

Praxisstudie zu Katathym Imaginativer Psychotherapie, Hypnosepsychotherapie und Autogener Psychotherapie

Eine naturalistische Untersuchung zu Behandlungstechnik, moderierenden Faktoren und Wirksamkeit – Projektbericht

Christian Sell, Svenja Taubner, Heidi Möller

Einleitung

Die vorliegende Studie wurde von Januar 2012 bis März 2016 vom Institut für Psychologie der Universität Kassel in Kooperation mit der *Österreichischen Gesellschaft für angewandte Tiefenpsychologie und allgemeine Psychotherapie* (ÖGATAP) durchgeführt. Die Projektleitung hatte Frau Prof. Dr. phil. Heidi Möller (Abteilung für Theorie und Methodik der Beratung, Institut für Psychologie, Universität Kassel), die Projektkoordination erfolgte durch Frau Prof. Dr. phil. Svenja Taubner (Institut für Psychosoziale Prävention, Universitätsklinikum Heidelberg). Mit der Durchführung des Projekts war Herr Dr. phil. Christian Sell betraut. Herr Sell war während der Studiendauer Drittmittelbediensteter am Institut für Psychologie der Universität Kassel. Für seine Anstellung hat die Universität Kassel Projektmittel von der ÖGATAP erhalten. Die ÖGATAP hatte keinen Einfluss auf das Studiendesign, die Analysen, die Interpretation der Ergebnisse und die Entscheidung zur Publikation. Die Ethikkommission der Universität Kassel, sowie die lokalen Ethikkommissionen in Innsbruck und Wien beurteilten die Durchführung der Studie als ethisch unbedenklich. Die hier vorzustellenden Ergebnisse wurden zuvor bereits in zwei Artikeln in wissenschaftlichen Fachzeitschriften (peer-reviewed) veröffentlicht. Entsprechende Hinweise auf diese Artikel wurden finden sich an den relevanten Stellen.

Das übergeordnete Ziel der Studie bestand in der Evaluation der therapeutischen Wirksamkeit von drei trance- und imaginationsbasierten tiefenpsychologischen Methoden – Katathym Imaginative Psychotherapie (KIP), Hypnosepsychotherapie (HY) und Autogene Psychotherapie (ATP) – anhand einer Reihe von

unterschiedlichen Ergebnismaßen. Zur Frage der differentiellen Indikation wurde zusätzlich versucht, Moderatoren des Behandlungserfolgs zu identifizieren.

Im Folgenden sollen die drei zu untersuchenden Psychotherapiemethoden zunächst kurz anhand ihrer theoretischen Verortung vorgestellt werden. Dabei wird jeweils der bisherige Stand der Prozess- und Ergebnisforschung zusammengefasst, anschließend werden die konkreten forschungsleitenden Fragestellungen abgeleitet. Nach der Vorstellung von Studiendesign, Messinstrumenten, Einschlusskriterien und den Charakteristika der durchgeführten Behandlungen, werden dann in separaten Sektionen die Ergebnisse zu moderierenden Faktoren und Wirksamkeit berichtet. Abschließend sollen die Ergebnisse insgesamt zusammenfassend betrachtet werden. Dies beinhaltet auch eine Diskussion der Limitationen der Aussagekraft der vorliegenden Studie.

Wir danken den Patientinnen und Patienten, den Therapeutinnen und Therapeuten, sowie dem Vorstand der ÖGATAP für ihre Kooperation im Rahmen der Durchführung dieser Studie. Besonderen Dank möchten wir Frau Mag. Eveline Schöpfer-Mader aussprechen. Sie hat in ihrer Zeit als Vorstandsvorsitzende der ÖGATAP durch ihren großartigen Leitungs- und Organisationseinsatz maßgeblichen Anteil daran gehabt, dass die Zusammenarbeit zwischen den Mitgliedern der ÖGATAP und uns als Forschern so produktiv in Gang gekommen ist. Dadurch ist die Studie in dieser Form überhaupt erst möglich geworden. Bei der Entwicklung des Fragebogens zur Erfassung der therapeutischen Interventionen haben insbesondere folgende Personen einen inhaltlichen Beitrag bei der Erstellung der Items und bei methodenspezifischen Fragestellungen geleistet: Dr. Bernhard Brömmel, Dr. Barbara Burian-Langegger, Dr. Wilfried Dieter, Dr. Magdalena Grünewald, Dr. Doris Hönlgl, Dr. Mathilde Pichler, Dr. Michael Rosner, Dr. Franz Sedlak, Dr. Brigitte Bischof, Dr. Michael Harrer, Dr. Hans Kanitschar, Eva Pollani MSc und Mag. Wolfgang Oswald. Weiterhin danken wir Prof. Dr. Johannes Zimmermann und Prof. Dr. Morten Moshagen für ihre Beratung in Fragen der Datenauswertung, sowie Dr. Johannes C. Ehrenthal für seine hilfreichen Anmerkungen zu verschiedenen Teilen des Textes.

2 Untersuchungsgegenstand & Erkenntnisinteresse

Die Wirksamkeit tiefenpsychologischer Therapien kann für eine Vielzahl von Settings und Störungen als gut belegt angesehen werden (Ehrenthal, Dinger & Nikendei, 2014; Leichsenring & Rabung, 2011; Shedler, 2012), darunter Depression, Angststörungen und Persönlichkeitsstörungen (Barber, Muran, McCarthy & Kneefe, 2013). Allerdings finden sich unter dem Verfahren Tiefenpsychologie eine Reihe von Therapieformen, die sich zwar als »psychoanalytisch begründet« verstehen lassen, welche jedoch behandlungstechnische Besonderheiten aufweisen, deren Bedeutung für das

Therapieergebnis noch weitgehend unverstanden ist. Naturalistische Forschung erscheint als Methodologie der Wahl, um die spezifischen Effekte dieser verschiedenen Varianten psychodynamischer Therapie wie sie in verschiedenen Gesundheitsversorgungssystemen Anwendung finden zu untersuchen. Behandlungstechnische Besonderheiten innerhalb von psychodynamischen Verfahren haben in früheren Studien sehr wenig Aufmerksamkeit erfahren und stellen eine Lücke im bisherigen Wissen über diese Verfahren dar. Im Sinne der empirisch gestützten Weiterentwicklung psychodynamischer Psychotherapie erscheint es bedeutsam, diese Lücke schrittweise zu schließen, damit eigenständige Weiterentwicklungen und Abwandlungen vom bekannten Vorgehen in ihrer spezifischen Bedeutsamkeit erfasst werden können. Zu den genannten Verfahren gehören solche Methoden welche systematisch auf geführte Imagination und Trancen zurückgreifen. Drei solcher Verfahren, die Katathym Imaginative Psychotherapie (KIP), die Hypnosepsychotherapie (HY) und die Autogene Psychotherapie (ATP) sollen hier untersucht werden.

Die spezifischen Auswirkungen der Anwendung von Imaginationen und Trancen auf Prozess und Ergebnis der Therapie in diesen Verfahren sind bislang nur vereinzelt systematisch untersucht worden (Übersicht bei Stigler & Pokorny, 2012). Wir gehen davon aus, dass der Anschluss dieser Verfahren an die gegenwärtigen Paradigmen der Psychotherapieforschung gleichermaßen ein Gewinn für die Verfahren wäre, wie auch eine Bereicherung des Wissens über die Wirkweise tiefenpsychologischer Psychotherapie im Allgemeinen.

2.1 Therapiemethoden und Stand der Forschung

KIP, HY und ATP zählen zu den staatlich anerkannten wissenschaftlich-psychotherapeutischen Methoden in Österreich. Laut dem österreichischen Bundesministerium für Gesundheit sind sie den Methoden mit psychodynamisch-tiefenpsychologischer Orientierung zuzuordnen. Alle drei Verfahren werden in Österreich seit 1969 durch die *Österreichische Gesellschaft für angewandte Tiefenpsychologie und allgemeine Psychotherapie* (ÖGATAP) vertreten und gelehrt. Katathymes Bilderleben und Autogenes Training als therapeutische Techniken werden in der Schweiz und in Deutschland ebenfalls gelehrt und finden stationär wie ambulant Anwendung – in Deutschland in der Regel im Rahmen tiefenpsychologisch fundierter Psychotherapien. Die für die Hypnosepsychotherapie charakteristische Integration von Hypnoanalyse und neueren Ansätzen findet sich in dieser Ausprägung nur in Österreich.

KIP, HY und ATP werden im Rahmen der österreichischen Gesundheitsversorgung sowohl als Kurzzeit- (< 25 Sitzungen), als auch als Langzeittherapie (> 50 Sitzungen) eingesetzt. Die mittlere Behandlungsdauer liegt laut der Selbstauskunft der Fachgesellschaft für alle drei Verfahren bei ca. 50 Sitzungen im Einzelsetting. Die Stundenfrequenz beträgt in der Regel eine Sitzung pro Woche, kann aber im

Bedarfsfall sowohl verringert, als auch auf 2x/Woche gesteigert werden. In der ATP und der KIP werden zudem häufig ambulante Therapiegruppen angeboten (für die KIP wurden Gruppentherapien nicht in die Studie miteinbezogen).

2.1.1 *Katathym Imaginative Psychotherapie (KIP)*

Pichler (2011) definiert die KIP als »eine tiefenpsychologisch fundierte Psychotherapiemethode, bei der Imaginationen einen zentralen Stellenwert im therapeutischen Geschehen einnehmen« (S. 109). Das technische Vorgehen in der Behandlung basiert maßgeblich auf den Arbeiten Hanscarl Leuners (1994) zum Katathymen Bilderleben. Dieter (2001) fasst es wie folgt zusammen: »In einem leichten Entspannungszustand wird der Patient zu Imaginationen (Tagträumen) angeregt, die dem Therapeuten unmittelbar berichtet werden und die zum Gegenstand einer dialogischen Untersuchung, Erforschung und Intervention werden« (S. 15). Eine Behandlung im Rahmen der KIP besteht charakteristischerweise zum einen aus Stunden, in denen Imaginationen zu vorgegebenen Motiven stattfinden (»Erlebnisraum«; Kottje-Birnbacher 1992), und zum anderen aus Stunden, in denen die Imaginationen therapeutisch aufgearbeitet werden sollen (»Verarbeitungsraum«, ebd.). Diese Nachbereitung soll sich an psychodynamischen Grundprinzipien orientieren und die Arbeit mit Übertragung, Gegenübertragung, Abwehr und Widerstand umfassen (Pichler 2011).

Größere kontrolliert-randomisierte (RCT-)Studien gibt es für ambulante KIP-Behandlungen bislang nicht. Es wurden allerdings zwei naturalistische Wirksamkeitsstudien vorgelegt. Wietersheim, Wilke, Röser und Meder (2003) haben in einer deutschen Stichprobe (N = 66) große Effekte für die Verbesserung somatischer Symptome, depressiver Symptome und der Lebenszufriedenheit gefunden, sowie mittlere Effekte für die Verbesserung von Angstsymptomen. Ebenfalls in einer deutschen Stichprobe (N = 25) fanden Sachsse, Imruck und Bahrke (2016) sehr große Effekte für eine Reduktion der Symptombelastung und große Effekte für die Verbesserung interpersoneller Probleme, Depressions- und Angstsymptomen. Weiterhin liegt eine Reihe von Prozess- und Einzelfallstudien vor (z. B. Wilke 1982; Stigler & Pokorny 2001; Übersicht bei Stigler & Pokorny 2012), sowie einzelne Studien für spezifische psychosomatische Syndrome, sowie Essstörungen. Esplen, Garfinkel, Olmsted, Gallop und Kennedy (1998) etwa konnten zeigen, dass in der Behandlung von Bulimia nervosa Kurzzeit-KIP-Behandlungen signifikant effektiver waren als eine wöchentliche Gesprächsgruppe.

2.1.2 *Hypnosepsychotherapie (HY)*

Analog zur KIP beschreibt Kanitschar (2009) die HY als ein »integratives tiefenpsychologisches Verfahren« (S. 1). Dieses wird erläutert, als eine Methode, welche

ebenfalls in der therapeutischen Arbeit psychodynamischen Grundannahmen verpflichtet ist, in der aber dennoch »alle bekannten Anwendungsformen und Techniken psychotherapeutischer Hypnose Platz haben« (ebd.). Es sollen demnach auch solche hypnosetherapeutischen Techniken integriert werden, welche nicht genuin tiefenpsychologisch sind. Der historischen Entwicklung nach fußt das Verfahren auf der tiefenpsychologischen Hypnoanalyse nach Watkins und Fromm (Watkins 1992) und wurde später durch ericksonsche, hypnosystemische und hypnobehaviorale Ansätze (Robertson 2012) ergänzt. In Unterscheidung zur weiter verbreiteten Hypnotherapie nach Milton Erickson (Erickson & Rossi 1981) versucht die HY also Techniken aus verschiedenen hypnosetherapeutischen Traditionen mit einer explizit tiefenpsychologischen Grundhaltung zu verbinden. Die systematische Induktion von und Arbeit mit hypnotischen Trancen beim Patienten – verstanden als entspannte Wachzustände, welche sich u. a. durch eine auf wenige Inhalte fokussierte Aufmerksamkeit und eine erhöhte Suggestibilität auszeichnen – stellt dabei ein zentrales Element dar (Harrer 2008). Nach einer »Anfangsphase«, in der neben der Diagnose- und Indikationsstellung v. a. die hypnotischen Fähigkeiten des Patienten erarbeitet werden, zielt die eigentliche »Arbeitsphase« der Therapie auf drei »Interventionsmodi« ab: die Aktualisierung und den Aufbau von Ressourcen, Konfliktbearbeitung und korrigierende emotionale Erfahrung, sowie Zukunfts- und Lösungsorientierung (Kanitschar 2009).

Der Großteil der bisherigen Forschung zu klinischer Hypnose allgemein stammt aus dem medizinischen Bereich. RCT- und naturalistische Wirksamkeitsstudien in ausreichender Zahl finden sich zu Schmerztherapie und Stressreduktion im Zusammenhang mit medizinischen Eingriffen (Hagl 2013). Hinsichtlich der psychotherapeutischen Behandlung psychischer Störungen zeigte die Metaanalyse von Revenstorf (2006) große bis sehr große Effekte bei Patienten mit Angststörungen, chronischen Schmerzsyndromen und einigen Formen von Abhängigkeitserkrankungen – sowohl auf allgemeinen, als auch auf störungsspezifischen Ergebnismaßen. In einer neueren Metaanalyse berichtet Hagl (2014), dass neuere Hypnoseforschung keine klinischen Studien zu trancebasierter Einzelpsychotherapie umfasste. HY als Methode in der oben beschriebenen Form wurde bislang weder durch naturalistische noch durch RCT-Wirksamkeitsstudien untersucht.

2.1.3 Autogene Psychotherapie (ATP)

Die ATP wird von Sedlak (2012) als tiefenpsychologisch fundierte Methode charakterisiert und zwar insofern, als dass sie »Äußerungen des Unbewussten durch Erinnern, Wiederholen, Durcharbeiten aufgreift« (S. 12). Das Verfahren wird von Bischof (2012) in drei Therapieelementen, namentlich Grundstufe, Mittelstufe und (analytische) Oberstufe, beschrieben. Die Grundstufe basiert dabei auf dem Erlernen und Üben des Autogenen Trainings (Schultz 1930). Auf Basis der

so erworbenen Grundfähigkeit des Patienten zum »seelisch-körperlichen Spannungsausgleich« (Sedlak 2012, S. 12), wird in der Mittelstufe mit der Problemkonfrontation begonnen. Im Fokus stehen die während des Autogenen Trainings oft bildhaft aufsteigenden, problembesetzten Gedanken und Gefühlen. In der analytischen Oberstufe schließlich, sollen diese Inhalte dann nach tiefenpsychologischen Prinzipien benannt und »durchgearbeitet« werden.

Zur ATP als eigenständiger Psychotherapie im oben beschriebenen Sinne liegen bislang keine Wirksamkeitsnachweise vor. Das Autogene Training nach Schultz hingegen, welches die ATP in eine tiefenpsychologische Psychotherapie integriert hat, wurde relativ umfangreich beforscht. Die Ergebnisse zur Wirksamkeit des Autogenen Trainings wurden in einer Metastudie von Linden (1994), in einer Metaanalyse mit neueren RCT-Studien von Stetter und Kupper (2002), sowie einer Aktualisierung letzterer durch Bregenzer (2015) zusammengetragen. Die hier zusammengefassten Studien beziehen sich durchweg auf das Autogene Training als Entspannungsverfahren wie es in ambulanten Gruppen oder als begleitende Therapie in somatischen, psychosomatischen und Rehabilitationseinrichtungen eingesetzt wird. Die ATP im oben beschriebenen Sinne ist bislang nicht im Rahmen größerer prospektiver Gruppenstudien untersucht worden. Zum Autogenen Training berichten die Autoren einheitlich mittlere bis große Effekte bei der Verbesserung des allgemeinen Wohlbefindens, der Stimmung und der kognitiven Leistungsfähigkeit. Weiterhin wurden mittelgroße Effekte bei der symptomatischen Verbesserung somatoformer Störungen, Asthma bronchiale, Spannungskopfschmerz, funktioneller Hypertonie, koronarer Herzerkrankung, Schlafstörungen, Angststörungen, sowie leichter bis mittelgradiger Depressionen gefunden.

2.2 Forschungsfragen und Hypothesen

2.2.1 Moderatoren des Behandlungserfolgs

Zur Frage nach Prädiktoren des Behandlungserfolgs oder besonders geeigneten Gruppen von Patientinnen und Patienten liegt zu KIP, HY und ATP eine Reihe von konzeptuellen, oft auf Einzelfallanalysen basierten Arbeiten vor. Ein prominenter Fokus in allen drei Verfahren ist dabei die These, dass die Arbeit mit Imaginationen und Trancen einen wichtigen Wirkfaktor darstellt. Stigler und Pokorny (2001) konnten anhand einer inhaltsanalytischen Untersuchung des in einer KIP-Therapie verwendeten Vokabulars zeigen, dass während der Imaginationen mehr und deutlich positivere Emotionen aktiviert sind als in den verbalen Sitzungsphasen. Weiterhin erscheinen Therapeutin oder Therapeut sowie Patientin oder Patient während der Imaginationen empathischer und kongruenter auf einander eingestimmt. Diese Arbeit mit Imaginationen oder Trancen wird von einigen Autorinnen und Autoren in engem Zusammenhang mit Prozessen der

Symbolisierung und Mentalisierung gesehen. Pichler (2011) etwa schreibt, dass »[d]urch die Imaginationen [...] eine Umsetzung unbewusster oder vorbewusster Inhalte in ›innere Bilder‹ ermöglicht« wird (S. 113), welche dann in einem weiteren Schritt verstanden und bearbeitet werden können. Dies wird ebenfalls durch empirische Befunde aus der Prozessforschung gestützt: Pokorny und Stigler (2006) konnten in einer späteren Studie anhand einer Clusteranalyse zeigen, dass die Eigenschaften der in der KIP imaginierten Objekte im Zusammenhang stehen mit jenen der realen Beziehungspersonen, über die die Patienten im Verarbeitungsraum sprechen. Diesen Prozess der Symbolisierung von psychischen Inhalten stellt Dieter (2001) in den Kontext der Forschung zur Mentalisierungstheorie. Diese legt nahe, dass die Fähigkeit mit Symbolen zu arbeiten zusammenhängt mit der Fähigkeit und Bereitschaft einer Person, eigene und fremde Gefühle zu erleben und über ihre Bedeutung nachzudenken (Allen, Fonagy, Bateman 2011). Bislang liegen jedoch keine empirischen Befunde vor, inwiefern diese Disposition tatsächlich einen positiven Einfluss auf das Ergebnis von KIP-, HY- oder ATP-Behandlung hat.

Ebenso wenig liegen für HY, KIP oder ATP bislang Daten darüber vor, inwieweit sich das Ausmaß pathologischer Persönlichkeitszüge der Patientinnen und Patienten auf die Erfolgsaussichten der Behandlung auswirkt. Die konzeptuelle Literatur, besonders zu KIP und HY, weist darauf hin, dass die Verfahren ein Behandlungsangebot machen, von dem strukturell gestörte Patientinnen und Patienten spezifisch profitieren können. Die Methoden seien »in besonderer Weise auch für Störungen mit Beeinträchtigung der Ich-Struktur (z. B. Persönlichkeitsstörungen)« geeignet (Dieter 2001, S. 115; für HY vgl. Zindel 2001). Ergebnisstudien hierzu liegen bislang nicht vor. Die Befundlage bei tiefenpsychologischer Therapie von Persönlichkeitsstörungen (Übersicht bei Leichsenring 2013) zeigt übereinstimmend, dass allgemein eine gute Wirksamkeit bei diesen Störungen angenommen werden kann, die Veränderung bei komorbider Persönlichkeitsstörung aber deutlich langsamer erfolgt als bei Patientinnen und Patienten, die nur eine Achse-I-Störung haben (Tyrer & Simmonds 2003).

Vorbereitend für die Testung von Psychological Mindedness und Persönlichkeitspathologie als Moderatoren des Behandlungserfolgs soll zunächst die Hypothese geprüft werden, dass sich die Symptombelastung der Patienten nach einem Jahr Behandlung in KIP, HY oder ATP statistisch signifikant verbessert. Diese Hypothese knüpft an die bisherige Befundlage in Bezug auf die Wirksamkeit von tiefenpsychologischen Verfahren an. Die Reduktion der Symptombelastung innerhalb des ersten Jahres einer Psychotherapie erscheint für den Nachweis moderierender Faktoren geeignet, da sie sich als eine gleichsam universelle – über alle Psychotherapiemethoden hinweg stabil beobachtbare – Verbesserung durch Psychotherapie darstellt (vgl. Ogles 2013).

