

DAS SCREENING FÜR SOMATOFORME STÖRUNGEN (SOMS): ENTWICKLUNG EINES NEUEN BEEINTRÄCHTIGUNGSINDEX

MARIA KLEINSTÄUBER¹, WINFRIED RIEF¹ & WOLFGANG HILLER²

¹ Philipps-Universität Marburg,
Abteilung für Klinische Psychologie und Psychotherapie
² Johannes Gutenberg-Universität Mainz,
Abteilung für Klinische Psychologie und Psychotherapie

ZUSAMMENFASSUNG: **Fragestellung:** Im Rahmen der vorliegenden Studie soll ein neuer Auswertungsindex (SOMS-PSDI-1/-2) für das Screening für Somatoforme Störungen (SOMS-7T) vorgestellt werden und in drei unabhängigen Stichproben mit den bereits bestehenden Indizes des Messinstruments verglichen werden. **Methode:** Die Stichproben umfassen jeweils Patienten mit einer diagnostizierten somatoformen Störung aus einer randomisiert-kontrollierten Studie ($N = 82$), aus einer konsekutiv-ambulanten Behandlung ($N = 235$) und aus einer konsekutiv-stationären Therapie ($N = 239$). Als Kennwerte für das Therapieergebnis wurden Effektstärken und der Reliable Change Index ermittelt. **Ergebnisse:** Im Vergleich zu den klassischen Indizes des SOMS-7T unterlagen – in Abhängigkeit von der für die Berechnung verwendeten Standardabweichung – die Effektstärken basierend auf dem SOMS-PSDI-2 innerhalb der Stichproben starken Schwankungen. Die Rate der Patienten, die den Reliable Change Index erfüllten, lagen für den SOMS-PSDI-1/-2 am höchsten. Korrelationsstatistische Analysen zeigen, dass die traditionellen Indizes des SOMS-7T stichprobenübergreifend sehr hoch interkorrelieren, während sie mit dem SOMS-PSDI-1/2 deutlich geringere Zusammenhänge aufweisen. **Schlussfolgerungen:** Der SOMS-PSDI stellt eine wichtige neue Auswertungsmethode alternativ zu den bestehenden Indizes des SOMS-7T dar. Die Ergebnisse weisen jedoch auch daraufhin, dass breit-angelegte, unspezifische Beschwerdelisten als primäres Ergebnismaß in der Therapie somatoformer Störungen kritisch zu hinterfragen sind und ggf. durch spezifischere, änderungssensitivere Instrumente ergänzt werden sollten.

SCHLÜSSELWÖRTER: Screening für Somatoforme Störungen, somatoforme Störungen, Distress, Effektstärke, Reliable Change Index

THE SCREENING OF SOMATOFORM DISORDERS (SOMS): DEVELOPMENT OF A NEW DISTRESS INDEX

ABSTRACT: **Aims:** The aim of the current study is to introduce a new distress index (SOMS-PSDI-1/-2) for the Screening of Somatoform Disorders (SOMS-7T). The new index shall be compared with existing scores of the SOMS-7T in three independent samples. **Methods:** The samples include patients diagnosed with a somatoform disorder from a randomized-controlled trial ($N = 82$), from a consecutive-outpatient treatment ($N = 235$), and from a consecutive-inpatient therapy ($N = 239$). Effect sizes and the reliable change index were calculated. **Results:** In contrast to the classical indices of the SOMS-7T, the effect sizes based on the SOMS-PSDI-2 varied considerably in each sample in dependence of the standard deviation used for the calculations. The rates of patients fulfilling the RCI were highest for the SOMS-PSDI-1/-2. Results of correlational analyses demonstrated that the traditional scores of SOMS-7T were highly intercorrelated, however they had considerably lower associations with the new index SOMS-PSDI. **Conclusions:** The SOMS-PSDI constitutes an important new score of the SOMS-7T. The current findings encourage however to question the general use of broadly conceived, unspecific symptom lists as primary outcome in the therapy of somatoform disorders. Maybe more specific and responsive instruments should be considered for measuring therapy efficacy.

KEY WORDS: Screening for Somatoform Disorders, somatoform disorders, distress, effect size, reliable change index

1 EINFÜHRUNG

Multiple unerklärte Körpersymptome stellen in der Allgemeinbevölkerung keine Seltenheit dar. Hiller und Rief (2006) ermittelten in einer deutschen Repräsentativstichprobe eine Punktprävalenz von 81,6% für unerklärte Körperbeschwerden einhergehend mit einer leichten Beeinträchtigung. Die Häufigkeit krankheitswertiger somatoformer Symptome mit hohem Beeinträchtigungsgrad hingegen ist deutlich niedriger. Eine besonders schwerwiegende Form – die Somatisierungsstörung – tritt laut epidemiologischen Untersuchungen mit einer Lebenszeitprävalenz unter 1% nur sehr selten auf (z.B. Robins & Regier, 1991). Diese Befunde verdeutlichen, dass eine klare diagnostische Einordnung der Körperbeschwerden in klinisch relevante Symptome bzw. klinisch nicht relevante

Alltagsphänomene eine wichtige Rolle in der psychotherapeutischen Praxis spielt.

Als unterstützendes Instrument für die diagnostische Abklärung somatoformer Störungen kommen meist sogenannte Beschwerdelisten zum Einsatz. Beispiele dafür sind u.a. das Screening für Somatoforme Störungen (SOMS; Rief & Hiller, 2008), die Somatisierungsskala des Patient Health Questionnaire (PHQ-15; Kroenke, Spitzer & Williams, 2002) oder die Somatisierungsskala der Symptom-Checkliste (SCL-90-R; Derogatis, 1983). Gemeinsam ist diesen Instrumenten eine Auflistung von körperlichen Beschwerden, die von den Betroffenen auf einer Skala dahingehend eingeschätzt werden sollen, wie stark sie Leidensdruck bzw. Beeinträchtigungen hervorgerufen haben. Als Auswertungsindex wird dann in der Regel ein Summen- oder Mittelwert aus allen Items berechnet.

Diese Art der Auswertung birgt ein Problem, das am Beispiel des SOMS-7T (Rief & Hiller, 2008) verdeutlicht werden soll. Bei diesem Fragebogen wird der Betroffene instruiert, den Beeinträchtigungsgrad in Bezug auf verschiedene Körperbeschwerden über die vergangenen sieben Tage einzuschätzen. Die Autoren (Rief & Hiller, 2008) leiten zwei Indizes ab: die Beschwerdenanzahl und den Belastungsgrad (Intensitätsindex). Das erste Ergebnismaß basiert auf der Anzahl der Beschwerden, die mindestens mit einer mittelmäßigen Beeinträchtigung bewertet wurden. Beim zweiten Index wird ein Summen- oder Mittelwert über die erhaltenen Intensitätsscores aller Items berechnet. Problematisch am Intensitätsindex ist, dass dieser die Gesamtbelastung eines Patienten von der Zahl der berichteten Symptome abhängig macht. Dies stimmt jedoch nicht mit der klinischen Wirklichkeit überein. So kann bereits das Vorhandensein von ein oder zwei schwerwiegenden Symptomen zu massiven Beeinträchtigungen bis hin zur Arbeitsunfähigkeit oder vorzeitigen Berentung führen. Ein Beispiel wäre ein Patient mit schwersten Kopf- und Rückenschmerzen, der beide Symptome mit 4 bewertet und somit einen Intensitätsindex von 8 erreichen würde. Den gleichen Wert im Intensitätsindex hätte ein anderer Patient mit vier mittelgradig ausgeprägten (mit 2 kodierten) Symptomen, obwohl bei diesem Patienten womöglich eine deutlich geringere Gesamtbelastung besteht. Ein ähnliches Problem ergibt sich für die Beschwerdenanzahl, da mit dieser lediglich die Anzahl der Symptome aber nicht deren Schweregrad erfasst wird. Bezüglich des vorliegenden Beispiels würde dementsprechend die Beschwerdenanzahl von 4 für den zweiten Patienten zu Unrecht eine wesentlich stärker ausgeprägte Störung indizieren als beim ersten Patienten

mit einem Score von nur 2. Ähnliche Verzerrungen können sich beim Vergleich von therapiebedingten Veränderungen ergeben.

