

## **Studienbericht Novego Depression**

Titel der klinischen Studie:

*Untersuchung der Wirksamkeit eines Online-basierten Unterstützungsprogramms zur Behandlung von Depression (Novego – „Depressionen bewältigen“<sup>®</sup>) – Eine randomisiert-kontrollierte Studie*

### **Name und Kontaktangaben des Sponsors oder des Vertreters des Sponsors:**

Herr Dr. Norbert Paas, IVPNetworks GmbH, Conventstr. 8-10  
22089 Hamburg Deutschland

### **Identifizierung des Studienplans:**

DRKS00027459

### **Studienleiter:**

Hr. Prof. Dr. Steffen Moritz, Universitätsklinikum Hamburg-Eppendorf, Martinistraße 52,  
20246 Hamburg, Deutschland

Operative Studiendurchführung: Fr. Anna Baumeister, Universitätsklinikum Hamburg-Eppendorf,  
Martinistraße 52, 20246 Hamburg, Deutschland

### **Ethikvotum:**

Positives Votum/Zustimmende Bewertung durch die Lokale Psychologische Ethikkommission des  
Universitätsklinikums Hamburg-Eppendorf (Ethikkommission Vorlage-Nr.: LPEK-0405)

### **Datum des Berichts:**

07. September 2023

### **Verfasser des Berichts:**

Dipl.-Psych. Anna Baumeister und Prof. Dr. Steffen Moritz

## Inhalt

1. Zusammenfassung .....	5
1.1 Einleitung .....	5
1.2 Zweck der Studie.....	5
1.3 Beschreibung der untersuchten Population .....	5
1.4 Ablauf.....	5
1.5 Ergebnisse .....	5
1.6 Schlussfolgerung .....	6
2. Einleitung .....	6
3. Studienprodukt .....	6
4. Studienplan.....	8
4.1 Zielstellungen .....	8
4.2 Design.....	8
4.3 Erhebungen und Endpunkte der klinischen Studie.....	8
4.4 Studienpopulation.....	10
4.5 Rekrutierung und Randomisierung.....	11
5. Statistische Analysen.....	12
5.1 mITT-Analysen .....	15
5.2 Sensitivitätsanalysen .....	16
6. Ergebnisse .....	16
6.1 Stichprobenbeschreibung .....	16
6.2 Wiedererreichungsquote.....	19
6.3 Interventionsnutzung und -zufriedenheit.....	19
6.4 mITT- Analysen .....	19
6.5 Sensitivitätsanalysen .....	25
6.5.1 J2R-Analysen .....	25
6.5.2 CC-Analysen .....	27
6.5.3 PP-Analysen .....	28
6.6 Klinische Relevanz.....	28
6.7 Sicherheit.....	31
6.8 Moderationsanalysen.....	33
7. Diskussion .....	34
Literatur .....	36
Anhang .....	40
A Baseline-Charakteristika der ITT-Population.....	40
B Ergebnisse J2R.....	42
C Ergebnisse ITT mit MMRM und MI .....	43
D Sensitivitätsanalysen der ITT-Population (CC- und PP-Analysen).....	48
E Moderation mit ITT-Population.....	50

## Abkürzungsverzeichnis

ANCOVA	Kovarianzanalyse
AV	Abhängige Variable
BDI-II	Beck-Depressionsinventar-II; Messinstrument zur Erhebung des Schweregrades der depressiven Symptomatik
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
BPtK	Bundespsychotherapeutenkammer
CAU	Care as usual
CC	Complete Cases; Analyse vollständiger Datensätze
DiGA	Digitale Gesundheitsanwendung
DiGAV	Digitale Gesundheitsanwendungen-Verordnung
DRKS	Deutsches Register Klinischer Studien
DSGVO	Datenschutzgrundverordnung
DVG	Digitale-Verordnung-Gesetz
EMA	European Medicines Agency
FDA	Food and Drug Administration
IBI	Internetbasierte Intervention
IG	Interventionsgruppe
ITT	Intention to Treat; Analyse von allen Datensätzen nach Imputation fehlender Daten
J2R	Jump to Reference; Imputationsmethode
KI	Konfidenzintervall
LS Means	Least squares mean differences
MAR	missing at random; Annahme über fehlende Daten
MCMC	Markov Chain Monte Carlo Methode
MI	multiple Imputation; Imputationsmethode
mITT	modifizierte Intention to Treat; modifizierte Analyse von allen plausiblen Datensätzen nach Imputation fehlender Daten
MMRM	mixed effects model for repeated measures
MNAR	missing not at random; Annahme über fehlende Daten
MVZ	Medizinisches Versorgungszentrum

PANEPS-I	Positive and Negative Effects of Psychotherapy Scale for Internet-based Interventions; Messinstrument zur Erfassung von positiven und negativen Effekten der Intervention
PHQ-9	Patient Health Questionnaire - 9 items depression module
PP	Per Protocol; Analyse von Datensätzen derjenigen Teilnehmer:innen, von denen T0 und T2 vorliegt; Zusatzkriterium für Interventionsgruppe: haben sich mind. viermal in das Programm eingeloggt
RSES	Rosenberg Self Esteem Scale; Messinstrument zur Erfassung des Selbstwertes
TEX-Q	Treatment Expectation Questionnaire; Messinstrument zur Erfassung der Behandlungserwartung
UV	Unabhängige Variable
WHOQOL-100	World Health Organization Quality of Life - 100
WHO-QOL-BREF	World Health Organization Quality of Life Questionnaire; Messinstrument zur Erfassung der Lebensqualität
WKG	Wartekontrollgruppe
WSQ	Web Screening Questionnaire; Messinstrument zur Erfassung häufiger psychischer Störungen
ZUF-8	Fragebogen zur Patientenzufriedenheit; Messinstrument zur Erhebung der Zufriedenheit mit der Behandlung nach Beendigung der Behandlung

## 1. Zusammenfassung

### 1.1 Einleitung

Depressionen zählen zu den häufigsten psychischen Erkrankungen in Deutschland. Sie gehen mit einer hohen Krankheitslast einher und trotz wirksamer Behandlungsmethoden wie kognitiver Verhaltenstherapie, bleiben ca. ein Drittel der Betroffenen unversorgt. Online-basierte Programme wie Novego – „Depressionen bewältigen“<sup>®</sup> versprechen, diese Behandlungslücke zu adressieren. Das Programm wurde in der vorliegenden randomisiert-kontrollierten Studie untersucht hinsichtlich seiner Wirksamkeit auf die Reduktion depressiver Symptombelastung (primärer Endpunkt), Erhöhung des Selbstwertes und Steigerung der Lebensqualität (sekundäre Endpunkte). Dabei wurde die Nutzung des Programms über 12 Wochen gegen eine Wartekontrollbedingung verglichen.

### 1.2 Zweck der Studie

Der positive Versorgungseffekt des Programms wurde mit der vorliegenden Studie im Rahmen des Prüfverfahrens beim Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) zur Aufnahme in das Verzeichnis für Digitale Gesundheitsanwendungen (DiGA) überprüft. Der angestrebte medizinische Nutzen des Programms umfasst die Verbesserung des Gesundheitszustandes, Verbesserung des Selbstwertes und Verbesserung der Lebensqualität.

### 1.3 Beschreibung der untersuchten Population

In die Studie wurden Personen mit einer depressiven Störung mit milder bis schwerer Symptomatik zwischen 18 und 75 Jahren eingeschlossen, die zum Zeitpunkt des Einschlusses einen Summenwert von mindestens 5 im Patient-Health-Questionnaire - 9 items depression module (PHQ-9) aufwiesen.

### 1.4 Ablauf

Bei der Studie handelte es sich um eine zweiarmige randomisiert-kontrollierte Studie mit einer Interventionsgruppe (IG) und einer Wartekontrollgruppe (WKG). Die Teilnehmenden der IG erhielten Zugang zu dem Online-Unterstützungsprogramm Novego – „Depressionen bewältigen“<sup>®</sup>. Die WKG erhielt erst nach Abschluss des Interventionszeitraumes (12 Wochen) Zugang zum Programm. Novego – „Depressionen bewältigen“<sup>®</sup> basiert auf der deutschen S3-Richtlinie für die Behandlung unipolarer Depression. Außerdem hatten beide Studiengruppen uneingeschränkter Zugang zur Regelversorgung. Die Erhebungen erfolgten zum Baseline-Messzeitpunkt, zum Interims-Messzeitpunkt (6 Wochen nach Baseline) und zum Post-Messzeitpunkt (12 Wochen nach Baseline).

### 1.5 Ergebnisse

Die Ergebnisse zeigen in der relevanten (modifizierten) Intention-to-treat-Analyse (mITT) eine signifikant stärkere Reduktion in der Depressionssymptomatik über 12 Wochen. Erste positive Ergebnisse in Bezug auf die Verbesserung der depressiven Symptomatik konnten bereits nach 6 Wochen festgestellt werden. Die Ergebnisse konnten im Rahmen von Sensitivitätsanalysen bestätigt werden.

## 1.6 Schlussfolgerung

Das Programm kann in der Gesamtbetrachtung in Bezug auf die Reduktion von depressiven Symptomen bei Patient:innen mit einer gesicherten, primären Depressionsdiagnose gegenüber einer Wartekontrolle als überlegen erachtet werden.

## 2. Einleitung

Am 19. Dezember 2019 trat das Digitale-Versorgung-Gesetz (DVG) in Kraft. Damit besteht für Patient:innen nun auch rechtlich ein Anspruch auf die Versorgung mit DiGA, die von Behandler:innen verordnet und von den Krankenkassen erstattet werden („App auf Rezept“). DiGA bieten besonders im Rahmen von Psychiatrie und Psychotherapie die Chance, die bestehende Versorgungslücke (v.a. Wartezeiten auf Behandlungsbeginn von durchschnittlich fast 5 Monaten laut der Bundespsychotherapeutenkammer; BPtK, 2018) zu adressieren und den Patient:innen so schnelle, niedrighschwellige und erstattungsfähige Unterstützung anzubieten. Internetbasierte Interventionen (IBI) können aber ebenso in sogenannten „blended“ (engl., „vermischt“) Ansätzen angewendet werden, bei denen den Patient:innen zusätzlich zur Therapeut:innen-vermittelten (face-to-face) Therapie die IBI zur Verfügung steht (Erbe et al., 2017). Der *blended* Ansatz kann dabei helfen, Therapieabbrüche zu verringern, im Rahmen der Therapie Zeit zu sparen (indem Psychoedukation auf die IBI verlagert wird) und die Effekte der Therapie zu stabilisieren (Erbe et al., 2017). Das Programm Novego – „Depressionen bewältigen“<sup>®</sup> beabsichtigt, eine solche Unterstützung für Patient:innen mit Depression zu sein.

Das Programm Novego – „Depressionen bewältigen“<sup>®</sup> wurde zuvor in drei Studien mit unterschiedlichen Patient:innengruppen untersucht (Beiwinkel et al., 2017; Miegel et al., 2019; Moritz et al., 2016). Beiwinkel et al. (2017) untersuchten das Programm Novego - „Depressionen bewältigen“<sup>®</sup> in einer 2013 bis 2014 durchgeführten Studie unter dem damaligen Namen „HelpID“ an einer Stichprobe von 180 Patient:innen mit milder bis moderater Depression. Dabei zeigte sich eine signifikante Verbesserung der depressiven Symptome gegenüber der Kontrollgruppe. Miegel et al. (2019) konnten in einer Untersuchung von Schmerzpatient:innen mit im Mittel moderaten, sekundären depressiven Symptomen und Moritz et al. (2016) an einer Stichprobe von 58 schizophrenen Patient:innen (moderate bis schwere depressive Symptome) ebenfalls signifikante Verbesserungen der depressiven Symptome im Vergleich zur Kontrollgruppe nachweisen. Die vorliegende Studie untersuchte das Programm Novego – „Depressionen bewältigen“<sup>®</sup> an einer Stichprobe von erwachsenen Personen mit einer gesicherten primären, milden bis schweren depressiven Störung ohne psychotische Symptome (F32.x, F33.x [ohne F32.3 und F33.3], F34.1).

## 3. Studienprodukt

Das Programm Novego – „Depressionen bewältigen“<sup>®</sup> basiert auf etablierten Methoden der kognitiven Verhaltenstherapie sowie auf Ansätzen der systemischen Therapie und der Achtsamkeitstherapie. Es besteht aus zwölf wöchentlichen Modulen, deren Bearbeitungsdauer jeweils zwischen ca. 45 und 60 Minuten liegt. Die Module umfassen psychoedukative Texte, Audiodateien, Video-Moderationen,

Übungen, Illustrationen und Bilder sowie mehrere Tagebuch-Vorlagen. Darüber hinaus können die Nutzer:innen wählen, ob sie motivierende E-Mail- und SMS-Erinnerungen erhalten möchten. Das gesamte Programm wird den Teilnehmer:innen noch ein Jahr nach Abschluss der regulären Laufzeit von 12 Wochen zur Verfügung gestellt und viele Arbeitsmaterialien, Dokumente und Übungen können heruntergeladen werden, sodass sie langfristig genutzt werden können.

Die Titel der Module lauten wie folgt:

- (1) Der Weg aus der Depression,
- (2) Depression verstehen,
- (3) Angenehmes im Alltag,
- (4a) Aktivität entdecken,
- (4b) Sich selbst erkennen,
- (5) Gedankenspiralen durchbrechen,
- (6) Gemeinsam gegen Depression,
- (7a) Entspannung gegen Depressionen,
- (7b) Kognitive Fähigkeiten stärken,
- (7c) Einfluss von Licht auf das Wohlbefinden,
- (8) Umgang mit negativen Gefühlen,
- (9a) Das gönne ich mir,
- (9b) Mein Herz und ich,
- (10) Auf den Körper hören,
- (11) Vor Rückfällen schützen,
- (12) Therapeutische Unterstützung.

Anhand eines Eingangsfragebogens werden aus diesem Pool von 16 Modulen 12 Module automatisch zusammengestellt. Auch die Inhalte sowie die Bilder werden individuell zusammengestellt und auf die persönliche Situation der Teilnehmer:innen zugeschnitten. Die Zusammensetzung erfolgt beispielsweise anhand von Angaben zur Aktivität (weniger aktive Teilnehmer:innen erhalten Modul 4a „Aktivität entdecken“, aktive Teilnehmer:innen erhalten Modul 4b „Sich selbst erkennen“), zur Konzentration oder zu Begleiterkrankungen wie Herzerkrankungen oder chronischen Schmerzen. Zudem wird u.a. die Ansprache der Teilnehmer:innen daran angepasst, welche Erwartungen sie an das Programm haben, und Beispiele werden auf Alter und Geschlecht der Teilnehmer:innen zugeschnitten.

Das Programm Novego – „Depressionen bewältigen“<sup>®</sup> wird mit Unterstützung von einer telefonischen Erinnerungsfunktion, einer schriftlichen Begleitung für Verständnisfragen und einer Krisenhotline angeboten.