In den konzeptuellen Arbeiten zu allen drei Verfahren, wie auch in den Ergebnissen der KIP-Prozessforschung, wird als spezifische Besonderheit im

therapeutischen Vorgehen der explizite Fokus auf Prozesse der Symbolisierung und Mentalisierung deutlich. Das Konzept der Psychological Mindedness (Appelbaum 1973) wurde vorgeschlagen, um die Bereitschaft, Gefühle zuzulassen und über eigene wie fremde Gedanken und Gefühle nachzudenken auf einen Begriff zu bringen. Mit Allen, Fonagy und Bateman (2011) verstehen wir Psychological Mindedness als »Mentalisierungsfähigkeit oder [...] Disposition zu mentalisieren« (S. 91). Aufgrund des spezifischen Fokus beider Methoden auf Symbolisierungs- und Mentalisierungsprozesse wird also vorhergesagt, dass Psychological Mindedness in einem positiven Zusammenhang steht mit dem Ausmaß der Symptomreduktion. Zur Frage nach den Auswirkungen von pathologischen Persönlichkeitszügen lassen bisherige Forschungsergebnisse vermuten, dass Patientinnen und Patienten mit starken strukturellen Defiziten deutlich langsamer von der Behandlung profitieren werden. Wir formulieren daher die Hypothese, dass das Ausmaß pathologischer Persönlichkeitszüge – innerhalb der hierfür betrachteten 12 Monate – in negativem Zusammenhang mit der therapeutischen Veränderung steht. Für beide Moderatoren wird angenommen, dass beide Variablen jeweils einen Effekt auf die Symptomveränderung haben, auch, wenn für die jeweils andere Variable kontrolliert wird.

2.2.2 Wirksamkeit und Verlauf der Veränderungen

Die zweite Fragestellung der vorliegenden Studie bezieht sich auf die Erfassung langfristiger therapeutischer Veränderung durch KIP, HY und ATP mittels einer Reihe verschiedener Ergebnismaße; inklusive Selbst- und Fremdbeobachtung. An vorangegangene Forschung zu den Methoden oder ihren Varianten anknüpfend wurden das globale Funktionsniveau, die Symptombelastung und -schwere, sowie die Lebenszufriedenheit als Indikatoren dafür ausgewählt, inwieweit eine psychische Störung sich direkt negativ auf das Wohlbefinden von Patientinnen und Patienten auswirkt. Zudem wurden interpersonelle Probleme und Persönlichkeitsorganisation erfasst. Dies erscheint notwendig, da psychodynamische Methoden zwar zum einen an der Symptomreduktion arbeiten, zum anderen aber auch weiterreichende Ziele wie Persönlichkeitsentwicklung und die Veränderung interpersoneller Muster verfolgen. Den Empfehlungen von Siefert und DeFife (2012) folgend soll dies auch in der Wahl der Ergebnismaße berücksichtigt werden. Da neben der Persönlichkeitspathologie der Patientinnen und Patienten auch ihre Psychological Mindedness als Moderator des Behandlungserfolgs diskutiert wird (s. 2.2.1), wurde letztere auch als Ergebnismaß erfasst, um zu prüfen, ob die therapeutische Arbeit mit KIP, HY und ATP diese Disposition auch verbessern kann. Als gesundheitsökonomische Maße wurden darüber hinaus Krankenstandstage, Arztbesuche und Krankenhausaufenthalte erfasst.

3. Studiendesign & Methoden

3.1 Studiendesign und Ablauf der Erhebung

Die Studie ist eine prospektive, multimodale und multimethodale Längsschnittstudie mit naturalistischem Ein-Gruppen-Design. Die Studienteilnehmer waren Patientinnen und Patienten, welche im Rahmen der österreichischen Gesundheitsversorgung psychotherapeutische Hilfe aufsuchten. Laut Studiendesign sollten alle Teilnehmerinnen und Teilnehmer für 30 Monate in der Studie verbleiben – unabhängig von der Dauer der Behandlung. Die Datenerhebung erfolgte zu Beginn der Behandlung und dann alle sechs Monate, so dass sich insgesamt sechs Messzeitpunkte ergeben. Eine individuelle Datenerhebung am Ende jeder Behandlung erwies sich als nicht durchführbar, da das Forschungsteam keine Möglichkeit hatte, zeitnah vom Abschluss der einzelnen Behandlungen zu erfahren. Von Seiten der Therapeutinnen und Therapeuten erfolgte eine Fremdbeurteilung der Patientinnen und Patienten zu Beginn und am Ende der Behandlung.

Die Rekrutierung der Patientinnen und Patienten erfolgte in Kooperation mit der ÖGATAP und niedergelassenen Psychotherapeutinnen und Psychotherapeuten in Österreich. Während einer Rekrutierungsphase von einem Jahr wurden Patientinnen und Patienten welche eine neue Therapie begannen von ihrer Therapeutin oder ihrem Therapeuten auf die Studie angesprochen, ausführlich aufgeklärt und erhielten einen Eingangsfragebogen. Ihre Entscheidung für oder gegen die Teilnahme an der Studie teilten die Patientinnen und Patienten direkt der Studienleitung mit, den Therapeutinnen und Therapeuten war nicht bekannt, ob sich ihre Patientinnen und Patienten für oder gegen die Teilnahme entschieden hatten. Jene Patientinnen und Patienten, die sich zur Teilnahme bereit erklärt hatten, erhielten alle 6 Monate einen weiteren Verlaufsfragebogen per Post von der Studienleitung zugesandt. Die Rücksendung erfolgte jeweils im verschlossenen Umschlag an die Studienleitung, ohne Umweg über die Psychotherapeuten.

Unsere Bemühungen eine Kontrollgruppe aus vergleichbaren Patientinnen und Patienten zu gewinnen, die sich nicht in psychotherapeutischer, sondern in allgemeinärztlicher Behandlung befinden, erbrachten leider nicht genügend Teilnehmer. Trotz verschiedener Versuche, an niedergelassene Hausärzte heranzutreten, und trotz des Angebots einer finanziellen Aufwandsentschädigung, konnten nur 12 Ärzte für die Teilnahme gewonnen werden, welche insgesamt weniger als 10 Patienten rekrutieren konnten – deutlich zu wenige für eine repräsentative Kontrollgruppe. Entsprechend bleibt die Untersuchung in Bezug auf ein Außenkriterium für den Erfolg der Behandlungen auf Angaben zur klinischen Signifikanz der Ergebnisse beschränkt.

3.2 Messinstrumente

Der Eingangsfragebogen für die Patientinnen und Patienten zu Beginn der Behandlung (T₁) umfasste diagnostische Maße (PHQ, ADP-IV), klinische Ergebnismaße (BSI, IIP, IPO-r6, FLZ^M, PMS), gesundheitsökonomische Maße, sowie Fragen zu soziodemographischen Informationen. Alle Patientinnen und Patienten die ihr schriftliches Einverständnis zur Studienteilnahme gaben, erhielten direkt vom Forschungsteam einen kürzeren Verlaufsfragebogen (BSI) alle sechs Monate nach Beginn der Behandlung (T₂–T₅). 30 Monate nach Beginn der Behandlung (T₆) erhielten die Patientinnen und Patienten einen Fragebogen zum Therapieende, welcher erneut die oben genannten klinischen Ergebnismaße und gesundheitsökonomischen Fragen enthielt. Die Therapeutinnen und Therapeuten schätzten zu Beginn und am Ende der Behandlung die Symptomschwere (CGI-S) und das globale Funktionsniveau (GAF) ihrer Patientinnen und Patienten ein. Als Maß für Adhärenz wurden die Therapeutinnen und Therapeuten darüber hinaus gebeten, zu Beginn der Behandlung anzugeben, welche Interventionen und therapeutischen Techniken sie einzusetzen planen, und am Ende der Behandlung, welche Interventionen und Techniken tatsächlich eingesetzt wurden (CPPS; PIT).

Diagnostische Instrumente (Patienteneinschätzung)

PRIME-MD Patient Health Questionnaire (PHQ). Der Gesundheitsfragebogen für Patienten (PHQ; Spitzer, Kroenk & Williams 1999; deutsche Version: Löwe, Spitzer, Zipfel & Herzog 2002) ist die Selbstauskunftsversion des PRIME-MD. Es handelt sich um ein Screening-Instrument für eine Reihe von psychischen Störungen gemäß der Achse I des DSM-IV. Die vollständige Version wie sie hier eingesetzt wurde enthält 78 Items verschiedener Skalierung die direkt von den DSM-IV-Kriterien abgeleitet wurden. Einige der durch den PHQ erfassten Störungen haben einen subklinischen Grenzwert, d. h. sie umfassen weniger Symptome als für die korrespondierende DSM-IV-Diagnose notwendig wären. Nichtsdestotrotz wurden die psychometrischen Eigenschaften der deutschen Version des PHQ, einschließlich der Kriteriumsvalidität, als gut beurteilt (Gräfe, Zipfel, Herzog & Löwe 2004). Die interne Konsistenz lässt sich nur für die Subskalen Depressivität und Somatisierung sinnvoll berechnen. In der vorliegenden Stichprobe waren die Werte gut: Cronbach's $\alpha = .87$ für Depressivität und Cronbach's $\alpha = .79$ für Somatisierung.

Fragebogen zur Erfassung von DSM-IV Persönlichkeitsstörungen (ADP-IV). Der ADP-IV (Schotte, De Doncker, Vankerckhoven, Vertommen & Cosyns 1998; deutsche Version: Doering et al. 2007) ist ein 94-Item Selbstauskunftsfragebogen zur Erfassung von Persönlichkeitsstörungen gemäß der Achse II des DSM-IV. Jedes Item umfasst sowohl Eigenschafts- (Trait) als auch Einschränkungskarakteristika

(distress/impairment) eines DSM-IV-Kriteriums. Die Versuchsperson gibt zunächst auf einer siebenstufigen Skala (Range: 1–7) an, wie typisch eine bestimmte Persönlichkeitseigenschaft für sie ist (Beispielitem: »Es ist typisch für mich, dass meine Haltung gegenüber anderen zwischen offener Feindseligkeit und reumütigen Versprechungen, mich zu bessern wechselt.«). Wenn eine Eigenschaft als typisch (≥ 5) eingeschätzt wurde, schätzt die Versuchsperson in einem zweiten Schritt auf einer dreistufigen Skala (Range: 1–3) ein, ob diese Eigenschaft bei ihr oder anderen bereits zu Schaden oder Leid geführt hat. Der ADP-IV bietet sowohl die Möglichkeit zur kategorialen, als auch zur dimensional Diagnostik. Für die Beschreibung der Stichprobe wurde anhand der im Manual beschriebenen Cut-offs jede DSM-IV-Persönlichkeitsstörung als entweder vorhanden oder nicht vorhanden diagnostiziert. Für die Moderatoranalyse wurde ein dimensionaler Score aus der Summe aller Trait-Items gebildet, um auch pathologische Persönlichkeitszüge von Patienten abzubilden, bei denen der Cut-off-Wert für die Vergabe einer Persönlichkeitsstörung nicht erreicht wird (vgl. Renn, Höfer, Schüßler, Rumpold, Smrekar, Janecke & Doering 2008). Dieser Gesamtsummenscore wird als allgemeines Maß für pathologische Persönlichkeitszüge verstanden. Zur Validität des ADP-IV: Der Fragebogen erreichte mittlere Übereinstimmungen mit SKID-Diagnosen und Expertenratings (Doering et al. 2007). Es wurde gezeigt, dass der ADP-IV eine sehr hohe Sensitivität als Screening-Instrument hat, allerdings nur eine geringe Spezifität (Renn, Höfer, Schüßler, Rumpold, Smrekar, Janecke & Doering 2008). Die interne Konsistenz (Cronbach's α) für die einzelnen Subskalen lag im vorliegenden Datensatz zwischen akzeptabel (0,66) und gut (0,85).

Klinische Ergebnismaße (Patienteneinschätzung)

Brief Symptom Inventory (BSI). Das BSI (Derogatis 1993; deutsche Version: Franke 2000) ist ein Selbsteinschätzungsinstrument für Patientinnen und Patienten mit dem die aktuell wahrgenommene Beeinträchtigung durch körperliche und psychische Beschwerden gemessen wird (Itembeispiel: »Wie sehr litten Sie in den letzten sieben Tagen unter Verletzlichkeit in Gefühlsdingen?«). Es handelt sich um die Kurzform der weit verbreiteten Symptom-Checkliste-90-R. Anhand von 53 fünfstufig skalierten Items (Range: 0–4) werden Beschwerden aufgrund von Somatisierung, Zwanghaftigkeit, Depressivität, Ängstlichkeit, Aggressivität, sozialer Unsicherheit, paranoidem Denken und Psychotizismus erfasst. Als globales Maß für die Symptombelastung wurde der Global Severity Index (GSI) als arithmetisches Mittel der Einzelitems berechnet. Die Validität des BSI, auch hinsichtlich der Korrespondenz zur etablierten Langform, kann insgesamt als gut angesehen werden, die Test-Retest-Reliabilität liegt bei $r_{xx} = .90$ (Franke 2000). In den vorliegenden Messungen ergaben sich für die interne Konsistenz sehr gute Werte: Cronbach's $\alpha_{T1} = 0,97$, Cronbach's $\alpha_{T2} = 0,97$, Cronbach's $\alpha_{T3} = 0,98$. Zur Definition von

klinischen und nicht-klinischen Wertebereichen wurden die folgenden Stichproben herangezogen: Im Testmanual wird eine deutsche Stichprobe von $N = 600$ Erwachsenen beschrieben, die sich selbst als weder körperlich noch psychisch krank beschrieben – diese wurde hier als nicht-klinische Normstichprobe verwendet. Der mittlere GSI lag bei $M = 0,31$ ($SD = 0,23$). Für die klinische Normstichprobe wurde anstatt der im Manual beschriebenen US-amerikanischen Stichprobe ambulanter Psychatriepatientinnen und -patienten eine neuere deutsche Stichprobe von Geisheim und Kollegen (2002) verwendet. Diese umfasst $N = 1252$ ambulante Psychotherapiepatientinnen und -patienten, für die ein mittlerer GSI von $M = 1,14$ ($SD = 0,67$) berichtet wird. Der Cut-off-Punkt zwischen der klinischen und der nicht-klinischen Population liegt bei $c = .73$.

Inventar interpersoneller Probleme (IIP). Das IIP (Horowitz, Rosenberg, Baer, Ureno & Villasenor 1988; deutsche Version: Horowitz, Strauß & Kordy 2000) ist ein 64-Item Selbstauskunftsfragebogen. Auf einer fünfstufigen Skala (Range: 0–4) geben die Versuchspersonen an, inwieweit sie bestimmte Schwierigkeiten in Beziehungen zu anderen erleben (Beispielitem: »Es fällt mir schwer jemandem zu sagen, dass er mich nicht weiter belästigen soll«). Der Mittelwert über alle Items wurde berechnet und als globales Maß für interpersonelle Probleme interpretiert. Die Konstruktvalidität des IIP ist gut belegt (Gurtman 1996; Horowitz, Strauß & Kordy 2000) und es wurde als sensitiv für Veränderungen durch Psychotherapie ausgewiesen (Huber, Henrich, Klug 2007). Siefert und DeFife (2012) benennen das IIP als eines von zehn Maßen welche speziell dafür geeignet sind, die Veränderungsdimensionen auf welche psychodynamische Therapien abzielen zu erfassen. Die mittlere Test-Rest-Reliabilität liegt bei $r_{xx} = .86$. Die interne Konsistenz in der vorliegenden Stichprobe reichte von akzeptabel bis sehr gut ($.77 \leq \text{Cronbach's } \alpha \leq .96$). Der Wertebereich der nicht-klinischen Population wurde anhand der Stichprobe von Brähler, Horowitz, Kordy, Schumacher und Strauß (1999) definiert. Diese Stichprobe galt zum Zeitpunkt ihrer Erfassung als repräsentativ für die deutsche Bevölkerung im Alter von 14 bis 92 ($N = 2798$). Der IIP-Gesamtscore für diese Stichprobe liegt bei $M = 1.29$ ($SD = .52$). Der Wertebereich der klinischen Population wurde definiert durch eine Kombination aus vier klinischen Stichproben (neurotische Störungen, Persönlichkeitsstörungen, Essstörungen, psychosomatische Störungen) derselben Autoren (ebd.). Gewichtete Mittelwerte und Varianzen wurden berechnet: $N_{\text{Total}} = 565$, $M_w = 1.76$ ($SD_w = .54$). Der Cut-off-Punkt zwischen der klinischen und der nicht-klinischen Population liegt bei $c = 1.53$.

Kurzversion des Inventars zur Persönlichkeitsorganisation (IPO-16). Das IPO-16 (Zimmermann et al. 2013) ist ein 16-Item Selbstauskunftsfragebogen, welcher auf dem Inventar zur Persönlichkeitsorganisation (IPO; Lenzenwegger, Clarkin, Kernberg & Foelsch 2001) basiert, welches wiederum auf Kernbergs Theorie der Persönlichkeitsorganisation fußt. Auf einer fünfstufigen Skala (Range: 1–5) geben die Versuchspersonen an, inwieweit eine bestimmte Aussage auf sie zutrifft

(Beispielitem: »Es fällt mir schwer, Menschen zu vertrauen, da sie sich so oft gegen mich wenden oder mich betrügen.«). Die dreifaktorielle Struktur des IPO konnte für das IPO-16 repliziert werden (Zimmermann et al. 2013). Der Gesamtmittelwert ist als allgemeines Maß für Persönlichkeitspathologie interpretierbar. Trotz seiner relativen Kürze weist das IPO-16 insgesamt gute psychometrische Eigenschaften auf (ebd.). Die Test-Retest-Reliabilität liegt bei $r_{xx} = .85$. Die interne Konsistenz war in der vorliegenden Stichprobe gut (Cronbach's $\alpha_{T1} = .84$; Cronbach's $\alpha_{T6} = .85$). Der Wertebereich der nicht-klinischen Population wurde anhand der repräsentativen deutschen Normstichprobe (Zimmermann, Benecke, Hörz-Sagstetter & Dammann 2015) definiert: $N = 2502$, $M = 1.87$ ($SD = .62$). Der Wertebereich der klinischen Population basiert auf einer ambulanten Psychotherapiestichprobe wie sie von Zimmermann und Kollegen (2013) beschrieben wurde: $N_{Total} = 1036$, $M = 2.10$ ($SD = .59$). Der Cut-off-Punkt zwischen der klinischen und der nicht-klinischen Population liegt bei $c = 2.00$.

Fragebogen zur Lebenszufriedenheit (FLZ^M). Der FLZ^M (Henrich & Herschbach 2000) ist ein 32-Item Selbstausskunftsforschungsbogen. Er besteht aus zwei Modulen (mit jeweils acht Items zu verschiedenen Aspekten des jeweiligen Bereichs): »Allgemeine Lebenszufriedenheit« (ALZ; Beispielitems: »Einkommen/Finanzen«, »Familie/Kinder«) und gesundheitsbezogene Lebenszufriedenheit (Gesundheit; Beispielitems: »Körperliche Verfassung/Fitness«, »Unabhängigkeit von Unterstützung/Pflege«). Die Versuchspersonen schätzen jedes Items zweimal ein, einmal für die subjektive Bedeutsamkeit eines Aspekts und einmal für die aktuelle Zufriedenheit mit diesem Aspekt in ihrem Leben. Beide Angaben werden dann zu einem gewichteten Zufriedenheitswert zusammengefasst. Die Summe dieser gewichteten Werte für jedes Modul wird als globales Maß für allgemeine und gesundheitsbezogene Lebenszufriedenheit verstanden (Range: $-96 - 160$). Der FLZ^M hat gute psychometrische Eigenschaften, inklusive Sensitivität für Veränderung durch Psychotherapie (ebd.). Die Test-Retest-Reliabilität liegt bei $r_{xx} = .87$ für ALZ und $r_{xx} = .85$ für Gesundheit (ebd.). In der vorliegenden Stichprobe reichte die interne Konsistenz der Skalen von fragwürdig bis gut für ALZ ($.66 \leq \text{Cronbach's } \alpha \leq .84$) und von akzeptabel bis gut für Gesundheit ($.79 \leq \text{Cronbach's } \alpha \leq .86$). Der Wertebereich der nicht-klinischen Population wurde anhand von Henrich and Herschbach's (2000) repräsentativer Stichprobe der deutschen Bevölkerung definiert: $N_{ALZ} = 2534$, $M_{ALZ} = 60.49$ ($SD = 37.31$), $N_{Gesundheit} = 2218$, $M_{Gesundheit} = 74.39$ ($SD = 41.54$). Der Wertebereich der klinischen Population wurde anhand einer Stichprobe stationärer psychosomatischer Patientinnen und Patienten definiert, welche Ende der 1990er Jahre in Deutschland erhoben wurden (G. Henrich, persönliche Kommunikation, 6. Februar 2016; Henrich & Herschbach 2000): $N_{ALZ} = 1044$, $M_{ALZ} = 9.49$ ($SD = 33.66$), $N_{Gesundheit} = 1027$, $M_{Gesundheit} = -6.60$ ($SD = 35.76$). Diese einzig verfügbare Stichprobe ist suboptimal, da sie stationäre statt ambulante Patientinnen und Patienten umfasst und damit sehr wahrscheinlich zu

einer Überschätzung des Unterschieds zwischen klinischer und nicht-klinischer Population führt. Der Cut-off-Punkt zwischen der klinischen und der nicht-klinischen Population liegt bei $c = 34.99$ für ALZ und bei $c = 33.90$ für Gesundheit.