Als primäres Ziel der vorliegenden Studie soll daher ein neuartiger Beeinträchtigungsindex für den SOMS-7T vorgestellt werden, mit dem die genannten Auswertungsprobleme gelöst werden können. Diesbezüglich orientieren wir uns am „Positive Symptom Distress Index (PSDI)“ des Brief Symptom Inventory (BSI; Derogatis, 1992; Franke, 2000), der in der internationalen Forschungsliteratur eine sehr große Akzeptanz gefunden hat. Der hier neu vorgestellte Auswertungskennwert für den SOMS-7T soll des Weiteren in drei unabhängigen Stichproben von Patienten mit somatoformer Symptomatik mit den anderen bereits existierenden Indizes verglichen werden. Dieser Vergleich soll basierend auf dem Effektstärkenmaß Cohens d (Cohen, 1977) sowie dem Reliable Change Index (RCI; Jacobson & Truax, 1991) erfolgen.

2

METHODIK

2.1

STICHPROBEN UND SETTING

Die Auswertung des SOMS-7T in der vorliegenden Studie erfolgt für drei unabhängige Stichproben von Patienten mit chronischen, medizinisch unerklärten Körperbeschwerden. Übergreifend für alle Stichproben gilt, dass lediglich Patienten aufgenommen wurden, die entsprechend dem DSM-IV (Saß, Wittchen, Zaudig & Houben, 2003) oder der ICD-10 (World Health Organization, 1992) die Kriterien einer Somatisierungsstörung, einer undifferenzierten somatoformen Störung, einer somatoformen autonomen Funktionsstörung, einer so-

matoförmigen Schmerzstörung, einer Konversionsstörung, einer nicht näher bezeichneten somatoförmigen Störung oder einer Hypochondrie erfüllten. Komorbide psychische Störungen wurde nicht als Ausschlusskriterium definiert. Ebenfalls übergreifend für alle drei Stichproben wurde ein ausführliches diagnostisches Interview – basierend auf dem Strukturierten Klinischen Interview für Psychische Störungen (SKID; Wittchen, Zaudig & Fydrich, 1997) bzw. den Internationalen Diagnosen Checklisten für DSM-IV und ICD-10 (IDCL; Hiller, Zaudig & Mombour, 1997) – zur Abklärung der somatoforme(n) bzw. ggf. komorbiden Diagnose(n) zu Therapiebeginn durchgeführt.

Die erste Stichprobe (*RCT-Stichprobe*) schließt 82 Patienten ein, die an der Poliklinischen Institutsambulanz für Psychotherapie der Johannes-Gutenberg-Universität Mainz im Rahmen einer randomisiert-kontrollierten Studie zur Evaluation einer 20-stündigen ambulanten kognitiven Verhaltenstherapie (KVT) speziell für medizinisch unerklärte Körperbeschwerden rekrutiert wurden. Das manualisierte Therapieprogramm (Kleinstäuber, Thomas, Witthöft & Hiller, 2012), auf dessen Basis die Behandlung erfolgte, umfasst Elemente zur Psychoedukation, zur Stressbewältigung, zu Ablenkungsstrategien, zu kognitiven Techniken und zum Abbau von dysfunktionalem Krankheitsverhalten. Die Mainzer Hochschulambulanz ist eine nach DIN-EN ISO 9001 zertifizierte Hochschulambulanz und dient der Psychotherapeutenausbildung sowie der Forschung und Lehre (Hiller, Bleichhardt et al., 2006).

Die zweite Stichprobe (*konsekutiv-ambulante Stichprobe*) umfasst 235 Patienten, die im Rahmen des fortlaufenden Routinebetriebs dieser Ambulanz zwischen 2001 und 2012 behandelt wurden, jedoch nicht Teil der oben genannten

kontrollierten Therapiestudie waren. Die Patienten wurden einzeltherapeutisch oder in einer Kombination mit Gruppentherapie in Orientierung am kognitiv-verhaltenstherapeutischen Konzept von Rief und Hiller (2011) behandelt. Sie erhielten durchschnittlich 36 Therapiesitzungen à 50 Minuten ($SD = 18,40$).

Die dritte Stichprobe (konsekutiv-stationäre Stichprobe) schließt 239 Patienten ein, die 1995 bis 1996 im Rahmen einer Kosten-Nutzen-Therapiestudie in der psychosomatischen Klinik Roseneck in Prien untersucht und behandelt wurden (Hiller, Fichter & Rief, 2003). Die Teilnehmer erhielten ebenfalls kognitiv-verhaltenstherapeutische Interventionen, die Einzel- und Gruppentherapiesitzungen, Entspannungstraining und Biofeedback, Selbstsicherheitstraining und ggf. bei Vorliegen komorbider Störungen weitere spezifische Interventionen umfassten.

Die demografischen und störungsbezogenen Merkmale aller drei Stichproben werden in Tabelle 1 beschrieben. Ein post-hoc Scheffé-Test ergab hinsichtlich des Alters, dass die RCT-Stichprobe signifikant jünger ist als die konsekutiv-stationäre Stichprobe (Mittlere Differenz = 6,37; $p < ,001$). Des Weiteren ergab sich ein signifikanter Unterschied zwischen den Gruppen hinsichtlich des Bildungsstatus sowie der Verteilung der Diagnosen somatoformer Störungen (siehe Tabelle 1). Die Teilnehmer in beiden ambulanten Stichproben wiesen ein höheres Bildungsniveau auf als die stationären Patienten. Des Weiteren schlossen beide konsekutive Stichproben deutlich mehr Patienten mit einer Somatisierungsstörung, mit einer somatoförmigen Schmerzstörung sowie mit einer Hypochondrie ein als die RCT-Gruppe.

Tabelle 1: Demographische Variablen und Störungsdiagnosen in den drei Stichproben

