das Alphaniveau wurde für alle Analysen auf 0.05 (zweiseitig) festgelegt. Zwischenauswertungen waren nicht vorgesehen. Fehlende Werte wurden durch ein *Baseline-Observation-Carried-Forward*-Verfahren ersetzt und die Wirksamkeit der Intervention im Rahmen von *Intention-to-Treat*-Analysen überprüft.

## Wirksamkeitsanalyse primärer Endpunkt

Zur Analyse des Unterschieds der depressiven Symptomatiken (Baseline- und Post-Messzeitpunkt) wurden die Daten entsprechend dem Studienprotokoll in Abhängigkeit von der Verteilung mit folgenden Verfahren analysiert:

- Bei Normalverteilung: t-Test für abhängige Stichproben
- Bei Nicht-Normalverteilung: Wilcoxon-Vorzeichen-Rang-Test

Da der t-Test relativ robust gegenüber Verletzungen der Normalverteilung ist [34, 39, 59] und eine ausreichend große Stichprobe vorlag ( $n > 30$ ) [89], wurden zwecks direkter Vergleichbarkeit für alle Endpunkte t-Tests gerechnet und ggf. um non-parametrische Tests zur Ergebnisüberprüfung ergänzt, falls eine Normalverteilung der Daten nicht sicher festgestellt werden konnte. Alle t-Tests wurden darüber hinaus mit Bootstrapping-Verfahren von 2000 Samples überprüft und Bias-korrigierte Konfidenzintervalle, BCa 95% CI, angegeben.

Da bei t-Tests für abhängige Stichproben die Verwendung des t-Werts bei der Effektstärkenberechnung zu einer Überschätzung der Populations-Effektstärke führen kann [27], wurde der *standardized response mean* (SRM) als Cohen's  $d$  berechnet. Hierfür wurde der Depressionswert der Baseline-Messung vom Depressionswert der Postmessung subtrahiert und das Ergebnis durch die Standardabweichung der Baseline-Messung geteilt; ein Vorgehen, das Morris und DeShon (2008) empfehlen, da der Baseline-Wert (im Gegensatz zur Standardabweichung der Postmessung oder einer gepoolten Standardabweichung) nicht durch die Intervention beeinflusst wird [72]. Effektgrößen wurden ab Cohen's  $d = 0.2$  als klein, ab 0.5 als mittel und ab 0.8 als groß klassifiziert.

### Wirksamkeitsanalyse sekundärer Endpunkte

Für die sekundären Endpunkte wurden die gleichen Analyseverfahren und Berechnungsgrundlagen angewendet, wie für den primären Endpunkt beschrieben. In den sekundären Wirksamkeitsanalysen wurden folgenden Maße berücksichtigt:

- Persönliche Belastbarkeit (Resilienz): Connor-Davidson Resilience Scale (CD-RISC-10)
- Beeinträchtigung der Alltagsfunktionalität durch die Erkrankung: Daily Functionality Impairment Scale (DFI)
- Subjektive Symptomveränderungen von Modul 1 zu Modul 7 der *My7steps App* als gemittelte Differenz

### Response-Analysen

Um die Responseraten und die klinische Signifikanz der Behandlung zu analysieren, wurden die Symptomveränderungen vom Baseline- zum Post-Messzeitpunkt für den primären Endpunkt Depression berechnet und folgende Definitionen auf alle Fälle mit einer klinisch signifikanten Depressionsausprägung (PHQ-9  $> 9$  zu Beginn der Studie) [65, 78] angewendet:

- **Minimal clinically important difference (MCID):** Vergleichende Studien empfehlen die Verwendung eines *Reliable Change Indexes* (RCI) zur Bestimmung des minimalen klinisch bedeutsamen Unterschieds (MCID) [43, 69]. Der RCI ist die Differenz zwischen den Post- und Baseline-Messungen, geteilt durch den Standardmessfehler des Messinstruments ( $SE_{meas}$ ), der als Standardabweichung der Baseline-Messung multipliziert mit der Quadratwurzel aus eins minus dem Reliabilitätskoeffizient gebildet wird. Um einen intraindividuellen MCID zu erhalten, der das 95%-Konfidenzintervall widerspiegelt, wurde der Standardmessfehler mit 1.96 multipliziert. Wenn der je Fall berechnete RCI größer als 1.96 ist, gilt die Veränderung als zuverlässig. Dieses Vorgehen stellt sicher, dass jede als reliabel klassifizierte Veränderung der Messergebnisse den Messfehler des eingesetzten Messinstruments übersteigt [41]. Als Referenzwerte einer deutschen Stichprobe für die Berechnung des RCIs wurden Normdaten einer Studie mit 5018 Fällen aus dem Jahr 2013 herangezogen [51]: Die Standardabweichung des PHQ-9 betrug in der Studie  $SD = 3.52$ , die Reliabilität wurde mit  $\alpha = 0.87$  angegeben, erfüllt aber nicht die RCI-Bedingung, dass es sich um eine echte Test-Retest-Reliabilität handelt, sodass stattdessen eine für die „schlechteste“ (d.h. konservativste) Fallstichprobe ermittelte  $rtt = 0.81$  aus einer weiteren deutschen Studie herangezogen wurde [61]. Das Ergebnis dieser Berechnung ergab einen RCI-Schwellenwert von 3.60 Punkten Veränderung auf dem PHQ-9-Score und lag damit im Wertebereich bereits publizierter RCIs für den PHQ-9 [10, 61].
- **Response:** Da der RCI den Nachteil aufweist, dass er bei hoher Reliabilität des Messinstruments hohe Werte annehmen kann, selbst wenn das Ausmaß der Änderung gering ist [40, 67], und der vorliegende RCI unter der 5-Punkte-Veränderung lag, wurde im Sinne eines konservativen Testens zur Bestimmung einer klinisch reliablen Veränderung auf eine für den PHQ-9 etablierte Response-Berechnung zurückgegriffen [54, 61, 69]: Als „*Responder mit reliabler Verbesserung*“ wurden im Folgenden alle Studienteilnehmer:innen mit einer 5-Punkte-Reduktion vom Baseline- zum Post-Messzeitpunkt definiert. Als „*Responder mit reliabler Verschlechterung*“ wurden im Folgenden alle Studienteilnehmer:innen mit einer 5-Punkte-Zunahme vom Baseline- zum Post-Messzeitpunkt definiert.
- **Klinische Signifikanz (partielle Remission und Remission):** Ein PHQ-9-Summenwert von 10 ist als Cut-Off-Wert für eine klinisch signifikante Depressionsausprägung etabliert [65, 78]. Für Studienteilnehmer:innen, die einen PHQ-Baselinewert von  $\geq 10$  aufwiesen und nach der Behandlung eine reliable Verbesserung nach der oben ausgeführten Response-Definition sowie einen PHQ-Postmesswert von  $\leq 9$  zeigten, kann von klinischer Signifikanz ausgegangen werden [69]. Berücksichtigt man, dass im Wertebereich von 5 bis 9 des PHQ beim Patienten noch vereinzelte Symptome vorliegen, kann argumentiert werden, dass erst ab PHQ-9-Werten von  $\leq 4$  ein asymptotischer Befund vorliegt und dieser zusätzlich ausgewiesen werden sollte – auch wenn die Forschungsgemeinschaft zu dieser und anderen Auswertungsfragen des PHQ-9 noch zu keinem abschließenden Urteil gelangt ist [69]. Der Vollständigkeit halber wurden in der vorliegenden Untersuchung beide Kriterien in Kombination mit der oben ausgeführten Response-Definition angewendet. Das Erreichen des etablierten Cut-Off-Werts  $\leq 9$  samt einer 5-Punkte-Reduktion wurde hierbei konservativ als „partielle Remission“ und das Erreichen des verschärften Cut-Off-Werts  $\leq 4$  samt einer 5-Punkte-Reduktion als „Remission“ ausgewiesen.

Das Effektstärkemaß für die Response wurde, wie im Abschnitt „Wirksamkeitsanalyse primärer Endpunkt“ dieses Kapitels beschrieben, als Cohen's  $d$  berechnet.

## Sensitivitätsanalysen

Um die Robustheit der Ergebnisse für den primären Endpunkt hinsichtlich bekannter potenzieller Einflussfaktoren im Rahmen der Sensitivitätsanalysen zu überprüfen, wurde eine mehrfaktorielle ANCOVA (Analysis of Covariance) mit Messwiederholungen gerechnet. Hierbei wurden Geschlecht, Zusatzbehandlung und Bildung als between-subjects Faktoren berücksichtigt, sowie die Variable Alter als Kovariate ins Modell eingebunden. Die vom Modell postulierte lineare Beziehung zwischen Kovariate und primären Endpunkt wurde mit Hilfe von Streudiagrammen überprüft. Um die Varianzhomogenität (Homoskedastizität) zu überprüfen, wurde der Levene-Test auf Gleichheit der Fehlervarianzen herangezogen. Als Interventionseffekt wurde der within Effekt der Messwiederholung des primären Endpunkts (PHQ-9) berichtet. Effektstärken wurden als partielles Eta-Quadrat ( $\eta^2_p$ ) ausgegeben sowie zu Vergleichszwecken in Cohen's  $d$  umgerechnet. Die Umrechnung in Cohen's  $d$  erfolgt nach der von Cohen veröffentlichten Formel für den Fall, dass die Anzahl der Beobachtungen in jeder Gruppe gleich ist und wie von ihm empfohlen über Cohen's  $f$  [17]:

$$\eta^2 = f^2 / (1 + f^2)$$

$$d = 2 * f$$

Effektgrößen wurden ab Cohen's  $d = 0.2$  als klein, ab 0.5 als mittel und ab 0.8 als groß klassifiziert. Im Falle von signifikanten Interaktionen wurden Simple Effects Analysen unter Verwendung von Bonferroni-Korrekturen zur Aufklärung der dahinterliegenden Muster durchgeführt.

Als weitere Sensitivitätsanalysen zur kritischen Überprüfung der Wirksamkeit wurden die primären Wirksamkeitsanalysen zum einen nur für Studienteilnehmer:innen „ohne Zusatzbehandlung“ und zum anderen „Per-Protokoll“ durchgeführt:

- Wirksamkeit ohne Zusatzbehandlung: Alle Studienteilnehmer:innen, die während der Studienteilnahme eine zusätzliche Behandlung erhalten hatten, wurden aus der Analyse ausgeschlossen.
- Wirksamkeit Per-Protokoll: Alle Studienteilnehmer:innen, die nicht mindestens das vorletzte (sechste) Modul der *My7steps App* abgeschlossen hatten, wurden aus der Analyse ausgeschlossen.

Die Testverfahren dieser beiden Sensitivitätsanalysen entsprachen der oben beschriebenen primären Wirksamkeitsanalyse (t-Tests unter Berücksichtigung der Verteilung). Die Analyse „ohne Zusatzbehandlung“ wurde nachträglich beschlossen, um aktualisierten Anforderungen des DiGA-Zulassungsverfahrens zu entsprechen.

## Subgruppenanalysen

Zur gesonderten Betrachtung von Subgruppen bzw. zwecks Überprüfung des Einflusses von Gruppenvariablen wurden folgende Analysen durchgeführt:

Subgruppenanalyse Indikationen: Um die Gültigkeit der Ergebnisse für den primären Endpunkt hinsichtlich der vorab spezifizierten Indikationsgruppen (vgl. 7.9.2) zu überprüfen, wurde eine einfaktorielle ANOVA (Analysis of Variance) mit Messwiederholungen unter Einbeziehung der

Diagnosegruppen als Faktor gerechnet. Die statistische Vorgehensweise entsprach bei dieser Subgruppenanalysen der Überprüfung und Konfiguration der im vorherigen Abschnitt beschriebenen ANCOVA-Sensitivitätsanalyse.

Subgruppenanalyse leichte Depression: Um die Gültigkeit der Ergebnisse auch für Studienteilnehmer:innen mit subklinischen bis leichten Depressionen (PHQ-9 < 10) zu überprüfen, wurde eine gesonderte Wirksamkeitsanalyse für diese Gruppe durchgeführt. Das Testverfahren entsprach der oben beschriebenen primären Wirksamkeitsanalyse (t-Tests unter Berücksichtigung der Verteilung). Diese Subgruppenanalyse wurde nachträglich beschlossen, um aktualisierten Anforderungen des DiGA-Zulassungsverfahrens gerecht zu werden.

### **Explorative Analysen**

Die von den Studienteilnehmer:innen selbst gewählte Behandlungsdauer wurde exploriert, um eventuell Hinweise für zukünftige Studien und Behandlungsempfehlungen ableiten zu können. Hierfür wurde die Behandlungsdauer mit Stichprobenmerkmalen, Endpunkten und der App-Bewertung durch die Studienteilnehmer:innen explorativ korreliert. Dropouts konnten aufgrund der nicht genau zu bestimmenden Behandlungsdauer in dieser Analyse nicht berücksichtigt werden, die explorative Analyse wurde entsprechend als Per-Protokoll-Verfahren durchgeführt.

Um einen eventuell quadratischen Zusammenhang überprüfen zu können, wurden drei Ausreißer der Behandlungsdauer ausgeschlossen, ein quadratischer Term der Behandlungsdauer-Variable erstellt und eine Regression mit den Depressions-Change-Scores als Regressand und den beiden Behandlungsdauer-Variablen (einfach und quadratisch) als Regressoren gerechnet sowie graphisch ausgewertet.

### **Sicherheitsanalysen**

Zur Einschätzung der Sicherheit der Behandlung mit *My7steps App* wurden folgende Parameter herangezogen und deskriptiv ausgewertet:

1. Alle Vorkommnisse, wie unerwünschte Ereignisse und unerwünschte Wirkungen.
2. Die Risikoevaluation im Rahmen der technischen Überwachung der Webanwendung.
3. Der Anteil der Studienteilnehmer:innen mit klinisch reliabler Verschlechterung der depressiven Symptomatik.
4. Die Ergebnisse der abschließenden Befragung der Studienteilnehmer:innen zur subjektiven Bewertung der *My7steps App* unter besonderer Berücksichtigung negativer bzw. stark negativer Bewertungen.

## 7.9.5 Analyse und fehlende Werte

Laut Studienprotokoll soll nach dem *Intention-to-Treat*-Prinzip (ITT) ausgewertet werden, was in einer kontrollierten randomisierten Studie (RCT) der zu bevorzugende Standard ist, um die Robustheit der gefundenen Effekte hinsichtlich eines möglichen systematischen Dropouts zu überprüfen [36, 68]. Im Studienprotokoll ist des Weiteren vermerkt, dass fehlende Werte im Rahmen der vorliegenden Pilotstudie nicht zu ersetzen sind. Dies steht im Widerspruch zu einem *Intention-to-Treat*-Verfahren, bei dem es einer geeigneten Strategie im Umgang mit fehlenden Werten bedarf. Eine solche Strategie wurde dementsprechend nachträglich beschlossen: Aufgrund des ökonomischen Anspruchs einer Pilotstudie, der überschaubaren Stichprobengröße, der verhältnismäßig geringen Anzahl an Dropouts (ca. 10%) und dem Vorliegen nur einer Post-Baseline-Messung war ein *Baseline-Observation-Carried-Forward*-Verfahren angezeigt, bei dem die zur Baseline gemessenen Werte ggf. auch als Werte der Postmessung herangezogen werden [58]. Dies entspricht inhaltlich einer „Nichtveränderungsannahme“ für die betroffenen Fälle.

## 7.10 Festlegung der Fallzahl

### 7.10.1 Ausgangssituation und Fallzahl

Auf Basis der statistischen Minimalgröße für einen Prä-Post-Gruppenvergleich und unter Berücksichtigung der Forderung des BfArMs, alle drei Indikationsgruppen abzubilden, ergab sich eine Fallzahl von 60 bis 80 Studienteilnehmer:innen, die laut Studienprotokoll erfolgreich die Studie durchlaufen sollten. Da keine exakte Powerkalkulation für die Pilotstudie vorgesehen war, wurde die benötigte Fallzahl auf Grundlage wissenschaftlicher Erfahrung geschätzt: Die minimale Subgruppengröße sollte berücksichtigen, dass eventuell weitere Variablen (wie z.B. Geschlecht oder Alter) innerhalb der Subgruppen differenziert werden können müssen, und wurde dementsprechend auf  $n = 20$  gesetzt. Aufgrund der präspezifizierten drei Subgruppen ergab dies eine geschätzte Fallzahluntergrenze von 60 Studienteilnehmer:innen für die Studie. Da keine Stratifizierung vorgesehen war und Studienteilnehmer:innen mit den gesuchten ICD-10-Diagnosen nicht gezielt rekrutiert werden konnten, war es nötig diese Zahl etwas zu erhöhen, damit die Subgruppen ausreichend Fallzahlen erreichen konnten: Entsprechend wurde die zu bevorzugende Fallzahl der Studie auf 80 Studienteilnehmer:innen festgesetzt. Bei einer angenommenen Drop-out-Rate von ca. 20% wurde für die Untersuchung angesetzt, dass insgesamt ca. 100 Studienteilnehmer:innen rekrutiert werden sollen.

## 7.11 Änderungen des Prüfplans

Änderungen des Prüfplans sowie der darin festgelegten Voraussetzungen waren aus Gründen der Gewährleistung vergleichbarer Studienbedingungen sowie zum Erhalt der statistischen Validität und Güte der Datenauswertung generell nicht vorgesehen. Nur nach Abstimmung des Sponsors mit dem wissenschaftlichen Leiter und dem Leiter der klinischen Studie sowie den im Prüfplan involvierten Prüfern und Psychotherapeuten konnte in Ausnahmefällen eine Änderung des Prüfplans unter besonderen Umständen erwogen werden.

### **Folgende nachträgliche Änderungen am Prüfplan wurden dokumentiert:**

- Es wurde keine Änderungen am Prüfplan dokumentiert.

**Folgende Änderungen der Dokumentation und Korrekturen an Prüfbögen liegen vor:**

- Es wurde keine Abweichung von den implementierten Dokumentationsmaßnahmen festgestellt.

**Folgende Korrekturen an Prüfbögen waren notwendig:**

- Es waren keine Korrekturen von Prüfbögen notwendig.

**Folgende Änderungen der statistischen Auswertung wurden beschlossen:**

Nachträglich beschlossene Verfahren aufgrund nicht explizierter Freistellen des Studienprotokolls:

- *Baseline-Observation-Carried-Forward*-Verfahren zur Ersetzung fehlender Werte im Rahmen der *Intention-to-Treat*-Analysen

Nachträglich beschlossene Analysen, um aktualisierten Anforderungen des DiGA-Zulassungsverfahrens zu entsprechen:

- Sensitivitätsanalyse "Ohne Zusatzbehandlung", um die Wirksamkeit der Intervention bei Abwesenheit jedweder psychotherapeutischer Zusatzbehandlung zu überprüfen.
- Subgruppenanalyse "leichte Depression", um die Wirksamkeit gesondert für subklinische bis leichte Depressionen zu überprüfen.
- Analysen mit externen Kontrollgruppen (literaturbasierter Kontrollarm), um die gefundenen Ergebnisse der Interventionsgruppe mit Kontrollgruppenverläufen vergleichbarer Stichproben zu kontrastieren, sodass die Effekte einer care-as-usual als auch des natürlichen Krankheitsverlaufs bei der Bewertung der Ergebnisse berücksichtigt werden können.

Die Notwendigkeit der hier aufgeführten Ergänzungen des statistischen Analyseplans wurde erst nach Abschluss der Erhebung festgestellt. Da es sich nicht um inhaltliche Änderungen der Auswertungsstrategie oder der Hypothesen handelt, sondern um eine Harmonisierung von Protokollvorgaben und Zulassungsanforderungen, wurde auf die Erstellung eines nachträglichen Protokoll-Amendments verzichtet. Die in Absprache mit Sponsor und Studienleitung beschlossenen Ergänzungen der statistischen Auswertung wurden im vorliegenden Abschlussbericht dokumentiert, begründet und werden durch die Unterzeichnung des Berichts durch Sponsor und Studienleitung formell bestätigt.

## 8 Stichprobe

### 8.1 Stichprobe der Studie

Im Rahmen der Rekrutierung von November 2021 bis Februar 2022 bekundeten 93 Personen ihr Interesse an der Studie und wurden als potenzielle Studienteilnehmer:innen erfasst. Zwei Personen erfüllten nicht die Ein- und Ausschlusskriterien und wurden im Rahmen des diagnostischen Aufnahmegesprächs von der Teilnahme ausgeschlossen und auf alternative Behandlungsangebote hingewiesen. Zwei weitere Studienteilnehmer:innen brachen die Zuordnung vor der Registrierung ab und beendeten die Teilnahme; die Fälle wurden entsprechend dem Studienprotokoll im Folgenden von der Dokumentation ausgeschlossen. Die Daten von 9 weiteren Studienteilnehmer:innen wurden von der *Per-Protokoll*-Analyse ausgeschlossen, da sie das im Studienprotokoll definierte Kriterium der Nichtadhärenz an die vorgesehene Intervention erfüllten, indem sie das vorletzte (sechste) Moduls der *My7steps App* nicht abschlossen und ihre Teilnahme vorzeitig beendeten; in die *Intention-to-treat*-Analyse wurden sie einbezogen. Die tatsächliche Stichprobe umfasste somit 89 Studienteilnehmer:innen, die Dropout-Rate betrug 10,11% (9 von 89). Es gab keinen Antrag auf Datenlöschung, entsprechend gingen die Daten von 89 Studienteilnehmer:innen in die ITT-Analysen und die Daten von 80 Studienteilnehmer:innen in die PP-Analysen ein (vgl. Abbildung 2).

