

(gemäß Nr. 14.1 ANBest-IF)

Konsortialführung:	Universitätsklinikum Hamburg-Eppendorf (UKE)
Förderkennzeichen:	01NVF16018
Akronym:	RECOVER
Projekttitel:	Modell der sektorenübergreifend-koordinierten, schweregradgestuften, evidenzbasierten Versorgung psychischer Erkrankungen
Autoren:	<p>Aufgaben und Institute:</p> <p>VERSORGUNGSFORSCHUNG Stiftungslehrstuhl Klinische Versorgungsforschung, Institut und Poliklinik für Medizinische Psychologie, UKE Dr. Judith Peth, Prof. Dr. Holger Schulz</p> <p>GESUNDHEITSÖKONOMIE Institut für Gesundheitsökonomie und Versorgungsforschung, UKE Dr. Hannah König, PD Dr. Alexander Konnopka, Prof. Dr. Hans-Helmut König</p> <p>MEDIZINISCHE BIOMETRIE Institut für Medizinische Biometrie und Epidemiologie, UKE Dr. Ann-Kathrin Ozga, M. Sc. Amra Hot, Prof. Dr. Antonia Zapf</p>
Förderzeitraum:	1. Juli 2017 – 30. Juni 2021

Inhaltsverzeichnis

I. ABKÜRZUNGSVERZEICHNIS..... 3

II. ABBILDUNGSVERZEICHNIS 4

III. TABELLENVERZEICHNIS 4

 1. Ziele der Evaluation 5

 2. Darstellung des Evaluationsdesigns 5

2.1 STUDIENDESIGN..... 5

2.2 ZIELPOPULATION (EINSCHLUSS- UND AUSSCHLUSSKRITERIEN) 6

2.3 FALLZAHLEN (UNTERSUCHTE STICHPROBE) INKL. DROP-OUT..... 6

2.4 RANDOMISIERUNG UND VERBLINDUNG 7

2.5 COVID-19 PANDEMIE 8

2.6 HYPOTHESEN UND ENDPUNKTE..... 9

 2.6.1 Primäre Hypothesen und Endpunkte..... 9

 2.6.2 Sekundäre Hypothesen und Endpunkte..... 10

 2.6.3 Sekundärdaten 11

2.7 SCHWERWIEGENDE UNERWÜNSCHTE EREIGNIS (SAE) 12

2.8 STATISTISCHE AUSWERTUNG 12

 2.8.1 Fehlende Daten 12

 2.8.2 Deskription der Baseline-Daten..... 13

 2.8.3 Primärer Endpunkt 1: Versorgungskosten (gesundheitsökonomische Evaluation)..... 14

 2.8.4 Primärer Endpunkt 2: Psycho-Soziales-Funktionsniveau (Patientenrelevanter Endpunkt; Effektevaluation)..... 15

2.8.5	Primärer Endpunkt 3: Kosteneffektivität (gesundheitsökonomische Evaluation)	16
2.8.6	Sekundäre Endpunkte.....	17
2.8.7	Sekundäre Analysen	18
2.8.8	Auswertung Qualitätssicherungsstudie.....	18
3.	Ergebnisse der Evaluation.....	19
3.1	STICHPROBENREKRUTIERUNG UND STUDIENEINSCHLÜSSE.....	19
3.2	PRIMÄRER ENDPUNKT 1: VERSORGUNGSKOSTEN (GESUNDHEITSÖKONOMISCHE EVALUATION).....	21
3.2.1	Primärer Endpunkt 1: Subgruppenanalyse	26
3.2.2	Primärer Endpunkt 1: Sensitivitätsanalysen	26
3.3	PRIMÄRER ENDPUNKT 2: PSYCHO-SOZIALES-FUNKTIONSNIVEAU (PATIENTENRELEVANTER ENDPUNKT; EFFEKTEVALUATION) 27	
3.3.1	Primärer Endpunkt 2: Subgruppenanalysen.....	27
3.3.2	Primärer Endpunkt 2: Sensitivitätsanalysen.....	28
3.4	PRIMÄRER ENDPUNKT 3: KOSTENEFFEKTIVITÄT (GESUNDHEITSÖKONOMISCHE EVALUATION)	28
3.4.1	Primärer Endpunkt 3: Subgruppenanalyse.....	30
3.4.2	Primärer Endpunkt 3: Sensitivitätsanalysen.....	31
3.5	SEKUNDÄRE HYPOTHESEN UND ENDPUNKTE.....	32
3.6	ZUSÄTZLICHE ANALYSEN	37
3.7	QUALITÄTSSICHERUNGSSTUDIE (TRANSFER NACH ITZEHOE)	38
4.	Schlussfolgerungen und Empfehlungen der drei evaluierenden Institute	39
5.	Literaturverzeichnis	45
6.	Anhang	47
7.	Anlage.....	47

I. Abkürzungsverzeichnis

ACT	Assertive Community Treatment Team
ANCOVA	Kovarianzanalyse
AU	Arbeitsunfähigkeit
CC	Complete Cases
CGI-S	Clinical Global Impressions- Severity Scale
CM	Case Management
CRF	Case Report Form
CRT	Crisis Resolution Team
DGPPN	Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde
DiGA	Digitale Gesundheitsanwendung
EIS	Early Intervention Services
FIMA	Fragebogen zur Inanspruchnahme von medizinischen und nicht-medizinischen Leistungen bei älteren Menschen
FIMPsy	Fragebogen zur Bewertung der Inanspruchnahme medizinischer und nicht medizinischer Ressourcen bei psychischen Störungen
G-BA	Gemeinsamen Bundesausschuss
GAF	Global Assessment of Functioning Scale
HEALTH 49	Hamburger Module zur Beurteilung der psychosozialen Gesundheit
ICC	ICC Intra-Klassen-Korrelation (englisch intraclass correlation)
ICER	Inkrementelle Kosten-Effektivitäts-Relation
ITT	Intention-To-Treat
KIS	Krankenhausinformationssystem
MCO	Managed Care Organisation
OECD	Organisation for Economic Co-operation and Development
QALY	Qualitätskorrigiertes Lebensjahr (englisch quality-adjusted life year)
RCT	Randomised Controlled Trial
Re-QOL-20	Recovering Quality of Life Measure
SAE	Serious Adverse Event
SAP	Statistischer Analyseplan
SF-12	Short Form Health-Questionnaire
SMI	Severe Mental Illness
t0, t6 ,t12	Baseline, Verlaufserhebung 6 Monate, Verlaufserhebung 12 Monate
TAU	Treatment-As-Usual
UKE	Universitätsklinikum Hamburg Eppendorf

II. Abbildungsverzeichnis

Abbildung 1. Studiendesign RECOVER.....	6
Abbildung 2. Flow-Chart RECOVER RCT; Anmerkung: ¹ Baseline imputierte ITT-Stichprobe (Datensatz II); ² vollständig imputierte Stichprobe (Datensatz I).	19
Abbildung 3. Subgruppenanalysen für das primäre Outcome der Effektevaluation (Psycho-Soziales-Funktionsniveau). Für Psycho-Soziales-Funktionsniveau zu Baseline adjustierte Analyse.	28
Abbildung 5. Kosteneffektivitätsakzeptanzkurven für die Subgruppen (Stufe 1-4).....	31

III. Tabellenverzeichnis

Tabelle 1. Beschreibung Stichprobe (Baseline).....	19
Tabelle 3. Unadjustierte Kosten aus gesellschaftlicher Perspektive, jeweils im zurückliegenden 6-Monatszeitraum zu t0, t6 und t12 (in € und Preisen von 2019).....	22
Tabelle 4. Adjustierte Kosten im 12-Monats-Follow-up-Zeitraum nach Kostenkategorien (in € nach Preisen 2019).....	23
Tabelle 5. Ergebnisse patientenrelevanter Endpunkt: Psycho-Soziales-Funktionsniveau (t12-t0)	27
Tabelle 6. Adjustierte Schätzer für QALYs über den 12-Monatszeitraum	28
Tabelle 7. Unadjustierte ICERs	30
Tabelle 8. Ergebnisse sekundäre Endpunkte (kontinuierliche Daten).....	33
Tabelle 9. Ergebnisse sekundäre Endpunkte (binäre Daten).....	35
Tabelle 10. Ergebnisse sekundäre Endpunkte (negative Binomialregression für Zähldaten) .	37
Tabelle 11. Stichprobe Itzehoe (Baseline).....	38
Tabelle 12. Veränderungen über die Zeit (Qualitätssicherungsstudie RECOVER).....	39
Tabelle 2. SAE RECOVER Projekt	21

1. Ziele der Evaluation

Das Ziel von RECOVER war, in der großstädtischen Region Hamburg (Modellphase) und in der ländlich-kleinstädtischen Region Kreis Steinburg (Itzehoe; Transferphase) ein sektoren-übergreifend-koordiniertes, schweregradgestuftes und evidenzbasiertes Versorgungsmodell für Menschen mit psychischen Erkrankungen zu implementieren und wissenschaftlich auf Effektivität und Effizienz zu prüfen. Übergeordnetes Ziel war es beispielhaft für Deutschland zu zeigen, wie sich, unter Nutzung der Regelversorgung und bestehender Selektivverträge, durch eine bessere sektorenübergreifende Zusammenarbeit, Steuerung und Koordination, sowie eine systematische und evidenzbasierte Ergänzung des Versorgungssystems, Behandlungsqualität und Effizienz verbessert. Das Versorgungsziel für Patienten und Angehörige war eine vom Schweregrad unabhängige höhere Chance auf rasche und umfassende Gesundung (RECOVERY). Das Versorgungsziel für das Gesundheitssystem war die Verfügbarkeit eines Referenzmodells, welches moderne und innovative Therapieansätze integriert und Leistungserbringer innerhalb eines Netzwerks aktiv an der Umstrukturierung hin zu einer evidenzbasierten, wohnortnahen und gesellschaftlich eingebundenen Versorgung beteiligt.

Das RECOVER Modell verbindet für die erfolgreiche Umsetzung dieser Ziele neun innovative Komponenten innerhalb eines Versorgungsmodells: (1) Verbesserung von Steuerung und Koordination im Versorgungssystem durch Managed Care; (2) Verbesserung von Diagnostik, Indikationsstellung und ambulanter Krisenintervention durch interdisziplinäre Zentren (Diagnostik, Beratung, psychotherapeutischen Sprechstunden, Home Treatment u.a.); (3) Integration hausärztlicher Versorgung durch e-Health, Schnittstellenmanagement und übergreifende Tätigkeit; (4) Verbesserung der evidenzbasierten Behandlung durch Implementierung eines schweregrad-gestuftes Modells mit vier Stufen und jeweils zugeordneten evidenzbasierten Interventionsmodulen; (5) Integration und Flexibilisierung der ambulanten Psychotherapie durch sechs Anreizsysteme, (6) Integration von bereits entwickelten und neuen E-Mental-Health Modulen inkl. eines Online-Therapieraums; (7) Integration von Supported Employment zur sofortigen und umfassenden Ausbildungs- und Arbeitsintegration; (8) Integration kultur- und sprachsensibler Versorgung für Migrant:innen und Geflüchtete durch Expertengestützte Versorgungsansätze und (9) Integration der Patient:innen- und Angehörigenbeteiligung durch Zusammenarbeit in Organisation, Behandlung (Peer-Arbeit) und Forschung.

2. Darstellung des Evaluationsdesigns

2.1 Studiendesign

Das RECOVER Modell wurde in Hamburg im Rahmen einer prospektiven, randomisiert-kontrollierten Studie (randomised controlled trial, RCT) evaluiert. Die Interventionsgruppe (IG) konnte für 12 Monate die RECOVER Behandlung nutzen, die Kontrollgruppe (KG) erhielt die Regelversorgung (treatment as usual, TAU). Datenerhebungen in beiden Gruppen erfolgten vor der Randomisierung (Baseline, t0), nach 6 Monaten (t6) und nach 12 Monaten (t12). Die Datenerhebung erfolgte durch Personen, die hinsichtlich der Gruppenzuordnung verblindet waren. Die Rekrutierung erfolgt über die Regelversorgung (bspw. UKE, ambulante Kooperationspraxen) und die Studienhomepage (www.recover-hamburg.de). Studienteilnehmer:innen erhielten eine Aufwandsentschädigung für den erhöhten Zeitaufwand während der Follow-up-Visiten. Daten wurden zwischen dem 01.01.2018 und dem 31.01.2021 erhoben (vgl. Abb. 1).

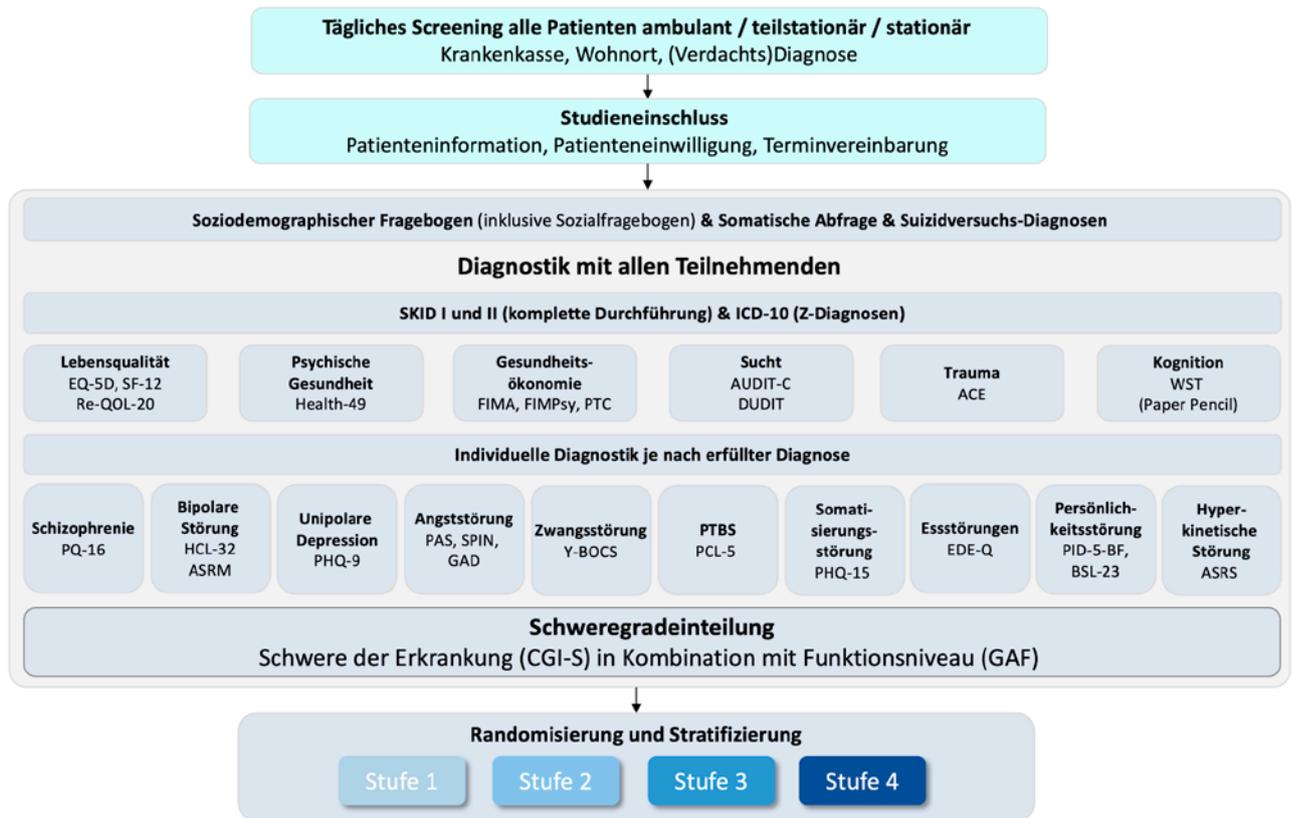


Abbildung 1. Studiendesign RECOVER

2.2 Zielpopulation (Einschluss- und Ausschlusskriterien)

Die Zielgruppe der RECOVER Studie waren Menschen im Einzugsgebiet des UKE (Wohnort max. 8 km vom UKE entfernt) im Alter von ≥ 16 Jahren, versichert bei einer der zu Beginn vier und im Verlauf zwanzig beteiligten Krankenkassen mit mindestens einer relevanten psychischen Erkrankung (ICD-10): schizophrener Formenkreis (F20, F22, F23, F25), bipolare Störung (F31), unipolare Depression (F32, F33), Angststörung (F40, F41), Zwangsstörung (F42), posttraumatische Belastungsstörung (F43.1), Anpassungsstörung (F43.2), somatoforme Störung (F50), Essstörung (F50), Persönlichkeitsstörung (F60, F61) und hyperkinetische Störung (F90). Ausschlusskriterien waren organische Störungen (F00-09) oder Suchtstörungen (F10-19) als Hauptdiagnose. Komorbide Suchtstörungen wurden einbezogen. Eine Teilnahme an der Studie wurde zudem ausgeschlossen für Menschen mit schwerer oder schwerster Intelligenzminderung (F72, F73), unzureichenden Deutschkenntnissen und nicht zu korrigierender Seh- und/oder Hörbeeinträchtigung.

2.3 Fallzahlen (untersuchte Stichprobe) inkl. Drop-Out

Die Stichprobengröße im Erstantrag basierte ursprünglich auf einer Powerkalkulation zur statistisch abgesicherten Unterschiedstestung der Gruppen (IG vs. KG) hinsichtlich des primären Endpunkt „Psycho-Soziales-Funktionsniveau“ (vgl. 2.5.1 Primäre Hypothesen und Endpunkte). Die Stichprobengröße war geeignet, um einen kleinen bis mittleren standardisierten Mittelwertunterschied zwischen den Gruppen (Cohen's f von 0.175) nach 12 Monaten (t_{12}) festzustellen. Bei einer Power von mind. 80% mit einer Typ-1 Fehlerrate von 5% bei einem 2-seitigen

Test und 10% erklärter Varianz durch den Baselinewert, waren mind. 233 Studienteilnehmer:innen nötig, um diesen Effekt zu zeigen. Für die Hypothesen, die Wechselwirkung der Intervention mit den stratifizierten Subgruppen (Schweregradstufen 1-4) einzubeziehen, erhöhte sich die Stichprobengröße auf 383 Teilnehmer. Diagnostik und Therapie sollte in ca. 50 Clustern mit jeweils ca. 21 Teilnehmer:innen erfolgen. Bezogen auf die primären, kontinuierlichen Zielgrößen wurde eine Intra-Klassen-Korrelation (englisch intraclass correlation, ICC) von 0.05 angenommen, so dass sich ein Designeffekt von annähernd 2.0 ergab und 1070 Patient:innen (incl. 30% Dropout) eingeschlossen werden sollten. Bei einer geschätzten Teilnahmequote von 50% hätten 2140 Patient:innen angesprochen werden müssen. Im Verlauf der Studie wurde aufgrund einer Unterrekrutierung von Proband:innen der Stufe 1 (leichter Schweregrad) eine Anpassung der Powerkalkulation vorgenommen, da sich die Rekrutierung der vorab festgelegten 383 Proband:innen in dieser Stufe im gegebenen Projektzeitraum als nicht realisierbar abzeichnete und somit für die notwendige Power eine Überrekrutierung der anderen Schweregradstufen 2-4 geplant wurde. Zwischenzeitlich durchgeführte Hochrechnungen konnten die erfolgreiche Rekrutierung und den Studieneinschluss von 890 Patient:innen in die Studie bestätigen. Es wurde festgelegt, dass von diesen Patient:innen mindestens 60 für die Auswertung in Stufe 1 zur Verfügung stehen sollten. Die Dropout-Rate lag damals - wie im Antrag prognostiziert - bei knapp 30%, so dass davon auszugehen war, dass ca. 623 Patient:innen für die Auswertung zur Verfügung stehen würden. Unter diesen Bedingungen ergab eine post-hoc durchgeführte Poweranalyse, dass diese Fallzahl es ermöglicht, die primären Hypothesen zum Outcome wie kalkuliert mit einer Power von mindestens 80% und einer Mindesteffektstärke von $f=0.175$ zu testen. Bei den sekundären Hypothesen zur Wechselwirkung zwischen den Faktoren Gruppe (IG, KG) und stratifizierten Schweregraden (Stufe 1-4) war mit den geplanten Fallzahlen davon auszugehen, dass mit einer Power von 80% eine Interaktion mit einer Effektstärke von mindestens $f=.20$ nachzuweisen war. Diese lag somit weiterhin, wie anfangs kalkuliert, im Bereich einer kleinen bis mittleren Effektstärke.

Die ursprünglich Planung cluster-randomisiert vorzugehen, wurde im Laufe der Projektplanung und -umsetzung korrigiert, da deutlich weniger klinische Einrichtungen als ursprünglich geplant am Projekt teilnahmen. Dadurch fokussierte sich die Behandlung auf wenige Zentren, sodass ein Cluster-Randomisierung nicht notwendig war, sondern eine individuelle Randomisierung durchgeführt werden konnte. Daraus folgte auch, dass bei der Auswertung der Daten potentielle Cluster nicht modelliert werden mussten.