## 4. Studienplan

Die vorliegende Studie wurde von der Lokalen Psychologischen Ethikkommission des Universitätsklinikums Hamburg-Eppendorf geprüft und als ethisch unbedenklich eingeschätzt (Vorlage-Nr.: LPEK-0405). Die Studie wurde im Deutschen Register Klinischer Studien (DRKS) registriert: DRKS00027459 (Registrierungsdatum: 23.12.2021). Datum des Beginns der Studie (first patient in) war der 06.01.2022, Datum des Abschlusses (last patient out) war der 06.06.2023.

### 4.1 Zielstellungen

Das primäre Ziel der Studie war es, die Wirksamkeit der Intervention Novego – „Depressionen bewältigen“<sup>®</sup> zur Behandlung depressiver Symptome über 12 Wochen im Vergleich zu einer WKG zu untersuchen.

### 4.2 Design

Bei der vorliegenden Studie handelte es sich um eine zweiarmige randomisierte kontrollierte Studie mit einer IG und einer WKG. Die Erhebungen fanden online über die Umfragesoftware Qualtrics<sup>®</sup> statt. Die Randomisierung erfolgte automatisch am Ende der Baseline-Erhebung durch einen Randomisierungsalgorithmus der Software. Es wurden insgesamt 311 Personen randomisiert.

### 4.3 Erhebungen und Endpunkte der klinischen Studie

Die Messzeitpunkte lagen wie folgt:

- T0: Baseline-Erhebung
- T1: 6 Wochen nach Baseline (Interims-Erhebung)
- T2: Primärer Messzeitpunkt, 12 Wochen nach Baseline (Post-Erhebung)

Als primärer Endpunkt zur Überprüfung des medizinischen Nutzens der Intervention wurde die Veränderung des Schweregrads der depressiven Symptome von T0 zu T2 mit Hilfe des Beck-Depressionsinventars II (BDI-II; Beck, Steer & Brown, 1996) festgelegt. Der BDI-II ist ein valides und reliables Instrument zur Erfassung depressiver Symptome (García-Batista et al., 2018, Subica et al., 2014). Es umfasst 21 Items, über die ein Summenwert gebildet wird. Der Summenwert kann zwischen 0 und 63 liegen, wobei höhere Werte eine höhere Symptombelastung anzeigen (0-8 keine Depression, 9-13 minimale Depression, 14-19 leichte Depression, 20-28 mittlere Depression und 29-63 schwere Depression).

Als sekundäre Endpunkte dienten die folgenden Instrumente:

*PHQ-9*. Der PHQ-9 (Gräfe et al., 2004; Kroenke et al., 2001) erhebt den depressiven Schweregrad in Selbstauskunftsfragebogen über die letzten zwei Wochen mit neun Items, die auf einer vierstufigen Likert-Skala beantwortet werden. Dieser zu bildende Summenwert zwischen 0 und 27 annehmen, wobei höhere Werte eine höheren Depressionsschwere anzeigen. Kategorisiert wird in keine oder minimale (0-4), milde (5-9), mittlere (10-14) und schwere (15-27) Depression. Der PHQ-9 weist eine hohe interne

Konsistenz auf (Cronbach's  $\alpha = .86-.89$ ; Kroenke et al., 2001). Die Veränderung des Summenwertes des PHQ-9 zwischen T0 und T2 wird in dieser Studie als sekundärer und zwischen T0 und T1 als explorativer Endpunkt verwendet, um die im primären Endpunkt mit dem BDI-II erhobene depressive Psychopathologie zu bestätigen und die Veränderung nach der hälftigen Interventionszeit explorativ zu erfassen.

*Rosenberg Self-Esteem Scale (RSES)*. Die RSES (Rosenberg, 1965) erhebt in Selbstauskunft das Selbstwertgefühl über 10 Items, die auf einer vierstufigen Likert-Skala von (1) „stark zustimmend“ bis (4) „stark ablehnend“ beantwortet werden. Die Ausprägung des Selbstwertgefühls wird über den Summenwert ermittelt, der entsprechende Werte zwischen 10 und 40 annehmen kann. Ein höherer Summenwert zeigt dabei ein stärkeres Selbstwertgefühl an. Die interne Konsistenz der Skala (Cronbach's  $\alpha = .77$  bis  $.88$ ) ist als gut zu bewerten (Rosenberg, 1965).

*World Health Organization Quality of Life – BREF (WHO-QOL-BREF)*. Der WHO-QOL-BREF (Whoqol Group, 1998) erhebt mit 26 Items die subjektive Lebensqualität. Er stellt die Kurzversion des Fragebogens World Health Organization Quality of Life – 100 (WHOQOL-100) dar und umfasst vier Subskalen (psychisches Wohlbefinden, physisches Wohlbefinden, soziale Beziehungen und Umwelt). Das Antwortformat ist eine fünfstufige Likert-Skala (1 bis 5). Für die vier Subskalen können jeweils Mittelwerte gebildet werden, wobei höhere Werte für eine bessere Qualität stehen. Über alle Items hinweg kann durch Umrechnung ( $(\text{Subskalenmittelwert}-4) \times (100/16)$ ) ein Gesamtwert zwischen 0 und 100 gebildet werden, wobei höhere Werte für eine bessere Lebensqualität stehen. Die interne Konsistenz (Cronbach's  $\alpha$  zwischen  $.68$  und  $.82$  für die vier Subskalen) ist akzeptabel bis gut (Skevington et al., 2004).

Die explorativen Endpunkte sind Tabelle 1 zu entnehmen.

**Tabelle 1**

Explorative Ziele und Endpunkte der Studie mit Beschreibung der Messinstrumente.

<b>Ziel</b>	<b>Endpunkt</b>	<b>Beschreibung</b>
Überprüfung von Effekten auf die depressive Symptomatik nach der Hälfte des Interventionszeitraumes	Verringerung der depressiven Symptomatik (Summenwert BDI-II und PHQ-9) zu T1 im Vergleich zu T0	<i>s. Beschreibungen der Instrumente</i>
Überprüfung des Einflusses der Behandlungserwartung auf die Veränderung der depressiven Symptome	Mittelwert der positiven Behandlungserwartung zum Zeitpunkt T0, gemessen durch den TEX-Q (Alberts et al., 2020)	<i>Der TEX-Q ist ein valides Instrument zur Erfassung der Behandlungserwartung (Alberts et al., 2020)</i>

<b>Ziel</b>	<b>Endpunkt</b>	<b>Beschreibung</b>
Überprüfung von nicht intendierten positiven und negativen Effekten (auch Nebenwirkungen) der Intervention	Häufigkeit weiterer positiver und negativer Effekte (einschließlich Nebenwirkungen) der Intervention zum Zeitpunkt T2, gemessen über die PANEPS-I (Baumeister & Moritz, 2022)	<i>Die PANEPS-I ist das bisher einzige Instrument, das Nebenwirkungen spezifisch für IBI erfragt</i>
Überprüfung des Einflusses von komorbiden psychischen Störungen auf die Veränderung depressiver Symptome	Vorliegen von weiteren, häufigen psychischen Störungen zum Zeitpunkt T0, erhoben durch den WSQ (Donker et al., 2009)	<i>Der WSQ ist ein kurzes, valides Screening-Instrument zur Erfassung von Komorbiditäten (Donker et al., 2009)</i>
Überprüfung der Zufriedenheit der Nutzer:innen der Intervention	Summenwert der Zufriedenheit der Teilnehmer:innen mit dem Programm zum Zeitpunkt T2, erhoben über den ZUF-8 (Schmidt et al., 1989)	<i>Der ZUF-8 ist ein ökonomisches und valides Verfahren (Cronbach's <math>a = .88-.92</math>; Kriz et al., 2008) zur Erfassung der Patientenzufriedenheit und liegt bereits in für internetbasierte Interventionen adaptiert vor</i>

Die Eignung der o. g. Auswahl primärer und sekundärer Endpunkte im spezifischen Indikationsgebiet wurde im Rahmen einer Beratung nach § 23 DiGAV durch das BfArM bestätigt (BfArM, 24.03.2021).

#### 4.4 Studienpopulation

Kriterien für den Einschluss in die Studie waren

- a) eine der folgenden, durch Fachärzt:innen oder Psychotherapeut:innen gestellten Diagnosen nach ICD- 10: F32.0, F32.1, F32.2, F32.8, F32.9, F33.0, F33.1, F33.2, F33.8, F33.9, F34.1 (Diagnosen aus dem depressiven Formenkreis ohne psychotische Symptome),
- b) ein Summenwert im PHQ-9 zum Zeitpunkt der Baseline-Befragung von mindestens 5 (mindestens milde Ausprägung einer Depression),
- c) ein Alter zwischen 18 und 75 Jahren,
- d) die Einwilligung zur Studienteilnahme nach Aufklärung,
- e) Zugang zum Internet,
- f) ausreichende Kenntnisse der deutschen Sprache,

- g) die Bereitschaft zur Teilnahme an drei anonymen Online-Befragungen sowie zur selbstständigen Nutzung von Novego – „Depressionen bewältigen“<sup>®</sup>,
- h) Patient:innen müssen mindestens 4 Wochen vor Randomisierung eine stabile Therapiesituation aufweisen (Pharmakotherapie und/oder Psychotherapie oder Wartesituation, d.h. keine aktive Therapie) und sollten diese über den Studienzeitraum möglichst beibehalten.

Kriterien für den Ausschluss von der Teilnahme waren

- a) das Vorhandensein einer psychotischen oder bipolaren Störung (Lebenszeit),
- b) akutes suizidales Verhalten innerhalb der letzten 4 Wochen,
- c) Demenz,
- d) Alkohol- oder Substanzabhängigkeit innerhalb der letzten 6 Monate,
- e) aktuelle Einnahme von hochpotenten und/oder atypischen Antipsychotika,
- f) Vorliegen einer schwerwiegenden Erkrankung, die nach Ansicht des/der Prüfarztes/Prüfärztin eine Teilnahme des/der Patienten/Patientin an der Studie bis zum Studienende ausschließt.

Die Eignung der o. g. Ein- und Ausschlusskriterien wurde im Rahmen einer Beratung nach § 23 DiGAV durch das BfArM bestätigt (BfArM, 24.03.2021).

#### 4.5 Rekrutierung und Randomisierung

Die Rekrutierung erfolgte über Kooperationen mit dem Medizinischen Versorgungszentrum (MVZ) Falkenried, dem MVZ am Ostebogen und dem MVZ Colonnaden. Die Patient:innen erhielten durch standardisierte Interviews mit Fachärzt:innen und/oder Psychotherapeut:innen eine nach ICD-10 kodierte Diagnose. Darunter waren sowohl Patient:innen, die bereits eine Behandlung in Anspruch nahmen, als auch Personen, die nur diagnostisch im MVZ vorstellig wurden und zum entsprechenden Zeitpunkt noch keine Therapie erhalten hatten bzw. auf diese warteten. Die E-Mail-Adressen und Diagnoseschlüssel von Patient:innen, die studieninteressiert waren und eingewilligten, über Studien entsprechend ihrer Diagnose informiert zu werden, wurden an die Studienmitarbeiter:innen weitergeleitet (DSGVO-konform). Patient:innen mit einer den Einschlusskriterien entsprechenden Diagnose wurden per E-Mail angeschrieben und zur Studienteilnahme eingeladen. Außerdem wurde studieninteressierten Patient:innen durch Fachärzt:innen und/oder Psychotherapeut:innen der MVZ ein Informationsblatt mit einem Online-Link zur Studie ausgehändigt. Die E-Mail und auch das Informationsblatt enthielten grundlegende Studieninformationen sowie einen Link zur informierten Einwilligungserklärung (online). Wurde der Teilnahme zugestimmt, gelangte der/die Teilnehmende direkt zur Baseline-Befragung. Wurde der Teilnahme nicht zugestimmt, erfolgte automatisch ein Abbruch der Online-Befragung.

Geeignete Personen, die die Baseline-Befragung abschlossen, wurden am Ende dieser automatisch über einen Algorithmus der Umfragesoftware Qualtrics<sup>®</sup> einer der beiden Gruppen (IG oder WKG) im

Verhältnis 1:1 zugeordnet. Teilnehmende der IG erhielten daraufhin eine E-Mail mit ihrem persönlichen Zugangscode für das Interventionsprogramm sowie einer Anleitung, wie dieser eingelöst werden kann. Teilnehmende der WKG wurden erneut darüber informiert, dass sie im Anschluss an den Studienzeitraum (12 Wochen) Zugang zum Programm erhalten. Die Teilnehmenden beider Studiengruppen hatten außerdem auch während des Studienzeitraumes uneingeschränkten Zugang zur Regelversorgung.

Das Studienprotokoll sah ursprünglich eine Rekrutierung von 200 Patient:innen vor. Im Verlauf der Studie wurde eine erneute Fallzahlkalkulation durchgeführt, um einer wahrscheinlichen Abbruchquote von 20% Rechnung zu tragen und um mehr statistische Sicherheit für weitere Moderationsanalysen zu gewinnen. Basierend auf einem angenommenen Effekt von  $d = 0,4$  wurde die Fallzahlkalkulation auf der um die erwartete Anzahl fehlender Werte reduzierten Effektstärke von  $d^* = d^*(1-0,2) = 0,32$  durchgeführt. Hieraus ergab sich eine Fallzahl von  $N = 310$ , d.h.  $n = 155$  pro Behandlungsarm und eine entsprechende Anpassung des Erprobungszeitraums in Abstimmung mit dem BfArM (BfArM, i.A. Waider, 2022). Die Rekrutierungswege sowie der Ablauf der Studie änderten sich dabei nicht und verliefen weiter protokollkonform.

## 5. Statistische Analysen

In der vorliegenden Studie wurde die Hypothese verfolgt, dass die IG gegenüber der WKG eine signifikant stärkere Verbesserung nach 12 Wochen sowohl im primären Endpunkt (Veränderung im Summenwert des BDI-II) als auch in den sekundären Endpunkten (Veränderung in den Summenwerten des PHQ-9 und RSES und des Mittelwerts des WHO-QOL-BREF) aufweist.

Für die Analysen wurden blind zu den Ergebnissen und der Gruppenzuweisung 8 Teilnehmer:innen (WKG  $n = 5$ , IG  $n = 3$ ) ausgeschlossen, da sie zu T2 unplausible Angaben bezüglich ihrer depressiven Symptomatik machten. Als unplausibel wurde es gewertet, wenn die z-standardisierten T2-Scores der beiden Depressionsmaße BDI-II und PHQ-9 mindestens 1,5 SD voneinander abwichen<sup>1</sup> sowie die Veränderung im BDI-II mind. 1 SD von der mittleren Veränderung abwich. Bei diesen Teilnehmer:innen zeigten sich konträre Verläufe mit Verbesserungen in der einen und Verschlechterungen in der anderen Depressionsskala. Die nachfolgenden Ergebnisse werden ohne diese 8 Teilnehmer:innen als mITT dargestellt und diskutiert. Die Ergebnisse inklusive der 8 Teilnehmer:innen sind als Intention to Treat- (ITT) Analysen im Anhang A - E dargestellt.

