

UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Zentrum für Innere Medizin
Institut und Poliklinik für Psychosomatische Medizin und Psychotherapie
Prof. Dr. med. Bernd Löwe

**Psychometrische Evaluation der deutschen Kurzversion des
Brief Symptom Inventory (BSI-18)**

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Sebastian Hammer
aus Hamburg

Hamburg 2016

**Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 07.07.2016**

**Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.**

Prüfungsausschuss, der Vorsitzende: Prof. Dr. Bernd Löwe

Prüfungsausschuss, zweiter Gutachter: PD Dr. Ingo Schäfer

Inhaltsverzeichnis

1 Arbeitshypothese und Fragestellung	6
2 Einleitung	7
2.1 Depression, Somatisierung und Angst	8
2.1.1 Depressivität, depressives Syndrom und depressive Störungen	8
2.1.2 Somatisierung und somatoforme Störungen	9
2.1.3 Angst, Angstsyndrome und Angststörungen.....	12
2.2 Diagnostische Verfahren für Depression, Somatisierung und Angst	14
2.3 Das BSI-18 und seine Vorgängerinstrumente.....	15
2.3.1 Die Vorgängerverfahren.....	16
2.3.2 Die Entwicklung des BSI-18.....	19
2.3.3 Die psychometrischen Gütekriterien des BSI-18	21
2.3.3.1 Reliabilität	21
2.3.3.2 Faktorielle Struktur.....	23
2.3.3.3 Itemanalyse und Zusammenhänge mit soziodemographischen Merkmalen.....	27
2.3.3.4 Diskriminante und konvergente Validität	28
2.3.4 Cut-Off-Werte, Sensitivität und Spezifität des BSI-18	32
2.3.5 Verbreitung und Anwendung des BSI-18	33
2.4 Schlussfolgerungen für diese Untersuchung und Fragestellungen	35
3 Material und Methoden	37
3.1 Studiendesign und Durchführung.....	37
3.2 Beschreibung der Untersuchungsstichprobe	37
3.3 Instrumente	40
3.3.1 Short-Form 12 Health Survey (SF-12).....	40
3.3.2 Patient Health Questionnaire (PHQ-D).....	41
3.4 Testtheorie und statistische Analysen.....	41
3.4.1 Klassische Testtheorie und Testgütekriterien	41
3.4.2 Objektivität.....	43
3.4.3 Reliabilität	44
3.4.4 Itemanalyse.....	45

3.4.5	Validitätsanalyse	48
3.4.5.1	Kriteriumsbezogene Validität.....	48
3.4.5.2	Sensitivität, Spezifität und Cut-Off-Werte	48
3.4.5.3	Faktorielle Validität.....	49
3.4.5.4	Statistische Analyse.....	50
4	Ergebnisse	52
4.1	Deskriptive Kennwerte und Ergebnisse der Itemanalyse.....	52
4.2	Reliabilitätsanalyse.....	55
4.3	Auswertung der Testergebnisse.....	55
4.3.1	Korrelation der Skalen	56
4.3.2	Geschlechtsunterschiede und Alterseffekte	56
4.3.3	Zusammenhänge mit dem Bildungsstand	58
4.3.4	Vergleich von Patienten mit und ohne Migrationshintergrund.....	58
4.3.5	Zusammenhänge mit Arbeitsunfähigkeit und Berentung.....	59
4.3.6	Zusammenhänge mit Familienstand und Wohnsituation	60
4.3.7	Einfluss von Arztkontakten und Psychotherapieerfahrung.....	61
4.4	Kriteriumsbezogene Validität.....	63
4.4.1	Selbsteinschätzung	63
4.4.2	BSI-18-Skalenwerte in Abhängigkeit der Diagnose nach ICD-10	64
4.4.3	Vergleich zwischen BSI-18, PHQ-D und SF-12.....	67
4.5	Faktorielle Validität.....	68
4.6	Diskriminante Validität, Sensitivität und Spezifität des BSI-18	71
5	Diskussion.....	75
5.1	Inhaltliche Validität.....	75
5.2	Teststatistische Kennwerte	76
5.3	Reliabilität	79
5.4	Kriteriumsbezogene Validität.....	81
5.5	Faktorielle Validität.....	84
5.6	Sensitivität und Spezifität des Fragebogens	85
5.7	Methodenkritik und Ausblicke hinsichtlich möglicher zukünftiger Forschung, Implikationen und Schlussfolgerungen für die Praxis.....	87

6	Zusammenfassung	90
7	Abkürzungsverzeichnis	91
8	Literaturverzeichnis	92
9	Tabellenverzeichnis	104
10	Abbildungsverzeichnis	106
11	Anhang	107
12	Danksagung	110
13	Eidesstattliche Erklärung	110