2.4 Randomisierung und Verblindung

Nach Aufnahme von Patient:innen in die Studie erfolgte eine umfassende Baseline-Diagnostik, anhand derer eine Einteilung der Patient:innen in einen der vier Schweregrade vorgenommen wurde (für Details siehe im Anhang Tabelle A1): Stufe 1 (Leichter Schweregrad), Stufe 2 (Mittlerer Schweregrad), Stufe 3 (Mittlerer bis hoher Schweregrad), Stufe 4 (sehr hoher Schweregrad). Entsprechend dieser vier Schweregrade sollten vier annähernd gleich große Strata gebildet werden (es ist hier - wie bereits im Abschnitt 2.3 dargestellt - darauf hinzuweisen, dass es sich im Verlauf der Studie zeigte, dass sich die geplante Rekrutierung für Stufe 1, d.h. leichter Schweregrad, innerhalb der vorgesehenen Erhebungszeiträume im Projekt als nicht realisierbar abzeichnete). Nach der Diagnostik wurde einem Mitarbeiter des Institut für Medizinische Psychologie bzw. einer Mitarbeiterin des Instituts für Gesundheitsökonomie die pseudonymisierte Patient:innennummer sowie der Schweregrad (Stufe 1-4) per E-Mail oder SMS mitgeteilt. Beide Mitarbeiter:innen waren hinsichtlich Erhebung und Auswertung der Daten verblindet und nahmen anhand einer vorher mit Hilfe des Programms STATA 14 für jeden Schweregrad (Stratum) generierten Liste mit Zufallszahlen die Einteilung auf IG (RECOVER) oder KG

(TAU) vor (Prozedur 'ralloc' mit variabler zufälliger Blocklänge von 2, 4 oder 6 innerhalb jedes Stratums) und übermittelte das Ergebnis zurück an das Diagnostikteam der Evaluation. Damit die Mitarbeiter:innen der Datenerhebung ebenfalls hinsichtlich der randomisierten Gruppe (IG, KG) verblindet blieben, erfolgte die Mitteilung der Gruppenzuordnung an die Patient:innen durch eine:n andere:n Mitarbeiter:in, die/der nicht an der Datenerhebung für diese Person beteiligt war. Für die Verlaufserhebungen wurde darauf geachtet, dass die Datenerhebung ebenfalls nur durch Personen erfolgte, die hinsichtlich der Gruppenzuordnung verblindet waren. Die Evaluator:innen waren bei der Erstellung des statistischen Analyseplans sowie der Analyse der Daten ebenfalls verblindet. Die Entblindung der Gruppen (IG oder KG) erfolgte erst nach Vorliegen aller primären und sekundären Analysen.

Die Datenerhebungen zu Baseline (t0) und den Verlaufserhebungen nach sechs Monaten (t6) sowie zwölf Monaten (t12) erfolgten sowohl durch ein standardisiertes Interview aus relevanten Fragen (sog. Case Report Form, CRF), sowie die Bearbeitung von Fragebögen. Die Person, welche das Interview führte, war verblindet gegenüber der Gruppenzuordnung der Proband:innen. Bei den Fragebögen wurden generische Instrumente, welche allen Studienteilnehmer:innen vorgelegt wurden, von störungsspezifischen Fragebögen, welche nur bei Vorliegen bestimmter Diagnosen vorgegeben wurden, unterschieden.

2.5 Covid-19 Pandemie

Aufgrund der Maßnahmen der Bundesregierung zur Bewältigung der Corona-Pandemie beginnend im Frühjahr 2020, erfolgte eine Anpassung der Datenerhebung zu den Zeitpunkten t6 und t12. Die Erhebung der Baseline-Daten (t0) war zu diesem Zeitpunkt bereits abgeschlossen und nicht von einer Anpassung betroffen. Ab dem 12.03.2020 bis August 2020 erfolgte die Erhebung der Daten über Telefoninterviews, um die Kontakte vor Ort und damit die Gefahr einer Infektion zu reduzieren. Zu diesem Zweck wurden den Studienteilnehmer:innen die Selbsteinschätzungsfragebögen zum Ausfüllen postalisch zugestellt und zurückgesendet. Die Erfassung der fremdbeurteilten Maße erfolgte dann durch ein Telefongespräch. Ab August 2020 war ein Treffen vor Ort wieder offiziell möglich. Um potentielle Verzerrungen der Ergebnisse durch die Erhebungsmethode zu kontrollieren erfolgte ab diesem Zeitpunkt eine randomisierte Zuordnung zum Erhebungsformat der Verlaufsdaten (vor Ort vs. Telefoninterview). Ab Dezember 2020 wurde erneut eine Verordnung erlassen, die Treffen vor Ort untersagte, sodass die letzten Erhebungen bis Januar 2021 ausschließlich über Telefoninterviews stattfanden. Um mögliche Verzerrungen der Ergebnisse für den primären Endpunkt „Psycho-Soziales-Funktionsniveau“ (vgl. Abschnitt 2.6.1) zu untersuchen wurde für jede Person die Art der Visite (alle Verlaufserhebungen (t6/t12) in Präsenz vs. teilweise oder ganz telefonisch) dokumentiert. Der primäre Endpunkt 2 (Psycho-Soziales-Funktionsniveau, siehe 2.6.1) ist für Verzerrungen möglicherweise vulnerabler, da er auf einem Kombinationsmaß (Composite-Score) aus Selbstbeurteilungen und Fremdeinschätzungen hinsichtlich patientenrelevanter Endpunkte errechnet wird, sodass das Erhebungsformat insbesondere für die Fremdeinschätzungen relevant sein könnte.

2.6 Hypothesen und Endpunkte

2.6.1 Primäre Hypothesen und Endpunkte

Es wurden drei primäre Hypothesen formuliert und mit unterschiedlichen primären Endpunkten operationalisiert. Alle Hypothesen bezogen sich auf den Vergleich der Veränderung zwischen den Gruppen (IG vs. KG) von Baseline zum Follow-up 12 Monate nach Einschluss. Die im Folgenden beschriebenen Hypothesen und Endpunkte sind im Studienprotokoll (Lambert et al., 2020) und der Registrierung bei ClinicalTrials.gov (NCT03459664) vorab festgelegt worden. Eventuell notwendige Abweichungen vom ursprünglichen Evaluationskonzept werden an den entsprechenden Stellen erläutert.

1) Primärer Endpunkt 1: Versorgungskosten (gesundheitsökonomische Evaluation)

RECOVER zeigt geringere durchschnittliche Versorgungskosten im 12-Monats-Follow-up-Zeitraum im Vergleich zu TAU. Verschiedene Kostenbereiche wurden mit dem Fragebogen zur Bewertung der medizinischen und nicht-medizinischen Ressourcennutzung bei psychischen Störungen (FIMPsy) (Helen Grupp, König, Riedel-Heller, & Konnopka, 2017) und dem Fragebogen zur Inanspruchnahme medizinischer und nicht-medizinischer Leistungen im Alter (FIMA) (Seidl et al., 2015) erfasst und zu einem Maß "Versorgungskosten" addiert.

2) Primärer Endpunkt 2: Psycho-Soziales-Funktionsniveau (Patientenrelevanter Endpunkt; Effektevaluation)

RECOVER zeigt eine größere Verbesserung des „Psycho-Soziales-Funktionsniveau“ im Vergleich zu TAU. Das „Psycho-Soziales-Funktionsniveau“ wurde als kombiniertes Ergebniskriterium bewertet (Aiken, 1987) Zu diesem Zweck wurden die Werte der Hamburger Module zur Beurteilung der psychosozialen Gesundheit (HEALTH 49) (Rabung, Harfst, Koch, Wittchen, & Schulz, 2007) der fremdeingeschätzte Wert der Global Assessment of Functioning Scale (GAF) (Gold, 2014) und die psychische Summenskala des Short Form Health-Questionnaire (SF-12) (Ware, Kosinski, & Keller, 1996) linear transformiert und zu einem Maß "Psycho-Soziales-Funktionsniveau" addiert. Die geplanten Auswertungen wurden im entsprechenden statistischen Analyseplan (SAP) festgehalten.

3) Primärer Endpunkt 3: Kosteneffektivität (gesundheitsökonomische Evaluation)

RECOVER zeigt eine Verbesserung der Kosteneffektivität im Vergleich zu TAU. Es wurde die inkrementelle Kosten-Effektivitäts-Relation (ICER) berechnet, die dem Verhältnis zwischen der Kostendifferenz zwischen RECOVER und TAU und der Differenz der qualitätsbereinigten Lebensjahre (QALYs) zwischen RECOVER und TAU entspricht.

2.6.2 Sekundäre Hypothesen und Endpunkte

Es wurden 18 sekundäre Endpunkte für die RECOVER Studie definiert:

- 1) **Schweregrad der Erkrankung:** RECOVER führt im Vergleich zu TAU zu einer stärkeren Verringerung des Schweregrads der Erkrankung. Die Erfassung erfolgte mit der Clinical Global Impression - Severity Scale (CGI-S) (Guy, 1976), bei der der Schweregrad der Krankheit des Patient:innen zum Zeitpunkt der Bewertung im Vergleich zu den Schweregraden von Patient:innen mit derselben Diagnose bewertet wurde. Folgende Bewertungen waren möglich: 0 = nicht zutreffend; 1 = normal, überhaupt nicht krank; 2 = grenzwertig psychisch krank; 3 = leicht krank; 4 = mäßig krank; 5 = ausgeprägt krank; 6 = schwer krank; 7 = zu den extrem kranken Patient:innen gehörend.
- 2) **Globale Funktionsfähigkeit:** RECOVER führt im Vergleich zu TAU zu einer größeren Verbesserung der globalen Funktionsfähigkeit, die mit der Global Assessment of Functioning Scale (GAF) (Gold, 2014) bewertet wurde. Die GAF-Werte reichten von 1 (niedrigstes Funktionsniveau) bis 100 (höchstes Funktionsniveau).
- 3) **Verbesserte Lebensqualität:** RECOVER führt im Vergleich zu TAU zu einer höheren Lebensqualität, die mit dem Re-QOL-20 Fragebogen (Keetharuth et al., 2018) sowie dem EQ-5D-5L gemessen wurde. Die Werte des Re-QOL-20 reichten von 0 bis 80, wobei 0 für die schlechteste Lebensqualität und 80 für die höchste Lebensqualität steht. Die Werte des EQ-5D-5L (Janssen et al., 2013) wurden in einen Indexwert zwischen -0,661 für den schlechtmöglichen Gesundheitszustand und 1 für den bestmöglichen Gesundheitszustand überführt (Ludwig, Graf von der Schulenburg, & Greiner, 2018).
- 4) **Symptomatische Remission:** RECOVER führt im Vergleich zu TAU zu einem höheren Anteil an symptomatischer Remission (Anzahl Patient:innen). Symptomatische Remission war nach Guy et al. (1976) definiert mit einem Wert von ≤ 3 Punkten im Schweregrad-Score der CGI-S (nicht größer als "mild") für ≥ 6 Monate. Zusätzlich wurde explorativ ein diagnoseübergreifender Wert basierend auf Fragebogenbasierten Symptomeinschätzungen der Patient:innen errechnet. Dieser Wert wurde durch Standardisierung der Summenwerte in diagnose-spezifischen Symptomfragebögen errechnet.
- 5) **Funktionelle Remission:** RECOVER führt im Vergleich zu TAU zu einer höheren funktionellen Remission (Anzahl Patient:innen). Funktionelle Remission war nach dem Kriterium von Albert et al. (2012), gemessen mit der GAF (Gold, 2014) erfüllt, wenn ein Wert von ≥ 60 Punkten für ≥ 6 Monate anhielt.
- 6) **Rate der psychiatrischen Krankenhausaufenthalte:** RECOVER führt im Vergleich zu TAU zu einer größeren Reduktion der psychiatrischen Hospitalisierungen (Anzahl der Hospitalisierungen in Interview erfasst).
- 7) **Dauer psychiatrischer Krankenhausaufenthalte:** RECOVER führt im Vergleich zu TAU zu einer größeren Verringerung der Dauer psychiatrischer Hospitalisierungen (Anzahl Hospitalisierungstage in Interview erfasst)

- 8) **Arbeitsunfähigkeit:** RECOVER führt im Vergleich zu TAU zu einer größeren Verringerung der Arbeitsunfähigkeit (Anzahl Arbeitsunfähigkeitstage in Interview erfasst)
- 9) **Verzögerung der Behandlung:** RECOVER führt im Vergleich zu TAU zu einer stärkeren Reduktion des diagnosespezifischen Behandlungsbeginnes (Anzahl Tage).
- 10) **Evidenzbasierte Interventionen:** RECOVER führt im Vergleich zu TAU zu einer größeren Zunahme der leitliniengerechten Behandlung, je nach spezifischer Diagnose und Schweregrad (Anzahl Patient:innen mit leitliniengerechter Behandlung).
- 11) **Ambulante psychotherapeutische Interventionen bei schweren psychischen Störungen:** RECOVER führt im Vergleich zu TAU zu einem größeren Anstieg an Patient:innen mit schweren psychischen Störungen, die ambulant-psychotherapeutische Interventionen erhalten (Anzahl Patient:innen die in Fragebogen Erhalt von ambulanter Therapie angeben)
- 12) **Psychotherapeutische Gruppeninterventionen:** RECOVER führt im Vergleich zu TAU zu einem größeren Anstieg der psychotherapeutischen Gruppeninterventionen (Anzahl Patient:innen, die in Fragebogen Erhalt von ambulanter Gruppentherapie angeben).
- 13) **Psychotherapeutische Kurzzeitinterventionen:** RECOVER führt im Vergleich zu TAU zu einer größeren Zunahme der psychotherapeutischen Kurzzeitinterventionen (Anzahl Patient:innen, die in Fragebogen Erhalt von ambulanter Kurzzeittherapie angeben).
- 14) **Spezifische psychiatrische Interventionen bei schweren psychischen Störungen:** RECOVER führt im Vergleich zu TAU zu einem größeren Anstieg spezifischer psychiatrischer Interventionen für schwere psychische Störungen (Anzahl Kontakte zur Krisenbewältigung/ assertive community treatment, die in Fragebogen angegeben werden).
- 15) **Behandlungskontinuität:** RECOVER führt im Vergleich zu TAU zu einer größeren Verbesserung der Behandlungserhaltungsrate (Tage bis zum Abbruch einer Behandlung).
- 16) **Behandlungsabbruch:** RECOVER führt im Vergleich zu TAU zu einer größeren Reduktion Behandlungsabbrüchen während der Behandlung (Anzahl Behandlungsabbrüche, die in Fragebogen angegeben werden).
- 17) **Web-basierte Therapie:** RECOVER führt im Vergleich zu TAU zu einer häufigeren Nutzung webbasierter Therapieangebote (Anzahl Nutzer:innen, die in Fragebogen Anwendung von webbasierter Therapie angeben).
- 18) **Peer-Unterstützung:** RECOVER führt im Vergleich zu TAU zu einer häufigeren Nutzung von Peer-Support (Anzahl Nutzer:innen, die in Fragebogen Erhalt von Peer-Support angeben).

2.6.3 Sekundärdaten

Zu Beginn der Datenerhebungen wurden nur Patient:innen eingeschlossen, die in der Krankenkasse BARMER, DAK-Gesundheit, HEK oder AOK Rheinland/Hamburg versichert waren. Hintergrund war, dass die Versorgung der Patient:innen in Stufe 4 in RECOVER im Rahmen der Besonderen Versorgung nach § 140a SGB V erfolgte und anfangs nur mit den zuvor genannten Krankenkassen Selektivverträge bestanden. Im Studienverlauf haben zusätzlich die folgenden zwölf Betriebskrankenkassen an der RECOVER-Studie teilgenommen: BKK Salzgitter, BKK Public, WMF Betriebskrankenkasse, BKK RWE, BKK Mobil Oil, BKK Linde, Salus BKK, BKK Technoform, BKK VerbundPlus, TUI BKK, Heimat BKK, BKK Continentale.

Ursprünglich war angedacht, zusätzlich zur Hauptanalyse mit Primärdaten auch Krankenkassendaten in die Analysen einzuschließen. Auf die ergänzende Analyse wurde unter anderem verzichtet, da sich mit den zusätzlichen Patient:innen aus den Betriebskrankenkassen die Patient:innenstichproben der Primärdatenanalyse (alle beteiligten Krankenkassen) und Sekundärdatenanalyse (nur die Krankenkassen BARMER, DAK-Gesundheit, HEK und AOK Rheinland/Hamburg) unterscheiden hätten und somit keine vergleichenden Analysen möglich gewesen wären. Hinzu kam, dass sich eine Analyse von Krankenkassendaten innerhalb der Projektlaufzeit aufgrund der Vielzahl der beteiligten Kassen und der Fristen für die Datenbereitstellung aufgrund des erforderlichen Datennachlaufs als unrealistisch herausstellte.

2.7 Schwerwiegende unerwünschte Ereignis (SAE)

In dieser Studie waren folgende unerwünschte Ereignisse denkbar: Symptomverstärkung, Auftreten neuer Symptome, Auftreten von passiven bzw. aktiven Suizidgedanken oder Suizidplänen, Probleme in der Patient-Therapeut-Beziehung, Probleme in der Therapiegruppe und andere nachteilige Vorfälle (wie zum Beispiel eine Grippe, ein Beinbruch, ein Sport- oder Verkehrsunfall etc.).

Serious (ernsthafte) Adverse Events (SAE) sind definiert als unerwünschte medizinische Ereignisse, die den Tod einer Person zur Folge haben, unmittelbar lebensbedrohend sind (z.B. Suizidversuch), eine stationäre Behandlung oder eine Verlängerung des stationären Aufenthalts des Prüfungsteilnehmers erforderlich machen, zu bleibenden oder signifikanten Schäden/Behinderungen führen oder eine angeborene Missbildung bzw. einen Geburtsfehler darstellen.

Das Auftreten von SAE wurde zu den Verlaufsvisiten (t6 und t12) bei allen Studienteilnehmer:innen aktiv erfragt und dokumentiert. Das SAE-Protokoll wurde Teil der Patientenakte.

2.8 Statistische Auswertung

Die Datenauswertung für die Effektevaluation wurde durch das Institut für Medizinische Biometrie und Epidemiologie, UKE (Prof. Dr. Antonia Zapf) in enger Abstimmung mit dem Institut für Medizinische Psychologie des UKE (Dr. Judith Peth, Prof. Dr. Holger Schulz) durchgeführt. Für die gesundheitsökonomische Auswertung war das Institut für Gesundheitsökonomie und Versorgungsforschung, UKE (Prof. Hans-Helmut König) zuständig. Die Datenauswertung erfolgte erst nach Abschluss der letzten Verlaufsdatenerhebung (im Januar 2021).

Alle angewendeten statistischen Tests waren zweiseitig und wurden mit einem Signifikanzniveau von 5% durchgeführt. Alle Konfidenzintervalle waren 95% und zweiseitig. Die Auswertung der sekundären Endpunkte ist deskriptiv. Eine Adjustierung für multiples Testen wurde nicht durchgeführt.

2.8.1 Fehlende Daten

Die Datensätze der primären und sekundären Analysen basieren grundsätzlich auf der Intention-to-treat (ITT) Stichprobe, die aus allen zur Baseline-Erhebung (t0) randomisierten Patient:innen besteht. Fehlende Werte in den Baselinedaten (t0) wurden mit EM-Algorithmus (expectation maximization) (10) imputiert und fließen in alle primären und sekundären Analysen

ein. Darauf aufbauend wurden für die vorliegenden Fragestellungen zwei ITT-Datensätze erstellt, die sich hinsichtlich des Ersetzens von fehlenden Datenpunkten zu t6 bzw. t12 unterscheiden. Eine Definition der verwendeten Datensätze wird im Folgenden dargestellt:

- I. Vollständig imputierte ITT-Stichprobe (Datensatz I, gesundheitsökonomische Evaluation): Enthält alle randomisierten Personen. In diesem Datensatz wurden fehlende Datenpunkte zu t6 und t12 mittels EM-Algorithmus ersetzt.
- II. Baseline imputierte ITT-Stichprobe (Datensatz II, Effektevaluation): Enthält alle randomisierten Personen, für die keine fehlenden Daten zu t6 und t12 vorlagen (d.h. es wurden keine fehlenden Werte zu t6 bzw. t12 ersetzt).

Entsprechend den a priori definierten Auswertungen wurde die vollständig imputierte ITT-Stichprobe für die primäre Analyse der Endpunkte 1 und 3 verwendet und die Baseline imputierte ITT-Stichprobe für die Sensitivitätsanalyse der Endpunkte 1 und 3. Für die primäre Auswertung von Endpunkt 2 wurde die Baseline imputierte ITT-Stichprobe zugrunde gelegt und entsprechend die vollständig imputierte ITT-Stichprobe für eine Sensitivitätsanalyse des Endpunkts 2.

Darüber hinaus wurden drei weitere Datensätze für Sensitivitätsanalysen für den primären Endpunkt 2, d.h. die Effektevaluation, genutzt:

- III. Per Protokoll Stichprobe (Datensatz III): Dient für eine Sensitivitätsanalyse für den primären Endpunkt der Effektevaluation. Datenpunkte, deren Erhebung mehr als sechs Wochen nach dem errechneten Erhebungstermin stattfanden, wurden nicht in die Auswertung mit einbezogen. Auch Patient:innen mit Therapieabbruch wurden hierbei ausgeschlossen. Es erfolgt keine Ersetzung fehlender Werte zu t6 und t12.
- IV. Complete Case Stichprobe (Datensatz IV): Dient für eine Sensitivitätsanalyse für den primären Endpunkt der Effektevaluation. Enthält alle randomisierten Personen, für die zu Baseline (t0) und t12 alle für die Analyse benötigten Werte vollständig vorlagen (d.h. es erfolgte zu keinem Zeitpunkt eine Ersetzung fehlender Werte).
- V. Multiple Imputation Stichprobe (Datensatz V): Dient für eine Sensitivitätsanalyse für den primären Endpunkt der Effektevaluation. Hierbei wurden fehlende Werte in der Baseline mit EM-imputierten Werten und Werte zu t6 und t12 durch multiple Imputation (10) ersetzt.

2.8.2 Deskription der Baseline-Daten

Für die gesamte Stichprobe erfolgte eine Deskription der Baseline-Daten (t0). Zu den beschriebenen Patient:innen-Charakteristika gehörten: Alter, Geschlecht, Psycho-Soziales-Funktionsniveau, Schulabschluss, Hauptdiagnose, Arbeitsstatus, Familienstatus, Wohnverhältnisse, stationäre oder ambulante Behandlung zu Baseline, Anzahl an Hospitalisierungen zu Baseline, Anzahl von Selbstmordversuchen zu Baseline, Anzahl an psychiatrischer Verwahrung zu Baseline, Familienanamnese, Trauma-Geschichte, Komorbiditäten. Kategoriale Daten wurden über absolute und relative Häufigkeiten beschrieben. Für kontinuierliche Daten wurden Mittelwert und Standardabweichung angegeben. Die Anzahl der beobachteten Werte sowie der fehlenden Werte wurden gesamt und pro Gruppe (IG, KG) beschrieben. Statistische Tests für den Vergleich der Baseline-Werte zwischen den beiden Gruppen wurden nicht durchgeführt.