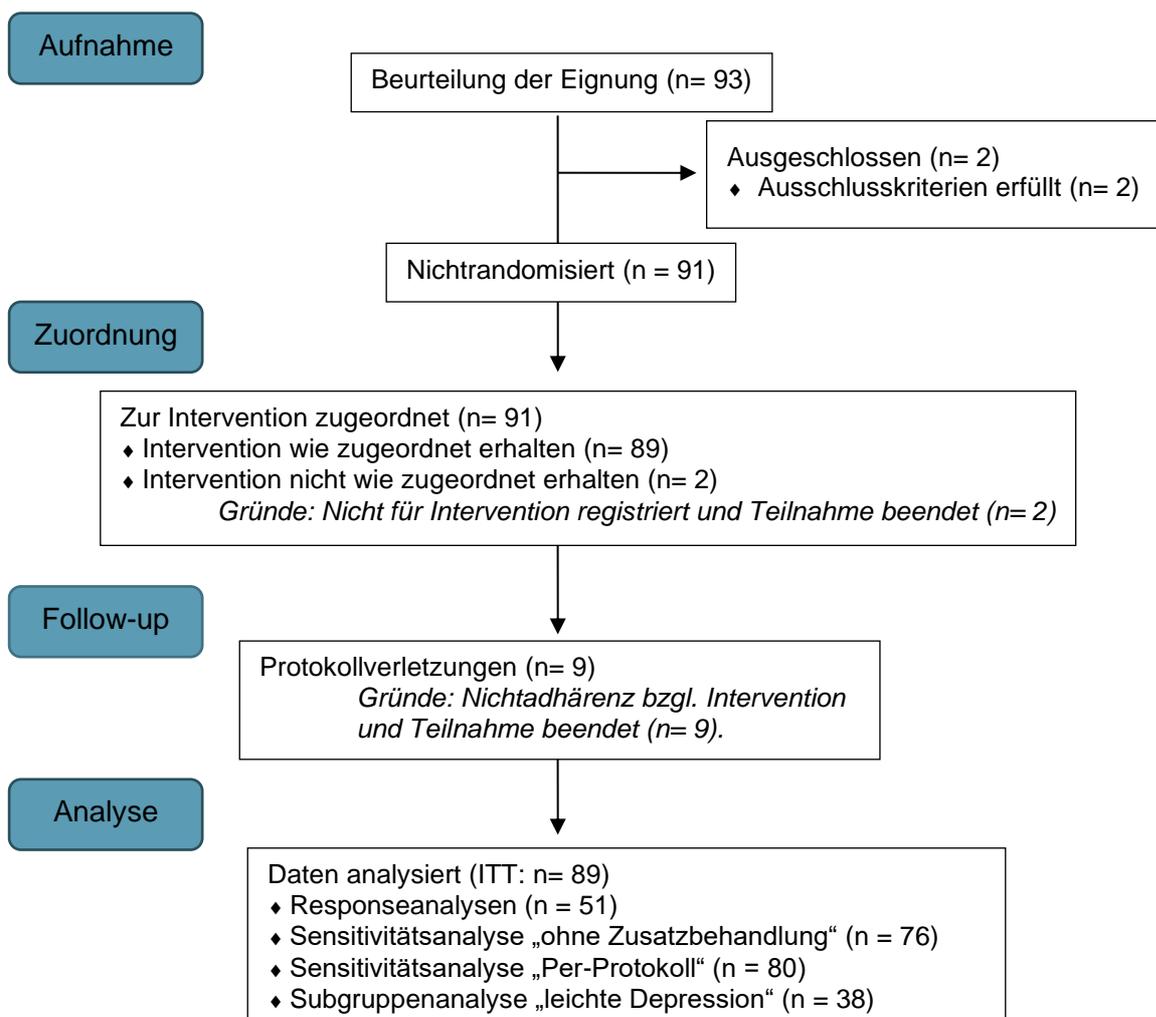


Abbildung 2: Flussdiagramm der Registrierung, Zuordnung und Analyse

## 8.2 Protokollverletzungen im Rahmen der Patientenrekrutierung und -erhebung

Es kam im Rahmen der Studie zu keinen Protokollverletzungen bezüglich der Vorgaben an Ein- und Ausschluss, Widerruf, oder sonstigen Aspekten der Rekrutierung und Betreuung von Studienteilnehmer:innen.

Es kam zu folgenden Protokollverletzungen bezüglich des Abbruchs der Behandlung:

- Neun Studienteilnehmer:innen erfüllten das Kriterium der Nichtadhärenz an die vorgesehene Intervention, indem sie das vorletzte (sechste) Moduls der *My7steps App* nicht abschlossen und ihre Teilnahme vorzeitig beendeten. Diese Fälle wurden in der *Per-Protokoll*-Analyse nicht berücksichtigt, sind aber Teil aller anderen auf dem ITT-Verfahren basierenden Analysen.

## 9 Ergebnisse: Wirksamkeitsevaluation

### 9.1 Demographische Angaben zur Stichprobe

Im Rahmen der Studie wurden relevante Merkmale der Stichprobe dokumentiert. In Tabelle 3 finden sich folgende Angaben zur Stichprobe zum Baseline-Messzeitpunkt: Alter, Geschlecht, Interviewsprache, Familienstand, Bildungsniveau, vorherige psychotherapeutische Erfahrungen, ICD-10-Diagnosen, und aktuelle Schwere der depressiven Symptomatik nach den PHQ-9-Kategorien.

Tabelle 3: Demographische Angaben zur Stichprobe

<b>Merkmal</b>	<b>Werte</b>	<b>N (fehlend)</b>
<b>Soziodemographische Merkmale</b>		
Alter, Durchschnitt (SD)	30.48 (7.77)	89 (0)
Alter, Minimum - Maximum (Range)	19 – 54 (35)	89 (0)
Geschlecht: weiblich, <i>n</i> (%)	66 (75.9)	87 (2)
<b>Interviewsprache</b>		<b>89 (0)</b>
Deutsch, <i>n</i> (%)	25 (28.1)	
Englisch, <i>n</i> (%)	21 (23.6)	
Arabisch, <i>n</i> (%)	17 (19.1)	
Dari/Farsi, <i>n</i> (%)	26 (29.2)	
<b>Familienstand</b>		<b>89 (0)</b>
Ledig, <i>n</i> (%)	38 (42.7)	
Verheiratet, <i>n</i> (%)	27 (30.3)	
In einer Beziehung lebend, <i>n</i> (%)	19 (21.3)	
Geschieden, <i>n</i> (%)	5 (5.6)	
<b>Bildung nach Schulabschluss</b>		<b>89 (0)</b>
Grundschule (~ 6 Jahre Schule), <i>n</i> (%)	1 (1.1)	
Sekundarschule (~ 9 Jahre Schule), <i>n</i> (%)	1 (1.1)	
Mittelschule (~ 10 Jahre Schule), <i>n</i> (%)	2 (2.2)	
Abitur / High School (~ 12 Jahre Schule), <i>n</i> (%)	22 (24.7)	
Universitätsabschluss, <i>n</i> (%)	63 (70.8)	

<b>Vorherige Psychotherapie</b>	<b>89 (0)</b>
keine Psychotherapie vor der Teilnahme, <i>n</i> (%)	42 (47.2)
zuvor in ambulanter Behandlung, <i>n</i> (%)	46 (51.7)
zuvor in teilstationärer Behandlung, <i>n</i> (%)	0 (0.0)
zuvor in stationärer Behandlung, <i>n</i> (%)	1 (1.1)
<b>ICD-10 Diagnose</b>	<b>89 (0)</b>
leichte Depression (F32.0), <i>n</i> (%)	15 (16.9)
mittelschwere Depression (F32.1), <i>n</i> (%)	17 (19.1)
rezidivierende leichte Depression (F33.0), <i>n</i> (%)	18 (20.2)
rezidivierende mittelschwere Depression (F33.1), <i>n</i> (%)	18 (20.2)
Anpassungsstörung (F43.2), <i>n</i> (%)	21 (23.6)
<b>Depressionswerte (PHQ-9-Scores) zum Baseline-Messzeitpunkt</b>	<b>89 (0)</b>
Subklinisch (< 5), <i>n</i> (%)	4 (4.5)
Leicht (< 10), <i>n</i> (%)	34 (38.2)
Moderat (< 15), <i>n</i> (%)	31 (34.8)
Mittelschwer (< 20), <i>n</i> (%)	16 (18.0)
Schwer (> 19), <i>n</i> (%)	4 (4.5)

*SD = Standardabweichung; ICD-10 International Statistical Classification of Diseases and Related Health Problems (10th revision); PHQ-9 = Patient Health Questionnaire.*

## 9.2 Dokumentation der weiteren Kontakte zwischen Studienteilnehmer:innen und Prüfer

Im Rahmen der Studie konnte es – neben der diagnostischen Visite und den beiden Studienvisiten – zu weiteren Kontakten zwischen Mitgliedern des Studienteams und Studienteilnehmer:innen kommen.

Dies war zum einen im Rahmen der vorgesehenen Checkup-Calls (vgl. Abschnitt 7.5.1) möglich. Die Kontakte wurden dokumentiert und die Prüfer gaben die Länge der Telefongespräche auf der Grundlage ihrer Telefondaten am Ende der Erhebung an:

Aufgrund der Checkup-Calls kam es zu insgesamt 146 telefonischen Kontakten im Studienzeitraum. Die Anzahl der Kontakte je Studienteilnehmer:in betrug damit  $M = 1.64$  ( $SD = 0.87$ ), mit einem Minimum von 0 und einem Maximum von 4 Checkup-Calls. Im Schnitt betrug die Dauer der Checkup-Calls  $M = 4.44$

Minuten ( $SD = 4.22$ ) mit einem Minimum von 0 und einem Maximum von 15 Minuten. Es wurden keine Abweichungen von den Vorgaben des Gesprächsleitfadens festgestellt.

Darüber hinaus waren Kontakte im Rahmen von Sicherheitsaktionen zur Lösung von Warnungen seitens des Nutzerreise Monitoring möglich (vgl. Abschnitt 7.5.2). Die Kontakte wurden dokumentiert und die Prüfer gaben die Länge der Telefongespräche auf der Grundlage ihrer Telefondaten am Ende der Erhebung an:

Innerhalb des Studienzeitraums gab das Monitoring-System der App 135 Warnungen für Studienteilnehmer:innen aus, von denen der Großteil telefonisch geklärt wurde. Bei Vorliegen einer Warnung wurde i.d.R. erst einmal eine E-Mail oder SMS versendet, um ein Gespräch zu vereinbaren – zum Teil kam es hier schon zu einer Klärung. Die Anzahl der größtenteils telefonischen Kontakte im Rahmen der Nachverfolgung von Warnungen betrug je Studienteilnehmer:in  $M = 1.52$  ( $SD = 1.64$ ), mit einem Minimum von 0 und einem Maximum von 14 Warnungen. Die durchschnittliche Dauer der telefonischen Kontakte im Rahmen der Nachverfolgung von Warnungen betrug  $M = 3.44$  Minuten ( $SD = 3.01$ ), mit einem Minimum von 0 und einem Maximum von 15 Minuten. Die Nachverfolgungen resultierten alle in Entwarnungen; es musste in keinem Fall die Empfehlung ausgesprochen werden, professionelle Hilfe außerhalb der Intervention hinzuzuziehen.

Schließlich bestand theoretisch die Möglichkeit, dass es während der Evaluation zu einem vom Studienteilnehmer:in initiierten persönlichen oder telefonischen Kontakten kam, z.B. im Rahmen auftretender Symptome und Beschwerden oder notfalliger Vorstellungen. Zu den Kontakten zwischen Mitgliedern des Studienteams und Studienteilnehmer:innen lagen am Ende des Nachbeobachtungszeitraum keine Daten vor, da diese nicht erfolgten.

### 9.3 Analysen zur Wirksamkeit

Zur Überprüfung der Normalverteilung der Differenzen der Endpunktmessungen wurden Histogramme und Q-Q-Plots verwendet. Da die visuelle Inspektion teilweise uneindeutige Ergebnisse lieferte, wurden Kolmogorov–Smirnov-Tests für alle Endpunkte durchgeführt. Es wurden signifikante Ergebnisse für den primären Endpunkt Depression sowie den sekundären Endpunkt Resilienz gefunden; für beide Endpunkte konnte das Vorliegen einer Normalverteilung nicht angenommen werden (alle anderen  $ps > .05$ ). Da der t-Test aber relativ robust gegenüber Verletzungen der Normalverteilung ist [34, 39, 59] und mit 89 gültigen Fällen eine ausreichend große Stichprobe vorlag ( $n > 30$ ) [89], wurden zwecks direkter Vergleichbarkeit für alle Endpunkte t-Tests gerechnet und für die beiden betroffenen Endpunkte non-parametrische Tests zur Ergebnisüberprüfung ergänzt.

Sofern durch vollständige Datensätze gewährleistet, wurden Summenwerte für die Analysen herangezogen, um eine direkte Interpretation der Skalenwerte zu ermöglichen. Bei einzelnen fehlenden Item-Antworten wurde – solange mindestens 50% der Items beantwortet waren – alternativ der Mittelwert der Antworten verwendet. Die Verwendung von Mittelwerten war lediglich für den sekundären Endpunkt „Beeinträchtigung der Alltagsfunktionalität“ (DFI) notwendig, für Depression (PHQ-9) und Resilienz (CD-Risc-10) lagen hingegen für alle Fälle vollständige Item-Antworten vor, es konnten die vorgesehenen Summenwerte verwendet werden. Für den sekundären Endpunkt „Subjektive Symptomveränderungen“ lag dagegen per Definition nur eine gemittelte Variable vor, da die Studienteilnehmer:innen eigenständig bestimmen konnten, ob sie ein Symptom oder zwei Symptome als dominant klassifizieren und über die Interventionsmodule 1 und 7 überwachen lassen wollten. Dieser

Endpunkt ergab sich also je nach Fall aus zwei zusammengefassten Mittelwertdifferenzen oder nur einer Mittelwertdifferenz.

Alle Analysen wurden nach dem *Intention-to-Treat*-Verfahren durchgeführt, falls nicht anders ausgewiesen.

In Tabelle 4 sind für beide Messzeitpunkte die Mittelwerte und Standardfehler aller Endpunkte sowie die Mittelwertdifferenzen als auch die dazugehörigen Standardfehler und Effektstärken angegeben. Zur korrekten Berechnung von Konfidenzintervallen für Effektstärken speziell bei gepaarten Stichproben findet sich in der einschlägigen Literatur leider nur wenig; die Möglichkeit einer verlässlichen Bestimmung wird für den Fall gepaarter Stichproben generell angezweifelt [32]. Auf eine nicht verlässliche Berechnung der Konfidenzintervalle wurde dementsprechend verzichtet.

**Tabelle 4:** Mittelwerte und Effektstärken aller Endpunkte über die Messzeitpunkte (primäre Wirksamkeitsanalysen)

	<i>Baseline</i>	<i>Post</i>	<i>Differenz</i>		
	<i>M (SE)</i>	<i>M (SE)</i>	<i>MD</i>	<i>SE</i>	<i>ES</i>
<b>PHQ-9</b>	10.81 (0.47)	6.70 (0.46)	-4.11	0.42	-0.93
<b>CD-RISC-10</b>	19.83 (0.62)	24.11 (0.63)	4.28	0.54	0.73
<b>DFI</b>	1.90 (0.09)	1.12 (0.08)	-0.78	0.09	-0.95
<b>Symptome</b>	77.72 (1.76)	48.26 (2.50)	-29.46	2.27	-1.89

*PHQ-9 = Patient Health Questionnaire; CD-RISC-10 = Connor-Davidson Resilience Scale; DFI = Daily Functionality Impairment Scale; Symptome = subjektive Symptomstärken in Modul 1 bzw. Modul 7 der My7steps App; M = Mittelwert; SE = Standardfehler; MD = Mittelwertdifferenz; ES = Effektstärke (Cohen's d).*

### 9.3.1 Wirksamkeitsanalyse primärer Endpunkt

#### Ergebnis des t-Tests für abhängige Stichproben für den primären Endpunkt „Depression“

Im Durchschnitt wiesen die Studienteilnehmer:innen ( $N = 89$ ) zum Post-Messzeitpunkt nach der Intervention niedrigere Depressions-Summenwerte auf ( $M = 6.70$ ,  $SE = 0.46$ ) als zum Baseline-Messzeitpunkt ( $M = 10.81$ ,  $SE = 0.47$ ). Dieser Unterschied,  $M_{diff} = -4.11$ , BCa 95% CI [-4.98, -3.27], war in einem t-Test für abhängige Stichproben signifikant  $t(88) = -9.90$ ,  $p < .001$ , und repräsentiert eine große Effektstärke,  $d = -0.93$ . Die Berechnung ergab eine negative Effektstärke, wobei niedrigere Werte bessere Ergebnisse auf dem primären Endpunkt repräsentieren. Der Interquartilsabstand auf der Skala des Messinstruments betrug zur Baseline 7 Punkte, bei einem Minimum von 3 und einem Maximum von 21 Punkten ( $R = 18$ ;  $SD = 4.42$ ), während er zur Post-Messung 5 Punkte betrug, bei einem Minimum von 1 und einem Maximum von 21 Punkten ( $R = 20$ ;  $SD = 4.31$ ).

Da eine Normalverteilung für den primären Endpunkt Depression nicht sicher angenommen werden konnte, wurde das Ergebnis des t-Tests mit einem non-parametrischen Verfahren überprüft: Der

Wilcoxon-Vorzeichen-Rang-Test bestätigte, dass sich die Depressionswerte zur Baseline-Messung ( $Mdn = 10.00$ ) und Postmessung ( $Mdn = 6.00$ ) signifikant voneinander unterschieden,  $z = -7.19$ ,  $p < .001$ .

### 9.3.2 Wirksamkeitsanalysen sekundäre Endpunkte

#### **Ergebnis des t-Tests für abhängige Stichproben für den sekundären Endpunkt „Resilienz“**

Im Durchschnitt wiesen die Studienteilnehmer:innen ( $N = 89$ ) zum Post-Messzeitpunkt nach der Intervention höhere Resilienz-Summenwerte auf ( $M = 24.11$ ,  $SE = 0.63$ ) als zum Baseline-Messzeitpunkt ( $M = 19.83$ ,  $SE = 0.62$ ). Dieser Unterschied,  $M_{diff} = 4.28$ , BCa 95% CI [3.25, 5.44], war in einem t-Test für abhängige Stichproben signifikant  $t(88) = 7.93$ ,  $p < .001$ , und repräsentiert eine mittlere Effektstärke,  $d = 0.73$ . Der Interquartilsabstand auf der Skala des Messinstruments betrug zur Baseline 7 Punkte, bei einem Minimum von 6 und einem Maximum von 33 Punkten ( $R = 27$ ;  $SD = 5.83$ ), während er zur Post-Messung 8 Punkte betrug, bei einem Minimum von 3 und einem Maximum von 37 Punkten ( $R = 34$ ;  $SD = 5.97$ ).

Da eine Normalverteilung für den sekundären Endpunkt Resilienz nicht sicher angenommen werden konnte, wurde das Ergebnis des t-Tests mit einem non-parametrischen Verfahren überprüft: Der Wilcoxon-Vorzeichen-Rang-Test bestätigte, dass sich die Depressionswerte zur Baseline-Messung ( $Mdn = 24.00$ ) und Postmessung ( $Mdn = 20.00$ ) signifikant voneinander unterschieden,  $z = -6.31$ ,  $p < .001$ .

#### **Ergebnis des t-Tests für abhängige Stichproben für den sekundären Endpunkt „Beeinträchtigung der Alltagsfunktionalität“**

Im Durchschnitt wiesen die Studienteilnehmer:innen ( $N = 89$ ) zum Post-Messzeitpunkt nach der Intervention niedrigere Mittelwerte der Alltagsbeeinträchtigung auf ( $M = 1.12$ ,  $SE = 0.08$ ) als zum Baseline-Messzeitpunkt ( $M = 1.90$ ,  $SE = 0.09$ ). Dieser Unterschied,  $M_{diff} = -0.78$ , BCa 95% CI [-0.95, -0.62], war in einem t-Test für abhängige Stichproben signifikant,  $t(88) = -9.03$ ,  $p < .001$ , und repräsentiert eine große Effektstärke,  $d = -0.95$ . Die Berechnung ergab eine negative Effektstärke, wobei niedrigere Werte bessere Ergebnisse auf dem sekundären Endpunkt repräsentieren. Der Interquartilsabstand der gemittelten Antworten betrug zur Baseline 1.09, bei einem Minimum von 0 und einem Maximum von 3.75 ( $R = 3.75$ ;  $SD = 0.82$ ), während er zur Post-Messung 1.10 betrug, bei einem Minimum von 0 und einem Maximum von 3.75 ( $R = 3.75$ ;  $SD = 0.79$ ).

#### **Ergebnis des t-Tests für abhängige Stichproben für den sekundären Endpunkt „Subjektive Symptomveränderungen“**

Im Durchschnitt wiesen die Studienteilnehmer:innen ( $N = 84$ ) in Modul 7 am Ende der Intervention niedrigere Symptom-Mittelwerte auf ( $M = 48.26$ ,  $SE = 2.50$ ) als in Modul 1 zu Beginn der Intervention ( $M = 77.72$ ,  $SE = 1.76$ ). Dieser Unterschied,  $M_{diff} = -29.46$ , BCa 95% CI [-33.87, -24.96], war in einem t-Test für abhängige Stichproben signifikant  $t(83) = -12.98$ ,  $p < .001$ , und repräsentiert eine große Effektstärke,  $d = -1.89$ . Die Berechnung ergab eine negative Effektstärke, wobei niedrigere Werte bessere Ergebnisse auf dem sekundären Endpunkt repräsentieren. Der Interquartilsabstand der gemittelten Antworten betrug

zur Baseline 28.13, bei einem Minimum von 37.00 und einem Maximum von 100.00 ( $R = 63.00$ ;  $SD = 15.57$ ), während er zur Post-Messung 33.88 betrug, bei einem Minimum von 9.50 und einem Maximum von 100.00 ( $R = 90.50$ ;  $SD = 22.74$ ).