1 Arbeitshypothese und Fragestellung

Depressionen, Angsterkrankungen, Panikstörungen und somatoforme Störungen sind Erkrankungen mit einer weltweit hohen Prävalenz, die für den Patienten mit einem hohen Leidensdruck verbunden sind. Sie spielen vor allem im psychosomatischen, psychotherapeutischen und psychiatrischen Bereich eine Rolle, sind jedoch auch sehr häufig Komorbiditäten von chronischen somatischen Erkrankungen (Leichsenring 2010). Epidemiologische Untersuchungen zeigen die hohe Verbreitung dieser Erkrankungen in der Allgemeinbevölkerung und in der Primärversorgung (Hanel et al. 2009, Jacobi et al. 2004). Gerade in nicht psychiatrischen/psychosomatischen Fachbereichen werden diese Erkrankungen und ihre Komplikationen von dem behandelnden Team häufig übersehen oder erst in einer akuten Krisensituation erkannt (Weisman und Worden 1976).

Zum Screening eignen sich besonders Verfahren, die sich als ökonomisch und zuverlässig erwiesen haben. Ein anwendungsökonomisches Instrument ist das *Brief Symptom Inventory 18* (BSI-18). Als Selbstbeurteilungsinstrument ist es einfach und wenig zeitintensiv in der Durchführung.

In vielen Studien, besonders im angloamerikanischen Raum, fanden sich solide Hinweise für eine gute Validität und Reliabilität des Instruments. Der Fragebogen ist für den deutschen Sprachraum jedoch nur für bestimmte Teilstichproben evaluiert. Ziel dieser Dissertation ist die psychometrische Evaluation des BSI-18 im deutschen Sprachraum anhand eines Patientenkollektivs des Zentrum für Innere Medizin, Institut und Poliklinik für Psychosomatische Medizin und Psychotherapie der Universität Hamburg.

Teile der Ergebnisse dieser Dissertation wurden im Rahmen einer Ko-Autorenschaft unter dem Titel: „Die Kurzform des Brief Symptom Inventory (BSI-18): Erste Befunde zu den psychometrischen Kennwerten der deutschen Version“ im Fachblatt „Fortschritte der Neurologie – Psychiatrie“, dem wissenschaftlichen Organ des Berufsverbandes Deutscher Nervenärzte publiziert (Spitzer et al. 2011).

Es sei darauf hingewiesen, dass im Folgenden aus Gründen der Vereinfachung und Lesbarkeit zur Bezeichnung von Personen und Untersuchungsgruppen teilweise nur die maskuline Form verwendet wird, grundsätzlich sind jedoch sowohl Frauen als auch Männer gemeint.

2 Einleitung

Obwohl die psychiatrisch-psychosomatische Terminologie wie alle Fach- und Wissenschaftssprachen darauf abzielt, ihre Sachverhalte so präzise und eindeutig wie möglich zu beschreiben und abzubilden, ist ihre Begrifflichkeit oft unscharf und klärungsbedürftig. Dies trifft insbesondere dann zu, wenn die interessierenden Sachverhalte auch als normalpsychologische Phänomene aufzufassen sind und sich im alltäglichen Sprachgebrauch wiederfinden. Diese Überlegungen sind für diese Arbeit insofern relevant, als dass das Brief Symptom Inventory 18 (BSI-18) Merkmale erfasst, die gemeinhin unter den Oberbegriffen *Depression*, *Somatisierung* und *Angst* subsumiert werden. Die genannten Begriffe sind jedoch keineswegs so eindeutig wie sie vordergründig scheinen. Vielmehr lassen sie viele Fragen offen, etwa ob sie ganz breit allgemein menschliche Erlebens- und Verhaltensweisen bezeichnen, ob sie einen pathologischen Zustand oder gar eine psychische Krankheit charakterisieren. Es erscheint daher sinnvoll, in den nachfolgenden Abschnitten zunächst kurz Begriffsklärungen vorzunehmen.

In der psychiatrisch-psychopathologischen Nomenklatura wird – wie in der Medizin im Allgemeinen – zwischen *Symptom*, *Syndrom* und *Krankheit* bzw. *Störung* differenziert. Als *Symptom* versteht man ein objektivierbares oder angegebenes Krankheitszeichen. Ein *Syndrom* ist eine Gruppe von Krankheitszeichen, die für ein bestimmtes Krankheitsbild mit meist uneinheitlicher Ätiopathogenese charakterisch sind. Während eine *Krankheit* eine definierbare Einheit typischer ätiologisch, morphologisch, symptomatisch und nosologisch beschreibbarer Erscheinungen ist (Pschyrembel 2002).