2.8.3 Primärer Endpunkt 1: Versorgungskosten (gesundheitsökonomische Evaluation)

Für die Analyse der Versorgungskosten wurden die Inanspruchnahme von Versorgungsleistungen sowie Arbeitsunfähigkeitstage mittels Fragebögen zu t0, t6 und t12 erfasst. Diese gemessene Inanspruchnahme wurde mit durchschnittlichen Kostensätzen multipliziert, um die direkten Kosten zu erhalten (Bock et al., 2015; H. Grupp, König, & Konnopka, 2017). Indirekte Kosten durch Produktivitätsverluste wurden mit dem Human-Kapital-Ansatz ermittelt. Dabei wurden die erfassten Arbeitsunfähigkeitstage mit durchschnittlichen Bruttolöhnen, ergänzt um die Arbeitgeberbeiträge zur Sozialversicherung, multipliziert (Statistisches Bundesamt, 2018, 2019). Alle durchschnittlichen Kostensätze wurden auf Preise aus dem Jahr 2019 inflationiert. Da die Inanspruchnahme zu den Erhebungszeitpunkten jeweils retrospektiv über einen Zeitraum von sechs Monaten erfasst wurde, konnten die Versorgungskosten für den gesamten Behandlungszeitraum auf Grundlage der zu t6 und t12 gemessenen Inanspruchnahme abgebildet werden. Die zu t0 retrospektiv gemessene Inanspruchnahme diente zur Berechnung der Baseline-Kosten, die als Kontrollvariable verwendet wurden (s.u.). Die Gesamtkosten wurden durch Summation aller Einzelkosten errechnet.

Als Basisanalyse der Daten wurde eine gesellschaftliche Perspektive angesetzt, bei der alle Kosten berücksichtigt wurden. Die Analyse der Kosten aus gesellschaftlicher Perspektive umfasst die folgenden Leistungen, für die eine Ersetzung fehlender Einzelwerte durchgeführt werden konnte: Stationäre Aufenthalte, ambulante Arztkontakte, ambulante Heilmittelerbringer, Medikamente, Pflege/Hilfe, informelle Pflege, Reha, Beratung & Betreuung, berufliche Integration, Krisen-Interventionen, webbasierte Angebote, Peerbegleitung und indirekte Kosten durch Arbeitsunfähigkeits(AU)-Tage.

Die Analyse des primären Endpunkts 1) Versorgungskosten erfolgte zunächst als unadjustierter Vergleich der direkten, indirekten und gesamten Kosten zu den Messzeitpunkten t0, t6 und t12 zwischen den Gruppen (IG, KG), ergänzt um eine Berechnung der durchschnittlichen Kostendifferenz zwischen den beiden Gruppen. Zur Prüfung von Hypothese 1) wurden adjustierte Vergleiche der gesamten Kosten und einzelnen Kostenkategorien für den Behandlungszeitraum von 12 Monaten durchgeführt. Die Adjustierung erfolgte mit generalisierten linearen Modellen mit Poisson-Verteilung und Log-Link Funktion. Dieses Vorgehen ist besonders gut geeignet für Analysen von schiefen Verteilungen, wie man sie üblicherweise bei Kostendaten findet. Im Rahmen der Regression wurde für verschiedene Variablen kontrolliert (Alter, Geschlecht, Baseline-Kosten, Somatische Erkrankungen, Krankheitsschwere, Erwerbstätigkeit).

2.8.3.1 Primärer Endpunkt 1: Subgruppenanalysen

In Subgruppenanalysen wurden adjustierte Vergleiche der gesamten Kosten und einzelnen Kostenkategorien für den Behandlungszeitraum von 12 Monaten nach Schweregraden (Stufe 1-4) durchgeführt. Die Subgruppenanalyse erfolgte für jede Stufe einzeln, nach derselben Methodik wie in der Basisanalyse.

2.8.3.2 Primärer Endpunkt 1: Sensitivitätsanalysen

Für den primären Endpunkt 1 wurde als Sensitivitätsanalyse eine Auswertung in der Baseline imputierten ITT-Stichprobe (Datensatz II), durchgeführt. Eine adjustierte Auswertung einzelner Kostenarten war für diesen Datensatz jedoch nicht möglich, da die Regressionsmodelle aufgrund der geringen Stichprobengröße oftmals nicht mehr konvergierten.

2.8.4 Primärer Endpunkt 2: Psycho-Soziales-Funktionsniveau (Patientenrelevanter Endpunkt; Effektevaluation)

Die Analysen für den primären Endpunkt 2) Psycho-Soziales-Funktionsniveau folgten dem statistischen Analyseplan (SAP), welcher vor Entblindung der auswertenden Institution (Institut für Medizinische Biometrie und Epidemiologie, UKE) erstellt wurde. Für den primären Endpunkt Psycho-Soziales-Funktionsniveau wurde die Differenz zwischen t0 und t12 in den beiden Gruppen (IG, KG) verglichen. Das Psycho-Soziale-Funktionsniveau war hierbei ein kombinierter gleichgewichteter Wert aus A) Symptom- und Krankheitsschwere (Modul A des HEALTH-49) (Rabung et al., 2007), B) globalem Funktionsniveau (GAF) (Gold, 2014) und C) gesundheitsbezogener psychischer Lebensqualität (Psychische Summenskala des SF 12) (Ware et al., 1996). Es wurden die Fragen von Komponente A auf einer fünfstufigen Skala (Stufen 0-4) gemessen, die Komponente B wurde auf einer Skala von 1 bis 100 gemessen und die Komponente C war eine Mischung aus zwei- bis sechsstufigen Antwortskalen, die gewichtet in die Berechnung des Wertes der Psychischen Summenskala eingehen und von 0-100 reichen können. Da die drei Maße unterschiedliche Wertebereiche aufweisen wurden sie zunächst standardisiert, d.h. für jeden patientenbezogenen Score einer Komponente wurde der Studienmittelwert der Komponente abgezogen und diese Differenz durch die entsprechende Standardabweichung dividiert. Anschließend konnten die standardisierten Scores summiert werden (die drei Komponenten wurden dabei gleich gewichtet). Dies erfolgte für alle Zeitpunkte. Für den primären Endpunkt wurde die Differenz zwischen t0 und t12 berechnet. Mit diesem Differenzwert wurde ein lineares Modell (ANCOVA) berechnet mit den Kovariaten Gruppe (IG, KG), Schweregrad (Stufe 1-4) und Baselinewert des Psycho-Sozialen-Funktionsniveaus. Interaktion zwischen Gruppe und Schweregrad wurden ebenfalls berücksichtigt und nur dann ausgeschlossen, wenn der Interaktions-p-Wert $\geq .05$ war.

Folgende Subgruppen wurden auf Basis der Baseline imputierten ITT-Stichprobe (Datensatz II) analysiert:

- Geschlecht (männlich, weiblich)
- Schweregrad (Stufe 1, 2, 3, 4)
- Alter (≤ 18 vs. > 18 Jahre)
- Diagnosegruppe ICD-10 (F2, F3, F4, F6 vs. andere)
- Datenerhebung Verlaufstermine (t6/t12): alle in Präsenz vs. teilweise oder ganz telefonisch
- Covid-19 Pandemie: Erhebung vor Lockdown vs. nach Lockdown
- Stationäre Vorbehandlung: Ja vs. Nein

Zur Analyse wurde ein lineares Modell (ANCOVA) verwendet, wobei die Kovariaten Gruppe (IG, KG), Schweregrad (Stufe 1-4) und Baselinewert des Psycho-Sozialen-Funktionsniveau, sowie die Interaktion zwischen diesen Subgruppen und der Behandlungsgruppe einbezogen wurden.

2.8.4.1 Primärer Endpunkt 2: Sensitivitätsanalyse

Zusätzlich wurden folgende Sensitivitätsanalysen analog zu der im Abschnitt 2.7.4 beschriebenen Analyse durchgeführt:

- Analyse basierend auf vollständig mit dem EM-Algorithmus imputierten Daten (Vollständig EM-imputierte ITT-Stichprobe, Datensatz I)
- Analyse basierend auf multiple imputierten Daten zu t6 und t12 (Datensatz V)
- Zur Gewichtung der Komponenten des zusammengesetzten primären Endpunkts wurde die sogenannte „Optimal weights method“ (Govindarajulu, 1988) verwendet (Datensatz II)
- Analyse basierend auf der Per-Protocol Stichprobe (Datensatz III)
- Analyse basierend auf der Complete Case Stichprobe (Datensatz IV), d.h. ohne ersetzte Werte zu Baseline (t0) oder Follow-Up (t12)

2.8.5 Primärer Endpunkt 3: Kosteneffektivität (gesundheitsökonomische Evaluation)

Für die Analyse der Kosteneffektivität wurden die Kostendaten verwendet, welche bereits im Rahmen der Analyse des primären Endpunktes 1) ermittelt wurden. Als Basisanalyse wurde analog zu der Analyse des primären Endpunkts 1) eine gesellschaftliche Perspektive angesetzt. Zusätzlich wurden als Maß für die Effektivität qualitätsgewichtete Lebensjahre (QALYs) berechnet. Zu diesem Zweck wurde zu t0, t6 und t12 der EQ-5D-5L Lebensqualitätsfragebogen (Janssen et al., 2013) eingesetzt. Auf Grundlage der gemessenen EQ-5D-5L-Gesundheitszustände wurden die zugehörigen EQ-5D-Index-Werte (Utilities) bestimmt, die auf Präferenzen der deutschen Allgemeinbevölkerung basieren (Ludwig et al., 2018). Zur Berechnung von QALYs für den 12-Monats-Behandlungszeitraum wurden die EQ-5D-Index-Werte zu t0, t6 und t12 linear interpoliert (area under the curve Methode).

Als eigentliches Outcome des primären Endpunktes 3) wurde die inkrementelle Kosteneffektivitäts-Relation (ICER) ermittelt, bei der die Differenz in den durchschnittlichen Kosten zwischen den Gruppen (IG, KG) zur Differenz in den durchschnittlichen QALYs zwischen beiden Gruppen in Relation gesetzt werden. Bei der ICER handelt es sich um einen unadjustierten Punktschätzer. Um die statistische Unsicherheit der ICER abzuschätzen, wurden Kosten-Effektivitäts-Akzeptanz-Kurven ermittelt. Zu diesem Zweck wurden für verschiedene Zahlungsbereitschaften pro QALY Net-Benefits berechnet und mittels OLS Regression der Schätzer für den Gruppeneffekt (Unterschied zwischen IG und KG) auf den Net-Benefit ermittelt, wobei für Alter, Geschlecht, Baseline-Kosten, Baseline-Utilities, somatische Erkrankungen, Krankheits-schwere und Erwerbstätigkeit kontrolliert wurde. Der p-Wert des Schätzers für den Gruppeneffekt wurde verwendet, um die Kosten-Effektivitäts-Akzeptanz-Kurven zu ermitteln, welche in Abhängigkeit von der angesetzten Zahlungsbereitschaft die Wahrscheinlichkeit aufzeigen, dass die IG im Vergleich zur KG kosteneffektiv ist (Zethraeus, Johannesson, Jönsson, Löthgren, & Tambour, 2003). Die Kosten-Effektivitäts-Akzeptanz-Kurven wurden für die gesellschaftliche Perspektive und zusätzlich für die GKV-Perspektive durchgeführt. Bei der GKV-Perspektive wurden die folgenden Leistungen berücksichtigt: stationäre Aufenthalte, ambulante Arztkontakte, ambulante Heilmittelerbringer, Medikamente, ambulante psychiatrische Pflege, bezahlte Haushaltshilfe, Krisen-Interventionen, Webbasierte Angebote, Peerbegleitung.

2.8.5.1 Primärer Endpunkt 3: Subgruppenanalysen

In Subgruppenanalysen wurden Kosten-Effektivitäts-Akzeptanz-Kurven nach Schweregraden (Stufe 1-4) aus der gesellschaftlichen Perspektive ermittelt. Die Subgruppenanalyse erfolgte für jede Stufe einzeln, nach derselben Methodik wie in der Basisanalyse.

2.8.5.2 Primärer Endpunkt 3: Sensitivitätsanalysen

In der ersten Sensitivitätsanalyse wurden aus gesellschaftlicher Perspektive zusätzlich Kostenkategorien berücksichtigt, bei denen keine fehlenden Werte mittels EM-Imputation ersetzt werden konnten. Dafür wurde für die fehlenden Werte die Annahme getroffen, dass die entsprechende Leistung nicht in Anspruch genommen wurde, was einem Wert von 0 entspricht. In der zweiten Sensitivitätsanalyse wurden alle Szenarien alternativ in der Baseline imputierten ITT-Stichprobe (Datensatz II) ausgewertet.

2.8.6 Sekundäre Endpunkte

Alle sekundären Endpunkte wurden auf Basis der Baseline imputierten ITT-Stichprobe (Datensatz II) analysiert. Für die sekundären Endpunkte 1) Schweregrad der Erkrankung (CGI-S) (Guy, 1976), 2) Globale Funktionsfähigkeit (GAF) (Gold, 2014), 3) Verbesserte Lebensqualität (ReQOL-20; EQ-5D) sowie die drei einzelnen Skalenwerte des primären Endpunkts wurde die Differenz zwischen t0 und t12 berechnet und zwischen den Gruppen (IG, KG) verglichen. Ebenso wurde für den weiteren sekundären Endpunkt „Diagnosespezifische Symptomstärke“ vorgegangen. Dabei wurden lineare Regressionsmodelle (ANCOVA) verwendet und Gruppe (IG, KG), Schweregrad (Stufe 1-4) und der entsprechende Ausgangswert für die Ergebnisvariable als Kovariaten einbezogen. Die Interaktion zwischen Schweregrad und Gruppe wurde ebenfalls berücksichtigt und entfernt, wenn der p-Wert der Interaktion über .05 lag.

Für die sekundären Endpunkte 4) Symptomatische Remission (CGI-S), 5) Funktionelle Remission (GAF), 7) Dauer psychiatrischer Krankenhausaufenthalte (Anzahl Tage), 8) Arbeitsunfähigkeit (Anzahl Tage) wurden Vergleiche zwischen den Gruppen (IG, KG) berechnet, wobei Baseline-Korrekturen erfolgten (s.u.).

Für die weiteren sekundären Endpunkte wurden die Werte zu t12 zwischen den Gruppen (IG, KG) verglichen, wobei keine Baseline-Korrektur (aufgrund fehlender Erfassung) vorgenommen wurde. Dies betrifft die sekundären Endpunkte: 6) Rate der psychiatrischen Krankenhausaufenthalte (Anzahl Aufenthalte), 11) Ambulante psychotherapeutische Interventionen bei schweren psychischen Störungen (Anzahl Patient:innen), 12) Psychotherapeutische Gruppeninterventionen (Anzahl Patienten), 13) Psychotherapeutische Kurzzeitinterventionen (Anzahl Patient:innen), 14) Spezifische psychiatrische Interventionen bei schweren psychischen Störungen (Anzahl Kontakte zur Krisenbewältigung / assertive community treatment), 16) Behandlungsabbruch (Anzahl Behandlungsabbrüche), 17) Web-basierte Therapie (Anzahl Nutzer), 18) Peer-Unterstützung (Anzahl Nutzer).

Zur Analyse der sekundären Endpunkte 4 -18 (s.o.) wurden folgenden Modelle verwendet:

- Für binäre Endpunkte wurde eine logistische Regression zur Analyse verwendet. Dabei wurden Gruppe und Schweregrad als Kovariaten einbezogen (Baseline-Anpassung, wenn Ausgangswerte gegeben). Die Interaktion zwischen Schweregrad und Gruppe wurde ebenfalls berücksichtigt und entfernt, wenn der p-Wert der Interaktion über .05 lag. Die adjustierten Raten und Odds Ratios wurden mit zweiseitigen 95% Konfidenzintervallen angegeben.
- Für Zähldaten wurde die negative Binomialregression verwendet. Dabei wurden Gruppe und Schweregrad sowie der Ausgangswert (sofern vorhanden) als Kovariaten einbezogen. Die Interaktion zwischen Schweregrad und Gruppe wurde ebenfalls berücksichtigt und entfernt, wenn die Interaktion einen p-Wert über .05 aufwies. Die Offset-Variable war der individuell gültige logarithmierte Beobachtungszeitraum.

Für die ursprünglich geplanten sekundären Endpunkte 9) Verzögerung der Behandlung, 10) Evidenzbasierte Interventionen und 15) Behandlungskontinuität erfolgte keine Auswertung, da keine Daten vorlagen. Die sekundären Endpunkte wurden explorativ ohne Anpassung für multiples Testen analysiert.

2.8.7 Sekundäre Analysen

Als zusätzliche Auswertungen für den primären Endpunkt 2 Psycho-Soziales-Funktionsniveau und die sekundären Endpunkte wurden Analysen unter Berücksichtigung des Zeitverlaufs durchgeführt. Im Gegensatz zu den vorher beschriebenen Auswertungen wurden hier nun auch die Daten zu t6 berücksichtigt. Alle Endpunkte, bei denen Messungen zu Studienbeginn (Baseline) und t6 bzw. t12 angegeben waren, wurden anhand der entsprechenden gemischten Regressionsmodelle analysiert. In dem Modell wurde Gruppe (IG, KG) und Schweregrad (Stufe 1-4) sowie der entsprechende Ausgangswert für die Ergebnisvariable und eine Interaktion zwischen Zeit (t6 und t12) und Gruppe einbezogen. Die Interaktion zwischen Zeit, Gruppe und Schweregrad wurde nur einbezogen, wenn der entsprechende Interaktions-p-Wert zwischen Schweregrad und Gruppe in den oben beschriebenen Modellen ohne Zeit-Interaktion unter .05 lag. Die Probandennummer wurde als zufälliger Effekt, d.h. als ein zufälliger Intercept, einbezogen.

2.8.8 Auswertung Qualitätssicherungsstudie

Die Ergebnisse der Implementierung des RECOVER Modells in der ländlichen Region Kreis Steinburg (Itzehoe) wurden anhand deskriptiver Analysen der soziodemographischen und diagnostischen Daten ausgewertet. Es wurden zu Beginn der RECOVER-Behandlung (Baseline t0) und nach 6 Monaten (t6) sowie nach 12 Monaten (t12) ebenfalls mittels Interview und Fragebögen verschiedene klinische Angaben untersucht, zum Beispiel Selbsteinschätzungen zur Lebensqualität mittels SF-12 (Ware et al., 1996) und Fremdeinschätzungen zum Schweregrad der Erkrankung durch CGI-S (Guy, 1976) und die globale Funktionsfähigkeit mittels GAF (Gold, 2014).

3. Ergebnisse der Evaluation

3.1 Stichprobenrekrutierung und Studieneinschlüsse

Die finalen Fallzahlen für die Rekrutierung und Studieneinschlüsse in das RCT sind in Abb. 2 dargestellt. Bei der Rekrutierung erfolgte die Zuführung für 77,5 % der Personen über das UKE, für 12,2 % durch Eigeninitiative und 10,3 % durch externe Kooperationspartner:innen

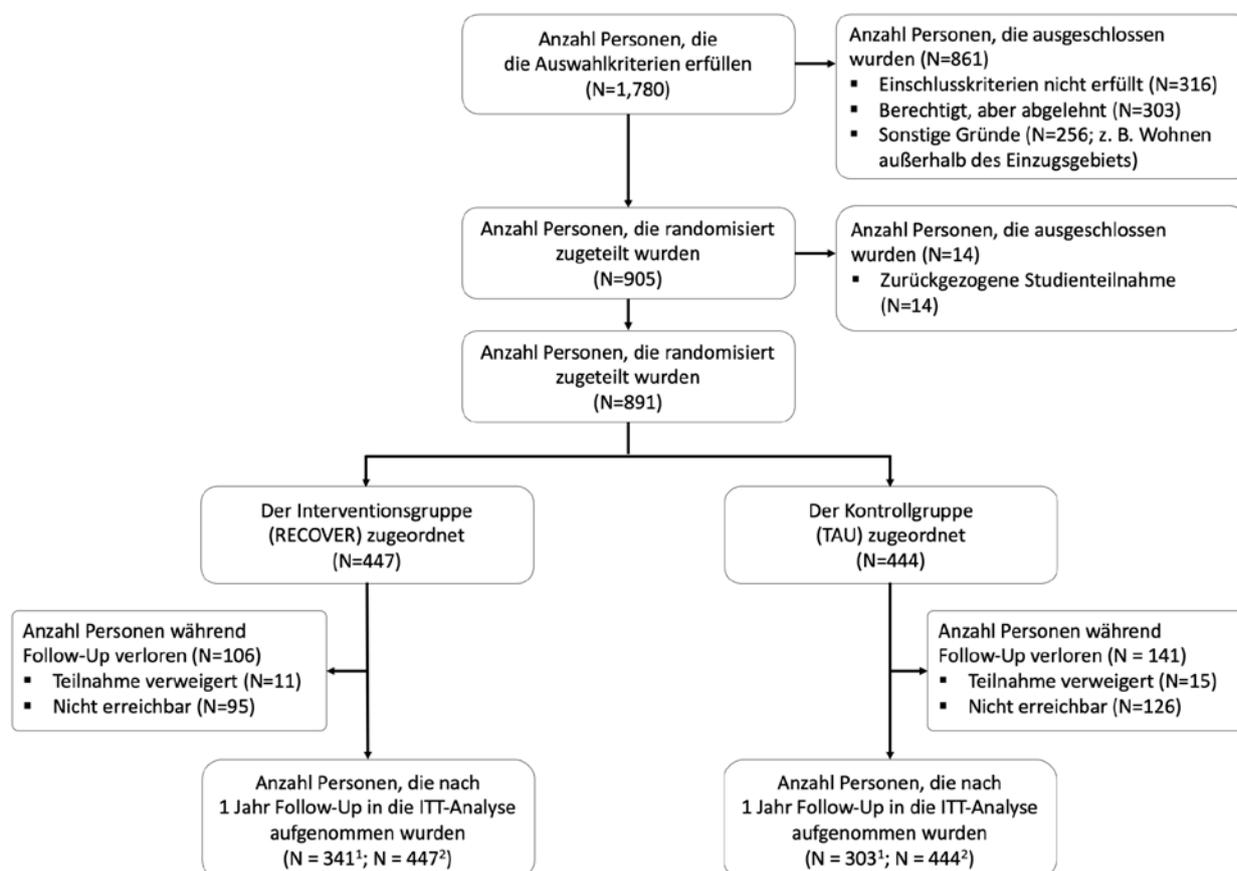


Abbildung 2. Flow-Chart RECOVER RCT; Anmerkung: ¹ Baseline imputierte ITT-Stichprobe (Datensatz II); ² vollständig imputierte Stichprobe (Datensatz I).