### 9.3.3 Response-Analysen

Von 89 Studienteilnehmer:innen wiesen 57.3% ( $N = 51$ ) zur Baseline einen Wert über 9 auf dem primären Endpunkt Depression (gemessen mit dem PHQ-9) auf und wurden in die Response-Analysen aufgenommen. In Tabelle 5 sind für alle vier Responsevariablen und ihre Ausprägungen die Häufigkeiten sowie prozentualen Anteile, Mittelwertdifferenzen und Standardabweichungen sowie die jeweiligen Effektstärken (als Cohen's  $d$ ) dargestellt.

Eine Reduktion der Depressionssymptome, die mindestens die *minimal clinically important difference* (MCID) erfüllte, definiert als RCI größer 1.96, wiesen 64.7% (33/51) der betrachteten Fälle auf. Diese statistisch reliablen Veränderungen auf dem primären Endpunkt waren alle negativ, stellten also ausschließlich Verbesserungen der Depressionswerte dar.

Eine klinisch reliable Response, definiert als mindestens 5-Punkte-Veränderung der depressiven Symptomatiken von der Baseline- zur Post-Messung, zeigten 60.8% (31/51) der analysierten Studienteilnehmer:innen. Hierbei handelte es sich ausschließlich um Responder mit reliabler Verbesserung; eine reliable Verschlechterung trat nicht auf.

Klinische Signifikanz in Form einer partiellen Remission, definiert als Erreichen des PHQ-9-Cut-Off-Werts  $\leq 9$  samt einer 5-Punkte-Reduktion des Depressionswerts, wiesen 54.9% (28/51) der analysierten Fälle auf. Eine vollständige Remission, definiert als das Erreichen des verschärften PHQ-9-Cut-Off-Werts  $\leq 4$  samt einer 5-Punkte-Reduktion des Depressionswerts, wiesen 21.6% (11/51) der analysierten Fälle auf.

Tabelle 5: Response-Analysen für den primären Endpunkt

	<i>n</i> (%)	<i>MD</i> ( <i>SD</i> )	<i>Effektstärke</i>
<b>Minimal clinically important difference (MCID)</b>			
verschlechtert	0 (0.0%)	<i>n. a.</i>	<i>n. a.</i>
unverändert	18 (35.3%)	-1.00 (1.28)	-0.30
verbessert	33 (64.7%)	-7.94 (3.03)	-2.59
<b>Response (5-Punkte-Veränderung)</b>			
verschlechtert	0 (0.0%)	<i>n. a.</i>	<i>n. a.</i>
unverändert	20 (39.2%)	-1.30 (1.51)	-0.36
verbessert	31 (60.8%)	-8.19 (2.95)	-2.84
<b>Partielle Remission</b>			
keine	23 (45.1%)	-1.87 (2.07)	-0.50
part. Remission	28 (54.9%)	-8.46 (2.98)	-3.22
<b>Remission</b>			
keine	40 (78.4%)	-4.35 (3.87)	-1.31
Remission	11 (21.6%)	-9.64 (2.38)	-4.24

Von 89 Studienteilnehmer:innen wiesen  $N = 51$  (57.3%) zur Baseline einen Wert über 9 auf dem primären Endpunkt (PHQ-9) auf und wurden in die Response-Analysen aufgenommen. Die minimal clinically important difference (MCID) wurde auf Basis des Reliable Change Index (RCI) berechnet.

$n$  = Anzahl der Studienteilnehmer:innen in der jeweiligen Responsekategorie; % = prozentualer Anzahl der Studienteilnehmer:innen in der jeweiligen Responsekategorie; MD = Mittelwertdifferenzen des PHQ-9; SD = Standardabweichung; *n. a.* = nicht anwendbar; ES = Effektstärke (Cohen's  $d$ ).

### 9.3.4 Sensitivitätsanalysen primärer Endpunkt

#### Mehrfaktorielle ANCOVA unter Berücksichtigung von Geschlecht, Bildung, Zusatzbehandlung und Alter

Für die Überprüfung von möglichen Einflussvariablen in einer mehrfaktoriellen ANCOVA (Analysis of Covariance) mit Messwiederholungen mussten die beiden Faktoren Bildung und Zusatzbehandlung aufgrund einiger zu gering besetzter Zellen dichotomisiert werden; Bildung wurde hierbei aufgrund des hohen Niveaus in der Stichprobe als „mit Universitätsabschluss“ versus „ohne Universitätsabschluss“ kodiert. 14.6% (13/89) der Studienteilnehmer:innen gaben an, während der Studien eine psychotherapeutische Behandlung begonnen zu haben. Da nur zwei Messzeitpunkte vorlagen, war die Annahme der Sphärizität gegeben und musste nicht überprüft werden. Der Boxsche M-Test bestätigte die

Gleichheit der Kovarianzmatrizen, der Levene-Test die Gleichheit der Fehlervarianzen (alle  $p$ s > .05). Die resultierende 2x2x2 ANCOVA mit Messwiederholungen war somit für die Faktoren Geschlecht, Bildung und Zusatzbehandlung sowie für die Kovariate Alter adjustiert. In zwei Fällen lagen keine Angaben zum Geschlecht vor, diese konnten in der Analyse nicht berücksichtigt werden. Tabelle 15 und Tabelle 15 im Anhang zeigen die Mittelwerte, Standardabweichungen und Zellbesetzungen für alle aufgenommenen Variablen zu beiden Messzeitpunkten. Abbildung 3, Abbildung 4 und Abbildung 5 im Folgenden zeigen die Baseline- und Post-Depressionswerte der drei aufgenommenen dichotomen Faktoren, ergänzt durch lineare Regressionsgerade für die sich jeweils ergebenden Subgruppen. Die visuelle Inspektion der Datenpunkte und Regressionsgeraden zeigte nur geringe Unterschiede in allen drei Subgruppenvergleichen, erkennbar an den vergleichbaren Steigungen der Regressionsgeraden.

Das Ergebnis der mehrfaktoriellen ANCOVA bestätigte die primäre Wirksamkeitsanalyse unter Berücksichtigung der Einflussfaktoren: Die Studienteilnehmer:innen ( $N = 87$ ) wiesen zum Post-Messzeitpunkt ( $M = 6.70$ ;  $SD = 4.31$ ) einen signifikant geringeren Schweregrad depressiver Symptome auf als zur Baseline ( $M = 10.81$ ;  $SD = 4.42$ ),  $F(1, 79) = 13.04$ ,  $p < .01$ ,  $\eta^2_p = 0.14$  bzw. Cohen's  $d = 0.81$ , was eine große Effektstärke darstellt. Alle anderen Faktoren, Interaktionen und paarweisen Vergleiche waren nicht signifikant (alle  $p$ s > .05; vgl. Tabelle 6). Die Wirksamkeit der Intervention wurde dementsprechend nicht signifikant von den berücksichtigten Drittvariablen modifiziert.

**Tabelle 6:** Kennwerte der mehrfaktoriellen ANCOVA mit Messwiederholungen zur Überprüfung möglicher Einflussfaktoren

	<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>p</b>	<b><math>\eta^2_p</math></b>
<b>Treatment</b>	94.36	1	94.36	13.04	.001	.14
<b>Treatment x Age</b>	26.39	1	26.39	3.65	.060	.04
<b>Treatment x Thera</b>	19.64	1	19.64	2.71	.103	.03
<b>Treatment x Edu</b>	15.24	1	15.24	2.10	.151	.03
<b>Treatment x Sex</b>	0.02	1	0.02	0.00	.960	.00
<b>Treatment x Edu x Thera</b>	2.97	1	2.97	0.41	.524	.00
<b>Treatment x Sex x Thera</b>	3.56	1	3.56	0.49	.485	.01
<b>Treatment x Sex x Edu</b>	25.69	1	25.69	3.55	.063	.04
<b>Error</b>	571.68	79	7.24			

*Treatment = Behandlungseffekt auf den primären Endpunkt (PHQ-9); Age = Kovariate Alter; Thera = Faktor Zusatzbehandlung; Edu = Faktor Bildung; Sex = Faktor Geschlecht; Error = Variabilität innerhalb der Gruppen (nicht aufgeklärte, zufällige Fehlervarianz); SS = Summe der Quadrate; df = Freiheitsgrade; MS = Mittel der Quadrate; F = empirisch ermittelter F-Wert; p = p-Wert (Signifikanz);  $\eta^2_p$  = partielles Eta-Quadrat (Effektstärke).*



Abbildung 3: Baseline- und Post-Depressionswerte aufgeteilt nach Geschlecht (n = 87)

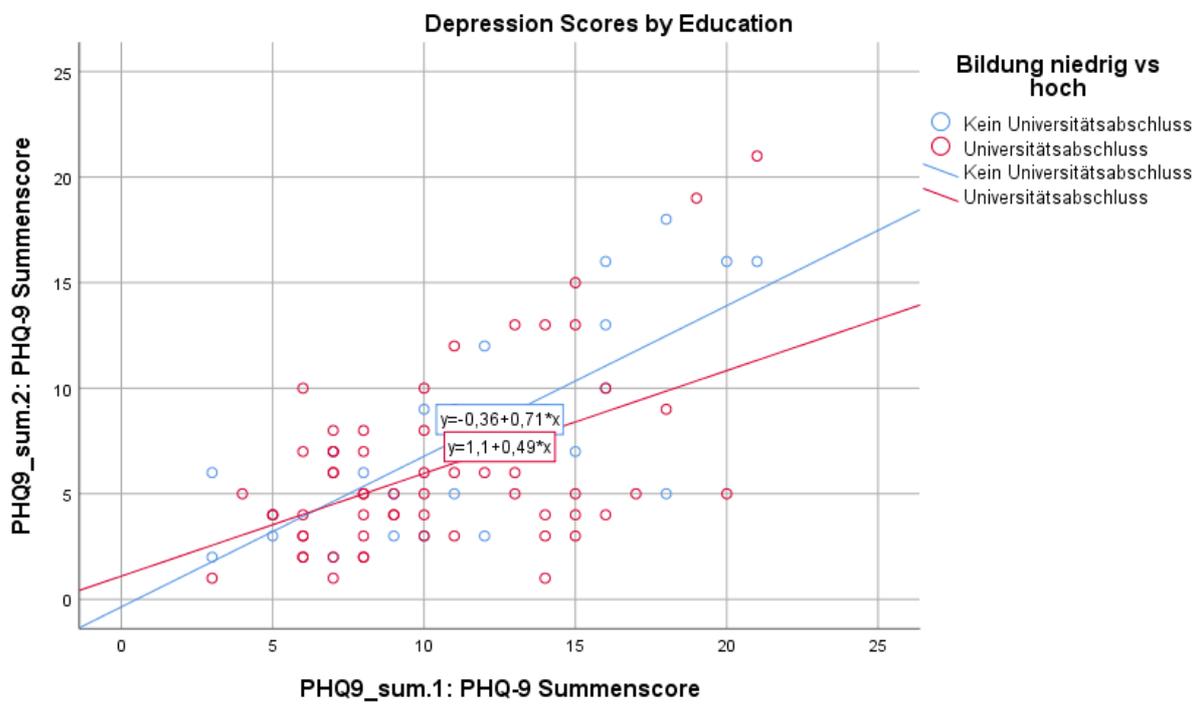


Abbildung 4: Baseline- und Post-Depressionswerte aufgeteilt nach Bildung (n = 89)

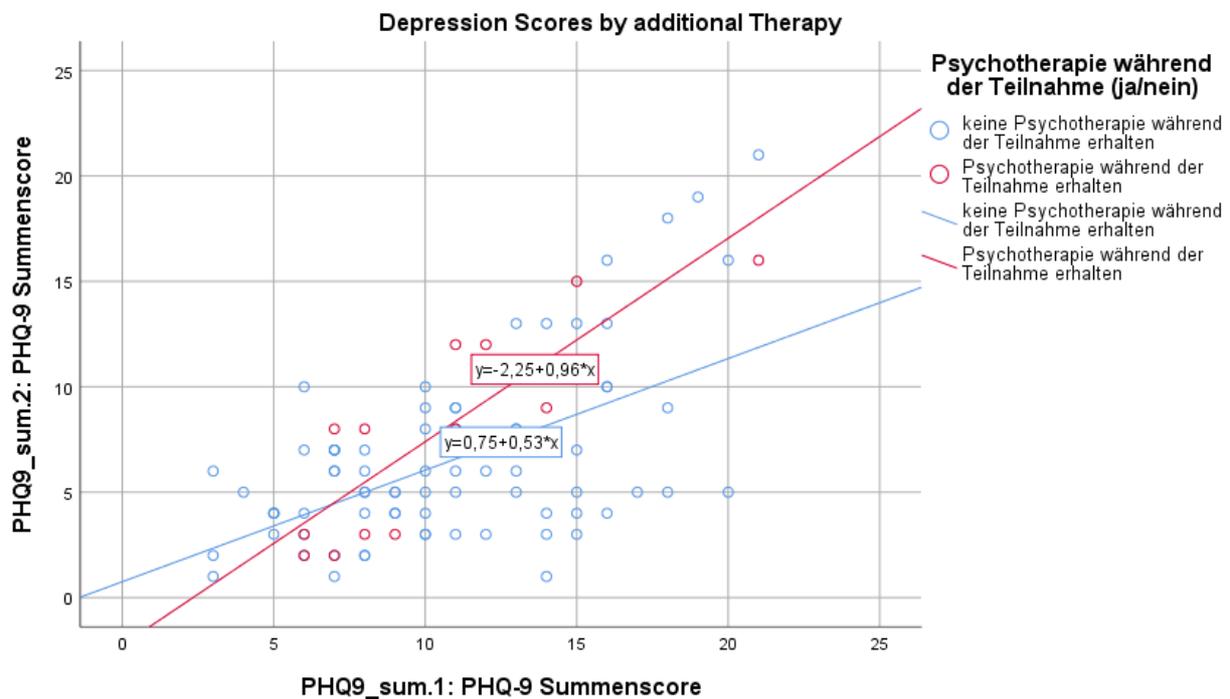


Abbildung 5: Baseline- und Post-Depressionswerte aufgeteilt nach Zusatzbehandlung ( $n = 89$ )

### t-Tests für abhängige Stichproben für den primären Endpunkt „Depression“ ohne Zusatzbehandlung

Alle Studienteilnehmer:innen, die während der Studienteilnahme eine zusätzliche Behandlung in Anspruch genommen hatten, wurden aus der Analyse ausgeschlossen: Dies betraf 13 Fälle. Im Durchschnitt wiesen die Studienteilnehmer:innen ohne Zusatzbehandlung ( $N = 76$ ) zum Post-Messzeitpunkt nach der Intervention niedrigere Depressions-Summenwerte auf ( $M = 6.51$ ,  $SE = 0.47$ ) als zum Baseline-Messzeitpunkt ( $M = 10.88$ ,  $SE = 0.51$ ). Dieser Unterschied,  $M_{diff} = -4.37$ ,  $BCa$  95% CI [-5.30, -3.44], war in einem t-Test für abhängige Stichproben signifikant  $t(75) = -9.39$ ,  $p < .001$ , und repräsentiert eine große Effektstärke,  $d = -0.98$ . Die Berechnung ergab eine negative Effektstärke, wobei niedrigere Werte bessere Ergebnisse auf dem primären Endpunkt repräsentieren. Der Interquartilsabstand auf der Skala des Messinstruments betrug zur Baseline 7 Punkte, bei einem Minimum von 3 und einem Maximum von 21 Punkten ( $R = 18$ ;  $SD = 4.46$ ), während er zur Post-Messung 4 Punkte betrug, bei einem Minimum von 1 und einem Maximum von 21 Punkten ( $R = 20$ ;  $SD = 4.20$ ).

Da eine Normalverteilung für den primären Endpunkt nicht sicher angenommen werden konnte, wurde das Ergebnis des t-Tests mit einem non-parametrischen Verfahren überprüft: Der Wilcoxon-Vorzeichen-Rang-Test bestätigte, dass sich die Depressionswerte der Studienteilnehmer:innen ohne Zusatzbehandlung zur Baseline-Messung ( $Mdn = 10.00$ ) und Postmessung ( $Mdn = 5.00$ ) signifikant voneinander unterschieden,  $z = -6.73$ ,  $p < .001$ .

### Ergebnis des t-Tests für abhängige Stichproben für den primären Endpunkt „Depression“ Per-Protokoll

Alle Studienteilnehmer:innen, die nicht mindestens das vorletzte (sechste) Moduls der *My7steps App* abgeschlossen hatten, wurden aus der Per-Protokoll-Analyse ausgeschlossen: Dies betraf neun Fälle. Im Durchschnitt wiesen die adhärennten Studienteilnehmer:innen ( $N = 80$ ) zum Post-Messzeitpunkt nach der Intervention niedrigere Depressions-Summenwerte auf ( $M = 5.90$ ,  $SE = 0.38$ ) als zum Baseline-Messzeitpunkt ( $M = 10.48$ ,  $SE = 0.47$ ). Dieser Unterschied,  $M_{diff} = -4.58$ , BCa 95% CI [-5.44, -3.71], war in einem t-Test für abhängige Stichproben signifikant  $t(79) = -10.58$ ,  $p < .001$ , und repräsentiert eine große Effektstärke,  $d = -1.09$ . Die Berechnung ergab eine negative Effektstärke, wobei niedrigere Werte bessere Ergebnisse auf dem primären Endpunkt repräsentieren. Der Interquartilsabstand auf der Skala des Messinstruments betrug zur Baseline 7 Punkte, bei einem Minimum von 3 und einem Maximum von 21 Punkten ( $R = 18$ ;  $SD = 4.21$ ), während er zur Post-Messung 5 Punkte betrug, bei einem Minimum von 1 und einem Maximum von 16 Punkten ( $R = 15$ ;  $SD = 3.37$ ).

Da eine Normalverteilung für den primären Endpunkt nicht sicher angenommen werden konnte, wurde das Ergebnis des t-Tests mit einem non-parametrischen Verfahren überprüft: Der Wilcoxon-Vorzeichen-Rang-Test bestätigte, dass sich die Depressionswerte der adhärennten Studienteilnehmer:innen zur Baseline-Messung ( $Mdn = 10.00$ ) und Postmessung ( $Mdn = 5.00$ ) signifikant voneinander unterschieden,  $z = -6.73$ ,  $p < .001$ .

In Tabelle 7 sind für beide Messzeitpunkte die Mittelwerte und Standardfehler bzgl. des primären Endpunkts für das PP- und das ITT-Verfahren gegenübergestellt. Ergänzend wurden die Mittelwertdifferenzen zwischen den Messzeitpunkten sowie die dazugehörigen Standardfehler und Effektstärken verzeichnet.

**Tabelle 7:** Mittelwerte und Effektstärken des primären Endpunkts über die Messzeitpunkte (PP vs. ITT)

	<i>Baseline</i>	<i>Post</i>	<i>Differenz</i>		
	<i>M (SE)</i>	<i>M (SE)</i>	<i>MD</i>	<i>SE</i>	<i>ES</i>
<b>PHQ-9 (PP, <math>N = 80</math>)</b>	10.48 (0.47)	5.90 (0.38)	-4.58	0.43	-1.09
<b>PHQ-9 (ITT, <math>N = 89</math>)</b>	10.81 (0.47)	6.70 (0.46)	-4.11	0.42	-0.93

*PHQ-9 = Patient Health Questionnaire; M = Mittelwert; SE = Standardfehler; MD = Mittelwertdifferenz; ES = Effektstärke (Cohen's d).*

### 9.3.5 Subgruppenanalysen

#### Einfaktorielle ANOVA mit Messwiederholungen unter Berücksichtigung der Indikationsgruppe

Um die Gültigkeit der Ergebnisse für den primären Endpunkt hinsichtlich der Indikationsgruppen zu überprüfen, wurde eine einfaktorielle ANOVA (Analysis of Variance) mit Messwiederholungen und der dreistufigen Diagnosegruppe als Faktor gerechnet. Da nur zwei Messzeitpunkte vorlagen, war die Annahme der Sphärizität gegeben und musste nicht überprüft werden. Der Boxsche M-Test bestätigte die Gleichheit der Kovarianzmatrizen, der Levene-Test die Gleichheit der Fehlervarianzen (alle  $ps > .05$ ).

Das Ergebnis der Analyse ( $N = 89$ ) bestätigte die primäre Wirksamkeitsanalyse ( $F(1, 86) = 89.62; p < .001$ ), während die Interaktion von Behandlung und Diagnosegruppe nicht signifikant wurde ( $p > .05$ ): Die Diagnosegruppen unterschieden sich demnach nicht signifikant hinsichtlich des primären Endpunkts und scheinen die in der Wirksamkeitsanalyse gefundene Reduktion der depressiven Symptomatik nicht beeinflusst zu haben. Auf *Post-Hoc*-Tests wurde dementsprechend verzichtet. Tabelle 8 zeigt die Mittelwerte und Standardabweichungen der Depressionswerte für beide Messzeitpunkte sowie die dazugehörigen Fallzahlen je Diagnosegruppe; in Tabelle 9 sind die Kennwerte der einfaktoriellen ANOVA verzeichnet.