In diesem Sinne wird z.B. ein depressiver Affekt oder eine Schlafstörung als Symptom verstanden. Kommen mehrere typische Symptome zusammen, kann von einem depressiven Syndrom gesprochen werden, wobei dieses im Rahmen unterschiedlicher psychischer Störungen und Erkrankungen (z.B. einer bipolaren Störung oder einer Schizophrenie) sowie als normaler, d.h. nicht krankhafter emotionaler Prozess vorkommen kann.

Die Diagnose einer Depression bzw. einer affektiven Störung schließlich folgt streng definierten Ein- und Ausschlusskriterien, die beobachtet bzw. exploriert werden können. Die modernen Klassifikationssysteme *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) und *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) ermöglichen eine operationalisierte, d.h. exakten Ein- und Aus-

schlusskriterien sowie diagnostischen Entscheidungsregeln und Verknüpfungen folgende, und kriteriumsorientierte Diagnostik. Diagnosen im psychiatrisch-psychotherapeutischen Kontext sind häufig weniger scharf bzw. eindeutig als somatische Diagnosen. So sind psychiatrische Diagnosen häufig prototypischer Natur. Daher kann es bei einzelnen Patienten nötig sein, mehrere Diagnosen zu vergeben, um ihre Problematik hinreichend abzubilden; dies nennt man Komorbiditätsprinzip (Berking und Rief 2012).

An dieser Stelle sei bereits erwähnt, dass mit kurzen Selbstbeurteilungsverfahren wie dem hier untersuchten Brief Symptom Inventory mit 18 Items (BSI-18) die Diagnose z.B. einer Depression formal nicht gestellt werden kann, da nicht alle nötigen Kriterien hinreichend erfasst werden. Vielmehr wird das Vorhandensein von Symptomen erfasst, welche im Rahmen der zu erfassenden psychischen Erkrankungen vorkommen können.

2.1 Depression, Somatisierung und Angst

2.1.1 Depressivität, depressives Syndrom und depressive Störungen

Der eher unscharfe Begriff der Depressivität bezeichnet in der Regel eine Veränderung der Stimmung (Affektivität) in Richtung Traurigkeit und Niedergeschlagenheit, die von leichten Formen bis zu sehr schweren Ausprägungen reichen kann. Ein depressives Syndrom ist hingegen durch das Nebeneinander von mehreren Symptomen, die typisch für eine Depression sind, gekennzeichnet. Neben der gedrückten Stimmung können u.a. folgende Symptome auftreten: Interessensverlust und Freudlosigkeit, Antriebsverminderung, erhöhte Ermüdbarkeit, Konzentrations- und Aufmerksamkeitsstörungen, vermindertes Selbstwertgefühl und Selbstvertrauen, psychomotorische Hemmung oder Agitiertheit, Schuldgefühle, negative und pessimistische Zukunftsperspektiven, Suizidgedanken, Schlafstörungen und verminderter Appetit (Scharfetter 2002). Depressive Syndrome kommen im Rahmen einer depressiven Episode (F32), aber auch bei anderen psychischen Erkrankungen vor, etwa bei Zwangs- und Angststörungen, Suchterkrankungen, Somatisierungs- und Persönlichkeitsstörungen und der Schizophrenie. Häufig begleitet ein depressives Syndrom auch eine somatische Erkrankung.

Die depressiven Störungen als krankhafte Zustände zählen zu den affektiven Störungen (ICD-10: F30-F39). Diese sind vornehmlich durch eine Veränderung der Stimmung definiert, wobei sowohl eine krankhaft gehobene Stimmung (Manie) als auch ein krankhaft gedrückter Affekt (Depressivität) auftreten können (Möller et al. 2009). Der Schweregrad einer depressiven Störung kann stark variieren. Die Grenze zwischen einer „depressiven

Verstimmung“ und einer klinisch bedeutenden depressiven Störung kann nicht immer scharf gezogen werden.

In den aktuellen Klassifikationssystemen wird zwischen einer ersten bzw. einzigen depressiven Episode (dafür sieht die ICD-10 die Diagnose „depressive Episode [F32]“ vor) und wiederholten Episoden unterschieden (wofür die ICD-10 die Diagnose „rezidivierende depressive Störung [F33]“ vorsieht). Davon abgegrenzt werden anhaltende affektive Störungen (F34), bei denen zwischen Zykllothymia (F34.0) und Dysthymia (F34.1) unterschieden wird. Die Zykllothymia ist durch eine anhaltende Stimmungsinstabilität gekennzeichnet, bei der es zu einem Wechsel von hypomanen und depressiven Phasen kommt. Hingegen stellt die Dysthymia eine chronisch depressive Verstimmung dar (Dilling et al. 2011).