	RCT-Stichprobe N = 82		Konsekutiv-ambulante Stichprobe N = 235		Konsekutiv-stationäre Stichprobe N = 239		Zwischen- gruppenunter- schiede Teststatistiken
Anteil Frauen, n, %	52	63,4	159	67,7	155	64,9	Chi ² (2, N = 556) = 0,66; p = ,718
Alter, M, SD	39,84	12,44	39,61	12,00	45,97	10,69	F (2,553) = 20,41; p < ,001
Bildung, n, %							Chi ² (12, N = 550) = 98,73; p < ,001
Noch in Schulausbildung	2	2,4	2	0,9	0	0,0	
Haupt-/ Sonderschulabschluss	12	14,6	51	21,7	97	40,6	
Realschulabschluss	21	25,6	59	25,1	51	21,3	
Abitur	17	20,7	49	20,9	11	4,6	
Anderer Schulabschluss	10	12,2	12	5,1	16	6,7	
Kein Abschluss	0	0,0	5	2,1	27	11,3	
Hochschul-/ Universitätsabschluss	20	24,4	57	24,3	31	13,0	
Keine Angaben	0	0,0	0	0,0	6	2,5	
Somatoforme Diagnosen, n, %							Chi ² (12, N = 556) = 172,04; p < ,001
Somatisierungsstörung	7	8,5	43	18,3	56	23,4	
Undifferenzierte somatoforme Störung	67	81,7	53	22,6	87	36,4	
Somatoforme autonome Funktionsstörung	3	3,7	0	0,0	0	0,0	
Somatoforme Schmerzstörung	3	3,7	66	28,1	82	34,3	
Konversionsstörung	0	0,0	1	0,4	4	1,7	
Somatoforme Störung NNB	1	1,2	6	2,6	0	0,0	
Hypochondrie	1	1,2	66	28,1	10	4,2	
Komorbiditäten, n, %							
Weitere somatoforme Störung	0	0,0	16	6,8	64	26,8	
Affektive Störung	29	35,4	159	67,7	184	77,0	
Angststörung	10	12,2	83	35,3	106	44,4	
Substanzbezogene psychische Störung	2	2,4	13	5,5	27	11,3	
Psychotische Störungen	0	0,0	0	0,0	2	0,8	
Persönlichkeitsstörung	0	0,0	26	11,1	57	23,9	
Andere psychische Störungen	0	0,0	14	6,0	17	7,1	

2.1 SCREENING FÜR SOMATOFORME STÖRUNGEN (SOMS)

Das Screening für Somatoforme Störungen (SOMS; Rief & Hiller, 2008), ein im deutschsprachigen Raum besonders häufig genutztes Instrument zur Erfassung somatoformer Beschwerden, umfasst 53 Items, die alle körperlichen Symptome berücksichtigen, welche für eine Somatisierungsstörung nach DSM-IV (Saß et al., 2003) bzw. ICD-10 (World Health Organization, 1992) sowie für die somatoforme autonome Funktionsstörung nach ICD-10 relevant sind. Es liegt in zwei Versionen vor: In der Traitversion SOMS-2 wird die Beeinträchtigung durch körperliche Beschwerden während der letzten zwei Jahre erfasst und in der State-Version SOMS-7T wird sie während der vergangenen sieben Tagen gemessen. Für die vorliegende Studie wurde das SOMS-7T zugrunde gelegt. Wie stark die Patienten unter den einzelnen Symptomen gelitten haben, müssen sie für jedes Item auf einer 5-stufigen Skala (0 = gar nicht; 4 = sehr stark) einschätzen. Wie bereits erwähnt stellt das SOMS-7T zwei Auswertungsindizes zur Verfügung – die Beschwerdenanzahl und den Intensitätsindex. Das SOMS-7T weist eine zufriedenstellende 4-Monats-Retestreliabilität auf (Rief & Hiller, 2008): Für den Intensitätsindex ergab sich ein Wert von $r_{tt} = 0,76$ und für die Beschwerdenanzahl $r_{tt} = 0,71$.

Bei der Auswertung des SOMS erscheint nicht einheitlich geregelt und daher problematisch der Umgang mit fehlenden Werten sowie mit den sechs geschlechterspezifischen Items. Ein nicht selten zu beobachtendes Problem besteht darin, dass beim Nichtvorhandensein bestimmter Symptome Items nicht mit „Null“ beantwortet, sondern ganz

ausgelassen werden und daher der Eindruck von fehlenden Werten entsteht. In der vorliegenden Arbeit wurde entsprechend den Hinweisen im Fragebogenmanual (Rief & Hiller, 2008) ein Plausibilitätscheck durchgeführt. Bei Patienten, die keines der Items mit „Null“ bewertet hatten, wurden sämtliche fehlende Werte nicht als solche gezählt und durch eine „Null“ ersetzt. Bei Patienten, die vereinzelt Items auch mit Null bewertet hatten, wurden fehlende Werte als solche gekennzeichnet. Bei einer Missingrate von mehr als 10 Items wurde auf die Berechnung des Index verzichtet. Des Weiteren erscheint problematisch, dass von den 53 Items des SOMS-7T eines männerspezifisch und fünf frauenspezifisch sind. Uns erschien die methodisch sauberste Lösung, mit diesem Problem umzugehen, im Ausschluss der geschlechterspezifischen Items aus der Indexberechnung. Um die Berechnung der Indizes jedoch möglichst eng an den Auswertungsvorgaben des Fragebogen-Manuals (Rief & Hiller, 2008) auszurichten, wurde in der vorliegenden Studie für den Intensitätsindex ein Mittelwert über alle Items (ausschließlich der fehlenden Werte) unabhängig vom Geschlecht berechnet. Die gegengeschlechtsspezifischen Items erhielten dabei immer den Wert Null.

In allen einbezogenen Stichproben wurde der SOMS-7T zu zwei Messzeitpunkten erhoben: Als Prämessung gilt der Erhebungszeitpunkt unmittelbar vor der Therapie. Die Postmessung erfolgte in der RCT-Stichprobe nach der 20. Therapiesitzung und in der konsekutiv-ambulanten bzw. -stationären Stichprobe unmittelbar bei oder nach Therapieabschluss.

2.2 SOMS-7T:

EIN NEUER BEEINTRÄCHTIGUNGSINDEX

Wie bereits erwähnt, wird in der vorliegenden Studie ein neuer Beeinträchtigungsindex für das SOMS-7T in Anlehnung an das Brief Symptom Inventory (BSI; Derogatis, 1992; Franke, 2000) berechnet. Das BSI umfasst 53 Items zur Messung von psychischer Belastung in den vergangenen 7 Tagen, auf deren Basis drei globale Kennwerte berechnet werden können. Der sogenannte *Global Severity Index (GSI)* misst die grundsätzliche psychische Belastung und ist bezüglich der Berechnung identisch zum Intensitätsindex des SOMS-7T (berechnet als Mittelwert). In den folgenden Analysen wird daher der Intensitätsindex mit *SOMS-GSI* abgekürzt.

Der *Positive Symptom Total (PST)* des BSI gibt Auskunft über die Anzahl der Symptome, bei denen eine Belastung vorliegt. Dieser Kennwert wird genauso wie der Index „Beschwerdenanzahl“ des SOMS-7T berechnet (SOMS-PST). Hierbei ergibt sich lediglich die Frage, ab wann ein Symptom als „belastend“ definiert wird. Beim BSI werden die Items vergleichbar zum SOMS-7T auf einer 5-stufigen Skala (0 = überhaupt nicht; 4 = sehr stark) bewertet. Für den PST des BSI werden Items gezählt, die mindestens mit einem Wert 1 (= ein wenig Leidensdruck) eingeschätzt wurden. Beim SOMS-7T hingegen werden – nach den Empfehlungen von Hiller und Rief (2008) sowie dem Hinweis, dass auch bei gesunden Menschen die Prävalenz leichter Körpersymptome sehr hoch ist (Hiller, Rief et al., 2006) – nur Items mit einem Mindestwert von 2 (= mittelmäßige Beeinträchtigung) berücksichtigt. In der vorliegenden Studie wurde für den SOMS-7T der Index „Beschwerdenanzahl“ zweimal berechnet: entweder ba-

sierend auf Items, die mindestens mit 1 bewertet wurden (*SOMS-PST-1*), oder basierend auf Items, die mindestens mit 2 eingestuft wurden (*SOMS-PST-2*).