**Tabelle 8:** Depressions-Mittelwerte und -Standardabweichungen sowie Gruppengrößen der Diagnosegruppen

Diagnosegruppe	Baseline		Post		n
	M	SD	M	SD	
Leichte/mittelschwere Depression (F32.0, F32.1)	11.25	4.18	7.22	4.05	32
Rez. leichte/mittelschwere Depression (F33.0, F33.1)	10.81	4.50	6.50	4.25	36
Anpassungsstörung (F43.2)	10.14	4.75	6.24	4.89	21
Gesamt	10.81	4.42	6.70	4.31	89

*M = Mittelwert; SD = Standardabweichung; n = Anzahl Studienteilnehmer:innen in der jeweiligen Diagnosekategorie.*

**Tabelle 9:** Kennwerte der einfaktoriellen ANOVA mit Messwiederholungen zur Überprüfung der Indikationsgruppen

	SS	df	MS	F	p	$\eta^2_p$
<b>Treatment</b>	702.58	1	702.58	89.62	.000	.51
<b>Treatment x Group</b>	1.23	2	0.61	0.78	.925	.00
<b>Error</b>	674.21	86	7.84			

*Treatment = Behandlungseffekt auf den primären Endpunkt (PHQ-9); Treatment x Group = Interaktion des Behandlungseffekts auf den primären Endpunkt und der ICD-10-Indikationsgruppen; Error = Variabilität innerhalb der Gruppen (nicht aufgeklärte, zufällige Fehlervarianz); SS = Summe der Quadrate; df = Freiheitsgrade; MS = Mittel der Quadrate; F = empirisch ermittelter F-Wert; p = p-Wert (Signifikanz);  $\eta^2_p$  = partielles Eta-Quadrat (Effektstärke).*

### Ergebnis des t-Tests für abhängige Stichproben für den primären Endpunkt bei leichten Depressionen

Alle Studienteilnehmer:innen, die zur Baseline mindestens eine klinisch signifikante Depressionsausprägung ( $PHQ-9 > 9$ ) aufwiesen, wurden aus der Analyse ausgeschlossen: Dies betraf 51 Fälle. Im Durchschnitt wiesen die Studienteilnehmer:innen mit subklinischer bis leichter Depression ( $N = 38$ ) zum Post-Messzeitpunkt nach der Intervention niedrigere Depressions-Summenwerte auf ( $M = 4.45$ ,  $SE = 0.36$ ) als zum Baseline-Messzeitpunkt ( $M = 6.71$ ,  $SE = 0.28$ ). Dieser Unterschied,  $M_{diff} = -2.26$ ,  $BCa$  95%

CI [-3.03, -1.40], war in einem t-Test für abhängige Stichproben signifikant  $t(37) = -5.46$ ,  $p < .001$ , und repräsentiert eine große Effektstärke,  $d = -1.34$ . Die Berechnung ergab eine negative Effektstärke, wobei niedrigere Werte bessere Ergebnisse auf dem primären Endpunkt repräsentieren. Der Interquartilsabstand auf der Skala des Messinstruments betrug zur Baseline 2 Punkte, bei einem Minimum von 3 und einem Maximum von 9 Punkten ( $R = 6$ ;  $SD = 1.69$ ), während er zur Post-Messung 3 Punkte betrug, bei einem Minimum von 1 und einem Maximum von 10 Punkten ( $R = 9$ ;  $SD = 2.19$ ).

Da eine Normalverteilung für den primären Endpunkt nicht sicher angenommen werden konnte, wurde das Ergebnis des t-Tests mit einem non-parametrischen Verfahren überprüft: Der Wilcoxon-Vorzeichen-Rang-Test bestätigte, dass sich die Depressionswerte der Studienteilnehmer:innen mit subklinischer bis leichter Depression zur Baseline-Messung ( $Mdn = 7.00$ ) und Postmessung ( $Mdn = 4.00$ ) signifikant voneinander unterschieden,  $z = -4.15$ ,  $p < .001$ .

### 9.3.6 Explorative Analysen

#### Behandlungsdauer

Die Exploration der selbst gewählten Behandlungsdauer der adhärennten Teilnehmer:innen ( $N = 80$ ) zeigte, dass sich die Studienteilnehmer:innen durchschnittlich fünf Wochen für die sieben Module der Intervention Zeit nahmen ( $M = 5.04$ ,  $SD = 2.91$ ). Die kürzeste Behandlungsdauer betrug 0.7 Wochen, die längste 14.9 Wochen. Die Überprüfung linearer Zusammenhänge ergab kein signifikantes Ergebnis: Die Behandlungsdauer wies weder Korrelationen mit Merkmalen der Stichprobe (Alter, Geschlecht, Bildung oder Baseline-Depression) auf noch mit den Change Scores der Endpunkte oder der Bewertung der App durch die Studienteilnehmer:innen (alle  $ps > .05$ ).

Eine quadratische Regression mit den Depressions-Change-Scores als Regressand sowie der einfachen und der quadrierten Variable Behandlungsdauer als Regressoren war ebenfalls nicht signifikant ( $F(2,77) = 2.51$ ;  $p > .05$ ; vgl. Tabelle 10 **Fehler! Verweisquelle konnte nicht gefunden werden.**). Das Modell verfügt mit einem  $R^2 = .06$  (korrigiertes  $R^2 = .04$ ) nach Cohen nur über eine geringe Anpassungsgüte [17]. Entsprechend scheint im Rahmen der vorliegenden Daten weder ein linearer noch ein quadratischer Zusammenhang von Behandlungsdauer und Wirksamkeit vorzuliegen.

**Tabelle 10:** Kennwerte der Regressions-ANOVA zum quadratischen Zusammenhang von Behandlungsdauer und Wirksamkeit

	<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>p</b>
<b>Regression</b>	0.89	2	0.45	2.51	.088
<b>Residuen</b>	13.69	77	0.18		
<b>Gesamt</b>	14.59	79			

*Abhängige Variable: Depressions Changescor; Prädiktoren: Konstante, Behandlungsdauer, quadrierte Behandlungsdauer; SS = Summe der Quadrate; df = Freiheitsgrade; MS = Mittel der Quadrate; F = empirisch ermittelter F-Wert; p = p-Wert (Signifikanz);  $\eta^2_p$  = partielles Eta-Quadrat (Effektstärke).*

Abbildung 6 zeigt die Behandlungsdauer und die entsprechenden Change-Scores des primären Endpunkts in einem Streudiagramm. Es wird deutlich, dass im Bereich der Langzeitanwender:innen eventuell nicht ausreichend Fälle vorliegen, um eine verlässliche Beurteilung des Zusammenhangs von Behandlungsdauer und Behandlungseffekt vorzunehmen. Auf Grundlage der vorliegenden Daten zeigt die Behandlungsdauer aber keinen Einfluss auf die Wirksamkeit.

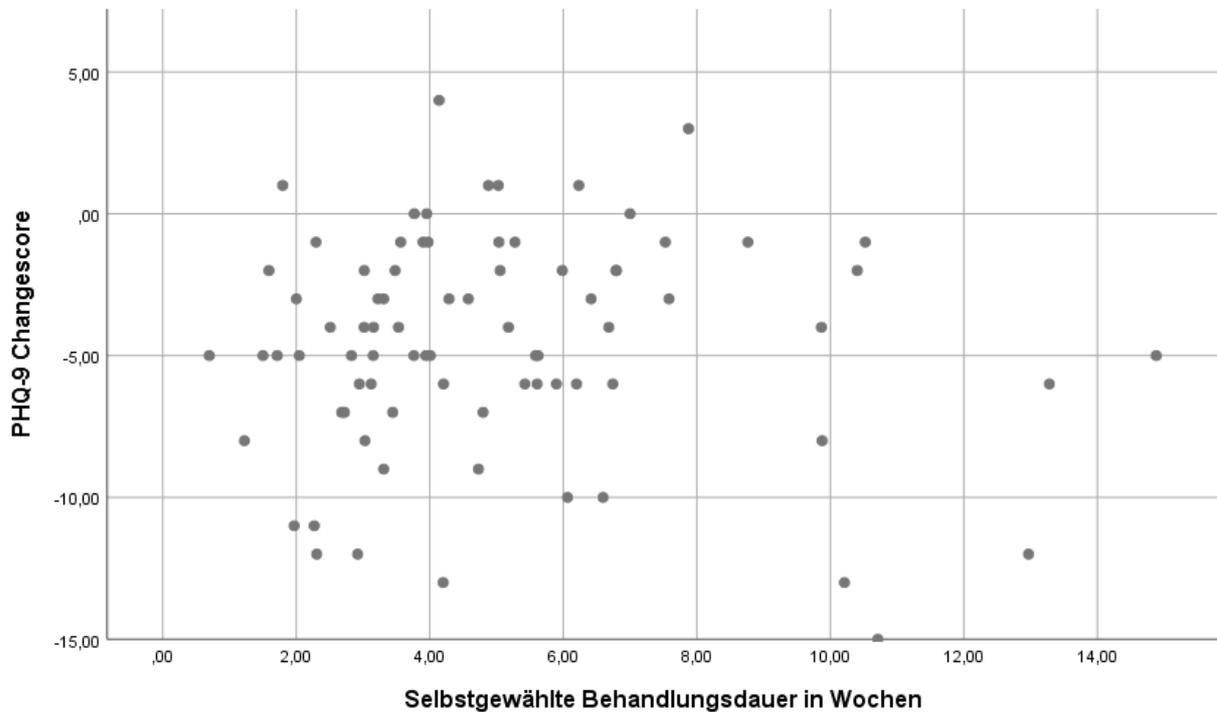


Abbildung 6: Streudiagramm Behandlungsdauer und Change-Scores des primären Endpunkts ( $n = 80$ )

## 10 Sicherheitsevaluation

### 10.1 Auswertung der Sicherheitsparameter

#### 10.1.1 Vorkommnisse und unerwünschte Ereignisse

Im Rahmen der Studie und im Nachbeobachtungszeitraum von vier Wochen nach der letzten Studienvisite wurden kein Vorkommnis und kein unerwünschtes Ereignis bzw. Adverse Events (AE) verzeichnet.

#### 10.1.2 Laufende Risikoevaluation

Die durch den Sponsor ausgeführte Risikoanalyse ergab keine als "nicht akzeptabel" bewerteten Restrisiken. Im Zeitraum der Studie kam es außerdem zu keiner Anpassung der im Studienprotokoll dargelegten Risikobewertung der Webanwendung als auch der Studie. Bei einer geringfügigen Veränderung wären die Informationen für Studienteilnehmer:innen sowie die Einwilligungserklärung aktualisiert und der Vorgang implementiert sowie dokumentiert worden. Dieser Fall ist nicht eingetreten.

#### 10.1.3 Studienteilnehmer:innen mit reliabler Verschlechterung

Zur weiteren Beurteilung von Sicherheitsaspekten wurde die Anzahl der Studienteilnehmer:innen als relevant definiert, die zum Post-Messzeitpunkt eine reliable Verschlechterung der depressiven Symptomatik (operationalisiert durch den PHQ-9) gemäß der in Abschnitt 7.9.4 unter *Response-Analysen* dargelegten Kriterien aufwies. Diese Sicherheitsanalyse wurde nicht (wie die Responseanalysen) auf Studienteilnehmer:innen mit einem klinischen Baseline-Depressionswert (PHQ-9 > 9) beschränkt, sondern für alle 89 Studienteilnehmer:innen ausgeführt. Eine reliable Verschlechterung lag nach dem Kriterium der *minimal clinically important difference* (RCI > 1.96) für einen Fall vor (1.3%), nach dem Response-Kriterium (Change Score des PHQ-9 > 4) für keinen Fall. Das Ausmaß der Verschlechterung des gefundenen Falls betrug RCI = 2.18, was einer absoluten Veränderung von 4 Punkten auf dem primären Endpunkt entspricht. Die Veränderung war somit etwas geringer als eine Symptome-Schwerekatgorie des Messinstruments.

#### 10.1.4 Bewertung der Intervention durch die Studienteilnehmer:innen

##### **Allgemeine Bewertung**

Nach der Behandlung wurden die Studienteilnehmer:innen gebeten, die Intervention zu bewerten: Die *My7steps App* wurde im Durchschnitt auf einer Skala von „sehr negativ“ (-2) bis „sehr positiv“ (2) als „positiv“ ( $M = 0.91$ ,  $SD = 0.43$ ) bewertet. Des Weiteren wurde die App auf einer Skala von „gar nicht hilfreich“ (-2) bis „sehr hilfreich“ (2) durchschnittlich als „eher hilfreich“ eingestuft ( $M = 0.85$ ,  $SD = 0.68$ ).

Gefragt, ob Veränderungen festgestellt wurden, die auf die Nutzung der App zurückzuführen sind, gaben die Studienteilnehmer:innen auf einer Skala von „keine Veränderungen“ (0) bis „sehr starke Veränderungen“ (4) durchschnittlich „einige persönliche Veränderungen“ an ( $M = 1.79$ ,  $SD = 0.93$ ). 68,4% (56/76) der Studienteilnehmer:innen gaben an, dass sie die *My7steps App* weiterempfehlen würden.

### **Negative Bewertungen**

Im Rahmen der Bewertung der Intervention gaben 6.3% (5/80) an, dass die App „wenig hilfreich“ gewesen sei und 8.8% (7/78) konnten keine auf die Intervention zurückzuführende Veränderung bei sich feststellen.

## **10.2 Sicherheitsschlussfolgerung**

Im Rahmen der Sicherheitsevaluation konnte kein Vorkommnis, kein unerwünschtes Ereignis, kein als „nicht akzeptabel“ bewertetes Restrisiko und keine Veränderung der Risikobewertung festgestellt werden. Die Anzahl der Studienteilnehmer:innen mit reliabler Verschlechterung (1.3%) und der Studienteilnehmer:innen, die die App abschließend negativ bewerteten (6.3%), konnten im Vergleich mit Zahlen anderen Untersuchungen im Bereich Psychotherapie und IMIs [19, 46, 55] als „unterdurchschnittlich“ eingeordnet und dementsprechend als „nicht auffällig“ klassifiziert werden.

## 11 Diskussion und Schlussfolgerungen

### 11.1 Wirksamkeits- und Sicherheitsergebnisse

Die vorliegende Studie ist eine erste Untersuchung zur Wirksamkeit der webbasierten Selbsthilfe-Intervention *My7steps App* für Personen mit leichten bis mittelgradig depressiven Episoden als auch für Personen mit rezidivierenden depressiven Störungen mit gegenwärtig leichten bis mittelgradigen Episoden und für Personen mit Anpassungsstörungen. Die Ergebnisse der Wirksamkeitsanalysen lassen den Schluss zu, dass die Intervention in Form einer Reduktion depressiver Symptomatiken und einer Verbesserung der Resilienz und Alltagsfunktionalität wirksam ist: Die depressiven Symptomatiken waren zum Post-Messzeitpunkt signifikant reduziert gegenüber der Baseline-Messung ( $d = -0.93$ ). Die Wirksamkeit zeigte sich ebenfalls in den sekundären Endpunkten durch eine erhöhte Resilienz ( $d = 0.73$ ), eine geringere Beeinträchtigung der Alltagsfunktionalität ( $d = -0.95$ ) sowie eine deutliche Reduktion der subjektiven Symptomstärken ( $d = -1.89$ ). In den primären Wirksamkeitsanalysen konnten somit für alle primären und sekundären Endpunkte große Effektstärken ermittelt werden – mit Ausnahme des sekundären Endpunkts Resilienz, für den sich eine mittlere Effektstärke ergab.

In den Sensitivitätsanalysen konnten keine generellen Wirksamkeitsunterschiede hinsichtlich Geschlecht, Bildung, Alter und Zusatzbehandlungen festgestellt werden. Die Sensitivitätsanalyse zur Überprüfung von möglichen Einflussvariablen bestätigten dabei die primären Wirksamkeitsanalysen: Die ermittelte Effektstärke auf den primären Endpunkt war ebenfalls groß – wenn auch etwas kleiner ( $d = 0.81$ ; nicht negativ, weil aus  $\eta^2_p$  berechnet) als die aus den primären Wirksamkeitsanalysen abgeleitete Effektstärke ( $d = -0.93$ ) – und stellt einen konservativeren Schätzer unter Berücksichtigung relevanter Einflussvariablen dar. Der Befund der Reduktion depressiver Symptomatiken durch die *My7steps App* bestätigte sich des Weiteren auch nach Ausschluss aller nicht adhärenenten Studienteilnehmer:innen in einer *Per-Protokoll*-Analyse mit einer höheren Effektstärke ( $d = -1.09$ ) sowie nach Ausschluss aller Studienteilnehmer:innen, die von einer Zusatzbehandlung während des Studienzeitraums profitiert hatten ( $d = -0.98$ ). Psychotherapeutische Zusatzbehandlungen scheinen dementsprechend keine geeigneten Alternativerklärung für die Befunde zu sein.

In den Subgruppenanalysen konnten weder generelle Wirksamkeitsunterschiede hinsichtlich der untersuchten ICD-10-Diagnosen noch für Studienteilnehmer:innen mit subklinischen bis leichten Depressionen festgestellt werden: Eine Überprüfung der verschiedenen Indikationsgruppen hinsichtlich des Behandlungseffekts auf den primären Endpunkt ergab keine signifikanten Unterschiede für diese Gruppen, sodass auf Basis der vorliegenden Daten angenommen werden kann, dass die Behandlung für alle fünf untersuchten Indikation (F32.0, F32.1, F33.0, F33.1, F43.2) gleichermaßen wirksam ist. Die gesonderte Betrachtung der Studienteilnehmer:innen mit subklinischer bis leichter Depression ergab, wie in der primären Wirksamkeitsanalyse ermittelt, eine große Effektstärke der Intervention ( $d = -1.34$ ), sodass auch für diese weniger schweren Fälle von einer deutlichen Reduktion der depressiven Symptomatik ausgegangen werden kann.

Eine explorative Untersuchung der selbstgewählten Behandlungsdauer, der Stichprobenmerkmale und der Wirksamkeit konnte dagegen weder signifikante Zusammenhänge noch erkennbare Verlaufsmuster aufzeigen, die Behandlungsdauer zeigte keinen Einfluss auf die Wirksamkeit. Dementsprechend können aus der vorliegenden Untersuchung keine neuen Empfehlungen hinsichtlich einer optimalen Behandlungsdauer abgeleitet werden. Die selbstgewählte Behandlungsdauer variierte in der untersuchten Stichprobe erwartungsgemäß stark (von 0.7 Wochen bis 14.9 Wochen), war aber auch in den Extrembereichen nicht mit einer signifikanten Verringerung oder Zunahme des Behandlungseffekts

assoziiert. Die selbstbestimmte Wahl der Behandlungsdauer durch die Studienteilnehmer:innen scheint daher aus einer Behandlungsperspektive unkritisch. Trotz des potenziellen Vorteils einer bedarfsgerechteren „Passung“ für die Studienteilnehmer durch diesen Ansatz, sollte die Varianz der Behandlungsdauer bei zukünftigen, kontrollierten Designs standardisiert werden, um eine bessere Vergleichbarkeit zu erreichen.

Im Rahmen der Sicherheitsevaluation konnten keine besonderen Risiken oder Belastungen festgestellt werden. Die subjektive Evaluation der Behandlung durch die Studienteilnehmer:innen unterstützt diese Befundlage und die der Wirksamkeitsanalysen: Die Intervention wurde durch die Teilnehmenden durchschnittlich als „positiv“ bewertet, und die Studienteilnehmer:innen attribuierten im Mittel „einige persönliche Veränderungen“ auf die Behandlung.

Zusammenfassend lässt sich festhalten: Die webbasierte Selbsthilfe-Intervention *My7steps App* scheint nach den vorliegenden Daten eine wirksame Intervention bei der Behandlung von leichten bis mittelgradigen depressiven Symptomatiken und Anpassungsstörungen zu sein. Damit hat die Intervention das Potential, depressive Symptomatiken zeitnah nach der Diagnose erfolgreich zu behandeln und die erste Stufe eines "stepped care"-Ansatzes zu bilden.

## 11.2 Verhältnis von Risiken und Nutzen

Aufgrund der Daten der vorliegenden Studie lässt sich festgehalten, dass die Behandlung mit der webbasierten Selbsthilfe-Intervention *My7steps App* höchstwahrscheinlich wirksam ist und bei Personen mit leichten bis mittelgradig depressiven Episoden, rezidivierenden depressiven Störungen und Anpassungsstörungen mittlere bis große Effektstärken aufweist, welche neben der depressiven Symptomatik auch die Resilienz und Alltagsfunktionalität der Betroffenen verbessern können. Von einem deutlichen gesundheitlichen Nutzen der Intervention kann somit ausgegangen werden.

Im Rahmen der Sicherheitsevaluation konnte kein Vorkommnis, kein unerwünschtes Ereignis, kein als "nicht akzeptabel" bewertetes Restrisiko und keine Veränderung der Risikobewertung während der Studie festgestellt werden. Die Anzahl der Studienteilnehmer:innen mit reliabler Verschlechterung (0%, wenn nur Studienteilnehmer:innen mit klinisch signifikanten Depressionsausprägung zur Baseline betrachtet werden; 1.1% bei Betrachtung der gesamten Stichprobe) und der Studienteilnehmer:innen, die die App abschließend negativ bewerteten (6.3%), war darüber hinaus im Vergleich mit anderen Untersuchungen gering. Auch in den Analysen des literaturbasierten Kontrollarms (vgl. Kapitel 12) konnten keine Symptomverschlechterungen im Vergleich mit den externen Kontrollgruppen und den damit einbezogenen Krankheitsverläufen festgestellt werden.

Insgesamt kann das Nutzen-Risiko Verhältnis der Intervention als positiv bewerten werden.

## 11.3 Einordnung der Ergebnisse in den wissenschaftlichen Kontext

Es wurde gezeigt, dass die Behandlung mit der *My7steps App* bei der Reduktion depressiver Symptomatiken für alle fünf untersuchten Indikationen (F32.0, F32.1, F33.0, F33.1, F43.2) höchstwahrscheinlich gleichermaßen wirksam ist und die Intervention eine große Effektstärke von  $d = 0.81$  bis  $d = 0.93$  auf den primären Endpunkt aufweist. Die in der vorliegenden Studie ermittelte

Effektstärke ist damit – je nach Schätzer – höher bzw. deutlich höher als jene mittlere Gesamteffektstärke, die in einer aktuellen Metaanalyse bezüglich digitaler Interventionen zur Behandlung von Depressionen berichtet wurde (dort angegeben als Hedge's  $g$ , welches hier direkt mit Cohen's  $d$  vergleichbar ist:  $g = 0.52$ ) und auf 83 Studien mit über 15 000 Studienteilnehmer:innen basiert [73]. Die gefundene Effektstärke ist darüber hinaus höher bzw. deutlich höher als die in einer anderen Metaanalyse berichtete Wirksamkeit computergestützter Formen der kognitiven Verhaltenstherapie bei schweren depressiven Störungen, für die eine kleine bis mittlere Gesamteffektstärke ( $g = 0.23$  für unbegleitete Interventionen;  $g = 0.50$  gesamt) ermittelt wurde [100] sowie höher als die in einer weiteren Metaanalyse publizierte mittlere Effektstärke ( $d = 0.69$ ) für allgemeine, psychologische Behandlungen von Depressionen [22].