Weltweit finden sich deutliche Unterschiede in den Prävalenzraten für affektive Erkrankungen. Es wurden Daten zur Prävalenz der depressiven Erkrankungen im niedrigen einstelligen Prozentbereich (Nigeria) bis zu einem Viertel der Bevölkerung (Frankreich) publiziert (Andrade et al. 2003). Im In- und Ausland findet sich eine Lebenszeitprävalenz für depressive Störungen zwischen 13 und 20%. Die Einjahresprävalenz für eine depressive Episode liegt in Europa bei ca. 4 bis 11% (Alonso et al. 2004, Jacobi et al. 2004).

In der Primärversorgung ergibt sich eine Depressionsrate von 5 bis 10%. Bei den stationär versorgten Patienten liegt die Erkrankungshäufigkeit mit ca. 20% noch höher (Mulrow et al. 1995). Erhöhte Raten finden sich bei Personen weiblichen Geschlechts, im mittleren Lebensalter, bei geringerem Einkommen und bei sozialen Randgruppen. Hohe Punktprävalenzen von 18,7% wurden auch im hohen Lebensalter beschrieben (Meller et al. 1996).

Die höchste Komorbiditätsrate psychischer Erkrankungen besteht zwischen Depressionen und Angstsyndromen (Fawcett und Kravitz 1983). Ein komorbider Substanzmissbrauch ist ebenfalls häufig (Kessler et al. 2005). Depressive Syndrome respektive Störungen begleiten häufig als Komorbidität andere körperliche und psychische Erkrankungen (Paykel et al. 2005).

2.1.2 Somatisierung und somatoforme Störungen

Somatisierung beschreibt üblicherweise „eine Tendenz, körperliche Beschwerden und Symptome, für die keine erklärenden pathologischen Befunde bestehen, zu erleben und zu kommunizieren, sie einer somatischen Krankheit zuzuschreiben und dafür medizinische Hilfe in Anspruch zu nehmen“, wobei ein Zusammenhang mit psychosozialen Stress an-

genommen wird (Lipowski 1988). Darüber hinaus finden sich weitere Bedeutungen (Kirmayer und Robbins 1991): (i) Rein deskriptiv als medizinisch ungeklärte Symptome, (ii) als eine übermäßige Beschäftigung mit den eigenen Körperfunktion und (iii) als eine somatische Manifestation einer bekannten psychischen Störung, insbesondere bei der Depression. Andere Autoren verwenden den Begriff weniger deskriptiv-phänomenologisch als theoretisch-erklärend, etwa als einen Mechanismus, durch den Affekte zu körperlichen Beschwerden führen (Henningsen 1998).

Bei vielen sogenannten funktionellen Störungen wie z.B. Fibromyalgie, multiple Nahrungsmittelunverträglichkeit, Fatiguesyndrome, Colon irritabile, phobischer Schwindel, atypischer Brustschmerz und schwere prämenstruelle Syndrome spielt das Konzept der Somatisierung eine Rolle.

Auch in der Psychiatrie sind somatoforme Aspekte in einer Vielzahl von Diagnosen enthalten. Beispiele hierfür sind z.B.: die somatoforme Störung, die Konversionsstörungen, die Neurasthenie und somatische Symptome bei Patienten mit Angststörungen und Depressionen. Überlappungen sind häufig und die diagnostische Einteilung Gegenstand wissenschaftlicher Diskussionen.

Somatoforme Störungen (ICD-10: F45) imponieren durch die wiederholte Empfindung, Schilderung und Darbietung von körperlichen Symptomen, die nach ausführlicher Diagnostik somatisch nicht ausreichend zu erklären sind. Damit verbunden sind in der Regel viele Arztkontakte und wiederholte Forderungen nach somatischer Diagnostik. Bei dem Vorliegen somatischer Befunde erklären diese nicht hinreichend das Ausmaß der psychischen und körperlichen Symptome. Häufig existiert ein zeitlicher und ursächlicher Bezug zu belastenden Lebensereignissen, Schwierigkeiten oder Konflikten. Es besteht nicht selten eine innere Ablehnung, psychische Umstände als mögliche Ursachen der körperlichen Symptome zu erkennen, auch wenn depressive und ängstliche Symptome eminent erscheinen. Unterschiede im Verständnis der Krankheitsursache führen häufig zu Konflikten zwischen den behandelnden Ärzten und dem Patienten (Dilling et al. 2011).

Nach ICD 10 werden unterschiedliche, konzeptuell jedoch sehr verwandte Diagnosen unter der Ziffer F45 zusammengefasst: Somatisierungsstörung (F45.0), undifferenzierte Somatisierungsstörung (F45.1), hypochondrische Störung (F45.2), somatoforme autonome Funktionsstörung (F45.3) und anhaltende somatoforme Schmerzstörung (F45.4).