Als dritter Index für das BSI gibt es den *Positive Symptom Distress Index (PSDI)*, der den Grad der Belastung über alle berichteten Symptome hinweg misst. Daher wird beim PSDI der Summenwert aller Items nur durch die Anzahl der Items, die mindestens mit 1 bewertet wurden, dividiert, und nicht durch die Gesamtzahl aller 53 Items. Ein Äquivalent zum PSDI liegt bisher für den SOMS-7T nicht vor. Er stellt jedoch eine Indexvariante dar, mit der die erläuterten Probleme der bereits bestehenden SOMS-Indizes gelöst werden könnten. In der vorliegenden Studie soll der PSDI daher auf den SOMS-7T übertragen werden. Analog zum SOMS-PST wird im Folgenden der PSDI ebenfalls zweimal berechnet, jeweils unter Berücksichtigung von Items mit den Mindestwerten 1 (*SOMS-PSDI-1*) bzw. 2 (*SOMS-PSDI-2*).

2.4 STATISTISCHE ANALYSEN

Alle Analysen wurden mit Hilfe von IBM SPSS Statistics 20[®] sowie Microsoft Office Excel 2010[®] durchgeführt. Basierend auf den Veränderungen zwischen der Prä- und Postmessung bezüglich der jeweiligen SOMS-Indizes wurden Effektstärken mit dazugehörigen 95%-Konfidenzintervallen ermittelt. In Orientierung an der Formel nach Cohen (1977) dividiert man die Differenz aus dem Prä- und Postmittelwert durch ein Streuungsmaß, bezüglich dessen jedoch Uneinigkeit besteht. Während z.B. Kazis, Anderson und Meenan (1989) die Standardabweichung der Prämessung präferieren, bevorzugen Hedges und Olkin (1985) bzw. Hartmann und Herzog (1995) die Verwen-

dung einer gepoolten Prä-Post-Streuung. In der vorliegenden Arbeit wurden beide Versionen berechnet. Zur Interpretation wurde die Konvention nach Cohen (1992) herangezogen, wonach Effekte ab 0,20 als klein, ab 0,50 als moderat und ab 0,80 als groß interpretiert werden.

Als weiterer Kennwert wurde der Reliable Change Index (RCI) nach Jacobson, Follette und Revenstorf (1984) für die verschiedenen SOMS-Indizes berechnet. Zur Ermittlung des RCI wird die Differenz des Prä- und des Postmesswerts durch den Standardmessfehler der Differenzwerte geteilt. Für die Berechnung des Standardmessfehlers werden die Standardabweichung der Referenzpopulation sowie die Reliabilität des Messinstruments benötigt. Bezüglich der Bestimmung der Reliabilität wird empfohlen, ein Maß für die interne Konsistenz zu nutzen, da Retest-Reliabilitätswerte zu einer Unterschätzung aufgrund von Zustandsänderungen zwischen den Messzeitpunkten führen können (Lambert & Ogles, 2009). Für den SOMS-7T liegen jedoch nur Retest-Reliabilitätswerte vor (Rief & Hiller, 2008), auf die für die Berechnung des RCI in der vorliegenden Studie zurückgegriffen wurde (siehe Abschnitt 2.1). Für den RCI des SOMS-PST-1 wird in der vorliegenden Studie die Retest-Reliabilität der Beschwerdenanzahl angewendet, für den RCI des SOMS-PSDI-1 und -2 wird der Reliabilitätswert des Intensitätsindex verwendet.

Die Berechnung der Kennwerte erfolgte erstens auf Basis eines Datensatzes, in dem nur Patienten mit vollständigen Daten zu beiden Messzeitpunkten berücksichtigt wurden (Completer-Datensatz). Zweitens wurden die Kennwerte für einen Datensatz, bei dem fehlende Postwerte anhand der Last observation carried forward (LOCF)-Methode durch die Prä-Werte ersetzt wurden, ermittelt. Zur Bestimmung statistischer Zusam-

menhänge zwischen den Indices des SOMS-7T wurden Pearson Produkt-Moment-Korrelationskoeffizienten berechnet.

3

ERGEBNISSE

3.1

MITTELWERTE UND EFFEKTSTÄRKEN

Für den Datensatz, der ausschließlich vollständige Prä-Post-Messungen einschließt, werden die Mittelwerte, Standardabweichungen und Effektstärken (ES) in Tabelle 2 zusammengefasst. Die Effektgrößen – basierend auf Cohens d , das mit der gepoolten Prä-Post-Standardabweichung berechnet wurde – fielen für die verschiedenen SOMS-Indizes innerhalb einer Stichprobe überwiegend homogen aus. Während in der RCT-Stichprobe ($0,68 \leq d \leq 0,83$) die ES im Übergangsbereich zwischen mittleren und großen Effekten bzw. in der konsekutiv-stationären Stichprobe ($0,35 \leq d \leq 0,51$) an der Grenze zwischen kleinen und mittleren Effekten lagen, fanden sich für die konsekutiv-ambulante Stichprobe ES ausschließlich im kleinen Bereich ($0,31 \leq d \leq 0,47$). Unter Verwendung der Prä-Standardabweichung zur Berechnung von Cohens d traten beim SOMS-PSDI-2 über alle drei Stichproben deutliche Veränderungen der Effekte auf. So entwickelte sich in der RCT-Stichprobe die mittelgroße ES ($d = 0,68$) zu einer sehr großen ES ($d = 1,26$). In der konsekutiv-ambulanten Stichprobe erhöhte sich der kleine Effekt ($d = 0,47$) zu einem großen Effekt ($d = 0,81$). In der konsekutiv-stationären Stichprobe konnte unter Verwendung der Prä-Standardabweichung eine mittlere ES ($d = 0,59$) im Vergleich zu einer kleinen ES ($d = 0,43$), basierend auf der gepoolten Standardabweichung, identifiziert werden. In allen Fällen war

Tabelle 2: Mittelwerte, Standardabweichungen und Effektstärken für jede Stichprobe basierend auf Completer-Datensätzen

	$M_{\text{prä}}$	$SD_{\text{prä}}$	M_{post}	SD_{post}	d_{gepoolt}	95%-KI _{gepoolt}	$d_{\text{Prä-SD}}$	95%-KI _{Prä-SD}
RCT-Stichprobe, N = 42								
SOMS-GSI	0,59	0,35	0,32	0,32	0,81	0,34-1,27	0,77	0,31-1,23
SOMS-PST-1	15,38	7,39	9,50	7,43	0,79	0,33-1,25	0,80	0,34-1,26
SOMS-PST-2	9,48	6,23	4,60	5,50	0,83	0,37-1,29	0,78	0,32-1,24
SOMS-PSDI-1	2,03	0,53	1,61	0,56	0,77	0,31-1,23	0,79	0,33-1,25
SOMS-PSDI-2	2,59	0,42	2,06	1,02	0,68	0,23-1,13	1,26	0,76-1,77
Konsekutiv-ambulante Stichprobe, N = 157								
SOMS-GSI	0,78	0,54	0,58	0,51	0,38	0,16-0,61	0,37	0,15-0,60
SOMS-PST-1	21,58	10,27	17,06	10,42	0,44	0,21-0,66	0,44	0,21-0,67
SOMS-PST-2	11,83	9,24	8,09	8,75	0,42	0,19-0,64	0,40	0,18-0,63
SOMS-PSDI-1	1,80	0,48	1,64	0,56	0,31	0,08-0,53	0,33	0,11-0,56
SOMS-PSDI-2	2,49	0,43	2,14	0,96	0,47	0,24-0,70	0,81	0,58-1,05
Konsekutiv-stationäre Stichprobe, N = 206								
SOMS-GSI	0,72	0,46	0,53	0,39	0,45	0,25-0,64	0,41	0,22-0,61
SOMS-PST-1	18,28	9,47	15,02	9,15	0,35	0,15-0,55	0,34	0,15-0,54
SOMS-PST-2	11,74	7,90	7,91	7,02	0,51	0,31-0,71	0,48	0,29-0,68
SOMS-PSDI-1	1,97	0,55	1,72	0,56	0,45	0,25-0,65	0,45	0,26-0,65
SOMS-PSDI-2	2,46	0,54	2,14	0,90	0,43	0,23-0,63	0,59	0,39-0,79