Die primäre Analyse als mITT wurde mit dem mixed-effects model for repeated measures (MMRM) sowie zusätzlich mit multipler Imputation (MI) durchgeführt. Dies entspricht nicht der präspezifizierten Imputation durch die Jump-to-Reference-Methode (J2R). Klassischerweise nutzen psychologische

---

<sup>1</sup> Eine große Abweichung zwischen dem BDI-II- und dem PHQ-9-Score ist als unplausibel zu werten, da beide validierten und weit verbreiteten Instrumente dasselbe Konstrukt messen und hoch miteinander korrelieren ( $r = .72-.73$ ,  $p < .001$ ; Titov et al., 2011)

Wirksamkeitsstudien multiple Imputationen unter der Annahme fehlender Daten als „missing at random“ (MAR). Unter dieser Prämisse folgen die fehlenden Datenpunkte, abhängig von beobachteten Variablen, der Verteilung der beobachteten Datenpunkte und lassen sich daher durch multiple Imputation auf deren Basis schätzen (Little & Rubin, 1978). Die Vorgabe zur referenzbasierten multiplen Imputation unter Verwendung der J2R-Methode trifft eine überaus strikte Annahme zur Entstehung fehlender Daten, die auch nach Aussage eines der Autoren der Methode (s.u.) nicht immer gegeben ist, wie auch in diesem Fall. Wenn davon auszugehen ist, dass die fehlenden Datenpunkte sich aufgrund unbeobachteter Variablen anders verhalten als die beobachteten Datenpunkte, spricht man von „missing not at random“ (MNAR). Im Gegensatz zu gängigen Methoden der multiplen Imputation stützt sich die J2R-Methode auf die Annahme fehlender Daten als MNAR und nicht als MAR. Für den Kontext randomisierter kontrollierter Studien geht dieser Algorithmus davon aus, dass Personen der IG nach dem Drop-out dem Profil der Kontrollgruppe folgen (Carpenter et al., 2013). Dies macht es zu einem geeigneten Instrument für Sensitivitätsanalysen, ist vor dem Hintergrund der Wirksamkeitsprüfung der vorliegenden psychologischen Intervention jedoch kritisch zu betrachten.

Gemäß dieser Hypothese würde eine Person, welche für die Datenerhebung aus unbekanntem Gründen nicht mehr zur Verfügung steht, sich immer eher dem Profil einer Person der Kontrollgruppe annähern als einer ähnlichen Person der ursprünglich zugeteilten Gruppe (Carpenter et al. 2013, Kenward, 2015). Die Konsequenz ist eine Imputation des fehlenden Wertes ohne Berücksichtigung möglicher oder bereits beobachteter Verbesserungen durch die Intervention. Diese Überlegung mag vor dem Hintergrund medikamentöser Interventionen mit kurzfristigen Wirkprofilen durchaus sinnvoll sein, jedoch lässt sich dies nicht analog auf psychologische Interventionen übertragen. Es ist davon auszugehen, dass Personen auch nach dem vorzeitigen Ausscheiden aus der Studie von der Intervention weiter partiell profitieren, welche sie absolviert haben. Vor diesem Hintergrund erscheint es nicht plausibel, davon auszugehen, den Verlauf dieser Proband:innen dem der Kontrollgruppe anzugleichen.

Dazu schreibt Kenward<sup>2</sup> (2015, S. 316):

„Such a change may be seen as extreme: the patient immediately reverts, following dropout, to the reference profile, losing any gain that might have been made under the randomized active treatment. Hence, using the reference group in this way might be used as a worst-case scenario in terms of reducing any treatment effect because patients on active treatment who have withdrawn will lose the effect of their period on active treatment. It may well not be an appropriate scenario for treatments with a long lasting, or even permanent, impact.

Hence, this scenario might be considered for treatments with only short-term effect.“

---

<sup>2</sup> Michael G. Kenward ist einer der Autoren des der J2R-Methode zugrundeliegenden Artikels (Carpenter et al., 2013)

Betrachtet man die Profile der Personen in der IG, welche vor Beendigung der Abschlussbefragung aus der Studie ausschieden ( $n = 47$ ), so zeigt sich, dass sich diese bis zu 42 Mal im Programm einloggen (Range 0-42,  $M = 7,87$ ,  $SD = 10,58$ ). Die bisherige Studienlage deutet daraufhin, dass psychologische (Online-) Interventionen durchaus einen Langzeiteffekt auf die depressive Symptomatik aufweisen (für eine Meta-Analyse s. Karyotaki et al., 2021). Auch bewegt sich der Prozentsatz der Dropouts in einem für Onlineinterventionsstudien gängigen Bereich (Karyotaki et al., 2015), so dass nicht davon auszugehen ist, dass die Anzahl fehlender Daten im Zusammenhang mit der spezifischen Intervention und ihrem Nutzen für die Proband:innen steht. Daher ist abschließend zu sagen, dass im Fall einer psychologischen Intervention die J2R-Methode unter der MNAR-Annahme nicht angemessen erscheint, um den wahren Nutzen und die Wirksamkeit der Intervention zu beurteilen. Es liegt nahe, dass es so zu einer starken Unterschätzung des tatsächlichen Interventionseffekts käme. Viel mehr bildet sie ein „worst-case“ Szenario ab, dass eher als Sensitivitätsanalyse zu betrachten ist, wie sie auch ursprünglich von Carpenter et al. (2013) beschrieben wurde. Diesen Überlegungen und den Aussagen des BfArM folgend, dass es für die Imputation fehlender Werte unterschiedliche Möglichkeiten gibt, die Baseline- und Interimswerte heranzuziehen (BfArM, 24.03.2021), wurde die J2R-Methode als das primäre Analyseverfahren verworfen und durch das MMRM sowie MI ersetzt. Die Ergebnisse der J2R-Methode sind der Vollständigkeit halber als Sensitivitätsanalyse in Abschnitt 6.5.1 zu finden..

Die statistischen Analysen wurden durchgeführt mit IBM SPSS Statistics 27 sowie R Version 4.3.2. Baseline-Unterschiede zwischen der WKG und der IG wurden mit  $t$ -Tests für unabhängige Stichproben und Chi-Quadrat-Tests überprüft.

Um mögliche Variablen zu identifizieren, die einen Einfluss auf den Effekt haben, wurden Moderationsanalysen mithilfe des SPSS®-Makros PROCESS (entwickelt von Andrew. F. Hayes) durchgeführt.

Gemäß dem Studienprotokoll wurde die klinische Relevanz anhand von Hedges'  $g$  überprüft. Es ist anzumerken, dass „standardisierte Effektmaße in erster Linie rein statistische Konzepte sind, die zur Vergleichbarkeit von Studien mit unterschiedlichen Messinstrumenten herangezogen werden“ (BfArM, i.A. Moellers, 2023). Im Kontext von DiGA fehlen zum aktuellen Zeitpunkt die methodologischen Grundlagen, um die Bedeutung dieser Konzepte im Hinblick auf die Interpretation der Ergebnisse zu bewerten. Im Zuge einer Beratung nach § 23 DiGAV kommt das BfArM zu dem Schluss, dass das genannte Vorgehen zur Beurteilung der klinischen Relevanz „nicht notwendigerweise erforderlich“ ist (BfArM, 24.03.2021). Vor diesem Hintergrund soll für die valide Evaluation und Interpretation die gezeigte statistisch signifikante Reduktion von Symptomen der Depression dienen. Zusätzlich wurde post hoc die klinische Relevanz auf der Originalskala für den BDI-II und den PHQ-9 bewertet und zwischen den Gruppen verglichen (vgl. Abschnitt 6.6). Die gewählten Kriterien zur Beurteilung klinisch relevanter Verbesserung stimmen mit denen aus Studien einer bereits endgültig im DiGA-Verzeichnis aufgenommen Intervention (Deprexis) überein.

Um dem Multiplizitätsproblem entgegenzuwirken, wurde in Abstimmung mit dem BfArM ein hierarchisches Testen für den primären und die drei sekundären Endpunkte festgelegt (BfArM, 24.03.2021). Bei jedem Schritt wird ein Behandlungseffekt als statistisch signifikant betrachtet, wenn der  $p$ -Wert unter .05 (zweiseitig) liegt und alle vorherigen Tests ebenfalls dieses Signifikanzlevel erreichen. Die Testhierarchie wird wie folgt festgelegt:

1. Verringerung der depressiven Symptomatik (Summenwert) zu T2 im Vergleich zu T0, gemessen durch das BDI- II (Beck, Steer, & Brown, 1996)
2. Verringerung der depressiven Symptomatik (Summenwert) zu T2 im Vergleich zu T0, gemessen durch den PHQ-9 (Gräfe et al., 2004; Kroenke et al., 2001)
3. Erhöhung des Selbstwertes (Summenwert) zu T2 im Vergleich zu T0, gemessen durch die RSES (Rosenberg, 1965)
4. Erhöhung der Lebensqualität (Mittelwert) zu T2 im Vergleich zu T0, gemessen durch den WHO-QOL-BREF (WHOQOL Group, 1998)

### 5.1 mITT-Analysen

Als primäre Analysen wurden mITT durchgeführt. Als Verfahren wurde dazu, entsprechend den Empfehlungen für klinische Studien der Food and Drug Administration (FDA; National Research Council, 2010), das MMRM, sowie zusätzlich MI, folgend der Empfehlung zum Umgang mit fehlenden Daten der European Medicines Agency (EMA; EMA, 2011) verwendet.

Im MMRM galten Gruppe, Zeitpunkt und Gruppe $\times$ Zeitpunkt als feste Effekte und Teilnehmende als zufälliger Effekt. Im Modell wurden zusätzlich die jeweiligen Baseline-Werte als Kovariate berücksichtigt. Als Ergebnis der MMRM werden Mittelwertdifferenzen (*least squares mean differences, LS Means*; inklusive des 95%-Konfidenzintervalls, KI) angegeben.

Die MI, durchgeführt mit SPSS 27<sup>®</sup>, folgte der iterativen Markov Chain Monte Carlo (MCMC) Methode mit 10 Iterationen. Für jede Iteration und jede zu imputierende Variable passt die Methode ein univariates Modell mit allen anderen Variablen im Modell als Prädiktoren an und imputiert dann die fehlenden Werte für die anzupassende Variable. Die Methode wird fortgesetzt, bis die maximale Zahl an Iterationen erreicht ist. Es gingen Gruppenzuweisung, Alter und Bildungsniveau als Prädiktoren für die Imputation in das Modell ein, wobei 100 Imputationen durchgeführt wurden. Für die MI-Ergebnisse werden die kombinierten  $p$ -Werte der Regressionsanalysen angegeben.

Die Effektstärken der Analysen wurden unter Verwendung von Hedges'  $g$  berechnet, in Anlehnung an die Parameter von Hedges (Hedges & Olkin, 1985) für kleine ( $g \approx 0,2$ ), mittlere ( $g \approx 0,5$ ) und große ( $g \approx 0,8$ ) Effektstärken. Zusätzlich werden Mittelwertdifferenzen (*least squares mean differences, LS Means*; inklusive des 95%-Konfidenzintervalls, KI) angegeben.

Zusätzlich zur Gesamtstichprobe wurden Subgruppen gebildet, die explorativ mit MMRM für den primären Endpunkt (Veränderung Summenscore BDI-II T0-T2) und den sekundären Endpunkt

(Veränderung Summenscore PHQ-9 T0-T2) analysiert wurden. Die Subgruppen unterschieden dabei zwischen Geschlecht (männlich, weiblich), Diagnose (leicht: F32.0, F33.0, mittel: F32.1, F33.1, schwer: F32.2, F33.2, Dysthymie: F34.1, sonstige: F32/33.8/9) und Care as usual (CAU) (mit begleitender Psychotherapie, ohne begleitende Psychotherapie, mit begleitender Psychopharmakaeinnahme, ohne begleitende Psychopharmakaeinnahme). Die Subgruppe Alter >65 Jahre konnte nicht gebildet werden, da in die Studie 6 Teilnehmer:innen über 65 Jahre eingeschlossen wurden, davon wurde eine Teilnehmerin der IG zugewiesen. Bei dieser Teilnehmerin handelt es sich um einen Studiendropout zu T1 und T2. Die statistischen Verfahren finden bei einer Subgruppe von  $n = 1$ , noch dazu ohne vorliegende Werte, keine Anwendung. Der Vollständigkeit halber wird die Subgruppe  $\leq 65$  Jahre dennoch berichtet.

## 5.2 Sensitivitätsanalysen

Zur Evaluation der Robustheit der Ergebnisse wurden die Wirksamkeitsanalysen zusätzlich mit der ursprünglich präspezifizierten J2R-Imputation durchgeführt, sowie mit der Complete Cases (CC; Daten von Teilnehmenden mit vollständigen Datensets) und der Per Protocol Population (PP; Daten von Teilnehmenden mit vollständigen Datensets, für Teilnehmende der IG zusätzlich: die sich mindestens viermal in das Programm einloggten). Berechnungen des primären und der sekundären Endpunkte wurden über Kovarianzanalysen (ANCOVA) durchgeführt. Die Differenzwerte zwischen Baseline- und Post-Erhebung (12 Wochen nach Baseline) dienten dabei als abhängige Variablen. Die Baseline-Werte gingen dabei als Kovariaten mit in die Analyse ein, um für eine Regression zur Mitte zu kontrollieren. Effektstärken wurden als partielles Eta-Quadrat berechnet ( $\eta^2_{\text{partiell}} \approx 0,01$ : kleiner Effekt,  $\eta^2_{\text{partiell}} \approx 0,06$ : mittlerer Effekt,  $\eta^2_{\text{partiell}} \approx 0,14$ : großer Effekt). Zusätzlich werden Mittelwertdifferenzen (*least squares mean differences*, *LS Means*; inklusive des 95%-Konfidenzintervalls, KI) angegeben.

## 6. Ergebnisse

Insgesamt nahmen 358 Personen an der Baseline-Befragung teil. Zu T0 mussten 47 Personen ausgeschlossen werden, weil sie entweder die Einschlusskriterien nicht erfüllten (21 Personen mit einem PHQ-9-Score von unter 5) oder ein Ausschlusskriterium erfüllten (7 Personen mit bipolarer Störung, 5 Person mit psychotischer Störung, 7 Personen mit Substanzabhängigkeit und 7 Personen mit suizidalen Handlungsabsichten). Somit wurden 311 Teilnehmer:innen randomisiert. Die eingeschlossenen Teilnehmer:innen wurden zufällig entweder der WKG ( $n = 155$ ) oder der IG ( $n = 156$ ) zugeteilt. Post-hoc wurden 8 weitere Teilnehmer:innen (WKG  $n = 5$ , IG  $n = 3$ ) aufgrund unplausibler Antworten ausgeschlossen (vgl. Kapitel 5). Die Analysepopulation umfasste somit 303 Teilnehmer:innen.