Psychological Mindedness Scale (PMS-D). Die PMS (Conte, Ratto & Karasu 1996; ins Deutsche Übersetzt von: Hamburger, Taubner & Huber 2012) ist ein 45-Item Selbstauskunftsfragebogen zur Messung der Psychological Mindedness. Psychological Mindedness (psychologische Sensibilität) wurde erstmals von Appelbaum (1973) als Konzept eingeführt und bezeichnet das Interesse und die Fähigkeit einer Person, über die eigenen Gefühle, Gedanken und Handlungen zu reflektieren. Das Konzept wurde später mit der Mentalisierungstheorie verknüpft und Allen, Fonagy und Bateman (2011) bezeichnen Psychological Mindedness als Disposition zur Mentalisierung. In der PMS-D schätzen die Versuchspersonen auf einer vierstufigen Skala (Range: 1–4) ein, inwieweit eine bestimmte Aussage auf sie zutrifft (Beispielitem: »Ich bin neugierig, aus welchen Gründen Menschen sich so verhalten, wie sie es tun«). Conte, Ratto und Karasu (1996) berichten gute psychometrische Eigenschaften für die PMS-D, die Test-Retest-Reliabilität liegt bei $r_{xx} = .92$. In späteren Studien erwies sich die ursprüngliche 5-Faktoren-Struktur der Skala als schlecht exakt zu replizieren. Die alternativen Faktorenlösungen wurden jedoch als konzeptuell ähnlich eingeschätzt (Shill & Lumley 2002). Die Summe aller (teilweise negativ kodierten) Itemwerte wird als allgemeines Maß für Psychological Mindedness interpretiert (ebd., Range: 45–180). In der vorliegenden Stichprobe ergaben sich gute interne Konsistenzen für T1 (Cronbach's $\alpha_{T1} = .86$) und T6 (Cronbach's $\alpha_{T6} = .86$). Im Fall der PMS-D konnten die Wertebereiche für die klinische und nicht-klinische Population nicht anhand von vorangegangenen Studien definiert werden, da die mittleren Werte in einer nicht-klinischen Stichprobe US-amerikanischer College-Studentinnen und -Studenten (Shill & Lumley 2002; $N = 397$, $M = 130.02$, $SD = 10.83$) sich nicht signifikant von ambulanten Psychiatriepatientinnen und -patienten (Conte, Ratto & Karasu 1996; $N = 116$, $M = 131.78$, $SD = 14.16$) unterschied. Als Kriterium für klinisch bedeutsame Veränderungen in der Psychological Mindedness wurde daher $+2SD$ vom Mittelwert zu T1 als Cut-off-Punkt verwendet.

Soziodemografische Informationen & gesundheitsökonomische Ergebnismaße

Zusätzlich zur Abfrage von soziodemographischen Informationen wurde abgefragt, wie häufig die Patientinnen und Patienten im vergangenen Jahr Ärzte besucht hatten oder medizinische Eingriffe wie Operationen oder Endoskopien vornehmen ließen. Weiterhin wurden Krankenstandstage, Medikation, Krankenhausaufenthalte und vorangegangene stationäre wie ambulante Psychotherapie abgefragt. Da für diese Einzelitems keine Reliabilitätsmaße zur Verfügung stehen, konnte der Reliable Change Index als Kriterium für Responder nicht berechnet werden. Die

Methode der prozentuellen Verbesserung ($\geq 50\%$ Verbesserung zu M_{T1} und $\geq 25\%$ individuelle Verbesserung zu $T1$) wurde stattdessen verwendet (Hiller, Schindler & Lambert 2012).

Klinische Ergebnismaße (Therapeuteneinschätzung)

Clinical Global Impression – Severity (CGI-S). Die CGI-S (Guy 1976) ist eines von zwei Items der Clinical Global Impression (CGI). Die Therapeutinnen und Therapeuten werden gebeten, die Schwere der Psychopathologie der Patientinnen und Patienten auf Basis aller verfügbaren Informationen und der Gesamtheit ihrer eigenen klinischen Erfahrung auf einer siebenstufigen Skala (Range: 1–7) einzuschätzen. Das zweite Item der CGI fragt nach dem Ausmaß der allgemeinen therapeutischen Verbesserung der Patientinnen und Patienten. Es wurde hier nicht verwendet, da keine Vorteile gegenüber einer Prä-Post-Messung mit dem Schweregrad-Item gezeigt werden konnte (Forkmann et al. 2011). Hinsichtlich der Validität korreliert die CGI-S mittelstark bis stark mit aufwendigeren störungsspezifischen Maßen für Depression und Angststörungen, sowie psychotischen Störungen (Bandelow, Baldwin, Dolberg, Andersen & Stein 2006). Die Interrater-Reliabilität wurde als gut berichtet: $ICC_{1,1} = .89$ (Honer et al. 1995, zitiert nach Johnson 2010). Da für die CGI-S keine adäquaten Normstichproben vorliegen, wurde $+2$ SD vom Mittelwert zu $T1$ als Cut-off-Punkt für klinisch bedeutsame Veränderungen gesetzt.

Global Assessment of Functioning (GAF). Die globale Erfassung des Funktionsniveaus (GAF) ist die Achse V im DSM-IV-TR (APA, 2000, deutsche Version: Saß, Wittchen, Zaudig & Houben 2003). Die Therapeutinnen und Therapeuten werden gebeten, ihren Gesamteindruck der sozialen, beruflichen und psychologischen Funktionsfähigkeit ihrer Patientinnen und Patienten wiederzugeben. Die GAF umfasst eine numerische Skala (Range: 1–100), sowie eine ausführliche Anleitung zum Einschätzen des Funktionsniveaus. Die Skala wurde wiederholt aufgrund von Problemen der konvergenten wie prädiktiven Validität kritisiert (Aas 2010). Für die vorliegende Studie wurde sich aufgrund der Kürze der Skala, sowie der weiten Verbreitung und einfachen Anwendung dennoch für die GAF entschieden. Zudem wurde in der Regel eine akzeptable bis gute Interrater-Reliabilität berichtet: $IIC_{1,1} = .84$ (Honer et al. 1995, zitiert nach Johnson 2010) – zumindest bei erfahrenen Klinikern (Söderberg, Tungenström & Armelius 2005). $+2$ SD vom Mittelwert zu $T1$ wurde als Cut-off-Punkt für klinisch bedeutsame Veränderungen gesetzt.

Instrumentenentwicklung & Adhärenzmessung

Comparative Psychotherapy Process Scale (CPPS) – Therapist Form. Die CPPS (Hilsenroth, Blagys, Ackerman, Bonge & Blais 2005) ist ein Messinstrument zur

Erfassung von therapeutischen Techniken und Interventionen in Psychotherapien. Es liegt eine Patientenversion, eine Therapeutenversion und eine Fremdbeobachtungsversion vor. Für die vorliegende Studie wurde lediglich die Therapeutenversion zu Beginn und am Ende der Behandlungen eingesetzt. Die CPPS umfasst 20 siebenstufige Items: Zehn Items beschreiben therapeutische Techniken und Interventionen die typisch für ein kognitiv-verhaltenstherapeutisches Vorgehen sind und zehn Items beschreiben ein typisch psychodynamisch-interpersonelles Vorgehen. Die CPPS hat eine gute Reliabilität und ist gut in der Lage, zwischen psychodynamischen und nicht-psychodynamischen Therapieformen zu unterscheiden (ebd.). Als globales Maß dafür, inwieweit eine gegebene Psychotherapie Elemente der beiden Richtungen enthält, wird das arithmetische Mittel der beiden Subskalen berechnet. Die von Hilsenroth, Blagys, Ackerman, Bonge und Blais, (2005) bekannte gute interne Konsistenz der beiden Subskalen konnte in der aktuellen Stichprobe bestätigt werden: $.82 \leq \text{Cronbach's } \alpha \leq .88$.

Prozessskala für imaginations- und trancebasierte tiefenpsychologische Methoden (PIT) – Therapeutenversion. Da KIP, HY und ATP einerseits psychodynamische Therapiemethoden sind, andererseits aber auch imaginations- und trancebasierte Techniken verwenden die im klassisch-psychodynamischen Vorgehen nicht enthalten sind, wurde im Rahmen der vorliegenden Studie eine Erweiterung der CPPS für imaginations- und trance-basierte psychodynamische Methoden entwickelt, die PIT (Sell, Schöpfer-Mader, Brömmel & Möller 2017). Die PIT wurde für den kombinierten Einsatz mit der CPPS entwickelt und erweitert die CPPS um 6 KIP-spezifische, 7 HY-spezifische und 4-ATP-spezifische Items. Die interne Konsistenz der einzelnen Subskalen lag in der vorliegenden Stichprobe zwischen akzeptabel und sehr gut: $.78 \leq \text{Cronbach's } \alpha \leq .92$.

CPSS und PIT sollen gemeinsam Aussagen darüber erlauben, inwieweit die Therapeutinnen und Therapeuten in den untersuchten Behandlungen ihre jeweiligen Methoden verfahrenstreu angewandt haben und wie typisch ihre Interventionen für verschiedene therapeutische Orientierungen sind. Als Baseline dienen hier zum einen die zu Beginn der Behandlungen geplanten Techniken und zum anderen die von Hilsenroth, Blagys, Ackerman, Bonge und Blais (2005) publizierten prototypischen Ratings. Hier wurden erfahrene psychodynamische und KVT-Therapeuten gebeten, CPPS-Ratings für prototypische Sitzungen in ihrem jeweiligen Verfahren abzugeben.

3.3 Einschlusskriterien & Stichprobe

Neben der informierten Einwilligung der Patientinnen und Patienten waren die Einschlusskriterien wie folgt: Alter zwischen 18 und 65 Jahren und gemäß dem diagnostischen Urteil der behandelnden Therapeutinnen und Therapeuten sollte keine psychotische Erkrankung, keine Suchterkrankung und keine schwere

neurologische Erkrankung im Vordergrund stehen. Eine F-Diagnose nach ICD-10 und ausreichende Deutschkenntnisse zum Verständnis der Fragebögen mussten gegeben sein. Als wichtige Besonderheit in Österreich ist zu nennen, dass Psychotherapien über eine Mischung aus Privatzahlungen (bei Bezuschussung durch die GKV in Höhe von 21.80 € pro Therapiestunde), privaten Zusatzversicherungen und einem festen, pro Bundesland begrenztem, Kontingent an GKV-finanzierten Therapiestunden finanziert werden (Dittrich 2012). Es ist daher zu erwarten, dass die Stichprobe – im Vergleich etwa zu Studien in Deutschland – in Richtung eines höheren sozioökonomischen Status verschoben sein wird.

Tabelle 1 stellt die soziodemographischen und klinischen Merkmale der Stichprobe zu T₁ dar. Die Gesamtstichprobe zu Beginn der Behandlungen (T₁) umfasste 353 Patientinnen und Patienten, 177 für KIP, 123 für HY und 53 für ATP. Durch Dropouts wurde die Stichprobe nach 6 Monaten (T₂) auf 209 Patientinnen und Patienten (115 für KIP, 68 für HY, 29 für ATP), nach 12 Monaten (T₃) auf 169 Patientinnen und Patienten (94 für KIP, 56 für HY, 19 für ATP), nach 18 Monaten (T₄) auf 146 Patientinnen und Patienten (82 für KIP, 48 für HY, 16 für ATP), nach 24 Monaten (T₅) auf 129 Patientinnen und Patienten (73 für KIP, 41 für HY, 15 für ATP) und nach 30 Monaten auf 109 Patientinnen und Patienten (61 für KIP, 36 für HY, 12 für ATP) reduziert – vgl. Kap. 6.2 für Dropout-Analysen und Vergleich der Stichproben.

Die Patientinnen und Patienten sind zu T₁ zu 93,7% österreichische Staatsbürgerinnen und Staatsbürger und im Mittel 37.2 (SD=13,1) Jahre alt. 70.3% sind weiblich, 28.7% männlich. 27.3% gaben als höchsten Schulabschluss ein Hochschulstudium an, 28.7% die Matura, 18.7% einen Fachschul- und 19.0% einen Pflichtschulabschluss. Lediglich 1% der Patientinnen und Patienten gaben an keinen Schulabschluss zu haben. 53.0% sind verheiratet oder leben in fester Partnerschaft, 30.0% sind ledig und 14.0% sind geschieden oder leben getrennt.

Tabelle 1: Informationen zu den Behandlungen, Patienteninformationen und klinische Merkmale der Stichprobe zu T₁

	KIP	HY	ATP
Informationen zu den Behandlungen			
N (Patienten)	177	123	53
N (Therapeuten)	51	21	9
Anzahl der Sitzungen pro Behandlung (M, SD)	63 (45.43)	53 (44.63)	20 (11.76)
Frequenz der Sitzungen (wöchentlich, zweiwöchentlich)	88.1% / 9.0%	69.3% / 28.4%	72.4% / 24.1%
Anzahl der Sitzungen mit Imaginationen / Trancen pro Behandlung (M, SD)	10 (8.01)	34 (34.55)	6 (.88)
Anzahl der noch andauernden Behandlungen zu T ₆ (n, %)	23 (13.0%)	4 (3.3%)	0 (0%)

	KIP	HY	ATP
Patienteninformationen			
Alter (M, SD)	37.0 (13.64)	37.46 (12.34)	47.80 (13.73)
Geschlecht (n, % weiblich)	117 (66.1 %)	94 (76.4 %)	33 (63.5 %)
Diagnosen (n, %)			
Somatoforme Störung (PHQ)	53 (29.9 %)	32 (26.0 %)	17 (32.1 %)
Depressive Störung (PHQ)	87 (49.2 %)	52 (42.3 %)	15 (28.3 %)
Angststörung (PHQ)	57 (32.2 %)	51 (41.4 %)	16 (30.2 %)
Essstörung (PHQ)	19 (10.7 %)	13 (10.5 %)	4 (7.6 %)
V.a. Alkoholmissbrauch (PHQ)	34 (19.2 %)	15 (12.2 %)	6 (11.3 %)
Cluster-A-Persönlichkeitsstörung (ADP-IV)	4 (2.3 %)	5 (4.1 %)	0 (0 %)
Cluster-B-Persönlichkeitsstörung (ADP-IV)	24 (13.6 %)	11 (8.9 %)	2 (3.8 %)
Cluster-C-Persönlichkeitsstörung (ADP-IV)	24 (13.6 %)	16 (13.0 %)	2 (3.8 %)
Chronische körperliche Erkrankung	76 (44.4 %)	50 (41.7 %)	24 (45.3 %)
Andere Behandlungen (n, %)			
Aktuelle psychiatrische Medikation	102 (57.6 %)	75 (61.0 %)	28 (54.4 %)
Vorangegangene Psychotherapie	108 (62.8 %)	70 (58.8 %)	18 (34.0 %)

Gemäß den mittels PHQ-D ermittelten Diagnosen (vgl. Tab. 1), kommen depressive Störungen mit 46.3 % in der Stichprobe am häufigsten vor. 36.0 % erfüllen die Kriterien einer Angst- oder Panikstörung, 28.3 % die einer somatoformen Störung. Bei 16.3 % ist von einem Alkoholmissbrauch zu sprechen, bei 10.6 % von einer Essstörung. Laut Selbstbericht im ADP-IV erfüllten insgesamt 23.7 % der Patientinnen und Patienten die Kriterien einer Persönlichkeitsstörung gemäß DSM-IV-R, wobei die selbstunsichere Persönlichkeitsstörung mit 8.6 % und die Borderline-Störung mit 10.1 % am häufigsten sind. 42.0 % der Patientinnen und Patienten in der Stichprobe gaben an, aufgrund zumindest einer chronischen somatischen Erkrankung regelmäßig in Behandlung zu sein. Der Bevölkerungsdurchschnitt laut Österreichischer Gesundheitsbefragung (Klimont & Baldaszeti 2015) liegt bei 36,0 %. Die mittlere Gesamtanzahl der (nicht-psychotherapeutischen) Arztbesuche pro Jahr liegt bei der hier untersuchten Patientengruppe mit 12,1 (SD = 12,0) deutlich über dem Bevölkerungsdurchschnitt von 6,8 (ebd.), ebenso wie die mittleren Krankenstandstage pro Jahr mit 23,1 (SD = 62,3) gegenüber einem Bevölkerungsdurchschnitt von 17,3 (ebd.).

3.4 Durchführung der Behandlungen

Die Patientinnen und Patienten in der vorliegenden Stichprobe wurden insgesamt von 81 Studientherapeutinnen und Studientherapeuten (51 für KIP, 21 für HY und

9 für ATP) behandelt. Gemäß den Prinzipien naturalistischer Forschung wurden die Therapeutinnen und Therapeuten gebeten, die Behandlungen gemäß ihrer üblichen Praxis durchzuführen. Dadurch finden sich in der Stichprobe Behandlungen unterschiedlicher Länge, Sitzungsfrequenz und, im Falle der ATP, sowohl Einzel- als auch Gruppentherapien. Die mittlere Sitzungsanzahl lag bei 63 (SD = 45.43) für KIP, 53 (SD = 44.63) für HY und 20 (SD = 11.76) für ATP. KIP-Behandlungen waren zu 88.1 % wöchentlich und zu 9.0 % zweiwöchentlich; HY-Behandlungen waren zu 69.3 % wöchentlich und zu 28.4 % zweiwöchentlich; ATP-Behandlungen waren zu 72.4 % wöchentlich und zu 24.1 % zweiwöchentlich. Eine durchschnittliche KIP-Behandlung umfasste 10 Sitzungen (SD = 8.01) mit begleiteter Imagination, eine durchschnittliche HY-Behandlung umfasste 34 Sitzungen (SD = 35.55) mit hypnotischen Trancen, eine durchschnittliche ATP-Behandlung umfasste 6 (SD = .88) Sitzungen mit autogenen Imaginationen. Nach Beendigung der Studie nach 30 Monaten dauern 23 (13.0 %) KIP-Behandlungen, 4 (3.3 %) HY-Behandlungen und keine ATP-Behandlung weiterhin an. Als Besonderheit der ATP-Behandlungen ist zu nennen, dass 24 der 53 zu T1 begonnenen Therapien Gruppenpsychotherapien im Rahmen der ATP-Grundstufe waren. Nach 30 Monaten waren 3 der 12 in der Stichprobe verbleibenden ATP-Patientinnen und -Patienten ausschließlich gruppentherapeutisch behandelt worden.

4. Ergebnisse I: Moderatoren des Behandlungserfolgs¹

Um mögliche Moderatoren des Behandlungserfolgs im Sinne der differentiellen Indikation zu identifizieren, wurde die Veränderung in der Symptombelastung innerhalb des ersten Jahres nach Beginn der Behandlung betrachtet. Mittels regressionsanalytischer Verfahren (Moderatoranalysen) wurde geprüft, ob die therapeutische Veränderung in den einzelnen Methoden unterschiedlich ausfällt, je nachdem, wie groß die Psychological Mindedness der Patienten ist und in welchem Ausmaß sie pathologische Persönlichkeitszüge aufweisen.

4.1 Datenauswertung

Die statistische Datenanalyse wurde mit *IBM SPSS Statistics Version 19.0.0* durchgeführt. Das Signifikanzniveau für alle statistischen Analysen wurde auf $\alpha = .05$ festgelegt. Es erfolgte eine explorative Datenanalyse und bei insgesamt drei

1 Die hier berichteten Ergebnisse wurden, mit freundlicher Genehmigung von Springer Nature, entnommen aus: Sell, C., Möller, H. & Taubner, S. (2017). Katathym Imaginativer Psychotherapie und Hypnosepsychotherapie: Symptomreduktion und Prädiktoren des Behandlungserfolgs. *Psychotherapeut*, 62, 547–559.

Merkmalsträgern wurde jeweils ein Ausreißer (z -Wert > 2.5) auf Ebene der globalen Skalenskennwerte entsprechend der Empfehlungen von Fidell and Tabachnick (2003) manuell adjustiert.

Da die ATP-Stichprobe besonders zu den späteren Messzeitpunkten sowohl auf Patientinnen- und Patienten- wie auch auf Therapeutinnen- und Therapeutenseite relativ geringe Fallzahlen aufwies, musste die Auswertungsstrategie für diese Stichprobe dahingehend angepasst werden. Daher wird die ATP-Stichprobe im Folgenden sowohl für die Auswertungsstrategien als auch für die Ergebnisse separat von den KIP- und HY-Stichproben behandelt.

