

Endbericht

Evaluierung fit2work klinisch-psychologische Einzelbehandlung im Vergleich zur Kombinationsbehandlung mit Kunsttherapie: 5. Fördervertrag (1.1.2019 bis 31.12.2019)

Reinhold Jagsch

Institut für Angewandte Psychologie:
Gesundheit, Entwicklung und Förderung

Fakultät für Psychologie
Universität Wien

Wien, Mai 2021

Inhaltsverzeichnis

Abstract	2
1. Einleitung	4
2. Methoden	6
2.1. Zu prüfende Fragestellungen.....	6
2.2. Erhebungsinstrumente.....	7
2.2.1. Fragebogen zur Erfassung des krankheitsspezifischen Status	7
2.2.2. Beck-Depressions-Inventar BDI-II (Hautzinger, Keller & Kühner, 2006)	8
2.2.3. Brief Symptom Inventory BSI-53 (Franke, 2000).....	9
2.2.4. Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire HAQ (Bassler, Potratz & Krauthauser, 1995)	10
2.3. Statistische Analyse	11
3. Beteiligung an der Evaluation	13
4. Darstellung der Ausgangssituation (Stichprobe der Ersttestung)	13
4.1. Vergleich soziodemographischer Charakteristika	13
4.2. Vergleich krankheitsspezifischer Charakteristika.....	14
5. Ergebnisse zur Veränderungsmessung sowie zur Behandlungszufriedenheit.....	16
5.1. Veränderungen in der Inanspruchnahme von Gesundheitsleistungen und im Gesundheitsverhalten	17
5.2. Veränderungen bezüglich der validierten klinischen Fragebögen.....	18
5.3. Bewertung der Behandlungszufriedenheit mittels HAQ.....	19
5.4. Direkter Vergleich der Bewertungen der Behandlungsarten	20
6. Dropout-Analyse	20
7. Zusammenfassung.....	21
Literatur	27
Anhang A. Abbildungen.....	28
Anhang B. Tabellen	34

Abstract

Theoretischer Hintergrund. Auf Basis der österreichischen Gesetzeslage mit einerseits Psychologengesetz und andererseits Psychotherapiegesetz sind Klinische Psychologinnen und Psychologen und Psychotherapeutinnen und Psychotherapeuten sowie Behandlerinnen und Behandler mit Doppelqualifikation berechtigt, Personen mit psychischen Störungen zu behandeln. Nachdem bereits belegt werden konnte, dass im Rahmen der Evaluation von fit2work Behandlungen durch die drei Teilgruppen hoch wirksam und gleichwertig sind, lag der Schwerpunkt der aktuellen Analyse auf dem Vergleich von klinisch-psychologischen Behandlungen, die im Einzel- bzw. im Kombinationsformat (Einzelbehandlung mit ergänzender Kunsttherapie) angeboten wurden. Zielpersonen mit Indikation für die Kombinationsbehandlung waren solche, die aktuell in einem Beschäftigungsverhältnis standen, dieses aber durch psychische Belastungen (vor allem durch Burnoutsymptomatik) stark gefährdet war.

Methode. Österreichweit nahmen insgesamt 1112 Personen (1046 in Einzel-, 66 in Kombinationsbehandlung), deren Arbeitsplatz entweder durch gesundheitliche Probleme gefährdet war bzw. bei denen der Arbeitsplatz bereits verlorengegangen war, an dieser Studie teil und wurden nach einer umfangreichen Anamnese einer der Behandlungsarten zugewiesen. Die Klientinnen und Klienten erhielten zu Beginn der Behandlung eine Testbatterie, die Fragebögen zu Depression (Beck-Depressions-Inventar-II, BSI-II), zur psychischen Beeinträchtigung (Brief Symptom Inventory-53, BSI-53) und einen Fragebogen zum arbeits- und gesundheitsbezogenen Status enthielt. Nach Abschluss der Behandlung wurden sie gebeten, diese Testbatterie noch einmal zu bearbeiten und die Zufriedenheit der Behandlung mittels Helping Alliance Questionnaire (HAQ) zu beurteilen. Ergänzt wurde die subjektive Einschätzung der Klientinnen und Klienten durch die Fremdeinschätzung der Behandlerinnen und Behandler, die online durchgeführt wurde und die auch eine Beurteilung mittels GAF-Skala¹ (Global Assessment of Functioning) beinhaltete.

¹ Fremdeinschätzungsskala von 1 bis 100 für Funktionsfähigkeit in Fertigkeiten des täglichen Lebens

Ergebnisse. Insgesamt absolvierten 898 Klientinnen und Klienten die Behandlungen und füllten zu beiden Zeitpunkten die Fragebögen aus (80.9%, 446 in Einzel- und 52 in Kombinationsbehandlung). Laut den Daten vor der Behandlung sind die Klientinnen und Klienten der beiden Behandlungsarten als zwei distinkte Zielgruppen der Maßnahme zu sehen, die sich sowohl in soziodemographischen (höherer Frauenanteil und höherer Akademikeranteil bei der Kombinationsbehandlung) als auch in arbeitsbezogenen Variablen grundsätzlich unterschieden, aber in Bezug auf psychische Belastungen durchaus ähnlichen Hintergrund aufwiesen. Nach der Behandlung wiesen beide Behandlungsarten maßgebliche Verbesserungen auf: In allen klinischen Skalen des BSI-53 lagen Klientinnen und Klienten beider Behandlungsarten über dem Cutoff-Wert² für Behandlungsbedürftigkeit, nach Abschluss der Behandlung in keiner der Subskalen. Neben einem hoch signifikanten Effekt der Dauer der Behandlung konnten im Rahmen der Veränderungsmessung simultan weder signifikante Gruppeneffekte noch signifikante Interaktionen gefunden werden. Auch das statistische Modell für das Depressionsausmaß zeigte mit signifikanten Reduktionen analoge Ergebnisse für beide Behandlungsarten. Die Ergebnisse der GAF-Skala zeigten ähnlich starke Verbesserungen für beide Behandlungsarten auf, mit dem Unterschied, dass das Niveau der Funktionsfähigkeit der Klientinnen und Klienten der Kombinationsbehandlung bereits vor der Behandlung ein höheres war. Bezüglich Reintegration in den Arbeitsprozess kann vermerkt werden, dass die Quote des Wiedereinstiegs relativ zu der Quote der Klientinnen und Klienten, die in der Zeit der Behandlung ihre Arbeit verloren hatten, für die klinisch-psychologische Einzelbehandlung signifikant höher lag. Die Interventionsmaßnahmen führten zusätzlich, ohne Unterschiede zwischen den Behandlungsarten, zu einer massiven Reduktion der Inanspruchnahme von Gesundheitsleistungen. So konnten bedeutsame Reduktionen bei Arztbesuchen, bei ärztlichen Untersuchungen, bei Krankenhausaufenthalten und bei der Einnahme von Psychopharmaka verzeichnet werden. Auch das Gesundheitsverhalten (Sport, Bewegung, Ernährung) konnte positiv verändert werden. Beziehungs- und

² Trennwert, bei dessen Überschreitung Behandlungsbedürftigkeit angenommen wird

Erfolgswufriedenheit scorten in beiden Behandlungsarten klientenseitig hoch, behandlerseitig erwiesen sie sich im Vergleich signifikant höher für die Kombinationsbehandlung. Im direkten Vergleich wurden die beiden Teile der Behandlung (klinisch-psychologische Behandlung, Kunsttherapie) von den Klientinnen und Klienten der Kombinationsbehandlung als hoch wirksam und hilfreich bewertet.

Schlussfolgerungen. Die Evaluation konnte belegen, dass zwei distinkte Zielgruppen von den auf sie zugeschnittenen Behandlungsarten angesprochen werden und von diesen profitieren konnten. Die Interventionen konnten zu hoch signifikanten Reduktionen in den klinischen Skalen und den Belastungen auf individueller Ebene führen wie auch auf gesamtgesellschaftlicher Ebene zu einer massiven Reduktion der Inanspruchnahme von Gesundheitsleistungen beitragen. Gleichzeitig berichteten die Klientinnen und Klienten von bedeutsamen Verbesserungen im arbeits- und gesundheitsbezogenen Bereich, verbunden mit einer hohen Behandlungs- und Erfolgswufriedenheit.

1. Einleitung

Die Studie war als Vergleich von drei unterschiedlichen Behandlungssettings geplant und wurde anfangs auch in dieser Form durchgeführt. Es wurden drei unterschiedliche klinisch-psychologische Behandlungsarten für Klientinnen und Klienten angeboten, einerseits die beiden bereits früher eingesetzten Formen der klinisch-psychologischen Einzel- und Gruppenbehandlung, zusätzlich neu diesmal aber auch eine kombinierte Form aus klinisch-psychologischer Behandlung und Kunsttherapie. Im Rahmen der Gruppenbehandlungen wurden drei unterschiedliche Formen eingerichtet, einerseits die „psychosozialen und Reha-Jahresgruppen“, andererseits eine Gruppe mit „Gruppentraining Soziale und Emotionale Kompetenz, SEK)“ sowie drittens Gruppen, bei denen die Bearbeitung von Schmerzen im Vordergrund stand („Schmerzgruppen“), diese aber in deutlich unterschiedlicher Zahl (eine Jahresgruppe, zwei „Schmerzgruppen“ sowie 14 SEK-Gruppen). Gruppen, die in der ersten Hälfte 2019 begonnen hatten, konnten erfolgreich abgeschlossen werden.

Durch das Aufkommen der Corona-Pandemie musste ab März 2020 ein Teil der Gruppenbehandlungen vorzeitig eingestellt werden, die laufenden Gruppenbehandlungen wurden unterbrochen, konnten aber bis zum Ende der Studie nicht wieder aufgenommen werden. Die betroffenen Klientinnen und Klienten bekamen ersatzweise Einzelbehandlungen, entweder im Rahmen der laufenden oder der darauffolgenden Evaluationsstudie. Durch die Maßnahmen infolge der Pandemie ergaben sich für die Evaluation unerwartete Konsequenzen: Während die Einzel- und die Kombinationsbehandlungen nach kurzfristigen Unterbrechungen in modifizierter Form fortgesetzt werden konnten, ergab sich aufgrund der beschriebenen Situation für die Gruppenbehandlung eine Verteilung mit heterogenen Teilgruppen (unterschiedliche Gruppenformen, teils abgeschlossene, teils abgebrochene Gruppen) mit teilweise extrem niedrigen Fallzahlen, was für letztere eine seriöse Datenanalyse nicht zuließ. Als Konsequenz fokussiert die Auswertung dieses Endberichts auf den Vergleich der Einzelbehandlungen mit der kombinierten Behandlung mit Kunsttherapie.