Anmerkungen: d_{gepoolt} = Prä-Post-Effektstärke berechnet mit gepoolter Prä-Post-Standardabweichung. $d_{\text{Prä-SD}}$ = Prä-Post-Effektstärke berechnet mit Prä-Standardabweichung. 95%-KI = 95%-Konfidenzintervall von d .

der Grund für den Effektzuwachs beim SOMS-PSDI-2, dass die Standardabweichung zum Post-Messzeitpunkt deutlich größer war als die zum Prä-Messzeitpunkt.

Die analogen Statistiken für den LOCF-Datensatz werden in Tabelle 3 zusammengefasst. Die ES basierend auf Cohens d , berechnet mit der gepoolten Standardabweichung, lagen über alle drei Stichproben hinweg im kleinen Be-

reich ($0,23 \leq d \leq 0,44$). Unter Verwendung der Prä-Standardabweichung zur Berechnung der ES ergaben sich erneut über alle drei Stichproben kaum Veränderungen, außer für den SOMS-PSDI-2. In der RCT-Stichprobe ($d = 0,69$) und in der konsekutiv-ambulanten Stichprobe ($d = 0,53$) stieg dieser zu einem mittleren Effekt an. In der konsekutiv-stationären Stichprobe lag er an der Grenze zum mittleren Effekt ($d = 0,48$).

Tabelle 3: Mittelwerte, Standardabweichungen und Effektstärken für jede Stichprobe basierend auf LOCF-Datensätzen

	$M_{\text{prä}}$	$SD_{\text{prä}}$	M_{post}	SD_{post}	d_{gepoolt}	95%-KI _{gepoolt}	$d_{\text{Prä-SD}}$	95%-KI _{Prä-SD}
RCT-Stichprobe, N = 82								
SOMS-GSI	0,65	0,41	0,51	0,43	0,33	0,02-0,64	0,34	0,03-0,65
SOMS-PST-1	16,37	8,25	13,35	9,11	0,35	0,04-0,66	0,37	0,05-0,68
SOMS-PST-2	10,55	7,01	8,05	7,50	0,34	0,03-0,65	0,36	0,05-0,67
SOMS-PSDI-1	2,03	0,49	1,82	0,55	0,40	0,09-0,72	0,43	0,12-0,74
SOMS-PSDI-2	2,58	0,39	2,31	0,81	0,42	0,11-0,74	0,69	0,37-1,02
Konsekutiv-ambulante Stichprobe, N = 235								
SOMS-GSI	0,79	0,52	0,65	0,51	0,27	0,09-0,45	0,27	0,09-0,45
SOMS-PST-1	22,17	10,30	19,15	10,79	0,29	0,10-0,47	0,29	0,11-0,48
SOMS-PST-2	12,02	9,13	9,52	9,03	0,28	0,09-0,46	0,27	0,09-0,46
SOMS-PSDI-1	1,78	0,45	1,67	0,51	0,23	0,05-0,41	0,24	0,06-0,43
SOMS-PSDI-2	2,46	0,45	2,22	0,84	0,36	0,17-0,54	0,53	0,35-0,72
Konsekutiv-stationäre Stichprobe, N = 239								
SOMS-GSI	0,71	0,45	0,54	0,39	0,40	0,22-0,59	0,38	0,20-0,56
SOMS-PST-1	18,15	9,48	15,34	9,24	0,30	0,12-0,48	0,30	0,12-0,48
SOMS-PST-2	11,56	7,81	8,26	7,08	0,44	0,26-0,63	0,42	0,24-0,61
SOMS-PSDI-1	1,95	0,54	1,74	0,55	0,39	0,20-0,57	0,39	0,21-0,57
SOMS-PSDI-2	2,43	0,56	2,16	0,87	0,37	0,19-0,55	0,48	0,30-0,67

Anmerkungen: d_{gepoolt} = Prä-Post-Effektstärke berechnet mit gepoolter Prä-Post-Standardabweichung.
 $d_{\text{Prä-SD}}$ = Prä-Post-Effektstärke berechnet mit Prä-Standardabweichung. 95%-KI = 95%-Konfidenzintervall von d .

3.2 RELIABLE CHANGE INDEX

Die Reliable Change Indizes für jede der drei Stichproben basierend auf dem Completer- wie auch dem LOCF-Datensatz werden in Tabelle 4 zusammengefasst. Wie zu erwarten, fiel der RCI stichprobenunabhängig höher aus, wenn

er basierend auf den vollständigen Datensätzen im Gegensatz zu den LOCF-Datensätzen berechnet wurde. Des Weiteren zeigte sich unabhängig vom Therapiesetting und dem Datensatz, dass die Rate an Patienten, die den RCI erfüllten, am höchsten für den SOMS-PSDI-1 sowie -2 ausfielen.

Tabelle 4: *Reliable Change Index für jede Stichprobe basierend auf Completer- bzw. LOCF-Datensätzen*

	Completer		LOCF	
	RCI	RCI %	RCI	RCI %
RCT-Stichprobe, N = 82				
SOMS-GSI	± 0,48	14,3%	± 0,55	3,7%
SOMS-PST-1	± 11,04	19,0%	± 12,31	7,3%
SOMS-PST-2	± 9,30	7,1%	± 10,46	3,7%
SOMS-PSDI-1	± 0,72	26,2%	± 0,66	13,4%
SOMS-PSDI-2	± 0,58	26,2%	± 0,53	14,6%
Konsekutiv-ambulante Stichprobe, N = 235				
SOMS-GSI	± 0,73	6,4%	± 0,73	4,7%
SOMS-PST-1	± 15,33	6,4%	± 15,38	4,3%
SOMS-PST-2	± 13,79	6,4%	± 13,62	4,3%
SOMS-PSDI-1	± 0,65	11,5%	± 0,62	9,8%
SOMS-PSDI-2	± 0,59	21,0%	± 0,61	13,6%
Konsekutiv-stationäre Stichprobe, N = 239				
SOMS-GSI	± 0,62	7,8%	± 0,61	7,1%
SOMS-PST-1	± 14,13	6,3%	± 14,15	5,4%
SOMS-PST-2	± 11,80	9,2%	± 11,65	10,5%
SOMS-PSDI-1	± 0,75	17,5%	± 0,73	15,5%
SOMS-PSDI-2	± 0,73	17,5%	± 0,76	14,6%

Anmerkungen: RCI = Reliable Change Index. RCI % = prozentualer Anteil an Patienten, die den RCI erfüllen.