Zur Baseline-Erhebung nahmen insgesamt N = 905 Personen an der RECOVER Studie teil und wurden zufällig einer der beiden Gruppen (IG, KG) zugeordnet. Aufgrund von Widerruf der Studienteilnahme reduzierte sich die Größe der Baseline-Stichprobe auf N = 881. Für Details zur Stichprobe siehe Tabelle 1. Im Verlauf reduzierte sich die Anzahl der erhobenen Datensätze, da nicht alle Teilnehmenden zu den Verlaufserhebungen bereit waren. Zum sechs Monate-Verlaufstermin wurde die Teilnahme am Termin abgelehnt (n = 26), waren die Personen nicht zu erreichen (n = 124) oder hatten die Teilnahme an der Studie widerrufen (n = 39). Zum zwölf Monate-Verlaufstermin waren die Personen nicht zu erreichen (n = 64) oder hatten die Teilnahme an der Studie widerrufen (n = 8).

Tabelle 1. Beschreibung Stichprobe (Baseline)

Variable	IG (N = 447)	KG (N = 444)
	n (%) / M (SD)	n (%) / M (SD)
Alter	36.92 (13.67)	37.49 (14.56)

Geschlecht (weiblich)	268 (59.96%)	263 (59.23%)
Berufstätig	210 (46.98%)	206 (46.40%)
Familienstand (Single)	231 (51.68%)	245 (55.18%)
Schulabschluss (Abitur, Fachabitur)	227 (50.78%)	235 (52,95%)
Wohnung (eigener Haushalt, allein lebend)	169 (37.89%)	208 (46.95%)
Behandlungsstatus (teil-/stationär)	109 (24.38%)	111 (25%)
Hauptdiagnose:		
ICD-10 F2	40 (8.95%)	35 (7.88%)
ICD-10 F3	209 (46.76%)	205 (46.17%)
ICD-10 F4	104 (23.27%)	91 (20.50%)
ICD-10 F5	0 (0.00%)	4 (0.90%)
ICD-10 F6	92 (20.58%)	106 (23.87%)
Andere	2 (0.45%)	3 (0.68%)
Schweregrad:		
Stufe 1	30 (6.71%)	30 (6.76%)
Stufe 2	165 (36.91%)	160 (36.04%)
Stufe 3	159 (35.57%)	158 (35.59%)
Stufe 4	93 (20.81%)	96 (21.62%)
Anmerkungen		
IG= Interventionsgruppe; KG = Kontrollgruppe (treatment-as-usual); N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; % = Prozent.		

Aufgrund der Corona-Pandemie wurde die Datenerhebung bei Verlauf Erhebungen teilweise telefonisch durchgeführt. Dies ist in beiden Gruppe gleichverteilt (IG: 29 Erhebungen in Präsenz, 145 Erhebungen telefonisch; KG: 21 Erhebungen in Präsenz, 143 Erhebungen telefonisch). Schwerwiegende unerwünschte Ereignisse (SAE) wurden erfasst (Details siehe Tabelle 2). Es wurden bei IG und KG keine Zusammenhänge zwischen SAE und Teilnahme an der Studie beobachtet.

Tabelle 1. SAE RECOVER Projekt

Zeitpunkt SAE-Erfassung	IG (n = 389)	KG (n = 329)
Nach 6 Monaten (t6)		
Medizinisch signifikant und somatisch	n = 4	n = 3
Medizinisch signifikant und psychiatrisch	n= 6	n= 15
Medizinisch signifikant und Zwangseinweisung	n = 5	n = 4
Verfehlter Suizidversuch	n= 6	n= 5
Anderer Grund	n = 3	n = 3
Nach zwölf Monaten (t12)	IG (n = 373)	KG (n = 326)
Medizinisch signifikant und somatisch	n = 5	n = 3
Medizinisch signifikant und psychiatrisch	n= 5	n= 15
Medizinisch signifikant und Zwangseinweisung	n = 6	n = 12
lebensbedrohlich	n = 1	n = 2
Verfehlter Suizidversuch	n= 3	n= 7
Anderer Grund	n = 0	n = 1

3.2 Primärer Endpunkt 1: Versorgungskosten (gesundheitsökonomische Evaluation)

Tabelle 3 zeigt die unadjustierten direkten, indirekten und gesamten Kosten für IG und KG nach Messzeitpunkten jeweils für den zurückliegenden 6-Monatszeitraum.

Es zeigte sich, dass die direkten, die indirekten und die gesamten durchschnittlichen Kosten zu t0 in der IG jeweils höher waren als in der KG. Zu t12 nahmen die direkten Kosten gegenüber t0 ab, in der IG jedoch stärker als in der KG (IG -3182€ vs. KG -800€). Für die indirekten Kosten fand sich ein Rückgang zu t12 vor allem in der IG (-1329€), während die indirekten Kosten in der KG geringfügig zunahmen (+52€).

Betrachtet man die Summe aus direkten und indirekten Kosten über den gesamten 12-Monats-Follow-up-Zeitraum (t6 plus t12), so fanden sich in der IG niedrigere Kosten als in der KG (-4774€), obwohl die Kosten zu t0 in der IG höher waren als in der KG (+1756€).

Tabelle 3. Unadjustierte Kosten aus gesellschaftlicher Perspektive, jeweils im zurückliegenden 6-Monatszeitraum zu t0, t6 und t12 (in € und Preisen von 2019)

	IG (N=447)	KG (N=444)	Differenz IG-KG
	M (SF)	M (SF)	
Direkte Kosten			
t0	8068 (546)	7033 (532)	1035
t6	6138 (351)	8412 (474)	-2274
t12	4886 (275)	6233 (378)	-1347
Indirekte Kosten			
t0	4050 (391)	3329 (341)	721
t6	4282 (469)	4775 (465)	-493
t12	2721 (385)	3381 (413)	-660
Gesamtkosten			
t0	12118 (706)	10362 (669)	1756
t6	10420 (605)	13187 (680)	-2767
t12	7607 (479)	9614 (576)	-2007
Anmerkungen			
Direkte Kosten: stationäre Aufenthalte, ambulante Arztkontakte, ambulante Heilmittelerbringer, Medikamente, Pflege/Hilfe, informelle Pflege, Reha, Beratung & Betreuung, Berufliche Integration, Krisen-Interventionen, Webbasierte Angebote, Peerbegleitung. Indirekte Kosten: Produktivitätsverluste durch AU-Tage. SF = Standardfehler, M = Mittelwert, N = Anzahl Probanden.			

Tabelle 4 zeigt die adjustierten Kosten im 12-Monats-Follow-up-Zeitraum, differenziert nach Kostenarten für die Gesamtstichprobe und nach Stufen (s. 3.3.1 Subgruppenanalysen). In der Gesamtstichprobe waren die Gesamtkosten in der IG statistisch signifikant niedriger als in der KG (-5039 €; $p < 0.001$). In Bezug auf die einzelnen Kostenarten zeigten sich deutliche Unterschiede: Die Kosten waren in der IG bei „Heilmittelerbringern“ (-72 €, $p = 0.019$), „stationären Aufenthalten“ (-2601 €, $p < 0.001$), „teilstationären Aufenthalten“ (-1264 €, $p < 0.001$) und „ambulant psychiatrischen Pflegediensten“ (-222 €, $p = 0.006$) statistisch signifikant niedriger als in der KG. Andererseits waren die Kosten in der IG bei der Nutzung von ambulanten „Ärzten/Psychotherapeuten“ (+926 €, $p < 0.001$), „Eignungsabklärung“ (+16 €, $p = 0.001$), „Unterstützter Beschäftigung“ (+31 €, $p = 0.001$), „Krisentelefon“ (+3 €, $p < 0.001$), „Webbasierten Angeboten“ (+38 €, $p = 0.001$) und „Peerbegleitung“ (+64 €, $p < 0.001$) statistisch signifikant höher als in der KG.

Werden die stationären Kosten nach Allgemeinkrankenhaus und Psychiatrie unterschieden, so zeigte sich, dass die Kosteneinsparungen in der IG im stationären und teilstationären Bereich in der Psychiatrie generiert wurden, während die Kosten für Allgemeinkrankenhäuser in der IG und der KG nahezu identisch waren.

Tabelle 4. Adjustierte Kosten im 12-Monats-Follow-up-Zeitraum nach Kostenkategorien (in € nach Preisen 2019)

Kostenkategorie	Alle			Stufe 1			Stufe 2			Stufe 3			Stufe 4		
	IG	KG	Δ	IG	KG	Δ									
Ärzte/Psychotherapeuten	3198	2272	** *	1779	1786		2815	2254	**	3771	2341	** *	3278	2404	**
Allgemeinmedizin	130	146		96	137		120	151	*	144	163		134	116	
Psychologe/Psychotherapeut	1566	1251	**	689	1056		1509	1337		1825	1250	**	1478	1200	
Psychiater	212	225		100	112		126	156		301	274		231	310	
PIA	124	131		21	185		91	98		131	153		200	159	
Sonstige Arztgruppen	1173	514	** *	842	387	** *	988	500	** *	1369	505	** *	1245	608	** *
Heilmittelerbringer¹	231	303	*	154	321		151	226	*	225	278		410	467	
stationäre Aufenthalte	2876	5477	** *	2767	1580		1452	3524	**	3412	5336	*	4503	10458	**
Allgemein-KH	514	590		951	115		269	439		539	683		796	837	
Psychiatrie	2377	4867	** *	1552	1638		1177	3104	**	2870	4659	*	3806	9473	**
teilstationäre Aufenthalte	1061	2325	** *	229	705		509	1945	** *	1246	2538	**	1876	3345	*
Allgemein-KH	231	250		84	70		35	188	**	200	110		652	673	
Psychiatrie	844	2063	** *	161	677		474	1755	** *	1050	2425	**	1290	2601	*
Rehabilitationen	119	292	**	45	139		111	324	*	153	362	*	98	174	
Pflege/Hilfen	800	1078	*	289	323		316	671	**	1018	923		1460	2258	
bezahlte Haushaltshilfe	141	106		201	87		101	112		201	113	*	102	81	
gesetzlicher Betreuer	88	80		19	25		44	25		57	87		252	174	
Sozialarbeiter/Begleitperson	396	451		31	43		83	232	*	500	476		843	968	
amb. psychiatr. Pflegedienst	188	410	**	34	360		89	311	**	261	254		279	987	*
Informelle Pflege	688	940		688	294		376	771	*	851	680		1210	1691	

Medikamente	786	1034		113	314		684	970		803	1187		1170	1133	
Beratung & Betreuung	315	390		76	206		143	313	**	403	352		597	575	
Beratungsstelle psych. Probleme/Sucht	92	131		26	15		77	132		94	132		115	204	
Sozialpsychiatrischer Dienst	134	179		15	171		27	116	**	178	147		348	288	
Tagesstätte/Kontaktstätten	89	82		43	57		32	95	*	139	71		134	122	
Berufliche Integration	495	439		423	62	**	311	378		590	349	*	731	730	
Integrationsfachdienst	45	26		0 ²	0 ²		54	29		52	5	**	30	57	
Eignungsabklärung	31	15	** *	39	0		30	13	*	44	14	** *	13	26	
Geschützte Werkstatt - Arbeitsbereich	206	186		103	16		59	121		202	96		632	416	
Unterstützte Beschäftigung, Berufsbegleitung	14	23		5	8		9	14		23	26		13	38	
Berufliche Weiterbildung	106	113		384	1		50	119		145	144		83	109	
Unterstützte Beschäftigung - Jobcoaching	100	70	**	61	42		109	91		124	63	** *	55	57	
Kriseninterventionen	176	131		117	143		67	133		300	138		206	97	*
Krisen-Telefon	4	2	** *	3	1		2	2		6	2	** *	7	2	*
Hausbesuche	172	129		113	141		65	131		294	136		200	96	*
Webbasierte Angebote	67	29	** *	72	66		128	29	** *	34	24		11	27	
Peerbegleitung	98	34	** *	69	19	*	105	35	**	114	35	**	58	37	
Indirekte Kosten durch AU-Tage	6867	8321		6380	1316 4		8996	1084 0		5498	7356		5340	4952	
Kosten Gesamt	1791 1	2295 0	** *	1344 3	1666 6		1635 6	2216 2	**	1846 2	2189 3		2067 7	28621	**

Anmerkungen

Δ = Signifikanz der Kostendifferenz zwischen IG (RECOVER) und KG (TAU), der besseren Übersicht halber wurden die exakten p-Werte nicht dargestellt, sondern nur signifikante Ergebnisse: * $p < 0.05$, ** $p < 0.01$, und *** $p < 0.001$. Adjustierung der Kosten für: Alter, Geschlecht, jeweilige Baseline-Kosten, Somatische Erkrankungen, Krankheitsschwere, Berechnung mit generalisierten linearen Modellen. Die Gesamtkosten, Kosten für berufliche Integration und Kosten durch AU-Tage wurden zusätzlich für Erwerbstätigkeit adjustiert.

¹ Heilmittelerbringer: Ergotherapie/Arbeitstherapie, Kunsttherapie, Massage/Lymphdrainage, Heilpraktiker, Krankengymnastik, Sport/Bewegungstherapie.

² Keine Beobachtungen in Stufe 1.

3.2.1 Primärer Endpunkt 1: Subgruppenanalyse

Betrachtet man stufenspezifische Kosten (Tabelle 4) so zeigt sich, dass die Kosten in der IG in Stufe 1 am niedrigsten waren (13443 €) und dann bis Stufe 4 (20677 €) kontinuierlich anstiegen. In der KG waren die Kosten in Stufe 1 ebenfalls am niedrigsten (16666 €) und in Stufe 4 am höchsten (28621 €), während die Kosten in Stufe 3 (21893 €) gegenüber Stufe 2 (22162 €) leicht zurückgingen.

Innerhalb von Stufe 1 fanden sich nur für „sonstige Arztgruppen“ (+455 €, $p=0.001$), „berufliche Integration“ (+361 €, $p=0.007$) und „Peerbegleitung“ (+50 €, $p=0.015$) statistisch signifikante Unterschiede zwischen IG und KG (in allen drei Fällen höhere Kosten in der IG). In den Stufen 2 bis 4 fanden sich deutlich öfter statistisch signifikante Unterschiede. So war die Nutzung ambulanter „Ärzte/Psychotherapeuten“ in allen drei Stufen in der IG statistisch signifikant teurer als in der KG (Stufe 2: +561 €, $p=0.002$; Stufe 3: +1430 €, $p<0.001$; Stufe 4: +873 €, $p=0.001$), während sowohl die Nutzung „stationärer Aufenthalte“ (Stufe 2: -2072 €, $p=0.001$; Stufe 3: -1924 €, $p=0.04$; Stufe 4: -5955 €, $p=0.002$), als auch die Nutzung „teilstationärer Aufenthalte“ (Stufe 2: -1436 €, $p<0.001$; Stufe 3: -1291 €, $p=0.007$; Stufe 4: -1468 €, $p=0.033$) in den Stufen 2 bis 4 in der IG mit statistisch signifikant niedrigeren Kosten assoziiert war. Wie in der Gesamtstichprobe zeigte sich, dass die Kosteneinsparungen im stationären und teilstationären Bereich in der Psychiatrie generiert wurden.

3.2.2 Primärer Endpunkt 1: Sensitivitätsanalysen

In der Sensitivitätsanalyse auf Grundlage der Baseline imputierten ITT-Stichprobe (Datensatz II) waren die Ergebnisse insgesamt ähnlich wie in der Basisanalyse (mit der vollständig imputierten ITT-Stichprobe (Datensatz I)).

Tabelle A2 im Anhang zeigt die unadjustierten direkten, indirekten und gesamten Kosten nach Messzeitpunkt. Tabelle A3 im Anhang zeigt die unadjustierten Kosten über den 12-Monats-Follow-up-Zeitraum, nach Kostenkategorien.

Im Gegensatz zur Basisanalyse waren die Kosten in der IG nur bei „stationären Aufenthalten“ und „teilstationären Aufenthalten“, sowie zusätzlich in geringem Umfang bei „Tagesstätten/Kontaktstätten“ niedriger als in der KG. Auch in der Sensitivitätsanalyse zeigte sich, dass in der Gesamtstichprobe die Kosten in der IG bei der Nutzung von ambulanten „Ärzten/Psychotherapeuten“, „Eignungsabklärung“, „Unterstützter Beschäftigung“, „Krisentelefon“, „Webbasierten Angeboten“ und „Peerbegleitung“ statistisch signifikant höher als in der KG waren.

Betrachtet man stufenspezifische Kosten so zeigte sich ebenfalls, dass die Kosten in der IG in Stufe 1 am niedrigsten waren, gefolgt von Stufe 2 und Stufe 3, während sie in Stufe 4 wiederum gegenüber Stufe 3 leicht zurückgingen. In der KG hingegen waren die Kosten in Stufe 1 und Stufe 2 nahezu identisch und stiegen erst in den Stufen 3 und 4 kontinuierlich an.

Innerhalb von Stufe 1 fand sich nur für „sonstige Arztgruppen“ ein statistisch signifikanter Unterschied zwischen IG und KG. Die Stichprobengröße war in der Baseline imputierten ITT-Stichprobe (Datensatz II) in Stufe 1 allerdings so niedrig, dass bei neun Vergleichen keine Inanspruchnahme von Leistungen vorlag. Insgesamt fanden sich in der Baseline imputierten ITT-Stichprobe (Datensatz II) in den stufenspezifischen Vergleichen deutlich weniger statistisch signifikante Unterschiede zwischen IG und KG, am meisten noch in Stufe 2.

3.3 Primärer Endpunkt 2: Psycho-Soziales-Funktionsniveau (Patientenrelevanter Endpunkt; Effektevaluation)

Für den primären Endpunkt Psycho-Soziales-Funktionsniveau wurde die Baseline imputierte ITT-Stichprobe (Datensatz II) verwendet. Es zeigte sich eine höhere Differenz und damit Verbesserung zwischen t12 und t0 in der IG (RECOVER) verglichen zur KG (TAU) (siehe Tabelle 5). Im Mittel beträgt dieser Unterschied zwischen den Gruppen in etwa -0.3 Punkte des Composite-Score (Spalte adjustierte Differenz in Tabelle 5).

Tabelle 5. Ergebnisse patientenrelevanter Endpunkt: Psycho-Soziales-Funktionsniveau (t12-t0)

IG			KG			KG vs. IG		
N	beo. M (SD)	adj. M (95% KI)	N	beo. M (SD)	adj. M (95% KI)	Adj. Diff. (95% KI)	p-Wert	Inter. p-Wert (Gr.,Sg)
34 1	0.06 (2.49)	0.22 (-0.04, 0.49)	30 3	-0.16 (2.24)	-0.09 (-0.36, 0.19)	-0.31 (-0.64, 0.03)	0.0751	0.4769
Anmerkungen								
n = Anzahl Proband:innen, beo. M = beobachteter Mittelwert; SD = Standardabweichung, adj. M = adjustierter Mittelwert, 95%KI = 95% Konfidenzintervall. Adj. Diff. = adjustierte Differenz, Inter. P-Wert = Interaktions p-Wert, Gr = Gruppe, Sg = Schweregrad). Adjustierungsvariablen: Schweregrad zu Baseline, Psycho-Soziales-Funktionsniveau zu Baseline. Anzahl ersetzter Beobachtungen zu t0: n = 129								

3.3.1 Primärer Endpunkt 2: Subgruppenanalysen

Abbildung 3 zeigt die Ergebnisse der Subgruppenanalyse für das primäre Outcome Psycho-Soziales-Funktionsniveau. Hierbei zeigte sich für fast alle untersuchten Subgruppenvergleiche kein Interaktions-p-Wert <0.05 mit dem Faktor IG vs. KG, d.h., dass die Behandlungsergebnisse nicht unterschiedlich für die einzelnen Subgruppen ausfielen. Allein für die Art der Datenerhebung (alle Verlaufserhebungen, d.h. t6 und t12 in Präsenz vs. teilweise oder ganz telefonisch) scheint ein unterschiedlicher Behandlungseffekt vorzuliegen: Der Vorteil der IG gegenüber der KG war in Verlaufserhebungen in Präsenz stärker ausgeprägt. Eine detaillierte Ergebnisdarstellung findet sich in Anhang (Tabelle A4).