Die Einzelbehandlung umfasste maximal bis zu 30 Behandlungseinheiten klinisch-psychologischer Behandlung pro behandelter Person. Einer definierten Teilgruppe von Klientinnen und Klienten, die zwar im Berufsleben stand, aber stark von Burnout-Gefährdung bedroht war, wurde zusätzlich zu 18 Einheiten klinisch-psychologischer Einzelbehandlung ein Kontingent von 12 Einheiten Kunsttherapie angeboten. Der Umfang der betreuten Klientinnen und Klienten wurde mit 780 (davon 80 in der Kombination mit Kunsttherapie) für die Einzelbehandlung festgelegt, war aufgrund der durch die Pandemie bedingten geänderten Rahmenbedingungen aber deutlich erhöht (N=1112). Im Unterschied zu vorangegangenen Evaluationen wurden die Behandlungen ausschließlich durch Klinische Psychologinnen und Psychologen (zum Teil mit Doppelqualifikation Psychotherapie) durchgeführt, die Behandlungen im Rahmen der Kunsttherapie durch eine Reihe von Diplomierten Kunsttherapeutinnen. Aus organisatorischen Gründen konnten die kunsttherapeutischen Behandlungen nicht in allen Fällen parallel mit den klinisch-psychologischen Einzelbehandlungen gestartet werden, sondern teilweise auch davor oder danach.

Allgemeines Ziel dieser Evaluation ist es, die Effektivität der Maßnahme zu überprüfen, d.h. eine Evaluierung im Sinne einer Ergebnisevaluation (Outcome Evaluation) durchzuführen. Aufbau und Konzept, Erhebungsinstrumente sowie Auswertungsmethoden entsprechen weitgehend denen der vorangegangenen Evaluationen (Jagsch, 2015, 2018, 2020). Die Darstellung der Ergebnisse erfolgt wieder, wie gewünscht, in kurzer und prägnanter Weise, die inhaltliche Darstellung punktuell kompakt im Fließtext, die ergänzenden Abbildungen und Tabellen finden sich zusammengefasst in den Anhängen.

2. Methoden

2.1. Zu prüfende Fragestellungen

Ziel der Evaluation ist es, im Rahmen der Studie die Wirksamkeit der Maßnahme im Sinne einer Ergebnisevaluation zu prüfen. Basis für diesen Vergleich bilden alle Klientinnen und Klienten, die die Behandlungseinheiten regulär absolviert und die vorgesehenen Fragebögen zu beiden Zeitpunkten (Behandlungsbeginn, Behandlungsabschluss) ausgefüllt und an das Evaluationsteam retourniert haben (Completer-Analyse). Gleichzeitig werden auch Daten der Behandlerinnen und Behandler in die Analyse miteinbezogen, die die dafür vorgesehenen Instrumente der Fremderhebung via Online-Erhebung bearbeitet haben. Teil der Evaluation ist auch eine Dropout-Analyse, um Charakteristika der Ausgangsstichprobe in Bezug auf die Repräsentativität der Gesamtstichprobe zu erheben.

Inhaltlich soll die Evaluation folgende Fragestellungen auf Basis der erhobenen empirischen Daten prüfen:

1. Veränderungen definierter Outcome-Variablen (Depression, psychische Beeinträchtigung, Funktionsfähigkeit, arbeits- und gesundheitsbezogene Variablen) im prä/post-Vergleich vor und nach Abschluss der Behandlung
2. Gleichwertigkeit der Behandlungsergebnisse zwischen Einzel- und Kombinationsbehandlung
3. Prüfung von Interaktionseffekten (Wechselwirkung von Behandlungsart und

Zeit)

4. Vergleich der Beziehungs- und Erfolgszufriedenheit mit der Behandlung zwischen den Behandlungsarten auf Ebene der Behandelten und der Behandlerinnen und Behandler
5. Prüfung der Gesamteffekte bezüglich Reduktion der Inanspruchnahme von Gesundheitsleistungen

2.2. Erhebungsinstrumente

Für die Erhebung kamen vier verschiedene Fragebögen zum Einsatz, ein speziell für diese Studie von der Academy for Value in Health GmbH erstellter Fragebogen zur Erfassung des krankheitsspezifischen Status, der in einer Selbstbeurteilungsform für die Klientinnen und Klienten und einer Fremdbeurteilungsform für die Behandlerinnen und Behandler vorliegt, und zwei validierte klinische Fragebögen zur Erhebung der Stärke der Depression (Beck-Depressions-Inventar) und der psychischen Beeinträchtigungen (Brief Symptom Inventory) jeweils in Selbstbeurteilungsform. Zusätzlich wurde nach Abschluss der Behandlung der Fragebogen Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire (HAQ) im Selbst- und Fremdbeantwortungsformat vorgegeben, um die Zufriedenheit mit dem Behandlungsergebnis von beiden Seiten abzuklären.

2.2.1. Fragebogen zur Erfassung des krankheitsspezifischen Status

Dieser Fragebogen dient dazu, einerseits soziodemographische (Alter, Geschlecht, Familienstand, höchste abgeschlossene Ausbildung, aktuelles Arbeitsverhältnis etc.) und andererseits arbeits- und krankheitsspezifische Daten (Zeitpunkt der Diagnosestellung, Beurteilung der Schwere der psychischen Belastung für den Bereich Arbeit und alltägliche Tätigkeiten, Umfang der Inanspruchnahme von Gesundheitsleistungen vor Aufnahme in die aktuelle Behandlung, Einnahme von Psychopharmaka etc.) zu erheben. Die Fremdbeurteilungsform, die in Form einer Online-Befragung vorgegeben wird, dient dazu, die durch die Behandlerinnen und Behandler erhobenen Diagnosen zu erfragen, die Bereiche der Beeinträchtigungen zu

beschreiben, die verschriebenen Medikamente aufzulisten sowie den Grad der Funktionseinschränkungen zu objektivieren.

Diagnosen werden nach der Version 10 der International Classification of Diseases (ICD) der Weltgesundheitsorganisation vergeben, die die Vergabe von zusätzlichen komorbiden Diagnosen erlaubt, d.h. den zu Diagnostizierenden können mehrere Diagnosen zugeordnet werden. Der Grad der Funktionseinschränkungen der Klientinnen und Klienten wird von den Behandlerinnen und Behandlern mittels des Global Assessment of Functioning (GAF-Skala; Saß, Wittchen, Zaudig & Houben, 2003) eingeschätzt. Mit Hilfe normierter Vorgaben von Beschreibungen von Einschränkungen können so Werte auf einer Skala von 1 bis 100 zur Beurteilung des aktuellen Funktionsniveaus getätigt werden, wobei 1 den höchsten Grad der Einschränkung und 100 keinerlei Einschränkung des Funktionsniveaus bedeutet.

2.2.2. Beck-Depressions-Inventar BDI-II (Hautzinger, Keller & Kühner, 2006)

Das BDI-II ist ein objektives, reliables und valides Instrument zur Erfassung der Schwere depressiver Symptomatik in Stichproben aus der Normalpopulation sowie aus klinischen Populationen. Die interne Konsistenz des Verfahrens wird für eine klinische Population mit Cronbach's Alpha .93 und für eine gesunde mit .90 angegeben. Das BDI-II als Selbstbeurteilungsverfahren beinhaltet 21 verschiedene Gruppen von Aussagen, die aus Sätzen bestehen, die typische depressive Symptome beschreiben. Die Aussagen umfassen jeweils vier (in den Inhaltsbereichen Appetit und Schlaf sieben) Sätze pro Gruppe und sind nach aufsteigender Schwere der Symptome von Aussage eins (0 = nicht vorhanden) bis Aussage vier (3=stark vorhanden) geordnet.

Die Befragten sollen in jeder der 21 Gruppen die Aussage ankreuzen, die am besten beschreibt, wie sie sich in den vergangenen zwei Wochen einschließlich des Testtages selbst gefühlt haben. Kreuzen sie bei einer Gruppe von Aussagen mehrere Aussagen an, so geht die Aussage mit der höchsten Zahl in den Summenwert ein. Der Summenwert aller Items, der zwischen 0 und 63 Punkten liegen kann, bildet den Gesamtwert der Stärke der depressiven Verstimmung. Summenwerte unter 14

Punkten werden als unauffällig interpretiert, Werte zwischen 14 und 19 weisen auf eine leichte Depression, Werte von 20 bis 28 auf eine mittelschwere Depression und Werte von 29 bis 63 Punkten auf eine schwere Depression hin.

2.2.3. Brief Symptom Inventory BSI-53 (Franke, 2000)

Das BSI ist eine der Kurzformen der Symptom-Checkliste SCL-90-R. Mit Hilfe dieses Selbstbeurteilungsverfahrens können psychische Beeinträchtigungen über die letzten sieben Tage durch die Beurteilung von insgesamt 53 Items erfasst werden. Die interne Konsistenz der Subskalen liegt mit Cronbach's Alpha zwischen .71 bis .85 im guten bis sehr guten Bereich. Die individuelle Stärke der psychischen Beeinträchtigung wird dabei mittels fünfstufiger Likert-Skala (0=überhaupt nicht, 1=ein wenig, 2=ziemlich, 3=stark, 4=sehr stark) beurteilt.

Der Fragebogen enthält insgesamt neun Subskalen:

- 1) *Somatisierung* (7 Items): Ausmaß von psychischen Beeinträchtigungen, die durch die Wahrnehmung von körperlichen Dysfunktionen entstehen
- 2) *Zwanghaftigkeit* (6 Items): Umfasst Gedanken, Impulse und Handlungen, die von der Person als konstant vorhanden, unveränderbar sowie als ich-fremd erlebt werden
- 3) *Unsicherheit im Sozialkontakt* (4 Items): Umfasst Fragen von leichter sozialer Unsicherheit bis hin zum Gefühl völliger persönlicher Unzulänglichkeit
- 4) *Depressivität* (6 Items): Umfasst Traurigkeit bis hin zu schwerer Depressivität
- 5) *Ängstlichkeit* (6 Items): Umfasst körperlich spürbare Nervosität bis hin zu tiefer Angst
- 6) *Aggressivität/Feindseligkeit* (5 Items): Umfasst Reizbarkeit und Unausgeglichenheit bis hin zu aggressiven Durchbrüchen
- 7) *Phobische Angst* (5 Items): Umfasst ein leichtes Gefühl von Bedrohung bis hin zu massiven phobischen Ängsten
- 8) *Paranoides Denken* (5 Items): Umfasst Misstrauen und Gefühle der

Minderwertigkeit bis hin zu starkem paranoiden Denken

9) *Psychotizismus* (5 Items): Umfasst das Gefühl von Isolation und Entfremdung bis hin zu Psychosen

Neben diesen einzelnen Skalen besteht auch die Möglichkeit, einen Globalwert (GSI) zu berechnen, der die grundsätzliche psychische Beeinträchtigung misst. Die Ausprägungen der Stärkegrade auf Einzelitemniveau werden jeweils aufaddiert und bilden so die Rohwerte für die Subskalen und den Globalwert. Anschließend können die Rohwerte der Subskalen und des Globalwertes, die unterschiedliche Itemanzahl und daher unterschiedliche Spannweiten aufweisen, mittels Normierung in *T*-Werte umgewandelt werden. *T*-Werte sind Werte einer Normwertskala, die auf einen Mittelwert (*M*) von 50 und eine Standardabweichung (*SD*) von 10 hin normiert sind, d.h. dass alle Werte im Umfang von maximal drei Standardabweichungen rund um den Mittelwert von 50 im Gesamtbereich von 20 bis 80 abgebildet werden können. Für das BSI liegen darüber hinaus für jede Subskala und den Globalwert *T*-Wert-Normen getrennt für Männer und Frauen vor, d.h. die Unterschiedlichkeit der Ausprägungen auf geschlechtsspezifischem Niveau, die für einige psychopathologische Bereiche vorliegt, kann damit ausgeschaltet werden. Laut Autorin des Verfahrens stellt ein *T*-Wert im GSI von 63 aufwärts eine Indikation für Behandlungsbedürftigkeit dar, das selbe gilt, wenn in mindestens zwei Subskalen des BSI ein *T*-Wert von 63 erreicht oder überschritten wird.