3.3

INTERKORRELATIONEN ZWISCHEN DEN SOMS-INDIZES

Unabhängig von der Stichprobe und vom Messzeitpunkt zeigten sich sehr hohe, signifikante Interkorrelationen der traditionellen Indices des SOMS (SOMS-GSI, SOMS-PST-1 und -2) zwischen $r = ,83$ und $r = ,97$ (p -Werte $< ,001$). Deut-

lich niedrigere Korrelationskoeffizienten ergaben sich für die Zusammenhänge zwischen dem traditionellen Score SOMS-GSI und dem neuartigen Index SOMS-PSDI-1 bzw.-2 in der konsekutiv-ambulanten wie auch konsekutiv-stationären Stichprobe ($,50 \leq r \leq ,78$; p -Werte $< ,001$). Besonders niedrig fielen diese Koeffizienten in der RCT-Stichprobe aus (SOMS-PSDI-1: $r = ,38$; $p < ,050$;

SOMS-PSDI-2: $r = ,28$; $p < ,069$). Auch der traditionelle Score SOMS-PST wies unabhängig von der Stichprobe und vom verwendeten Cut-off (d.h. SOMS-PST-1 wie auch -2) überwiegend deutlich niedrigere Assoziationen mit dem neuen Index SOMS-PSDI-1 bzw. -2 auf ($-,06 \leq r \leq ,52$; $,000 < p \leq ,699$). Hierbei stellte lediglich der Zusammenhang zwischen SOMS-PST-2 und SOMS-PSDI-1 eine Ausnahme dar, für den moderate bis hohe Korrelationskoeffizienten identifiziert werden konnten ($,42 \leq r \leq ,82$; $,000 < p \leq ,006$). SOMS-PSDI-1 und -2 korrelierten stichprobenübergreifend und unabhängig vom Messzeitpunkt hoch miteinander ($,70 \leq r \leq ,80$; $p < ,001$)

4

DISKUSSION

Als primäres Ziel ging die vorliegende Studie der Frage nach, inwieweit ein neuartiger Auswertungsindex (SOMS-PSDI-1/-2) und die bisher existierenden Auswertungsmethoden des SOMS-7T zu vergleichbaren Ergebnissen führen. Der Vergleich der verschiedenen Indizes erfolgte mit verschiedenen Auswertungskennwerten (Effektstärke und Reliable Change Index) in drei unabhängigen Stichproben.

In Bezug auf die Effektstärken, die mit einer gepoolten Prä-Post-Standardabweichung berechnet wurden, zeigte sich innerhalb der Patientengruppen ein homogenes Bild. Unabhängig vom Index ergaben sich für die spezifisch auf somatoforme Störungen zugeschnittene Kurzzeittherapie mittlere bis große Effekte, während für die konsekutiv-ambulante und -stationäre Stichprobe überwiegend kleine Effekte identifiziert wurden. Vor dem Hintergrund der relativ hohen Rate fehlender Postwerte erscheint es nicht verwunderlich, dass in der RCT-

Stichprobe die Effektstärken unter Verwendung des LOCF-Datensatzes auf ein niedriges Niveau absanken. Interessanterweise zeigte sich, dass für den neuen Index SOMS-PSDI-2 über alle drei Stichproben hinweg die Varianz der Messwerte zur Prämessung deutlich niedriger ausfiel als zur Postmessung. Dies führte beim SOMS-PSDI-2 dazu, dass bei der Berechnung der Effektstärken unter Verwendung der Prä-Standardabweichung die Effekte deutlich anstiegen. Der SOMS-PSDI-2 wies stichprobenunabhängig im Vergleich zum SOMS-GSI und SOMS-PSDI-1 eine größere Prä-Post-Mittelwertsdifferenz auf, die sich bei den Effektstärken, die mit der gepoolten Standardabweichung berechnet wurden, jedoch nicht in der Effektgröße widerspiegelte. Dies kann auf die – im Vergleich zu den anderen Indizes – relativ hohe Varianz der Postmessung zurückgeführt werden. Eine Erklärung für dieses Phänomen liegt im Wertebereich des Index. Der niedrigste Wert, den ein Patient bzgl. des SOMS-PSDI-2 erreichen kann, beträgt Null. Das ist der Fall, wenn vom Betroffenen jedes Item mit einem Wert < 2 angekreuzt wurde. Da nur Items in die Scoreberechnung integriert werden, die mindestens mit einem Wert 2 bewertet wurden, kann der SOMS-PSDI-2 keine Werte zwischen 0 und 2 erreichen. Der Wertebereich des Index liegt somit entweder bei Null oder zwischen einschließlich 2 und 4. Es kommt somit zu einer „künstlichen“ Erhöhung der Varianz für den SOMS-PSDI-2, wenn dieser Werte im unteren Bereich der Antwortskala annimmt. Somit lässt sich erklären, dass im Vergleich zur Prämessung, bei der eher hohe Skalenwerte zu erwarten sind, bei der Postmessung, bei der die Itemantworten der Patienten eher niedriger ausfallen, die Standardabweichung ansteigt. Die künstliche Erhöhung der Varianz stellt natürlich ein Auswertungs-

problem bzgl. des SOMS-PSDI-2 dar. Beim SOMS-PSDI-1 hingegen ist diese künstliche Varianzerhöhung eher gering, da hier nur der Wertebereich zwischen 0 und 1 ausgespart wird. Dieser Sachverhalt spiegelt sich darin wider, dass sich die Standardabweichungen des SOMS-PSDI-1 stichprobenübergreifend kaum zwischen den beiden Messzeitpunkten voneinander unterscheiden.

Ein vergleichbares Phänomen wie bei der Effektstärkenberechnung zeigte sich auch beim Reliable Change Index. Die Responseraten zum RCI lagen stichprobenübergreifend am höchsten für den SOMS-PSDI-2. Hierbei muss berücksichtigt werden, dass in die Berechnung des RCI die Standardabweichung der Prä- und nicht der Postmessung einfließt.

Obwohl die Befunde, insbesondere die Effektstärken, eher dafür sprechen, dass der neue SOMS-Index zu vergleichbaren Ergebnissen wie die traditionellen SOMS-Indices führen, weisen die Interkorrelationen zu den Indices darauf hin, dass der SOMS-PSDI-1 bzw. -2 und die traditionellen Scores sich inhaltlich anscheinend doch nicht vollständig überschneiden. Die Koeffizienten zu den Interkorrelationen der traditionellen SOMS-Indices liegen deutlich höher als die Zusammenhänge zwischen SOMS-GSI bzw. SOMS-PST-1/-2 mit dem SOMS-PSDI-1/-2.

Zusammenfassend kann aus den vorliegenden Befunden geschlussfolgert werden, dass der SOMS-PSDI eine wichtige neue Auswertungsmethode alternativ zu den bestehenden Indices des SOMS-7T darstellt. Seine Bedeutung liegt jedoch weniger darin, dass er zu anderen bzw. höheren Effekten als die herkömmlichen SOMS-Indices führt, sondern dass er inhaltlich besser zu interpretieren ist. Mit ihm kann konkret die mittlere Beeinträchtigungsintensität, die der Betroffene im Rahmen seiner somatoformen

Symptome erlebt, bestimmt werden. Er überwindet damit das zentrale Probleme der klassischen Auswertungsindizes des SOMS-7T: die Konfundierung zwischen der Anzahl der Beschwerden und deren Beeinträchtigungsgrad. Bezüglich der Frage, welcher Cut-off für die Berechnung des SOMS-PSDI gewählt werden sollte, liefern die vorliegenden Befunde eine klare Aussage. Aufgrund der künstlichen Erhöhung der Varianz im unteren Abschnitt des Wertebereiches sollten für die Indexkalkulation Items mit einem Mindestwert von 1 berücksichtigt werden (d.h. der SOMS-PSDI-1). Für eine umfassende Dateninterpretation sollte auch in zukünftigen Studien darauf geachtet werden, die Ergebnisse zu möglichst allen Indices des SOMS-7T und verschiedenen statistischen Kennwerten zu berichten. Diese Empfehlung basiert auf dem vorliegenden Befund, dass in Abhängigkeit vom verwendeten Index bzw. Kennwert die Ergebnisse deutlich voneinander abweichen können (v.a. beim Reliable Change Index).