### 6.1 Stichprobenbeschreibung

Die Stichprobe war überwiegend weiblich (71,9%), im Durchschnitt 40,7 Jahre alt und eher höher gebildet (62,0% mindestens Fachhochschulreife; s. Tab. 2). Die psychopathologische Untersuchung der Teilnehmer:innen zum Zeitpunkt der Baseline-Befragung zeigte durchschnittlich eine moderate depressive Symptombelastung (BDI-II, PHQ-9), einen mittleren Selbstwert-Score (RSES) sowie eine

mittlere Lebensqualität (WHO-QOL-BREF Globalwert; s. Tab. 2). Knapp 47% der Teilnehmer:innen befanden sich zum Zeitpunkt der Baseline-Befragung nicht in psychotherapeutischer Behandlung. Etwas mehr als 40% der Stichprobe (41,3%) nahmen regelmäßig Psychopharmaka ein, wobei sich die Anzahl derjenigen mit und ohne medikamentöse Behandlung in den beiden Gruppen signifikant voneinander unterschied ( $\chi^2(1) = 4,53, p = 0,033$ ; s. Tab. 2). Andere signifikante Gruppenunterschiede in den Baseline-Charakteristika gab es nicht (s. Tab. 2).

**Tabelle 2**

Häufigkeiten (Prozente), Mittelwerte (Standardabweichung) der soziodemografischen und psychopathologischen Variablen der mITT-Stichprobe zur Baseline-Erhebung.

Variable	Gesamt (N = 303)	WKG (n =150)	IG (n =153)	Statistik
<i>Demografische Charakteristika</i>				
Geschlecht (w/m/d)	218/81/4 (71,9/26,7/1,3%)	111/37/2 (74,0/24,7/1,3%)	107/44/2 (69,9/28,8/1,3%)	$\chi^2(2) = 0,65,$ $p = 0,723$
Alter in Jahren	40,68 (12,50)	40,54 (13,11)	40,82 (11,91)	$t(299) = 0,20,$ $p = 0,843$
Höhere Schulbildung (mind. Fachhochschulreife)	188 (62,0%)	88 (58,7%)	100 (65,4%)	$\chi^2(4) = 4,19,$ $p = 0,395$
<i>Psychopathologie</i>				
BDI-II	26,47 (9,52)	25,73 (10,08)	27,20 (8,68)	$t(301) = 1,36,$ $p = 0,175$
PHQ-9	13,53 (4,99)	13,03 (5,23)	14,03 (4,72)	$t(301) = 1,73,$ $p = 0,084$
RSES	14,19 (6,69)	14,53 (6,72)	13,86 (6,67)	$t(301) = 0,87,$ $p = 0,387$
WHO-QOL	39,56 (19,24)	41,42 (18,95)	37,75 (19,42)	$t(301) = 1,67,$ $p = 0,097$
<i>Diagnose Depression</i>				$\chi^2(4) = 2,72,$ $p = 0,606$
Leicht (F32.0, F33.0)	25 (8,3%)	15 (10,0%)	10 (6,5%)	
Mittel (F32.1, F33.1)	151 (49,8%)	73 (48,7%)	78 (51,0%)	
Schwer (F32.3, F33.3)	47 (15,5%)	23 (15,3%)	24 (15,7%)	

<b>Variable</b>	<b>Gesamt (N = 303)</b>	<b>WKG (n =150)</b>	<b>IG (n =153)</b>	<b>Statistik</b>
Dysthymie (F34.1)	21 (6,9%)	13 (8,7%)	8 (5,2%)	
Sonstige (F32.8/.9, F33.8/.9)	17 (5,6%)	10 (6,7%)	7 (4,6%)	
<i>Psychotherapie</i>				$\chi^2(4) = 6,32,$ $p = 0,176$
Keine	143 (47,2%)	76 (50,7%)	67 (43,8%)	
Teilstationär	2 (0,7%)	0 (0,0%)	2 (1,3%)	
Stationär	2 (0,7%)	1 (0,7%)	1 (0,7%)	
Ambulant	135 (44,6%)	67 (44,7%)	68 (44,4%)	
Anderes	21 (6,9%)	6 (4,0%)	15 (9,8%)	
<i>Psychopharmaka</i>				$\chi^2(1) = 4,53,$ $p = 0,033$
Keine Einnahme	178 (58,7%)	79 (52,7%)	99 (64,7%)	
Regelmäßige Einnahme	125 (41,3%)	71 (47,3%)	54 (35,3%)	
<i>Weitere</i>				
Anzahl komorbider Störungen (WSQ)	3,43 (1,26)	3,30 (1,20)	3,55 (1,30)	$t(301) = 1,73,$ $p = 0,084$
Behandlungs- erwartung (TEX-Q)	6,33 (1,09)	6,30 (1,00)	6,37 (1,17)	$t(301) = 0,59,$ $p = 0,559$

*BDI-II = Beck Depressionsinventar-II; PHQ-9 = Patient Health Questionnaire; RSES = Rosenberg Self-Esteem Scale; WHO-QOL = World Health Organization Quality of Life Kurzfragebogen.*

## 6.2 Wiedererreichungsquote

Insgesamt schlossen 219 Teilnehmer:innen (72,2%) die Post-Befragung (T2, 12 Wochen nach Baseline) ab. Die Wiedererreichungsquote unterschied sich nicht signifikant zwischen der IG (69,3%) und WKG (75,3%) ( $\chi^2(1) = 1,39, p = 0,239$ ). Studienabbrecher:innen (ohne Abschluss der Post-Befragung;  $M = 38,4, SD = 11,62$ ) waren jünger im Vergleich zu denjenigen, die die Post-Befragung beendeten ( $M = 41,57, SD = 12,73$ ),  $t(301) = 2,00, p = 0,047$ . Studienabbrecher:innen (absolvierte Psychotherapien vor Studienteilnahme:  $M = 2,67, SD = 2,42$ ) und Teilnehmer:innen mit abgeschlossener Post-Befragung ( $M = 3,75, SD = 8,20$ ) berichteten gleich viel Erfahrung mit Psychotherapie ( $t(298) = 1,18, p = 0,241$ ). Weiter zeigten sich keine signifikanten Unterschiede bezüglich des Geschlechts ( $\chi^2(2) = 3,08, p = 0,215$ ), des Bildungsstatus ( $\chi^2(4) = 6,25, p = 0,181$ ), oder in Bezug auf die depressive Symptombelastung zum Zeitpunkt der Baseline-Befragung (BDI-II; Abbrecher:innen:  $M = 24,89, SD = 9,85$ , Absolvent:innen:  $M = 27,07, SD = 9,19$ ;  $t(301) = 1,81, p = 0,071$ ).

Gründe für das Abbrechen konnten nur von einer Person erfasst werden. Diese berichtete verzögert, keine Zeit gehabt zu haben, an der Befragung teilzunehmen. Alle anderen Abbrecher:innen meldeten sich auf wiederholte Kontaktversuche nicht zurück.

## 6.3 Interventionsnutzung und -zufriedenheit

Die Teilnehmer:innen der IG loggten sich durchschnittlich 14-mal ( $M = 14,88, SD = 17,61$ ) im Programm ein (Range = 0 bis 142). Die Zufriedenheit, erhoben mit dem ZUF-8 (Range 8-32), liegt im Mittel bei 22,54 ( $SD = 5,18$ ). Sehr oder weitgehend zufrieden mit dem Programm („Wie zufrieden sind Sie mit dem Novego-Programm im Großen und Ganzen?“) waren 72,8% der Novego-Nutzer:innen, 74,7% würden das Programm weiterempfehlen.

## 6.4 mITT- Analysen

Das MMRM zeigte für den primären Endpunkt (Veränderung im BDI-II zum Zeitpunkt T2) eine signifikant stärkere Reduktion im Summenscore und somit Abnahme Patient:innen-berichteter depressiver Symptome in der IG im Vergleich zur WKG (s. Tab. 3).

**Tabelle 3**

Ergebnisse der mITT-Analysen mit MMRM für den BDI-II.

mITT	WKG	IG	Effektstärke [95 %-KI] p-Wert <sup>a</sup>
N	150	153	-
<b>Baseline</b>			
M (SD)	26,68 (9,80)	27,49 (8,52)	-

<b>mITT</b>	<b>WKG</b>	<b>IG</b>	<b>Effektstärke [95 %-KI] p-Wert<sup>a</sup></b>
<b>Studienende</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	-
<i>M (SD)</i>	22,31 (12,60)	20,45 (11,84)	
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	104/150 (69,3)	99/153 (64,7)	LS MD 2,653
<i>M (SD)</i>	-2,226 (7,659)	-5,153 (7,133)	[0,527; 4,780] 0,015
LS Mean (SE)	-2,189 (0,752)	-4,842 (0,776)	Hedges' <i>g</i> 0,358
95 %-KI	[-3,667; -0,710]	[-6,368; -3,316]	[0,071; 0,645]
<b>Veränderung zu T2</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	LS MD 2,537
<i>M (SD)</i>	-4,518 (8,313)	-6,905 (8,357)	[0,461; 4,610] 0,017
LS Mean (SE)	-4,507 (0,734)	-7,044 (0,758)	Hedges' <i>g</i> 0,304
[95 %-KI]	[-5,950; -3,064]	[-8,535; -5,553]	[0,055; 0,553]

<sup>a</sup>Hedges' *g*: Differenz der Mittelwerte geteilt durch die gepoolte Standardabweichung. Die Veränderungen der Punktwerte im Vergleich zum Ausgangswert plus 95 %-KI wurden mit MMRM zu jedem angegebenen Zeitpunkt auf der Grundlage der kleinsten Quadrate geschichtet analysiert. Die *p*-Werte beruhen auf einer ANCOVA der entsprechenden Effektmaße.

KI: Konfidenzintervall; ITT: Intention-to-Treat; LS MD: Mean difference of least squares; LS MW (SE): Mean of least squares (Standardfehler); *M (SD)*: Mittelwert (Standardabweichung); *N*: Teilnehmende mit Baseline-Werten; *n*: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse

Ebenso zeigte sich eine signifikant stärkere Reduktion im sekundären Endpunkt (Veränderung im Summenwert des PHQ-9 zum Zeitpunkt T2) in der IG im Vergleich zur WKG (s. Tab. 4).

**Tabelle 4**

Ergebnisse der mITT-Analysen mit MMRM für den PHQ-9.

mITT	WKG	IG	Effektstärke [95 %-KI] p-Wert <sup>a</sup>
<i>N</i>	150	153	-
<b>Baseline</b>			
<i>M (SD)</i>	13,06 (5,21)	14,01 (4,71)	-
<b>Studienende</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	
<i>M (SD)</i>	-1,876 (3,953)	-3,581 (4,004)	-
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	104/150 (69,3)	99/153 (64,7)	LS MD 0,677
<i>M (SD)</i>	-1,462 (3,549)	-2,398 (3,454)	[-0,288; 1,642] 0,168
LS Mean (SE)	-1,570 (0,341)	-2,248 (0,353)	Hedges' <i>g</i> 0,190
95 %-KI	[-2,241; -0,900]	[-2,941; -1,554]	[-0,082; 0,469]
<b>Veränderung zu T2</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	LS MD 1,651
<i>M (SD)</i>	-1,876 (3,953)	-3,581 (4,004)	[0,710; 2,591] 0,001
LS Mean (SE)	-1,913 (0,333)	-3,564 (0,343)	Hedges' <i>g</i> 0,415
[95 %-KI]	[-2,568; -1,259]	[-4,239; -2,889]	[0,179; 0,651]

<sup>a</sup>Hedges' *g*: Differenz der Mittelwerte geteilt durch die gepoolte Standardabweichung. Die Veränderungen der Punktwerte im Vergleich zum Ausgangswert plus 95 %-KI wurden mit MMRM zu jedem angegebenen Zeitpunkt auf der Grundlage der kleinsten Quadrate geschichtet analysiert. Die *p*-Werte beruhen auf einer ANCOVA der entsprechenden Effektmaße.

KI: Konfidenzintervall; ITT: Intention-to-Treat; LS MD: Mean difference of least squares; LS MW (SE): Mean of least squares (Standardfehler); *M (SD)*: Mittelwert (Standardabweichung); *N*: Teilnehmende mit Baseline-Werten; *n*: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse

Das Ergebnis wurde jeweils auch für die explorativen Endpunkte (Veränderung der Summenwerte im BDI-II und PHQ-9 zum Zeitpunkt T1) bestätigt (BDI-II: s. Tab. 3, PHQ-9: s. Tab. 4).

LS MD (mit 95%-KI) der explorativen Subgruppenanalysen für Alter ( $\leq 65$  Jahre), Geschlecht, Diagnose und CAU für BDI-II und PHQ-9 sind in den Abbildungen 1 und 2 dargestellt. Die Subgruppe Alter  $> 65$  J. konnte nicht gebildet werden.