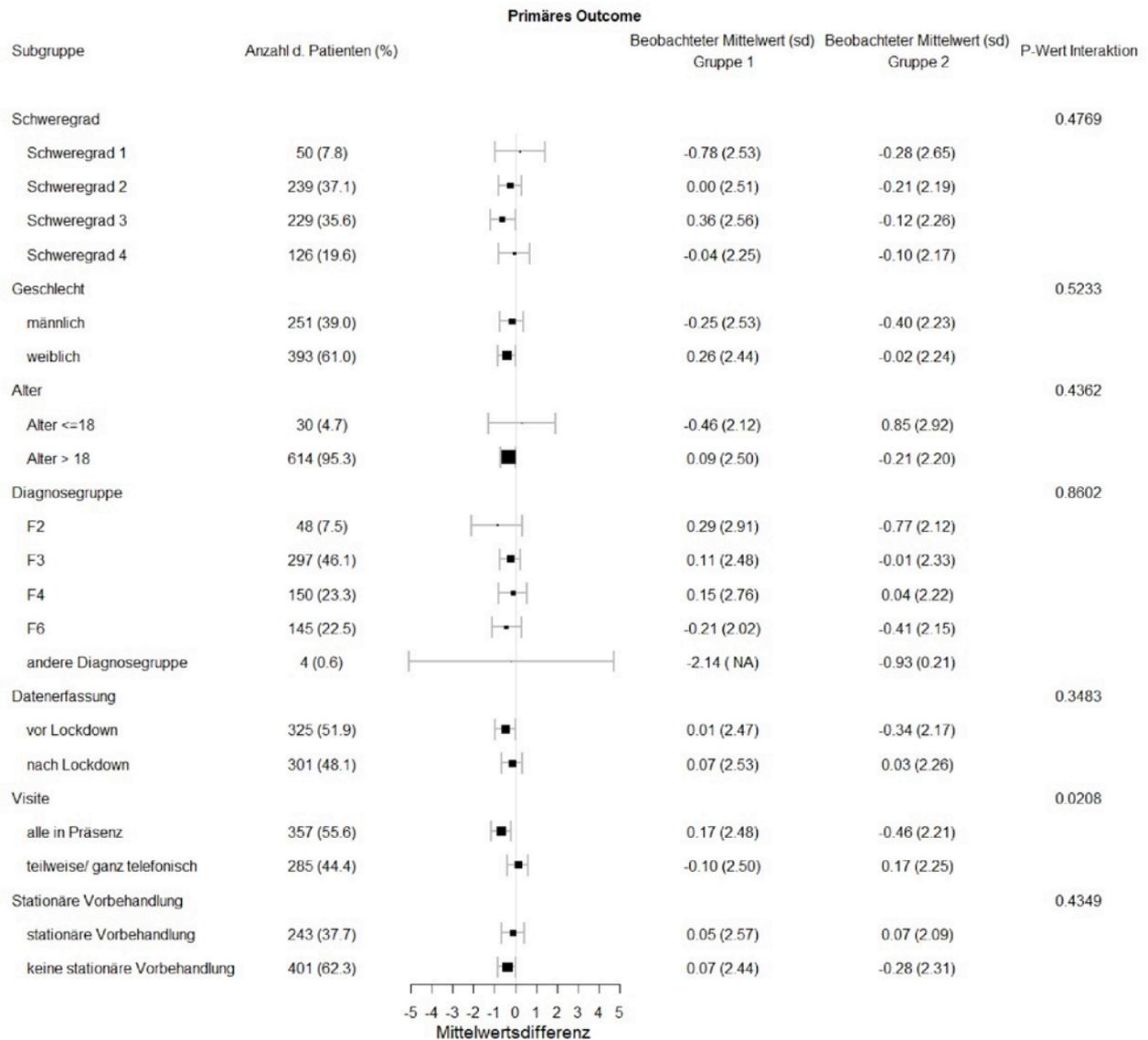


Abbildung 3. Subgruppenanalysen für das primäre Outcome der Effektevaluation (Psycho-Soziales-Funktionsniveau). Für Psycho-Soziales-Funktionsniveau zu Baseline adjustierte Analyse.

3.3.2 Primärer Endpunkt 2: Sensitivitätsanalysen

Die Ergebnisse der Sensitivitätsanalysen befinden sich im Anhang (Tabelle A5). Hierbei werden die Ergebnisse aus Tabelle 5 (Analyse basierend auf der Baseline imputierten ITT-Stichprobe, Datensatz II) überwiegend bestätigt.

3.4 Primärer Endpunkt 3: Kosteneffektivität (gesundheitsökonomische Evaluation)

Tabelle 6 zeigt die adjustierten Schätzer für QALYs über den 12-Monats-Follow-up-Zeitraum insgesamt und nach Stufen. Nach Adjustierung fanden sich in der IG 0,01 QALY mehr als in der KG, wobei die Unterschiede zwischen den Gruppen nicht statistisch signifikant waren.

Tabelle 6. Adjustierte Schätzer für QALYs über den 12-Monatszeitraum

	M IG	M KG	Differenz	p-Wert
Gesamt	0,72	0,71	0,01	0,141
Stufe 1	0,81	0,85	-0,04	0,092
Stufe 2	0,78	0,77	0,01	0,332
Stufe 3	0,68	0,66	0,02	0,241
Stufe 4	0,69	0,66	0,03	0,247
Anmerkungen				
M = Mittelwert. Adjustierung der Effekte für: Alter, Geschlecht, Baseline-Utilities, Somatische Erkrankungen, Krankheitsschwere, Berechnung mit OLS-Regression.				

Tabelle 7 zeigt die unadjustierten Punktschätzer für die Kosteneffektivitätsanalyse. Die IG war aus beiden Perspektiven (Gesellschaft und GKV) mit signifikant weniger Kosten und mit mehr QALYs (+0,02, nicht signifikant) verbunden, woraus sich als Punktschätzer für die ICER jeweils ein negativer Wert (d.h. Dominanz der IG gegenüber der KG) errechnet. Aus einer gesellschaftlichen Perspektive waren die Gesamtkosten in der IG im Beobachtungszeitraum um 4774€ geringer als in der KG (t6 + t12, siehe auch Tabelle 3), was in einer ICER von -238700€/QALY resultierte (Dominanz). Aus einer GKV-Perspektive waren die Gesamtkosten in beiden Gruppen deutlich niedriger, als bei einer gesellschaftlichen Perspektive, aber im Gruppenvergleich waren die Kosten auch hier in der IG geringer als in der KG (-3176€), entsprechend einer ICER von -158800€/QALY (Dominanz).

Die Kosteneffektivitätsakzeptanzkurven sind in Abbildung 4 dargestellt. Die Kurven zeigen für verschiedene Werte einer angenommenen Zahlungsbereitschaft eines Entscheidungsträgers für ein durch die IG im Vergleich zur KG zusätzlich gewonnenes QALY die Wahrscheinlichkeit dafür, dass die Intervention kosteneffektiv ist. Die IG war für die Basisanalyse für die gesellschaftliche Perspektive und die GKV-Perspektive für alle untersuchten Zahlungsbereitschaften von 0€ pro QALY bis 250000€ pro QALY mit einer Wahrscheinlichkeit von mehr als 95% kosteneffektiv.

Tabelle 7. Unadjustierte ICERs

	MW IG	MW KG	Differenz	ICER (€/QALY)
Gesellschaftliche Perspektive				
Kosten	18027 €	22801 €	-4774 €	Dominanz
QALY	0,73	0,71	0,02	
GKV-Perspektive				
Kosten	8901 €	12077 €	-3176 €	Dominanz
QALY	0,73	0,71	0,02	
Anmerkungen Kosten in € und Preisen von 2019				

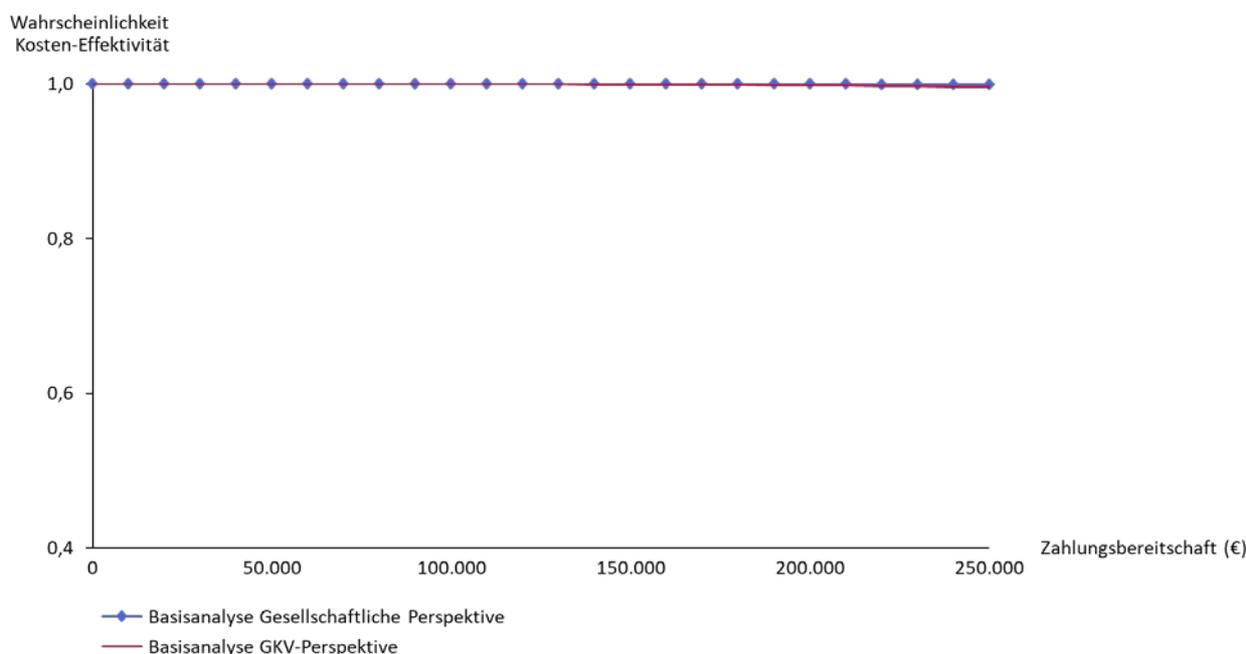


Abbildung 2. Kosteneffektivitätsakzeptanzkurven für die Basisanalysen

3.4.1 Primärer Endpunkt 3: Subgruppenanalyse

Auch in den analysierten Subgruppen fanden sich ausschließlich statistisch nicht-signifikante Unterschiede zwischen IG und KG in Bezug auf die Effekte (QALYs).

Abbildung 5 zeigt die Kosteneffektivitätsakzeptanzkurven für die Subgruppenanalysen. Die IG war für alle Subgruppenanalysen mit Ausnahme von Subgruppenanalyse 1 („nur Stufe 1“) für alle untersuchten Zahlungsbereitschaften von 0€ pro QALY bis 250000€ pro QALY mit einer Wahrscheinlichkeit von mehr als 95% kosteneffektiv. Grund für die Abweichung bei Subgruppenanalyse 1 war, dass der Punktschätzer für QALYs in Stufe 1 (und nur in Stufe 1) in der IG niedriger als in der KG war.

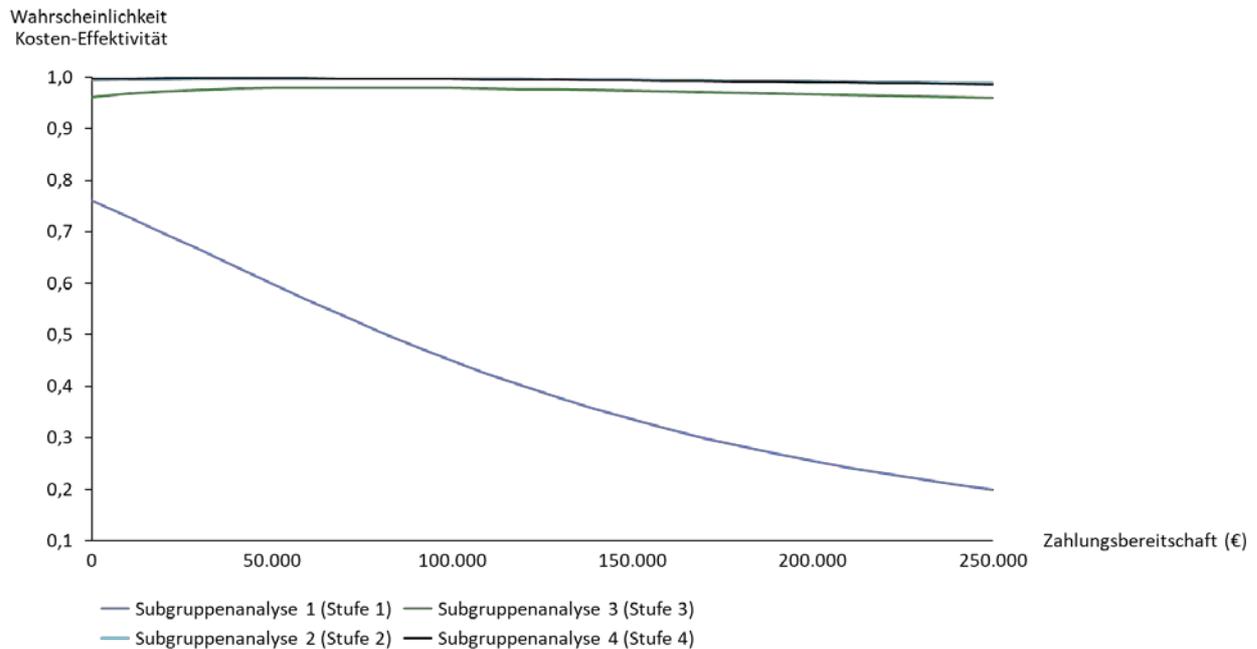


Abbildung 5. Kosteneffektivitätsakzeptanzkurven für die Subgruppen (Stufe 1-4)

3.4.2 Primärer Endpunkt 3: Sensitivitätsanalysen

Die erste Sensitivitätsanalyse, bei der zusätzliche Kostenkategorien in der Analyse der gesellschaftlichen Perspektive berücksichtigt wurden, zeigte ebenfalls keine Veränderung in Bezug auf die Kosten-Effektivität, s. Anhang Abbildung A1.

Die Ergebnisse der Sensitivitätsanalyse 2 mit der Baseline imputierten ITT-Stichprobe (Datensatz II) finden sich im Anhang in den Tabellen A6, A7 und in Abbildung A2. Während in der Basisanalyse und der Sensitivitätsanalyse mit der Baseline imputierten ITT-Stichprobe (Datensatz II) in der IG jeweils 0,02 QALYs (unadjustiert) mehr als in der KG gefunden wurden, zeigten sich Unterschiede zwischen den Datensätzen bei den Kosten, die auch in unterschiedlichen inkrementellen Kosten-Effektivitäts-Relationen (ICER) resultierten. So fand sich für die IG in der vollständig imputierten ITT-Stichprobe (Datensatz I) eine unadjustierte Kostenersparnis von -4774€ (ICER: -238700€/QALY), während sich in Datensatz II eine unadjustierte Kostenersparnis von -3303€ (ICER: -165150€/QALY) fand.

In der Baseline imputierten ITT-Stichprobe (Datensatz II) war die IG in der Basisanalyse (Gesellschaftliche Perspektive) und im Szenario „Sensitivitätsanalyse I“ über alle untersuchten Zahlungsbereitschaften hinweg mit einer Wahrscheinlichkeit von mehr als 95% kosteneffektiv. Aus der GKV-Perspektive war die IG bis zu einer Zahlungsbereitschaft von maximal 160000 € pro QALY und für Stufe 3 bis zu einer Zahlungsbereitschaft von maximal 20000 € pro QALY mit einer Wahrscheinlichkeit von mehr als 95% kosteneffektiv. In den verbleibenden Stufen 1, 2 und 4 hingegen war die IG für keine der untersuchten Zahlungsbereitschaften mit einer Wahrscheinlichkeit von mehr als 95% kosteneffektiv. Eine Betrachtung nach Stufen in der Baseline imputierten ITT-Stichprobe (Datensatz II) zeigt niedrigere QALYs in der IG in Stufe 1 und Stufe 4 (jeweils -0,02 QALYs), während die IG in Stufe 2 und Stufe 3 mit mehr QALYs als in der KG assoziiert war (jeweils +0,02 QALYs).

3.5 Sekundäre Hypothesen und Endpunkte

Die Tabellen 8-10 zeigen die Ergebnisse für die Analyse der sekundären Endpunkte (kontinuierliche Daten, binäre Daten, Zähldaten) in der Baseline imputierten ITT-Stichprobe (Datensatz II).

Zunächst zu den **kontinuierlichen Datenpunkten**. Wie in Tabelle 8 ersichtlich, zeigte sich in den Einzelkomponenten des kombinierten primären Endpunkts teils eine höhere Verbesserung im Vergleich zur KG. So zeigt die IG im Modul A des HEALTH-49 einen Vorteil gegenüber der KG. In der Psychischen Summenskala des SF-12 und im GAF waren die Verbesserungen von t0 zu t12 in beiden Gruppen vergleichbar. Die Veränderungen der Symptomschwere erhoben über den CGI-S zeigten keinen Unterschied zwischen den Gruppen. Auf das Maß der diagnosespezifischen Symptombelastung bezogen, war die IG der KG allerdings überlegen. Weiterhin zeigte sich in den zusätzlich zum SF-12 erhobenen weiteren Skalen zur Lebensqualität eine höhere Verbesserung zu Gunsten der IG im Re-QoL-20, wohingegen für die Indexwerte und die visuelle Analogskala des EQ-5D-5L keine Gruppenunterschiede der Differenzwerte gefunden wurden. Für keine dieser kontinuierlichen Endpunkte ergaben sich unterschiedliche Differenz je nach Zugehörigkeit zu einer der vier Schweregradgruppen (siehe Tabelle 8 Spalte Interaktion p-Wert; alle ≥ 0.05).

Für vier der acht untersuchten **binären Zielgrößen** zeigte sich eine Interaktion des zu t0 definierten Schweregrads (Stufe 1-4) mit der Behandlungsgruppe bzgl. des Outcomes. Somit unterschied sich bei diesen Endpunkten der Behandlungseffekt zwischen den Baseline-Schweregradgruppen (siehe Tabelle 9, Spalte Interaktion). Dies sind zum einen die beiden Endpunkte symptomatische und funktionelle Remission zu t12, bei denen sich Odds Ratios größer Eins, d.h. ein Vorteil der Kontrollgruppe, für die Schweregradgruppen 1 und 2 zeigt und Odds Ratios kleiner 1, d.h. Vorteile von RECOVER, für die beiden Gruppen mit einem hohen Schweregrad (3 und 4). Zum anderen zeigte sich für den Endpunkt „Ambulante psychotherapeutische Interventionen schwere psychische Störungen“ eine höhere Rate in der RECOVER-Gruppe gegenüber der Kontrollgruppe in den Schweregradgruppen 2-4, wohingegen in der Schweregruppe 1 der Anteil in der KG höher war. In der Variable Peer Support zeigte sich ein Unterschied zwischen den Gruppen (15.28% bei IG vs. 1.56% bei KG). Für die anderen vier binären Endpunkte zeigten sich keine Gruppenunterschiede.

Tabelle 8. Ergebnisse sekundäre Endpunkte (kontinuierliche Daten)

Zielgröße	IG			KG			KG vs. IG			Anz. ersetzter Beob. zu t0
	N	beo. M (SD)	adj. M (95% KI)	n	beo. M (SD)	adj. M (95% KI)	Adj. Diff. (95% KI)	p-Wert	Inter. p-Wert (Gr.,Sg)	
Modul A HEALTH 49 (inverted) (t12-t0)	360	0.43 (0.69)	0.47 (0.40, 0.54)	322	0.33 (0.64)	0.34 (0.27, 0.42)	-0.13 (-0.22, -0.03)	0.0077	0.3302	26
GAF (t12-t0)	373	9.96 (12.23)	10.49 (9.02, 11.96)	327	9.63 (12.62)	10.15 (8.64, 11.67)	-0.34 (-2.16, 1.49)	0.7181	0.5821	9
PSK12 (t12-t0)	341	10.05 (13.26)	11.08 (9.79, 12.38)	305	8.41 (12.23)	9.47 (8.13, 10.80)	-1.62 (-3.29, 0.06)	0.0586	0.7328	127
CGI-S (t12-t0)	373	-0.60 (1.11)	-0.61 (-0.73,-0.48)	327	-0.51 (1.08)	-0.52 (-0.65, -0.39)	0.09 (-0.07, 0.25)	0.2780	0.6427	9
Re-QOL-20 (t12-t0)	360	11.93 (17.08)	12.85 (11.17, 14.53)	323	9.27 (14.91)	9.96 (8.23, 11.69)	-2.89 (-5.04, -0.74)	0.0086	0.8864	23
EQ-5D Item 6 (t12-t0)	360	10.86 (22.08)	11.06 (8.86, 13.26)	322	9.96 (21.80)	10.01 (7.74, 12.28)	-1.05 (-3.88, 1.77)	0.4640	0.8620	32
EQ-5D Index	353	0.09 (0.24)	0.10	313	0.08 (0.25)	0.08	-0.02	0.1967	0.8140	53

(t12-t0)			(0.07, 0.13)			(0.05, 0.10)	(-0.05, 0.01)			
Diagnose-spezifische Symptombelastung	297	-0.20 (1.07)	-0.27 (-0.38, -0.16)	267	-0.05 (0.96)	-0.07 (-0.18, 0.04)	0.20 (0.06, 0.34)	0.0057	0.4612	102
<p>Anmerkungen</p> <p>Anmerkungen: n = Anzahl Probanden, beo. M = beobachteter Mittelwert; SD = Standardabweichung, adj. M = adjustierter Mittelwert, 95%KI = 95% Konfidenzintervall. Adj. Diff. = adjustierte Differenz, Inter. P-Wert = Interaktions p-Wert, Gr = Gruppe (IG vs. KG), Sg = Schweregradstufe (1-4), PSK12 = Psychische Summenskala des SF-12. Adjustierungsvariablen: Schweregrad zu Baseline, jeweilige Zielgröße zu Baseline; Anzahl ersetzter Beobachtungen.</p>										

Tabelle 9. Ergebnisse sekundäre Endpunkte (binäre Daten)

Zielgröße	Sg	IG	KG	KG vs. IG			Anz. ersetzter Beob. zu t0
		Zielgröße = Ja (%)	Zielgröße = Ja (%)	Adj. Odds Ratio (95% KI)	p-Wert	Interaktion p-Wert (Gr, Sg)	
Symptomatische Remission t12						<0.0001	9
	1	19/25 (76.00)	21/25 (84.00)	1.46 (0.35, 6.04)	0.6029		
	2	76/147 (51.70)	63/112 (56.25)	1.24 (0.74, 2.05)	0.4145		
	3	27/128 (21.09)	22/120 (18.33)	0.85 (0.45, 1.59)	0.6117		
	4	4/73 (5.48)	3/70 (4.29)	0.77 (0.17, 3.58)	0.7414		
Funktionelle Remission t12						0.0172	9
	1	19/25 (76.00)	23/25 (92.00)	3.37 (0.60, 18.80)	0.1666		
	2	90/147 (61.22)	74/112 (66.07)	1.28 (0.76, 2.16)	0.3454		
	3	44/128 (34.38)	37/120 (30.83)	0.85 (0.50, 1.45)	0.5524		
	4	12/73 (16.44)	9/70 (12.86)	0.75 (0.29, 1.91)	0.5462		
Ambulante psychoth. Interventionen schwere psychische Störungen t12 ¹	4	40/71 (56.34)	32/68 (47.06)	0.69 (0.35, 1.34)	0.2745	0.0040	

Peer-Unterstützung t12		55/360 (15.28)	5/321 (1.56)	0.09 (0.03, 0.22)	<0.001	0.0455	
Psychotherapeutische Gruppeninterventionen t12		21/178 (11.80)	23/136 (16.91)	1.47 (0.77, 2.80)	0.2469	0.4862	
Web-basierte Therapie t12		16/358 (4.47)	10/321 (3.12)	0.72 (0.32, 1.63)	0.4319	0.7013	
Psychotherapeutische Kurzzeitinterventionen t12		83/360 (23.06)	62/322 (19.25)	0.81 (0.56, 1.18)	0.2814	0.3781	
Behandlungsabbrüche t12		42/359 (11.70)	35/321 (10.90)	0.89 (0.55, 1.44)	0.6384	0.6989	
Anmerkungen Ergebnisse der logistischen Regression für binäre Endpunkte; CI= Konfidenzintervall; Sg = Schweregrad (Stufe 1-4); Adj. Odds-Ratio = Adjustiertes Odds-Ration; Gr = Gruppe (RECOVER, TAU). ¹ Es wurden alle vier Stufen im Modell berücksichtigt, aber entsprechend des Outcomes nur Stufe 4 hier dargestellt. Adjustierungsvariablen für alle: Schweregrad zu Baseline; zusätzliche Adjustierungsvariable für Symptomatische und Funktionelle Remission: Werte diese Variablen zu Baseline.							