Im direkten Vergleich mit Studien zu ähnlichen Programmen aus dem Bereich der digitalen Gesundheitsapplikationen und Depressionstherapie sind die in der vorliegenden Studie ermittelten within-group Effektstärken ( $d = 0.81$  und  $d = 0.93$ ) in drei Fällen höher ( $d = 0.39$ ,  $d = 0.55$  und  $d = 0.57$ , between-group Effektstärken) [10, 48, 71], in zwei Fällen vergleichbar ( $d = 0.89$  und  $d = 0.94$ , between-group Effektstärken) [6, 74] und in einem Fall kleiner ( $d = 1.38$ , within-group Effektstärke) [52] als die publizierten Effektstärken anderer KVT-basierter Online-Anwendungen auf dem deutschen Markt.

Der gefundene Effekt von  $d = 0.81$  bis  $d = 0.93$  liegt über der ursprünglich vom *National Institute of Clinical Excellence* (NICE, UK) vorgeschlagenen standardisierten mittleren Differenz (SMD oder Cohen's  $d$ ) von 0.50 zur Bestimmung klinischer Relevanz und damit ebenfalls deutlich über dem durch eine Forschergruppe später korrigierten Grenzwert von SMD = 0.24 [21].

Zusammenfassend lässt sich festhalten, dass die in der vorliegenden Studie gefundenen Effektstärken im Bereich vergleichbarer Studien liegen und teilweise höher ausfallen als die für andere Medizinprodukte auf dem deutschen Markt berichteten Wirksamkeiten.

## 11.4 Klinische Relevanz der Ergebnisse

Die in der vorliegenden Studie ermittelte klinische Response-Rate von 60.8% liegt im Bereich der Vergleichswerte einer Meta-Analyse, die Response-Raten von 59% bis 68% für Psychotherapien berichtet [19] und entspricht bzw. übersteigt auch die für vergleichbare Medizinprodukte berichteten Raten von 35.6% [48], 38.2% bis 53% [71], 43.3% [52], 63% [10] bzw. 62.5% bis 65% [6, 74]. Während eine partielle Remission für 54.9% der hier untersuchten Stichprobe festgestellt werden konnte, erreichten nur 21.6% vollständige Remission, was unter den Referenzwerten für Psychotherapien von 26% bis 34% liegt [19]. Wenn Remission aber ähnlich der hier verwendeten partiellen Remission definiert wurde, übersteigt der vorliegende Wert von 54.9% die berichteten 50% (identisch definiert als 5-Punkte-PHQ-9-Reduktion und Post-Depressionsscore unter 10, ermittelt an der *Per-Protokoll*-Stichprobe) [71], 28% (definiert als BDI-II-Score von 12 oder weniger) [52] bzw. 22.7% (definiert als Reduktion um  $\geq 8.99$  Punkte auf der ADS-Skala und einem Gesamtscore unter 16) [6] vergleichbarer KVT-basierter Online-Anwendungen.

Die ermittelten Effekte der *My7steps App* auf depressive Symptomatiken können wie dargelegt als klinisch relevant eingestuft werden. Insgesamt lässt sich festhalten, dass die Intervention eine geeignete Behandlungsalternative für Betroffene mit leichten bis mittelgradigen depressiven Episoden, rezidivierenden depressiven Störungen und Anpassungsstörungen zu sein scheint.

## 11.5 Implikationen für die Durchführung zukünftiger Studien

Die vorliegende Studie ist eine erste Untersuchung zur Wirksamkeit der webbasierten Selbsthilfe-Intervention *My7steps App* und damit Ausgangspunkt für eine differenziertere Evaluation und genauere Überprüfung der Intervention. Diese weitere Evaluation zur Bestätigung der Ergebnisse dieser Vorstudie soll im Rahmen der Erprobungsstudie im Rahmen des DiGA-Antragsverfahren mit einer randomisierten kontrollierten Studie, einer entsprechend größeren Fallzahl und einem konfirmatorischen Nachweis erfolgen.

Die Integration in die Regelversorgung muss aufgrund der noch überschaubaren Befundlage weiter forschungstechnisch begleitet werden. Zukünftige Studien sollten Beiträge zur Behandlungssteuerung liefern als auch Behandlungsempfehlungen entwickeln und überprüfen. Hierbei sind vor allem soziostrukturelle und individuelle Einflussvariablen, Nutzungsparameter und Zielgruppenmerkmale relevant. Zu den relevanten Forschungsfragen gehört weiterhin die differenzierte Untersuchung möglicher Einflüsse von Alter, Bildung und Geschlecht, die Untersuchung von verschiedenen Nutzungsstilen und Behandlungsverläufen als auch die Identifikation von eventuell behandlungsrelevanten Subgruppen der aus fünf diagnostischen Kategorien stammenden Zielgruppe. Diese Fragen sollten im Rahmen des konfirmatorischen Nachweises der positiven Versorgungseffekte vertieft werden.

Eine Herausforderung wird die selbstbestimmte Behandlungsdauer für die Konzeption eines Kontrollgruppendesigns darstellen. Trotz des potenziellen Vorteils einer bedarfsgerechteren „Passung“ für die Studienteilnehmer durch diesen Ansatz, sollte die Varianz der Behandlungsdauer bei zukünftigen, kontrollierten Designs standardisiert werden, um eine bessere Vergleichbarkeit zu erreichen.

## 11.6 Einschränkungen der Studien

Eine Einschränkung der vorliegenden Studie ist das quasi-experimentelle Design und Fehlen einer Kontrollgruppe: Alle inferenzstatistisch basierten Gruppenunterschiede und Effektstärken konnten nur unkontrolliert ermittelt werden. Um dieser Einschränkung zu begegnen, wurde ein literaturbasierter Kontrollarm ergänzt, in dem die *My7steps App* Interventionsgruppe vergleichbaren Kontrollgruppen gegenübergestellt wurde (vgl. Kapitel 12).

Mit Blick auf die untersuchte Stichprobe fällt vor allem das hohe Bildungsniveau auf (70% mit Universitätsabschluss), aber auch die erhöhte Anzahl weiblicher Studienteilnehmerinnen (ca. 74%). Beide Stichprobenmerkmale können die Generalisierbarkeit der Ergebnisse einschränken, gleichfalls aber auch Ausdruck einer Zielgruppe sein, die zunehmend an digitalen, selbstadministrierten Interventionen interessiert ist. Eine erhöhte Anzahl weiblicher Studienteilnehmerinnen ist in Studien zur Behandlung von Depression zudem keine Seltenheit, und sowohl die Variable Bildung als auch Geschlecht wurden in der Sensitivitätsanalyse zur Untersuchung möglicher Drittvariablen berücksichtigt, ohne dass ein signifikanter Einfluss dieser Variablen festgestellt werden konnte.

Die für die Intervention empfohlene Anwendungsdauer beträgt drei Monate, um ausreichend Zeit für die Integration persönlicher Verhaltensänderungen in das eigene Leben zu gewähren. Von dieser Empfehlung wurde im Rahmen der Studie abgewichen, um einer übermäßig starken Varianz der Behandlungsdauer vorzubeugen. Als Folge dieser studienbedingt verkürzten Behandlungsempfehlung, die einige

Studienteilnehmer:innen nicht von einer längeren Nutzung abhielt, ist eine geringere Integration persönlicher Verhaltensänderungen nicht auszuschließen, was in der Konsequenz zu einer reduzierten Wirksamkeit in der vorliegenden Untersuchung geführt haben könnte. Da kein linearer oder quadratischer Einfluss der Behandlungsdauer auf die Wirksamkeit festgestellt werden konnte, scheint dies für den Befund der Wirksamkeit aber nicht relevant. In der folgenden Studie sollte auf die vollumfängliche Untersuchung der empfohlenen Anwendungsdauer von drei Monaten geachtet und der Behandlungszeitraum standardisiert werden, auch um die Möglichkeit einer Reduktion und Unterschätzung der Wirksamkeit aufgrund verkürzter Behandlungszeiträume auszuschließen.

Da es sich bei der vorliegenden Untersuchung um eine Vorstudie handelt, deren Stichprobengröße auf einer erfahrungsbegründeten Fallzahlschätzung basierte, sind die gefundenen Ergebnisse nicht vollständig belastbar. Für die kontrollierte Erprobungsstudie sollte eine entsprechende powerbasierte Fallzahlkalkulation durchgeführt werden, um die Ergebnisse auf Grundlage der anzunehmenden Effektstärken und des abzuleitenden Stichprobenumfangs überprüfen und bestätigen zu können.

## **Fazit**

Die vorliegende Studie ist eine erste Untersuchung zur Wirksamkeit der webbasierte Selbsthilfe-Intervention *My7steps App* für Personen mit leichten bis mittelgradig depressiven Episoden als auch für Personen mit rezidivierenden depressiven Störungen mit gegenwärtig leichten bis mittelgradigen Episoden und für Personen mit Anpassungsstörungen. Die Ergebnisse der Wirksamkeitsanalysen lassen den Schluss zu, dass die Intervention in Form einer Reduktion depressiver Symptomatiken und einer Verbesserung der Resilienz und Alltagsfunktionalität wirksam ist. So zeigen alle Endpunkte nach Ende des Interventionszeitraums signifikante Verbesserungen im Vergleich zur Baseline-Erhebung mit Effektstärken im mittleren bis hohen Bereich. Eine Einordnung in den aktuellen wissenschaftlichen Kontext und eine Gegenüberstellung der Ergebnisse mit den Befunden vergleichbarer Studien ergab, dass die in der vorliegenden Studie gefundenen Ergebnisse auf eine den Befunden entsprechende und teilweise höher ausfallende Wirksamkeit hindeuten als die anderer Medizinprodukte auf dem deutschen Markt. Auch die Untersuchung der klinischen Relevanz unterstützt diese Schlussfolgerung: Die in der Studie ermittelten Response-Raten sowie Raten partieller und vollständiger Remission erreichen einschlägige Referenzwerten für Psychotherapien, vergleichbare Medizinprodukte und KVT-basierter Online-Anwendungen bzw. übertreffen diese teilweise. Folglich konnte in der Studie gezeigt werden, dass die *My7steps App* eine wirksame Intervention für Patienten mit Depressionen und Anpassungsstörungen ist. Dementsprechend kann auch davon ausgegangen werden, dass in der geplanten Erprobungsstudie durch die 12-wöchige Anwendung der *My7steps App* eine Verbesserung der depressiven Symptomatik und damit der angestrebte Versorgungseffekt erreicht werden kann.

## 12 Analysen mit externen Kontrollgruppen (literaturbasierter Kontrollarm)

Ein Prä-Post-Vergleich ist geeignet einen ersten, ökonomischen Wirksamkeitsnachweis zu erbringen und daher ein angemessenes Design im Rahmen einer Pilotstudie. Der Vergleich ohne Kontrollgruppe ist allerdings mit methodischen Limitationen behaftet, da der natürliche Krankheitsverlauf und unbekannte Confounder nur bedingt herausgerechnet werden können. Um die Ergebnisse zu plausibilisieren und den Limitationen zu begegnen, wurden die Ergebnisse der Studienteilnehmer:innen der *My7steps App* mit aktuellen kontrollierten Studien unterschiedlicher Studiendesigns kontrastiert. Das methodische Vorgehen und die Ergebnisse werden in den folgenden Abschnitten dargestellt und abschließend diskutiert.

### 12.1 Methodisches Vorgehen zur Erstellung des literaturbasierten Kontrollarms

Bei der Suche von Studien, deren Kontrollgruppen für einen Vergleich mit der vorliegenden Interventionsgruppe geeignet sind, wurde auf verschiedene Merkmale geachtet, um eine möglichst valide Aussagekraft zu gewährleisten. Die Kontrollgruppen der ausgewählten Studien sollten hinsichtlich unterschiedlicher Merkmale eine Vergleichbarkeit mit der vorliegenden Studienteilnehmer:innengruppe aufweisen, beispielsweise hinsichtlich soziodemographischer Einflussfaktoren wie Bildung, Alter, Geschlecht und Familienstand, aber auch in Bezug auf den Ausgangsschweregrad der Depressivität. Darüber hinaus wurden aktuelle und in Deutschland durchgeführte Studien bevorzugt, um Kohorten- sowie Umwelteffekte möglichst zu harmonisieren.

Zur Berechnung der Effektstärke wurde die sogenannte *Differenz des standardized mean change für Experimental- und Kontrollgruppen* [72] herangezogen, bei welcher die Mittelwerte und Standardabweichungen der Experimental- und Kontrollgruppen zum Baseline- und Postzeitpunkt, sowie die jeweilige Stichprobengröße mit einfließen (s. Formeln):

$$ES_{sm} = \frac{(\bar{X}_{G1post} - \bar{X}_{G2post}) - (\bar{X}_{G1pre} - \bar{X}_{G2pre})}{S_{pooled\ pre}}$$

$$S_{pooled\ pre} = \sqrt{\frac{(n_{G1} - 1)s_{G1pre}^2 + (n_{G2} - 1)s_{G2pre}^2}{n_{G1} + n_{G2} - 2}}$$

In einem ersten Schritt fand eine systematische Literatursuche in der Datenbank PsycInfo statt. Hierfür wurde anhand einer Suchsyntax nach Studien gesucht, die zum einen ebenfalls den PHQ-9 zur Messung der Depressivität verwendeten, zum anderen eine Form des Kontrollgruppensdesigns nutzten, wie etwa Wartelisten, Care- beziehungsweise Treatment-as-Usual. Die Suche wurde auf Studien beschränkt, die in den letzten fünf Jahren bzw. ab 2018 veröffentlicht wurden (eine der untersuchten Studien wies eine Vorveröffentlichung von 2017 auf, die aufgrund fehlender Informationen in der bereits eingeschlossenen Publikation ebenso herangezogen wurde).

Folgende Suchsyntax wurde zur Literatursuche herangezogen:

((phq9 or phq-9 or "Brief Patient Health Questionnaire")) AND ((control\* OR CG OR "compar\* group\*" OR "compar\* trial\*" OR "compar\* stud\*" OR "between\* group\*" OR rct OR random\* OR "treatment as usual" OR tau OR waitlist\* OR "attention control\*" OR "match\* sampling" OR "match\* control\*" OR "clinical trial\*" OR "experiment\* control\*" OR "random\* assign\*" OR "quasi experiment\*"))

Zum Zeitpunkt der Suche wurden durch die Suchsyntax 505 Studien identifiziert. Dieser Pool wurde mittels Abstract-Screening nach Studien selektiert, deren Kontrollgruppen hinsichtlich der oben beschriebenen Variablen eine potenzielle Vergleichbarkeit mit den Studienteilnehmer:innen der vorliegenden Studie aufwiesen. Neben der beschriebenen Literatursuche mittels Suchsyntax wurde ergänzend das DiGA-Verzeichnis des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM) nach vergleichbaren Studien durchsucht; dadurch konnte eine weitere Studie aus 2017 als vergleichsgeeignet identifiziert werden. Eine Übersicht der systematischen Suche ist in Abbildung 7 dargestellt.

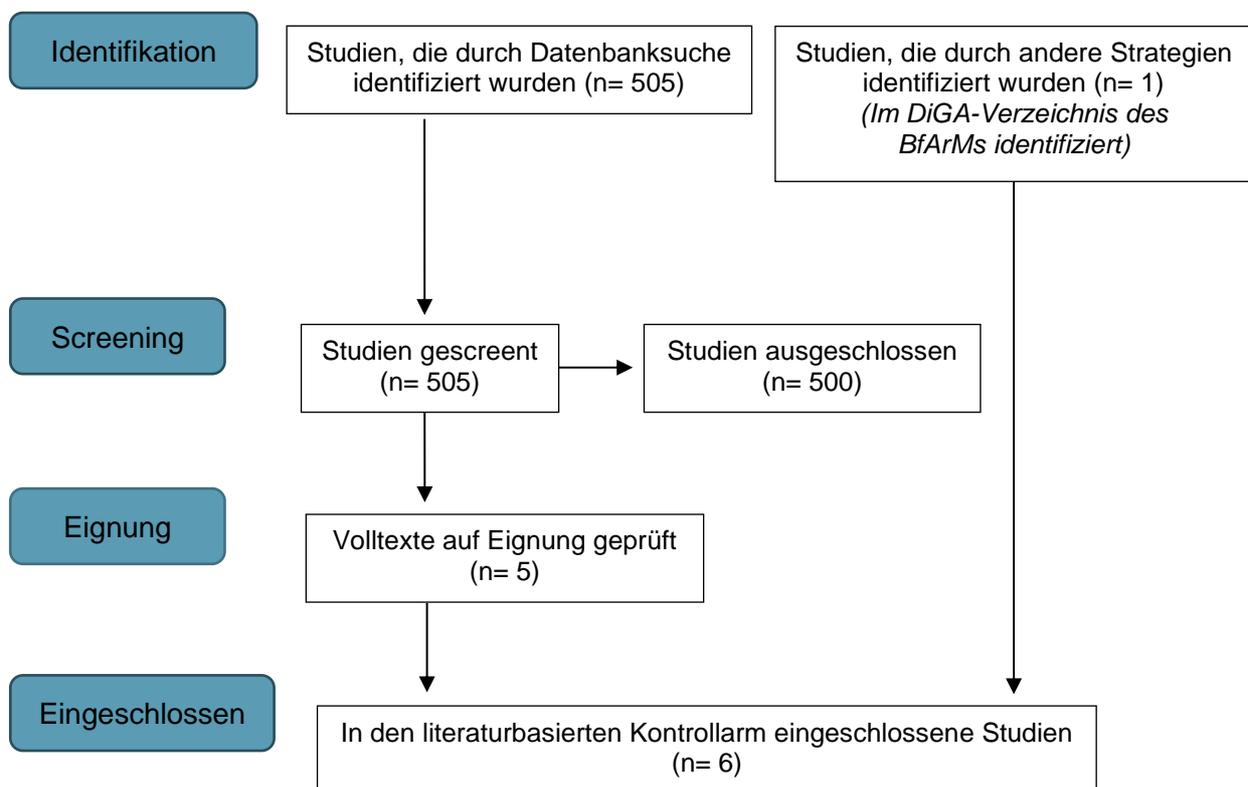


Abbildung 7: Methodik der systematischen Identifikation von Studien, die für den literaturbasierten Kontrollarm geeignet sind

Da es sich bei den Analysen mit externen Kontrollgruppen um Daten aus publizierten Studien handelt und um keine Berechnung mit einem personenbezogenen Datensatz, konnte die Vergleichbarkeit zwischen den gefundenen Kontrollgruppen und den Studienteilnehmer:innen der vorliegenden Studie nicht inferenzstatistisch untersucht werden, etwa durch eine ANOVA oder einen t-Test für unabhängige Stichproben. Entsprechend wurden mehrere Studien herangezogen, um anhand der ermittelten

Effektstärken einen umfassenden Eindruck des hypothetischen Effekts der *My7steps App* unter kontrollierten Studienbedingungen zu ermöglichen.

## 12.2 Ergebnisse des literaturbasierten Kontrollarms

### Beschreibung der gefundenen Studien

Insgesamt flossen sechs externe Publikationen zur Behandlung depressiver Erkrankungen in die literaturbasierte Kontrollgruppenanalyse ein. Eine Übersicht der analysierten Studien kann Tabelle 11 entnommen werden. Bis auf eine Studie aus Kanada [63] wurden alle eingeschlossenen Studien in Deutschland durchgeführt [10, 12, 38, 50, 62]. Fünf der sechs Studien untersuchten eine Form der web-basierten Intervention [10, 12, 50, 62, 63], eine Studie untersuchte die Wirksamkeit eines neuen Behandlungsansatzes für Ärzte in Allgemeinarztpraxen zur Betreuung von Menschen mit depressiven Erkrankungen [38]. Bei der Publikation von Lüdtke et al. [62] handelt es sich um die Pilotierung einer Intervention, die darauffolgend von Bruhns und Kollegen mit einer größeren, unabhängigen Stichprobe repliziert wurde [12] – beide Studien wurden berücksichtigt. Die Dauer der Intervention reichte in den gefundenen Studien von vier Wochen [12, 62] über eine bedarfsorientierte Dauer [38] bis hin zu drei Monaten [50, 63]. Somit liegt die durchschnittliche Interventionsdauer der vorliegenden Untersuchung in der Bandbreite der hier betrachteten Studien. Bis auf eine Intervention, deren Teilnahme bereits für Personen ab 16 Jahren möglich war [63], waren alle weiteren Interventionen für Erwachsene ab 18 Jahren zugelassen. Die Studie von Beiwinkel et al. [10] machte hierzu keine Angaben, jedoch kann anhand der Informationen im DiGA-Verzeichnis davon ausgegangen werden, dass auch hier ein Mindestalter von 18 Jahren vorlag (vgl. <https://diga.bfarm.de/de/verzeichnis/1110>). In Bezug auf die dargestellten Ein- und Ausschlusskriterien sei zu erwähnen, dass in der Darstellung aus Gründen der Übersichtlichkeit auf Vollständigkeit verzichtet wurde und Einschlusskriterien wie etwa einen Zugang zum Internet zu haben oder die Bereitschaft an der Studie teilzunehmen nicht extra aufgeführt wurden. Stattdessen wurde die Darstellung der Kriterien auf die Vergleichbarkeit mit der *My7steps App* beschränkt (in Tabelle 11 als „zentrale Einschlusskriterien“ bzw. „zentrale Ausschlusskriterien“ bezeichnet). Allen Studien ist gemein, dass sie für die Allokation der Studienteilnehmer:innen eine Form der Randomisierung verwendeten. Die Stichprobengröße unterschied sich zwischen den Studien dagegen deutlich: Die untersuchten Stichproben reichten von rund 100 Studienteilnehmer:innen [62, 63], bis hin zu Stichprobengrößen von 1000 Studienteilnehmer:innen [50]. Die Autoren der Studien nutzten für die Beschreibung ihrer Kontrollgruppendesigns unterschiedliche Terminologien, die aber alle eine Form des *Care-as-Usual* darstellten, was in den folgenden Abschnitten näher ausgeführt werden soll.