2.2.4. Zufriedenheit mit Behandlung und Beziehung: Helping Alliance Questionnaire HAQ (Bassler, Potratz & Krauthauser, 1995)

Das HAQ umfasst elf Fragen, mit deren Hilfe die Qualität der therapeutischen Beziehung von den Klientinnen und Klienten eingeschätzt werden kann. In der deutschen Übersetzung existiert darüber hinaus eine Fremdeinschätzungsform, mit der mit entsprechend umformulierten Fragen auch die therapeutenseitige Einschätzung erhoben werden kann. Auf Basis faktorenanalytischer Untersuchungen lassen sich neben dem Gesamtwert der Zufriedenheit (Cronbach's Alpha: .91) zwei Subskalen unterscheiden: Skala 1 (6 Items, Cronbach's Alpha: .95) erhebt die

Beziehungszufriedenheit, Skala 2 (5 Items) die Erfolgszufriedenheit (Cronbach's Alpha: .91). Jedes der elf Items kann mit Hilfe einer Likert-Skala (von 1=sehr unzutreffend bis 6=sehr zutreffend) beurteilt werden.

2.3. Statistische Analyse

Zur Beschreibung der Stichprobe werden deskriptivstatistisch für kontinuierliche Daten Mittelwert (M), Standardabweichung (SD) und Spannweite (Minimum; Maximum) verwendet, Häufigkeitsdaten werden mittels absoluten und relativen Häufigkeiten beschrieben. Für die Vergleiche von Gruppenmittelwerten kommen statistische Verfahren zum Einsatz, die auf Basis von Varianzzerlegungen Aussagen darüber erlauben, ob Unterschiede zwischen verschiedenen Gruppen als bedeutsam (signifikant) beurteilt werden können. Für die Veränderungsmessung bei kontinuierlichen Daten wurde nach Abklärung der notwendigen Voraussetzungen (Normalverteilung der Teilstichproben, Homogenität der Varianzen und der Kovarianzen) das Allgemeine Lineare Modell (ALM) für Messwiederholungen angewendet. Dieses ist in der Lage, gleichzeitig den Haupteffekt Zeit (Veränderungen prä/post), den Haupteffekt Gruppe (Effekt der Behandlungsart) sowie die Wechselwirkung zwischen Zeit und Gruppe zu überprüfen. Um einer Kumulation des Alphafehlers³ entgegenzuwirken, kam bei signifikanten Ergebnissen die Alphafehler-Korrektur nach Bonferroni-Holm zum Einsatz.

Vergleiche bei Häufigkeitsdaten werden mit dem χ^2 -Test nach Pearson oder mittels Fisher's Exact Test durchgeführt. Um Veränderungen von dichotomen Daten (z.B. beschäftigt vs. beschäftigungslos) analysieren zu können, wird der McNemar-Test verwendet, der die Häufigkeit von Wechsleren in gegenläufige Richtungen (z.B. von beschäftigungslos zu in Beschäftigung vs. von in Beschäftigung zu beschäftigungslos) in Beziehung setzen kann (bei Konstanthaltung der Personen, die zu beiden Zeitpunkten den gleichen Beschäftigungsgrad aufweisen). Im Rahmen einer Dropout-

³ Kumulation des Alphafehlers bedeutet, dass aufgrund von mehrfachen Analysen (z.B. bei mehreren Subskalen eines Testverfahrens) die Tendenz Richtung Signifikanz potentiell ansteigt, der mittels Methoden zur Alphafehler-Korrektur entgegengewirkt werden muss.

Analyse werden im ersten Schritt die maßgeblichen Variablen mit entsprechenden statistischen Methoden auf signifikante Unterschiede hin geprüft, im zweiten Schritt werden die dabei ausfindig gemachten Variablen einer binär-logistischen Regressionsanalyse zugeführt, um Prädiktoren für die Frage ausfindig zu machen, welche charakteristischen Variablen vorhersagen können, welche Klientinnen und Klienten trotz vorhandener Ersttestung nicht an der Abschlusstestung teilgenommen haben. Für direkte statistische Vergleiche der beiden Behandlungsarten im Rahmen der Kombinationsbehandlung wurden der *t*-Test für abhängige Stichproben bzw. der Wilcoxon-Test herangezogen.

Für alle statistischen Analysen wurde ein Signifikanzniveau von $p < .05$ festgelegt. Aufgrund der Tatsache, dass Angaben im Sinne des ausschließlichen Anführens von Kennwerten zur Absicherung statistischer Signifikanz bei stark unterschiedlich großen Teilstichproben zu verzerrten Ergebnissen führen können, empfiehlt es sich, zusätzlich Kennwerte der inhaltlichen Relevanz (Effektstärken) ergänzend anzugeben (Bortz & Döring, 2006). Diese haben den Vorteil, dass deren Ausmaß unabhängig von den vorhandenen Stichprobengrößen interpretiert werden kann. Für Vergleiche von zwei Stichproben wird Cohen's *d* verwendet (Cohen, 1988). Für diese Art der Effektstärke gilt:

Werte bis 0.20	kein Effekt
Werte von 0.21 bis 0.50	kleiner Effekt
Werte von 0.51 bis 0.80	moderater Effekt
Werte ab 0.81	großer Effekt

Für Vergleiche mittels ALM für Messwiederholung ist als Effektstärkemaß η^2 anzuführen (Cohen, 1988). Für diese Art der Effektstärke gilt:

Werte bis .010	kein Effekt
Werte von .011 bis .060	kleiner Effekt
Werte von .061 bis .140	moderater Effekt
Werte ab .141	großer Effekt

3. Beteiligung an der Evaluation

Insgesamt sind 1219 ausgefüllte Fragebogensets der Ersttestung (Beginn der Behandlung) beim Evaluationsteam der Fakultät für Psychologie der Universität Wien eingelangt, nach Abzug derjenigen der Gruppenbehandlung waren es 1112, davon 1046 von Einzel- und 66 von Kombinationsbehandlungen. Von Seiten der Behandlerinnen und Behandler lagen insgesamt 593 Online-Fremdbeurteilungen vollständig für Behandlungsbeginn und -ende vor, das entspricht einer Quote von 53.3%. Davon beziehen sich 557 auf die Einzel- (Quote: 53.3%) und 36 (Quote: 54.5%) auf die Kombinationsbehandlung. Für die Abschlusstestungen der Einzelbehandlungen lagen von 846 Klientinnen und Klienten vollständige Datensätze für die beiden Erhebungen vor, das sind 80.9% der 1046 Ersttestungen. Für die Kombinationsbehandlung lag die Quote ähnlich hoch (78.8%). Für eine vollständige Übersicht der Gesamt- sowie der beiden Teilpopulationen siehe die Venn-Diagramme in den Abbildungen 1, 2 und 3.

Ziel des Endberichts ist es, einen Vergleich der Effektivität von Einzel- und Kombinationsbehandlung durchzuführen. Als Stichtag für das Ende der Aufnahme von Fragebögen wurde als Termin der 12. Mai 2021 gewählt, um die benötigten Vorarbeiten (Verknüpfung der selbst- und fremdberichteten Daten, Data Cleaning, Erstellung von Summenscores, Transformation der Rohdaten des BSI) für die Datenanalyse durchführen und so die Darstellung der Ergebnisse zum Termin der Abgabe des Endberichts gewährleisten zu können.

4. Darstellung der Ausgangssituation (Stichprobe der Ersttestung)

4.1. Vergleich soziodemographischer Charakteristika

Die Gesamtstichprobe der Ersttestung lässt sich folgendermaßen charakterisieren: Der Großteil ist weiblich ($n=755$, 68.0%), das Durchschnittsalter liegt bei 44.78 Jahren ($SD=10.13$; Range 18–63). Der vorherrschende Zivilstatus ist verheiratet ($n=501$, 45.5%), gefolgt von ledig ($n=342$, 31.0%), geschieden ($n=249$, 22.6%) und verwitwet ($n=10$, 0.9%). Vom Bildungsstand her herrscht abgeschlossene Ausbildung

„Berufsschule/Lehre“ vor ($n=426$, 38.5%). Unterschiede zur Gesamtfallzahl von 1112 ergeben sich durch unvollständige Beantwortung der einzelnen Variablen. Bezüglich der soziodemographischen Charakteristika unterscheiden sich die zwei Behandlungsarten bezüglich Geschlecht ($\chi^2(1)=6.143$, $p=.013$) sowie im Bildungsstand ($\chi^2(6)=15.272$, $p=.018$). Die Kombinationstherapie wies einen höheren Frauenanteil sowie eine höhere Quote an akademischem Bildungsabschluss auf.

Was die berufsspezifischen Daten anlangt, waren insgesamt 734 Klientinnen und Klienten (66.1%) zum Zeitpunkt der Ersterhebung arbeitslos, von denen etwa zwei Drittel auf Nachfrage angaben, aufgrund von psychischen Beschwerden arbeitslos geworden zu sein. Die Durchschnittsdauer der Arbeitslosigkeit wurde mit 23.05 Monaten ($SD=35.73$; Median=9 Monate, Spannweite 1–284 Monate) angegeben. In Bezug auf die arbeitslosenspezifischen Daten der Klientinnen und Klienten zeigte sich ein statistisch hoch signifikanter Unterschied zwischen den zwei Behandlungsarten ($\chi^2(1)=104.937$, $p<.001$): Während in der Einzelbehandlung die Quote für „arbeitslos“ bei 69.7% lag, lag die entsprechende Quote für die Kombinationsbehandlung bei lediglich 7.7%, d.h. hier war die Gesamtheit der Klientinnen und Klienten bis auf wenige Ausnahmen im Gegensatz zur Einzelbehandlung berufstätig.