Bei etwa jedem fünften Patienten, der wegen eines somatischen Symptoms ambulant vorstellig wird, findet sich keine organische Ursache (Khan et al. 2003). Die Vergleichbar-

keit epidemiologischer Untersuchungen zur Somatisierungsstörung wird durch unterschiedliche Definitionen und diagnostische Einteilungen erschwert. Ungleiche Schwellendefinitionen im Sinne einer unterschiedlichen Anzahl von nötigen Symptomen, um z.B. eine somatoforme Störung zu diagnostizieren, tragen dazu bei. Weiterhin werden berichtete Symptome häufig somatischen Krankheitsbildern zugeordnet und der Anteil der Somatisierung unterschätzt.

Die 4-Wochen-Prävalenz in der allgemeinmedizinischen Praxis ist hoch (Schoepf et al. 2003). Führend dabei sind die somatoformen autonomen Funktionsstörungen (12,1%), gefolgt von undifferenzierten somatoformen Störungen (8,7%), Neurasthenie (8,2%) und schließlich Somatisierungsstörungen (1,8%).

Tabelle 1: Einjahres-Prävalenz funktioneller Beschwerden in der erwachsenen Allgemeinbevölkerung Mannheims (Franz und Schepank 1996)

Symptom	1983-85 in % (A-Studie)	1991-93 in % (B-Studie)
Kopfschmerzen	38,7	31,6
Oberbauchbeschwerden	31,7	27,1
Ermüdung/Erschöpfung	28,8	31,8
Muskel,- Skelettschmerzen	19,7	39,6
Herzschmerzen	17,8	12,9
Unterbauchbeschwerden	16,6	18,2
Appetit-Esstörungen	14,2	21,6
Palpitationen	13,7	14,0
Starkes Schwitzen	11,2	19,1
Schwindel, Ohnmacht	8,5	9,7
Schluckstörungen, Globusgefühl	7,0	6,1
Übelkeit/ Erbrechen	6,7	3,4
Atembeschwerden	4,2	6,6
Alibidinie	5,8	8,1

Mögliche Symptome einer somatoformen Störung fanden sich bei ca. 82% der deutschen Bevölkerung innerhalb einer Woche. Bei einem Viertel der Befragten war gleichzeitig ein schwerwiegender emotionaler Distress vergesellschaftet. Durchschnittlich berichteten die Personen 6,6 Symptome von 53 häufigen Symptomen einer somatoformen Störung (Hiller et al. 2006). In der Mannheimer Kohortenstudie (1983-85 und 1991-93) lagen die Punktprävalenzen für funktionelle Beschwerden bei 28,2% und 21,3%. In knapp der Hälfte der Fälle waren die Kriterien für eine somatoforme Störung erfüllt (11,7% und 8,5%). Auffällig waren polysymptomatische Beschwerdebilder mit durchschnittlicher Nennung von drei bis vier unterschiedlichen Beschwerden (Franz und Schepank 1996). Der prozentuale Anteil der Symptome ist in **Tabelle 1** aufgeführt.

Frauen berichten häufiger über funktionelle Symptome als Männer. Diese geben eher gastrointestinale Beschwerden an, während Frauen häufiger Herzsymptome, Rückenschmerzen sowie körperliche Symptome von Angst- und Depressionserkrankungen erleben (Bräutigam und Christian 1981).

2.1.3 Angst, Angstsyndrome und Angststörungen

Angst ist eine der Basisemotionen des Menschen und beschreibt einen unangenehmen Gefühlszustand der Bedrohung und Gefahr (Möller et al. 2009). Ein Angstsyndrom ist eine übertriebene und situationsinadäquate Angstaussprägung. Die Symptomatik kann auf emotionaler Ebene (Angstgefühl), vegetativer Ebene (z.B. Schwitzen, Zittern, trockener Mund), kognitiver Ebene (z.B. Erwartungsangst) und Verhaltensebene (Flucht und Vermeidung) manifest werden (Aigner et al. 2014). Ein Angstsyndrom kann im Rahmen einer isolierten Angst- oder Panikstörung oder als Aspekt einer psychiatrischen bzw. somatischen Erkrankung auftreten.

Unter den Angst- und Panikstörungen sind psychische Erkrankungen subsumiert, in denen krankhaft gesteigerte Symptome von Angst und Panik im Vordergrund stehen, wobei akute Gefahren oder Bedrohungen fehlen (Möller et al. 2009). Nach ICD-10 werden zwei Gruppen unterschieden, die phobischen Störungen (F40) und die anderen Angststörungen (F41).