Auswertungsstrategie für KIP- und HY-Daten

Die Veränderungen im Gesamtwert der Symptombelastung (BSI) über die drei Messzeitpunkte sowie der Einfluss der Prädiktoren Therapiemethode (KIP, HY), Psychological Mindedness (PMS-D) und pathologische Persönlichkeitszüge (ADP-IV-Gesamtscore) wurden im Rahmen eines hierarchischen linearen Wachstumsmodells mit Hilfe der vollen Maximum-Likelihood-Methode (ML) geschätzt. Alle Prädiktoren wurden für die Analyse Grand-Mean-zentriert. Einige Therapeutinnen und Therapeuten brachten mehrere Patientinnen oder Patienten in die Studie ein; entsprechend wurde durch Aufnahme einer 3. Modellebene für Therapeuteneffekte kontrolliert. Den Empfehlungen von Tasca und Gallop (2009a) folgend wurde das Modell schrittweise aufgebaut. Zur Beurteilung der Anpassungsgüte (Model-Fit) wurden *Akaiikes Informationskriterium* (AIC) und das *Schwarz-Bayes Kriterium* (SBC) herangezogen. Als Maß für die Effektgröße relativ zu einem Base-Model wurde $\sim R^2$ berechnet. Die Veränderung in der Symptombelastung zwischen den einzelnen Messzeitpunkten wurde zusätzlich durch Post-hoc-Tests (Šidák) über die modellbasiert geschätzten Randmittelwerte geprüft. Die Differenzen dieser Randmittel wurden anhand der gepoolten Streuung des BSI ($SD_{BSI, pooled} = .69$) standardisiert und als Effektmaß d interpretiert (Feingold 2009).

Es wurde a priori eine Poweranalyse durchgeführt, um die Sensitivität der geplanten statistischen Tests bei dem gegebenen $n_{T3} = 150$, $\alpha = .05$ und $1 - \beta = .90$ zu prüfen. Die Poweranalyse ergab, dass ein Effekt mindestens bei $d \geq .24$ liegen muss, um im statistischen Sinne entdeckt werden zu können.

Auswertungsstrategie für ATP-Daten

Um die Veränderung im Gesamtwert der Symptombelastung (BSI) anhand von Psychological Mindedness (PMS-D) und pathologischen Persönlichkeitszügen (ADP-IV-Gesamtscore) vorherzusagen, wurde eine Moderatoranalyse durchgeführt. Im ersten Schritt einer hierarchischen multiplen Regression wurde die Symptombelastung nach einem Jahr (T_3) mittels der Symptombelastung zu

Beginn der Therapie (T_1) vorhergesagt. Die im zweiten Schritt in die Regression aufgenommenen Variablen sagen entsprechend das Regressionsresiduum von Schritt 1, also die Veränderung von T_1 zu T_3 vorher. Alle Prädiktoren wurden für die Analyse Grand-Mean-zentriert. Als Effektgrößen wurden die standardisierten Regressionsgewichte (β) verwendet. Eine Kontrolle für Therapeuteneffekte ist hier nicht möglich.

4.2 Ergebnisse

Ergebnisse für KIP- und HY-Daten

Um die Veränderung in der Symptombelastung nach einem Jahr zu prüfen, und zur Testung möglicher Prädiktoren des Behandlungserfolgs, wurde schrittweise ein hierarchisches lineares Wachstumsmodell (ML) mit den folgenden Spezifikationen aufgebaut und getestet: 1. Ebene: Symptombelastung (BSI GSI) zu 3 Messzeitpunkten, 2. Ebene: Patientin oder Patient, 3. Ebene: behandelnde Therapeutin oder behandelnder Therapeut. **Tabelle 2** zeigt die einzelnen geschätzten Modelle im Vergleich. Als Grundlage für spätere Schätzungen der Anpassungsgüte der Modelle, wurde zunächst ein Base Model berechnet, in welchem die Symptombelastung über alle Messzeitpunkte hinweg lediglich auf Basis des (Random) Intercepts vorhergesagt wurde, um die totalen Between- und Within-Varianzanteile zu schätzen. Für dieses Modell ergab sich für die Varianz zwischen den einzelnen Behandlungen $\tau_{00} = .27$, die totale Residualvarianz (innerhalb einer Behandlung) beträgt $\delta_{2\text{Base}} = .21$ (AIC: 1180.12 / SBC: 1193.47). Die Intraklassenkorrelation (ICC) lag bei $r_{IC} = .562$. 56.2% der Varianz in der Symptombelastung lassen sich also Unterschieden zwischen den Patient-Therapeut-Paaren zurechnen.

Im zweiten Schritt wurde die lineare Komponente (Messzeitpunkt: 0, 6, 12 Monate) in das Modell aufgenommen (Unconditional Linear Model) um zu testen, ob über die drei Messzeitpunkte eine Veränderung in der Symptombelastung vorliegt. Dabei wurde ein Random Slope für die linear gestufte Variable Messzeitpunkt in das Modell aufgenommen, um Varianz in den Verläufen der einzelnen Patienten mitmodellieren zu können. Dieses Modell verbesserte den Model-Fit laut AIC und SBC im Vergleich zum Base Model ($\sim R^2 = .32$). Alternative Modelle mit quadratischem, bzw. piecewise-linearem Trend erbrachten keine weiteren signifikanten Verbesserungen des Model-Fits. Es wurde ein vorhersagekonformer Haupteffekt für die Variable Messzeitpunkt gefunden ($\gamma_{10} = -.209$ (SE = .023); $t(200) = -9.06$; $p < .001$). Die Symptombelastung der Patientinnen und Patienten verringerte sich im Laufe der drei Messzeitpunkte. Die Varianz der Ausgangswerte (T_1) war signifikant von null verschieden ($\tau_{00} = .302$; SE = .035; Wald $Z = 8.67$; $p < .001$), die Behandlungen begannen also bereits mit unterschiedlicher Symptombelastung. Die Steigung variierte nicht signifikant zwischen den

Behandlungen ($\tau_{10} = .019$ (SE < .012); Wald $Z = 1.60$; $p = 0.11$), es konnte also kein Hinweis auf unterschiedliche Verläufe gefunden werden. Um zu prüfen, zwischen welchen Messzeitpunkten signifikante Veränderungen stattfanden, wurden Šidák-Post-hoc-Tests über die durch das Modell geschätzten Randmittelwerte gerechnet. Dabei ergab sich für die Differenz von T1 ($M_E = 1.04$) zu T2 ($M_E = .72$) ein signifikanter Unterschied ($p < .001$; $d = .47$) und ebenso für die Differenz von T1 zu T3 ($M_E = .65$; $p < .001$; $d = .58$). Für die Differenz von T2 zu T3 konnte kein signifikanter Effekt nachgewiesen werden ($p = .34$; $d = .10$). Die Symptombelastung der Patienten verringerte sich also innerhalb der ersten 6 Monate nach Behandlungsbeginn statistisch signifikant und blieb bis 12 Monate nach Behandlungsbeginn auf diesem Niveau. Zwischen dem 7. und dem 12. Monat nach Beginn der Behandlung konnte keine weitere, statistisch bedeutsame Symptomreduktion nachgewiesen werden.

Um zu prüfen, ob Unterschiede im Ausmaß der Symptomreduktion zwischen KIP und HY vorliegen, wurden die Variable Therapiemethode (KIP, HY) und deren Interaktionsterm mit dem Messzeitpunkt ins Modell aufgenommen (Conditional Linear Model I). Es ergab sich keine Verbesserung des Model-Fits. Die Variable Therapiemethode ($\gamma_{13} = .014$; SE = .077); $t(346) = .19$; $p = .85$) und der Interaktionsterm von Messzeitpunkt und Therapiemethode ($\gamma_{14} = -.031$; SE = .049); $t(193) = -.63$; $p = .53$) waren keine signifikanten Prädiktoren der Symptombelastung, es wurden also keine Hinweise auf unterschiedliche Ausgangsbelastungen oder unterschiedliche Symptomveränderung in beiden Therapiemethoden gefunden. Hierzu ist wichtig, dass mit der gegebenen Teststärke der Studie lediglich Effekte $\geq .24$ nachweisbar wären. **Abbildung 1** zeigt den Verlauf der gemittelten Symptombelastung getrennt für KIP-, HY- und ATP-Behandlungen.

Im letzten Schritt wurde die Veränderung in der Symptombelastung durch die Prädiktoren Psychological Mindedness (PMS-D) und pathologische Persönlichkeitszüge (ADP-IV-Gesamtscore) vorhergesagt. Beide Variablen und ihr jeweiliger Interaktionsterm mit dem Messzeitpunkt wurden gemeinsam ins Modell aufgenommen, um ihren jeweils spezifischen Prädiktionswert ermitteln zu können. Da beide Prädiktorvariablen mit der Eingangsbelastung der Patientinnen und Patienten korreliert waren, wurde zusätzlich die Symptombelastung zu T1 und ihre Interaktion mit dem Messzeitpunkt ins Modell aufgenommen (Conditional Linear Model II). **Abbildung 2** zeigt für KIP- und HY-Daten den Verlauf der gemittelten Symptombelastung für Patientinnen und Patienten mit niedriger (-1SD), mittlerer (MW) und hoher (+1SD) Psychological Mindedness, **Abbildung 3** zeigt für KIP- und HY-Daten den Verlauf für Patientinnen und Patienten mit niedrigem (-1SD), mittlerem (MW) und hohem (+1SD) Anteil an pathologischen Persönlichkeitszügen.

Für das Model ergab sich – wie aufgrund der Autoregression zu erwarten – ein immenser Anstieg des Model-Fits ($\sim R^2 = .79$). Es wurde, wie angenommen, kein

Haupteffekt für Psychological Mindedness ($\gamma_{15} = -.014$ (SE $< .042$); $t(440) = -.32$, $p = .75$) und Persönlichkeitspathologie ($\gamma_{17} < .001$ (SE = $.001$); $t(437) = .35$, $p = .73$) gefunden. Der Interaktionsterm aus Psychological Mindedness und Messzeitpunkt war ein signifikanter Prädiktor für die Symptombelastung in vorhersagekonformer Richtung ($\gamma_{16} = -.228$; SE = $.078$; $t(229) = -2.91$; $p < .01$): Je höher die Psychological Mindedness der Patientinnen und Patienten zu T₁ war, desto stärker ging ihre Symptombelastung im Laufe der drei Messzeitpunkte zurück – bei statistischer Kontrolle für pathologische Persönlichkeitszüge, Therapiemethode und Eingangsbelastung. Für die Interaktion von Messzeitpunkt und struktureller Belastung wurde ebenfalls ein vorhersagekonformer Effekt gefunden ($\gamma_{18} = .001$; SE = $.001$; $t(207) = 3.43$; $p < .01$): Die Symptombelastung verringerte sich umso stärker, je geringer die strukturelle Belastung der Patientinnen und Patienten war – bei Kontrolle Psychological Mindedness, Therapiemethode und Ausgangsbelastung. Das Ausmaß der Persönlichkeitspathologie hat einen negativen Einfluss auf die Symptomreduktion – wenn für den Einfluss der Ausgangsbelastung kontrolliert wird.

Ergebnisse für ATP-Daten

Im ersten Schritt einer hierarchischen multiplen Regression wurde die Symptombelastung (BSI) zu T₃ durch die Symptombelastung durch die Symptombelastung (BSI) zu T₁ und das Intercept vorhergesagt. Das Regressionsresiduum (die Veränderung von T₁ zu T₃) wurde dann im zweiten Schritt durch Psychological Mindedness (PMS-D) und pathologische Persönlichkeitszüge (ADP-IV-Gesamtscore) vorhergesagt. Die Ergebnisse werden in **Tabelle 3** dargestellt. Die Hinzunahme der Moderatoren ergab keine signifikante Verbesserung des Model-Fits: Veränderung in $F(2, 15) = 1.869$; $p = .19$; $\Delta R^2 = .19$). Weder Psychological Mindedness ($B = -.89$ (SE = $.52$); $\beta = -.39$; $t(15) = -1.710$, $p = .05$), noch das Ausmaß pathologischer Persönlichkeitszüge ($B < -.01$ (SE $< .01$); $\beta = -.19$; $t(15) = -.828$, $p = .42$) waren signifikante Prädiktoren der Veränderung in der Symptombelastung. Für die ATP-Daten konnten keine Moderatoren des Behandlungserfolgs identifiziert werden.

Tabelle 2: Schätzungen der hierarchischen linearen Wachstumsmodelle (ML) zur Vorhersage der Symptombelastung (BSI GSI) für KIP- und HY-Behandlungen

Modell	Parameter	b	SE	t	df	P	~R ²	Modell-Fit (AIC/SBC)
<i>Base Model</i>							--	1180,12 / 1193,47
Intercept	γ_{00}	0,889	0,036	24,92	287	< 0,001		
<i>Unconditional linear Model</i>							0,32	1101,78 / 1124,03
Messzeitpunkt	γ_{10}	-0,209	0,023	-9,06	200	< 0,001		
<i>Conditional linear Model I</i>							0,32	1116,52 / 1129,85
Therapiemethode	γ_{13}	0,014	0,077	0,19	346	0,85		
Therapiemethode* Messzeitpunkt	γ_{14}	-0,031	0,049	-0,63	193	0,53		
<i>Conditional linear Model II</i>							0,79	235,42 / 248,71
PMS-D	γ_{15}	-0,014	0,042	-0,32	440	0,75		
PMS-D* Messzeitpunkt	γ_{16}	-0,228	0,078	-2,91	229	< 0,01		
ADP-IV Gesamt	γ_{17}	< 0,001	0,001	0,35	437	0,73		
ADP-IV Gesamt* Messzeitpunkt	γ_{18}	0,001	0,001	3,43	207	< 0,01		
BSI GSI zu T1	γ_{19}	0,942	0,020	47,07	436	< 0,001		
BSI GSI zu T1* Messzeitpunkt	γ_{110}	-0,414	0,038	-10,91	203	< 0,001		

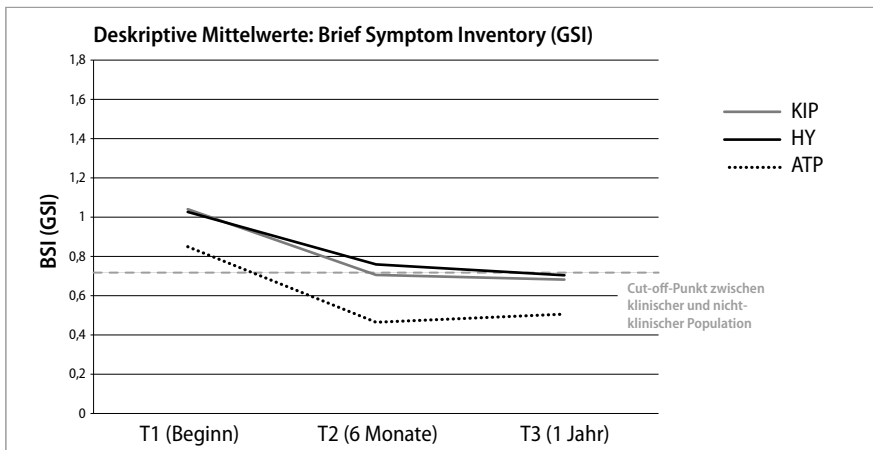


Abbildung 1: Symptombelastung (BSI GSI) für KIP-, HY- und ATP-Behandlungen

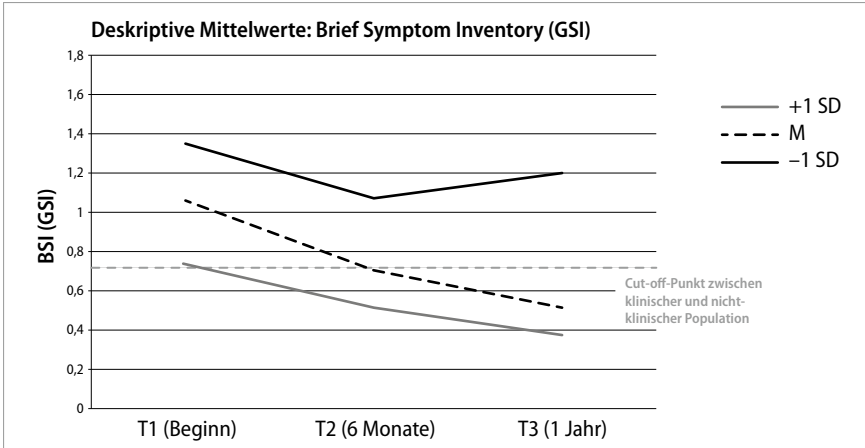


Abbildung 2: Symptombelastung (BSI GSI) für KIP- und HY-Patientinnen und Patienten mit niedriger (-1SD), mittlerer (M) und hoher (+1SD) Psychological Mindedness (PMS-D)

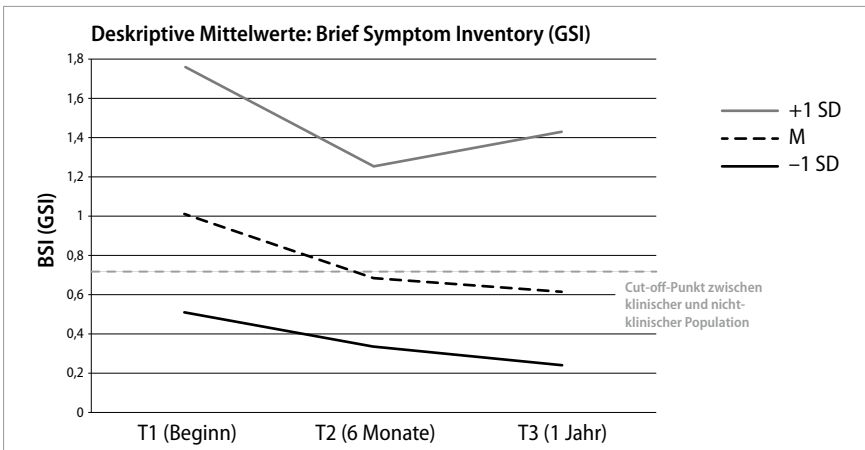


Abbildung 3: Symptombelastung (BSI GSI) für KIP- und HY-Patientinnen und Patienten mit niedrigem (-1SD), mittlerem (M) und hohem (+1SD) Anteil an pathologischen Persönlichkeitszügen (ADP-IV Gesamtscore)

Tabelle 3: Hierarchische multiple Regressionsanalyse zur Vorhersage der Symptombelastung (BSI GSI) zu T₃ durch Symptombelastung (BSI GSI) zu T₁, Psychological Mindedness (PMS-D) und pathologischen Persönlichkeitszügen (ADP-IV Gesamtscore) für ATP-Behandlungen

Variable	b	SE	β	t	P	F	ΔR ²
Schritt 1: Haupteffekt						15.440	.48***
Intercept	-.28	.24	--	-1.129	.28		
BSI GSI (T1)	.93	.24	.69	3.929	<.01		
Schritt 2: Moderatoren						1.869	.19
PMS-D	-.89	.52	-.39	-1.710	.05		
ADP-IV Gesamt	<-.01	<.01	-.19	-.828	.42		

Anmerkungen: n=19.; *** = p < .001.

4.3 Diskussion

Zusammenfassend weisen die Ergebnisse Psychological Mindedness und das Ausmaß pathologischer Persönlichkeitszüge als Moderatoren des Behandlungserfolgs in KIP- und HY-Behandlungen aus. Für ATP-Behandlungen konnten keine Moderatoren identifiziert werden.

Die Hypothese, dass eine spezifische Bedeutung der imaginations- und trancebasierten Methoden darin liegt, dass eine hoch ausgeprägte Psychological Mindedness die Behandlungsaussichten verbessert ließ sich für KIP und HY bestätigen. Die Ergebnisse deuten darauf hin, dass Patientinnen und Patienten, die es leichter haben über eigene und fremde Gefühlen nachzudenken und sich damit auseinanderzusetzen besser von einer therapeutischen Arbeit mit inneren Bildern und hypnotischen Trancen profitieren können als Patientinnen und Patienten ohne diese Disposition. Ebenfalls bestätigen ließ sich die Hypothese, dass sich eine starke Persönlichkeitspathologie negativ auf den Behandlungserfolg in KIP und HY auswirkt. Wie erwartet, finden sich zwischen KIP und HY keine Unterschiede im Ausmaß der Symptomreduktion. Für Fragen der Therapieplanung erscheint relevant, dass sich Patientinnen und Patienten mit mittlerer bis hoher Psychological Mindedness sowie Patientinnen und Patienten mit niedriger bis mittelstarker Persönlichkeitspathologiebelastung nach einem Jahr Behandlung im Mittel bereits im Wertebereich der nicht-klinischen Population befinden. Sie unterscheiden sich offenbar qualitativ von jenen Patientinnen und Patienten mit niedriger Psychological Mindedness, bzw. hohem Anteil an pathologischen Persönlichkeitszügen.

Die Tatsache, dass für ATP-Behandlungen keine Moderatoreffekte nachgewiesen werden konnten spricht nicht unbedingt dafür, dass diese nicht vorliegen. Aufgrund der kleinen Stichprobe ist die Teststärke gering und kleine bis mittelgroße

Effekte können mit den vorliegenden Daten nicht nachgewiesen werden. Zum gegenwärtigen Zeitpunkt muss die Frage nach Moderatoren des Behandlungserfolgs bei ATP-Behandlungen offen bleiben.

5 Ergebnisse II: Wirksamkeit & Verlauf der Veränderung²

Um die therapeutische Wirksamkeit von KIP, HY und ATP zu prüfen, wurde eine Reihe verschiedener Ergebnismaße eingesetzt. Die Patientinnen und Patienten schätzten Symptombelastung, interpersonelle Probleme, Persönlichkeitsorganisation, Lebenszufriedenheit, Psychological Mindedness, sowie gesundheitsökonomische Maße zu Beginn der Behandlung (T₁) und 30 Monate später (T₆) ein. Die Symptombelastung wurde zu Beginn der Behandlung und dann alle sechs Monate von den Patientinnen und Patienten eingeschätzt (T₁–T₆). Die Therapeutinnen und Therapeuten schätzten die Symptomschwere und das globale Funktionsniveau ihrer Patientinnen oder Patienten zu Beginn und am Ende der Behandlung ein. Die Verfahrenstreue (Adhärenz) der Behandlungen wurde mittels CPPS und PIT von Seiten der Therapeutinnen und Therapeuten eingeschätzt.