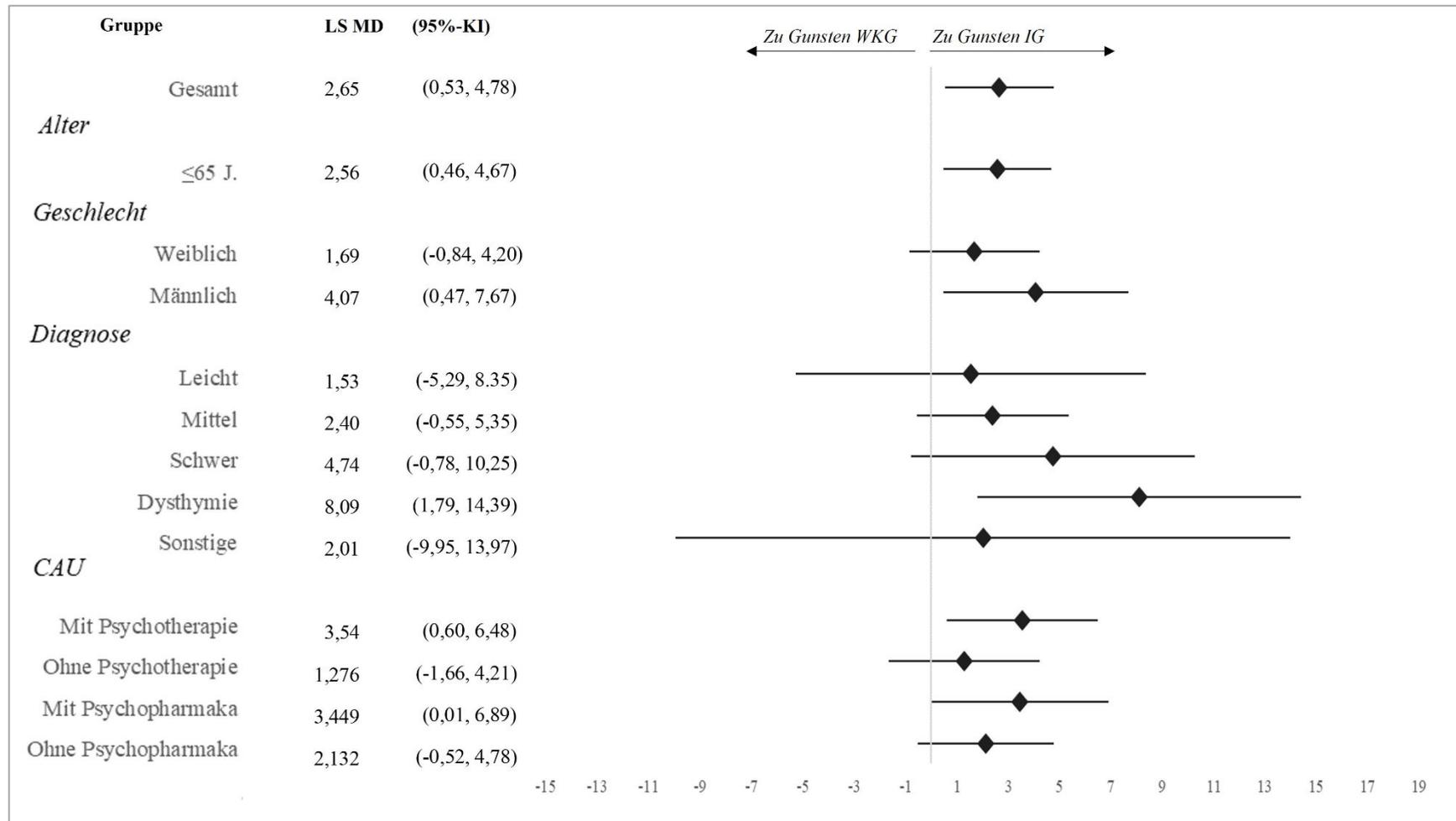


Abb.1 Forest Plot der LS MD und 95%-KI aus den explorativen mITT-Subgruppenanalysen (MMRM) des primären Endpunkts (Veränderung des BDI-II Summenscores T0-T2).

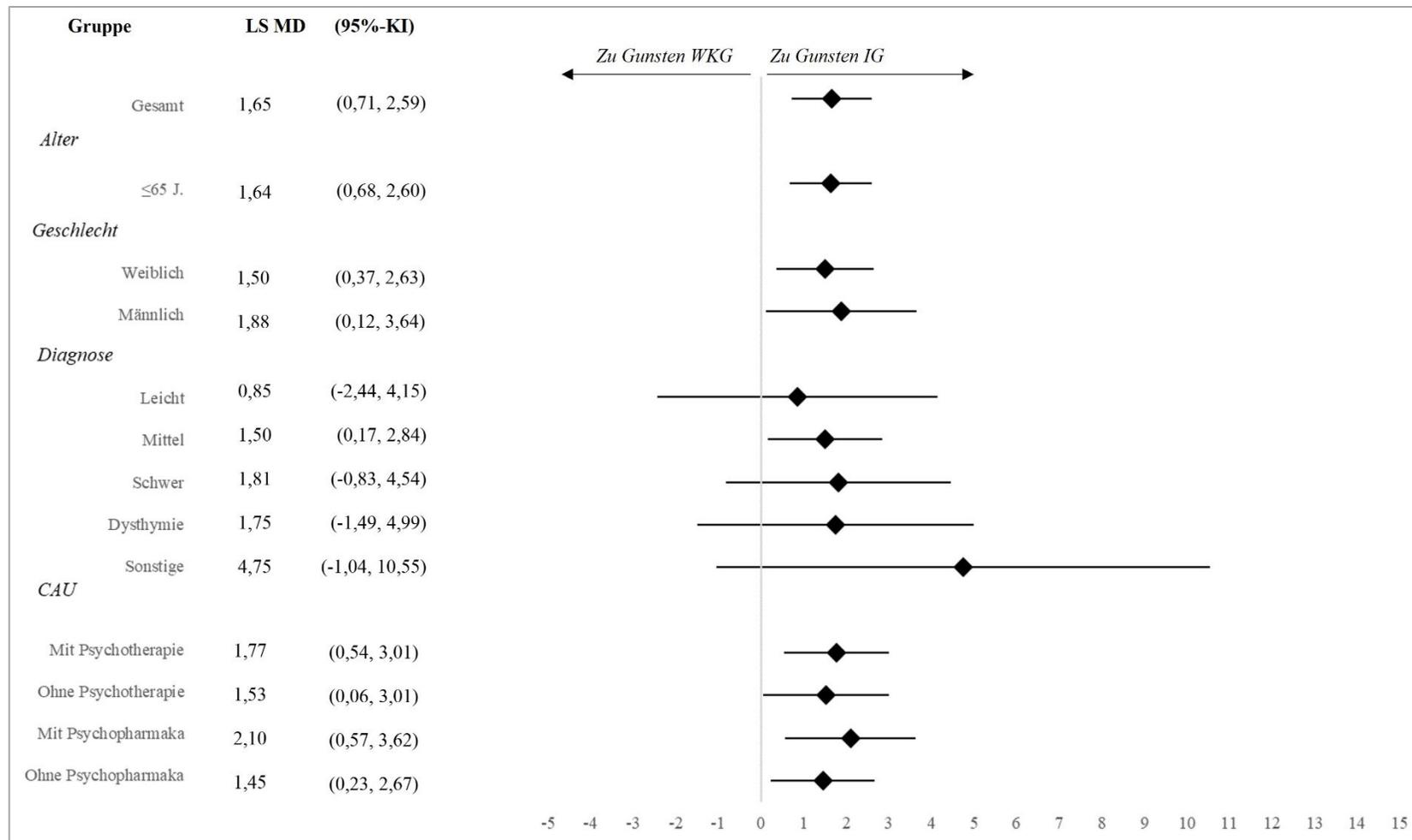


Abb.2 Forest Plot der LS MD und 95%-KI aus den explorativen mITT-Subgruppenanalysen (MMRM) des sekundären Endpunkts (Veränderung des PHQ-9 Summenscores T0-T2).

Unter Verwendung der MI (s. Tab. 5 und 6) zeigte sich eine signifikant größere Reduktion der Depressivität in der IG im Vergleich zur WKG zwischen T0 und T2 sowohl im BDI-II ( $p = 0,039$ ,  $g = 0,24$ ; s. Tab. 5) als auch im PHQ-9 ( $p = 0,002$ ,  $g = 0,39$ ; s. Tab. 6). In der RSES liegt lediglich ein Trend zu einem Vorteil der Intervention vor ( $p = 0,073$ ,  $g = 0,22$ ). Im WHO-QOL-BREF zeigte sich kein signifikanter Unterschied zwischen den Gruppen ( $p = 0,482$ ,  $g = 0,15$ ). Explorativ zeigte sich auch zwischen T0 und T1 ein signifikanter Unterschied in der Reduktion zwischen IG und WKG im BDI-II ( $p = 0,004$ ,  $g = 0,34$ ; s. Tab. 5), nicht aber für den PHQ-9 ( $p = 0,124$ ,  $g = 0,23$ ; s. Tab. 6).

**Tabelle 5**

Ergebnisse der mITT-Analyse mit MI für den BDI-II.

mITT	WKG	IG	Effektstärke [95 %-KI] $p$ -Wert <sup>a</sup>
<i>N</i>	150	153	-
<b>Baseline</b>			
<i>M (SD)</i>	26,68 (9,80)	27,49 (8,52)	-
<b>Studienende</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	-
<i>M (SD)</i>	22,31 (12,60)	20,45 (11,84)	-
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	104/150 (69,3)	99/153 (64,7)	LS MD 2,85 [0,878; 4,821]
<i>M (SD)</i>	1,94 (8,83)	4,90 (8,73)	$p = 0,004$  Hedges' $g = 0,34$
<b>Veränderung zu T2</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	LS MD 2,65 [0,424; 4,868]
<i>M (SD)</i>	4,48 (9,54)	6,89 (10,32)	$p = 0,039$  Hedges' $g = 0,24$

<sup>a</sup>Hedges'  $g$ : Differenz der Mittelwerte geteilt durch die gepoolte Standardabweichung; *M (SD)*: Mittelwert (Standardabweichung); *N*: Teilnehmende mit Baseline-Werten; *n*: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse; *KI*: Konfidenzintervall; *LS MD*: Mean difference of least squares

**Tabelle 6**

Ergebnisse der mITT-Analyse mit MI für den PHQ-9.

mITT	WKG	IG	Effektstärke [95 %-KI] p-Wert <sup>a</sup>
<i>N</i>	150	153	-
<b>Baseline</b>			
<i>M (SD)</i>	13,06 (5,21)	14,01 (4,71)	-
<b>Studienende</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	-
<i>M (SD)</i>	-1,876 (3,953)	-3,581 (4,004)	-
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	104/150 (69,3)	99/153 (64,7)	LS MD 0,710 [-0,198; 1,618]
<i>M (SD)</i>	1,39 (4,05)	2,33 (4,16)	<i>p</i> = 0,124  Hedges' <i>g</i> = 0,23
<b>Veränderung zu T2</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	LS MD 1,626 [0,060; 2,647]
<i>M (SD)</i>	1,86 (4,50)	3,62 (4,59)	<i>p</i> = 0,002  Hedges' <i>g</i> = 0,39

<sup>a</sup>Hedges' *g*: Differenz der Mittelwerte geteilt durch die gepoolte Standardabweichung; *M (SD)*: Mittelwert (Standardabweichung); *N*: Teilnehmende mit Baseline-Werten; *n*: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse; *KI*: Konfidenzintervall; *LS MD*: Mean difference of least squares

## 6.5 Sensitivitätsanalysen

### 6.5.1 J2R-Analysen

Wie bereits im Studienprotokoll und statistischen Analyseplan sowie in Kapitel 5 dieses Berichtes erläutert, wurde die J2R-Imputation nach ausführlicher Überlegung durch das MMRM und die MI ersetzt. Grund hierfür ist die Annahme der J2R-Imputation, dass Teilnehmende bei einem vorzeitigen Abbruch der Intervention in keinerlei Hinsicht mehr von dieser profitieren und sich daher dem Profil der Kontrollgruppe annähern würden. Jedoch ist für IBIs durchaus davon auszugehen, dass bei Abbruch der Intervention partielle langfristige Effekte bestehen bleiben und die Annahme eines „worst case Szenarios“ nach J2R-Imputation zu einer Unterschätzung des tatsächlichen Interventionseffektes führen würde. Daher werden die Ergebnisse der J2R-Imputation für den BDI-II und den PHQ-9 der Vollständigkeit halber in den folgenden Tabellen 7 und 8 als Sensitivitätsanalysen aufgeführt.

Bei Anwendung der J2R-Imputation ist eine signifikante Reduktion der depressiven Symptomatik im PHQ-9 zwischen T0 und T2 zugunsten der Intervention erkennbar ( $p = 0,026$ ;  $\eta_p^2 = 0,022$ ; s. Tab. 8), wenn auch nicht für den primären Endpunkt, den BDI-II zwischen T0 und T2, ( $p = 0,127$ ;  $\eta_p^2 = 0,011$ ; s. Tab. 7). Hier kann jedoch eine Verbesserung der depressiven Symptomatik zwischen T0 und T1 auf Trendniveau berichtet werden ( $p = 0,071$ ;  $\eta_p^2 = 0,015$ ; s. Tab. 7). Für die RSES und den WHO-QOL ergeben sich zwischen T0 und T2 keine signifikanten Effekte (RSES:  $p = 0,114$ ;  $\eta_p^2 = 0,013$ ; WHO-QOL:  $p = 0,371$ ;  $\eta_p^2 = 0,004$ ).

**Tabelle 7**

Ergebnisse der mITT-Analyse mit präspezifizierter J2R für den BDI-II

mITT	WKG	IG	Effektstärke [95 %-KI] $p$ -Wert <sup>a</sup>
<i>N</i>	150	153	-
<b>Baseline</b>			
<i>M (SD)</i>	26,68 (9,80)	27,49 (8,52)	-
<b>Studienende</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	-
<i>M (SD)</i>	22,31 (12,60)	20,45 (11,84)	-
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	104/150 (69,3)	99/153 (64,7)	LS MD 1,736 [0,206; 3,265]
<i>M (SD)</i>	2,21 (7,63)	4,11 (7,36)	$p = 0,071$  $\eta_p^2 = 0,015$
<b>Veränderung zu T2</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	LS MD 1,683 [-0,483; 3,848]
<i>M (SD)</i>	4,51 (8,26)	6,26 (8,37)	$p = 0,127$  $\eta_p^2 = 0,011$

<sup>a</sup>  $\eta_p^2$ : partielles Eta Quadrat; *M (SD)*: Mittelwert (Standardabweichung); *N*: Teilnehmende mit Baseline-Werten; *n*: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse; *KI*: Konfidenzintervall; *LS MD*: Mean difference of least squares

**Tabelle 8**

Ergebnisse der mITT-Analyse mit präspezifizierter J2R für den PHQ-9

<b>mITT</b>	<b>WKG</b>	<b>IG</b>	<b>Effektstärke [95 %-KI] p-Wert<sup>a</sup></b>
<i>N</i>	150	153	-
<b>Baseline</b>			
<i>M (SD)</i>	13,06 (5,21)	14,01 (4,71)	-
<b>Studienende</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	-
<i>M (SD)</i>	-1,876 (3,953)	-3,581 (4,004)	-
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	104/150 (69,3)	99/153 (64,7)	LS MD 0,529 [-0,310; 1,368]
<i>M (SD)</i>	1,36 (3,55)	2,11 (3,45)	<i>p</i> = 0,216  $\eta_p^2 = 0,007$
<b>Veränderung zu T2</b>			
<i>n/N (%)</i>	113/150 (75,3)	106/153 (69,3)	LS MD 1,133 [0,124; 2,124]
<i>M (SD)</i>	1,88 (3,98)	3,22 (4,05)	<i>p</i> = 0,026  $\eta_p^2 = 0,022$

<sup>a</sup>  $\eta_p^2$ : partielles Eta Quadrat; *M (SD)*: Mittelwert (Standardabweichung); *N*: Teilnehmende mit Baseline-Werten; *n*: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse; *KI*: Konfidenzintervall; *LS MD*: Mean difference of least squares

### 6.5.2 CC-Analysen

In der CC Population (WKG *n* = 113, IG *n* = 106) zeigte sich im primären Endpunkt eine signifikant größere Reduktion der depressiven Symptome in der IG, erhoben mit dem BDI-II, im Vergleich zur WKG ( $F(1;216) = 5,51, p = 0,020, \eta_p^2 = 0,025$ ; s. Tab. 9). Dieses Ergebnis wurde durch den sekundären Endpunktparameter, dem PHQ-9, bestätigt: auch hier zeigte sich ein signifikanter Gruppenunterschied in der Summenwertdifferenz zwischen Baseline- und Post-Befragung ( $F(1;215) = 9,88, p = 0,002, \eta_p^2 = 0,044$ ; s. Tab. 9). Für die RSES zeigte sich eine signifikant stärkere Verbesserung des Selbstwertes in der IG im Vergleich zur WKG von Baseline zu Post ( $F(1;213) = 4,13, p = 0,043, \eta_p^2 = 0,019$ ; s. Tab. 9). Für den WHO-QOL-BREF zeigte sich kein signifikanter Unterschied zwischen den Gruppen ( $F(1;215) = 0,55, p = 0,458, \eta_p^2 = 0,003$ ; s. Tab. 9).