Bei den phobischen Störungen (F40) handelt es sich um eine Gruppe von Störungen, bei denen ungefährliche Objekte oder Lebenssituationen ängstliche Gefühle hervorrufen. In der Regel werden die auslösenden Stimuli gemieden oder ängstlich ertragen. Die phobische Angst ist subjektiv und unterscheidet sich in ihrer physiologischen Reaktion nicht von anderen Angstformen und der Reaktion auf echte Gefährdungsmomente. Die Ausprägung variiert zwischen leichter Angst und panischen Reaktionen. Die Angst vor dem Tod und vor Kontrollverlust ist häufig. Die Antizipation eines angstausslösenden Stimulus führt häufig zu einer Erwartungsangst. Ängste, die sich vorwiegend auf den eigenen Körper beziehen, werden nach ICD-10 nicht als phobische Störung eingeordnet, da sich definitionsgemäß bei phobischen Störungen das auslösende Moment außerhalb des eigenen Körpers befindet. Häufig besteht eine Komorbidität mit einer depressiven Erkrankung. Eine Zunahme der depressiven Symptomatik geht nicht selten auch mit einer Zunahme der Ängste einher. Frauen sind, bis auf die sozialen Phobien, häufiger betroffen. Patienten mit einer Phobie erleben nicht zwangsläufig häufiger Ängste als Nichterkrankte, falls es ihnen gelingt den Stimulus zu vermeiden (Dilling et al. 2011).

Von den phobischen Angststörungen werden die anderen Angststörungen (F41) abgegrenzt. Hier stellen Angst oder Panik die zentralen Symptome dar, diese sind weniger abhängig von bestimmten Situationen oder Umgebungen (Dilling et al. 2011). **Tabelle 2** fasst die Symptomatik einiger Angst- und Panikerkrankungen zusammen.

Die körperlichen Symptome von Angst sind vielfältig, exemplarisch sind die häufigsten Symptome einer Panikattacke nach Buller et. al (1988) in **Tabelle 3** aufgeführt. Die Symptomatik einer generalisierten Angststörung ist gekennzeichnet durch pathologische Besorgnis, Probleme zu entspannen, Einschlafstörungen, Hypervigilanz, vegetative Hyperaktivität, häufige schmerzhafte Verspannungen vor allem der axialen Muskulatur, häufige Müdigkeit, innere Ruhelosigkeit und Agitiertheit mit gastrointestinalen Symptomen (Cameron et al. 1986). Im Vergleich zur Panikstörung treten sympathikotone Symptome seltener auf (Noyes et al. 1992).

In der Allgemeinbevölkerung gehören Angststörungen mit einer Lebenszeitprävalenz von 21% zu den häufigsten psychischen Erkrankungen (Jacobi et al. 2004). Sie verteilen sich wie folgt: Panikstörung mit oder ohne Agoraphobie (3,8%), Agoraphobie ohne Panikstörung (3,8%), generalisierte Angststörung (5,1%), soziale Phobie (5,8%), spezifische Phobien (13,2%). Komorbide Abhängigkeitserkrankungen sind häufig (Brawman-Mintzer et al. 1993).

Tabelle 2: Eine Auswahl an phobischen und anderen Angststörungen nach ICD-10

Agoraphobie (F40.0)	Gruppe von Phobien, die z.B. mit Ängsten das Haus zu verlassen, in Menschenmengen oder an öffentlichen Orten zu sein, einhergeht. Eine komorbide Panikstörung kommt häufig vor. Ein Vermeidungsverhalten der phobischen Situation steht oft im Vordergrund. Agoraphobiker erleben häufig nur wenig Angst, da phobische Situationen gemieden werden können.
Soziale Phobien (F40.1)	Im Zentrum steht die Angst vor kritischer Beurteilung durch andere Menschen und eine Vermeidung sozialer Situationen. Häufig besteht ein niedriges Selbstwertgefühl und eine geringe Fähigkeit Kritik zu empfangen.
Spezifische isolierte Phobien (F40.2)	Phobien, die auf definierte Stimuli wie bestimmte Tiere, Zahnarztbesuch oder ähnliches beschränkt sind. Obwohl die auslösende Situation streng begrenzt ist, können Panikzustände hervorgerufen werden.
Panikstörung (F41.0)	Kennzeichnend sind schwere isolierte Angstattacken (Panik), deren Auslöser nicht spezifische Situationen oder besondere Umstände sind und daher für den Betroffenen nicht vorhersehbar sind.
Generalisierte Angststörung (F41.1)	Die Angst ist "frei flottierend", generalisiert und anhaltend und nicht auf bestimmte Situationen beschränkt. Die Symptome sind variabel. Häufige Sorgen betreffen mögliche Unfälle oder Erkrankungen der eigenen Person oder von Angehörigen.