Bei der zusammenfassenden Betrachtung der Befunde über alle drei Stichproben fällt auf, dass sich für die Wirksamkeit kognitiver Verhaltenstherapie – basierend auf der Auswertung mit einer Beschwerdeliste – überwiegend nur kleine bis mittlere Effekte ergeben. Dieses Ergebnis zeigt sich relativ homogen über alle Indices des SOMS-7T, einschließlich dem SOMS-PSDI-1 und -2. Dieser Befund lässt sich gut in die Ergebnisse bestehender Überblicksarbeiten einordnen. Zunächst zeigen mehrere Reviews (z.B. Allen, Escobar, Lehrer, Gara & Woolfolk, 2002; Kroenke, 2007; Kroenke & Swindle, 2000) und eine Metaanalyse (Kleinstäuber, Witthöft & Hiller, 2011) zur Wirksamkeit von Psychotherapie bei somatoformen Störungen, dass Beschwerdelisten häufig als primäres Ergebnismaß in Studien eingesetzt wer-

den. Zugleich zeigt sich jedoch auffällig häufig, dass die Effektstärken basierend auf dieser Art von Selbst-rating-Skala im Vergleich zu anderen Outcomes wie z.B. Depressivität oder Angst meist nur klein bis sehr klein ausfallen (z.B. Kleinstäuber, Witthöft & Hiller, 2012; Thomson & Page, 2007; Zijdenbos, de Wit, van der Heijden, Rubin & Quartero, 2009).

Bezüglich des Reliable Change Index zeigt sich, dass die Responderraten im Vergleich zu anderen Störungsbildern eher gering ausfallen. In einer Studie von Gibbons et al. (2010) konnten in einer Stichprobe depressiver Patienten 61% eine klinisch relevante Verbesserung ihrer Symptomatik (basierend auf dem RCI) erreichen. Bei sozialer Phobie wurden RCI-basierte Responderraten von 51-57% (McEvoy, 2007), bei einer Patientenstichprobe mit generalisierter Angststörung 55% (Arntz, 2003) oder bei Patienten mit Schlafstörungen 57% (Currie, Wilson & Curran, 2002) nachgewiesen. Interessanterweise ergeben sich jedoch speziell bei kognitiven Verhaltenstherapien, bei denen körperliche Beschwerden im Vordergrund stehen, ebenfalls eher niedrigere RCI-Responderraten, die vergleichbar zu den Ergebnissen der vorliegenden Studie sind. In einer Untersuchung von Hesser et al. (2012) erreichten Patienten mit chronisch dekompensiertem Tinnitus in der kognitiv-verhaltenstherapeutischen Behandlung nur 29% eine klinisch relevante Symptomverbesserung. In einer Studie von Zaby, Heider und Schröder (2008) kam ebenfalls der SOMS-7T zum Einsatz. Hier erreichten die Teilnehmer einer KVT-Intervention in Abhängigkeit vom verwendeten Index nur zwischen 27,3 und 31,8% eine klinisch bedeutsame Veränderung.

Vor dem Hintergrund dieser Befunde stellt sich die Frage, inwieweit die Messung von Therapiewirksamkeit bei somatoformen Störungen über Be-

schwerdelisten möglicherweise kritisch hinterfragt werden sollte. Dabei sollte vor allem berücksichtigt werden, dass der Behandlungsfokus nicht primär auf der Veränderung der Körperbeschwerden bzw. deren Heilung liegt, sondern auf einem funktionalen Umgang mit den Symptomen (Kleinstäuber, Thomas et al., 2012; Rief & Hiller, 2011). Beschwerdelisten wie der SOMS-7T fragen zwar nach der Beeinträchtigung durch die Körpersymptome, jedoch auf eine sehr globale Art und Weise. Zudem bleibt unklar, ob die Betroffenen die aufgelisteten Beschwerden im Fragebogen wirklich hinsichtlich des Beeinträchtigungsmaßes und nicht eher bezüglich ihrer Intensität einschätzen. Im Vergleich z.B. zu psychosomatischen Beschwerden wie dem chronisch-dekompensierten Tinnitus, bei dem ebenfalls der verbesserte Umgang mit dem Ohrgeräusch und nicht die Heilung im Vordergrund der Therapie steht, wird als primäres Ergebnismaß nicht die subjektive Tinnitusintensität, sondern der Tinnitusdistress bzw. die Beeinträchtigung des alltäglichen Funktionsniveaus als primäres Outcome genutzt (z.B. Hesser, Weise, Westin & Andersson, 2011). Im Rahmen der Items entsprechender Skalen, wie z.B. dem Tinnitus Handicap Inventory (THI; Kleinjung et al., 2007; Newman, Jacobson & Spitzer, 1996), wird nicht nach dem Körpersymptom selbst und dessen Beeinträchtigungsgrad allgemein wie beim SOMS-7T gefragt, sondern nach konkreten Beeinträchtigungen in verschiedenen Lebensbereichen. Beispiele hierfür wären Fragen danach, inwieweit der Betroffene aufgrund des Ohrgeräusches sich nur schwer konzentrieren kann oder inwieweit er durch den Tinnitus im gesellschaftlichen Leben negativ beeinflusst wird.

Die kritische Beleuchtung des Einsatzes von Beschwerdelisten in der Therapieevaluation bei somatoformen Störun-

gen bedeutet jedoch nicht, dass diese im Allgemeinen keine bedeutsamen Messinstrumente darstellen. Die Gewichtung und der Einsatz sollte jedoch stark orientiert an der Fragestellung erfolgen. Während möglicherweise im Rahmen von Therapieevaluationsstudien Beschwerdelisten eher als ein Sekundärmaß genutzt werden könnten, können sie in experimentellen Studien (z.B. zur Untersuchung der subjektiven Symptomwahrnehmung) ein vorrangiges Messinstrument darstellen. In Anlehnung an psychosomatische Symptome, wie z.B. chronischer Tinnitus oder Schmerz, könnten Skalen, die konkrete Beeinträchtigungen durch das Körpersymptom erfragen, möglicherweise eine Alternative für ein Primäroucome in Therapiestudien mit somatoformen Patienten darstellen. Sie ermöglichen ggf. das Therapieergebnis inhaltlich valider und änderungssensitiver zu erfassen. Ein geeignetes Beispiel stellt dafür der Pain Disability Index (Dillmann, Nilges, Saile & Gerbershagen, 1994; Mewes et al., 2009) dar, der konkret erfragt, wie stark Patienten sich durch ihre Körperbeschwerden in verschiedenen Lebensbereichen (z.B. Beruf, soziale Aktivitäten, Sexualeben) beeinträchtigt fühlen.

DANKSAGUNG

Die Autoren danken Dr. Amrei Schindler sowie Dipl.-Psych. Ann-Christin Krämer für ihre Unterstützung bei der Zusammenstellung der Datensätze sowie der Datenauswertung.