Die Verbesserung der depressiven Symptome zeigte sich bereits zum Zeitpunkt der Interimserhebung (6 Wochen nach Baseline). Auch hier konnte ein signifikanter Unterschied zugunsten der IG im BDI-II ( $F(1;201) = 8,71, p = 0,004, \eta_p^2 = 0,042$ ) festgestellt werden. Dieser bestätigte sich jedoch nicht voll im PHQ-9 ( $F(1;201) = 2,82, p = 0,095, \eta_p^2 = 0,014$ ).

Für die CC-Analyse ergab sich für den primären Endpunkt (BDI-II) ein |Hedges'  $g$ | von 0,321, 95%-KI [0,054; 0,587]. Für den PHQ-9 zeigte sich  $|g| = 0,423$  [0,155; 0,691], in der RSES  $|g| = 0,303$  [0,035; 0,570] und für den WHO-QOL-BREF  $|g| = 0,163$  [-0,102; 0,428].

### 6.5.3 PP-Analysen

In der PP Population (WKG  $n = 118$ , IG  $n = 86$ ) zeigte sich im primären Endpunkt eine signifikant größere Reduktion der depressiven Symptome in der IG, erhoben mit dem BDI-II, im Vergleich zur WKG ( $F(1;193) = 9,90, p = 0,002, \eta_p^2 = 0,049$ ; s. Tab. 9). Dieses Ergebnis wurde durch den sekundären Endpunktparameter, erhoben mit dem PHQ-9, bestätigt ( $F(1;192) = 15,20, p < 0,001, \eta_p^2 = 0,073$ ; s. Tab. 9). Für die RSES zeigte sich eine signifikant stärkere Verbesserung des Selbstwertes in der IG im Vergleich zur WKG von Baseline zu Post ( $F(1;193) = 7,16, p = 0,008, \eta_p^2 = 0,036$ ; s. Tab. 9). Für den WHO-QOL-BREF zeigte sich kein signifikanter Unterschied zwischen den Gruppen ( $F(1;192) = 0,66, p = 0,418, \eta_p^2 = 0,003$ ; s. Tab. 9).

### 6.6 Klinische Relevanz

Zur Beurteilung der klinischen Relevanz wurde Hedges'  $g$  herangezogen. Folgend der Argumentation zur klinischen Relevanz nach Cohen (Cohen, 1988) ist der Effektschätzer selbst für die Beurteilung relevant. Somit wird ein Hedges'  $g$  über 0,2 als klinisch relevanter Effekt betrachtet. Die Ergebnisse zeigen entsprechend eine klinisch relevante Reduktion der depressiven Symptomatik zu T2 im primären (BDI-II) und sekundären Endpunkt (PHQ-9).

Zur Bewertung der klinischen Relevanz auf der Originalskala für den primären Endpunkt (BDI-II) wurde post hoc zusätzlich das in Berger et al. (2011; RCT zur DiGA Deprexis) herangezogene Kriterium verwendet. Dabei müssen die Teilnehmenden eine Verbesserung im BDI-II Score zum Postzeitpunkt von mindestens 8 Punkten sowie einen BDI-II Summenscore von unter 13 erreichen, um die Veränderung als klinisch relevant zu bewerten. Die Bewertung der klinischen Relevanz erfolgte ohne Personen, die unplausible Antworten lieferten (s. Abschnitt 5). In der IG erreichten mit 28 (26,9%) signifikant mehr Personen dieses Kriterium im Vergleich zur WKG mit 16 (13,9%;  $\chi^2(1) = 5,76, p = .016$ ).

Der gleiche Effekt zeigte sich bei der Bewertung der klinischen Relevanz für den sekundären Endpunkt (PHQ-9), definiert als eine 50% Verbesserung gegenüber dem Baseline-Wert und ein Summenscore zum Postzeitpunkt unter 10 (vgl. RCT zu Deprexis, Meyer et al., 2015). Dieses Kriterium erreichten 12 Teilnehmende aus der WKG (10,5%) und 22 Teilnehmende der IG (21,1%),  $\chi^2(1) = 4,67, p = .031$ .

Sowohl bei primärem als auch sekundärem Endpunkt erreichten in der Interventionsgruppe doppelt so viele Patienten eine klinisch relevante Verbesserung wie in der Kontrollgruppe.

**Tabelle 9**

Mittelwerte und Standardabweichungen (CC), Unterschiede zwischen den Gruppen über die Zeit (ANCOVAs) für CC und PP.

	WKG		IG		Unterschiede zwischen den Gruppen; ANCOVAs mit Baseline-Werten als Kovariaten	
	Baseline <i>M</i> ( <i>SD</i> )	Post <i>M</i> ( <i>SD</i> )	Baseline <i>M</i> ( <i>SD</i> )	Post <i>M</i> ( <i>SD</i> )	CC	PP
<i>Primär (T0-T2)</i>						
BDI-II	26,68 (9,80)	22,31 (12,60)	27,49 (8,52)	20,45 (11,84)	$F(1;216) = 5,51,$ $p = 0,020, \eta_p^2 = 0,025$	$F(1;193) = 9,90,$ $p = 0,002, \eta_p^2 = 0,049$
<i>Sekundär (T0-T2)</i>						
PHQ-9	13,49 (5,22)	11,62 (5,58)	13,77 (4,85)	10,21 (5,29)	$F(1;215) = 9,88,$ $p = 0,002, \eta_p^2 = 0,044$	$F(1;192) = 15,20,$ $p < 0,001, \eta_p^2 = 0,073$
RSES	14,34 (6,71)	15,14 (6,64)	13,50 (6,56)	15,68 (6,70)	$F(1;213) = 4,13,$ $p = 0,043, \eta_p^2 = 0,019$	$F(1;193) = 7,16,$ $p = 0,008, \eta_p^2 = 0,036$
WHO-QOL	41,29 (19,63)	46,88 (20,24)	38,09 (17,96)	46,70 (19,99)	$F(1;215) = 0,55,$ $p = 0,458, \eta_p^2 = 0,003$	$F(1;192) = 0,66,$ $p = 0,418, \eta_p^2 = 0,003$
	<b>Interim <i>M</i> (<i>SD</i>)</b>		<b>Interim <i>M</i> (<i>SD</i>)</b>			
<i>Explorativ (T0-T1)</i>						
BDI-II		2,10 (7,68)		5,26 (7,06)	$F(1;201) = 8,71,$ $p = 0,004, \eta_p^2 = 0,042$	$F(1;184) = 12,16,$ $p < 0,001, \eta_p^2 = 0,062$
PHQ-9		1,42 (3,54)		2,43 (3,46)	$F(1;201) = 2,82,$ $p = 0,095, \eta_p^2 = 0,014$	$F(1;184) = 4,21,$ $p = 0,041, \eta_p^2 = 0,022$

WKG: Wartekontrollgruppe; IG: Interventionsgruppe; BDI-II: Beck Depressionsinventar-II; PHQ-9: Patient Health Questionnaire 9; RSES: Rosenberg Self-Esteem Scale; WHO-QOL: World Health Organization Quality of Life Kurzfragebogen.

## 6.7 Sicherheit

In der Post-Befragung wurden durch die PANEPS-I positive und nicht intendierte negative Effekte, einschließlich potenzieller Nebenwirkungen der Intervention, erfasst. In diese Analysen sind alle Patient:innen mit vollständigen Post-Befragungen eingegangen. Die Häufigkeiten der Zustimmung je Item für die positiven Effekte sind in Tabelle 10, für die negativen Effekte in Tabelle 11 dargestellt.

**Tabelle 10**

Häufigkeiten in Prozent der Zustimmung der positiven Effekte.

<b>PANEPS-I Items 1–6</b>	<b><i>n</i></b>	<b>Davon Zustimmung in %</b>
Nach Abschluss des Online-Programms fasste ich Vertrauen in meine eigenen Fähigkeiten.	85	54,1
Ich habe während des Online-Programms gelernt, Verantwortung für mich selbst zu übernehmen.	85	60,0
Mein Leistungsvermögen hat sich durch das Online-Programm verbessert.	83	26,5
Ich bin stolz auf mich, weil ich das Online-Programm absolviert habe.	79	57,0
Durch das Online-Programm konnte ich mich aus belastenden Beziehungen lösen.	72	22,2
Durch das Online-Programm ist meine Bereitschaft für eine direkte Psychotherapie mit einem Therapeuten/einer Therapeutin gestiegen.	66	62,1

*Anmerkung. Fehlende Werte repräsentieren die Auswahl der Antwortmöglichkeit „nicht zutreffend“. Prozentangaben sind gültige Prozent.*

**Tabelle 11**

Häufigkeiten in Prozent der Zustimmung der negativen Effekte.

<b>PANEPS-I Items 7–27</b>	<b><i>n</i></b>	<b>Davon Zustimmung in %</b>
<i>Nebenwirkungen</i>		
Ich habe Angst, dass mein soziales Umfeld erfährt, dass ich ein Online-Programm gemacht habe.	101	5,0
Durch das Online-Programm hat sich mein Verhältnis zu meiner Familie/meinen Freunden verschlechtert.	96	3,1

<b>PANEPS-I Items 7–27</b>	<b><i>n</i></b>	<b>Davon Zustimmung in %</b>
Es belastet mich, dass mich meine Familie/Freunde anders behandeln, seit ich das Online-Programm gemacht habe.	92	0,0
Seit der Nutzung des Online-Programms, fühle ich mich als psychisch kranker Mensch <i>abgestempelt</i> .	94	8,5
Meiner Familie/meinen Freunden ist es peinlich, dass ich ein Online-Programm gemacht habe.	95	1,1
Mir ist es peinlich, ein Online-Programm gemacht zu haben.	100	4,0
Ich fürchte, dass ich bestimmte Übungen nicht richtig umgesetzt/ Inhalte nicht richtig verstanden habe und sich mein Zustand dadurch verschlechtert hat.	96	14,6
<i>Kunstfehler</i>		
Ich habe technische Schwierigkeiten mit dem Online-Programm erlebt, die mich störten.	96	11,5
Es wurde in dem Online-Programm nur auf „Beseitigung des Problems“ hingearbeitet, es fand keine positive Zielorientierung statt.	75	12,0
Durch das Online-Programm habe ich das Gefühl bekommen, ich sei schuld an meinen Problemen.	93	8,6
Aus meiner Sicht waren die angewandten therapeutischen Techniken des Online-Programms falsch.	80	7,5
Das Online-Programm hat meine persönlichen Probleme nicht (ausreichend) angesprochen.	87	40,2
Ich habe durch das Online-Programm Verhaltensweisen gelernt, die mir sehr geschadet haben (z.B. übermäßiges Nein-Sagen, sich abgrenzen).	90	8,9
Die Übungen/Informationen des Online-Programms stimmten nicht überein mit dem, was mein Therapeut/meine Therapeutin oder mein Arzt/meine Ärztin mir gesagt haben.	74	5,4
Ich hatte das Gefühl, dass meine Probleme in dem Online-Programm nicht ernst genommen werden.	90	11,1

<b>PANEPS-I Items 7–27</b>	<b><i>n</i></b>	<b>Davon Zustimmung in %</b>
Ich finde, dass ich mich durch das Online-Programm negativ verändert habe.	89	2,2
Ich habe mich durch das Online-Programm nicht angesprochen gefühlt (z.B. aufgrund fehlender geschlechtergerechter Sprache oder fehlender Individualität)	89	13,5
Das Online-Programm übte zu viel Zeit- und/oder Leistungsdruck auf mich aus.	97	48,5

#### *Unethische Inhalte*

Ich fühlte mich von dem Online-Programm unter Druck gesetzt, Dinge zu tun, die ich eigentlich gar nicht tun wollte.	92	23,9
Das Online-Programm erschien mir intolerant gegenüber meiner Herkunft/Religion/sexuellen Orientierung/Geschlecht	90	5,6
Ich hatte Bedenken bezüglich des Schutzes meiner Daten.	97	5,2

*Fehlende Werte repräsentieren die Auswahl der Antwortmöglichkeit „nicht zutreffend“. Prozentangaben sind gültige Prozent.*

Neben der mit der PANEPS-I erfassten nicht intendierten Effekte wurden keine negativen Ereignisse berichtet.

### 6.8 Moderationsanalysen

Moderationsanalysen wurden mit den Daten der mITT- und der PP-Population berechnet. Ziel war es, zu untersuchen, ob unterschiedliche Subpopulationen gleichermaßen von der Intervention profitieren. Die explorativen Endpunkte (Mittelwert der Behandlungserwartung, Häufigkeit weiterer positiver und negativer Effekte der Intervention [auch Nebenwirkungen], Vorliegen komorbider psychischer Störungen und Summenwert der Patient:innenzufriedenheit) gingen in die Analysen ein, um mögliche Variablen zu detektieren, die einen Einfluss auf die Veränderung der depressiven Symptomatik über die Zeit haben (unabhängige Variablen [UV]: Gruppenzuteilung sowie die Werte des jeweiligen explorativen Endpunktes; abhängige Variablen [AV]: Differenzwerte der BDI-II-Summenwerte T2-T0).

Für die Behandlungserwartung als Moderator ergab sich kein signifikanter Interaktionseffekt im mITT- ( $p = 0,306$ ) und PP-Sample ( $p = 0,773$ ). Das Auftreten negativer Effekte der Intervention (kombiniert für die Subskalen Nebenwirkungen, Kunstfehler und unethische Inhalte) zeigte ebenfalls keine signifikante Interaktion (nur für PP berechnet,  $p = 0,184$ ). Die Patient:innenzufriedenheit moderierte die

Abnahme der Symptombelastung ebenfalls nicht signifikant (nur für PP berechnet,  $p = 0,896$ ). Auch das Vorliegen komorbider Störungen zeigte keine signifikante Interaktion (mITT:  $p = 0,378$ ; PP:  $p = 0,617$ ).