5.1 Datenauswertung

Im Anschluss an eine explorative Datenanalyse wurden Single-Missing-Items mit Hilfe des EM-Algorithmus imputiert. Zur Adhärenzprüfung wurden Šidák-korrigierte multiple Vergleiche über die deskriptiven Mittelwerte der einzelnen CPPS- und PIT-Skalen für die drei Methoden berechnet.

Wie bereits in Kap. 4 erfordert die verhältnismäßig kleine ATP-Stichprobe für die weitergehende Auswertung auch hier eine andere Strategie als jene für die KIP- und HY-Behandlungen – da einige komplexere Berechnungen aufgrund zu geringer Fallzahl nicht lege artis durchführbar sind. Entsprechend wird im Folgenden stets angegeben, auf welche Stichproben sich ein Abschnitt bezieht.

Auswertungsstrategie für KIP- und HY-Daten

Mittels einer Reihe von Mehrebenenanalysen wurde die Veränderung von T₁ zu T₆ bei allen Ergebnismaßen berechnet, sowie der Verlauf der Veränderung über alle sechs Messzeitpunkte für die Symptombelastung (BSI). Den Parameterschätzungen

2 Die hier berichteten Ergebnisse für KIP und HY sind entnommen aus: Sell, C., Möller, H. & Taubner, S. (2017). Effectiveness of Integrative Imagination- and Trance-Based Psychodynamic Therapies: Guided Imagery Psychotherapy & Hypnotherapy. *Journal of Psychotherapy Integration*.

wurde die volle Maximum Likelihood-Methode (ML) sowie eine autoregressive Kovarianzstruktur, AR(1), zu Grunde gelegt. In allen Analysen lag der Messzeitpunkt auf Ebene 1, die individuellen Patientinnen und Patienten auf Ebene 2 und die jeweiligen Therapeutinnen und Therapeuten auf Ebene 3. Den Empfehlungen von Tasca und Gallop (2009a; 2009b) folgend wurden alle Modelle sequentiell aufgebaut: Zunächst wurde ein Base-Model (random-intercept-only model) getestet und Wachstumsterme, sowie Kovariaten (»Anzahl der Sitzungen« und »Therapie zu T6 bereits abgeschlossen?«) wurden in nachfolgenden Schritten aufgenommen. Alle Prädiktoren wurden grand-mean zentriert. $\sim R^2$ -Werte wurden jeweils für das Full-Model relativ zum jeweiligen Base-Model berechnet. Gemäß Feingold (2009) wurde die Effektgröße d aus den modellbasiert geschätzten Randmittelwerten anhand der Standardabweichung zu T1 berechnet. Die Effektgröße wurde dann anhand der Formel von Morris and DeShon (2002) für die Pretest-Posttest-Korrelation korrigiert.

Das Vorgehen zur Prüfung der klinischen Signifikanz der Veränderungen folgt den Empfehlungen von Jacobson und Truax (1991). Für jedes Ergebnismaß wurde ein Cut-off-Punkt zwischen klinischer und nicht-klinischer Population anhand von entsprechenden Stichproben aus der Literatur definiert (vgl. Kap. 3.2). Der Reliable Change Index wurde für jeden Merkmalsträger auf jedem Ergebnismaß berechnet. Hierfür wurde der Standardmessfehler des BSI anhand der im Manual angegebenen Streuung und Reliabilität geschätzt. Es wurde ein Konfidenzintervall von 95% festgelegt. Der Status aller einzelnen Patientinnen und Patienten wurde dann definiert als »klinisch signifikant verbessert« (reliable change und Patient liegt im nicht-klinischen Bereich), »verbessert« (reliable change, aber Patient liegt weiterhin im klinischen Bereich), »unverändert« (kein reliable change) oder »verschlechtert« (reliable change in negativer Richtung).

Auswertungsstrategie für ATP-Daten

Um die therapeutischen Veränderungen von T1 zu T6 zu erfassen wurde eine Reihe von t-Tests für adjustierte Mittelwerte durchgeführt. Dabei wurde statistisch für »Anzahl der Sitzungen« kontrolliert. Anders als bei KIP- und HY-Behandlungen war eine Kontrolle für weiterhin andauernde Therapien nicht notwendig, da alle Behandlungen zu T6 bereits beendet waren. Die Effektgröße d wurde aus den deskriptiven Mittelwerten anhand der Standardabweichung zu T1 berechnet. Die Korrektur der Effektgröße erfolgte dann anhand der Formel von Morris and DeShon (2002) für die Pretest-Posttest-Korrelation. Angaben zur klinischen Signifikanz der Veränderungen konnten, anders als bei KIP- und HY-Daten, nicht erfolgen, da der von Jacobson und Truax (1991) angegebene Mindeststichprobenumfang unterschritten wurde. Um den Verlauf der Veränderung in der Symptombelastung (BSI) zu erfassen, und zu ermitteln zwischen welchen

Messzeitpunkten Veränderungen stattgefunden haben, wurden – da aufgrund der geringen Stichprobengröße zu den späteren Messzeitpunkten kein Wachstumsmodell geschätzt werden konnte – ebenfalls t-Tests für adjustierte Mittelwerte – kontrolliert für »Anzahl der Sitzungen« – durchgeführt. Das Signifikanzniveau wurde gemäß Šidák korrigiert.

5.2 Ergebnisse

Allgemeine Ergebnisse für KIP-, HY- und ATP-Behandlungen

Dropout-Analyse und Vergleich der Stichproben. Die Dropout-Raten von 35 % bis 45 % zwischen T₁ und T₂, und von dort an noch einmal 46.9 % bis zu T₆, entsprechen in etwa denen vergleichbarer naturalistischer Studien mit ambulanten Behandlungen (z. B. Ward, Wood & Awal 2013). Die Daten lassen in begrenztem Umfang Hypothesen über die Ursachen der Dropouts zu: Zum hohen Dropout zwischen T₁ und T₂ ist zu sagen, dass die Patientinnen und Patienten den Eingangsfragebogen (T₁) direkt von ihren behandelnden Therapeutinnen oder Therapeuten erhielten, während die nachfolgenden Erhebungen rein postalisch durch das Forschungsteam administriert wurden. Es ist zu vermuten, dass für die Patientinnen und Patienten mit dem ersten Fall eine höhere Verbindlichkeit verbunden war als mit der rein postalischen Erhebung. Weiterhin zeigen die Daten, dass für jene Patientinnen und Patienten die die Therapie zu T₂ bereits beendet hatten, die Dropout-Rate mit 39 % mehr als doppelt so hoch ist wie für die zu diesem Zeitpunkt weiterhin laufenden Behandlungen. Dies legt die Hypothese nahe, dass die Patientinnen und Patienten generell nach dem Ende der Behandlung weniger bereit waren, weiterhin an der Studie teilzunehmen. Ebenfalls wurden Hinweise darauf gefunden, dass Unzufriedenheit mit der Behandlung mit Studiendropouts in Verbindung steht: Für jene Patientinnen und Patienten, die zu T₂ als Grund für das Therapieende Unzufriedenheit mit der Therapie angegeben hatten (22 %), ist die Dropout-Rate mit 66 % mehr als dreifach erhöht.

Um zu prüfen, ob die Dropouts zu systematischen Verzerrungen der Stichprobe geführt haben, wurden die Patientinnen und Patienten in der Stichprobe zu T₁ mit den zu T₆ in der Stichprobe verbliebenden Patientinnen und Patienten hinsichtlich der Eingangsmessungen verglichen. Weder für KIP-³ noch für

3 *Dropout-Analyse KIP*: Alter_{T1}: $t(236) = .933$, $p = .35$; Geschlecht_{T1}: $\chi^2(2) = 1.821$, $p = .18$; BSI GSI_{T1}: $t(236) = .19$, $p = .85$; IIP Gesamt_{T1}: $t(236) = .236$, $p = .81$; IPO-16 Gesamt_{T1}: $t(236) = .718$, $p = .47$; FLZ^M ALZ_{T1}: $t(236) = .401$, $p = .69$; FLZ^M Gesundheit_{T1}: $t(236) = .028$, $p = .98$; PMS-D Gesamt_{T1}: $t(236) = .96$, $p = .34$; Arztbesuche_{T1}: $t(236) = .113$, $p = .91$; Krankenstandstage_{T1}: $t(236) = .298$, $p = .77$; Krankenhaustage_{T1}: $t(236) = .864$, $p = .39$; CGI-S_{T1}: $t(236) = .811$, $p = .42$; GAF_{T1}: $t(236) = .931$, $p = .35$.

HY-Behandlungen⁴ wurden statistisch signifikante Unterschiede festgestellt. Das Verhältnis von KIP- und HY- und ATP-Behandlungen ($\chi^2(3) = 0,94$, $p = 0,63$), sowie von KIP- HY- und ATP-Therapeutinnen und -Therapeuten ($\chi^2(3) = 0,28$, $p = 0,87$) unterlag ebenfalls keiner statistisch signifikanten Veränderung. Insgesamt zeigen die untersuchten Variablen für KIP und HY keine Gruppenunterschiede zwischen Dropout- und Completer-Patientinnen und -Patienten. Wir interpretieren dies als Hinweis darauf, dass die Ergebnisse hinsichtlich dieser Variablen nicht systematisch durch Dropouts verzerrt wurden. Bezüglich aller Belastungsvariablen waren die KIP- und HY-Patienten zu T1 ebenfalls nicht statistisch signifikant verschieden,⁵ es wurden also keine Hinweise gefunden, dass die Stichproben in den beiden Verfahren sich hinsichtlich der Belastung zu Therapiebeginn unterscheiden. Für ATP-Behandlungen hingegen wurden statistisch signifikante Unterschiede zwischen Dropout- und Completer-Patientinnen und -Patienten gefunden.⁶ Die nach 30 Monaten noch in der Stichprobe verbleibenden Patientinnen und Patienten waren zu Beginn der Behandlung hinsichtlich ihrer Symptombelastung (BSI GSI) und ihrer Persönlichkeitsorganisation (IPO-16 Gesamt) signifikant geringer belastet als die Dropout-Patientinnen und -Patienten. Weiterhin haben die ATP-Completer im Mittel eine höhere Psychological Mindedness (PMS-D Gesamt) als die Dropouts. Anders als für KIP- und HY-Behandlungen hatten die Dropouts für die ATP-Behandlungen die Stichprobe also systematisch verzerrt: Solche Patientinnen und Patienten die eingangs (T1) bereits geringer belastet waren sind mit höherer Wahrscheinlichkeit zu T6 noch in der Stichprobe als eingangs stärker belastete Patientinnen und Patienten.

-
- 4 *Dropout-Analyse HY*: Alter_{T1}: $t(157) = 1,085$, $p = .28$; Geschlecht_{T1}: $\chi^2(2) = .008$, $p = .93$; BSI GSI_{T1}: $t(157) = 1,39$, $p = .17$; IIP Gesamt_{T1}: $t(157) = 0,891$, $p = .37$; IPO-16 Gesamt_{T1}: $t(157) = .401$, $p = .69$; FLZ^M ALZ_{T1}: $t(157) = .06$, $p = .95$; FLZ^M Gesundheit_{T1}: $t(157) = .83$, $p = .41$; PMS-D Gesamt_{T1}: $t(157) = .899$, $p = .37$; Arztbesuche_{T1}: $t(157) = .116$, $p = .91$; Krankenstandstage_{T1}: $t(157) = .188$, $p = .85$; Krankenhaustage_{T1}: $t(157) = .224$, $p = .82$; CGI-S_{T1}: $t(157) = .456$, $p = .65$; GAF_{T1}: $t(157) = .330$, $p = .74$.
- 5 *Vergleich der KIP- und der HY-Stichprobe zu T1*: Somatoforme Störungen_{PHQ}, T1: $\chi^2(1) = 0,55$, $p = 0,46$; Depressive Störungen_{PHQ}, T1: $\chi^2(1) = 1,38$, $p = 0,24$; Angst- u. Panikstörungen_{PHQ}, T1: $\chi^2(1) = 2,70$, $p = 0,10$; Alkoholabusus_{PHQ}, T1: $\chi^2(1) = 2,61$, $p = 0,11$; Essstörungen_{PHQ}, T1: $\chi^2(1) < 0,01$, $p = 0,96$; Persönlichkeitsstörungen_{ADP-IV}, T1: $\chi^2(1) = 0,34$, $p = 0,56$; Chronische somatische Erkrankungen_{T1}: $\chi^2(1) < 0,03$, $p = 0,86$; Arztbesuch_{T1}: $z(298) = 1,71$, $p = 0,09$; Krankenstandstage_{T1}: $t(298) = 0,64$, $p = 0,52$; BSI GSI_{T1}: $t(298) = 0,23$, $p = 0,82$; PMS-D_{T1}: $t(298) = 0,55$, $p = 0,58$; ADP-IV-Gesamt_{T1}: $t(298) = 1,27$, $p = 0,21$.
- 6 *Dropout-Analyse ATP*: Alter_{T1}: $t(59) = .949$, $p = .35$; Geschlecht_{T1}: $\chi^2(2) = .023$, $p = .93$; BSI GSI_{T1}: $t(59) = 2,879$, $p < .01$; IIP Gesamt_{T1}: $t(59) = 1,892$, $p = .06$; IPO-16 Gesamt_{T1}: $t(59) = 2,566$, $p < .05$; FLZ^M ALZ_{T1}: $t(59) = -.979$, $p = .33$; FLZ^M Gesundheit_{T1}: $t(59) = -1,855$, $p = .07$; PMS-D Gesamt_{T1}: $t(59) = 2,313$, $p < .02$; Arztbesuche_{T1}: $t(59) = .120$, $p = .91$; Krankenstandstage_{T1}: $t(59) = .608$, $p = .55$; Krankenhaustage_{T1}: $t(59) = 1,038$, $p = .30$; CGI-S_{T1}: $t(59) = .723$, $p = .47$; GAF_{T1}: $t(59) = -.214$, $p = .83$.
-

Adhärenzprüfung. Da keine der drei Therapiemethoden in manualisierter Form vorliegt, wurden die CPPS und die PIT als Indikatoren für die Verfahrenstreue (Adhärenz) verwendet. **Tabelle 4** enthält deskriptive Statistiken für die therapeutischen Techniken und Interventionen welche – laut Therapeutenurteil am Ende der Behandlung – in KIP-, HY- und ATP-Behandlungen eingesetzt wurden. Multiple Vergleiche (Šidák-Korrektur) zeigen, dass sowohl KIP- ($t(78) = 12.656$; $p < .001$), als auch HY-Therapeutinnen und Therapeuten ($t(59) = 3.259$; $p < .01$) im Mittel mehr psychodynamische Interventionen als KVT-Interventionen eingesetzt haben. Für ATP-Therapeutinnen und Therapeuten fand sich dieser Unterschied nicht ($t(26) = .822$; $p = .42$). Wie in den jeweiligen geplanten Behandlungen nutzten KIP-Therapeutinnen und -Therapeuten mehr KIP-Interventionen als HY- oder ATP-Interventionen ($t(139) = 8.707$; $p < .001$), HY-Therapeutinnen und -Therapeuten mehr HY-Interventionen als KIP- oder ATP-Interventionen ($t(139) = 2.385$; $p < .05$) und ATP-Therapeutinnen und -Therapeuten mehr ATP-Interventionen als KIP- oder HY-Interventionen ($t(87) = 7.466$; $p < .001$). Es erscheint erwähnenswert, dass HY-Behandlungen eklektischer waren als KIP- und ATP-Behandlungen: HY-Therapeutinnen und -Therapeuten nutzten mehr KVT-Techniken als KIP- ($t(139) = 9.815$; $p < .001$) oder ATP-Therapeutinnen und -Therapeuten ($t(87) = 3.892$; $p < .001$), aber weniger als in den idealen KVT-Sitzungen

Tabelle 4: Eingesetzte Interventionen und therapeutische Techniken [Range: 0–6]

Art der Therapie	Eingesetzte Interventionen und Techniken (CPPS & PIT)				
	Psycho- dynamische Technik	KVT- Technik	KIP- Technik	HY- Technik	ATP- Technik
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
KIP-Behandlungen (diese Stichprobe; n = 80)	4.39 (.70)	1.98 (.98)	4.63 (2.01)	0.97 (1.74)	0.41 (0.75)
HY-Behandlungen (diese Stichprobe; n = 61)	5.03 (.93)	4.00 (1.46)	3.93 (1.39)	4.71 (1.13)	1.98 (1.50)
ATP-Behandlungen (diese Stichprobe; n = 28)	3.18 (1.83)	2.88 (0.62)	0.83 (0.85)	1.78 (1.18)	3.76 (1.12)
Ideale KVT-Sitzung (Hilsenroth et al.; 2005, n = 13)	3.05 (.61)	4.75 (.61)	–	–	–
Ideale psychodynamische Sitzung (Hilsenroth et al., 2005; n = 30)	5.04 (.57)	1.43 (.61)	–	–	–
Geplante KIP-Behandlung (diese Stichprobe; n = 177)	4.86 (.62)	2.14 (.89)	5.29 (0.75)	1.47 (1.48)	1.27 (1.21)
Geplante HY-Behandlung (diese Stichprobe; n = 123)	4.95 (.67)	4.01 (1.17)	4.55 (1.00)	5.19 (0.71)	3.61 (1.57)
Geplante ATP-Behandlungen (diese Stichprobe; n = 53)	3.63 (1.20)	3.15 (0.87)	1.58 (1.36)	2.30 (1.37)	4.38 (1.19)

angegeben ($t(71) = 1.810$; $p < .05$). Weiterhin wurden in HY-Behandlungen signifikant mehr Nicht-HY-Techniken eingesetzt als in KIP- ($t(139) = 10.895$; $p < .001$) oder ATP-Behandlungen ($t(87) = 10.884$; $p < .001$). Insgesamt sprechen die Daten zur Adhärenzprüfung dafür, dass in den Behandlungen aller drei Methoden die jeweiligen methodenspezifischen Techniken tatsächlich prominent verwendet wurden. Weiterhin konnten KIP- und HY-Behandlungen ihrer Behandlungstechnik nach als vorrangig tiefenpsychologisch ausgewiesen werden, während in ATP-Behandlungen insgesamt weniger allgemeine Techniken zum Einsatz kamen als in den anderen beiden Methoden.

Spezifische Ergebnisse für KIP- und HY-Daten

Therapeutische Veränderung nach 30 Monaten. Für die KIP-Behandlungen zeigten die Mehrebenenanalysen für alle Ergebnismaße eine statistisch signifikante Veränderung nach 30 Monaten – bei statistischer Kontrolle für »Anzahl der Sitzungen« und »Therapie zu T6 bereits abgeschlossen?«. **Tabelle 5** fasst die Ergebnisse zusammen: Die Belastung der Patientinnen und Patienten nahm signifikant ab in Bezug auf Symptombelastung (BSI GSI: $t(52) = 5.515$; $p < .001$; $d = .78$), interpersonelle Probleme (IIP Gesamt: $t(52) = 2.842$; $p < .01$; $d = .39$), Persönlichkeitsorganisation (IPO-16 Gesamt: $t(49) = 3.744$; $p < .001$; $d = .50$), allgemeine Lebenszufriedenheit (FLZ^M – ALZ: $t(51) = 3.164$; $p < .01$; $d = .64$), gesundheitsbezogene Lebenszufriedenheit (FLZ^M – Gesundheit: $t(58) = 6.735$; $p < .001$; $d = .63$), Psychological Mindedness (PMS-D Gesamt: $t(55) = 2.185$; $p < .05$; $d = .71$), Arztbesuche im letzten Jahr ($t(44) = 3.534$; $p < .001$; $d = .48$), Krankenstandstage im letzten Jahr ($t(44) = 1.997$; $p < .05$; $d = .27$), Krankenhaustage im letzten Jahr ($t(44) = 2.316$; $p < .05$; $d = .12$), Symptomschwere (Therapeuteneinschätzung, CGI-S: $t(107) = 11.605$; $p < .001$; $d = 1.76$) und globales Funktionsniveau Therapeuteneinschätzung, GAF: $t(112) = 13.806$; $p < .001$; $d = 1.46$).

Die klinische Signifikanz der Veränderungen in KIP-Behandlungen stellt sich wie folgt dar: Laut der patienteneingeschätzten klinischen Ergebnismaße haben zwischen 18% und 36.7% der Patientinnen und Patienten eine positive Veränderung erfahren, zwischen 63.3% und 75.4% blieben nach 30 Monaten unverändert und zwischen 0% und 9.8% haben sich verschlechtert. Auf den gesundheitsökonomischen Maßen erfuhren zwischen 19.7% und 65.5% eine positive Veränderung, zwischen 10.9% und 67.2% blieben unverändert und zwischen 13.1% und 27.1% verschlechterten sich. Laut der therapeuteneingeschätzten klinischen Ergebnismaße veränderten sich zwischen 59.2% und 74.6% der Patientinnen und Patienten positiv, zwischen 20.5% und 38.4% blieben unverändert und zwischen 2.4% und 4.9% verschlechterten sich.