LITERATUR

- Allen, L. A., Escobar, J. I., Lehrer, P. M., Gara, M. A. & Woolfolk, R. L. (2002). Psychosocial treatments for multiple unexplained physical symptoms: A review of literature. *Psychosomatic Medicine*, 64, 939-950.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Currie, S. R., Wilson, K. G. & Curran, D. (2002). Clinical significance and predictors of treatment response to cognitive-behavior therapy for insomnia secondary to chronic pain. *Journal of Behavioral Medicine*, 25, 135-153.
- Derogatis, L. R. (1983). *The SCL-90-R manual. Scoring and administration procedures for SCL-90-R*. Baltimore, MA: John Hopkins University School of Medicine.
- Derogatis, L. R. (1992). *The Brief Symptom Inventory (BSI): Administration, scoring, and procedures manual II*. Baltimore, MD: Clinical Psychometric Research.
- Dillmann, U., Nilges, P., Saile, H. & Gerbershagen, H. U. (1994). Behinderungseinschätzung bei chronischen Schmerzpatienten. *Schmerz*, 8, 100-110.
- Franke, G. H. (2000). *Brief Symptom Inventory von L.R. Derogatis (Kurzform der SCL-90-R)*. Deutsche Version. Göttingen: Beltz.
- Gibbons, C. J., Fournier, J. C., Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P. & Beck, A. T. (2010). The clinical effectiveness of cognitive therapy for depression in an outpatient clinic. *Journal of Affective Disorders*, 125, 169-176.
- Hartmann, A. & Herzog, T. (1995). Varianten der Effektstärkenberechnung in Meta-Analysen: Kommt es zu variablen Ergebnissen? *Zeitschrift für Klinische Psychologie*, 24, 337-343.

- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.
- Hesser, H., Gustafsson, T., Lunden, C., Henrikson, O., Fattahi, K., Johnsson, E. et al. (2012). A randomized controlled trial of Internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. *Journal of Consulting and Clinical Psychology, 80*, 649-661.
- Hesser, H., Weise, C., Westin, V. Z. & Andersson, G. (2011). A systematic review and meta-analysis of randomized controlled trials of cognitive-behavioral therapy for tinnitus distress. *Clinical Psychology Review, 31*, 545-553.
- Hiller, W., Bleichhardt, G., Haaf, B., Legenbauer, T., Mauer-Matzen, K. & Rübler, D. (2006). Zertifizierung einer Psychotherapeutischen Hochschulambulanz nach DIN EN ISO 9001. *Zeitschrift für Klinische Psychologie und Psychotherapie, 35*, 225-233.
- Hiller, W., Fichter, M. M. & Rief, W. (2003). A controlled treatment study of somatoform disorders including analysis of healthcare utilization and cost-effectiveness. *Journal of Psychosomatic Research, 54*, 369-380.
- Hiller, W., Rief, W. & Brähler, E. (2006). Somatization in the population: From mild bodily misperceptions to disabling symptoms. *Social Psychiatry and Psychiatric Epidemiology, 41*, 704-712.
- Hiller, W., Zaudig, M. & Mombour, W. (1997). *IDCL – Internationale Diagnosen Checklisten für DSM-IV und ICD-10*. Göttingen: Hogrefe.
- Jacobson, N. S., Follette, W. C. & Revenstorf, D. (1984). Psychotherapy outcome research – methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336-352.
- 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*, 12-19.
- Kazis, L. E., Anderson, J. J. & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical Care, 27*, S178-S189.
- Kleinjung, T., Fischer, B., Langguth, B., Sand, P. G., Hajak, G., Dvorakova, J. et al. (2007). [Validation of the German-Version Tinnitus Handicap Inventory (THI)]. *Psychiatrische Praxis, 34*, 140-142.
- Kleinstäuber, M., Thomas, P., Witthöft, M. & Hiller, W. (2012). *Kognitive Verhaltenstherapie bei medizinisch unerklärten Körperbeschwerden und somatoformen Störungen*. Berlin: Springer.
- Kleinstäuber, M., Witthöft, M. & Hiller, W. (2011). Efficacy of short-term psychotherapy for multiple medically unexplained physical symptoms: A meta-analysis. *Clinical Psychology Review, 31*, 146-160.
- Kleinstäuber, M., Witthöft, M. & Hiller, W. (2012). Cognitive-behavioral and pharmacological interventions for premenstrual syndrome or premenstrual dysphoric disorder: A meta-analysis. *Journal of Clinical Psychology in Medical Settings, 19*, 308-319.
- Kroenke, K. (2007). Efficacy of treatment for somatoform disorders: a review of randomized controlled trials. *Psychosomatic Medicine, 69*, 881-888.
- Kroenke, K., Spitzer, R. L. & Williams, J. B. W. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine, 64*, 258-266.
- Kroenke, K. & Swindle, R. (2000). Cognitive-behavioral therapy for somatization and symptom syndromes: a critical review of controlled clinical trials. *Psychotherapy and Psychosomatics, 69*, 205-215.
- Lambert, M. J. & Ogles, B. M. (2009). Using clinical significance in psychotherapy outcome research: The need for a common procedure and validity data. *Psychotherapy Research, 19*, 493-501.

- McEvoy, P. M. (2007). Effectiveness of cognitive behavioural group therapy for social phobia in a community clinic: A benchmarking study. *Behaviour Research and Therapy*, 45, 3030-3040.
- Mewes, R., Rief, W., Stenzel, N., Glaesmer, H., Martin, A. & Brahler, E. (2009). What is "normal" disability? An investigation of disability in the general population. *Pain*, 142, 36-41.
- Newman, C. W., Jacobson, G. P. & Spitzer, J. B. (1996). Development of the tinnitus handicap inventory. *Archives of Otolaryngology-Head & Neck Surgery*, 122, 143-148.
- Rief, W. & Hiller, W. (2008). *SOMS – Das Screening für Somatoforme Störungen. Manual zum Fragebogen*. (4. vollst. überarb. u. neu norm. Aufl.). Bern, Schweiz: Huber.
- Rief, W. & Hiller, W. (2011). *Somatisierungsstörung (Band 1)*. Göttingen: Hogrefe.
- Robins, L. N. & Regier, D. A. (Hrsg.). (1991). *Psychiatric disorders in America: The Epidemiological Catchment Area Study*. New York: The Free Press.
- Saß, H., Wittchen, H.-U., Zaudig, M. & Houben, I. (Hrsg.). (2003). *Diagnostische und Statistisches Manual Psychischer Störungen. Textrevision. DSM-IV-TR*. Göttingen: Hogrefe.
- Thomson, A. B. & Page, L. A. (2007). Psychotherapies for hypochondriasis. *Cochrane Database of Systematic Reviews*, Issue 4. Art. No.: CD006520. doi: 10.1002/14651858.CD006520.pub2.
- Wittchen, H.-U., Zaudig, M. & Fydrich, T. (1997). *Strukturiertes Klinisches Interview für DSM-IV. Achse I und II*. Göttingen: Hogrefe.
- World Health Organization (1992). *International Statistical Classification of Diseases. 10th Revision (ICD-10)*. Genf, Schweiz: World Health Organization.
- Zijdenbos, I. L., de Wit, N. J., van der Heijden, G. J., Rubin, G. & Quartero, A. O. (2009). Psychological treatments for the management of irritable bowel syndrome. *Cochrane Database of Systematic Reviews*, Issue 1. Art. No.: CD006442. doi: 10.1002/14651858.CD006442.pub2.



DR. MARIA KLEINSTÄUBER
 PHILIPPS-UNIVERSITÄT
 ARBEITSGRUPPE KLINISCHE PSYCHOLOGIE
 UND PSYCHOTHERAPIE
 GUTENBERGSTRASSE 18
 D-35032 MARBURG
 E-MAIL: maria.kleinstaeuber
 @staff.uni-marburg.de