Für die vier als **Zähl**daten erfassten Endpunkte zeigt Tabelle 10 die Ergebnisse der Analysen, es ergaben sich keine Gruppenunterschiede.

Tabelle 10. Ergebnisse sekundäre Endpunkte (negative Binomialregression für Zähldaten)

Zielgröße	IG			KG		KG vs. IG			Anz. ersetzter Beob. zu t0
	Sg	N	beo. M (SD)	N	beo. M (SD)	Adj. Ratio (95% KI)	p-Wert	Interaktion p-Wert (Gr-Sg)	
Arbeitsunfähigkeit t12		366	10.99 (32.77)	321	12.21 (36.12)	1.33 (0.82, 2.15)	0.2473	0.6413	16
Rate der psychiatrischen Krankenhausaufenthalte t12		372	0.10 (0.50)	326	0.17 (0.58)	1.56 (0.90, 2.68)	0.1104	0.2995	
Dauer psychiatrischer Krankenhausaufenthalte t12		372	0.45 (2.26)	326	0.53 (3.44)	0.60 (0.24, 1.50)	0.2743	0.1672	10
Spezifische psychiatrische Interventionen bei schweren psychischen Störungen t12 ¹	4	71	0.92 (2.84)	68	0.07 (0.26)	0.08 (0.02, 0.30)	<0.001	0.8700	

Anmerkungen
 Ergebnisse der negativen binomial Regression für Zähl

daten; n= Anzahl in Analyse eingeschlossenen Beobachtungen pro Gruppe; sd=Standardabweichung; CI= Konfidenzintervall, Sg = Schweregrad (Stufe 1-4); Gr = Gruppe (RECOVER,TAU); beo M = beobachteter Mittelwert; adj. Odds-Ratio = adjustiertes Odds-Ratio. ¹ Es wurden alle vier Stufen im Modell berücksichtigt, aber entsprechend des Outcomes nur Stufe 4 hier dargestellt. Adjustierungsvariablen für alle: Schweregrad zu Baseline; zusätzliche Adjustierungsvariable für Arbeitsunfähigkeit und Dauer psychiatrischer Krankenhausaufenthalte; Werte diese Variablen zu Baseline.

3.6 Zusätzliche Analysen

Die Ergebnisse der gemischten Modelle, in denen beide Follow-up-Zeitpunkte gleichzeitig modelliert wurden, befinden sich im Anhang Tabellen A8-A11. Dabei zeigte sich, dass die Ergebnisse zu t6 vergleichbar zu t12 sind und damit die Zeitpunkte nicht unterschieden werden müssen.

3.7 Qualitätssicherungsstudie (Transfer nach Itzehoe)

Im Klinikum Itzehoe wurde eine Qualitätssicherungsstudie zum Transfer des RECOVER Modell durchgeführt. Die dort behandelte Stichprobe ist in Tabelle 11 genauer beschrieben und wird mit der RECOVER Stichprobe des RCT verglichen.

Tabelle 11. Stichprobe Itzehoe (Baseline)

Variable	RECOVER Itzehoe (N = 83)	RECOVER TAU (N = 447)
	n (%) / M (SD)	n (%) / M (SD)
Alter	38.66 (15.74)	36.92 (13.67)
Geschlecht (weiblich)	46 (55.4%)	268 (59.96%)
Berufstätig	56 (67.5%)	210 (46.98%)
Familienstand (Single)	34 (41%)	231 (51.68%)
Schulabschluss (Abitur, Fachabitur)	21 (24.3%)	227 (50.78%)
Wohnung (eigener Haushalt, allein lebend)	15 (18.1%)	169 (37.89%)
Behandlungsstatus (teil-/stationär)	4 (4.8%)	109 (24.38%)
Hauptdiagnose:		
ICD-10 F2	1 (1.2%)	40 (8.95%)
ICD-10 F3	42 (50.6%)	209 (46.76%)
ICD-10 F4	28 (33.7%)	104 (23.27%)
ICD-10 F5	2 (2.4%)	0 (0.00%)
ICD-10 F6	8 (9.6%)	92 (20.58%)
Andere	2 (2.4%)	2 (0.45%)
Schweregrad:		
Stufe 1	9 (10.8%)	30 (6.71%)
Stufe 2	28 (33.7%)	165 (36.91%)
Stufe 3	32 (38.6%)	159 (35.57%)
Stufe 4	14 (16.9%)	93 (20.81%)
Anmerkungen		
N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; % = Prozent.		

Es zeigen sich Ähnlichkeiten zwischen den Gruppen bzgl. der soziodemografischen Zusammensetzung, wobei die Patient:innen in Itzehoe deskriptiv einen geringeren Bildungsstand aufwiesen und seltener zum Zeitpunkt der Aufnahme in teil- oder stationärer Behandlung waren. In beiden Gruppen stellte ICD-10 Diagnosen der Kategorien F3, F4 und F6 die häufigste Hautdiagnose dar. Auch die Verteilung der Schweregrade (1-4) ist ähnlich, sodass von einer vergleichbaren Zusammensetzung der Patient:innen im städtischen und ländlichen Raum ausgegangen wird. Diese Einschätzung beruht allerdings allein auf den deskriptiven Werten.

In Itzehoe wurden Patient:innen im Rahmen der Qualitätssicherungsstudie nach 12 Monaten ebenfalls erneut mittels Interview und Fragebögen untersucht. Die Ergebnisse für Selbsteinschätzungen zur Lebensqualität (PSK12 des SF-12) sowie Fremdeinschätzungen zum Schweregrad der Erkrankung (CGI-S) und globaler Funktionsfähigkeit (GAF) sind in Tabelle 12 dargestellt.

Es zeigte sich deskriptiv für die RECOVER Patient:innen in Itzehoe ein Anstieg der selbsteingeschätzten Lebensqualität (Anstieg in der PSK12 um 10.86 Punkte), der globalen Funktionsfähigkeit (Anstieg im GAF um 15.15 Punkte) und eine Reduktion im Schweregrad der Erkrankung (Reduktion CGI-S um 1.32 Punkte). Somit scheinen Patient:innen in Itzehoe, die für 12 Monate die Möglichkeit auf die RECOVER Behandlung hatten, von diesem Angebot profitiert zu haben.

Tabelle 12. Veränderungen über die Zeit (Qualitätssicherungsstudie RECOVER)

Instrument	Baseline (t0)		t12	
	N	M (SD)	N	M (SD)
PSK12	45	21.61 (8.59)	53	32.47 (13.25)
GAF	83	52.82 (10.14)	76	67.97 (14.26)
CGI-S	83	4.20 (0.91)	76	2.88 (1.46)
Anmerkungen N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; % = Prozent.				

4. Schlussfolgerungen und Empfehlungen der drei evaluierenden Institute

Zunächst ist festzuhalten, dass die Evaluation eines innovativen Versorgungsmodells, welches die Behandlung für ein breites Spektrum von Patient:innen mit psychischen Störungen in der Versorgungsroutine kostengünstiger und kosteneffektiver gestalten sollte, erfolgreich wie geplant vorgenommen werden konnte.

Es sind dabei eine Vielzahl positiver Aspekte zu benennen:

- Die Studie wurde **vor Beginn der Datenerhebung** in einem internationalen Register (ClinicalTrials.gov) **registriert**, ein **Studienprotokoll** wurde publiziert. Ein **statistischer Analyseplan** für die Auswertung der Effektevaluation wurde vom auswertenden Institut für Medizinische Biometrie und Epidemiologie vor Erhalt der Daten erstellt.
- Es wurde mit erheblichem Aufwand eine **randomisiert kontrollierte Studie** mit einer ausreichend hohen Zahl eingeschlossener Patient:innen durchgeführt, wodurch erreicht wurde, dass sich bekannte, aber vor allem nicht bekannte mögliche Konfoundervariablen auf beide Gruppen gleichverteilen und somit nicht andere Faktoren als tatsächlich der Vergleich der beiden Versorgungsbedingungen, d.h. RECOVER vs. TAU, für die Interpretation der Ergebnisse herangezogen werden müssen.
- Es wurde ein **Vergleich mit einer TAU-Bedingung** vorgenommen, d.h. der klinischen Routine, womit die Wahrscheinlichkeit deutlich sinkt, sehr große Unterschiede in den untersuchten Outcomes nachweisen zu können. Allerdings wird es hierdurch möglich, im Falle hypothesenkonformer Ergebnisse eine hohe Relevanz für die Versorgung aufzuzeigen.
- Im RECOVER-Modell wurde **keine indikationsspezifische Stichprobe** untersucht, sondern Patient:innen mit einem breiten Spektrum von psychischen Störungen einbezogen, die für die Versorgung von hoher Relevanz sind. Demensprechend musste für die Effektevaluation ein generisches Outcome definiert werden, bei dem die Wahrscheinlichkeit hoch ist, klinisch relevante Veränderungen über alle Diagnosegruppen hinweg abbilden zu können. Dazu wurde ein generisches Composite Outcome vorab festgelegt, welches die selbsteingeschätzten Bereiche allgemeine Symptombelastung und gesundheitsbe-

zogene psychische Lebensqualität sowie die fremdeingeschätzte Funktionsfähigkeit umfasst. Bei diesem Vorgehen ist zu berücksichtigen, dass die nachzuweisenden Effekte des RECOVER-Modells weniger ausgeprägt ausfallen als bei Verwendung indikationsspezifische Maße, die eingesetzt werden können, wenn die Intervention auf nur eine einzige Diagnosegruppe abzielt (z.B. Patient:innen mit Depression). Um dieses Problem zu adressieren wurde im Rahmen des Projekts ergänzend ein innovatives Auswertungskonzept entwickelt, welches auch indikationsspezifische Maße einschließt und zumindest an einer Teilstichprobe untersucht werden konnte.

- Trotz der schwierigen Bedingungen durch die **COVID-19 Pandemie** ist es gelungen, eine ausreichend große Stichprobe von 881 Patient:innen zur Baseline in die Studie einzuschließen. Diese Stichprobe ist etwas reduziert gegenüber der ursprünglich angestrebten Stichprobengröße, erlaubt aber immer noch, die vorab festgelegte Effektstärke in den Outcomes mit einer ausreichend hohen Power von 80% inferenzstatistisch nachzuweisen und mit geringen Einschränkungen auch die geplanten Interaktionseffekte mit den vier Schweregradgruppen. Ausschlaggebend für die notwendige Fallzahlreduktion war die Schwierigkeit, die geplante Anzahl von Patient:innen mit einer leichten psychischen Störung, d.h. Stufe 1, innerhalb des gegebenen Zeitfensters in die Studie einzuschließen. Ein Grund hierfür könnte sein, dass bei vielen psychischen Störungen für diesen Schweregrad fachspezifische Interventionen zunächst nicht zwingend erforderlich sind und damit auch die Motivation zur Teilnahme an einer Interventionsstudie bei einer größeren Anzahl von Betroffenen nicht ausreichend hoch war. Für die Evaluation des RECOVER-Modells war es aber auch viel entscheidender, dass die weiteren Schweregradstufen 2-4 wie geplant rekrutiert werden konnten, was erfolgreich gelang.
- Die vorliegende Evaluation eines Versorgungsmodels wurde mit **sehr hohen methodischen Anforderungen** durchgeführt, bspw. die stringente Einhaltung des Studienprotokolls und des SAP. Dies wird aber auch einer Reihe weiterer Merkmale deutlich:
 - Es konnte eine Verblindung der Interviewer in allen Phasen der Erhebung im hohen Ausmaß realisiert werden,
 - alle Erhebungen und damit auch alle Fremdeinschätzungen erfolgten durch ein Team vom qualifizierten Psycholog:innen mit mindestens Diplom- oder Masterabschluss,
 - alle Fremdeinschätzungen wurden zu Beginn und im weiteren Verlauf supervidiert bzw. stichprobenartig auch durch Doppelbestimmungen überprüft,
 - das auswertende biometrische Institut hat die statistische Auswertung bezüglich der Gruppenzugehörigkeit verblindet vorgenommen,
 - es wurden eine Reihe von Sensitivitätsanalysen durchgeführt um die Interpretation der Ergebnisse zu unterstützen,
 - die für die Berechnung der QALYs verwendeten EQ-5D Indexwerte stammen aus einer aktuellen deutschen Stichprobe der Allgemeinbevölkerung,
 - die Erhebung der Daten für die gesundheitsökonomische Evaluation erfolgte mit publizierten standardisierten Instrumenten,
 - die monetäre Bewertung der direkten und indirekten Kosten wurde auf der Basis anerkannter publizierter Kostensätze vorgenommen.

Als **mögliche Einschränkungen** sind u.a. folgende Aspekte zu diskutieren:

- Ist die vorgenommene Einteilung der vier **Schweregradstufen** nachvollziehbar? Für ein Versorgungsmodell, welches einen Stepped Care Ansatz als ein zentrales Element beinhaltet, ist eine transparente, praktikable und klinisch relevante Zuordnung der Patient:innen von großer Bedeutung. Die Einteilung wurde im RECOVER-Projekt von denjenigen Testenden zur Baseline vorgenommen, die auch die Interviews und die Fremdeinschätzungen durchführten. Sie basierte mit dem GAF und CGI-S auf zwei standardisierten, international gut eingeführten Fremdeinschätzungsinstrumenten sowie ergänzend der Diagnose, die zur Baseline auf der Basis des diagnostischen Interviews gestellt wurde. Der Algorithmus beinhaltete eine Abstufung der Schweregrade 1 bis 3 in enger Anlehnung an die Bewertung der zehn Stufen des Funktionsniveaus der GAF, so dass bezogen auf die zurückliegenden vier Wochen der Wertebereich von 61-100 als leichter, von 51-60 als mittlerer und von 31-50 als mittlerer bis hoher Schweregrad eingestuft wurden. Die Stufe 4 wurde vergeben, wenn der GAF für die zurückliegenden sechs Monate als < 50 eingeschätzt wurde und eine emotional instabile Persönlichkeitsstörung, eine bipolare Störung oder eine psychotische Symptomatik vorlag. Die Entscheidung, für die Gruppe 4 nicht allein GAF und CGI-S für die letzten vier Wochen zu Grunde zu legen, sondern auch zwei besondere Diagnosegruppen einzubeziehen, spiegelt die Notwendigkeit wieder, die Gruppe der sehr schwer erkrankten Patient:innen, die gleichzeitig auch einen besonders hohen Therapiebedarf aufweisen, so zielgenau wie möglich zu identifizieren. Da eine ausschließliche dimensionale Einstufung nicht ausreicht, wurde hier um eine kategoriale Diagnostik ergänzt.
- **Selektivität der Stichprobe:** Hinsichtlich der Verteilung der soziodemographischen Merkmale wird vor allem deutlich, dass ca. die Hälfte der Patient:innen, die in die Studie eingeschlossen werden konnten, als höchsten Schulabschluss Abitur bzw. Fachabitur angaben, wohingegen dies nur für ein Drittel der Allgemeinbevölkerung zutrifft. Differenziert nach Altersgruppen und Gemeindegrößenklassen wird allerdings deutlich, dass in den für die RECOVER-Studie besonders häufigen Altersgruppen zwischen 20 und 50 Jahren für größere Städte (ab 100.000 Einwohner) der Anteil im Mittel zwischen 66,5% und 46,8% liegt (Statistisches Bundesamt (Destatis), 2020) und somit die untersuchte Stichprobe dieser Verteilung gut entspricht. Weiterhin ist zu diskutieren, dass die Stichprobe bezüglich der Verteilung der Diagnosen nicht repräsentativ für die Allgemeinbevölkerung ist, vor allem da zum einen keine organisch psychischen Störungen (F00) sowie keine psychischen Störungen durch psychotrope Substanzen (F10) einbezogen wurden und zum anderen durch die Stratifizierung nach Schweregradgruppen Patient:innen mit schwereren psychischen Störungen überrepräsentiert sind. Dies stellt aber keine Quelle für einen ausgeprägten Bias dar, da einerseits die genannten Störungen keine Zielpopulation für das Behandlungsmodell darstellen und deswegen auch nicht zu den Einschlusskriterien der Studie gehörten sowie andererseits für die Evaluation des RECOVER-Modells nicht die Verteilung in der Allgemeinbevölkerung maßgeblich war, sondern eher versorgungsepidemiologische Aspekte bei der Optimierung der Behandlung eine Rolle spielten und somit Gruppen mit besonders hohem Behandlungsbedarf im Fokus standen
- Für die gesundheitsökonomische Basisanalyse wurden **vollständig imputierte Daten** verwendet, da bei Kostenberechnungen schon ein einziger fehlender Wert in einer Leistungsart genügt, damit keine Summen mehr gebildet werden können. Es wurde für die maximale Transparenz der Ergebnisse entschieden, sowohl die primären Auswertungen

mit dem vollständig imputierten ITT-Datensatz (Datensatz I) zu berichten, aber als Sensitivitätsanalyse auch Datensatz II zu verwenden. Dies hatte für die Auswertung mit Datensatz II zur Folge, dass einerseits die Häufigkeit der Nutzung verschiedener, insgesamt selten genutzter Leistungen so stark reduziert wurde, dass Regressionsmodelle zur Testung auf einen adjustierten Kostenunterschied zwischen IG und KG nicht mehr angewendet werden konnten (sie sind nicht mehr konvergiert). Stattdessen mussten unadjustierte Signifikanztests durchgeführt werden, was jedoch zu Schätzern mit weniger Aussagekraft führte, als dies ein adjustierter Vergleich tun würde. Zum anderen wurde die statistische Power bei der Analyse der Kostendaten reduziert, was sich insbesondere auf den Verlauf der Kosten-Effektivitäts-Akzeptanz-Kurven auswirkte. So wird die Wahrscheinlichkeit, dass eine Intervention kosteneffektiv ist, vor allem bei niedriger Zahlungsbereitschaft wesentlich durch das Signifikanzniveau des Kostenunterschiedes beeinflusst. Insgesamt ergeben sich dadurch zwischen der Basisanalyse und der Sensitivitätsanalyse - vor allem aufgrund der eingeschränkten Power des nur Baseline imputierten ITT-Datensatzes (Datensatz II) - Abweichungen hinsichtlich der Signifikanzen.