Für die HY-Behandlungen ergaben die Mehrebenenanalysen eine statistisch signifikante Veränderung nach 30 Monaten für alle Ergebnismaße bis auf die Anzahl

der Krankenhaustage – bei statistischer Kontrolle für »Anzahl der Sitzungen« und »Therapie zu T6 bereits abgeschlossen?«. Die Ergebnisse sind in **Tabelle 6** zusammengefasst: Die Patientinnen und Patienten verbesserten sich in Bezug auf BSI GSI ($t(45) = 4.428$; $p < .001$; $d = .88$), IIP Gesamt ($t(45) = 3.510$; $p < .001$; $d = .67$), IPO-16 Gesamt ($t(28) = 5.623$; $p < .001$; $d = .95$), FLZ^M – ALZ ($t(44) = 4.716$; $p < .010$; $d = .83$), FLZ^M – Gesundheit ($t(40) = 4.310$; $p < .001$; $d = .81$), PMS-D Gesamt ($t(30) = 2.177$; $p < .05$; $d = .46$), Arztbesuche im letzten Jahr ($t(24) = 3.379$; $p < .01$; $d = .47$), Krankenstandstage im letzten Jahr ($t(23) = 1.767$; $p < .05$; $d = .26$), CGI-S ($t(75) = 11.233$; $p < .001$; $d = 1.58$), und GAF ($t(73) = 11.273$; $p < .001$; $d = 1.38$). Die Anzahl der Krankenhaustage im letzten Jahr ($t(18) = 1.374$; $p = .09$) verringerte sich nicht signifikant.

Zur klinischen Signifikanz der Veränderungen in HY-Behandlungen ist zu sagen, dass laut der patienteneingeschätzten Ergebnismaße zwischen 25 % und 52.7 % der Patientinnen und Patienten eine positive Veränderung nach 30 Monaten berichteten, zwischen 44.4 % und 75.0 % unverändert blieben und zwischen 0 % und 11.1 % sich verschlechterten. Gemäß der gesundheitsökonomischen Maße erfuhr zwischen 22.2 % und 61.3 % der Patientinnen und Patienten eine Verbesserung, zwischen 16.7 % und 58.8 % blieben unverändert und zwischen 16.1 % und 23.3 % verschlechterten sich. Laut der therapeuteneingeschätzten Ergebnismaße verbesserten sich zwischen 58.8 % und 82.5 % der Patientinnen und Patienten, zwischen 12.5 % und 41.3 % blieben unverändert und zwischen 0 % und 5 % verschlechterten sich.

Verlauf der Veränderung in der Symptombelastung. Um genauer zu bestimmen, zwischen welchen Messzeitpunkten die oben beschriebene therapeutische Veränderung stattfindet, konnten für die Symptomreduktion (BSI GSI) Mehrebenenanalysen über alle sechs Messzeitpunkte (T1–T6; alle sechs Monate) herangezogen werden – erneut mit statistischer Kontrolle für »Anzahl der Sitzungen« und »Therapie zu T6 bereits abgeschlossen?«. Die Ergebnisse werden in **Tabelle 7** dargestellt. Für KIP-Behandlungen waren sowohl ein linearer Term ($t(273) = 3.458$; $p < .001$), als auch ein quadratischer Term ($t(265) = 1.879$; $p < .05$) statistisch signifikant. Die Symptomreduktion folgt demnach einer quadratischen Wachstumskurve. Šidák-Post-hoc-Tests über die modellbasiert geschätzten Randmittelwerte ergaben, dass die mittlere Symptombelastung zunächst zwischen T1 und T2 ($t(304) = 3.319$; $p < .001$; $d = .52$) und dann erneut zwischen T3 und T6 ($t(284) = 2.471$; $p < .01$; $d = .56$) signifikant gesunken ist. Zwischen den dazwischenliegenden Messzeitpunkten wurden keine signifikanten Unterschiede gefunden.

Für HY-Behandlungen wurden vergleichbare Ergebnisse gefunden: linearer Effekt ($t(123) = 4.751$; $p < .001$), quadratisches Polynom ($t(132) = 3.559$; $p < .001$). Šidák-Post-hoc-Tests zeigten signifikante Differenzen zwischen den modellbasiert geschätzten Randmitteln von T1 und T2 ($t(112) = 2.867$; $p < .01$; $d = .51$) und erneut

zwischen T2 und T6 ($t(95) = 1.953$; $p < .05$; $d = .55$). Dazwischen wurden wiederum keine weiteren signifikanten Unterschiede gefunden.

Für KIP- und HY-Behandlungen gleichermaßen weisen die Ergebnisse damit zum einen auf eine schnelle initiale Verbesserung innerhalb der ersten sechs Monate nach Therapiebeginn hin und zum anderen auf eine zweite, langsamere Veränderung, welche bis mindestens 30 Monate nach Therapiebeginn andauern kann. Dieser Verlauf wird in **Abbildung 4** illustriert, in der die deskriptiven Mittelwerte in der Symptombelastung (BSI GSI) von T1 bis T6 dargestellt werden.

Spezifische Ergebnisse für ATP-Daten

Therapeutische Veränderung nach 30 Monaten. Für die ATP-Daten ergaben t-Tests für adjustierte Mittelwerte (bei statistischer Kontrolle für »Anzahl der Sitzungen«) 30 Monate nach Beginn der Behandlung eine statistisch signifikante Verbesserung auf allen klinischen Ergebnismaßen mit Ausnahme der Psychological Mindedness. Die Daten zu den gesundheitsökonomischen Items waren aufgrund zu vieler fehlender Werte nicht auswertbar. **Tabelle 8** fasst die Ergebnisse zusammen: Die Belastung der Patientinnen und Patienten nahm signifikant ab in Bezug auf Symptombelastung (BSI GSI: $t(11) = 2.480$; $p < .005$; $d = .65$), interpersonelle Probleme (IIP Gesamt: $t(11) = 2.916$; $p < .01$; $d = .93$), Persönlichkeitsorganisation (IPO-16 Gesamt: $t(11) = 2.091$; $p < .05$; $d = .61$), allgemeine Lebenszufriedenheit (FLZ^M – ALZ: $t(11) = 1.844$; $p < .05$; $d = .53$), gesundheitsbezogene Lebenszufriedenheit (FLZ^M – Gesundheit: $t(11) = 3.235$; $p < .01$; $d = .90$), Symptomschwere (Therapeuteneinschätzung, CGI-S: $t(27) = 1.838$; $p < .05$; $d = .58$) und globales Funktionsniveau (Therapeuteneinschätzung, GAF: $t(27) = 4.140$; $p < .001$; $d = .95$). Für Psychological Mindedness findet sich kein signifikanter Unterschied zwischen T1 und T6 (PMS-D Gesamt: $t(11) = -.108$; $p = .91$; $d = -.03$).

Verlauf der Veränderung in der Symptombelastung. Im schrittweisen Vergleich der Messzeitpunkte T1 bis T6 in Bezug auf die Symptombelastung ergaben geplante Kontraste (Šidák-Korrektur) mit t-Tests für adjustierte Mittelwerte eine signifikante Verbesserung von T1 zu T2 ($t(28) = 1.749$; $p < .05$; $d = .64$). Weitere signifikante Verbesserungen hin zu späteren Messzeitpunkten fanden sich, anders als bei KIP- und HY-Behandlungen, nicht. Der Verlauf der deskriptiven Mittelwerte für die Symptombelastung ist in **Abb. 4** dargestellt.

Tabelle 5: Veränderung in klinischen Ergebnismaßen nach 30 Monaten in KIP-Behandlungen: Mehrebenenanalysen (ML), kontrolliert für »Anzahl der Sitzungen« und »Therapie zu T6 bereits abgeschlossen«

	Veränderung nach 30 Monaten				Reliable Change und klinisch signifikante Veränderung					
	Beginn (T1/Prä) M (SD)	30 Monate (T6) M (SD)	p	~R ²	d	+++ (n, %)	++ (n, %)	+	o	- (n, %)
Klinische Ergebnismaße (Patientenbericht)										
BSI GSI [Range: 0–4]	1.05 (.72)	.51 (.49)	<.001	.52	.78	19 (31.7%)	3 (5.0%)	38 (63.3%)	0 (0%)	0 (0%)
IIP Gesamt [Range: 0–4]	1.37 (.58)	1.15 (.53)	<.01	.36	.39	13 (21.3%)	2 (3.3%)	44 (72.1%)	2 (3.3%)	2 (3.3%)
IPO-16 Gesamt [Range: 1–5]	2.11 (.58)	1.74 (.44)	<.001	.50	.50	12 (19.7%)	3 (4.9%)	46 (75.4%)	0 (0%)	0 (0%)
FLZ ^M – ALZ [Range: –96–160]	24.81 (34.90)	43.50 (31.37)	<.01	.32	.64	12 (19.7%)	2 (3.3%)	44 (72.1%)	3 (4.9%)	3 (4.9%)
FLZ ^M – Gesundheit [Range: –96–160]	31.58 (36.40)	63.03 (37.18)	<.001	.47	.63	19 (31.1%)	1 (1.6%)	41 (67.2%)	0 (0%)	0 (0%)
PMS Gesamt [Range: 45–180]	138.35 (14.17)	142.22 (14.59)	<.05	.09	.71	1 (1.6%)	10 (16.4%)	44 (72.1%)	6 (9.8%)	6 (9.8%)
Gesundheitsökonomische Ergebnismaße										
Arztbesuche im letzten Jahr	11.13 (8.87)	6.41 (7.00)	<.001	.14	.48	27 (49.1%)	9 (16.4%)	6 (10.9%)	13 (23.6%)	13 (23.6%)
Krankenstandstage im letzten Jahr	21.25 (49.07)	7.61 (16.37)	<.05	.19	.27	21 (35.6%)	1 (1.7%)	21 (35.6%)	16 (27.1%)	16 (27.1%)
Krankhaustage im letzten Jahr	3.07 (11.66)	0.77 (1.91)	<.05	.13	.12	12 (19.7%)	0 (0%)	41 (67.2%)	8 (13.1%)	8 (13.1%)
Klinische Ergebnismaße (Therapeutenbericht)										
CGI-S [Range: 1–7]	4.54 (.82)	3.04 (1.55)	<.001	.51	1.76	46 (37.7%)	45 (36.9%)	25 (20.5%)	6 (4.9%)	6 (4.9%)
GAF [Range: 1–100]	54.28 (11.48)	71.34 (14.93)	<.001	.63	1.46	46 (36.8%)	28 (22.4%)	48 (38.4%)	3 (2.4%)	3 (2.4%)

Anmerkungen: ~R²-Werte sind für das Full-Model relativ zum jeweiligen Base-Model (Random-Intercept-only Model); d-Werte sind die Differenz zwischen modellbasiert geschätzten Randmitteln, standardisiert anhand der SD zu T1, korrigiert für die Pretest-Posttest-Korrelation (Morris & DeShon 2002).
 Ergebnismaße (Patientenbericht): $\eta_1 = 177$; $\eta_6 = 61$; Ergebnismaße (Therapeutenbericht): $\eta_1 = 177$; $\eta_6 = 131$.
 +++ = »klinisch signifikant verbessert«; ++ = »verbessert«; + = »unverändert«; o = »verschlechtert«.

Tabelle 6: Veränderung in klinischen Ergebnismaßen nach 30 Monaten in HY-Behandlungen: Mehrebenenanalysen (ML, kontrolliert für »Anzahl der Sitzungen« und »Therapie zu T6 bereits abgeschlossen«)

	Veränderung nach 30 Monaten				Reliable Change und klinisch signifikante Veränderung				
	Beginn (T1/Prä) M (SD)	30 Monate (T6) M (SD)	p	$\sim R^2$	d	++ (n, %)	+	0	-
Klinische Ergebnismaße (Patientenbericht)									
BSI GSI [Range: 0–4]	1.03 (.74)	.52 (.51)	<.001	.34	.88	12 (33.3%)	7 (19.4%)	17 (47.2%)	0 (0%)
IIP Gesamt [Range: 0–4]	1.35 (.59)	1.00 (.59)	<.001	.22	.67	17 (47.2%)	0 (0%)	16 (44.4%)	3 (8.3%)
IPO-16 Gesamt [Range: 1–5]	1.99 (.52)	1.72 (.53)	<.001	.52	.94	8 (22.2%)	1 (2.8%)	27 (75.0%)	0 (0%)
FLZ ^M – ALZ [Range: –96–160]	29.97 (35.35)	59.39 (46.07)	<.001	.33	.83	11 (30.6%)	0 (0%)	24 (66.7%)	1 (2.8%)
FLZ ^M – Gesundheit [Range: –96–160]	31.74 (44.91)	62.10 (53.83)	<.001	.38	.81	9 (25.0%)	1 (2.8%)	26 (72.2%)	0 (0%)
PMS Gesamt [Range: 45–180]	139.08 (13.10)	140.47 (13.62)	<.05	.06	.46	0 (0%)	12 (33.3%)	20 (55.6%)	4 (11.1%)
Gesundheitsökonomische Ergebnismaße									
Arztbesuche im letzten Jahr	13.50 (15.28)	7.75 (10.91)	<.01	.14	.47	14 (45.2%)	5 (16.1%)	7 (22.6%)	5 (16.1%)
Krankenstandstage im letzten Jahr	19.99 (40.20)	8.42 (15.94)	<.05	.18	.26	15 (12.2%)	3 (10.0%)	5 (16.7%)	7 (23.3%)
Krankhaustage im letzten Jahr	5.39 (16.07)	0.78 (1.51)	=.09	.11	n.s.	8 (23.5%)	0 (0%)	20 (58.8%)	6 (17.6%)
Beginn (T1/Prä) Ende (Post)									
	M (SD)	M (SD)							
Klinische Ergebnismaße (Therapeutenbericht)									
CGI-S [Range: 1–7]	4.43 (1.06)	2.64 (1.46)	<.001	.57	1.58	36 (45.0%)	30 (37.5%)	10 (12.5%)	4 (5.0%)
GAF [Range: 1–100]	58.98 (12.37)	76.19 (13.28)	<.001	.62	1.38	28 (35.0%)	19 (23.8%)	33 (41.3%)	0 (0%)

Anmerkungen: $\sim R^2$ -Werte sind für das Full-Model relativ zum jeweiligen Base-Model (Random-Intercept-only Model); d-Werte sind die Differenz zwischen modellbasiert geschätzten Randmitteln, standardisiert anhand der SD zu T1, korrigiert für die Pretest-Posttest-Korrelation (Morris & DeShon, 2002).

Ergebnismaße (Patientenbericht): $n_1 = 123$; $n_6 = 36$; Ergebnismaße (Therapeutenbericht): $n_1 = 123$; $n_6 = 81$.

++ = »klinisch signifikant verbessert«; + = »verbessert«; 0 = »unverändert«; - = »verschlechtert«.

Tabelle 7: Deskriptive Statistiken und Veränderung in der Symptombelastung (BSI GSI) über sechs Messzeitpunkte. KIP- und HY-Behandlungen: Mehrebenenanalysen (ML, kontrolliert für »Anzahl der Sitzungen« und »Therapie zu T6 bereits abgeschlossen«); ATP-Behandlungen: t-Tests für adjustierte Mittelwerte (kontrolliert für »Anzahl der Sitzungen«)

	Beginn (T1/ Prä)		6 Monate (T2)		12 Monate (T3)		18 Monate (T4)		24 Monate (T5)		30 Monate (T6)		p (linear)	p (quad.)	~R ²	Šidák Post-hoc-Tests	
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)							
KIP																	
BSI GSI [Range: 0–4]	1.05 (.72)	.71 (.54)	.69 (.64)	.58 (.56)	.60 (.60)	.51 (.49)	.60 (.60)	.58 (.56)	.60 (.60)	.51 (.49)	.60 (.60)	.51 (.49)	< .001	< .05	.32	T1 > T2; T3 > T6	
HY																	
BSI GSI [Range: 0–4]	1.03 (.74)	.76 (.64)	.70 (.66)	.66 (.64)	.57 (.59)	.52 (.51)	.66 (.64)	.66 (.64)	.57 (.59)	.52 (.51)	.66 (.64)	.52 (.51)	< .001	< .001	.52	T1 > T2 > T6	
ATP																	
BSI GSI [Range: 0–4]	.85 (.60)	.47 (.59)	.51 (.81)	.30 (.29)	.31 (.43)	.37 (.63)	.30 (.29)	.30 (.29)	.31 (.43)	.37 (.63)	.37 (.63)	.37 (.63)	–	–	–	T1 > T2	

Anmerkungen: ~R²-Werte sind für das Full-Model relativ zum jeweiligen Base-Model (Random-Intercept-only Model)

KIP: $\eta_{11} = 177; \eta_{12} = 115; \eta_{13} = 94; \eta_{14} = 82; \eta_{15} = 73; \eta_{16} = 61;$

HY: $\eta_{11} = 123; \eta_{12} = 68; \eta_{13} = 56; \eta_{14} = 48; \eta_{15} = 41; \eta_{16} = 36;$

ATP: $\eta_{11} = 53; \eta_{12} = 29; \eta_{13} = 19; \eta_{14} = 16; \eta_{15} = 15; \eta_{16} = 12.$

Tabelle 8: Veränderung in klinischen Ergebnismaßen nach 30 Monaten in ATP-Behandlungen: t-Tests für adjustierte Mittelwerte (kontrolliert für »Anzahl der Sitzungen«)

	Veränderung nach 30 Monaten				
	Beginn (T1/Prä) M (SD)	30 Monate (T6) M (SD)	t	p	d
Klinische Ergebnismaße (Patientenbericht)					
BSI GSI [Range: 0–4]	.85 (.60)	.37 (.63)	2.480	<.05	.65
IIP Gesamt [Range: 0–4]	1.23 (.59)	.68 (.59)	2.916	<.01	.93
IPO-16 Gesamt [Range: 1–5]	1.99 (.52)	1.62 (0.69)	2.091	<.05	.61
FLZ ^M – ALZ [Range: –96–160]	38.51 (38.53)	62.93 (52.98)	1.844	<.05	.53
FLZ ^M – Gesundheit [Range: –96–160]	32.29 (42.84)	82.26 (68.53)	3.235	<.01	.90
PMS Gesamt [Range: 45–180]	134.91 (13.58)	131.47 (25.01)	–0.108	.91	–.03
	Beginn (T1/Prä) M (SD)	Ende (Post) M (SD)			
Klinische Ergebnismaße (Therapeutenbericht)					
CGI-S [Range: 1–7]	5.23 (7.03)	2.75 (1.56)	1.838	<.05	.58
GAF [Range: 1–100]	61.67 (13.33)	75.11 (14.92)	4.140	<.001	.95

Anmerkungen: d-Werte sind die Differenz zwischen den deskriptiven Mittelwerten, standardisiert anhand der SD zu T1, korrigiert für die Pretest-Posttest-Korrelation (Morris & DeShon, 2002).

Ergebnismaße (Patientenbericht): $n_{T1} = 53$; $n_{T6} = 12$; Ergebnismaße (Therapeutenbericht): $n_{T1} = 53$; $n_{T6} = 28$.

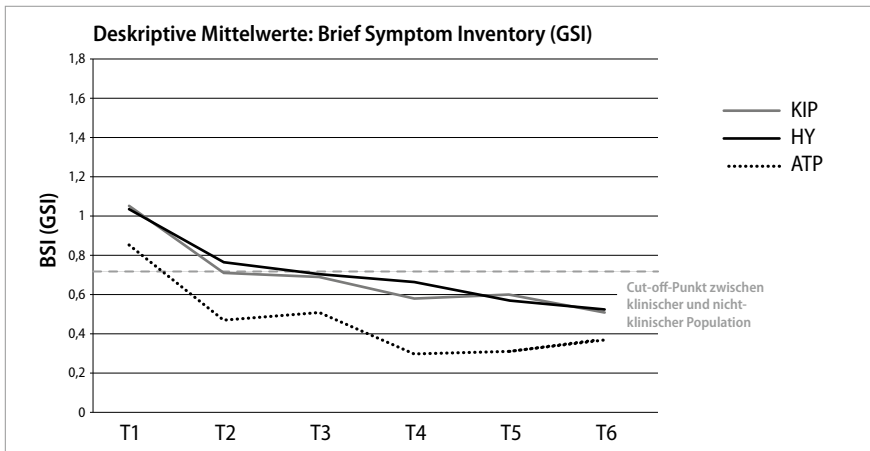


Abbildung 4: Deskriptive Mittelwerte für die Symptombelastung (BSI GSI) über sechs Messzeitpunkte in KIP-, HY- und ATP-Behandlungen

5.3 Diskussion

Anhand einer großen Bandbreite verschiedener Ergebnismaße – sowohl symptomals auch strukturbezogen, sowohl Patienten- wie auch Therapeutenbericht und zusätzliche gesundheitsökonomische Items – wurde die Wirksamkeit von KIP, HY und ATP in einem ambulanten naturalistischen Setting untersucht. Die Ergebnisse zeigen insgesamt eine gute Wirksamkeit der Methoden. Bezüglich der Verfahrenstreue (Adhärenz) der Behandlungen konnte gezeigt werden, dass die Behandlungen in allen drei Methoden in Übereinstimmung mit den jeweiligen spezifischen behandlungstechnischen Merkmalen durchgeführt wurden. KIP- und HY-Behandlungen waren darüber hinaus anhand der eingesetzten Techniken und Interventionen deutlich als tiefenpsychologische Therapien erkennbar.