Die Schweregradeinteilung nach PHQ-9 zeigte keine signifikante Interaktion mit der Reduktion der Symptombelastung (mITT:  $p = 0,315$ ; PP:  $p = 0,967$ ).

## 7. Diskussion

Ziel dieser zweiarmligen randomisiert-kontrollierten Studie war es, die Wirksamkeit in Bezug auf den medizinischen Nutzen des internetbasierten Programms Novego – „Depressionen bewältigen“<sup>®</sup> bei Erwachsenen mit einer depressiven Störung zu untersuchen. Es wurde die Hypothese verfolgt, dass Interventionsnutzer:innen nach einem Nutzungszeitraum von 12 Wochen eine signifikant größere Reduktion der Depressionssymptomatik sowie eine signifikant größere Steigerung des Selbstwertes und der Lebensqualität im Vergleich zu einer WKG zeigen.

### *Wirksamkeit*

Die primäre mITT-Analyse zeigte dabei, dass sich die Depressivität in der IG im Vergleich zur WKG nach 12 Wochen signifikant stärker verringerte, sowohl im primären Endpunktparameter (Summenwert des BDI-II, MMRM:  $p = 0,017$ , MI:  $p = 0,039$ ) als auch im sekundären Parameter (Summenwert der PHQ-9, MMRM:  $p = 0,001$ , MI:  $p = 0,002$ ), mit kleinen bis mittleren Effektstärken. Der Effekt zeigte sich im MMRM bereits nach 6 Wochen. Aus den Ergebnissen der explorativen Subgruppenanalysen für die Veränderung zu T2 sind keine Effektmodifikationen erkennbar. Die Veränderung im Selbstwert erreichte bei MI im Gruppenvergleich statistisches Trendniveau ( $p = 0,073$ ). Nach Durchführung der J2R-Imputation zeigte sich im sekundären Endpunkt, erhoben mit dem PHQ-9 zwischen T0 und T2, ein signifikanter Gruppenunterschied bei der Verbesserung der depressiven Symptomatik ( $p = 0,026$ ) sowie eine Verbesserung auf Trendniveau für den BDI-II zwischen T0 und T1 zugunsten der Intervention ( $p = 0,071$ ). Für die restlichen Endpunkte konnten keine signifikanten Ergebnisse erzielt werden. Es ist, wie bereits in Abschnitt 5 und 6.5.1 erläutert, anzumerken, dass die Ergebnisse auf Grundlage der J2R-Imputation ein „worst case Szenario“ abbilden, weshalb sie lediglich aus Gründen der Vollständigkeit berichtet werden. In den CC- und PP-Analysen zeigte sich ein signifikant stärkerer Zuwachs des Selbstwertes in der IG im Vergleich zur WKG (CC:  $p = 0,043$ , PP:  $p = 0,008$ ). Zudem bestätigten sich in den CC- und PP-Analysen die Ergebnisse bezüglich der Reduktion depressiver Symptomschwere mit insgesamt größeren Effektstärken, was die Effekte der Intervention unterstreicht.

Die Ergebnisse in Bezug auf die Reduktion der depressiven Symptomatik reihen sich ein in frühere Befunde zu verschiedenen Patient:innengruppen (Beiwinkel et al., 2017; Miegel et al., 2019; Moritz et al., 2016). Die Ergebnisse der vorliegenden Untersuchung zeigen nun, dass die Intervention Novego – „Depressionen bewältigen“<sup>®</sup> auch bei Patient:innen mit einer gesicherten, primären Depressionsdiagnose wirksam im Vergleich zu einer Wartekontrolle ist.

### *Klinische Relevanz*

Die Überlegenheit der Intervention gegenüber CAU wird ebenfalls unterstützt durch den Vergleich der Anteile der Teilnehmenden zwischen den Gruppen, die sich zum Postzeitpunkt klinisch relevant verbesserten. Mit über einem Viertel zeigten signifikant mehr Personen der IG eine klinisch relevante Reduktion ihrer depressiven Symptomatik auf der Originalskala (BDI-II, primärer Endpunkt). Dieser Anteil entspricht auch in etwa dem, der sich mit der unbegleiteten Version der DiGA Deprexis klinisch relevant verbesserten, wobei dasselbe Kriterium zugrunde gelegt wurde (28%; Berger et al., 2011).

### *Moderation*

Entgegen den Ergebnissen von Miegel et al. (2019), die bei einer Stichprobe von Schmerzpatient:innen mit depressiver Symptomatik einen moderierenden Effekt der Depressionsschwere (größere Reduktion bei schwererer Symptomatik zu Baseline) feststellen konnten, ergab sich in dieser Studie kein moderierender Effekt der Depressionsschwere. Ebenso wurde die Veränderung der depressiven Symptomatik nicht signifikant beeinflusst durch die Anzahl komorbider Störungen, die Summe der berichteten negativen Effekte und auch nicht durch die Behandlungserwartung. Somit ist davon auszugehen, dass alle Subpopulationen gleichermaßen profitieren.

### *Sicherheit*

Die standardisierte Erfassung von nicht-intendierten Effekten mithilfe der PANEPS-I diente zur Feststellung der Nutzer:innen-Sicherheit. Da Nebenwirkungen in der Psychotherapie allgemein, vor allem aber im Rahmen von Online-Interventionen ein vernachlässigtes Thema sind, können die berichteten Effekte nur deskriptiv eingeordnet werden. Über 80% der Teilnehmer:innen erlebten positive Ergebnisse bei der Nutzung der Intervention. Die am häufigsten berichtete positive Erfahrung war das Gefühl von Stolz über den Abschluss des Online-Programms. Berichtete negative Erlebnisse bezogen sich gehäuft auf die Sorge, Übungen oder Inhalte nicht korrekt umgesetzt zu haben und dadurch den eigenen Zustand verschlechtert zu haben oder sich unter Druck gesetzt gefühlt zu haben. In den Moderationsanalysen konnte kein Zusammenhang mit der Anzahl der berichteten unerwünschten Effekte und der Veränderung der Depressionsschwere festgestellt werden. Es ist daher davon auszugehen, dass das Erleben von nicht erwünschten Effekten während des Interventionszeitraumes keinen schädigenden Einfluss auf den Therapieverlauf genommen hat. Dass Nebenwirkungen generell auch bei Online-Interventionen auftreten, deckt sich mit den Ergebnissen anderer Studien über unerwünschte Wirkungen bei ähnlichen Interventionen (8-14% der Nutzer:innen berichten über Nebenwirkungen; Fenski et al., 2021, Böttcher et al., 2014).

Es bleibt bei der Interpretation zu beachten, dass nach dem Auftreten dieser Effekte aus rein subjektiver Sicht der Nutzer:innen gefragt wird und nicht, ob oder in welchem Ausmaß sich diese Ereignisse tatsächlich als negativ, belastend oder schädigend auswirkten. Außerhalb der standardisierten Erhebung wurden keine weiteren negativen Effekte von Teilnehmenden berichtet. Die Intervention kann damit als sicher beurteilt werden.

### *Schlussfolgerung*

In der Gesamtbetrachtung der primären Analyse sowie der Sensitivitätsanalysen erwies sich das Programm Novego – „Depressionen bewältigen“<sup>®</sup> gegenüber einer Wartekontrolle in Bezug auf die Reduktion der Symptomschwere als wirksam.

Zudem kann nach Betrachtung der Ergebnisse zu nicht intendierten Effekten das Programm als sicher beurteilt werden. Knapp drei Viertel der Nutzer:innen waren zufrieden mit dem Programm und würden es weiterempfehlen.

### Literatur

Alberts, J., Löwe, B., Glahn, M. A., Petrie, K., Laferton, J., Nestoriuc, Y., & Shedden-Mora, M. (2020). Development of the generic, multidimensional Treatment Expectation Questionnaire (TEX-Q) through systematic literature review, expert surveys and qualitative interviews. *BMJ open*, *10*(8), e036169.

Baumeister, A., & Moritz, S. (2022). *Skala zur Erfassung positiver und negativer Effekte in der Psychotherapie (Positive and Negative Effects of Psychotherapy Scale; PANEPS)*. In: Linden, M. & Strauss, B.: Erfassung von Risiken und Nebenwirkungen von Psychotherapie.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the beck depression inventory-II*. San Antonio, TX: Psychological Corporation, 1, 82.

Beiwinkel, T., Eißing, T., Telle, N. T., Siegmund-Schultze, E., & Roessler, 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.

Berger, T., Hämmerli, K., Gubser, N., Andersson, G., & Caspar, F. (2011). Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. *Cognitive behaviour therapy*, *40*(4), 251-266.

Böttcher, J., Rozentel, A., Andersson, G., & Carlbring, P. (2014). Side effects in Internet-based interventions for Social Anxiety Disorder. *Internet Interventions*, *1*(1).  
<https://doi.org/10.1016/j.invent.2014.02.002>

Bundespsychotherapeutenkammer (2018) *Ein Jahr nach der Reform der Psychotherapie-Richtlinie. Wartezeiten 2018*. BpTK, Berlin

Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) (24.03.2021). *Ergebnisprotokoll zur Beratung nach §23 DiGAV, für die DiGA Novego – „Depressionen bewältigen*

Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), i.A. Waider, L. (07.10.2022). *Verlängerung des Erprobungszeitraums*. Antrags-ID: 2018. GESCHZ: 96.01-5672-1820.

Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), i.A. Moellers, T. (06.02.2023). *Feststellung inhaltlicher Mängel*. Antrags-ID: 3907. GESCHZ: 96.03-5672-1820.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates

Carpenter, J. R., Roger, J. H., & Kenward, M. G. (2013). Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics*, 23(6), 1352-1371. doi: 10.1080/10543406.2013.834911.

Donker, T., van Straten, A., Marks, I. M., & Cuijpers, P. (2009). A brief Web-based screening questionnaire for common mental disorders: development and validation. *Journal of Medical Internet Research*, 11(3), e19.

European Medicines Agency (2011). *Guideline on Missing Data in Confirmatory Clinical Trials*. <https://www.ema.europa.eu/en/missing-data-confirmatory-clinical-trials-scientific-guideline> Zugriff am 05.07.2023)

Erbe, D., Eichert, H. C., Riper, H., & Ebert, D. D. (2017). Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. *Journal of Medical Internet Research*, 19(9), e306

Fenski, F., Rozental, A., Heinrich, M., Knaevelsrud, C., Zagorscak, P., & Boettcher, J. (2021). Negative effects in internet-based interventions for depression: A qualitative content analysis. *Internet Interventions*, 26, 100469.

García-Batista, Z. E., Guerra-Peña, K., Cano-Vindel, A., Herrera-Martínez, S. X., & Medrano, L. A. (2018). Validity and reliability of the Beck Depression Inventory (BDI-II) in general and hospital population of Dominican Republic. *PloS one*, 13(6), e0199750.

Gräfe, K., Zipfel, S., Herzog, W., & Löwe, B. (2004). Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)". *Diagnostica*, 50(4), 171-181.

Hedges, L. V., Olkin, I. (1985): *Statistical Methods for Meta-Analysis*. Cambridge, Massachusetts: Hedges, Academic Press. <https://doi.org/10.1016/C2009-0-03396-0>

Karyotaki, E., Efthimiou, O., Miguel, C., genannt BERPohl, F. M., Furukawa, T. A., Cuijpers, P., ... & Forsell, Y. (2021). Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA psychiatry*, 78(4), 361-371.

Karyotaki, E., Smit, F., Kleiboer, A., Mira Pastor, A., Berger, T., Cristina, B., de Graaf, L. E., Donker, T., Farrer, L., Meyer, B., Moritz, S., Spek, V., Vermark, K., & Cuijpers, P. (2015). Predictors of treatment dropout in self-guided web-based interventions for depression: An individual patient data meta-analysis. *Psychological Medicine*, *45*(13), 2717–2726.

Kenward, M. (2015). Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. *Clinical Investigation*, *5*. 311-320. 10.4155/cli.14.132.

Kriz, D., Nübling, R., Steffanowski, A., Rieger, J., & Schmidt, J. (2008). Patientenzufriedenheit: psychometrische reanalyse des ZUF-8. *DRV-Schriften*, *77*, 84-5.

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.

Little, R., Rubin, D. (1987). *Statistical Analysis With Missing Data*. New York: John Wiley & Sons.

Meyer, B., Bierbrodt, J., Schröder, J., Berger, T., Beevers, C. G., Weiss, M., ... & Klein, J. P. (2015). Effects of an Internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. *Internet Interventions*, *2*(1), 48-59.

Miegel, F., Gehlenborg, J., Bücken, L., Lion, D., & Moritz, S. (2019). Can an online intervention for depression alleviate emotional problems and pain? A randomized controlled study. *Verhaltenstherapie*, *29*(3), 166-181. <https://doi.org/10.1159/000501736>

Moritz, S., Schröder, J., Klein, J. P., Lincoln, T. M., Andreou, C., Fischer, A., & Arlt, S. (2016). Effects of online intervention for depression on mood and positive symptoms in schizophrenia. *Schizophrenia Research*, *175*(1-3), 216–222. <https://doi.org/10.1016/j.schres.2016.04.033>

National Research Council. (2010). *The prevention and treatment of missing data in clinical trials*. Washington, D.C.: The National Academies Press.

Rosenberg, M. (1965). *The measurement of self-esteem, society and the adolescent self-image*. Princeton, 16-36.

Schmidt, J., Lamprecht, F., & Wittmann, W. W. (1989). Zufriedenheit mit der stationären Versorgung. Entwicklung eines Fragebogens und erste Validitätsuntersuchungen. *PPmP: Psychotherapie Psychosomatik Medizinische Psychologie*.

Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL - BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research, 13*(2), 299-310.

Subica, A. M., Fowler, J. C., Elhai, J. D., Frueh, B. C., Sharp, C., Kelly, E. L., & Allen, J. G. (2014). Factor structure and diagnostic validity of the Beck Depression Inventory–II with adult clinical inpatients: Comparison to a gold-standard diagnostic interview. *Psychological Assessment, 26*(4), 1106.

Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., & Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive behaviour therapy, 40*(2), 126-136.

Whoqol Group. (1998). Development of the World Health Organization WHOQOL -BREF quality of life assessment. *Psychological Medicine, 28*(3), 551-558.

## Anhang

### A Baseline-Charakteristika der ITT-Population

**Tabelle A**

Häufigkeiten (Prozente), Mittelwerte (Standardabweichung) der soziodemografischen und psychopathologischen Variablen der Stichprobe zur Baseline-Erhebung.