4.2. Vergleich krankheitsspezifischer Charakteristika

Von insgesamt 873 Klientinnen und Klienten lagen Daten zu Diagnosen aus den Online-Fragebögen der Behandlerinnen und Behandler vor (Diagnosestellung laut International Classification of Diseases, ICD-10). Der Anteil der Klientinnen und Klienten mit Doppel- bzw. Mehrfachdiagnosen ($n=415$, 47.5%) lag knapp unter demjenigen mit lediglich einer einzigen Diagnose. Den Angaben zufolge können drei Hauptgruppen von Diagnosen unterschieden werden: Störungen aus der F4-Sektion (Neurotische, Belastungs- und somatoforme Störungen, $n=506$, 58.0%) stehen an erster Stelle, gefolgt von Affektiven Störungen (F3-Diagnosen, $n=478$, 54.8%) und Störungen bezüglich Z73 (Probleme bei der Lebensbewältigung, Burnout, $n=119$,

13.6%). Es konnten dabei zwei statistisch signifikante Unterschiede zwischen den zwei Behandlungsarten eruiert werden, es zeigten sich signifikant mehr Doppel- bzw. Mehrfachdiagnosen (66.7% vs. 46.1%; $\chi^2(1)=9.454$, $p=.002$) und signifikant mehr Z73-Diagnosen (63.3% vs. 10.0%; $\chi^2(1)=135.187$, $p<.001$) bei Klientinnen und Klienten der Kombinationsbehandlung.

Die Frage, wie lange die Erstdiagnose zurückliegt, wurde sehr unterschiedlich beantwortet. Die Minimaldauer lag bei einem Monat, die Maximaldauer bei 450 Monaten, der Median bei neun Monaten. Ähnlich inhomogen war das Antwortmuster bezüglich der Latenz vom Wahrnehmen der ersten Symptome bis zur Stellung der Erstdiagnose. Der Median lag bei 12 Monaten (Spannweite 0–600). Bezüglich dieser beiden Zeitmaße ließen sich keine Unterschiede zwischen den beiden Behandlungsarten nachweisen ($p>.05$).

Zu Variablen, die die gesundheitlichen Belastungen der Behandlungspopulation vor Beginn der Behandlung charakterisieren, lassen sich folgende Befunde nennen: Von den Klientinnen und Klienten, die zur Zeit des Beginns der Behandlung in einem aufrechten Arbeitsverhältnis waren, gaben 64.3% an, mindestens einmal in den letzten drei Monaten im Krankenstand gewesen zu sein. Insgesamt 614 aller Befragten (56.9%) antworteten, Medikamente wegen psychischer Beschwerden einzunehmen, 69.1% hatten im Zeitraum der letzten drei Monate mindestens einmal einen Arzt aufgesucht (Spannweite: 1 bis 30 Besuche), 60.7% ließen in diesem Zeitraum medizinische Untersuchungen vornehmen (Spannweite: 1 bis 20 Untersuchungen), und 8.0% hatten in den zurückliegenden drei Monaten Krankenhausaufenthalte (Spannweite: 1 bis 4 Aufenthalte). Die Auswirkungen der psychischen Beschwerden auf die täglichen Aktivitäten, die auf einer Skala von 0 (kein Einfluss) bis 10 (maximaler Einfluss) abgefragt wurden, zeigten bei Beginn der Behandlung einen mittleren Wert von 6.22 (SD=2.34) für die Gesamtstichprobe. Vergleicht man diese abgefragten Maße (Krankenstände, Medikamenteneinnahme, Arztbesuche, medizinische Untersuchungen, Krankenhausaufenthalte) zwischen den Behandlungsarten, so konnten keine signifikanten Unterschiede festgestellt werden ($p>.05$).

Durch Vorgabe der Fragebogenverfahren Beck-Depressions-Inventar (BDI-II) sowie des Brief Symptom Inventory (BSI-53) wurde die Stärke der psychischen Beeinträchtigung vor Beginn der Erhebung abgefragt. Bezüglich Stärke der Depression zeigte sich für die Gesamtstichprobe ein durchschnittlicher Summenscore von 27.76 (SD=11.42, Range 0–57), der als knapp unter dem Schwellenwert für eine schwere depressive Verstimmung interpretiert werden kann. Der Globalwert des BSI (GSI) erbrachte mit einem mittleren *T*-Wert von 71.49 (SD=10.22, Range 35–80) für die Gesamtstichprobe zu Beginn der Behandlung einen Score, der mehr als zwei Standardabweichungen vom theoretischen Mittelwert von 50 abweicht. Mittels Online-Erhebung bewerteten die Behandlerinnen und Behandler die Funktionsfähigkeit ihrer Klientinnen und Klienten mittels GAF-Skala, die einen Mittelwert von 54.20 (SD=10.52, Range 27–91) erbrachte, ein Wert, der laut Manual „moderaten Funktionsschwierigkeiten“ entspricht. Die Gruppenvergleiche zeigten in einem der drei klinischen Instrumente einen statistisch signifikanten Unterschied zwischen den zwei Behandlungsarten. In der Einschätzung der globalen Funktionsfähigkeit durch die Behandlerinnen und Behandler wurden Klientinnen und Klienten der Einzelbehandlung schlechter bewertet ($t(869)=-2.670, p=.008, d=0.36$).

5. Ergebnisse zur Veränderungsmessung sowie zur Behandlungszufriedenheit

Die Steuernden Psychologinnen und Psychologen wiesen den Behandlerinnen und Behandlern für die Klientinnen und Klienten der Einzelbehandlung ein Kontingent von jeweils bis zu 30 Behandlungseinheiten und für diejenigen der Kombinationsbehandlung ein Kontingent von jeweils 18 Behandlungseinheiten zu, ergänzt um 12 Einheiten Kunsttherapie. Je nach Behandlungsstand konnte die Standardform jeweils flexibel angepasst werden. Die Quote der Abbrüche lag bei der klinisch-psychologischen Behandlung insgesamt bei 24.6%, bei den Einzelbehandlungen mit 25.6% aber signifikant höher als bei der Kombinationsbehandlung mit 12.8% ($\chi^2(1)=7.044, p=.008$). Die insgesamt durchschnittlich verbrauchte Anzahl an Behandlungseinheiten lag bei 23.20 (SD=9.12,

Range 1–46), die statistische Analyse zeigte dabei einen bedeutsamen Unterschied zwischen den zwei Behandlungsarten (22.82 Einheiten (SD=9.06) für die Einzelbehandlungen, 29.32 Einheiten (SD=7.95) für die Kombinationsbehandlung; $t(76.048)=-6.387, p<.001, d=0.81$).

5.1. Veränderungen in der Inanspruchnahme von Gesundheitsleistungen und im Gesundheitsverhalten

Bezüglich der verschiedenen Formen der Inanspruchnahme von Gesundheitsleistungen wurden im Rahmen der Abschlusserhebung massive Änderungen berichtet, wobei in allen vier Variablen (Arztbesuche, medizinische Untersuchungen, Krankenhausaufenthalte, Medikamenteneinnahme) positive Effekte zu verzeichnen waren. Tabelle 1 zeigt die Änderungen in Richtung eines niedrigeren Bedarfs an Gesundheitsleistungen für die Gesamtstichprobe kompakt zusammengefasst im Überblick. Unterschiedlich hohe Fallzahlen ergeben sich wiederum durch unvollständige Beantwortung einzelner Items zum Fragenkomplex. In keiner der vier Variablen konnten signifikante Unterschiede zwischen den Behandlungsarten gefunden werden ($p>.05$).

Die Frage nach generellen Veränderungen im Gesundheitsverhalten während der Zeit der Behandlung beantwortete der Großteil der befragten Klientinnen und Klienten (94.8%) als positiv. Tabelle 2 zeigt die Bewertungen für die einzelnen Bereiche Sport, Bewegung, Ernährung und Rauchverhalten im Überblick. Auch hier überwiegen die positiven Bewertungen in den abgefragten Teilbereichen, insbesondere bei Sport und Bewegungsverhalten. Beim Rauchverhalten zeigt die Kategorie „keine Änderung“ die größten Häufigkeiten, bedingt auch dadurch, dass ein Großteil der Klientinnen und Klienten nicht rauchte. Etwa drei von vier aller Klientinnen und Klienten (73.3%) berichteten zudem von einer allgemeinen Verbesserung ihrer Leistungsfähigkeit durch die Behandlung. Auch in den Variablen zum Gesundheitsverhalten konnten keine signifikanten Unterschiede zwischen den Behandlungsarten belegt werden ($p>.05$).

Betrachtet man den Bereich Arbeit und Wiedereingliederung in den Arbeitsprozess,

so zeigte sich zu Beginn der Behandlung ein ungleiches Verhältnis von in Arbeit stehenden und arbeitslosen Klientinnen und Klienten im Verhältnis von etwa 2:1. Mittels McNemar-Test wurden die „Wechsler“ von ohne Arbeit vor der Behandlung in Richtung Arbeit nach der Behandlung denjenigen gegenübergestellt, deren Wechsel sich in die gegenläufige Richtung vollzog. Der Vergleich von Ersttestung und Folgetestung zeigte, dass 108 vorher arbeitslose Klientinnen und Klienten zwischenzeitlich ein neues Arbeitsverhältnis aufgenommen hatten, dem standen 81 Klientinnen und Klienten gegenüber, die in der Zwischenzeit ihren Arbeitsplatz verloren hatten. Bei der Gesamtstichprobe wurde mit einem $p=.058$ ein signifikantes Ergebnis knapp verfehlt. Nach Aufteilung nach den beiden Behandlungsarten zeigte sich für die Einzelbehandlung ein Verhältnis von „positiven“ und „negativen“ Wechslern von 107:76 ($p=.026$), während das Verhältnis für die Kombinationsbehandlung mit 1:5 invertiert, aber nicht signifikant ($p=.219$) ausfiel.

Abschließend wurden die Klientinnen und Klienten nach einer Gesamtbeurteilung der Behandlung befragt. Der Großteil ($n=849$, 97.9%) bewertete die Behandlung als sehr bzw. eher hilfreich, lediglich 18 Klientinnen und Klienten (2.1%) betrachteten die Behandlung als nicht hilfreich. Zwischen den zwei Behandlungsarten konnten diesbezüglich keine Unterschiede erhoben werden ($\chi^2(3)=1.427$, $p=.699$).

5.2. Veränderungen bezüglich der validierten klinischen Fragebögen

Im BDI-II zeigte sich für die Klientinnen und Klienten der Einzelbehandlung eine Reduktion der Depressionswerte um 12.63 Punkte ($SD=11.53$) und für diejenigen der Kombinationsbehandlung eine Reduktion um 12.83 Punkte ($SD=11.01$). Der Behandlungseffekt ($F(1, 844)=217.427$, $p<.001$, $\eta^2=.205$) war hoch signifikant, während Gruppeneffekt ($F(1, 844)=3.138$, $p=.077$, $\eta^2=.004$) und die Wechselwirkung aus Behandlungsdauer und Gruppe ($F(1, 844)=0.014$, $p=.907$, $\eta^2=.000$) sich als insignifikant erwiesen (siehe Tabelle 3).

In der GAF-Skala kam es für die Einzelbehandlung zu einer Zunahme der Werte um 15.69 Punkte ($SD=11.82$) und für die Kombinationsbehandlung zu einer solchen um 17.22 Punkte ($SD=9.78$). Der statistische Vergleich zeigte einen hoch signifikanten

Effekt für die Dauer der Behandlung ($F(1, 558)=267.111, p<.001, \eta^2=.312$), einen ebenfalls signifikanten Effekt für die Behandlungsart ($F(1, 558)=9.727, p=.002, \eta^2=.016$) sowie eine nicht signifikante Wechselwirkung zwischen Zeit und Behandlungsart ($F(1, 558)=0.576, p=.448, \eta^2=.001$).