Als zentrales Ergebnis kann gelten, dass Patientinnen und Patienten aller drei Methoden unter den realen Bedingungen der österreichischen Gesundheitsversorgung 30 Monate nach Beginn einer Behandlung deutlich geringer belastet sind als vor der Therapie – gemessen an nahezu allen eingesetzten Ergebnismaßen. Sowohl Maße die relativ direkt das Wohlbefinden der Patientinnen und Patienten abbilden (wie Symptombelastung und Lebenszufriedenheit), als auch die Qualität ihrer interpersonellen Beziehungen und das Level ihrer Persönlichkeitsorganisation verbesserten sich im Laufe der Behandlung. Bei KIP- und HY-Behandlungen konnte weiterhin gezeigt werden, dass die Psychological Mindedness – ein Persönlichkeitsmerkmal welches in positivem Zusammenhang mit dem Behandlungserfolg in KIP und HY steht (s. Kap. 4.2) – durch diese Therapiemethoden ebenfalls ansteigt. Des Weiteren weisen für KIP- und HY-Behandlungen drei (zwei von drei für HY) gesundheitsökonomische Maße darauf hin, dass Patientinnen und Patienten während und nach einer Psychotherapie weniger öffentliche Gesundheitsleistungen in Anspruch nehmen und weniger Krankenstandstage haben als vor ihrer Behandlung. In allen drei Methoden schätzten die Therapeutinnen und Therapeuten im Mittel sowohl die Symptomschwere als auch das globale Funktionsniveau ihrer Patientinnen und Patienten am Ende der Behandlung besser ein als zu Beginn.

In Bezug auf das Ausmaß der Veränderung auf den verschiedenen Skalen unterscheiden sich die Ergebnisse für die drei Methoden und werden daher im Folgenden einzeln diskutiert:

KIP: Während die Within-Group-Effektstärken für die therapeuteneingeschätzten Ergebnisse ($d_{\text{CGI-S}} = 1.76$; $d_{\text{CAF}} = 1.46$) innerhalb dessen lagen, was Leichsenring und Leibing (2003) diesbezüglich in einer Metaanalyse für andere Studien zu psychodynamischen Therapien fanden ($\bar{d} = 1.79$), waren die Effekte auf den patienteneingeschätzten Maßen wie Symptombelastung ($d = .78$), interpersonelle Probleme ($d = .39$), Lebenszufriedenheit ($\bar{d} = .64$) und Persönlichkeitsorganisation ($d = .50$) nicht so hoch wie die mittleren Effektgrößen in der Metaanalyse (ebd., $\bar{d} = 1.08$).

Die vorangegangenen Studien zur KIP legen nahe, dass diese Differenz in der gemessenen Effektstärke nicht spezifisch für KIP-Behandlungen im Vergleich zu anderen psychodynamischen Behandlungen, sondern sehr wahrscheinlich den spezifischen Rahmenbedingungen ambulanter Psychotherapie in Österreich geschuldet ist. In der KIP-Studie von Wietersheim, Wilke, Röser, und Meder (2003) wurden bedeutend größere Effekte für Lebenszufriedenheit ($d = .97$) gefunden, wie auch in der KIP-Studie von Sachsse, Imruck und Bahrke (2016) für Symptombelastung ($d = 1.41$) und interpersonelle Probleme ($d = .81$). Die beiden letztgenannten Studien untersuchten Behandlungen in Deutschland, welche vollständig von den gesetzlichen Krankenkassen finanziert wurden, wohingegen die Patientinnen und Patienten in der hier vorliegenden Stichprobe häufig zumindest für Teile ihrer Therapie selbst aufkommen mussten (s. Kap. 3.3). Durch diese sozioökonomische Beschränkung des Zugangs zu Psychotherapie sind sehr schwer kranke Patientinnen und Patienten in dieser Stichprobe erwartbarerweise deutlich seltener. Dies bestätigt sich anhand eines Vergleichs der Belastungswerte: Die Patientinnen und Patienten in den beiden deutschen Stichproben hatten zu Beginn ihrer Behandlung im Mittel signifikant höhere Belastungswerte als die österreichischen Studienteilnehmerinnen und -teilnehmer in der vorliegenden Stichprobe auf vergleichbaren Maßen. Die Kennwerte nach der Behandlung hingegen waren in der österreichischen Stichprobe vergleichbar oder besser als in den deutschen Studien. Dies bedeutet, dass der mittlere Patient in der vorliegenden Stichprobe zu Beginn der Behandlung weniger stark belastet war und daher weniger Raum für positive Veränderungen hatte.

Die Raten klinisch bedeutsamer Veränderungen deuten ebenfalls in diese Richtung. Sie sind in KIP- und HY-Behandlungen insgesamt vergleichbar. Auffällig ist bei beiden Methoden, dass ein relativ geringer Anteil der Patientinnen und Patienten für den gefundenen statistischen Effekt verantwortlich ist und dass viele der übrigen Patientinnen und Patienten bereits zu Beginn der Behandlung nicht übermäßig stark belastet sind: Gemessen am Mittel der patientenberichteten klinischen Ergebnismaße sind 42.7% der Patientinnen und Patienten bereits zu Therapiebeginn im Wertebereich der nicht-klinischen Population. Entsprechend konnte – je nach Skala – bei zwischen 44.4% und 75.4% der Patientinnen und Patienten auch keine statistisch reliable Veränderung festgestellt werden. Dennoch befanden sich 30 Monate nach Beginn der Behandlung zwischen 75.4% und 80.6% der Patientinnen und Patienten im Wertebereich der nicht-klinischen Population.

Die Analysen zum Verlauf der Veränderung in der Symptombelastung über sechs Messzeitpunkte zeigen – vergleichbar für KIP und HY – dass die klinische Verbesserung auf dieser Skala als zwei aufeinanderfolgende Schritte beschrieben werden kann: Eine schnelle, initiale Verbesserung innerhalb der ersten sechs Monate und eine zweite, langsamere Verbesserung innerhalb der verbleibenden 18 bis 24 Monate. Dies kann als Beleg für einen anhaltenden Veränderungsprozess in

diesen Behandlungen verstanden werden. Die Rate der Veränderung verringert sich über die Dauer der Behandlung, aber die Patientinnen und Patienten erfahren dennoch eine kontinuierliche Verbesserung – zumindest bis 30 Monate nach Beginn der Behandlung, wenn nicht länger, und erreichen schließlich zumeist Werte innerhalb der nicht-klinischen Population.

HY: Die mittlere Größe der patienteneingeschätzten Effekte ($\bar{d}=.77$) war für HY-Behandlungen etwas höher als für KIP-Behandlungen ($\bar{d}=.61$), während der mittlere Effekt laut Therapeuteneinschätzung etwas niedriger lag als bei der KIP (KIP: $\bar{d}=1.61$; HY: $\bar{d}=1.48$). Die Effektgröße für die Symptomreduktion ($d=.88$) war vergleichbar mit der mittleren Effektgröße aus Revenstorfs Metaanalyse zu hypnosebasierter Psychotherapie (2006; $\bar{d}=.95$). Die Effekte für interpersonelle Probleme ($d=.67$), Lebenszufriedenheit ($\bar{d}=.82$) und Persönlichkeitsorganisation ($d=.94$) können hier nur mit dem skalenübergreifenden Durchschnitt aus der Metaanalyse zu psychodynamischen Therapien von Leichsenring und Leibing (2003; $\bar{d}=1.08$) verglichen werden, da diese oder vergleichbare Maße bislang nicht in Studien zu hypnose- oder trancebasierten Therapien zum Einsatz gekommen sind. Die HY-Werte liegen, wie bei der KIP, im Mittel leicht unter den Ergebnissen der Metaanalyse. Die therapeuteneingeschätzte Wirksamkeit ($d_{\text{CGL-S}}=1.58$; $d_{\text{GAF}}=1.38$) ist ebenfalls – wenn auch nur geringfügig – niedriger als bei Leichsenring und Leibing (2003; $\bar{d}=1.79$).

Die für die KIP-Behandlungen diskutierte Unterschätzung des Behandlungseffekts durch die Selektivität der Stichprobe trifft mit hoher Wahrscheinlichkeit auf die Auswertung der HY-Behandlungen in vergleichbarer Weise zu. Die Interpretation zu den Raten klinisch signifikanter Veränderung und dem Verlauf der Symptomreduktion sind, wie oben ausgeführt, vergleichbar in HY- und KIP-Stichprobe.

ATP: Die gefundenen Within-Group-Effektgrößen für die therapeuteneingeschätzte Wirksamkeit der ATP-Behandlungen sind für das globale Funktionsniveau ($d=.95$) vergleichbar mit den mittleren unspezifischen Effekten bei psychischen Störungen in der Metaanalyse zu Autogenem Training von Stetter und Kupper (2002; $\bar{d}=1.08$), für die Symptomschwere ($d=.58$) allerdings deutlich niedriger. Die patienteneingeschätzte Wirksamkeit ist für Symptombelastung ($d=.65$), interpersonelle Probleme ($d=.93$), Persönlichkeitsorganisation ($d=.61$) und Lebenszufriedenheit ($\bar{d}=.72$) vergleichbar oder leicht höher als die, über mehrere Maße gemittelte, Effektstärke in der Metaanalyse (ebd., $\bar{d}=.67$). Dabei ist zu berücksichtigen, dass Stetter und Kupper Autogenes Training als therapeutische Technik betrachtet haben und nicht die hier untersuchte integrative Therapiemethode der Autogenen Psychotherapie. Wird daher als zweiter Vergleichsmaßstab die Metaanalyse zu psychodynamischen Therapien von Leichsenring und

Leibing (2003; $\bar{d}=1.08$) herangezogen, liegen die Effektgrößen der ATP zwischen denen des reinen Autogenen Trainings und denen der psychodynamischen Therapie. Wie zu KIP und HY diskutiert ist allerdings auch für die ATP-Daten eine Unterschätzung des tatsächlichen Behandlungseffekts durch die sozioökonomische Zugangsbeschränkung zu ambulanter Psychotherapie in Österreich wahrscheinlich. Im Vergleich zu KIP und HY ist die mittlere therapeuteneingeschätzte Wirksamkeit (ATP: $\bar{d}=.77$; KIP: $\bar{d}=1.61$, HY: $\bar{d}=1.48$) bei der ATP geringer, die mittlere patienteneingeschätzte Wirksamkeit (ATP: $\bar{d}=.60$; KIP: $\bar{d}=.61$, HY: $\bar{d}=.77$) ist vergleichbar mit der KIP und damit leicht geringer als bei der HY. Dabei ist zu berücksichtigen, dass die ATP-Behandlungen im Mittel deutlich kürzer waren (Anzahl Sitzungen: $M_{KIP}=63$; $M_{HY}=53$; $M_{ATP}=20$) als die KIP- und HY-Behandlungen.

Diese im Mittel deutlich geringere Therapiedauer erklärt auch, dass sich in den ATP-Behandlungen zwar – wie bei KIP und HY – eine initiale Reduktion der Symptombelastung innerhalb der ersten sechs Monate nach Behandlungsbeginn findet, aber – anders als bei den anderen beiden Methoden – keine zweite signifikante Verbesserung zu einem späteren Zeitpunkt. Ob sich ein solcher Effekt auch für ATP-Behandlungen nachweisen lässt – wenn etwa ein größerer Anteil der Patientinnen und Patienten einer Stichprobe auch im Rahmen der analytischen Oberstufe der Methode behandelt wird – kann mit dem vorliegenden Datensatz nicht beantwortet werden.

6 Diskussion der Gesamtergebnisse & Limitationen

6.1 Zusammenfassende Betrachtung der Ergebnisse

Die vorliegende Studie untersuchte die Wirksamkeit der imaginations- und tranzebasierten tiefenpsychologischen Psychotherapiemethoden Katathym Imaginative Psychotherapie (KIP), Hypnosepsychotherapie (HY) und Autogene Psychotherapie (ATP). Darüber hinaus wurde versucht, Moderatoren des Behandlungserfolgs zu identifizieren. Für die Studie konnte eine ökologisch valide Stichprobe gewonnen werden, welche die Behandlungsmodalitäten und Patientencharakteristika gemäß der aktuellen Versorgungspraxis in Österreich abbildet. Es wurden keine Hinweise auf eine systematische Verzerrung der Stichprobe durch Dropouts gefunden. Im Vergleich mit vorangegangenen Studien zu den drei Methoden waren Stichprobengröße und Studiendauer vergleichbar oder besser. Zur Beurteilung der Wirksamkeit wurde eine Reihe verschiedener Ergebnismaße ausgewählt, welche sowohl symptombezogene als auch allgemeinere Instrumente, beziehungs- und persönlichkeitsorientierte Methoden, sowie gesundheitsökonomische Maße umfasst.

In KIP- und HY-Behandlungen konnten sowohl Psychological Mindedness als auch Persönlichkeitspathologie als Moderatoren des Behandlungserfolgs

ausgewiesen werden. Je höher die Psychological Mindedness, desto stärker die Symptomreduktion innerhalb des ersten Jahres nach Therapiebeginn und je stärker die Persönlichkeitspathologie, desto geringer die Symptomreduktion. Für ATP-Behandlungen konnten diese Effekte nicht nachgewiesen werden. In Bezug auf die Wirksamkeit zeigte sich statistisch und klinisch signifikante Verbesserung auf allen Maßen für KIP, auf allen Maßen mit Ausnahme der Anzahl der Krankenhaustage für HY und auf allen Maßen mit Ausnahme der Psychological Mindedness für ATP (die gesundheitsökonomischen Daten waren hier nicht auswertbar). Die Effektgrößen für die nachgewiesenen Effekte reichten von $d = .12$ bis $d = 1.76$ für die verschiedenen Skalen. Die Anzahl der Patientinnen und Patienten mit Reliable Change reicht von 18% bis 82.5%. Bei allen Analysen wurde statistisch für die Anzahl der Sitzungen pro Therapie kontrolliert, sowie dafür, ob die Behandlungen innerhalb der Studienlaufzeit beendet wurden. Für die Symptombelastung (BSI) ergab sich für alle drei Verfahren eine signifikante Verbesserung innerhalb der ersten 6 Monate nach Beginn der Behandlung und für KIP und HY erneut innerhalb der folgenden 18 bis 24 Monate. Dies weist auf eine regressive, aber anhaltende Symptomreduktion im Laufe längerer Behandlungen hin.

Unter naturalistischen Bedingungen erweisen sich KIP, HY und ATP damit als wirksam in Bezug auf unterschiedliche Ergebnismaße. In KIP- und HY-Behandlungen wirkt sich die Psychological Mindedness der Patientinnen und Patienten positiv auf das Behandlungsergebnis aus, das Vorliegen einer ausgeprägten Persönlichkeitspathologie negativ.

6.2 Limitationen der Studie

Bezüglich der Untersuchungen zu Wirksamkeit und moderierenden Faktoren sind Einschränkungen der Aussagekraft zu nennen: Zunächst vermindert das Fehlen einer Kontrollgruppe die interne Validität der Studie. Am geeignetsten würde hier dem Anspruch naturalistischer Forschung gemäß eine Stichprobe aus Patientinnen und Patienten mit vergleichbaren Problemen und demographischen Merkmalen erscheinen, welche jedoch nicht psychotherapeutisch, sondern lediglich hausärztlich behandelt wird. Es wurden beträchtliche Anstrengungen unternommen um eine solche Stichprobe zu gewinnen (vgl. Kap. 3.1) – leider ohne hinreichenden Erfolg. Weiterhin zu Lasten der internen Validität geht es, dass für die meisten Ergebnismaße nur zwei Messzeitpunkte erhoben werden konnten, zu Beginn der Behandlung und 30 Monate später. Obschon statistisch für die Anzahl der Sitzungen und (im Fall von KIP und HY) auch dafür, ob die Behandlungen zum abschließenden Messzeitpunkt bereits beendet worden waren, kontrolliert werden konnte, besteht dennoch keine Möglichkeit, in den vorliegenden Daten zwischen Effekten zum Ende der Behandlung und Katamneseeffekten zu unterscheiden. Weiterhin unterlagen die untersuchten Behandlungen einer relativ großen Heterogenität im

Hinblick auf die Diagnosen. Dies ist zwar im Einklang mit dem naturalistischen Forschungsansatz, verunmöglicht aber – zumindest mit der gegebenen Stichprobengröße, welche für eine weitergehende Stratifizierung nicht ausreichte – Aussagen zu störungsspezifischen Effekten. Des Weiteren ist eine beträchtliche Anzahl von Dropouts zu verzeichnen. Obwohl die Dropout-Analysen (für KIP und HY) keine signifikanten Unterschiede zwischen der Intent-to-Treat-Stichprobe und der bis zum Ende der Erhebung verbleibenden Stichprobe hinsichtlich der Messungen zu Therapiebeginn ergaben, lässt sich dennoch nicht ausschließen, dass die geschätzten Parameter durch die Dropouts in einer so nicht erfassbaren Weise verzerrt worden sind.

Die Einschränkungen durch Dropouts sind im Besonderen ein Problem für die Aussagekraft der ATP-Daten: Die von Beginn an eher kleine ATP-Stichprobe schrumpfte (sowohl hinsichtlich der Anzahl der Patientinnen und Patienten, wie auch hinsichtlich der Anzahl unterschiedlicher Therapeutinnen und Therapeuten) durch die Dropouts auf ein Maß, bei welchem angezweifelt werden muss, ob die Daten als repräsentativ für Behandlungen in der untersuchten Methode allgemein gelten können. Zudem ergab die Dropout-Analyse bei den ATP-Daten eine systematische Verzerrung der Stichprobe durch die Dropouts. Aussagen über eine therapeutische Veränderung über die Zeit sind entsprechend mit Vorsicht zu gewichten. Weiterhin machte die kleine Stichprobe eine Anpassung des Vorgehens bei den statistischen Datenanalysen notwendig (s. Kap. 4.1 & 5.1), so dass ein Teil der Analysen – u. a. auch ein Vergleich von Einzel- und Gruppentherapien – nicht oder nicht in vollem Umfang durchgeführt werden konnte. Außerdem kann die Validität der regressionsanalytisch geschätzten Parameter (Kap. 4.2) nicht gewährleistet werden, da Verletzungen der Verteilungsvoraussetzungen bei derartig kleinen Stichproben schnell zu einer massiven Verzerrung der Schätzer führen können. Aus diesem Grund wurde für die Berechnung der Wirksamkeit auf dahingehend robustere Mittelwertvergleiche zurückgegriffen. Dies hat wiederum den Nachteil, dass so nur die Daten jener wenigen Patientinnen und Patienten berücksichtigt werden konnten, welche zu beiden Messzeitpunkten gemessen werden konnten. Die Ergebnisse zu den ATP-Daten sind daher insgesamt eher explorativ und als Trends zu betrachten, denn als Aussagen mit einer klar bezifferbaren Irrtumswahrscheinlichkeit.

Eine letzte Einschränkung betrifft wiederum die gesamte Stichprobe und zwar deren externe (ökologische) Validität: Wie beschrieben (s. Kap. 3.3) ist der Zugang zu Psychotherapie in Österreich zu unterschiedlichem Ausmaß von Wohnort und sozioökonomischem Status mitbestimmt. Dies bedeutet, dass die Stichprobe zwar mit hoher Wahrscheinlichkeit repräsentativ ist für Menschen die in Österreich psychotherapeutische Hilfe aufsuchen, hingegen aber ebenso wahrscheinlich nicht repräsentativ sein kann für alle Menschen, welche innerhalb Österreichs psychotherapeutischer Behandlung bedürfen.