### Beschreibung der Kontrollgruppen-Designs der gefundenen Studien

Laut der Deklaration von Helsinki [98], die ethische Standards für die Forschung am Menschen festlegt, gilt in Bezug auf die Untersuchung neuer Interventionen folgender Grundsatz: „Nutzen, Risiken, Belastungen und Wirksamkeit einer neuen Maßnahme müssen mit denjenigen der nachweislich besten Maßnahme(n) verglichen werden.“ (S.8). Daraus leitet sich für die medizinische Forschung die Verwendung sogenannter *Care-as-Usual* Designs ab, in welchen Studienteilnehmer:innen der Kontrollgruppe, die die zu untersuchende Intervention nicht erhalten, eine oft individuelle „Standard“-Behandlung bekommen. Dem gegenüber steht das sogenannte *Treatment-as-usual* Design, bei welchem

die gesamte Kontrollgruppe eine spezifische, meist bereits etablierte Parallelbehandlung erhält, beispielsweise in Form einer etablierten Gruppentherapie, mit der die neue Intervention kontrastiert werden soll – aber auch hier werden die Begrifflichkeiten teilweise synonym verwendet [33]. Im *Care-as-usual* Design ist oft nicht eindeutig definiert, was die „nachweislich beste Maßnahme“ umfasst, was zu einer sehr heterogenen Verwendung der Terminologie führt, mitunter auch zu unterschiedlichen Begriffsverwendungen wie „Wartelistenkontrollgruppe“, *Individual-Care-as-Usual* oder *Usual-Care* [91]. Hinzu kommt, dass sich *Care-as-usual* anders als *Treatment-as-usual* von Studienteilnehmer:in zu Studienteilnehmer:in innerhalb der Kontrollgruppe unterscheiden kann, je nachdem, was die Person als Behandlungsoption wählt. So kann es innerhalb der Kontrollgruppe zu großer Heterogenität kommen, die in der Forschung schwer konstant gehalten und somit kontrolliert werden kann, die allerdings oft die realen Gegebenheiten widerspiegelt, die außerhalb experimenteller Laborbedingungen und im Rahmen der Patientenversorgung herrschen. Dies zeigen auch die vorliegenden Analysen der ausgewählten Studien, deren Kontrollgruppendesign im Folgenden näher ausgeführt werden soll.

Klein et al. [50] nutzten ein *Care-as-usual* Design mit Warteliste, in welchem die Kontroll-Studienteilnehmer:innen jegliche Form von Zusatzbehandlung nutzen durften, inklusive Medikation und Psychotherapie, wobei in der Studie nicht aufgeführt wurde, welche Kontroll-Studienteilnehmer:innen welche Behandlungsoptionen nutzten. Erst im Anschluss an die Posterhebung erhielten die Kontroll-Studienteilnehmer:innen Zugang zur Intervention.

Für die Evaluation der Intervention von Beiwinkel et al. [10] wurde ebenso eine *Care-as-usual* Kontrollgruppe mit Warteliste verwendet. Zusätzlich erhielten alle Kontroll-Studienteilnehmer:innen eine text-basierte Psychoedukation zum Thema Depression. Laut der Stichprobenbeschreibung erhielten 45.00% eine Form von Medikation gegen ihre depressive Erkrankung, 20.00% waren zum Zeitpunkt der Intervention in psychotherapeutischer Behandlung.

In der Untersuchung von Lütke et al. [62] wurde ein Wartelisten-Design verwendet, das bei näherer Betrachtung ebenfalls als *Care-as-Usual* bezeichnet werden kann, da sich in den soziodemographischen Angaben zeigt, dass 50.00% der Studienteilnehmer:innen im Kontrollarm Psychopharmaka-Medikation verwendeten und weitere 38.64% in therapeutischer Behandlung waren.

Bruhns et al. [12] verglichen die Studienteilnehmer:innen der Interventionsgruppe mit einer Wartelistenkontrollgruppe, ohne diese näher zu definieren. Anhand der soziodemographischen Informationen und Diskussionspunkte kann aber abgeleitet werden, dass Zusatzbehandlungen in der Warteliste, wie etwa die Verwendung von Antidepressiva (8.50%) oder Erfahrung in Psychotherapie (45.50%) sowie die Nutzung von Hilfsdienstleistungen, eine Rolle in der Symptomreduktion der Kontroll-Studienteilnehmer:innen gespielt haben dürfte.

Härter und Kollegen [38] akquirierten ihre Studienteilnehmer:innen über Allgemeinarztpraxen, sodass die Studienteilnehmer:innen der Kontrollgruppe eine Behandlung durch den Allgemeinarzt erhielten, welche auch Überweisungen in Psychotherapie und ambulante oder stationäre Psychiatrie beinhalten konnte. Die Autoren sprechen in diesem Zusammenhang von *Treatment-as-usual*, da sich die Zusatzbehandlung aber von Patient zu Patient unterscheiden kann, ist auch hier eher von einem *Care-as-Usual* Design auszugehen. Genaue Angaben über die genutzten Angebote in der Kontrollgruppe gibt es auch in dieser Untersuchung nicht.

MacLean und Kollegen [63] nutzen für ihre Untersuchung ein sogenanntes *Individual Care-as-Usual* Design: Die Studienteilnehmer:innen der Kontrollgruppe erhielten neben individueller Behandlung durch

einen Allgemeinarzt, ein Hilfecenter, Beratungsangebote oder bei Bedarf eine psychiatrische Behandlung auch eine Broschüre, in welcher verschiedene web-basierte Unterstützungsangebote, inklusive der zu untersuchenden, aufgelistet waren. Eine deskriptive Darstellung, welche Kontrollgruppenteilnehmer:in welche möglichen Zusatzbehandlungen nutzte, wurde nicht veröffentlicht.

Zusammenfassend kann festgehalten werden, dass sich die *Care-as-usual* Designs zwischen den ausgewählten Studien hinsichtlich ihrer genauen Definition, der zugelassenen Zusatzbehandlungen und der Kontrolle derselben unterschieden. Des Weiteren wiesen die Kontrollgruppen sowohl Psychopharmaka-Medikation, therapeutischer Behandlungen als auch diverse alternative Behandlungsangebote auf und repräsentierten damit ein breites Spektrum der Versorgungsrealität. In Bezug auf die Rahmenbedingungen der *My7steps App* kann festgehalten werden, dass die Studien von Klein et al. [50], Lüttke et al. [62] und Bruhns et al. [12] am ehesten den Bedingungen der vorliegenden Untersuchungen entsprachen.

Tabelle 11: Übersicht der Studien des literaturbasierten Kontrollarms

<b>Publikation</b>	<i>Klein et al. 2017 [50]</i>	<i>Beiwinkel et al. 2017 [10]</i>	<i>Lüdtke et al. 2018 [62]</i>	<i>Bruhns et al. 2021 [12]</i>	<i>Härter et al. 2018 [38]</i>	<i>MacLean et al. 2020 [63]</i>
<b>Land</b>	Deutschland	Deutschland	Deutschland	Deutschland	Deutschland	Kanada
<b>Art der Intervention</b>	Web-basierte Intervention <i>Deprexis</i>	Web-basierte Intervention <i>HelpID</i>	Selbsthilfe App <i>Tu Dir Gut</i>	Selbsthilfe App <i>MCT &amp; More</i>	Alternativer Behandlungs-Ansatz <i>Stepped collaborative care</i>	Info-Broschüre + web-basiertes Programm <i>The Journal</i> + Telefoncoaching
<b>Dauer der Intervention</b>	12 Wochen	12 Wochen	4 Wochen	4 Wochen	12 Wochen und nach Bedarf	12 Wochen
<b>Population</b>	PHQ-9 Score 5-14	Leichte bis mittelgradige Depression (F32.0, F32.1, F.33.0, F.33.1) und Dysthymie (F34.1)	Betroffene mit subjektivem Interventionsbedarf bzgl. depressiver Symptomatik	Depressive Symptomatik (PHQ-9 > 0)	PHQ-9 Score $\geq 5$	Patienten mit diagnostizierter Depression oder Dysthymie
<b>Zentrale Einschlusskriterien</b>	Alter 18-65 Ausreichende Deutschkenntnisse	Krankheitsbedingte Fehlzeiten aufgrund von Depression	Alter 18-65	Deutsche Universitätsstudierende Alter $\geq 18$	Alter $\geq 18$	Behandlung am Royal Ottawa Health Centre Alter $\geq 16$
<b>Zentrale Ausschlusskriterien</b>	Akute Suizidalität Als lebenslang diagnostizierte bipolare oder schizophrene Erkrankung	PHQ-9 Score $\geq 20$ Akute Suizidalität	Hohe suizidale Tendenzen	Akute suizidale Tendenzen Aktuelle oder vergangene bipolare oder psychotische Erkrankung	Unzureichende Deutschkenntnisse Unzureichender Gesundheitszustand Komorbide mentale Störung (z.B. Trauma)	Unzureichende Englischkenntnisse Nicht krankenversichert Kognitive Beeinträchtigung. Proband am selben Ort
<b>Rekrutierung</b>	Medien, Depressionsforen, Krankenversicherungen, Kliniken und Ambulanzen	KKH Krankenversicherung, Screening der Mitglieder nach Depressionserkrankung	Onlineforen und einzelne Ambulanz	Online Plattformen und Foren	Hamburger Ärztekammer	Patienten mit Depression oder Dysthymie des Royal Ottawa Health Centre
<b>Allokation</b>	Stratifizierte Randomisierung Strata: PHQ-9 < 10 PHQ-9 $\geq 10$	Blockrandomisierung (Blockgröße: 10)	Randomisierung	Randomisierung mit Gleichverteilung	Cluster-Randomisierung: Allgemeinarztpraxen im Verhältnis 3:1	Randomisierung mit Gleichverteilung
<b>Stichprobengröße zur Baseline:</b>						
<b>IG</b>	509	100	45	208	569	47
<b>KG</b>	504	80	45	215	168	48
<b>Stichprobengröße zur Post-Messung:</b>						
<b>IG</b>	395	44	44	128	402	47
<b>KG</b>	399	44	44	135	123	46
<b>Art der Kontrollgruppe</b>	Care-as-usual mit Warteliste	Care-as-usual mit Warteliste und text-basierter Psychoedukation	Care-as-usual mit Warteliste	Warteliste	Treatment-as-usual (Standardversorgung nach individuellem Bedarf)	Individual care-as-usual + Broschüre mit Anlaufstellen für mentale Gesundheit

IG = Interventionsgruppe; KG = Kontrollgruppe.

## Beschreibung der Kontrollstichproben

Die sechs eingeschlossenen Kontrollstichproben wurden ausgewählt, weil sie eine hohe Vergleichbarkeit zur vorliegenden Stichprobe aufwiesen. Die Altersspannen der Stichproben waren dagegen nur eingeschränkt vergleichbar, da diese in den anderen Studien entweder geringer (22.84) oder höher waren (42.90, 44,57, ...) als in der vorliegenden Studie. Das Alter der Studienteilnehmer:innen von Bruhns et al. [12], bei welchen es sich ausschließlich um Studierende handelt, mit einem mittleren Alter von 22.84 Jahren ( $SD = 3.15$ ), ähnelte noch am ehesten den *My7steps App* Studienteilnehmer:innen, die ein mittleres Alter von 30.48 Jahren ( $SD = 7.77$ ) aufwiesen. In Bezug auf den Bildungsstand war abermals die Studie von Bruhns et al. [12], wie auch die Publikation von MacLean und Kollegen [63], der vorliegenden Untersuchung am ähnlichsten, da die Studienteilnehmer:innen in allen drei Studien fast ausschließlich einen akademischen Hintergrund aufwiesen. Bezüglich der Geschlechtsverteilung entsprachen die Studienteilnehmer:innen der Studien von Beiwinkel et al. [10], Lüdtkke et al. [62] und Härter und Kollegen [38] am ehesten den *My7steps App* Studienteilnehmer:innen: In allen vier Studien waren in etwa drei Viertel der Studienteilnehmer:innen weiblichen Geschlechts. In Bezug auf den Familienstand, bei welchem die *My7steps App* Studienteilnehmer:innen ungefähr zur Hälfte verheiratet beziehungsweise in Partnerschaft lebend waren, stimmte dieser Prozentsatz am ehesten mit den Teilstichproben von Bruhns et al. [12] und Härter et al. [38] überein. Im Hinblick auf die Ausgangswerte des primären Endpunkts entsprachen die Kontroll-Studienteilnehmer:innen von Klein et al. [50] mit einem Mittelwert von 10.34 ( $SD = 2.40$ ), Beiwinkel et al. [10] ( $M = 10.56$ ,  $SD = 4.53$ ) und Bruhns et al. [12] mit einem mittleren PHQ-9 Score von 11.10 ( $SD = 4.42$ ) am ehesten den Studienteilnehmer:innen der vorliegenden Studie, die zu Beginn der Erhebung einen mittleren PHQ-9 Score von 10.81 ( $SD = 4.42$ ) aufwiesen (vgl. hierzu auch Tabelle 13). Eine Auflistung der deskriptiven Daten der Kontrollstichproben kann Tabelle 12 entnommen werden.

Tabelle 12: Stichprobenmerkmale der vorliegenden Untersuchung und der literaturbasierten Kontrollgruppen

Publikation	<i>My7steps App</i>	<i>Klein et al. [50]</i>	<i>Beiwinkel et al. [10]</i>	<i>Lüdtkke et al. [62]</i>	<i>Bruhns et al. [12]</i>	<i>Härter et al. [38]</i>	<i>MacLean et al. [63]</i>
<b>Stichprobengröße: <math>n</math></b>	89	504	80	44	200	168	48
<b>Alter: <math>M</math> (<math>SD</math>)</b>	30.48 (7.77)	42.90 (11.00)	48.66 (11.59)	44.57 (10.69)	22.84 (3.15)	45.60 (15.50)	44.80 (13.70)
<b>Geschlecht: weiblich, <math>n</math> (%)</b>	66 (74.16)	345 (68.45)	57 (71.25)	33 (75.00)	181 (90.50)	127 (75.60)	28 (58.33)
<b>Familienstand (verheiratet / in Partnerschaft): <math>n</math> (%)</b>	46 (51.69)	222 (44.05)	46 (57.50)	/	106 (53.00)	88 (52.38)	22 (45.83)
<b>Bildungsstand (Abitur oder höherwertig): <math>n</math> (%)</b>	85 (95.51)	271 (53.77)	67 (83.75)	25 (56.81)	200 (100.00)	54 (32.14)	48 (100.00)

## Beschreibung der Effektstärken der Kontrollstichprobe

Um zu untersuchen, welche hypothetischen Effektstärken durch die *My7steps App* unter Einbezug der externen Kontrollgruppen erzielt worden wären, wurden die Mittelwerte, Standardabweichungen und Stichprobengrößen der Gruppen vor und nach der Intervention zur Berechnung einer Effektstärke nach Cohen's  $d$  herangezogen; Details hierzu können der Tabelle 13 entnommen werden. Diesen Schätzern liegen unterschiedliche inferenzstatistische Analysen zugrunde: Drei der sechs Studien nutzten zur Berechnung der Wirksamkeit ihrer Interventionen *Intention-to-Treat*-Analysen [10, 38, 50] und entsprachen somit dem Analyse-Verfahren im vorliegenden Bericht. Eine Evaluationsstudie verwendete

als Analyse *Per-Protokoll* [12], die restlichen rechneten mit den tatsächlichen Zellbesetzungen vor und nach der Intervention [62, 63].

Hinsichtlich des sogenannten within-group Effekts, der die Veränderungen innerhalb einer Gruppe über die Zeit untersucht, zeigte sich, dass alle Kontroll-Studienteilnehmer:innen der untersuchten Studien eine Verbesserung der Depressionssymptomatik über den Studienzeitraum erlebten – sei es im Rahmen des natürlichen Krankheitsverlaufs oder durch die jeweilige Kontrollgruppen-Versorgung. Das Ausmaß dieser Symptomreduktion in den Kontrollgruppen lag zwischen  $d = -0.21$  [12] und  $d = -0.62$  [10]. Vergleicht man diese Werte mit den vorliegenden within-group Effekten der *My7steps App* Interventionsgruppe ( $d = 0.81$  bis  $d = 0.93$ ), zeigt sich eine deutlich stärkere Verbesserung der Depressionssymptomatik für die Studienteilnehmer:innen der *My7steps App* Interventionsgruppe gegenüber den Kontrollgruppen ( $\Delta$  Cohen's  $d = 0.19$  bis  $0.72$ ), sodass davon ausgegangen werden kann, dass mindestens die Differenz der within-Effekte auf die *My7steps App* Intervention zurückzuführen ist.

Die Berechnung der between-Effektstärken unterstützt diesen Befund: In allen literaturbasierten Vergleichen erzielt die Interventionsgruppe der vorliegenden Untersuchung einen Effekt gegenüber der jeweiligen Kontrollgruppe. Die größte between-Effektstärke der *My7steps App* Intervention gegenüber einer Kontrollgruppe von  $d = -1.19$  zeigte sich bei der Gegenüberstellung mit den Kontroll-Studienteilnehmer:innen von Klein und Kollegen [50] – eine der Studien, die auch am ehesten den Bedingungen der vorliegenden Untersuchungen entsprachen. Die kleinste between-Effektstärke von  $d = -0.23$  wurde durch die Kontrastierung mit den Kontrollgruppenteilnehmern von MacLean et al. [63] erzielt. Die durchschnittliche between-Effektstärke betrug über die sechs Studien gemittelt  $d = -0.57$ , was nach den Konventionen einer mittleren Effektstärke entspricht.

Vergleicht man diese Effekte in einem weiteren Schritt mit den between-Effektstärken der Original-Interventionsgruppen gegenüber den Kontrollgruppen, so zeigt sich für alle untersuchten Studien – bis auf die Untersuchung von Beiwinkel et al. [10] – dass die Kontrastierung mit den Studienteilnehmer:innen der *My7steps App* höhere between-Effektstärken erzielte als mit den jeweiligen Teilnehmern der Original-Interventionsgruppen. Hierbei ergab sich die größte between-Effektstärkendifferenz zu Gunsten der vorliegenden Untersuchung von  $\Delta$  Cohen's  $d = -0.54$  in der Kontrastierung mit Klein et al. [50], die kleinste von  $\Delta$  Cohen's  $d = -0.12$  in der Kontrastierung mit MacLean et al. [63].

Tabelle 13: Mittelwerte und Effektstärken des primären Endpunkts über die Messzeitpunkte und Studien

	<i>Baseline</i>	<i>Post</i>	<i>ES within</i>	<i>ES between</i>	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>KG (<math>\Delta</math> My7steps)</i>	<i>IG</i>	<i>My7steps (<math>\Delta</math>)</i>
<b>Klein et al.<sup>a</sup></b>	10.34 (2.40)	9.56 (4.61)	-0.33 (0.48 – 0.60)	-0.65	-1.19 (-0.54)
<b>Beiwinkel et al.<sup>a</sup></b>	10.56 (4.53)	7.76 (3.63)	-0.62 (0.19 – 0.31)	-0.50	-0.29 (+0.21)
<b>Lüdtke et al.<sup>c</sup></b>	12.77 (6.40)	10.72 (6.05)	-0.32 (0.49 – 0.61)	-0.12	-0.40 (-0.28)
<b>Bruhns et al.<sup>b</sup></b>	11.10 (4.42)	10.17 (4.32)	-0.21 (0.60 – 0.72)	-0.22	-0.72 (-0.50)
<b>Härter et al.<sup>a</sup></b>	14.09 (4.92)	12.62 (5.33)	-0.30 (0.51 – 0.63)	-0.37	-0.56 (-0.19)
<b>MacLean et al.<sup>c</sup></b>	15.40 (5.40)	12.40 (6.40)	-0.56 (0.25 – 0.37)	-0.11	-0.23 (-0.12)

*M = Mittelwert; SD = Standardabweichung; ES = Effektstärke (Cohen's d); ES within KG = within-Effektstärke der Kontrollgruppe der Originalstichprobe;  $\Delta$  My7steps = Differenz zu den within-Effektstärken der vorliegenden Studie (als Betrag angegeben; Bezug sind  $d = -0,81$  und  $d = -0,93$ ); ES between IG = between-Effektstärke der Original-Interventionsgruppe zur Kontrollgruppe; ES between My7steps = between-Effektstärke der My7steps-App-Interventionsgruppe zur Kontrollgruppe;  $\Delta$  = Differenz; <sup>a</sup>= Intention-to-Treat; <sup>b</sup>= Per-Protokoll; <sup>c</sup>= tatsächliche Zellbesetzung.*

## 12.3 Diskussion der Ergebnisse des literaturbasierten Kontrollarms

Die vorliegende Studie wurde als intra-individueller Vergleich durchgeführt, eigene Daten eines Kontrollarms lagen nicht vor. Daher wurde zur Plausibilisierung der gezeigten Effekte ein literaturbasierter Kontrollarm aufgebaut, in den sechs im Rahmen einer systematischen Literaturrecherche identifizierten Studien einbezogen wurden. In all diesen Studien wurde eine vergleichbare Kohorte an Studienteilnehmer:innen untersucht, Depressivität mit einem etablierten Messinstrument erfasst – zumeist mit dem auch hier verwendeten PHQ-9 – und ausreichend Untersuchungsdetails publiziert, um den Studienablauf, die Stichprobenmerkmale und die ermittelten Effekte verlässlich mit den vorliegenden Befunden vergleichen zu können. Hierbei wurden die within Effektstärken der Kontrollgruppen zum einen den within Effektstärken der *My7steps App* gegenübergestellt, um die Wirksamkeit der Intervention gegenüber dem natürlichen Krankheitsverlauf und eventuellen Effekten einer Begleittherapie zu betrachten. Zum anderen wurden anhand der Kontrollgruppendaten hypothetische between Effektstärken für die *My7steps App* Intervention ermittelt und in einem zusätzlichen Schritt mit den Effektstärken der Original-Interventionsgruppen verglichen.