Im Global Severity Index des BSI-53 zeigte sich für die Einzelbehandlung eine Reduktion des T -Werts um 9.59 Punkte ($SD=12.37$) und für die Kombinationsbehandlung eine um 10.38 Punkte ($SD=12.30$). Der statistische Vergleich erbrachte einen hoch signifikanten Effekt der Dauer der Behandlung ($F(1, 831)=115.616, p<.001, \eta^2=.122$), einen nicht signifikanten Effekt für die Behandlungsart ($F(1, 831)=1.033, p=.310, \eta^2=.001$) sowie eine nicht signifikante Wechselwirkung zwischen Zeit und Behandlungsart ($F(1, 831)=0.184, p=.668, \eta^2=.000$).

Um diese Ergebnisse für die einzelnen Subskalen des BSI-53 zu verdeutlichen, sind in Abbildung 4 alle T -Werte ($M=50, SD=10$) aller neun Subskalen und des Globalwerts für beide Behandlungsarten dargestellt. Alle gemittelten T -Werte erreichten oder überschritten vor der Behandlung den Cutoff-Wert für Behandlungsbedürftigkeit. Abbildung 5 zeigt die durchschnittlichen Verbesserungen (Differenz aus t_1 und t_2 , daher die positive Ausrichtung) in jeder der einzelnen neun Subskalen und im Gesamtwert GSI im direkten Vergleich zwischen Einzel- und Kombinationsbehandlung. Abbildung 6 stellt das Niveau der zwei Behandlungsarten in den Skalen nach Abschluss der Intervention dar. Für jede einzelne Skala ist in der aktuellen Analyse zu erkennen, dass die gemittelten T -Werte nach der abgeschlossenen Behandlung unter dem Cutoff-Wert für Behandlungsbedürftigkeit liegen.

5.3. Bewertung der Behandlungszufriedenheit mittels HAQ

Mit der Selbstbeurteilungsform des HAQ wurde die Zufriedenheit mit der Behandlung in den zwei Subskalen Beziehungs- und Erfolgswufriedenheit sowie der Gesamtzufriedenheit zwischen Klientinnen und Klienten der zwei Behandlungsarten verglichen. Sowohl diejenigen der Einzelbehandlung als auch jene der

Kombinationsbehandlung zeigten dabei Ergebnisse am höheren Ende bei Beziehungs- und Erfolgszufriedenheit wie auch bei der Gesamtzufriedenheit mit der Therapie ($p > .05$). Behandlerseitig zeigten die Klinischen Psychologinnen und Psychologen der Kombinationsbehandlung ein höheres Maß an Zufriedenheit in allen drei Skalen des Instruments als diejenigen der Einzelbehandlung (siehe Tabelle 4).

5.4. Direkter Vergleich der Bewertungen der Behandlungsarten

Nach Abschluss der beiden Teile der Kombinationsbehandlung wurden den Klientinnen und Klienten jeweils drei Fragen in gleichlautender Formulierung gestellt, die Aufschluss über deren Bewertung geben können. Auf die Frage nach den Auswirkungen der jeweiligen Behandlungsart gaben 80.6% (klinisch-psychologische Behandlung) bzw. 77.4% (Kunsttherapie) an, dass sich eine allgemeine Verbesserung eingestellt hat ($z = -1.000$, $p = .317$). Die Auswirkungen der psychischen Beschwerden auf die täglichen Aktivitäten wurden nach Abschluss der jeweiligen Behandlungsart von den Befragten ($n = 31$) gleich bewertet, es zeigten sich diesbezüglich keine Unterschiede ($t(30) = -1.196$, $p = .241$). Die Frage, wie hilfreich die jeweilige Behandlungsart für die Bewältigung der Symptome war, wurde für die klinisch-psychologische Behandlung von insgesamt 58.1% (klinisch-psychologische Behandlung) bzw. 83.9% (Kunsttherapie) als „sehr hilfreich“ beschrieben, der Unterschied erreichte aber nicht Signifikanzniveau ($z = -1.807$, $p = .071$).

6. Dropout-Analyse

Eine wichtige Frage, die sich im Anschluss an die Darstellung der Ergebnisse stellte, war die Frage nach der Repräsentativität der Ergebnisse. Warum haben Klientinnen und Klienten, die an der Ersttestung zu Beginn der Erhebung teilgenommen und die Fragen der Testbatterie beantwortet haben, dies nach Abschluss der Behandlung nicht mehr getan? Da im Rahmen dieses Projektes keine Nachbefragung möglich war, lässt sich die Frage nach der Motivation, an der Evaluation des Projektes mitzuarbeiten, nicht direkt beantworten. Es ist jedoch möglich, auf Basis der Daten

aus der Erstbefragung Variablen zu extrahieren, die es ermöglichen, einen Überblick über spezifische Charakteristika der Klientinnen und Klienten zu bekommen. So können Prädiktoren ermittelt werden, die es zulassen vorherzusagen, wer sich an der Abschlusserhebung nicht mehr beteiligt hat, und zwar mittels binär-logistischer Regressionsanalyse (Backhaus, Erichson, Plinke & Weiber, 2000).

An Variablen, in denen sich Completer (Klientinnen und Klienten, die sich sowohl an der Erst- wie auch der Abschlusserhebung der Evaluation beteiligt haben) von Dropouts (Klientinnen und Klienten, die nur an der Ersterhebung der Evaluation teilgenommen haben) unterschieden haben, konnten in dieser Studie aber lediglich eine gefunden werden:

Erhaltene Behandlungseinheiten (Completer: $M=25.15$, $SD=7.92$, Dropouts: $M=15.03$, $SD=9.31$; $t(290.808)=14.682$, $p<.001$, Cohen's $d=1.12$)

Dieses Ergebnis weist darauf hin, dass Klientinnen und Klienten, die im Rahmen der Behandlung ein hohes Maß an Einheiten erhalten hatten, mit höherer Wahrscheinlichkeit auch nach Abschluss der Behandlung die Testbatterie bearbeitet und an das Evaluationsteam retourniert haben. Ein Bias auf Basis soziodemographischer, arbeits- oder krankheitsbezogener Variable konnte hingegen nicht belegt werden.

7. Zusammenfassung

Für die aktuelle Evaluation langten im Rahmen der klinisch-psychologischen Behandlung insgesamt 1219 vollständige Datensets der Ersttestung (Beginn der Behandlung) ein. Nach Ausschluss der pandemiebedingten Abbrüche der Gruppenbehandlungen standen für den statistischen Vergleich von Einzel- (1046 Datensets) und Kombinationsbehandlung mit Kunsttherapie (66 Datensets) insgesamt 1112 Datensätze der Ersttestung zur Verfügung. Die Completerrate war diesmal aufgrund von Veränderungen in der Zustellung der Fragebogensets vergleichsweise hoch (80.8%), insgesamt konnten 898 Datensätze im Rahmen der Veränderungsmessung verarbeitet werden (846 (80.9%) für die Einzelbehandlung, 52

(78.8%) für die Kombinationsbehandlung). Die Quote der Online-bezogenen Fremdbeurteilungen der Klinischen Psychologinnen und Psychologen übergreifend für Beginn und Abschluss der Behandlung lag bei 53.3% (557 Datensets (53.3%) für die Einzelbehandlung, 36 Datensets (54.5%) für die Kombinationsbehandlung). 36 Klientinnen und Klienten, die im Rahmen der Kombinationsbehandlung beide Arten der Behandlung abgeschlossen hatten, füllten beide Fragebogenbatterien aus und bildeten damit die Basis für den direkten statistischen Vergleich.

Im Mittelpunkt der Evaluierung stand diesmal die Frage, ob Interventionen, die in Form von Einzel- bzw. Kombinationsbehandlung mit Kunsttherapie im Rahmen des Projekts fit2work angeboten wurden, das Befinden der Klientinnen und Klienten insgesamt signifikant verbessern können. Operationalisiert wurde „Befinden“ im Rahmen dieses Projekts durch die Testverfahren BDI-II (Depression), BSI-53 (psychische Beeinträchtigung) sowie durch die GAF-Skala (Funktionsfähigkeit), beim BSI durch einen globalen Summenwert (GSI) bzw. durch die Ergebnisse der insgesamt neun Subskalen des Tests. Zusätzlich wurden als Outcome-Variablen die Items des Fragebogens zur Erfassung des krankheitsspezifischen Status miteinbezogen, die Fragen zur Arbeitsfähigkeit und zum Gesundheitszustand und deren Veränderungen zwischen den beiden Erhebungszeitpunkten erhoben haben. Im Rahmen der Abschlusstestung wurde auch nach der Zufriedenheit mit dem Interventionsprozess gefragt, befragt wurden sowohl die Klientinnen und Klienten als auch die Behandlerinnen und Behandler.

Die Signifikanzprüfung zwischen den Behandlungsarten musste bei den vorliegenden Fragebogendaten durch den direkten Vergleich der Testdaten zu den beiden Zeitpunkten erfolgen, da für die Evaluierung keine unbehandelte Kontrollgruppe zur Verfügung stand. Die klinischen validierten Fragebogenverfahren weisen aber Werte für die Einstufung nach „klinisch relevanter Ausprägung“ auf (entweder in Form von Cutoff- oder von *T*-Werten), sodass das Evaluationsergebnis im zeitlichen Vergleich zusätzlich auch auf Basis des Kriteriums Behandlungsbedürftigkeit dargestellt werden kann. Mit Hilfe des Allgemeinen Linearen Modells (ALM) für Messwiederholung war

es möglich, gleichzeitig die Haupteffekte Zeit (Vergleich zwischen Erst- und Abschlusstestung) und Gruppe (Vergleich der Behandlungsarten) bei Betrachtung etwaiger Wechselwirkungen zwischen diesen beiden Haupteffekten zu prüfen.