Literatur

- Aas, I. H. M. (2010): Global Assessment of Functioning (GAF): Properties and frontier of current knowledge. *Annals of General Psychiatry*, 9 (2)
- Allen, J. G., Fonagy, P. & Bateman, A. W. (2011): *Mentalisieren in der psychotherapeutischen Praxis*. Klett-Cotta: Stuttgart
- American Psychiatric Association (2000): *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)*. Washington, D. C.: American Psychiatric Association
- Appelbaum, S. A. (1973): Psychological Mindedness: Word, concept and essence. *International Journal of Psycho-Analysis*, 54: 35–46
- Bandelow, B., Baldwin, D. S., Dolberg, O. T., Andersen, H. F. & Stein, D. J. (2006): What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *Journal of Clinical Psychiatry*, 67: 1428–1434
- Barber, J. P., Muran, C., McCarthy, K. S. & Keefe, J. R. (2013): Research on dynamic therapies. In M. J. Lambert (Hrsg.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (S. 443–494). New Jersey: John Wiley & Sons
- Bischof, B. (2012): *Autogene Psychotherapie im Vergleich*. In Österreichische Gesellschaft für angewandte Tiefenpsychologie und allgemeine Psychotherapie, In ÖGATAP (Hrsg.), *Der ATP-Reader: Grundsatzbeiträge zur Theorie und Praxis der Autogenen Psychotherapie* (S. 69–76). Wien: ÖGATAP
- Brähler, E., Horowitz, L. M., Kordy, H., Schumacher, J. & Strauß, B. (1999): Zur Validierung des Inventars Interpersonaler Probleme (IIP): Ergebnisse einer Repräsentativbefragung in Ost- und Westdeutschland. *PPmP: Psychotherapie Psychosomatik medizinische Psychologie*, 49: 422–431
- Bregenger, M. (2015): *Wirksamkeitsnachweis und Indikation des Autogenen Trainings – eine Metanalyse nach den Cochrane Kriterien*. Dissertation, Universität Regensburg
- Conte H. R., Ratto R. & Karasu, T. B. (1996): The Psychological Mindedness Scale: Factor structure and relationship to outcome of psychotherapy. *The Journal of Psychotherapy Practice and Research*, 5: 250–259.
- Derogatis, L. R. (1993): *BSI Brief Symptom Inventory: Administration, scoring, and procedure manual*. Minneapolis, MN: National Computer Systems
- Dieter, W. (2001): Die Katathym Imaginative Psychotherapie – eine tiefenpsychologische Behandlungsmethode. *Imagination*, 23 (3): 5–41
- Dittrich, K. (2012): *Zur Lage der Psychotherapie in Europa: Österreich*. *Psychotherapeutenjournal*, 1: 24–27
- Doering, S., Renn, D., Höfer, S., Rumpold, G., Smrekar, U., Janecke, N. & Schüssler, G. (2007): Validierung der deutschen Version des Fragebogens zur Erfassung von DSM-IV Persönlichkeitsstörungen (ADP-IV). *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 53 (2): 111–128
- Ehrenthal, J. C., Dinger, U. & Nikendei, C. (2014): Aktuelle Entwicklungen der psychodynamischen Psychotherapieforschung. *Psychotherapeut*, 59: 212–218
- Erickson, M. H. & Rossi, E. L. (1981): *Experiencing Hypnosis: Therapeutic Approaches to Altered States*. New York: Irvington Pub
- Esplen, M. J., Garfinkel, P. E., Olmsted, M., Gallop, R. M. & Kennedy, S. (1998): A randomized controlled trial of guided imagery in bulimia nervosa. *Psychological Medicine*, 28: 1347–1357

- Feingold, A. (2009): Effect sizes for growth-modelling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, 14 (1): 43–53
- Fidell, L.S. & Tabachnick, B.G. (2003): Preparatory data analysis. In: J.A. Schinka & W.F. Velicer (Hrsg.): *Research methods in psychology* (Vol. 2: *Handbook of Psychology*, S. 115–141). Hoboken, NJ: Wiley
- Flora, D.B. & Curran, P.J. (2004): An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9: 466–491
- Forkmann, T., Scherer, A., Boecker, M., Pawelzik, M., Jostes, R. & Gauggel, S. (2011): The Clinical Global Impression Scale and the influence of patient or staff perspective on outcome. *BMC Psychiatry*, 11: 83
- Franke, G.H. (2000): *BSI: Brief Symptom Inventory – Deutsche Version, Manual*. Göttingen: Beltz
- Geisheim, C., Hahlweg, K., Fiegenbaum, W., Frank, M., Schröder, B. & von Witzleben, I. (2002): Das Brief Symptom Inventory (BSI) als Instrument zur Qualitätssicherung in der Psychotherapie. *Diagnostica*, 48: 28–36
- Gräfe, K., Zipfel, S., Herzog, W. & Löwe, B. (2004): Screening psychischer Störungen mit dem Gesundheitsfragebogen für Patienten (PHQ-D). *Diagnostica*, 50: 171–181
- Gurtman, M.B. (1996): Interpersonal problems and the psychotherapy context: The construct validity of the Inventory of Interpersonal Problems. *Psychological Assessment*, 8: 241–255
- Guy, W. (1976): *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: US Department of Health, Education, and Welfare Public Health Service Alcohol, Drug Abuse, and Mental Health Administration
- Hagl, M. (2013): Zur Wirksamkeit von Hypnose und Hypnotherapie. Eine Studienübersicht für die Jahre 2010–2012. *Hypnose – Zeitschrift für Hypnose und Hypnotherapie*, 8: 145–181
- Hagl, M. (2014): Studien zur Wirksamkeit von Klinischer Hypnose und Hypnotherapie im Jahr 2013. *Hypnose – Zeitschrift für Hypnose und Hypnotherapie*, 9: 147–168
- Harrer, M. (2008): Wirkkonzepte in der Hypnosepsychotherapie. *Imagination*, 30 (1): 5–37
- Henrich, G. & Herschbach, P. (2000): Questions on Life Satisfaction (FLZ^M) – A short questionnaire for assessing subjective quality of life. *European Journal of Psychological Assessment*, 16: 150–159
- Hiller, W., Schindler, A.C. & Lambert, M.J. (2012): Defining response and remission in psychotherapy research: a comparison of the RCI and the method of percent improvement. *Psychotherapy Research*, 22: 1–11
- Hilsenroth, M.J., Blagys, M.D., Ackerman, S.J., Bonge, D.R. & Blais, M.A. (2005): Measuring psychodynamic-interpersonell and cognitive-behavioral techniques: Development of the comparative psychotherapy process scale. *Psychotherapy: Theory, Research, Practice, Training*, 42 (3): 340–356
- Horowitz, L.M., Rosenberg, S.E., Baer, B.A., Ureno, G. & Villasenor, V.S. (1988): Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56: 885–892
- Horowitz, L.M., Strauß, B. & Kordy, H. (2000): *Inventar zur Erfassung interpersonaler Probleme – Deutsche Version*. Göttingen: Beltz
- Huber, D., Henrich, G. & Klug, G. (2007): The Inventory of Interpersonal Problems (IIP): Sensitivity to change. *Psychotherapy Research*, 17: 474–481
- Jacobson, N.S. & Truax, P. (1991): Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59 (1): 12–19

- Johnson, D. L. (2010): A compendium of psychosocial measures: Assessment of people with serious mental illness in the community. New York: Springer
- Kanitschar, H. (2009): Hypnosepsychotherapie, ein integratives, tiefenpsychologisch fundiertes Verfahren. *Hypnose – Zeitschrift für Hypnose und Hypnotherapie*, 4: 1–23
- Klimont, J. & Baldaszeti, E. (2015): Österreichische Gesundheitsbefragung 2014: Hauptergebnisse des Austrian Health Interview Survey (ATHIS) und methodische Dokumentation. Statistik Austria: Wien
- Kottje-Birnbacher, L. (1992): Strukturierende Faktoren des Katathymen Bilderlebens. *Praxis der Psychotherapie und Psychosomatik*, 37: 164–173
- Leichsenring, F. & Leibling, E. (2003): The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: a meta-analysis. *American Journal of Psychiatry*, 160: 1223–1232
- Leichsenring, F., & Rabung, S. (2011): Long-term psychodynamic psychotherapy in complex mental disorders: update of a meta-analysis. *British Journal of Psychiatry*, 199: 15–22
- Leichsenring, F. (2013): Empirische Wirksamkeitsnachweise psychodynamischer Psychotherapie bei Persönlichkeitsstörungen: Ein Überblick. In J. F. Clarkin, P. Fonagy & G. O. Gabbard (Hrsg.), *Psychodynamische Psychotherapie der Persönlichkeitsstörungen: Handbuch für die klinische Praxis* (S. 418–434). Stuttgart: Schattauer
- Lenzenwegger, M. F., Clarkin, J. F., Kernberg, O. F. & Foelsch, P. A. (2001): The Inventory of Personality Organization: Psychometric properties, factorial composition, and criterion relations with affect, aggressive dyscontrol, psychosis proneness, and self-domains in a non-clinical sample. *Psychological Assessment*, 13: 577–591
- Leuner, H. (1994): *Lehrbuch der Katathym Imaginativen Psychotherapie*. Bern: Huber
- Linden, W. (1994): Autogenic training: a narrative and quantitative review of clinical outcome. *Biofeedback and Self-Regulation*, 19: 227–264
- Löwe, B., Spitzer, R. L., Zipfel, S. & Herzog, W. (2002): Gesundheitsfragebogen für Patienten (PHQ-D). Kompletversion. Karlsruhe: Pfizer
- Morris, S. B. & DeShon, R. P. (2002): Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7: 105–125
- Ogles, B. M. (2013): Measuring Change in Psychotherapy Research. In M. J. Lambert (Hrsg.), *Bergin and Garfield's Handbook of psychotherapy and behavior change* (S. 134–166). New Jersey: John Wiley & Sons
- Pichler, M. (2011): Katathym Imaginative Psychotherapie. In G. Stumm (Hrsg.), *Psychotherapie, Schulen und Methoden: Eine Orientierungshilfe für Theorie und Praxis* (S. 109–117). Wien: Falter Verlag
- Pokorny, D. & Stigler, M. (2006): Beziehungsschemata in der realen und der imaginierten Welt: Mit der Clusteranalyse der Verschiebung auf der Spur. In: L. Kottje-Birnbacher, E. Wilke, K. Krippner & W. Dieter (Hrsg.): *Mit Imaginationen therapieren* (S. 108–123). Lengerich: Pabst R Core Team (2015): R: A language and environment for statistical computing. Wien: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Renn, D., Höfer, S., Schüßler, G., Rumpold, G., Smrekar, U., Janecke, N. & Doering, S. (2008): Dimensionale Diagnostik mit dem Fragebogen zur Erfassung von DSM-IV-Persönlichkeitsstörungen (ADP-IV). *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 54: 214–226
- Revenstorf D. (2006): Expertise zur Beurteilung der wissenschaftlichen Evidenz des Psychotherapieverfahrens Hypnotherapie. *Hypnose – Zeitschrift für Hypnose und Hypnotherapie*, 1: 7–164

- Robertson, D. (2012): The practice of cognitive-behavioural hypnotherapy: A manual for evidence-based clinical hypnosis. London: Karnac
- Sachsse, U., Imruck, B.H. & Bahrke, U. (2016): Evaluation ambulanter Behandlungen mit Katathym Imaginativer Psychotherapie KIP: Eine naturalistische Studie. *Ärztliche Psychotherapie*, 11: 87–92
- Saß, H., Wittchen, H.-U., Zaudig, M. & Houben, I. (2003): Diagnostisches und statistisches Manual psychischer Störungen, DSM-IV-TR, Textrevision. Göttingen: Hogrefe
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003): Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8: 23–74
- Schotte, C. K. W., De Doncker, D., Vankerckhoven, C., Vertommen, H. & Cosyns, P. (1998): Self-report assessment of the DSM-IV personality disorders. Measurement of trait and distress characteristics: the ADP-IV. *Psychological Medicine*, 28: 1179–1188
- Schultz, J. H. (1930): Das Autogene Training (konzentrierte Selbstentspannung). Versuch einer klinisch-praktischen Darstellung. Leipzig: Thieme
- Sedlak, F. (2012): Ein Steckbrief der ATP. In Österreichische Gesellschaft für angewandte Tiefenpsychologie und allgemeine Psychotherapie, In ÖGATAP (Hrsg.), *Der ATP-Reader: Grundsatzbeiträge zur Theorie und Praxis der Autogenen Psychotherapie* (S. 12). Wien: ÖGATAP
- Sell, C., Möller, H. & Taubner, S. (2017): Katathym Imaginative Psychotherapie und Hypnosepsychotherapie: Symptomreduktion und Prädiktoren des Behandlungserfolgs. *Psychotherapeut*, 62: 547–559
- Sell, C., Möller, H. & Taubner, S. (2017): Effectiveness of Integrative Imagination- and Trance-Based Psychodynamic Therapies: Guided Imagery Psychotherapy & Hypnopsychotherapy. *Journal of Psychotherapy Integration*
- Sell, C., Schöpfer-Mader, E., Brömmel, B. & Möller, H. (2017): Therapeutisches Vorgehen und Interventionen in Katathym Imaginativer Psychotherapie, Hypnosepsychotherapie und Autogener Psychotherapie: Entwicklung und Validierung der Therapeutenversion der ›Prozessskala für imaginations- und trancebasierte tiefenpsychologische Methoden‹ (PIT). *Psychotherapie Forum*
- Siefert, J. S. & DeFife, J. A. (2012): Ten diverse outcome measures for psychodynamic psychotherapy research. In R. A. Levy, J. S. Ablon, & H. Kächele (Hrsg.) *Psychodynamic psychotherapy research: Evidence-based practice and practice-based evidence* (S. 529–552). NY: Humana Press
- Shedler, J. (2012): The Efficacy of Psychodynamic Psychotherapy. In: R. A. Levy, J. S. Ablon, & H. Kächele (Hrsg.), *Psychodynamic Psychotherapy Research: Evidence-Based Practice and Practice-Based Evidence* (S. 9–25). New York: Humana Press
- Shill, M. A. & Lumley, M. A. (2002): The Psychological Mindedness Scale: factor structure, convergent validity and gender in a non-psychiatric sample. *Psychology and Psychotherapy*, 75: 131–50
- Söderberg, P., Tungström, S. & Armelius, B. A. (2005): Reliability of Global Assessment of Functioning ratings made by clinical psychiatric staff. *Psychiatric Services*, 56: 434–438
- Spitzer, R. L., Kroenke, K. & Williams, J. B. W. (1999): Validation and utility of a self-report version of PRIME-MD. *Journal of the American Medical Association*, 282: 1737–1744
- Stetter, F. & Kupper, S. (2002): Autogenic training: a meta-analysis of clinical outcome studies. *Applied Psychophysiology and Biofeedback*, 27 (1): 45–98
- Stigler, M., & Pokorny, D. (2001): Emotions and primary process in guided imagery psychotherapy: Computerized text-analytic measures. *Psychotherapy Research*, 11: 415–431

- Stigler, M. & Pokorny, D. (2012): Eine Dekade der KIP-Prozessforschung im Überblick. In: H. Ullmann & E. Wilke (Hrsg.) *Handbuch Katathym Imaginative Psychotherapie* (S. 122–145). Bern: Huber
- Tasca, G. & Gallop, R. (2009a): Multilevel modeling of longitudinal data for psychotherapy researchers: I. The basics. *Psychotherapy Research*, 19: 429–437
- Tasca, G. & Gallop, R. (2009b): Multilevel modeling of longitudinal data for psychotherapy researchers: II. The complexities. *Psychotherapy Research*, 19: 438–452
- Thompson-Brenner, H. & Westen, D. (2005): A Naturalistic Study for Bulimia Nervosa, Part 2: Therapeutic Interventions in the Community. *Journal of Nervous and Mental Disease*, 193: 585–595
- Tyrer, P. & Simmonds, S. (2003): Treatment models for those with severe mental illness and comorbid personality disorder. *British Journal of Psychiatry*, 182(44): 15–18
- Walfish, S., McAlister, B., O'Donnell, P. & Lambert, M.J. (2012): An investigation of self-assessment bias in mental health providers. *Psychological Reports*, 110: 639–644
- Ward, A., Wood, B. & Awal, M.A. (2013): Naturalistic psychodynamic psychotherapy study: Evaluating outcome with a patient perspective. *British Journal of Psychiatry*, 29: 292–314
- Watkins, J.G. (1992): *Hypnoanalytic techniques. The practice of clinical hypnosis, volume II.* New York: Irvington
- Wietersheim, J.v., Wilke, E., Röse, M. & Meder, G. (2003): Ergebnisse der Katathym-imaginativen Psychotherapie: Die Effektivität der Katathym-imaginativen Psychotherapie in einer ambulanten Längsschnittstudie. *Psychotherapeut*, 48: 173–178
- Wilke, E. (1982): Eine psychotherapeutische Behandlung von Colitis ulcerosa mit dem Katathymen Bilderleben. In: H. Leuner (Hrsg.), *Psychotherapie mit dem Tagtraum – Katathymes Bilderleben: Ergebnisse II* (S. 255–268). Bern: Huber
- Wiltink, J., Edinger, J., Haselbacher, A., Imruck, B.H. & Beutel, M.E. (2010): Adherence and Competence in der Psychotherapieforschung: Neuere Entwicklungen. *Klinische Diagnostik und Evaluation*, 3: 76–93
- Zimmermann, J., Benecke, C., Hörz, S., Rentrop, M., Peham, D., Bock, A., Wallner, T., Schauenburg, H., Frommer, J., Huber, D., Clarkin, J.F. & Dammann, G. (2013): Validierung einer deutschsprachigen 16-Item-Version des Inventars der Persönlichkeitsorganisation (IPO-16). *Diagnostica*, 59: 3–16
- Zimmermann, J., Benecke, C., Hörz-Sagstetter, S. & Dammann, G. (2015): Normierung der deutschsprachigen 16-Item-Version des Inventars der Persönlichkeitsorganisation (IPO-16). *Zeitschrift für Psychosomatische Medizin und Psychotherapie*, 61: 5–18
- Zindel, J.P. (2001): Hypnose mit frühgestörten und Borderline-Patienten. In D. Revenstorf & B. Peter (Hrsg.), *Hypnose in Psychotherapie, Psychosomatik und Medizin* (S. 488–498). Berlin, Heidelberg, New York: Springer

Zusammenfassung

Der Projektbericht beschreibt eine von 2012 bis 2016 durchgeführte naturalistische Psychotherapiestudie zu Katathym Imaginativer Psychotherapie (KIP), Hypnosepsychotherapie (HY) und Autogener Psychotherapie (ATP).

Hintergrund: Die Wirksamkeit tiefenpsychologischer Psychotherapie ist allgemein gut belegt. Allerdings ist bislang wenig über die spezifische Wirksamkeit von integrativen Methoden bekannt, die sich im Rahmen tiefenpsychologischer Behandlungen Techniken der therapeutisch begleiteten Imagination und der Induktion hypnotischer Trancen bedienen.

Fragestellung: In einer prospektiven Längsschnittstudie mit naturalistischem Ein-Gruppen-Design wurde die Wirksamkeit der tiefenpsychologischen Methoden KIP, HY und ATP anhand einer Reihe unterschiedlicher Ergebnismaße untersucht. Dabei konnten weiterhin auch Moderatoren des Behandlungserfolgs identifiziert werden.

Methoden: Eine gemischte ambulante Stichprobe innerhalb der österreichischen Gesundheitsversorgung wurde zu Beginn der Behandlungen ($N = 353$) und nach 30 Monaten ($N = 109$) mittels des *Inventars Interpersoneller Probleme* (IIP), der *Kurzform des Inventars zur Persönlichkeitsorganisation* (IPO-16), des *Fragebogens zur Lebenszufriedenheit* (FLZ^M), der *Psychological Mindedness Scale* (PMS-D), sowie gesundheitsökonomischer Maße untersucht. Weiterhin wurde das Ausmaß der Symptombelastung (BSI, *Brief Symptom Inventory*) zu Beginn der Therapie und anschließend alle 6 Monate erfasst. Die Therapeutinnen und Therapeuten schätzten die Belastung der Patientinnen und Patienten mittels *Clinical Global Impression* (CGI-S) und *Globaler Erfassung des Funktionsniveaus* (GAF) am Anfang und Ende der Behandlung ein. Als Moderatoren des Behandlungserfolgs wurden die Psychological Mindedness (PMS-D) und das Ausmaß ihrer Persönlichkeitspathologie (ADP-IV, *Fragebogen zur Erfassung von DSM-IV Persönlichkeitsstörungen*) geprüft.

Ergebnisse: In KIP- und HY-Behandlungen konnten sowohl Psychological Mindedness als auch Persönlichkeitspathologie als Moderatoren des Behandlungserfolgs ausgewiesen werden. Je höher die Psychological Mindedness, desto stärker die Symptomreduktion innerhalb des ersten Jahres nach Therapiebeginn und je stärker die Persönlichkeitspathologie, desto geringer die Symptomreduktion.

Für ATP-Behandlungen konnten diese Effekte nicht nachgewiesen werden. In Bezug auf die Wirksamkeit zeigten sich statistisch und klinisch signifikante Verbesserungen in allen Maßen für KIP, auf allen Maßen mit Ausnahme der Anzahl der Krankenhaustage für HY und auf allen Maßen mit Ausnahme der Psychological Mindedness für ATP (die gesundheitsökonomischen Daten waren hier nicht auswertbar). Die Effektgrößen für die nachgewiesenen Effekte reichten von $d = .12$ bis $d = 1.76$ für die verschiedenen Skalen. Die Anzahl der Patientinnen und

Patienten mit Reliable Change reichte von 18% bis 82,5%. Bei allen Analysen wurde statistisch für die Anzahl der Sitzungen pro Therapie kontrolliert, sowie dafür, ob die Behandlungen innerhalb der Studienlaufzeit beendet worden waren. Für die Symptombelastung (BSI) ergab sich für alle drei Verfahren eine signifikante Verbesserung innerhalb der ersten sechs Monate nach Beginn der Behandlung, für KIP und HY erneut innerhalb der folgenden 18 bis 24 Monate. Dies weist auf eine regressive, aber anhaltende Symptomreduktion im Laufe längerer Behandlungen hin.

Schlussfolgerungen: Unter naturalistischen Bedingungen erweisen sich KIP, HY und ATP als wirksam in Bezug auf unterschiedliche Ergebnismaße. In KIP- und HY-Behandlungen wirkt sich die Psychological Mindedness der Patientinnen und Patienten positiv auf das Behandlungsergebnis aus, das Vorliegen einer ausgeprägte Persönlichkeitspathologie negativ.

Autoren

Dr. phil. Christian Sell
Abteilung Klinische Psychologie und Psychotherapie
Institut für Psychologie, Universität Kassel
34127 Kassel, Holländische Straße 36–38
Kontakt: csell@uni-kassel.de

Prof. Dr. phil. Svenja Taubner
Institut für Psychosoziale Prävention
Universitätsklinikum Heidelberg
69115 Heidelberg, Bergheimer Straße 54

Prof. Dr. phil. Heidi Möller
Abteilung Theorie und Methodik der Beratung
Institut für Psychologie, Universität Kassel
34127 Kassel, Holländische Straße 36–38