### Diskussion der Vergleichbarkeit

Hinsichtlich der Vergleichbarkeit der Studienteilnehmer:innen gibt es Unterschiede: Keine Kontrollgruppe der externen Studien gleicht in jedem der untersuchten Merkmale den Teilnehmenden der *My7steps App* Interventionsgruppe und doch sind alle Stichproben entsprechend der angewendeten Suchstrategie in vielen Aspekten vergleichbar. Im Folgendem sollen die herangezogenen Studien noch einmal zusammengefasst sowie hinsichtlich der Vergleichbarkeit der Stichproben diskutiert werden:

- Die Untersuchung von Klein et al. [50] hatte zum Ziel, die Wirkung einer web-basierten Intervention gegen eine *Care-as-usual* Design mit Warteliste zu kontrastieren. Die Nutzung von Medikation und Therapie war für die Kontroll-Studienteilnehmer:innen explizit gestattet, wurde jedoch nicht weiter dargestellt. Die Autoren verwendeten zur Allokation eine stratifizierte Randomisierung durch Verwendung der PHQ-9 Ausgangswerte und als Analyseverfahren ITT. Die Kontrollarmteilnehmer waren im Schnitt älter, weniger weiblich und zeigten einen niedrigeren Bildungsstand im Vergleich mit der vorliegenden Untersuchung. Hinsichtlich der Ausgangswerte der Depressivität im PHQ-9 waren die Kontrollgruppenteilnehmer mit denen der vorliegenden Untersuchung vergleichbar. Hier zeigte sich ein hypothetischer between-group Interventionseffekt von  $d = -1.19$ .
- Beiwinkel et al. [10] untersuchten ebenfalls eine web-basierte Anwendung unter Verwendung einer blockrandomisierten Wartelistengruppe, die eine text-basierte Psychoedukation über Depression erhielt und Zusatzbehandlungen in Anspruch nehmen durfte. Die Studienteilnehmer:innen waren auch hier im Schnitt etwas älter, wiesen aber hinsichtlich

Geschlecht, Familien- und Bildungsstand sowie Ausgangsniveau der Depressivität vergleichbare Werte zur *My7steps App* Interventionsgruppe auf. Auch hier wurde zur Analyse ITT verwendet. Es ergab sich ein hypothetischer between-group Interventionseffekt von  $d = -0.29$ .

- Lütcke et al. [62], sowie Bruhns et al. [12] untersuchten anhand zweier unabhängiger Stichproben dieselbe web-basierte Intervention zur Verbesserung von Depression. Beide nutzten ein Wartelistendesign mit Randomisierung der Teilnehmer, im Falle von Lütcke et al. mit eindeutigem *Care-as-usual* Elementen, im Falle von Bruhns konnte dies aufgrund der Diskussion in Bezug auf die Symptomverbesserung und soziodemographischen Information abgeleitet werden. Die Kontroll-Studienteilnehmer:innen bei Lütcke et al. [62] waren etwas älter und wiesen einen geringeren Bildungsstand auf, waren aber hinsichtlich der Geschlechterverteilung vergleichbar mit den Studienteilnehmer:innen der vorliegenden Untersuchung. In den Ausgangswerten des PHQ-9 wiesen sie etwas höhere Werte auf. Zur Analyse wurde die tatsächliche Zellbesetzung zur Baseline- und Posterhebung verwendet. Die Berechnung des hypothetischen between-group Interventionseffekts ergab ein  $d = -0.40$ . Die Studienteilnehmer:innen von Bruhns et al. [12] waren ähnlichen Alters, Familien- und Bildungsstands, jedoch in überwiegender Mehrheit weiblichen Geschlechts. Hinsichtlich der PHQ-9 Baseline-Werte waren auch sie mit den *My7steps App* Studienteilnehmer:innen vergleichbar. Die Autoren verwendeten *Per-Protokoll*-Analysen zur Berechnung der Interventionseffekte. Es ergab sich ein hypothetischer between-group Interventionseffekt von  $d = -0.72$ .
- Härter et al. [38] untersuchten einen neuen Behandlungsansatz, der in der ärztlichen Regelversorgung durch praktizierende Mediziner mittels *Care-as-usual* Design erprobt werden sollte. Die Teilnehmer der Kontrollgruppe, die durch Cluster-Randomisierung zugewiesen wurden, waren zwar etwas älter und weniger gebildet, zeigten aber eine ähnliche Verteilung hinsichtlich Geschlecht und Familienstand. Hier zeigte sich deutlich ein höherer Ausgangswert in den PHQ-9 Ausprägungen. Die Analyse wurden nach dem ITT-Verfahren angewendet; es ergab sich ein hypothetischer between-group Interventionseffekt von  $d = -0.56$ .
- Die einzige nicht-deutsche Untersuchung von MacLean et al. [63] untersuchte eine web-basierte Intervention in Kanada anhand eines *Care-as-usual* Designs mit zusätzlicher minimaler Intervention unter Verwendung einer Hilfsbroschüre. Auch hier waren die Teilnehmer der Kontroll-Studienteilnehmer:innen, die mittels Randomisierung der Gruppen zugewiesen wurden, etwas älter als die *My7steps App* Studienteilnehmer:innen, es waren weniger weibliche Studienteilnehmer:innen beteiligt, die Gruppen jedoch hinsichtlich Familien- und Bildungsstands vergleichbar. Die Ausgangswerte der Depressivität mittels PHQ-9 Baseline-Messung zeigte auch hier deutliche höhere Werte der Kontroll-Studienteilnehmer:innen. Ein hypothetischer between-group Interventionseffekt ergab  $d = -0.23$ .

Stellt man die Studienteilnehmer:innen der Kontrollgruppen der externen Studien den Studienteilnehmer:innen der *My7steps App* gegenüber, zeigt sich die höchste Übereinstimmung hinsichtlich der soziodemographischen bzw. Baseline-Werte des primären Endpunkts mit den Kontroll-Studienteilnehmer:innen von Beiwinkel et al. [10] und Bruhns et al. [12]. Wie in Kapitel 12.2 (Abschnitt „Beschreibung der Kontrollgruppen-Designs der gefundenen Studien“) bereits ausgeführt, waren hinsichtlich der Zusatzangebote, die die Kontroll-Studienteilnehmer:innen der externen Studien nutzen durften, die Untersuchungen von Klein et al. [50], Lütcke et al. [62] und Bruhns et al. [12] am ehesten mit denen der vorliegenden Untersuchungen vergleichbar.

## Diskussion der Ergebnisse

Allen Studienteilnehmer:innen der Kontrollgruppen der externen Interventionsstudien war gemein, dass sie zwischen Baseline und Posterhebung eine Reduktion ihrer Symptomatik erlebten, die sich in Effektstärken zwischen  $d = -0.21$  und  $d = -0.62$  widerspiegeln. Dies ist im Feld der medizinischen Depressionsforschung nichts ungewöhnliches und wird auf unterschiedliche Faktoren zurückgeführt, wie etwa die Ausgangsschwere der Depression oder stattgefundenen Zusatzbehandlungen [70]. Auch die Autoren der verwendeten externen Studien vermuten hinter den Symptomverbesserungen innerhalb des Kontrollarms eine Wechselwirkung unterschiedlicher Faktoren, wie etwa Medikation und Therapie, was aber nicht in allen Fällen dokumentiert wurde [10, 12]. Dies ist unter anderem eine Folge der Verwendung ethisch notwendiger *Care-as-Usual* Designs. Der direkte Vergleich der within Effektstärken der externen Kontrollgruppen und der vorliegenden Untersuchung bezüglich des primären Endpunkts deutet bereits einen positiven Interventionseffekt der *My7steps App* an, der in allen Fällen über die in den sechs Kontrollgruppen festgestellten Veränderungen der depressiven Symptomatik hinausgeht: Die Höhe der Differenz der within Effekte reicht von  $\Delta$  Cohen's  $d = 0.19$  bis  $\Delta$  Cohen's  $d = 0.72$  zu Gunsten der vorliegenden Intervention. Der gefundene Interventionseffekt der *My7steps App* scheint damit deutlich über die verlaufs- oder begleittherapiebedingten Veränderungen der Kontrollgruppen hinauszugehen, was besonders dadurch an Bedeutsamkeit gewinnt, da unter den Kontrollgruppendesigns auch solche waren, die explizit die Verwendung von Medikation und Therapie zuließen und zum Teil sogar eine alternative Intervention anboten [10, 63]. Entsprechend kann davon ausgegangen werden, dass die Verbesserungen der *My7steps App* Interventionsgruppe nicht alleine auf natürliche Remission, simultane Medikation oder Psychotherapie zurückgeführt werden kann, sondern dass höchstwahrscheinlich ein substantieller zusätzlicher Interventionseffekt vorliegt.

Die Berechnung der Effektstärken zwischen den Studienteilnehmer:innen der vorliegenden Untersuchung und den Kontrollgruppenwerten der externen Studien bezüglich des primären Endpunkts zeigte in allen Fällen einen hypothetischen, positiven Interventionseffekt der *My7steps App*, wobei die Höhe des ermittelten between Effekts von  $d = -0.23$  bis  $d = -1.19$  reichte und im Mittel bei  $d = -0.57$  lag. Diese anhand des literaturbasierten Kontrollarms ermittelten Effektstärken wurden in einem weiteren Schritt mit den Effektstärken der Originalstichproben verglichen: In fünf von sechs Vergleichen übertraf die Effektstärke des vorliegenden Berichts die der Originalstichproben um mindestens  $d = 0.13$  bis  $d = 0.54$ . Hierbei war auffällig, dass die größeren Effektstärkendifferenzen ( $d \geq 0.5$ ) im Vergleich mit den drei Studien erzielt wurden, die ähnlich der *My7steps App* web-basierte Interventionen anboten und deren Studienteilnehmer:innen auch außerhalb ärztlicher Kontexte rekrutiert wurden, während die kleineren Effektstärkendifferenzen ( $d \leq 0.18$ ) durch Kontrastierung mit Studien zustande kamen, deren Kontroll-Studienteilnehmer:innen zum einen deutlich höhere Ausgangswerte des PHQ-9 aufwiesen und zum anderen im Kontext ärztlicher Hilfenahmen rekrutiert wurden und währenddessen eine ärztliche Standardversorgung erhielten, deren Umfang und therapeutische Wirkung tendenziell als umfassender einzuschätzen sind als Umfang und Wirkung von Kontrolldesigns, deren Studienteilnehmer:innen ausschließlich online betreut wurden. Die Kontroll-Studienteilnehmer:innen der Studie von Beiwinkel et al. [10] stammten zwar ebenfalls aus einer Studie, die eine web-basierte Intervention untersuchten; aufgrund des zusätzlichen psychoedukativen Angebots für die Kontrollgruppe und der Sachlage, dass fast die Hälfte der Studienteilnehmer:innen während der Intervention entweder Medikamente oder Psychotherapie erhielten, kann davon ausgegangen werden, dass die starke Symptomreduktion innerhalb dieser Kontrollgruppe die Höhe des hypothetischen between-Interventionseffekts eingeschränkt haben könnte. Dementsprechend sind die vorliegenden sechs positiven between-group Effektstärken des

literaturbasierten Kontrollarms – gerade mit Blick auf das weitreichende therapeutischen Zusatzangebot – vielversprechende Hinweise auf eine Wirksamkeit der *My7steps App* unter kontrollierten Bedingungen.

## Fazit

Um die Limitationen einer einarmigen Studie zu adressieren und die Ergebnisse zu plausibilisieren wurden die Effekte der Intervention in einem literaturbasierten Kontrollarm mit externen Kontrollgruppen verglichen. Die ermittelten Effekte der *My7steps App* auf depressive Symptomatiken können auch anhand des Vergleichs mit sechs vergleichbaren, externen Kontrollgruppen als bedeutsam und klinisch relevant eingestuft werden, da die vorliegenden Interventionseffekte deutlich über die verlaufs- oder begleittherapiebedingten Veränderungen der depressiven Symptomatiken der Kontrollgruppen hinausgehen ( $\Delta$  Cohen's  $d = 0.19$  bis  $\Delta$  Cohen's  $d = 0.72$ ) und die hypothetisch ermittelten between Effektstärken von  $d = -0.23$  bis  $d = -1.19$  ausnahmslos zu Gunsten der vorliegenden Intervention ausfielen. Somit kann davon ausgegangen werden, dass die in der Studie erzielten Effekte durch die Anwendung der *My7steps App* und nicht durch etwaige Begleitbehandlungen, wie Psychotherapie und Medikation, erreicht oder durch den natürlichen Krankheitsverlauf bedingt sind. Damit ist anzunehmen, dass die Intervention einen nachweisbaren Beitrag zur Verbesserung der Versorgung von Depressions- und Anpassungsstörungs-Patienten leisten kann und auch im Rahmen einer kontrollierten Erprobungsstudie ein entsprechender Nachweis des postulierten Versorgungseffekts erreicht werden kann.

## 13 Literaturverzeichnis

### References

1. Aherne, D., Fitzgerald, A., Aherne, C., Fitzgerald, N., Slattery, M., & Whelan, N. (2017). Evidence for the treatment of moderate depression: a systematic review. *Irish journal of psychological medicine*, 34(3), 197–204. doi:10.1017/ipm.2017.10.
2. American Psychological Association (2019). Clinical practice guideline for the treatment of depression across three age cohorts.
3. Andersson, G., Paxling, B., Roch-Norlund, P., Östman, G., Norgren, A., Almlöv, J., et al. (2012). Internet-based psychodynamic versus cognitive behavioral guided self-help for generalized anxiety disorder: a randomized controlled trial. *Psychotherapy and psychosomatics*, 81(6), 344–355. doi:10.1159/000339371.
4. Antúnez, J. M., Navarro, J. F., & Adan, A. (2015). Circadian typology is related to resilience and optimism in healthy adults. *Chronobiology international*, 32(4), 524–530. doi:10.3109/07420528.2015.1008700.
5. Ayoughi, S., Missmahl, I., Weierstall, R., & Elbert, T. (2012). Provision of mental health services in resource-poor settings: a randomised trial comparing counselling with routine medical treatment in North Afghanistan (Mazar-e-Sharif). *BMC psychiatry*, 12, 14. doi:10.1186/1471-244X-12-14.
6. Balzus, L., Feiler, M., Stephani, V., Groos, H., Heber, E., & Ebert, D. D. (2021). *HelloBetter Diabetes und Depression: Online-Intervention zur Bewältigung von depressiver Stimmung bei Personen mit Diabetes mellitus Typ 1 oder Typ 2*.
7. Bangasser, D. A., & Valentino, R. J. (2014). Sex differences in stress-related psychiatric disorders: neurobiological perspectives. *Frontiers in neuroendocrinology*, 35(3), 303–319. doi:10.1016/j.yfrne.2014.03.008.
8. Barak, A., Hen, L., Boniel-Nissim, M., & Shapira, N. (2008). A Comprehensive Review and a Meta-Analysis of the Effectiveness of Internet-Based Psychotherapeutic Interventions. *Journal of Technology in Human Services*, 26(2-4), 109–160. doi:10.1080/15228830802094429.
9. Barak, A., Klein, B., & Proudfoot, J. G. (2009). Defining internet-supported therapeutic interventions. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*, 38(1), 4–17. doi:10.1007/s12160-009-9130-7.
10. Beiwinkel, T., Eißing, T., Telle, N.-T., Siegmund-Schultze, E., & Rössler, W. (2017). Effectiveness of a Web-Based Intervention in Reducing Depression and Sickness Absence: Randomized Controlled Trial. *Journal of medical Internet research*, 19(6), e213. doi:10.2196/jmir.6546.
11. Brandt, W. A., Loew, T., Heymann, F. von, Stadtmüller, G., Georgi, A., Tischinger, M., et al. (2015). Das ICD-10 Symptomrating (ISR): Validierung der Depressionsskala an einer klinischen Stichprobe. *Psychotherapie, Psychosomatik, medizinische Psychologie*, 65(6), 234–237. doi:10.1055/s-0034-1398542.
12. Bruhns, A., Lüdtke, T., Moritz, S., & Bücker, L. (2021). A Mobile-Based Intervention to Increase Self-esteem in Students With Depressive Symptoms: Randomized Controlled Trial. *JMIR mHealth and uHealth*, 9(7), e26498. doi:10.2196/26498.
13. Callan, J. A., Wright, J., Siegle, G. J., Howland, R. H., & Kepler, B. B. (2017). Use of Computer and Mobile Technologies in the Treatment of Depression. *Archives of psychiatric nursing*, 31(3), 311–318. doi:10.1016/j.apnu.2016.10.002.
14. Campbell-Sills, L., Forde, D. R., & Stein, M. B. (2009). Demographic and childhood environmental predictors of resilience in a community sample. *Journal of psychiatric research*, 43(12), 1007–1012. doi:10.1016/j.jpsychires.2009.01.013.

15. Carta, M. G., Balestrieri, M., Murru, A., & Hardoy, M. C. (2009). Adjustment Disorder: epidemiology, diagnosis and treatment. *Clinical practice and epidemiology in mental health : CP & EMH*, 5, 15. doi:10.1186/1745-0179-5-15.
16. Casey, P. (2009). Adjustment disorder: epidemiology, diagnosis and treatment. *CNS drugs*, 23(11), 927–938. doi:10.2165/11311000-000000000-00000.
17. COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hove, London: Lawrence Erlbaum Associates.
18. Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depression and anxiety*, 18(2), 76–82. doi:10.1002/da.10113.
19. 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. doi:10.1111/acps.13335.
20. Cuijpers, P., Kleiboer, A., Karyotaki, E., & Riper, H. (2017). Internet and mobile interventions for depression: Opportunities and challenges. *Depression and anxiety*, 34(7), 596–602. doi:10.1002/da.22641.
21. Cuijpers, P., Turner, E. H., Koole, S. L., van Dijke, A., & Smit, F. (2014). What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depression and anxiety*, 31(5), 374–378. doi:10.1002/da.22249.
22. Cuijpers, P., van Straten, A., Warmerdam, L., & Smits, N. (2008). Characteristics of effective psychological treatments of depression: a meta-regression analysis. *Psychotherapy research : journal of the Society for Psychotherapy Research*, 18(2), 225–236. doi:10.1080/10503300701442027.
23. Davidson, J. R. T. (2018). *Connor-Davidson Resilience Scale (CD-RISC) Manual*.
24. Domhardt, M., & Baumeister, H. (2018). Psychotherapy of adjustment disorders: Current state and future directions. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry*, 19(sup1), S21-S35. doi:10.1080/15622975.2018.1467041.
25. Domhardt, M., Geßlein, H., Rezori, R. E. von, & Baumeister, H. (2019). Internet- and mobile-based interventions for anxiety disorders: A meta-analytic review of intervention components. *Depression and anxiety*, 36(3), 213–224. doi:10.1002/da.22860.
26. Donker, T., Batterham, P. J., Warmerdam, L., Bennett, K., Bennett, A., Cuijpers, P., et al. (2013). Predictors and moderators of response to internet-delivered Interpersonal Psychotherapy and Cognitive Behavior Therapy for depression. *Journal of affective disorders*, 151(1), 343–351. doi:10.1016/j.jad.2013.06.020.
27. Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1(2), 170–177. doi:10.1037/1082-989X.1.2.170.
28. Ebert, D. D., Buntrock, C., Lehr, D., Smit, F., Riper, H., Baumeister, H., et al. (2018). Effectiveness of Web- and Mobile-Based Treatment of Subthreshold Depression With Adherence-Focused Guidance: A Single-Blind Randomized Controlled Trial. *Behavior therapy*, 49(1), 71–83. doi:10.1016/j.beth.2017.05.004.
29. Ebert, D. D., van Daele, T., Nordgreen, T., Karekla, M., Compare, A., Zarbo, C., et al. (2018). Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health. *European Psychologist*, 23(2), 167–187. doi:10.1027/1016-9040/a000318.
30. Ebert, D. D., Zarski, A.-C., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M., et al. (2015). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a

- meta-analysis of randomized controlled outcome trials. *PLoS one*, 10(3), e0119895. doi:10.1371/journal.pone.0119895.
31. Fernández, A., Mendive, J. M., Salvador-Carulla, L., Rubio-Valera, M., Luciano, J. V., Pinto-Meza, A., et al. (2012). Adjustment disorders in primary care: prevalence, recognition and use of services. *The British journal of psychiatry : the journal of mental science*, 201, 137–142. doi:10.1192/bjp.bp.111.096305.
  32. Fitts, D. A. (2020). Commentary on "A review of effect sizes and their confidence intervals, Part I: The Cohen's d family": The degrees of freedom for paired samples designs. *The Quantitative Methods for Psychology*, 16(4), 281–294. doi:10.20982/tqmp.16.4.p281.
  33. Freedland, K. E., Mohr, D. C., Davidson, K. W., & Schwartz, J. E. (2011). Usual and unusual care: existing practice control groups in randomized controlled trials of behavioral interventions. *Psychosomatic medicine*, 73(4), 323–335. doi:10.1097/PSY.0b013e318218e1fb.
  34. Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, 42(3), 237–288. doi:10.3102/00346543042003237.
  35. Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)". *Diagnostica*, 50(4), 171–181. doi:10.1026/0012-1924.50.4.171.
  36. Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in clinical research*, 2(3), 109–112. doi:10.4103/2229-3485.83221.
  37. Harrer, M., Adam, S. H., Fleischmann, R. J., Baumeister, H., Auerbach, R., Bruffaerts, R., et al. (2018). Effectiveness of an Internet- and App-Based Intervention for College Students With Elevated Stress: Randomized Controlled Trial. *Journal of medical Internet research*, 20(4), e136. doi:10.2196/jmir.9293.
  38. Härter, M., Watzke, B., Daubmann, A., Wegscheider, K., König, H.-H., Brettschneider, C., et al. (2018). Guideline-based stepped and collaborative care for patients with depression in a cluster-randomised trial. *Scientific reports*, 8(1), 9389. doi:10.1038/s41598-018-27470-6.
  39. Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics*, 17(4), 315–339. doi:10.3102/10769986017004315.
  40. Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior therapy*, 15(4), 336–352. doi:10.1016/S0005-7894(84)80002-7.
  41. Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of consulting and clinical psychology*, 59(1), 12–19. doi:10.1037//0022-006x.59.1.12.
  42. Johansson, R., Hesser, H., Ljótsson, B., Frederick, R. J., & Andersson, G. (2012). Transdiagnostic, affect-focused, psychodynamic, guided self-help for depression and anxiety through the internet: study protocol for a randomised controlled trial. *BMJ open*, 2(6). doi:10.1136/bmjopen-2012-002167.
  43. Jones, S. M., Crane, P. K., & Simon, G. (2019). A comparison of individual change using Item Response Theory and sum scoring on the A Comparison of Individual Change using Item Response Theory and Sum Scoring on the Patient Health Questionnaire-9: implications for Measurement-Based Care. *Ann Depress Anxiety*, 6(1).
  44. Josephine, K., Josefine, L., Philipp, D., David, E., & Harald, B. (2017). Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. *Journal of affective disorders*, 223, 28–40. doi:10.1016/j.jad.2017.07.021.