Die Zuordnung der Klientinnen und Klienten zur Kombinationsbehandlung mit Kunsttherapie erfolgte nach folgenden Vorgaben: Im Gegensatz zur Einzelbehandlung sollten diese zum Zeitpunkt der Zuordnung in einem regulären Arbeitsverhältnis stehen, im Rahmen ihrer Beschäftigung aber mit besonderen Belastungen konfrontiert und daher Burnout-gefährdet sein. Ziel im Rahmen der Kombinationsbehandlung sollte sein, in dieser Situation spezifische Unterstützung zu bieten. In der empirischen Analyse unterschieden sich die beiden Gruppen dementsprechend in den Variablen aktuelles Arbeitsverhältnis ($p < .001$) und in der Quote der Z73-Diagnosen ($p < .001$) in der vorgegebenen Richtung. Zusätzlich zeigten sich aber auch signifikante Unterschiede bezüglich der soziodemographischen Variablen Geschlecht und Ausbildungsniveau. Die Kombinationsbehandlung mit Kunsttherapie zeigte eine höhere Frauenquote ($p = .013$) sowie einen höheren Akademikeranteil ($p = .018$). Neben dem höheren Anteil an Burnout-Diagnosen zeigten Klientinnen und Klienten der Kombinationsbehandlung auch eine signifikant höhere Quote an Doppel- und Mehrfachdiagnosen ($p = .002$), wurden aber von den Behandlerinnen und Behandlern in der GAF-Skala signifikant funktionsfähiger eingestuft ($p = .008$). In Bezug auf alle weiteren gesundheitsbezogenen Variablen konnten hingegen keine Unterschiede zwischen den beiden Behandlungsarten festgestellt werden, auch die beiden Fragebögen zu Depression und psychischer Belastung/Beeinträchtigung zeigten keine Unterschiede. Damit kann festgestellt werden, dass es sich bei Klientinnen und Klienten, die im Rahmen von fit2work der Kombinationsbehandlung mit Kunsttherapie zugeteilt wurden, um eine grundsätzliche andere Ausgangspopulation handelte als diejenige, die eine Einzelbehandlung erhielten, von einer Warte der psychischen Beeinträchtigungen aber eine durchaus miteinander vergleichbare Ausgangssituation aufwiesen.

Für die Klientinnen und Klienten der Einzelbehandlung stand in der Standardform ein

Kontingent von jeweils bis zu 30 Behandlungseinheiten klinisch-psychologische Behandlung zur Verfügung, für diejenigen der Kombinationsbehandlung eines von jeweils 18 Behandlungseinheiten, ergänzt um 12 Einheiten Kunsttherapie, die je nach Notwendigkeit aber flexibel angepasst werden konnten. Die insgesamt durchschnittlich absolvierte Anzahl an Behandlungseinheiten lag bei 23.20 (SD=9.12, Range 1–46), Klientinnen und Klienten der Kombinationsbehandlung bekamen mit zusammengekommen 29.32 Einheiten (SD=7.95) signifikant mehr Behandlungseinheiten als diejenigen der Einzelbehandlung (22.82 Einheiten (SD=9.06; $p<.001$).

Die Veränderungsmessung erbrachte für die klinischen Fragebögen und die GAF-Skala eindeutige und konsistente Ergebnisse, was die Veränderungen durch die Behandlung betrifft; übergreifend waren die erzielten Verbesserungen hoch signifikant mit moderaten bis hohen Effektstärken. Was die Gruppeneffekte betrifft, war lediglich derjenige für die GAF-Skala signifikant, d.h. auf unterschiedlich hohem Ausgangsniveau zeigten sich gleich große Zuwächse für beide Behandlungsarten. Wechselwirkungseffekte konnten bei keiner der drei Analysen belegt werden. Was die Subskalen der psychischen Beeinträchtigungen im BSI-53 betrifft, lagen alle Verbesserungen im Bereich von etwa 5 bis 11 *T*-Wert-Punkten, also Verbesserungen im Bereich von einer halben bis gut einer ganzen Standardabweichung. Alle Mittelwerte der Subskalen, die vor der Intervention über dem Cutoff-Wert ($T=63$) für Behandlungsbedürftigkeit lagen, befanden sich nach der Abschlusserhebung unter diesem Trennwert. Zweitens zeigte sich, dass sich die beiden Behandlungsarten in keiner der Subskalen signifikant voneinander unterschieden, d.h. dass sie gleich gute Ergebnisse erzielten (Haupteffekt Gruppe), und drittens, dass auch die Wechselwirkungen zwischen Zeit und Gruppe (nach Durchführung der Alphafehler-Korrektur nach Bonferroni-Holm) keine signifikanten Ergebnisse aufwiesen.

Als besonderes Ergebnis zeigte sich im Rahmen dieser Evaluation, dass sich infolge der Einzelbehandlungen positive Auswirkungen auf die Beschäftigungssituation zeigten: Die Quote der Wechsler von Klientinnen und Klienten, die vor der

Einzelbehandlung arbeitslos und während der Behandlung in den Arbeitsprozess zurückgekehrt waren, war signifikant größer als diejenige, die den umgekehrten Verlauf zeigte, d.h. von in Arbeit vorher zu arbeitslos danach ($p=.026$). Dieser Erfolg deutet einen ersten Schritt in Richtung des übergeordneten Ziels der beruflichen Reintegration durch die Teilnahme an den Angeboten von fit2work an.

Für das Gesamtkollektiv der Klientinnen und Klienten zeigten sich maßgebliche Veränderungen im Sinne einer umfassenden Reduktion der Inanspruchnahme von Gesundheitsleistungen, wobei insbesondere folgende Punkte zu nennen sind:

- Reduktion von Krankenhausaufenthalten (bei 62%), von Arztbesuchen (bei 52%), von ärztlichen Untersuchungen (bei 49%) und von Einnahme von Psychopharmaka (bei 40%),
- Positive Veränderungen im Bewegungsverhalten (bei 79%), im Sportverhalten (bei 54%) und im Ernährungsverhalten (bei 49%),
- 69% der Klientinnen und Klienten schätzten die erhaltene Behandlung als „sehr hilfreich“ ein, 29% als „eher hilfreich“.

Zusätzlich zu diesen Beurteilungen zeigten die Ergebnisse des Fragebogens Helping Alliance Questionnaire (HAQ), dass sowohl die Beziehungs- als auch die Erfolgszufriedenheit nach Beendigung der Intervention hoch ausgeprägt waren. Auf Seiten der Klientinnen und Klienten zeigten sich in der Zufriedenheit zwischen den beiden Behandlungsarten dabei keine Unterschiede, die Klinischen Psychologinnen und Psychologen stuften sowohl Beziehungs- als auch Erfolgszufriedenheit im Rahmen der Kombinationsbehandlung signifikant besser als bei der Einzelbehandlung ein.

Erfreulich ist abschließend zu bemerken, dass durch eine Veränderung des Zustellungsmodus (alle Fragebogenbatterien werden durch die Behandlerinnen und Behandler gesammelt und dem Evaluationsteam zugeschickt statt wie früher durch die Klientinnen und Klienten selbst) die Quote der Beteiligung an der Abschlussbefragung maßgeblich gesteigert werden konnte. Im Sinne von Limitationen

muss darauf hingewiesen werden, dass nicht alle Klientinnen und Klienten die Behandlungsarten in systematischer Weise hintereinander erhalten haben, sondern pragmatischerweise unterschiedlich, wie es der individuelle Fall erlaubte. Zusätzlich dauerten die Behandlungen durch die von der Pandemie erzwungenen Unterbrechungen im Vergleich mit früheren Evaluationen im Schnitt länger, außerdem war nicht ersichtlich, welche Behandlungen in Präsenz und welche mittels digitaler Kommunikationstechnik abgehalten wurden.

Danksagung. Dank gebührt Alina Dorsch, BSc BFA für die Layoutierung des Textes sowie Dr. Dorothea König für die Programmierung der Online-Erhebung.

Literatur

- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2000). *Multivariate Analysemethoden – eine anwendungsorientierte Einführung* (9., überarbeitete und erweiterte Auflage). Berlin: Springer.
- Bassler, M., Potratz, B. & Krauthauser, H. (1995). Der “Helping Alliance Questionnaire” (HAQ) von Luborsky. *Psychotherapeut*, 40, 23–32.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler* (4. Auflage). Berlin: Springer.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Franke, G. H. (2000). *BSI: Brief Symptom Inventory von L.R. Derogatis (Kurzform der SCL-90-R) – deutsche Version: Manual*. Göttingen: Beltz Test.
- Hautzinger, M., Keller, F. & Kühner, C. (2006). *BDI-II; Beck Depressions-Inventar Revision*. Frankfurt am Main: Harcourt Test Services.
- Jagsch, R. (2015). *Endbericht – Evaluierung des arbeitsbezogenen Therapieerfolgs der klinisch-psychologischen und psychotherapeutischen Behandlung im Rahmen des Projekts „fit2work“*. Unveröffentlichter Forschungsbericht.
- Jagsch, R. (2018). *Endbericht – Evaluierung der Einzelbehandlungen im Rahmen des Projekts fit2work klinisch-psychologische und psychotherapeutische Behandlung 3. Förderjahr*. Unveröffentlichter Forschungsbericht.
- Jagsch, R. (2020). *Endbericht – Evaluierung Einzel- versus Gruppenbehandlungen fit2work klinisch-psychologische und psychotherapeutische Behandlung im 4. Förderjahr*. Unveröffentlichter Forschungsbericht.
- Saß, H., Wittchen, H.-U., Zaudig, M. & Houben, I. (2003). *Diagnostische Kriterien des Diagnostischen und Statistischen Manuals psychischer Störungen. DSM-IV-TR*. Göttingen: Hogrefe.

Anhang A. Abbildungen

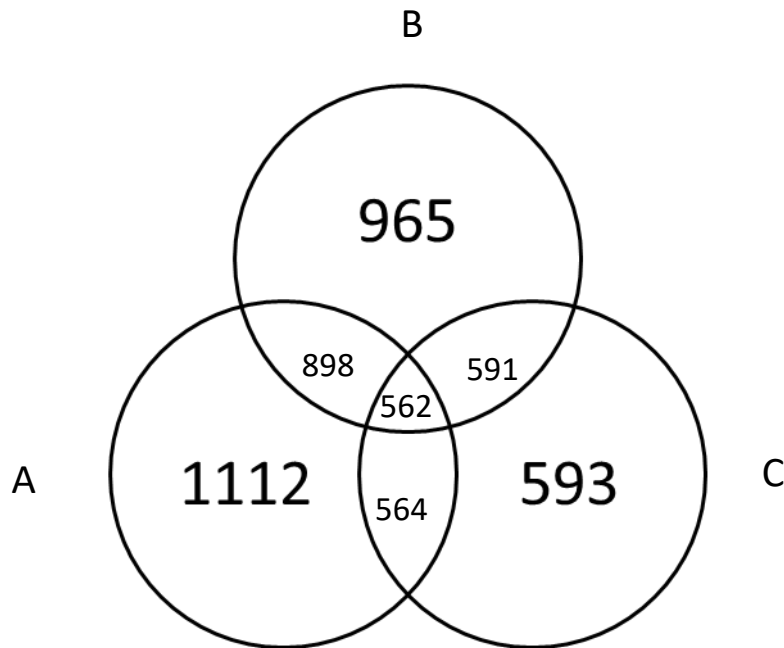


Abbildung 1. Venn-Diagramm zu den vorliegenden Datensätzen der beiden Teilstichproben (Einzel- und Kombinationsbehandlungen).