Tabelle 3: Die häufigsten Symptome einer Panikattacke nach Buller et. al. (1998)

Symptom	Prozent
Tachykardie	83,5 %
Hitzewallungen	81,4%
Beklemmungsgefühle	78,4%
Zittern, Beben	78,4%
Benommenheit	75,3%
Schwitzen	72,2%
Schmerzen in der Brust	62,9%
Atemnot	55,7%
Angst zu sterben	51,5%
Angst vor Kontrollverlust	49,5%
Abdominelle Beschwerden	45,4%
Ohnmachtsgefühle	43,3%
Parästhesien	42,3%
Depersonalisation	37,1%

2.2 Diagnostische Verfahren für Depression, Somatisierung und Angst

Der erste Fragebogen zur psychologischen Selbstbeurteilung wurde 1917 von Robert S. Woodworth entwickelt, ihm liegt die Annahme zu Grunde, dass eine Testperson *sich selbst interviewen* kann. Der Ansatz in standardisierter Form Testpersonen Aussagen über ihre psychische Gesundheit oder andere Aspekte machen zu lassen, wurde als *self-report* bekannt und erfolgreich. Die Einordnung der erhobenen Daten erfolgt mathematisch im Vergleich zur Norm und in Kenntnis der Befunde von Erkrankten bzw. anderen Merkmalsträgern. Diese Form der Diagnostik ist ökonomisch und inzwischen der Standard in unterschiedlichen Beratungs-, Betreuungs- und Behandlungssystemen. Die Selbstbeurteilungsverfahren erlauben es dem Patienten, sein Verhalten, Erleben und seine Merkmale auf vorgegeben Skalen anzugeben. Ein Nachteil der Verfahren sind bewusste oder unbewusste Verfälschungstendenzen (z.B. eine Dissimulationstendenz). Das Gros der Instrumente erfasst nur einzelne Erkrankungen bzw. Syndrome. In **Tabelle 4** findet sich eine Auswahl häufig genutzter Selbstbeurteilungsverfahren für Depressivität, Ängstlichkeit und Somatisierung.

Im Gegensatz zu den Selbstbeurteilungsverfahren gibt es Fremdbeurteilungsverfahren. Hier werden die Aussagen und Merkmale des Patienten von einem Untersucher eingestuft und bewertet. Während bei Selbstbeurteilungsverfahren eine verzerrte Selbstwahrnehmung Untersuchungsergebnisse beeinflussen kann, unterliegen Fremdbeurteilungsverfahren einer möglichen Beeinflussung durch den Untersucher.

Tabelle 4: Häufig genutzte Selbstbeurteilungsverfahren für Depressivität, Ängstlichkeit und Somatisierung

Skala	Einsatz	Skalen	Cronbachs α
HADS-D - Hospital Anxiety and Depression Scale (Herrmann-Lingen et al. 1995)	Ausprägung ängstlicher und depressiver Symptomatik während der vergangenen Woche.	zwei Subskalen mit je sieben Items und einem Gesamtwert	$\alpha = 0,80$
ADS - Allgemeine Depressions-skala (Hautzinger und Bailer 2003)	Beeinträchtigung durch depressive Symptome innerhalb der letzten Woche.	Langform mit 20 Items (ADS-L) Kurzform mit 15 Items (ADS-K)	$\alpha = 0,88 - 0,95$.
BAI - Beck Angst-Inventar (Margraf et al. 2007)	Erfassung der Schwere von Angst auf einer vierstufigen Skala in den letzten 7 Tagen.	Summenskala aus 21 Items	$\alpha = 0,86 - 0,92$
BDI – Beck-Depressionsinventar (Hautzinger et al. 2006)	Erfassung der Schwere von depressiven Symptomen auf einer vierstufigen Skala in den letzten 7 Tagen.	Summenskala	$\alpha = 0,90 - 0,93$.
SOMS - Screening für Somatoforme Störungen (Rief et al. 1997)	Screening für Somatoforme Störungen. Verfahren zur Statusmessung (SOMS-2) und Veränderungsmessung (SOMS-7T)	SOMS-2: 1) Index nach DSM-IV, 2) Index nach ICD-10 3) SAD-Index (somatoforme autonome Funktionsstörung) 4) globaler Somatisierungsindex	$\alpha = 0,73 - 0,88$
PHQ-D - Gesundheitsfragebogen für Patienten (Löwe et al. 2002)			Siehe Kapitel 3.3.2
BSI - Brief Symptom Inventory (Franke und Derogatis 2000)			Siehe Kapitel 2.3
SCL-90-R - Symptom Checklist 90-R (Franke und Derogatis 2002)			Siehe Kapitel 2.3
SF-12 / SF-36 - Short Form Gesundheitsfragebogen (Bullinger und Kirchberger 1998)			Siehe Kapitel 3.3.1

2.3 Das BSI-18 und seine Vorgängerinstrumente

Sowohl das BSI-18 als auch seine Vorgänger – die *Symptom Checklist-90* (SCL-90) und das *Brief Symptom Inventory* (BSI) – sind Fragebogenverfahren, mit denen bestimmte Symptome, die für psychische Erkrankungen relevant sind, erfasst werden können. Die erfassten Merkmale können zu Syndromen zusammengefasst werden und erlauben eine Einschätzung des Schweregrades; eine kategoriale Diagnostik, die für den Einzelfall ent-

scheidet, ob eine psychische Erkrankung vorliegt oder nicht, ist hingegen nicht möglich. Damit dienen das BSI-18 und seine Vorläufer lediglich der Orientierung im Sinne eines Screening.