<b>Variable</b>	<b>Gesamt (N = 311)</b>	<b>WKG (n =155)</b>	<b>IG (n =156)</b>	<b>Statistik</b>
<i>Demografische Charakteristika</i>				
Geschlecht (w/m/d)	222/85/4 (71,4/27,3/1,3%)	115/38/2 (74,2/24,5/1,3%)	107/47/2 (68,6/30,1/1,3%)	$\chi^2(2) = 1,24,$ $p = 0,538$
Alter in Jahren	40,88 (12,59)	40,82 (13,12)	40,94 (12,08)	$t(307) = 0,09,$ $p = 0,931$
Höhere Schulbildung (mind. Fachhochschulreife)	193 (62,1%)	91 (58,7%)	102 (65,4%)	$\chi^2(4) = 4,16,$ $p = 0,385$
<i>Psychopathologie</i>				
BDI-II	26,55 (9,39)	25,88 (10,11)	27,22 (8,61)	$t(309) = 1,2,$ $p = 0,207$
PHQ-9	13,54 (4,98)	13,06 (5,21)	14,01 (4,71)	$t(309) = 1,70,$ $p = 0,091$
RSES	14,26 (6,72)	14,60 (6,70)	13,92 (6,75)	$t(307) = 0,89,$ $p = 0,374$
WHO-QOL	39,31 (19,27)	41,05 (19,15)	37,58 (19,29)	$t(309) = 1,59,$ $p = 0,113$
<i>Psychotherapie</i>				$\chi^2(4) = 6,32,$ $p = 0,176$
Keine	146 (46,9%)	77 (49,7%)	69 (44,2%)	
Teilstationär	2 (0,6%)	0 (0,0%)	2 (1,3%)	
Stationär	2 (0,6%)	1 (0,6%)	1 (0,6%)	
Ambulant	140 (45,0%)	71 (45,8%)	69 (44,2%)	

<b>Variable</b>	<b>Gesamt (N = 311)</b>	<b>WKG (n =155)</b>	<b>IG (n =156)</b>	<b>Statistik</b>
Anderes	21 (6,8%)	6 (3,9%)	15 (9,6%)	
<i>Psychopharmaka</i>				$\chi^2(1) = 4,50,$ $p = 0,034$
Keine Einnahme	183 (58,8%)	82 (53,6%)	101 (64,7%)	
Regelmäßige Einnahme	128 (41,2%)	73 (47,1%)	55 (35,3%)	
<i>Weitere</i>				
Anzahl komorbider Störungen (WSQ)	3,41 (1,25)	3,28 (1,19)	3,54 (1,30)	$t(309) = -1,80,$ $p = 0,072$
Behandlungserwartung (TEX-Q)	6,35 (1,08)	6,30 (0,99)	6,39 (1,18)	$t(309) = -0,66,$ $p = 0,511$

*WKG: Wartekontrollgruppe; IG: Interventionsgruppe; BDI-II = Beck Depressionsinventar-II; PHQ-9 = Patient Health Questionnaire; RSES = Rosenberg Self-Esteem Scale; WHO-QOL = World Health Organization Quality of Life Kurzfragebogen; WSQ = Web Screening Questionnaire; TEX-Q = Treatment Expectation Questionnaire.*

## B Ergebnisse J2R

**Tabelle B**

Ergebnisse der präspezifizierten ITT-Analyse mit J2R-Imputation, Daten unter Einbezug der unplausiblen Datensets

<b>J2R ITT (N = 311)</b>			
		<i>M</i> <sub>Change</sub> (SD)	Statistik
BDI-II T0-T2	WKG	4,75 (8,49)	$F(1;4349,470) = 2,22, p = 0,136, \eta_p^2 = 0,009$
	IG	6,41 (8,49)	
PHQ-9 T0-T2	WKG	1,89 (3,88)	$F(1;4097,992) = 5,47, p = 0,019, \eta_p^2 = 0,022$
	IG	3,22 (4,01)	
RSES T0-T2	WKG	-0,79 (4,45)	$F(1;1843,907) = 2,93, p = 0,087, \eta_p^2 = 0,014$
	IG	-1,90 (4,64)	
WHO-QOL T0-T2	WKG	-5,98 (17,61)	$F(1;3631,226) = 0,95, p = 0,331, \eta_p^2 = 0,005$
	IG	-9,46 (19,61)	
BDI-II T0-T1	WKG	2,15 (7,56)	$F(1;3279,316) = 4,20, p = 0,040, \eta_p^2 = 0,018$
	IG	4,23 (7,63)	
PHQ-9 T0-T1	WKG	1,29 (3,51)	$F(1;3249,866) = 1,89, p = 0,169, \eta_p^2 = 0,009$
	IG	2,09 (3,56)	

WKG: Wartekontrollgruppe; IG: Interventionsgruppe; BDI-II = Beck Depressionsinventar-II; PHQ-9 = Patient Health Questionnaire; RSES = Rosenberg Self-Esteem Scale; WHO-QOL = World Health Organization Quality of Life Kurzfragebogen

C Ergebnisse ITT mit MMRM und MI

**Tabelle C.1**

Ergebnisse der ITT-Analyse für den BDI-II mit MMRM

ITT	WKG	IG	Effektstärke [95 %-KI] <i>p</i> -Wert <sup>a</sup>
<i>N</i>	155	156	-
<b>Baseline</b>			
<i>M (SD)</i>	25,88 (10,11)	27,22 (8,61)	-
<b>Studienende</b>			
<i>n/N (%)</i>	118/155 (76,1%)	109/156 (69,9%)	-
<i>M (SD)</i>	22,31 (12,55)	20,36 (11,90)	
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	109/155 (70,3%)	102/156 (65,4%)	LS MD 2,825 [0,704; 4,947] 0,009
<i>M (SD)</i>	-2,225 (7,545)	-5,250 (7,449)	
LS Mean (SE)	-2,124 (0,745)	-4,950 (0,780)	Hedges' <i>g</i> 0,376 [0,093; 0,660]
95 %-KI	[-3,589; -0,660]	[-6,483; -3,417]	

ITT	WKG	IG	Effektstärke [95 %-KI] p-Wert <sup>a</sup>
<b>Veränderung zu T2</b>			
n/N (%)	118/155 (76,1%)	109/156 (69,9%)	LS MD 2,241 [0,704; 4,310]
M (SD)	-4,808 (8,521)	-6,907 (8,414)	0,034
LS Mean (SE) [95 %-KI]	-4,796 (0,725) [-6,223; -3,369]	-7,037 (0,762) [-8,535; -5,538]	Hedges' g 0,265 [0,083; 0,588]

<sup>a</sup>Hedges' g: Differenz der Mittelwerte geteilt durch die gepoolte Standardabweichung. Die Veränderungen der Punktwerte im Vergleich zum Ausgangswert plus 95 %-KI wurden mit MMRM zu jedem angegebenen Zeitpunkt auf der Grundlage der kleinsten Quadrate geschichtet analysiert. Die p-Werte beruhen auf einer ANCOVA der entsprechenden Effektmaße.

KI: Konfidenzintervall; ITT: Intention-to-Treat; LS MD: Mean difference of least squares; LS MW (SE): Mean of least squares (Standardfehler); M (SD): Mittelwert (Standardabweichung); N: Teilnehmende mit Baseline-Werten; n: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse

**Tabelle C.2**

Ergebnisse der ITT-Analyse für den PHQ-9 mit MMRM

<b>ITT</b>	<b>WKG</b>	<b>IG</b>	<b>Effektstärke [95 %-KI] p-Wert<sup>a</sup></b>
<i>N</i>	155	156	-
<b>Baseline</b>			
<i>M (SD)</i>	13,06 (5,21)	14,01 (4,71)	-
<b>Studienende</b>			
<i>n/N (%)</i>	118/155 (76,1%)	109/156 (69,9%)	-
<i>M (SD)</i>	11,61 (5,53)	10,20 (5,24)	
<b>Veränderung zu T1</b>			
<i>n/N (%)</i>	109/155 (70,3%)	102/156 (65,4%)	LS MD 0,789
<i>M (SD)</i>	-1,351 (3,533)	-2,410 (3,447)	[-0,153; 1,731] 0,101
LS Mean (SE)	-1,471 (0,330)	-2,259 (0,347)	Hedges' g 0,226
95 %-KI	[-2,120.; -0,821]	[-2,941; -1,578]	[-0,044; 0,496]

ITT	WKG	IG	Effektstärke [95 %-KI] p-Wert <sup>a</sup>
<b>Veränderung zu T2</b>			
n/N (%)	118/155 (76,1%)	109/156 (69,9%)	LS MD 1,614 [0,697; 2,531]
M (SD)	-1,910 (3,871)	-3,579 (3,972)	0,001
LS Mean (SE) [95 %-KI]	-1,943 (0,322) [-2,575.; -1,310]	-3,557 (0,338) [-4,221.; -2,893]	Hedges' g 0,412 [0,178; 0,646]

<sup>a</sup>Hedges' g: Differenz der Mittelwerte geteilt durch die gepoolte Standardabweichung. Die Veränderungen der Punktwerte im Vergleich zum Ausgangswert plus 95 %-KI wurden mit MMRM zu jedem angegebenen Zeitpunkt auf der Grundlage der kleinsten Quadrate geschichtet analysiert. Die p-Werte beruhen auf einer ANCOVA der entsprechenden Effektmaße.

KI: Konfidenzintervall; ITT: Intention-to-Treat; LS MD: Mean difference of least squares; LS MW (SE): Mean of least squares (Standardfehler); M (SD): Mittelwert (Standardabweichung); N: Teilnehmende mit Baseline-Werten; n: Anzahl der Teilnehmenden mit Baseline-Werten in der Analyse

**Tabelle C.3**

Ergebnisse der ITT-Analyse mit multipler Imputation

	<b>WKG (n = 155)</b>	<b>IG (n = 156)</b>	<b>Unterschiede zwischen den Gruppen</b>
	<i>M</i> <sub>Change</sub> ( <i>SD</i> )	<i>M</i> <sub>Change</sub> ( <i>SD</i> )	<b>ITT (MI)</b>
<i>Primär</i>			
BDI-II T0-T2	4,60 (9,58)	6,97 (10,36)	$p = 0,040, g = 0,24$
<i>Sekundär</i>			
PHQ-9 T0-T2	1,87 (4,44)	3,61 (4,55)	$p = 0,002, g = 0,39$
RSES T0-T2	-0,80 (5,06)	-2,06 (5,51)	$p = 0,047, g = 0,24$
WHO-QOL T0-T2	-6,16 (20,10)	-9,35 (22,28)	$p = 0,422, g = 0,15$
<i>Explorativ</i>			
BDI-II T0-T1	1,88 (8,80)	4,98 (8,85)	$p = 0,003, g = 0,35$
PHQ-9 T0-T1	1,32 (4,07)	2,23 (4,14)	$p = 0,082, g = 0,25$

WKG: Wartekontrollgruppe; IG: Interventionsgruppe; BDI-II = Beck Depressionsinventar-II; PHQ-9 = Patient Health Questionnaire; RSES = Rosenberg Self-Esteem Scale; WHO-QOL = World Health Organization Quality of Life Kurzfragebogen

D Sensitivitätsanalysen der ITT-Population (CC- und PP-Analysen)

**Tabelle D**

Mittelwerte und Standardabweichungen (CC) der ITT-Population, Unterschiede zwischen den Gruppen über die Zeit (ANCOVAs) für CC und PP.

	WKG		IG		Unterschiede zwischen den Gruppen; ANCOVAs mit Baseline-Werten als Kovariaten	
	Baseline <i>M (SD)</i>	Post <i>M (SD)</i>	Baseline <i>M (SD)</i>	Post <i>M (SD)</i>	CC	PP
<i>Primär (T0-T2)</i>						
BDI-II	25,88 (10,11)	22,31 (12,55)	27,22 (8,61)	20,36 (11,90)	$F(1;224) = 5,40, p = 0,021,$ $\eta_p^2 = 0,024$	$F(1;201) = 9,80, p = 0,002,$ $\eta_p^2 = 0,047$
<i>Sekundär (T0-T2)</i>						
PHQ-9	13,06 (5,21)	11,61 (5,53)	14,01 (4,71)	10,20 (5,24)	$F(1;223) = 10,29, p = 0,002,$ $\eta_p^2 = 0,044$	$F(1;200) = 15,73, p < 0,001,$ $\eta_p^2 = 0,073$
RSES	14,60 (6,70)	15,09 (6,58)	13,92 (6,75)	15,83 (6,80)	$F(1;221) = 5,24, p = 0,023,$ $\eta_p^2 = 0,023$	$F(1;198) = 8,57, p = .004,$ $\eta_p^2 = 0,042$
WHO-QOL	41,05 (19,15)	46,47 (20,36)	37,58 (19,29)	46,67 (20,39)	$F(1;223) = 0,75, p = 0,388,$ $\eta_p^2 = 0,003$	$F(1;200) = 0,87, p = .352,$ $\eta_p^2 = 0,004$

	<b>Interim M (SD)</b>	<b>Interim M (SD)</b>		
<i>Explorativ (T0-T1)</i>				
BDI-II	24,31 (11,94)	22,40 (10,62)	$F(1;208) = 9,73, p = 0,002,$ $\eta_p^2 = 0,045$	$F(1;191) = 13,32, p < 0,001,$ $\eta_p^2 = 0,065$
PHQ-9	11,83 (4,88)	11,72 (5,24)	$F(1;208) = 3,75, p = 0,054,$ $\eta_p^2 = 0,018$	$F(1;191) = 5,25, p = 0,023,$ $\eta_p^2 = 0,027$

*WKG: Wartekontrollgruppe; IG: Interventionsgruppe; BDI-II: Beck Depressionsinventar-II; PHQ-9: Patient Health Questionnaire 9; RSES: Rosenberg Self-Esteem Scale; WHO-QOL: World Health Organization Quality of Life Kurzfragebogen.*

## E Moderation mit ITT-Population

**Tabelle E.1**

Ergebnisse der Moderationsanalysen der ITT-Population.

	<b>t</b>	<b>df</b>	<b>p</b>
Behandlungserwartung (TEX-Q)	1,122	207,788	0,263
Anzahl Komorbiditäten (WSQ)	0,952	208,259	0,342
Schweregrad Depression (PHQ-9, kategorial)	-1,016	244,729	0,310

*TEX-Q: Treatment Expectation Questionnaire; WSQ: Web Screening Questionnaire; PHQ-9: Patient Health Questionnaire.*

**Tabelle E.2**

Ergebnisse der Moderationsanalysen für PP in der ITT-Population.

	<b>t</b>	<b>df</b>	<b>p</b>
Behandlungserwartung (TEX-Q)	0,674	200	0,501
Anzahl Komorbiditäten (WSQ)	-0,468	200	0,641
Schweregrad Depression (PHQ-9, kategorial)	-0,361	200	0,719
Zufriedenheit (ZUF-8)	-0,087	87	0,931
Unerwünschte Effekte (UE; PANEPS-I)	1,441	87	0,153

*TEX-Q: Treatment Expectation Questionnaire; WSQ: Web Screening Questionnaire; PHQ-9: Patient Health Questionnaire; ZUF-8: Fragebogen zur Patientenzufriedenheit; PANEPS-I: Positive and Negative Effects of Psychotherapy Scale for Internet-based Interventions.*