Anmerkung. A Fragebogensets der Klientinnen und Klienten zu Behandlungsbeginn, B Fragebogensets der Klientinnen und Klienten zu Behandlungsende, C Online-Fragebogensets der Behandlerinnen und Behandler zu Behandlungsbeginn und -ende

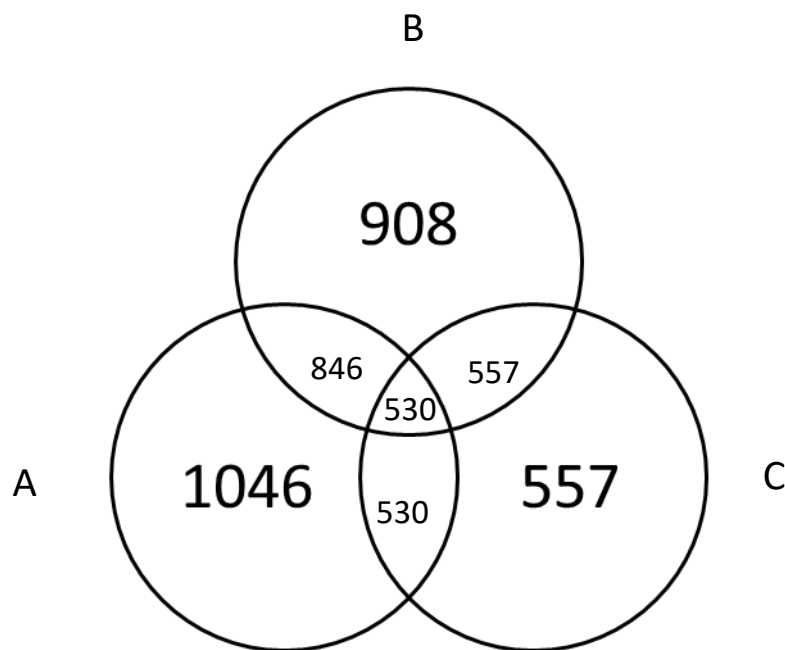


Abbildung 2. Venn-Diagramm zu den vorliegenden Datensätzen der Einzelbehandlung.
Anmerkung. A Fragebogensets der Klientinnen und Klienten zu Behandlungsbeginn, B Fragebogensets der Klientinnen und Klienten zu Behandlungsende, C Online-Fragebogensets der Behandlerinnen und Behandler zu Behandlungsbeginn und -ende

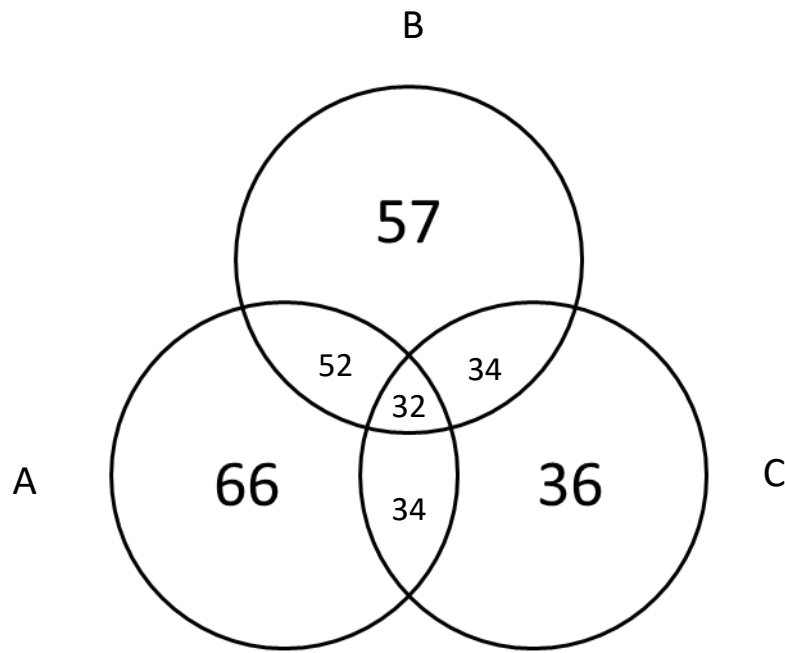


Abbildung 3. Venn-Diagramm zu den vorliegenden Datensätzen der Kombinationsbehandlung.

Anmerkung. A Fragebogensets der Klientinnen und Klienten zu Behandlungsbeginn, B Fragebogensets der Klientinnen und Klienten zu Behandlungsende, C Online-Fragebogensets der Behandlerinnen und Behandler zu Behandlungsbeginn und -ende

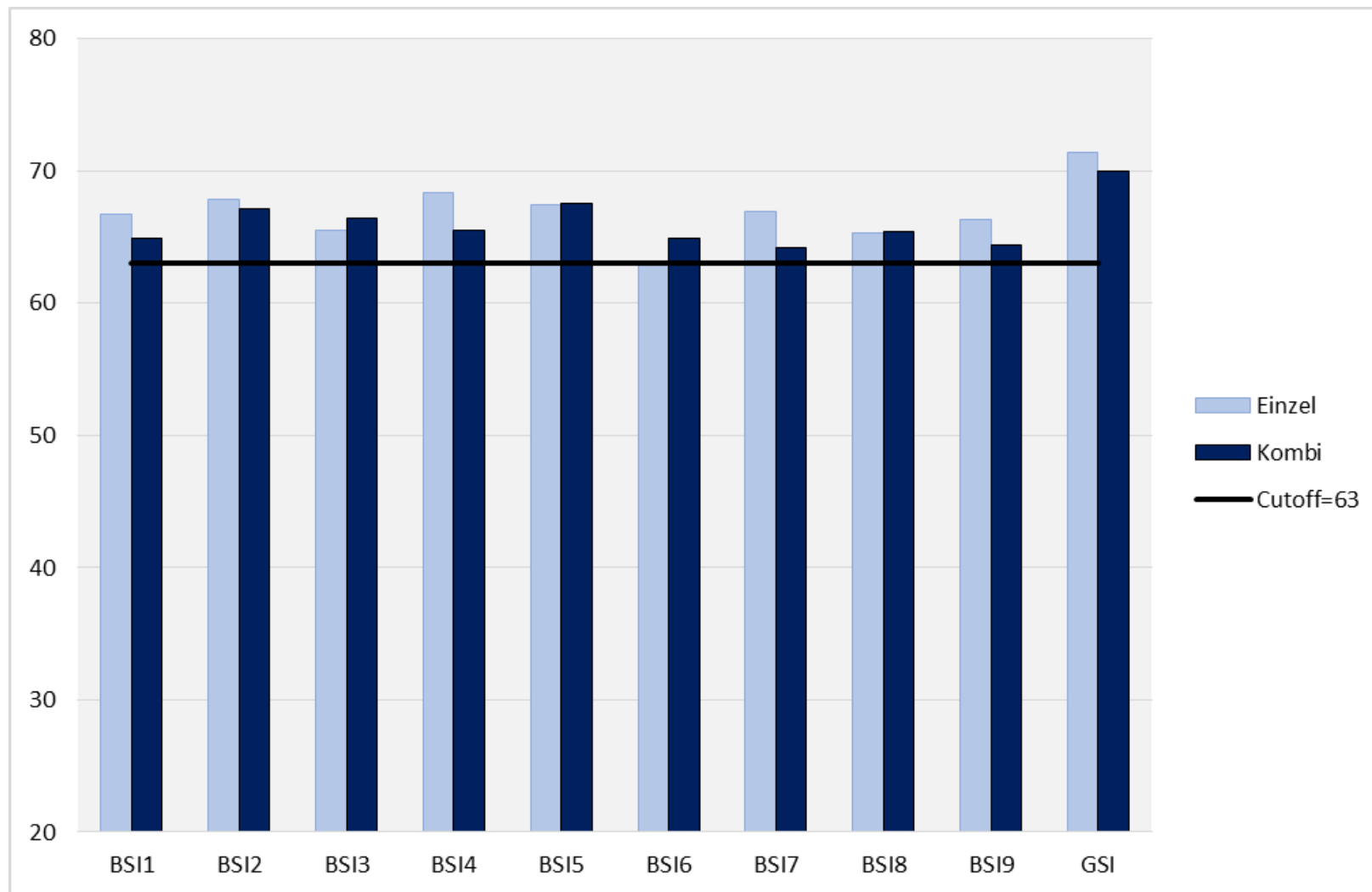


Abbildung 4. Vergleich der zwei Behandlungsarten im Brief Symptom Inventory (BSI) zu Behandlungsbeginn. Entsprechend den Angaben des BSI-Manuals weist ein T-Wert ab 63 (waagrechte Linie) auf Behandlungsbedürftigkeit hin.

Anmerkung. BSI1=Somatisierung, BSI2=Zwanghaftigkeit, BSI3=Unsicherheit im Sozialkontakt, BSI4=Depressivität, BSI5=Ängstlichkeit, BSI6=Aggressivität/Feindseligkeit, BSI7=Phobische Angst, BSI8=Paranoides Denken, BSI9=Psychotizismus, GSI=Globalwert (Global Severity Index)