45. Kaltenbach, E., Härdtnner, E., Hermenau, K., Schauer, M., & Elbert, T. (2017). Efficient identification of mental health problems in refugees in Germany: the Refugee Health Screener. *European journal of psychotraumatology*, 8(sup2), 1389205. doi:10.1080/20008198.2017.1389205.
46. Karyotaki, E., Kemmeren, L., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., et al. (2018). Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. *Psychological medicine*, 48(15), 2456–2466. doi:10.1017/S0033291718000648.
47. Kleiboer, A., Donker, T., Seekles, W., van Straten, A., Riper, H., & Cuijpers, P. (2015). A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. *Behaviour research and therapy*, 72, 63–71. doi:10.1016/j.brat.2015.06.013.
48. Klein, J. P., Berger, T., Schröder, J., Späth, C., Meyer, B., Caspar, F., et al. (2016). Effects of a Psychological Internet Intervention in the Treatment of Mild to Moderate Depressive Symptoms: Results of the EVIDENT Study, a Randomized Controlled Trial. *Psychotherapy and psychosomatics*, 85(4), 218–228. doi:10.1159/000445355.
49. Klein, J. P., Knaevelsrud, C., Bohus, M., Ebert, D. D., Gerlinger, G., Günther, K., et al. (2018). Internetbasierte Selbstmanagementinterventionen : Qualitätskriterien für ihren Einsatz in Prävention und Behandlung psychischer Störungen. *Der Nervenarzt*, 89(11), 1277–1286. doi:10.1007/s00115-018-0591-4.
50. Klein, J. P., Späth, C., Schröder, J., Meyer, B., Greiner, W., Hautzinger, M., et al. (2017). Time to remission from mild to moderate depressive symptoms: One year results from the EVIDENT-study, an RCT of an internet intervention for depression. *Behaviour research and therapy*, 97, 154–162. doi:10.1016/j.brat.2017.07.013.
51. Kocalevent, R.-D., Hinz, A., & Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General hospital psychiatry*, 35(5), 551–555. doi:10.1016/j.genhosppsych.2013.04.006.
52. Krämer, R., Köhne-Volland, L., Schumacher, A., & Köhler, S. (2022). Efficacy of a Web-Based Intervention for Depressive Disorders: Three-Arm Randomized Controlled Trial Comparing Guided and Unguided Self-Help With Waitlist Control. *JMIR formative research*, 6(4), e34330. doi:10.2196/34330.
53. Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, 16(9), 606–613. doi:10.1046/j.1525-1497.2001.016009606.x.
54. Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2010). The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *General hospital psychiatry*, 32(4), 345–359. doi:10.1016/j.genhosppsych.2010.03.006.
55. Ladwig, I., Rief, W., & Nestoriuc, Y. (2014). Welche Risiken und Nebenwirkungen hat Psychotherapie? - Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP). *Verhaltenstherapie*, 24(4), 252–263. doi:10.1159/000367928.
56. Lauridsen, L. S., Willert, M. V., Eskildsen, A., & Christiansen, D. H. (2017). Cross-cultural adaptation and validation of the Danish 10-item Connor-Davidson Resilience Scale among hospital staff. *Scandinavian journal of public health*, 45(6), 654–657. doi:10.1177/1403494817721056.
57. Lin, J., Lüking, M., Ebert, D. D., Buhrman, M., Andersson, G., & Baumeister, H. (2015). Effectiveness and cost-effectiveness of a guided and unguided internet-based Acceptance and Commitment Therapy for chronic pain: Study protocol for a three-armed randomised controlled trial. *Internet interventions*, 2(1), 7–16. doi:10.1016/j.invent.2014.11.005.
58. Liu-Seifert, H., Zhang, S., D'Souza, D., & Skljarevski, V. (2010). A closer look at the baseline-observation-carried-forward (BOCF). *Patient preference and adherence*, 4, 11–16. doi:10.2147/ppa.s8135.

59. Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test. *Review of Educational Research*, 66(4), 579–619. doi:10.3102/00346543066004579.
60. Löwe, B., Kroenke, K., Herzog, W., & Gräfe, K. (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *Journal of affective disorders*, 81(1), 61–66. doi:10.1016/S0165-0327(03)00198-8.
61. Löwe, B., Unützer, J., Callahan, C. M., Perkins, A. J., & Kroenke, K. (2004). Monitoring depression treatment outcomes with the patient health questionnaire-9. *Medical care*, 42(12), 1194–1201. doi:10.1097/00005650-200412000-00006.
62. Lüdtke, T., Pult, L. K., Schröder, J., Moritz, S., & Bücker, L. (2018). A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. *Psychiatry research*, 269, 753–762. doi:10.1016/j.psychres.2018.08.113.
63. MacLean, S., Corsi, D. J., Litchfield, S., Kucharski, J., Genise, K., Selaman, Z., et al. (2020). Coach-Facilitated Web-Based Therapy Compared With Information About Web-Based Resources in Patients Referred to Secondary Mental Health Care for Depression: Randomized Controlled Trial. *Journal of medical Internet research*, 22(6), e15001. doi:10.2196/15001.
64. Mak, W. W. S., Chan, A. T. Y., Cheung, E. Y. L., Lin, C. L. Y., & Ngai, K. C. S. (2015). Enhancing Web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. *Journal of medical Internet research*, 17(1), e8. doi:10.2196/jmir.3746.
65. Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 184(3), E191-6. doi:10.1503/cmaj.110829.
66. Marshall, J. M., Dunstan, D. A., & Bartik, W. (2019). The Digital Psychiatrist: In Search of Evidence-Based Apps for Anxiety and Depression. *Frontiers in psychiatry*, 10, 831. doi:10.3389/fpsy.2019.00831.
67. McAleavey, A. A. (2021). *When (Not) to Rely on the Reliable Change Index* .
68. McCoy, C. E. (2017). Understanding the Intention-to-treat Principle in Randomized Controlled Trials. *The western journal of emergency medicine*, 18(6), 1075–1078. doi:10.5811/westjem.2017.8.35985.
69. McMillan, D., Gilbody, S., & Richards, D. (2010). Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods. *Journal of affective disorders*, 127(1-3), 122–129. doi:10.1016/j.jad.2010.04.030.
70. Mekonen, T., Ford, S., Chan, G. C. K., Hides, L., Connor, J. P., & Leung, J. (2022). What is the short-term remission rate for people with untreated depression? A systematic review and meta-analysis. *Journal of affective disorders*, 296, 17–25. doi:10.1016/j.jad.2021.09.046.
71. Meyer, B., Bierbrodt, J., Schröder, J., Berger, T., Beevers, C. G., Weiss, M., et al. (2015). Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. *Internet interventions*, 2(1), 48–59. doi:10.1016/j.invent.2014.12.003.
72. Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364–386. doi:10.1177/1094428106291059.
73. Moshe, I., Terhorst, Y., Philippi, P., Domhardt, M., Cuijpers, P., Cristea, I., et al. (2021). Digital interventions for the treatment of depression: A meta-analytic review. *Psychological bulletin*, 147(8), 749–786. doi:10.1037/bul0000334.
74. Nobis, S., Lehr, D., Ebert, D. D., Baumeister, H., Snoek, F., Riper, H., et al. (2015). Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. *Diabetes care*, 38(5), 776–783. doi:10.2337/dc14-1728.

75. O'Donnell, M. L., Metcalf, O., Watson, L., Phelps, A., & Varker, T. (2018). A Systematic Review of Psychological and Pharmacological Treatments for Adjustment Disorder in Adults. *Journal of traumatic stress*, 31(3), 321–331. doi:10.1002/jts.22295.
76. Orang, T. M., Missmahl, I., Thoele, A.-M., Valensise, L., Brenner, A., Gardisi, M., et al. (2022). New directions in the mental health care of migrants, including refugees-A randomized controlled trial investigating the efficacy of value-based counselling. *Clinical psychology & psychotherapy*. doi:10.1002/cpp.2728.
77. Patra, B. N., & Sarkar, S. (2013). Adjustment disorder: current diagnostic status. *Indian journal of psychological medicine*, 35(1), 4–9. doi:10.4103/0253-7176.112193.
78. Pfizer (n.d.). *Instruction manual: Instructions for Patient Health Questionnaire (PHQ) and GAD-7 measures*.
79. Pinto-Meza, A., Serrano-Blanco, A., Peñarrubia, M. T., Blanco, E., & Haro, J. M. (2005). Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *Journal of general internal medicine*, 20(8), 738–742. doi:10.1111/j.1525-1497.2005.0144.x.
80. Rief, W., Nanke, A., Klaiberg, A., & Braehler, E. (2004). Base rates for panic and depression according to the Brief Patient Health Questionnaire: a population-based study. *Journal of affective disorders*, 82(2), 271–276. doi:10.1016/j.jad.2003.11.006.
81. RKI (2017). *Depressive Symptomatik bei Erwachsenen in Deutschland* : RKI-Bib1 (Robert Koch-Institut).
82. Robinson, E., Titov, N., Andrews, G., McIntyre, K., Schwencke, G., & Solley, K. (2010). Internet treatment for generalized anxiety disorder: a randomized controlled trial comparing clinician vs. technician assistance. *PLoS one*, 5(6), e10942. doi:10.1371/journal.pone.0010942.
83. Sander, L., Ebert, D. D., & Baumeister, H. (2017). Internet- und mobilebasierte Psychotherapie der Depression. *Fortschritte der Neurologie-Psychiatrie*, 85(1), 48–58. doi:10.1055/s-0042-123540.
84. Sarubin, N., Gutt, D., Giegling, I., Bühner, M., Hilbert, S., Krähenmann, O., et al. (2015). Erste Analyse der psychometrischen Eigenschaften und Struktur der deutschsprachigen 10- und 25-Item Version der Connor-Davidson Resilience Scale (CD-RISC). *Zeitschrift für Gesundheitspsychologie*, 23(3), 112–122. doi:10.1026/0943-8149/a000142.
85. Schäfer, J., Wittchen, H.-U., Höfler, M., Heinrich, A., Zimmermann, P., Siegel, S., et al. (2015). Is trait resilience characterized by specific patterns of attentional bias to emotional stimuli and attentional control? *Journal of behavior therapy and experimental psychiatry*, 48, 133–139. doi:10.1016/j.jbtep.2015.03.010.
86. Skruibis, P., Eimontas, J., Dovydaityene, M., Mazulyte, E., Zelviene, P., & Kazlauskas, E. (2016). Internet-based modular program BADI for adjustment disorder: protocol of a randomized controlled trial. *BMC psychiatry*, 16, 264. doi:10.1186/s12888-016-0980-9.
87. Spitzer, R. L., Williams, J. B., Kroenke, K., & et al. (2014). Test Review: Patient Health Questionnaire–9 (PHQ-9). *Rehabilitation Counseling Bulletin*, 57(4), 246–248. doi:10.1177/0034355213515305.
88. Stiles-Shields, C., Montague, E., Kwasny, M. J., & Mohr, D. C. (2019). Behavioral and cognitive intervention strategies delivered via coached apps for depression: Pilot trial. *Psychological services*, 16(2), 233–238. doi:10.1037/ser0000261.
89. Stone, E. R. (2010). t Test, Paired Samples. In N. J. Salkind (Ed.), *Encyclopedia of research design: Volume 1 - 3* (pp. 1560–1565). Los Angeles, London, New Delhi, Singapore, Washington, D.C.: SAGE.
90. Sucala, M., Cuijpers, P., Muench, F., Cardoso, R., Soflau, R., Dobrea, A., et al. (2017). Anxiety: There is an app for that. A systematic review of anxiety apps. *Depression and anxiety*, 34(6), 518–525. doi:10.1002/da.22654.

91. Thompson, B. T., & Schoenfeld, D. (2007). Usual care as the control group in clinical trials of nonpharmacologic interventions. *Proceedings of the American Thoracic Society*, 4(7), 577–582. doi:10.1513/pats.200706-072JK.
92. Tritt, K., Heymann, F. von, Zaudig, M., Zacharias, I., Söllner, W., & Loew, T. (2008). Entwicklung des Fragebogens "ICD-10-Symptom-Rating" (ISR). *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 54(4), 409–418. doi:10.13109/zptm.2008.54.4.409.
93. Van, H. L., Schoevers, R. A., & Dekker, J. (2008). Predicting the outcome of antidepressants and psychotherapy for depression: a qualitative, systematic review. *Harvard review of psychiatry*, 16(4), 225–234. doi:10.1080/10673220802277938.
94. van Ameringen, M., Turna, J., Khalesi, Z., Pullia, K., & Patterson, B. (2017). There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. *Depression and anxiety*, 34(6), 526–539. doi:10.1002/da.22657.
95. Weisel, K. K., Zarski, A.-C., Berger, T., Krieger, T., Schaub, M. P., Moser, C. T., et al. (2019). Efficacy and cost-effectiveness of guided and unguided internet- and mobile-based indicated transdiagnostic prevention of depression and anxiety (ICare Prevent): A three-armed randomized controlled trial in four European countries. *Internet interventions*, 16, 52–64. doi:10.1016/j.invent.2018.04.002.
96. Wingo, A. P., Briscione, M., Norrholm, S. D., Jovanovic, T., McCullough, S. A., Skelton, K., et al. (2017). Psychological resilience is associated with more intact social functioning in veterans with post-traumatic stress disorder and depression. *Psychiatry research*, 249, 206–211. doi:10.1016/j.psychres.2017.01.022.
97. Wollny, A. I., & Jacobs, I. (2021). Validity and reliability of the German versions of the CD-RISC-10 and CD-RISC-2. *Current psychology (New Brunswick, N.J.)*, 1–12. doi:10.1007/s12144-021-01670-2.
98. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. doi:10.1001/jama.2013.281053.
99. Wright, J. H., Mishkind, M., Eells, T. D., & Chan, S. R. (2019). Computer-Assisted Cognitive-Behavior Therapy and Mobile Apps for Depression and Anxiety. *Current psychiatry reports*, 21(7), 62. doi:10.1007/s11920-019-1031-2.
100. Wright, J. H., Owen, J. J., Richards, D., Eells, T. D., Richardson, T., Brown, G. K., et al. (2019). Computer-Assisted Cognitive-Behavior Therapy for Depression: A Systematic Review and Meta-Analysis. *The Journal of clinical psychiatry*, 80(2). doi:10.4088/JCP.18r12188.
101. Zelviene, P., & Kazlauskas, E. (2018). Adjustment disorder: current perspectives. *Neuropsychiatric disease and treatment*, 14, 375–381. doi:10.2147/NDT.S121072.

## 14 Anhang

### 14.1 Mittelwerte, Standardabweichungen, Zellbesetzungen und Streudiagramme für die Faktoren der Sensitivitätsanalyse

Tabelle 14: Baseline-Depressions-Mittelwerte und -Standardabweichungen sowie Zellbesetzungen für die Faktoren der Sensitivitätsanalyse

			<i>M</i>	<i>SD</i>	<i>N</i>
Weiblich	Kein Universitätsabschluss	keine Zusatzbehandlung	12.27	4.148	15
		Zusatzbehandlung	14.00	5.099	4
		Total	12.63	4.271	19
	Universitätsabschluss	keine Zusatzbehandlung	10.59	4.455	41
		Zusatzbehandlung	8.83	3.545	6
		Total	10.36	4.356	47
	Total	keine Zusatzbehandlung	11.04	4.402	56
		Zusatzbehandlung	10.90	4.771	10
		Total	11.02	4.422	66
Männlich	Kein Universitätsabschluss	keine Zusatzbehandlung	8.43	5.884	7
		Total	8.43	5.884	7
	Universitätsabschluss	keine Zusatzbehandlung	11.83	3.738	12
		Zusatzbehandlung	7.50	0.707	2
		Total	11.21	3.786	14
	Total	keine Zusatzbehandlung	10.58	4.788	19
		Zusatzbehandlung	7.50	0.707	2
		Total	10.29	4.638	21
	Total	Kein Universitätsabschluss	keine Zusatzbehandlung	11.05	4.971
Zusatzbehandlung			14.00	5.099	4
Total			11.50	5.006	26
Universitätsabschluss		keine Zusatzbehandlung	10.87	4.301	53
		Zusatzbehandlung	8.50	3.071	8
		Total	10.56	4.217	61
Total		keine Zusatzbehandlung	10.92	4.474	75
		Zusatzbehandlung	10.33	4.519	12
		Total	10.84	4.459	87

**Tabelle 15:** Post-Depressions-Mittelwerte und -Standardabweichungen sowie Zellbesetzungen für die Faktoren der Sensitivitätsanalyse

			<i>M</i>	<i>SD</i>	<i>N</i>
Weiblich	Kein Universitätsabschluss	keine Zusatzbehandlung	8.00	4.259	15
		Zusatzbehandlung	10.00	5.477	4
		Total	8.42	4.451	19
	Universitätsabschluss	keine Zusatzbehandlung	6.49	4.313	41
		Zusatzbehandlung	6.50	4.930	6
		Total	6.49	4.338	47
	Total	keine Zusatzbehandlung	6.89	4.314	56
		Zusatzbehandlung	7.90	5.174	10
		Total	7.05	4.425	66
Männlich	Kein Universitätsabschluss	keine Zusatzbehandlung	6.29	5.559	7
		Total	6.29	5.559	7
	Universitätsabschluss	keine Zusatzbehandlung	5.25	2.417	12
		Zusatzbehandlung	5.00	4.243	2
		Total	5.21	2.517	14
	Total	keine Zusatzbehandlung	5.63	3.760	19
		Zusatzbehandlung	5.00	4.243	2
		Total	5.57	3.696	21
	Total	Kein Universitätsabschluss	keine Zusatzbehandlung	7.45	4.647
Zusatzbehandlung			10.00	5.477	4
Total			7.85	4.756	26
Universitätsabschluss		keine Zusatzbehandlung	6.21	3.978	53
		Zusatzbehandlung	6.13	4.518	8
		Total	6.20	4.012	61
Total		keine Zusatzbehandlung	6.57	4.192	75
		Zusatzbehandlung	7.42	4.981	12
		Total	6.69	4.287	87

## 14.2 Leitfaden Check-Up Calls

My7steps Pilotstudie– Leitfaden Check-Up Calls

Version 1.3

### Check-Up Calls

#### Hintergrund und Zweck der Check-Up Calls

Da die Vorstudie zur My7steps App komplett online abgewickelt wird, stehen wir vor der Herausforderung, trotz des Fehlens eines persönlichen Kontakts technische und organisatorische Probleme schnell zu erfassen und anstehende Termine mit den Probanden abzustimmen. Deswegen sind im Rahmen der Erhebung zweiwöchige Check-Up Calls vorgesehen, die den Probanden die Möglichkeit geben, organisatorische und technische Fragen zur App oder zum Studienablauf mit uns zu klären. Des Weiteren dient der letzte Anruf der Terminfindung für die zweite Studienvisite. Wichtig: Die Check-Up Calls dürfen NICHT für therapeutische oder medizinische Fragen oder zur Motivation der Probanden genutzt werden. Die Gespräche müssen sich auf die genannten organisatorischen und technischen Fragen beschränken und sollten in jedem Fall möglichst kurz gehalten werden. Im Rahmen des Onboardings werden die Probanden über die Check-Up Calls aufgeklärt, der rein organisatorische Zweck wird erläutert, und es wird ausdrücklich darauf hingewiesen, dass die Anrufe nicht Teil der Behandlung sind und keine therapeutischen Fragen besprochen werden dürfen. Die Probanden stimmen dieser Regelung zu – bitte erinnern Sie die Probanden an diese Zustimmung, falls abweichende Themen angesprochen werden. Kontraindizierte Situationen werden über die App gemonitort. Bei Rückfragen zu dieser Regelung wenden Sie sich bitte an die Studienleitung. Um den Zweck zu erfüllen und die Studienqualität nicht zu beeinträchtigen, bitten wir Sie sich strikt an folgenden Gesprächsleitfaden zu halten. Vielen Dank!

#### Gesprächsleitfaden für die Check-Up Calls

1. Vorstellung und Begrüßung
2. Erinnerung an den Zweck des Anrufs: „Dieser Anruf dient ausschließlich der Klärung von organisatorischen und technischen Fragen.“
3. Fragen:
  - a. „Haben oder hatten Sie technische Schwierigkeiten bei der Verwendung der App?“
  - b. „Haben Sie Fragen zum weiteren Studienablauf?“
4. Wenn Schritt 6 (Step 6) der Behandlung abgeschlossen wurde:
  - a. „Wäre es möglich, dass wir jetzt einen Termin für die zweite Studienvisite vereinbaren?“
  - b. Wiederholung des vereinbarten Termins
5. Danksagung und Verabschiedung

#### Dokumentation der Check-Up Calls

Jeder Checkup-Call muss im „Counselor Protokoll“ des Falls händisch dokumentiert werden.

Berichte über technische Schwierigkeiten geben Sie bitte umgehend per E-Mail an die Studienleitung weiter.