- Obwohl durch die Erhebung der Inanspruchnahme mit Fragebögen viele Gesundheitsleistungen und damit auch Versorgungskosten erfasst wurden, konnten nicht alle wichtigen Kostenkategorien, wie die **Kosten für die Implementierung und das Management des RECOVER-Modells**, abgebildet werden. Eine präzise Ermittlung der Implementierungs- und Managementkosten gestaltet sich schwierig, da diese teilweise in den Kosten für erbrachte Leistungen enthalten sind (z.B. Sozialarbeiter, der auch als Case-Manager tätig ist), teilweise durch „normale“ Klinikaktivitäten gedeckt werden und teilweise einmalige Kosten darstellen, deren Anteil pro Patient:in im Zeitverlauf gegen 0 tendiert. Um dennoch eine Aussage zu den Kosten, die durch die verschiedenen personellen und finanziellen Umstrukturierungen durch die Einrichtung des RECOVER-Modells am UAE entstanden sind, machen zu können, wurde eine so genannte Threshold-Analyse durchgeführt. Dabei wurden die Kosten pro Studienteilnehmer:in in der IG schrittweise prozentual erhöht, bis der Kostenunterschied zwischen IG und KG nicht mehr statistisch signifikant war. Demnach könnten die Kosten in der IG um bis zu 13% oder durchschnittlich circa 2.210€ pro Teilnehmer höher sein als sie es im Erhebungszeitraum tatsächlich waren, und die gesellschaftlichen Kosten in der IG wären dann immer noch gerade statistisch signifikant niedriger als in der KG (20.237€ vs. 22.954€). Dies entspräche Gesamtkosten von 987.870€ für die gesamte IG-Stichprobe. Aus einer GKV-Perspektive könnten analog die Kosten in der IG um 18% bzw. 1.605€ pro Teilnehmer höher sein, was 717.435€ bezogen auf die betrachtete IG-Stichprobe entspricht.

Bezogen auf die primären Hypothesen und dazugehörigen sekundären Endpunkte zeigten sich zusammenfassend folgende Ergebnisse:

- 1) **Kostenreduktion: RECOVER** ist im Vergleich zur Regelversorgung mit **signifikant niedrigeren jährlichen Kosten** verbunden (primärer Endpunkt 1). Die Kosten aus der gesellschaftlichen Perspektive umfassten neben direkten Kosten auch indirekte Kosten durch Arbeitsunfähigkeitstage und waren in der IG (RECOVER) mit insgesamt 17.911€ vs. 22.950€ deutlich niedriger als in der KG (Regelversorgung). Insbesondere die Krankenhauskosten konnte in der IG statistisch signifikant um 50% reduziert werden. In den vier Schweregraden waren die Kosten in der IG ebenfalls niedriger als in der KG, wobei der Kostenunterschied nur in Stufe 2 und 4 statistisch signifikant war.

- 2) Bei Hochkostenpatienten (Stufe 4) zeigte sich in der IG eine signifikante Reduktion der Gesamtkosten um etwa 30% (20.677€ vs. 28.621€) und ebenfalls eine deutliche, signifikante Reduktion der Krankenhauskosten um 54%.
- 3) **Psycho-Soziales-Funktionsniveau: RECOVER** erreicht eine **stärkere Verbesserung im Vergleich zur Regelversorgung** (Primärer Endpunkt 2). Wenngleich dies nicht ausreichend konfirmatorisch belegt werden konnte, bestätigen doch alle Sensitivitätsanalysen, für die eine höhere Power vorhanden ist (z.B. Analyse basierend auf der vollständig imputierten Stichprobe), den gefundenen Unterschied und seine Höhe. Das zusammengesetzte Outcome aus selbsteingeschätzter allgemeiner Symptombelastung und gesundheitsbezogener psychischer Lebensqualität sowie fremdeingeschätzte Funktionsfähigkeit verbessert sich demnach in der IG (RECOVER) in höherem Ausmaß im Vergleich zur KG (Regelversorgung). Bei der Betrachtung der Höhe des Unterschieds ist zu berücksichtigen, dass der verwendete Composite-Score ein Summenwert der drei verwendeten z-transformierten Skalen ist und somit nicht direkt als Effektstärke verwendet werden kann. Wird dies entsprechend berücksichtigt, kann von einem besseren Behandlungsergebnis von RECOVER im Vergleich zur Regelversorgung analog einer Differenz mit etwas weniger als einer kleinen Effektstärke ausgegangen werden. Da trotz ausreichender Power kein Interaktionseffekt zwischen den Faktoren Gruppe (IG vs. KG) und Schweregradstufe nachgewiesen werden konnte, ist davon auszugehen, dass die Behandlungsergebnisse nicht unterschiedlich für die einzelnen Schweregradgruppen ausfielen.
- 4) **Kosteneffektivität:** Das RECOVER Modell ist im Vergleich zu der Regelversorgung kosteneffektiv (primärer Endpunkt 3). RECOVER ist im Vergleich zur Regelversorgung mit einer **Wahrscheinlichkeit von >95% kosteneffektiv**, unabhängig von der Zahlungsbereitschaft des Entscheidungsträgers (zwischen 0 bis 250.000 pro gewonnenem QALY) – mit Ausnahme von Stufe 1. Der Punktschätzer der **Kosteneffektivitätsrelation zeigt Dominanz des RECOVER Modells** im Vergleich zu der Regelversorgung.

Bezogen auf die **Kosten-Effektivitäts-Akzeptanz-Kurven** zeigte sich für die vollständig imputierte ITT-Stichprobe, dass die RECOVER-Intervention mit einer sehr hohen Wahrscheinlichkeit kosteneffektiv ist. Diese Wahrscheinlichkeit wird vor allem durch die Kosteneinsparungen getrieben. Bei den stufenspezifischen Subgruppenanalysen war das Ergebnis weniger deutlich, vor allem in der Baseline imputierten ITT-Stichprobe (Sensitivitätsanalyse mit Datensatz II). Hier muss jedoch beachtet werden, dass die Stichprobengrößen in den stufenspezifischen Analysen teilweise so klein waren, dass insbesondere bei den Kosten signifikante Gruppenunterschiede teilweise nicht zu erwarten waren. Dies beeinflusst die Wahrscheinlichkeit für Kosten-Effektivität jedoch erheblich, da diese auf der Grundlage des p-Werts für den Unterschied zwischen IG und KG ermittelt wird.

Der Effekt für den **primären Endpunkt Psycho-Soziales-Funktionsniveau** in der Gesamtpopulation wurde in den meisten Subgruppen ebenfalls beobachtet. Allerdings zeigte sich ein Unterschied im Effekt bei Patient:innen, welche komplett vor Ort befragt werden konnten und solchen, die im Zuge der Pandemie nur teilweise oder gar komplett telefonisch kontaktiert wurden. Der größere Vorteil für das RECOVER-Modell gegenüber der Regelversorgung zeigte sich hier für die Gruppe, die ihre Kontakte nur in Präsenz hatten. Da der primäre Endpunkt 2 ein Kompositionswert aus zwei Selbsteinschätzungen und einer Fremdeinschätzung ist,

scheint das Format der Fremdeinschätzung bedeutsam zu sein. Dies wird auch dadurch bestätigt, dass für die beiden Selbsteinschätzungen zur Lebensqualität (PSK SF-12) und Symptombelastung (Modul A Health-49) ein Vorteil des RECOVER-Model ersichtlich wird.

Für die **sekundären Hypothesen** zeigte sich selten ein Vorteil des RECOVER Model gegenüber der Regelversorgung. Bezogen auf spezifische Interventionen konnte allein für den Peer Support eine erhöhte Anwendung in RECOVER gezeigt werden. Auch bezogen auf Selbsteinschätzungen zu Lebensqualität (Re-QoL) und Symptombelastungen (Modul A Health-49 und diagnosespezifische Symptombelastung) zeigte sich ein Vorteil für 12 Monate RECOVER-Behandlung. Alle weiteren sekundären Endpunkte zeigten vergleichbare Ausprägungen zu t12 zwischen den beiden Gruppen.

Im Rahmen des Transfers von RECOVER in die kleinstädtische Region Itzehoe wurden wesentliche Elemente des RECOVER-Modells erfolgreich implementiert: Zum Beispiel die diagnostikgestützte Schweregradeinteilung im Rahmen einer Indikationssprechstunde, Kriseninterventionen, Netzwerkgespräche mit Betroffenen und Angehörigen, schweregradgestufte Behandlungskomponenten (z.B. Home-Treatment). Der erfolgreiche Transfer zeigt sich in deskriptiv verbesserten Werten in den zentralen Endpunkten der RECOVER-Behandlungsgruppe, wobei hier aufgrund der fehlenden Vergleichsgruppe keine Aussagen zu deren Vorteil ggü. der Regelversorgung getätigt werden können (und auch nicht intendiert waren). Ein Vergleich der Stichproben RECOVER Hamburg und Itzehoe zeigt ähnliche Zusammensetzungen der Gruppen. Dies deutet auf ähnliche Behandlungsbedarfe in den beiden unterschiedlichen Regionen hin, welche jeweils positiv adressiert wurden. Insgesamt zeigen die Ergebnisse des Transfers nach Itzehoe, dass wesentliche Elemente des RECOVER-Modells erfolgreich übertragen werden konnten.

Fazit:

Für das RECOVER-Model konnte gezeigt werden, dass bei gleichbleibenden bzw. sogar geringfügig besseren Behandlungsergebnissen eine erhebliche Kostenersparnis erzielt werden konnte. Es wird ein großes Potential für die Implementierung in anderen

5. Literaturverzeichnis

- Aiken, L. R. (1987). Testing with multiple-choice items. *Journal of Research & Development in Education*, 20(4), 44-58.
- Allison, P. D. (2001). Missing Data. *Thousan Oaks, CA: Sage citation*.
- Bock, J. O., Brettschneider, C., Seidl, H., Bowles, D., Holle, R., Greiner, W., & König, H. H. (2015). [Calculation of standardised unit costs from a societal perspective for health economic evaluation]. *Gesundheitswesen*, 77(1), 53–61. doi:10.1055/s-0034-1374621
- Gold, L. H. (2014). DSM-5 and the Assessment of Functioning: The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *Journal of the American Academy of Psychiatry and the Law Online*, 42(2), 173. Retrieved from <http://jaapl.org/content/42/2/173.abstract>
- Govindarajulu, Z. (1988). Alternative methods for combining several test scores. 48, 53-60. doi:10.1177/001316448804800107
- Grupp, H., König, H.-H., Riedel-Heller, S., & Konnopka, A. (2017). FIMPsy – Fragebogen zur Inanspruchnahme medizinischer und nicht medizinischer Versorgungsleistungen bei psychischen Erkrankungen: Entwicklung und Verwendung. *Psychiatrische Praxis*, 45. doi:10.1055/s-0042-118033
- Grupp, H., König, H. H., & Konnopka, A. (2017). [Calculation of Standardised Unit Costs for the Economic Evaluation of Mental Disorders]. *Gesundheitswesen*, 79(1), 48-57. doi:10.1055/s-0035-1555950
- Guy, W. (1976). *ECDEU assessment manual for psychopharmacology*: US Department of Health, Education, and Welfare, Public Health Service
- Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., . . . Busschbach, J. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*, 22(7), 1717-1727. doi:10.1007/s11136-012-0322-4
- Keetharuth, A. D., Brazier, J., Connell, J., Bjorner, J. B., Carlton, J., Taylor Buck, E., . . . Barkham, M. (2018). Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *The British Journal of Psychiatry*, 212(1), 42-49. doi:10.1192/bjp.2017.10
- Lambert, M., Karow, A., Gallinat, J., Ludecke, D., Kraft, V., Rohenkohl, A., . . . Schulz, H. (2020). Study protocol for a randomised controlled trial evaluating an evidence-based, stepped and coordinated care service model for mental disorders (RECOVER). *BMJ Open*, 10(5), e036021. doi:10.1136/bmjopen-2019-036021
- Ludwig, K., Graf von der Schulenburg, J. M., & Greiner, W. (2018). German Value Set for the EQ-5D-5L. *Pharmacoeconomics*, 36(6), 663-674. doi:10.1007/s40273-018-0615-8
- Rabung, S., Harfst, T., Koch, U., Wittchen, H. U., & Schulz, H. (2007). „Hamburger Module zur Erfassung allgemeiner Aspekte psychosozialer Gesundheit für die therapeutische Praxis (HEALTH)“- psychometrische Überprüfung eines neuen Selbstbeurteilungsinstruments zur multidimensionalen Erfassung psychosozialer Gesundheit. [“Hamburg Modules for the Assessment of Psychosocial Health (HEALTH)“- Psychometric Analysis of a New Self-Report Questionnaire for the

Akronym: RECOVER

Förderkennzeichen: 01NVF160018

Multidimensional Assessment of Psychosocial Health]. *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin*, 17(03), 133-140.

Seidl, H., Bowles, D., Bock, J. O., Brettschneider, C., Greiner, W., König, H. H., & Holle, R. (2015). FIMA – Fragebogen zur Erhebung von Gesundheitsleistungen im Alter: Entwicklung und Pilotstudie. [FIMA – Questionnaire for Health-Related Resource Use in an Elderly Population: Development and Pilot Study]. *Gesundheitswesen*, 77(01), 46-52.

Statistisches Bundesamt. (2018). *Verdienste und Arbeitskosten. Arbeitskosten im Produzierenden Gewerbe und im Dienstleistungsbereich- Ergebnisse für Deutschland - 2016*. Retrieved from Wiesbaden:

Statistisches Bundesamt. (2019). *Verdienste und Arbeitskosten. Arbeitnehmerverdienste 2019*. Retrieved from Wiesbaden:

Statistisches Bundesamt (Destatis). (2020). *Bildungsstand der Bevölkerung- Ergebnisse des Mikrozensus 2019*.

Ware, J., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*, 34(3), 220-233. doi:10.1097/00005650-199603000-00003

Zethraeus, N., Johannesson, M., Jönsson, B., Löthgren, M., & Tambour, M. (2003). Advantages of using the net-benefit approach for analysing uncertainty in economic evaluation studies. *Pharmacoeconomics*, 21(1), 39-48. doi:10.2165/00019053-200321010-00003

Akronym: RECOVER
Förderkennzeichen: 01NVF160018

6. Anhang

Anhang 1: Detailergebnisse

7. Anlage

Nicht zutreffend.

Anhang 1 Detailliergebnisse

Tabelle A1. Operationalisierung der Indikation zu den Schweregradstufen 1-4 inklusive Diagnose und Schweregrad der Beeinträchtigung des Funktionsniveaus

Operationalisierung	Schweregradstufen im RECOVER-Modell			
	1	2	3	4
	Leichter Schweregrad	Mittlerer Schweregrad	Mittlerer bis hoher Schweregrad	Sehr hoher Schweregrad
Hauptsächliche Störungsbilder nach ICD-10	F32, F40, F41, F43.2, F45, F90	F32, F40, F41, F42, F43.1, F43.2, F45, F50, F90	F20, F22, F23, F25, F31, F33, F41, F42, F43.1, F45, F50, F60, F61	F20, F22, F23, F25, F31, F32.3, F33.3, F60
Indikation nach Funktionsniveau GAF („Beeinträchtigung hinsichtlich der sozialen, beruflichen oder schulischen Leistungsfähigkeit“)	GAF 61-100 Punkte: <i>keine oder leichte Beeinträchtigung in den letzten 4 Wochen</i>	GAF 51-60 Punkte: <i>mäßig ausgeprägte Beeinträchtigung in den letzten 4 Wochen</i>	GAF 31-50 Punkte: <i>ernste oder starke Beeinträchtigung in den letzten 4 Wochen oder</i>	GAF ≤ 50 Punkten über mindestens 6 Monate: <i>dauerhafte ernste oder starke Beeinträchtigung</i>
Indikation nach klinischem Schweregrad CGI („Schwere der psychischen Erkrankung“)	CGI 1-3 <i>Nicht vorhanden bis leichter Schweregrad der psychischen Erkrankung in den letzten 4 Wochen vor Beginn einer diagnosespezifischen Behandlung</i>	CGI 3-4 <i>Leichter bis mittlerer Schweregrad der psychischen Erkrankung in den letzten 4 Wochen vor Beginn einer diagnosespezifischen Behandlung</i>	CGI 4-6 <i>Mittlerer bis hoher Schweregrad der psychischen Erkrankung in den letzten 4 Wochen vor Beginn einer diagnosespezifischen Behandlung</i>	CGI 5-7 <i>Hoher Schweregrad der psychischen Erkrankung in den letzten 4 Wochen vor Beginn einer diagnosespezifischen Behandlung</i>

GAF = Global Assessment of Functioning Scale.

Tabelle A2. Darstellung unadjustierter Kosten (in €) aus gesellschaftlicher Perspektive in Datensatz II

	IG (N=242)	KG (N=221)	Differenz IG-KG
	M (SF)	M (SF)	
Direkte Kosten			
t0	6997 (705)	6053 (654)	944
t6	4682 (440)	6577 (696)	-1895
t12	3097 (245)	3201 (323)	-104
Indirekte Kosten			
t0	3555 (487)	3490 (482)	65
t6	3598 (573)	4336 (669)	-738
t12	2311 (488)	2875 (618)	-564
Gesamtkosten			
t0	10552 (918)	9543 (920)	1009
t6	8279 (737)	10913 (988)	-2634
t12	5408 (577)	6076 (743)	-668
Anmerkungen.			
<p>Direkte Kosten: stationäre Aufenthalte, ambulante Arztkontakte, ambulante Heilmittelerbringer, Medikamente, Pflege/Hilfe, informelle Pflege, Reha, Beratung & Betreuung, Berufliche Integration, Krisen-Interventionen, Webbasierte Angebote, Peerbegleitung. Indirekte Kosten: Produktivitätsverluste durch AU-Tage. SF = Standardfehler, M = Mittelwert, N = Anzahl Probanden.</p>			

Tabelle A3. Unadjustierte Kosten über den 12-Monatszeitraum, nach Kostenkategorien in Datensatz II (in € 2019)

Kostenkategorie	Alle			Stufe 1			Stufe 2			Stufe 3			Stufe 4		
	IG n=242	KG n=221	Δ	IG n=18	KG n=24	Δ	IG n=104	KG n=80	Δ	IG n=81	KG n=81	Δ	IG n=39	KG n=36	Δ
Ärzte/Psychotherapeuten	3102	1932	***	1682	1685		2791	1919	***	3838	1929		3059	2132	*
Allgemeinmedizin	116	134		112	117		118	148		122	137		103	110	
Psychologe/Psychotherapeut	1635	1150	***	695	1104		1551	1230		2031	1066		1468	1192	
Psychiater	181	157		110	103		105	106		275	177		222	262	
PIA	90	95		0	97		64	56		128	114		122	139	
Sonstige Arztgruppen	1080	395	***	764	263	**	954	378	***	1281	435		1144	429	***
Heilmittelerbringer	158	232		82	266		99	113		148	269		372	389	
Stationäre Aufenthalte	1717	3572	*	35	751		840	1309		3019	5196		2129	6827	
Allgemein-KH	384	314		35	0		349	342		409	424		588	212	
Psychiatrie	1333	3258	*	0	751		491	967		2611	4771		1541	6615	*
Teilstationäre Aufenthalte	545	1771	***	54	305		251	1511	*	648	2315		1341	2102	
Allgemein-KH	128	0	¹	0	0	n.a.	0	0	n.a.	0	0	n.a.	796	0	¹
Psychiatrie	417	1771	***	54	305		251	1511	*	648	2315		545	2102	
Rehabilitationen	35	179		0	0	n.a.	62	212		25	279		0	0	n.a.
Pflege/Hilfen	370	443		86	0		139	393		583	425		672	886	
bezahlte Haushalts-hilfe	88	51		86	0		87	90		124	50		17	0	
gesetzlicher Be-treuer	28	10		0	0	n.a.	1	3		17	21		136	8	

Sozialarbeiter/Be- gleitperson	226	288		0	0	n.a.	18	143		442	354		438	649	
amb. psychiatr. Pfl- gedienst	27	94		0	0	n.a.	33	157		0	0	n.a.	80	229	
Informelle Pflege	470	568		0	545		248	295		869	528		451	1283	
Medikamente	793	639		116	301		896	706		710	473		1002	1088	
Beratung & Betreu- ung	132	219		7	112		89	193		141	163		286	477	
Beratungsstelle psych. Prob- leme/Sucht	81	103		7	10		89	123		48	110		164	109	
Sozialpsychiatri- scher Dienst	50	109		0	103		0	69		93	38		120	359	
Tagesstätte/Kon- taktstätten	0	8	*	0	0	n.a.	0	1		0	15		2	9	
Berufliche Integration	255	140		359	19		179	84	*	316	145		286	333	
Integrationsfach- dienst	46	19		0	0	n.a.	40	44	*	78	0	**1	18	19	
Eignungsabklärung	36	12	***	13	0		30	14		56	11		23	19	
Geschützte Werk- statt - Arbeitsbe- reich	39	42		0	0	n.a.	0	0	n.a.	0	0	n.a.	239	259	
Unterstützte Be- schäftigung - Berufs- begleitung	7	14		0	0	n.a.	5	8		13	16		0	30	
Berufliche Weiter- bildung	57	38		334	0	¹	16	1		74	104		0	0	n.a.
Unterstützte Be- schäftigung - Jobcoaching	71	14	***	13	19		87	17	***	95	14		6	6	

Kriseninterventionen	50	41		21	0		24	99		79	0		76	31	
Krisen-Telefon	3	0	***	0	0		1	0		7	0		5	0	
Hausbesuche	47	41		20	0		23	99		72	0	*1	71	31	
Webbasierte Angebote	86	36	**	129	77		143	29	***	40	34		12	26	
Peerbegleitung	64	7	***	60	0		65	3	***	86	16		20	3	
Indirekte Kosten durch AU-Tage	5909	7211		7951	10876		6957	7929		4993	6019		4071	5856	
Kosten Gesamt	13687	16990		10581	14937		12784	14794		15495	17792		13776	21433	

Anmerkungen

Δ = Signifikanz der Kostendifferenz zwischen IG (RECOVER) und KG (TAU), * p<0.05, ** p<0.01, und *** p<0.001. n.a.= Signifikanzniveau nicht auswertbar, da in beiden Gruppen keine Leistungen in Anspruch genommen wurden (Kosten = 0 €). ¹ Das Signifikanzniveau für die Kostenunterschiede wurde mittels t-test berechnet, da in einer Gruppe keine Leistungen in Anspruch genommen wurden (Kosten = 0 €). Heilmittelerbringer: Ergotherapie/Arbeitstherapie, Kunsttherapie, Massage/Lymphdrainage, Heilpraktiker, Krankengymnastik, Sport/Bewegungstherapie.