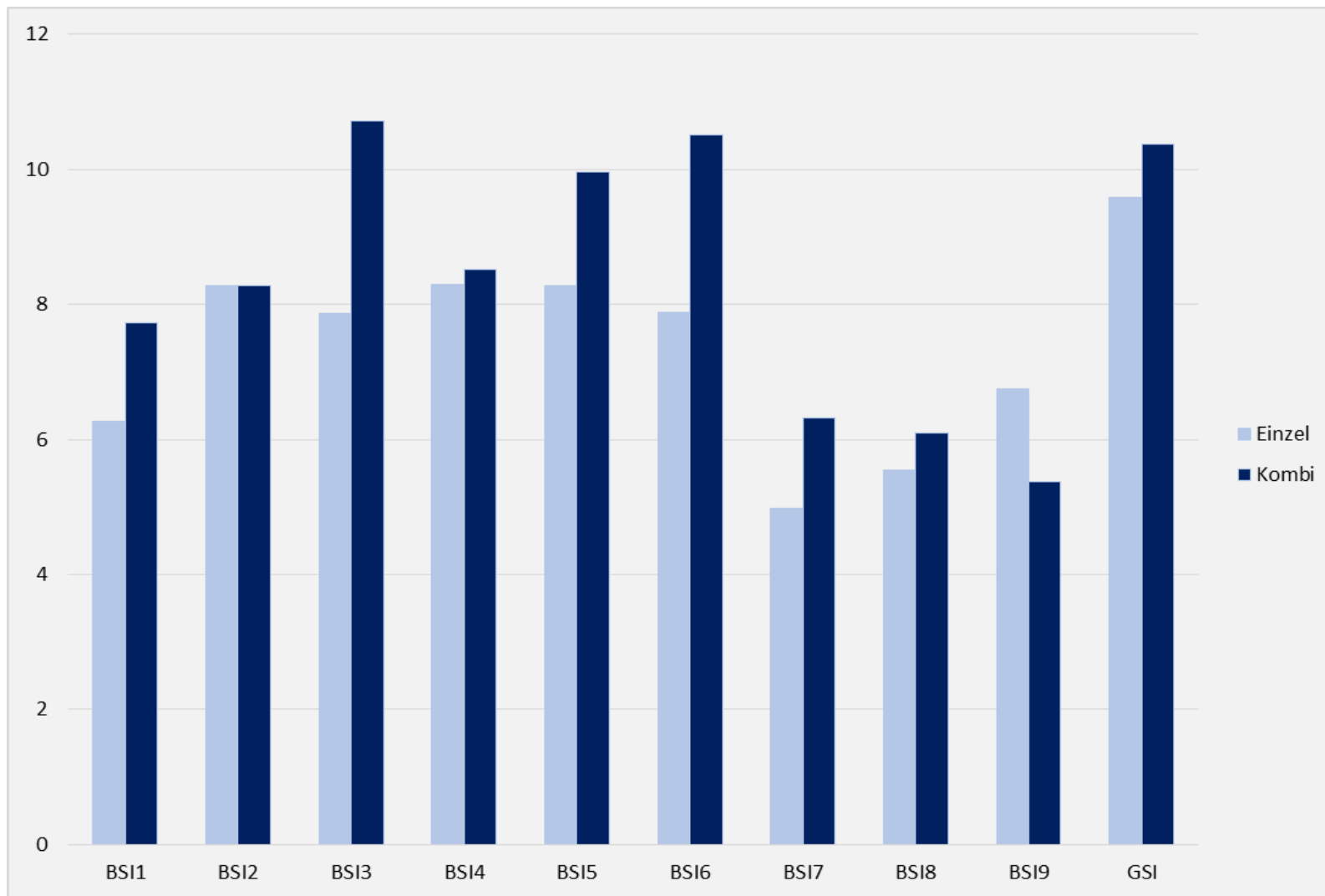


Abbildung 5. Vergleich der Verbesserungen (Differenzen in *T*-Werten) im Brief Symptom Inventory (BSI) zwischen den zwei Behandlungsarten.
Anmerkung. BSI1=Somatisierung, BSI2=Zwanghaftigkeit, BSI3=Unsicherheit im Sozialkontakt, BSI4=Depressivität, BSI5=Ängstlichkeit, BSI6=Aggressivität/Feindseligkeit, BSI7=Phobische Angst, BSI8=Paranoides Denken, BSI9=Psychotizismus, GSI=Globalwert (Global Severity Index)

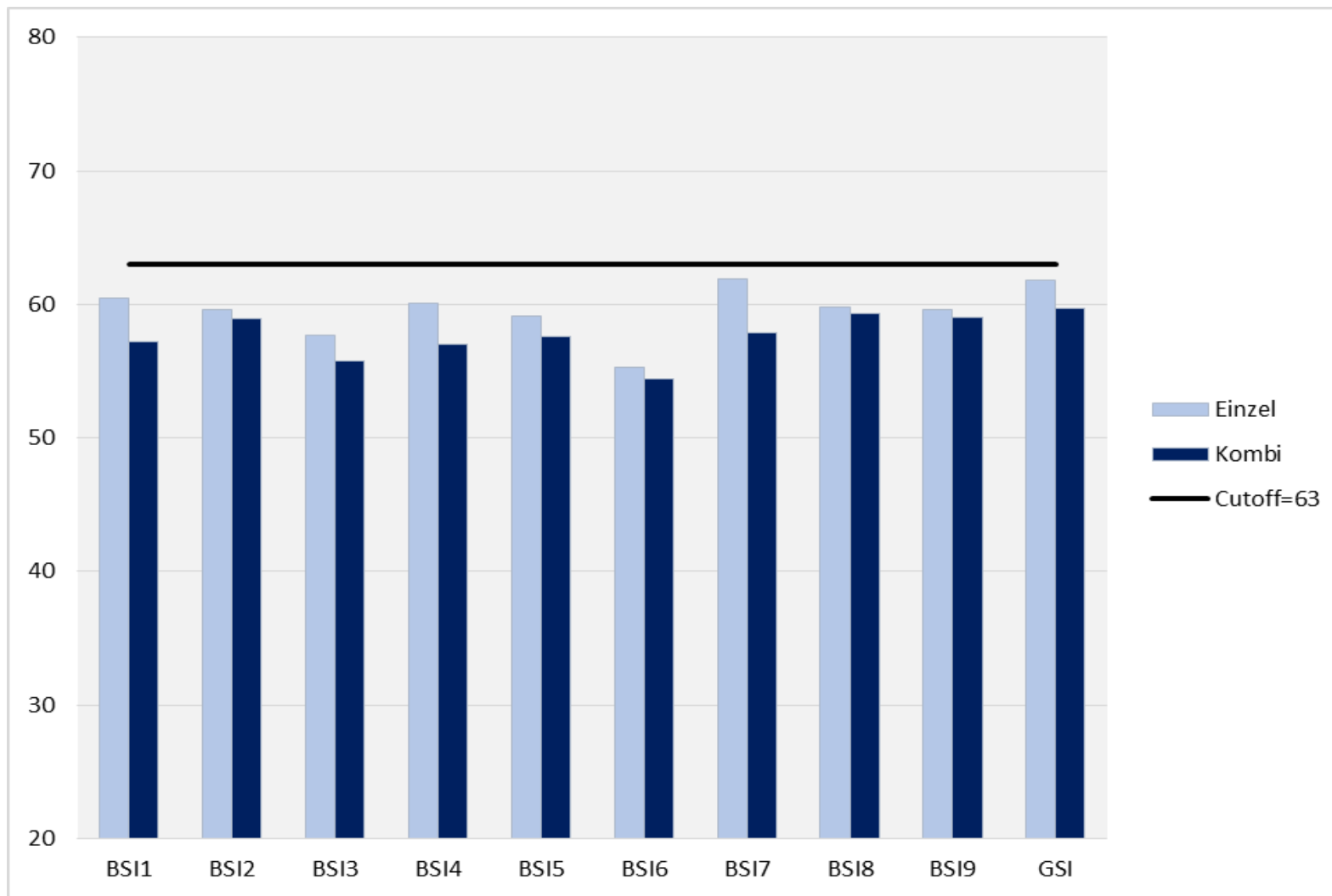


Abbildung 6. Vergleich der zwei Behandlungsarten im Brief Symptom Inventory (BSI) zu Behandlungsende. Entsprechend den Angaben des BSI-Manuals weist ein T-Wert ab 63 (waagrechte Linie) auf Behandlungsbedürftigkeit hin.

Anmerkung. BSI1=Somatisierung, BSI2=Zwanghaftigkeit, BSI3=Unsicherheit im Sozialkontakt, BSI4=Depressivität, BSI5=Ängstlichkeit, BSI6=Aggressivität/Feindseligkeit, BSI7=Phobische Angst, BSI8=Paranoides Denken, BSI9=Psychotizismus, GSI=Globalwert (Global Severity Index)

Anhang B. Tabellen

Tabelle 1. *Absolute und relative Veränderungen in Variablen der Inanspruchnahme von Gesundheitsleistungen*

Art der Änderung n (%)	Ärztliche Untersuchungen	Arztbesuche wegen psychischer Probleme	Krankenhaus-aufenthalte	Medikamenten-einnahme
Abnahme	358 (48.8%)	379 (52.1%)	287 (61.6%)	262 (40.4%)
keine	312 (42.6%)	288 (39.6%)	164 (35.2%)	276 (42.6%)
Zunahme	63 (8.6%)	60 (8.3%)	15 (3.2%)	110 (17.0%)

Tabelle 2. *Absolute und relative Veränderungen im Bereich des Gesundheitsverhaltens*

Art der Änderung n (%)	Sport-verhalten	Bewegungs-verhalten	Ernährungs-verhalten	Rauch-verhalten
positiv	350 (53.5%)	518 (79.2%)	316 (49.2%)	114 (24.2%)
keine	274 (41.8%)	113 (17.3%)	301 (46.9%)	338 (71.8%)
negativ	31 (4.7%)	23 (3.5%)	25 (3.9%)	19 (4.0%)

Tabelle 3: Vergleich der Evaluationsergebnisse zwischen Einzel- und Kombinationsbehandlung vor der und nach Abschluss der Behandlung

	Einzelbehandlung		Kombinationsbehandlung		p_Z	p_B	p_{Z*B}
	$t1$ M (SD)	$t2$ M (SD)	$t1$ M (SD)	$t2$ M (SD)			
BDI-II	27.92 (11.51)	15.29 (12.07)	25.30 (11.43)	12.47 (9.95)	<.001	.077	.907
GAF	53.55 (10.04)	69.25 (13.85)	58.42 (8.95)	75.64 (12.03)	<.001	.002	.448
BSI (T-Werte)							
Somatisierung	66.71 (11.49)	60.44 (12.94)	64.96 (11.86)	57.24 (12.27)	<.001	.106	.385
Zwanghaftigkeit	67.90 (10.54)	59.62 (13.05)	67.19 (10.67)	58.91 (12.45)	<.001	.651	.998
Unsicherheit im Sozialkontakt	65.55 (11.78)	57.68 (12.71)	66.45 (10.39)	55.75 (12.50)	<.001	.755	.124
Depressivität	68.33 (9.92)	60.04 (12.81)	65.57 (11.36)	57.06 (11.95)	<.001	.056	.896
Ängstlichkeit	67.42 (11.58)	59.15 (13.26)	67.57 (10.72)	57.62 (11.79)	<.001	.676	.332
Aggressivität/Feindseligkeit	63.16 (11.41)	55.28 (12.66)	64.91 (9.35)	54.40 (11.50)	<.001	.780	.137
Phobische Angst	66.90 (12.88)	61.93 (13.79)	64.21 (12.42)	57.89 (13.42)	<.001	.059	.461
Paranoides Denken	65.30 (10.84)	59.75 (11.84)	65.40 (9.89)	59.32 (11.87)	<.001	.917	.729
Psychotizismus	66.30 (11.47)	59.56 (12.58)	64.38 (11.62)	59.00 (11.93)	<.001	.437	.424
Global Severity Index	71.38 (10.40)	61.79 (15.07)	70.04 (10.93)	59.66 (14.20)	<.001	.310	.668

Anmerkung. p_Z Zeit, p_B Behandlungsart, p_{Z*B} Interaktion Zeit *Behandlungsart

Tabelle 4. Vergleich der Ergebnisse im Helping Alliance Questionnaire (HAQ) zwischen den zwei Behandlungsarten (Einzel- versus Kombinationsbehandlung) aus Sicht der Klientinnen und Klienten sowie der Klinischen Psychologinnen und Psychologen

	Einzelbehandlung M (SD)	Kombinations- behandlung M (SD)	t	p
HAQ				
Klientenseitig				
Beziehungszufriedenheit	5.60 (0.53)	5.50 (0.62)	1.226	.220
Erfolgswufriedenheit	4.94 (0.84)	4.97 (0.73)	-0.262	.793
Gesamtwufriedenheit	5.30 (0.60)	5.25 (0.60)	0.550	.582
Behandlerseitig				
Beziehungszufriedenheit	5.42 (0.52)	5.64 (0.43)	-2.539	.011
Erfolgswufriedenheit	4.96 (0.79)	5.42 (0.54)	-4.842	<.001
Gesamtwufriedenheit	5.21 (0.60)	5.54 (0.43)	-4.419	<.001
(Wertebereich jeweils von 1=sehr unzufrieden bis 6=sehr zufriede)				