	F2	48 (7.5)	26 (7.6)	0.29 (2.91)	0.16 (-0.68, 0.99)	22 (7.3)	-0.77 (2.12)	-0.72 (-1.63, 0.18)	-0.88 (-2.11, 0.35)	0.1611	
	F3	297 (46.1)	164 (48.1)	0.11 (2.48)	0.24 (-0.09, 0.58)	133 (43.9)	-0.01 (2.33)	-0.00 (-0.37, 0.37)	-0.24 (-0.74, 0.25)	0.3352	
	F4	150 (23.3)	78 (22.9)	0.15 (2.76)	0.51 (0.03, 1.00)	72 (23.8)	0.04 (2.22)	0.40 (-0.11, 0.90)	-0.11 (-0.81, 0.58)	0.7496	
	F6	145 (22.5)	72 (21.1)	-0.21 (2.02)	-0.64 (-1.15, -0.14)	73 (24.1)	-0.41 (2.15)	-1.07 (-1.58, -0.56)	-0.43 (-1.13, 0.28)	0.2386	
	Andere	4 (0.6)	1 (0.3)	-2.14 (NA)	-0.00 (-4.27, 4.27)	3 (1.0)	-0.93 (0.21)	-0.19 (-2.66, 2.27)	-0.19 (-5.11, 4.73)	0.9383	
Art Visite											0.0208
	Nur Präsenz	357 (55.6)	198 (58.1)	0.17 (2.48)	0.18 (-0.12, 0.49)	159 (52.8)	-0.46 (2.21)	-0.51 (-0.85, -0.16)	-0.69 (-1.15, -0.23)	0.0033	
	Präsenz telefo- nisch/n ur tele- fonisch	285 (44.4)	143 (41.9)	-0.10 (2.50)	-0.01 (-0.37, 0.35)	142 (47.2)	0.17 (2.25)	0.11 (-0.25, 0.47)	0.12 (-0.39, 0.63)	0.6432	
Covid 19											0.3483
	Vor Lock- down	325 (51.9)	181 (54.0)	0.01 (2.47)	0.08 (-0.24, 0.40)	144 (49.5)	-0.34 (2.17)	-0.39 (-0.75, -0.03)	-0.47 (-0.96, 0.01)	0.0539	
	Nach Lock- down	301 (48.1)	154 (46.0)	0.07 (2.53)	0.11 (-0.24, 0.45)	147 (50.5)	0.03 (2.26)	-0.04 (-0.39, 0.32)	-0.14 (-0.64, 0.35)	0.5723	
Stationäre Vorbehand- lung											0.4349
	Nein	243 (37.7)	137 (40.2)	0.05 (2.57)	0.40 (0.03, 0.78)	106 (35.0)	0.07 (2.09)	0.29 (-0.13, 0.71)	-0.11 (-0.67, 0.44)	0.6898	

	Ja	401 (62.3)	204 (59.8)	0.07 (2.44)	-0.09 (-0.40, 0.21)	197 (65.0)	-0.28 (2.31)	-0.49 (-0.79, -0.18)	-0.39 (-0.82, 0.04)	0.0730	
--	----	---------------	---------------	----------------	------------------------	---------------	-----------------	-------------------------	------------------------	--------	--

Anmerkungen.

IG= Interventionsgruppe; KG = Kontrollgruppe (treatment-as-usual); N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; % = Prozent.

Tabelle A5. Ergebnisse der Sensitivitätsanalysen (Primärer Endpunkt 2) „Psycho-soziales Funktionsniveau“)

Daten- grundlage	IG			KG			KG vs. IG		
	N	beo. M (SD)	adj. M (95% KI)	N	beo. M (SD)	adj. M (95% KI)	Adj. Diff. (95% KI)	p-Wert	Inter. p-Wert (Gr.,Sg)
Multiple Imputation	447	0.18 (2.42)	0.33 (0.09, 0.57)	444	-0.18 (2.28)	-0.05 (-0.28, 0.19)	-0.38 (-0.68, -0.08)	0.0139	0.6557
Vollständig impu- tierte ITT-Stichprobe (Datensatz I)	447	0.13 (2.40)	0.29 (0.06, 0.52)	444	-0.13 (2.27)	-0.00 (-0.23, 0.23)	-0.29 (-0.57, -0.01)	0.0426	0.4678
Optimal weights me- thod	341	0.01 (0.83)	0.07 (-0.02, 0.16)	303	-0.05 (0.74)	-0.03 (-0.12, 0.06)	-0.10 (-0.21, 0.02)	0.0931	0.5427
Per-Protokoll	300	0.08 (2.51)	0.25 (-0.03, 0.53)	267	-0.13 (2.24)	-0.08 (-0.37, 0.20)	-0.33 (-0.69, 0.03)	0.0745	0.6008
Complete Case	294	0.04 (2.46)	0.26 (-0.02, 0.55)	264	-0.18 (2.25)	-0.09 (-0.38, 0.19)	-0.36 (-0.72, 0.01)	0.0534	0.7036
Anmerkungen									
IG= Interventionsgruppe; KG = Kontrollgruppe (treatment-as-usual); N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; KI = Konfidenz-Intervall, Gr = Gruppe (IG,KG), Sg = Schweregrad (Stufe 1-4).									

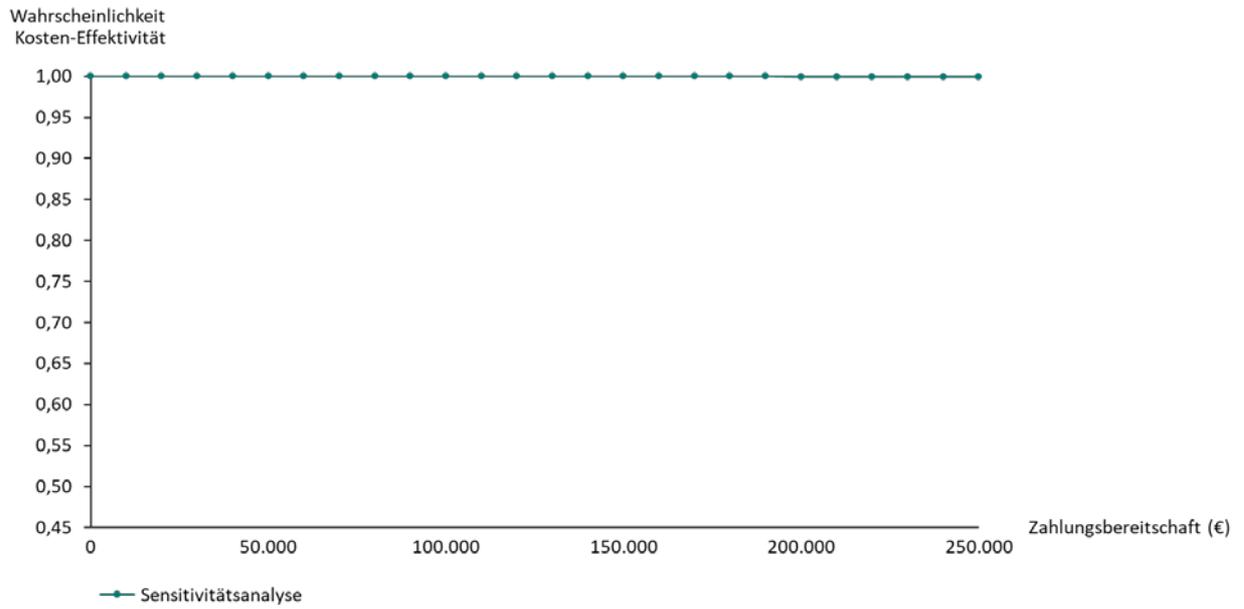


Abbildung A6. Kosteneffektivitätsakzeptanzkurve für Sensitivitätsanalyse 1 (Primärer Endpunkt 3)

Tabelle A7. Adjustierte Schätzer für QALYs über den 12-Monatszeitraum in Datensatz II

	M IG	M KG	Differenz	p-Wert
Gesamt	0.75	0.74	0.01	0.331
Stufe 1	0.85	0.87	-0.02	0.439
Stufe 2	0.80	0.78	0.02	0.201
Stufe 3	0.69	0.67	0.02	0.464
Stufe 4	0.70	0.71	-0.02	0.588
Anmerkungen				
M = Mittelwert. Adjustierung der Effekte für: Alter, Geschlecht, Baseline-Utilities, Somatische Erkrankungen, Krankheitsschwere, Berechnung mit OLS-Regression.				

Tabelle A8. Unadjustierte ICER für Datensatz II (Gesellschaftliche Perspektive)

	M IG	M KG	Differenz	ICER (€/QALY)
Kosten	13687	16990	-3303	Dominanz
QALY	0.76	0.74	0.02	
Anmerkungen. Kosten in €, 2019				

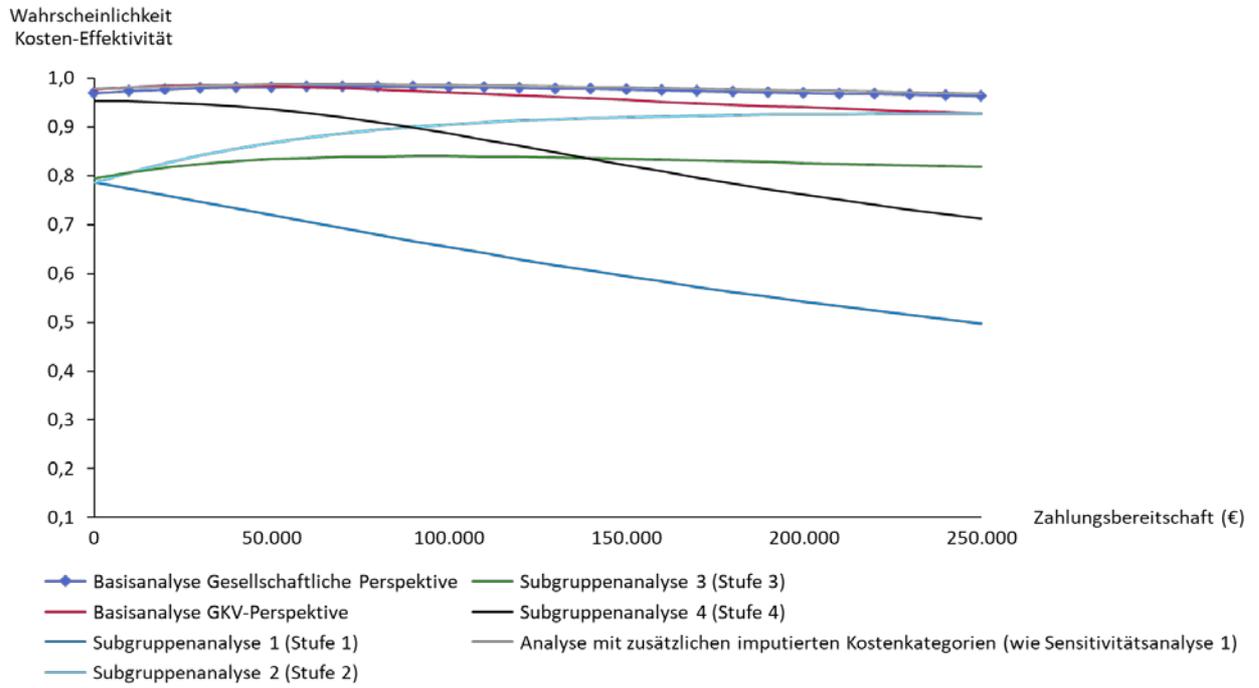


Abbildung A9. Kosteneffektivitätsakzeptanzkurven für alle Szenarien der Sensitivitätsanalyse 2 (Primärer Endpunkt 3)

Tabelle A10. Ergebnisse der linearen gemischten Modelle

Zielgröße	IG			KG			KG vs. IG		
	N	Beobachter Mittelwert (sd)	Adjustierter Mittelwert (95% KI)	N	Beobachter Mittelwert (sd)	Adjustierter Mittelwert (95% KI)	Adjustierte Differenz (95% KI)	p-Wert	Interaktions p-Wert (Gruppe-Zeit)
Primärer Endpunkt 2									0.8475
t0	447	0.03 (2.23)		444	-0.03 (2.16)		-0.29 (-0.58, 0.00)	0.0519	
t6-t0	364	0.11 (2.30)	0.16 (-0.08, 0.39)	302	-0.17 (2.37)	-0.13 (-0.38, 0.12)			
t12-t0	341	0.06 (2.49)	0.18 (-0.06, 0.42)	303	-0.16 (2.24)	-0.11 (-0.36, 0.14)			
HEALTH 49 (invertiert)									0.6540
t0	447	2.46 (0.74)		444	2.42 (0.76)		-0.12 (-0.20, -0.04)	0.0041	
t6-t0	382	0.37 (0.65)	0.40 (0.34, 0.47)	320	0.29 (0.65)	0.29 (0.22, 0.36)			
t12-t0	360	0.43 (0.69)	0.46 (0.39, 0.52)	322	0.33 (0.64)	0.34 (0.27,0.41)			
HEALTH standardisiert									0.6443
t0	447	0.03 (0.99)		444	-0.03 (1.01)		-0.14 (-0.24, -0.05)	0.0041	
t6-t0	382	0.03 (0.83)	0.08 (-0.00, 0.16)	320	-0.06 (0.83)	-0.07 (-0.15, 0.01)			
t12-t0	360	0.04 (0.89)	0.08 (-0.01, 0.16)	322	-0.08 (0.82)	-0.07 (-0.15,0.01)			

t0	396	0.06 (0.98)		393	0.05 (0.96)				
t6-t0	327	-0.13 (0.96)	-0.17 (-0.26, -0.07)	263	-0.11 (0.92)	-0.12 (-0.22, -0.01)	0.05 (-0.08, 0.18)	0.4766	
t12-t0	297	-0.20 (1.07)	-0.22 (-0.33, -0.12)	267	-0.05 (0.96)	-0.05 (-0.15, 0.06)	0.18 (0.05, 0.31)	0.0085	

Anmerkungen.

IG= Interventionsgruppe; KG = Kontrollgruppe (treatment-as-usual); N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; % = Prozent, KI = Konfidenz-Intervall.

Tabelle A11. Ergebnisse der gemischten logistischen Regression

Zielgröße	IG	KG	KG vs. IG		
	Zielgröße =Ja (%)	Zielgröße =Ja (%)	adjustierted odds ratio (95% CI)	p-Wert	Interaktions p-Wert (Gruppe - Zeit)
Symptomatische Remission					0.4708
t0	60/447 (13.42)	60/444 (13.51)	0.95 (0.42, 2.13)	0.9041	
t6	109/391 (27.88)	89/331 (26.89)			
t12	126/373 (33.78)	109/327 (33.33)			
Funktionale Remission					0.9740
t0	52/447 (11.63)	49/444 (11.04)	1.11 (0.44, 2.82)	0.8215	
t6	134/391 (34.27)	117/331 (35.35)			
t12	165/373 (44.24)	143/327 (43.73)			
Psychotherapeutische Gruppeninterventionen					0.0229
t6	40/191 (20.94)	18/136 (13.24)	0.25 (0.02, 3.12)	0.2845	
t12	21/178 (11.80)	23/136 (16.91)	10.46 (0.48, 226.16)	0.1344	
Web-basierte Therapie					<0.001
t6	43/382 (11.26)	12/318 (3.77)	0.21 (0.02, 2.15)	0.1893	
t12	16/358 (4.47)	10/321 (3.12)	95.51 (4.48, 2036.17)	0.0035	
Peer-Unterstützung					0.0343
t6	88/382 (23.04)	13/320 (4.06)	0.13 (0.02, 0.75)	0.0226	
t12	55/360 (15.28)	5/321 (1.56)	0.01 (0.00, 0.14)	0.0015	

Psychotherapeutische Kurzzeitinterventionen					0.8497
t6	100/380 (26.32)	70/320 (21.88)	0.76 (0.53, 1.08)	0.1275	
t12	83/360 (23.06)	62/322 (19.25)			
Ambulante psychoth. Interventionen schwere psychische Störungen					0.7805
t6	231/380 (60.79)	157/320 (49.06)	0.46 (0.30, 0.70)	0.0003	
t12	218/360 (60.56)	159/322 (49.38)			
Behandlungsabbrüche					0.7819
t6	47/381 (12.34)	34/316 (10.76)	0.88 (0.37, 2.12)	0.7771	
t12	42/359 (11.70)	35/321 (10.90)			

Tabelle A12. Ergebnisse der gemischten negativen binomial Regression

Zielgröße	IG		KG		KG vs. IG		
	n	Beobachteter Mittelwert (sd)	n	Beobachteter Mittelwert (sd)	adjustiertes Ratio (95% CI)	p-Wert	Interaktions p-Wert (Gruppenzeit)
Dauer psychiatrischer Krankenhausaufenthalte							0.8904
t0	447	0.43 (1.67)	444	0.70 (2.81)	1.08 (0.34, 3.43)	0.9000	
t6	391	0.29 (1.31)	331	0.37 (1.87)			
t12	372	0.45 (2.26)	326	0.53 (3.44)			
Arbeitsunfähigkeit							0.9610
t0	447	16.30 (32.37)	444	13.11 (27.18)	1.05 (0.57, 1.91)	0.8849	
t6	380	16.67 (38.93)	327	17.17 (38.55)			
t12	366	10.99 (32.77)	321	12.21 (36.12)			
Rate der psychiatrischen Krankenhausaufenthalte							0.1498
t6	390	0.11 (0.40)	330	0.28 (0.60)	2.31 (1.53, 3.51)	<0.001	
t12	372	0.10 (0.50)	326	0.17 (0.58)			
Spezifische psychiatrische Interventionen bei schweren psychischen Störungen							0.3537
t6	382	0.96 (3.58)	320	0.14 (0.81)	0.12 (0.06, 0.23)	<0.001	
t12	360	0.49 (1.89)	322	0.04 (0.21)			
Anmerkungen.							
IG= Interventionsgruppe; KG = Kontrollgruppe (treatment-as-usual); N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; % = Prozent.							

Tabelle A13. Ergebnisse mit 3-fach Interaktion

Zielgröße		RECOVER	TAU	TAU vs RECOVER				
	Base-line Schweregrad	Zielgröße=Ja (%)	Zielgröße=Ja (%)	Adjustiertes odds ratio (95% CI)	p-Wert	3-fach Interaktion p-Wert	Interaktions p-Wert (Zeit – Gruppe)	Interktions p-Wert (Schweregrad- Gruppe)
Symptomatische Remission						0.0552	0.2494	0.5163
t0	1	26/30 (86.67)	30/30 (100.00)					
t6	1	19/28 (67.86)	20/28 (71.43)	1.05 (0.06, 20.06)	0.9723			
t12	1	19/25 (76.00)	21/25 (84.00)	8.08 (0.17, 376.37)	0.2864			
t0	2	32/165 (19.39)	28/160 (17.50)					
t6	2	75/151 (49.67)	50/117 (42.74)	0.45 (0.08, 2.48)	0.3576			
t12	2	76/147 (51.70)	63/112 (56.25)	1.79 (0.30, 10.57)	0.5213			
t0	3	2/159 (1.26)	2/158 (1.27)					
t6	3	13/138 (9.42)	16/121 (13.22)	2.75 (0.43, 17.41)	0.2821			
t12	3	27/128 (21.09)	22/120 (18.33)	0.61 (0.16, 2.30)	0.4696			
t0	4	0/93 (0.00)	0/96 (0.00)					
t6	4	2/74 (2.70)	3/65 (4.62)	3.31 (0.08, 129.31)	0.5220			
t12	4	4/73 (5.48)	3/70 (4.29)	0.53 (0.03, 9.31)	0.6670			
Funktionale Remission						0.0110	0.1454	0.3505
t0	1	26/30 (86.67)	29/30 (96.67)					
t6	1	19/28 (67.86)	20/28 (71.43)	2.09 (0.04, 118.79)	0.7199			

t12	1	19/25 (76.00)	23/25 (92.00)	93.82 (0.25, 35701.38)	0.1341			
t0	2	26/165 (15.76)	20/160 (12.50)					
t6	2	75/151 (49.67)	50/117 (42.74)	0.74 (0.13, 4.36)	0.7374			
t12	2	90/147 (61.22)	74/112 (66.07)	6.95 (0.49, 97.50)	0.1505			
t0	3	0/159 (0.00)	0/158 (0.00)					
t6	3	13/138 (9.42)	16/121 (13.22)	11.25 (0.73, 173.47)	0.0830			
t12	3	44/128 (34.38)	37/120 (30.83)	0.68 (0.13, 3.57)	0.6489			
t0	4	0/93 (0.00)	0/96 (0.00)					
t6	4	2/74 (2.70)	3/65 (4.62)	14.72 (0.09, 2307.66)	0.2971			
t12	4	12/73 (16.44)	9/70 (12.86)	0.66 (0.04, 12.04)	0.7797			
Ambulante psychoth. Interventionen schwere psychische Störungen						0.2794	0.1697	0.0439
t6	1	9/27 (33.33)	13/26 (50.00)	3.53 (0.57, 21.84)	0.1754			
t12	1	7/25 (28.00)	16/25 (64.00)	15.65 (2.23, 109.87)	0.0057			
t6	2	100/147 (68.03)	54/112 (48.21)	0.25 (0.11, 0.58)	0.0012			
t12	2	92/144 (63.89)	56/110 (50.91)	0.41 (0.18, 0.93)	0.0340			

t6	3	85/134 (22.37)	58/117 (49.57)	0.36 (0.16, 0.83)	0.0165			
t12	3	79/120 (65.83)	55/119 (46.22)	0.26 (0.11, 0.62)	0.0024			
t6	4	37/72 (51.39)	32/65 (49.23)	0.84 (0.28, 2.53)	0.7532			
t12	4	40/71 (56.34)	32/68 (47.06)	0.50 (0.17, 1.52)	0.2243			

Anmerkungen.

IG= Interventionsgruppe; KG = Kontrollgruppe (treatment-as-usual); N = Anzahl Probanden; M= Mittelwert; SD = Standardabweichung; % = Prozent.