Um die Entstehungsgeschichte des BSI-18 etwas genauer zu verstehen, folgt ein kurzer Überblick über seine Vorgängerinstrumente und deren psychometrische Eigenschaften.

2.3.1 Die Vorgängerverfahren

1977 stellte Derogatis die von ihm entwickelte *Symptom-Checklist-90* (SCL-90) vor. Der Fragebogen erfasst als Selbstbeurteilungsverfahren subjektive Beeinträchtigungen durch körperliche und psychische Symptome in den vergangenen sieben Tagen. Mit seinen 90 Items bildet der Fragebogen neun Syndrome ab: *Depressivität, Somatisierung, Ängstlichkeit, Zwanghaftigkeit, Unsicherheit im Sozialkontakt, Aggressivität, phobische Angst, paranoides Denken* und *Psychotizismus*. Neben diesen neun Skalen ermöglicht das Verfahren die Berechnung von drei globalen Indices. Der *Global Severity Index* (GSI) erfasst die allgemeine psychische Belastung durch Summation aller Itemwerte, der *Positive Symptom Total* (PST) summiert die Anzahl der Symptome, bei denen eine möglich Belastung vorliegt und der *Positive Symptom Distress Index* (PSDI) gibt Auskunft über die Intensität der positiv getesteten Symptome.

Der Fragetypus der einzelnen Items unterscheidet sich nicht und alle Items werden von folgender genereller Einleitung begleitet: „*Wie sehr litten Sie in den letzten 7 Tagen bis heute unter ...*“, woraufhin dann die inhaltliche Spezifizierung erfolgt. Die Items werden auf einer fünfstufigen Likert-Skala von ‚*überhaupt nicht*‘ (entspricht einem Punktwert von 0) bis ‚*sehr stark*‘ (entspricht einem Punktwert von 4) durch Ankreuzen beantwortet. Die deutsche Version wurde von Franke übersetzt, normiert und die Erkenntnisse in einem deutschen Manual veröffentlicht (Franke 1994). Der Fragebogen eignet sich zum Screening und zur Verlaufsdagnostik psychischer Belastungen. Die Durchführung ist sowohl im Gruppen- als auch im Einzelsetting möglich. Das Instrument ist in elektronischer Form und auf Papier erhältlich. Die Bearbeitungszeit beträgt 10 bis 15 Minuten. Mit ausreichenden Sprachkenntnissen eignet sich der Fragebogen ab einem Alter von 14 Jahren zur Diagnostik der o.g. Syndrome.

Die ermittelten Rohwerte werden in T-Werte transformiert, wobei Geschlecht und Bildung berücksichtigt werden. Grundlage bildet eine bevölkerungsrepräsentative Stichprobe

(N = 2141). Laut Manual besteht eine deutliche psychische Belastung bei T-Werten größer/gleich 63 in mindestens zwei der neun Subskalen oder dem GSI.

Die SCL-90-R hat sich in den vergangenen 30 Jahren zu einem Standardinstrument der Forschung und Diagnostik im medizinischen, psychologischen, psychiatrischen und psychosozialen Bereich entwickelt (Franke und Derogatis 2002). Sowohl im deutschen Sprachraum als auch international zählt das Instrument zu den bestetablierten und am meisten eingesetzten Selbstbeurteilungsskalen zur Messung aktueller allgemeiner Psychopathologie. Deutlich wird dies an der großen Anzahl an Publikationen: Alleine für den Begriff „SCL-90-R“ fanden sich bei einer Internetrecherche in der Online-Datenbank pubmed 1230 Einträge, davon 708 in den letzten 10 Jahren (Recherche am 21.12.2014). Die SCL-90-R zeichnet sich durch gute psychometrische Kennwerte aus und wurde in vielen wissenschaftlichen Arbeiten gut rezensiert (z.B. Green et al. 1975, Schmitz et al. 1999).

Um den Test zu ökonomisieren und die Bearbeitungszeit zu reduzieren, erarbeitete Derogatis eine verkürzte Version des Instruments und nannte es *Brief Symptom Inventory* (BSI). Unter Beibehaltung der neun Subskalen wurde die Itemzahl von 90 auf 53 Items reduziert, kurze Zeit später folgte eine Version und ein Manual für den deutschen Sprachraum (Franke und Derogatis 2000). Die Kürze des Instruments erlaubt eine ökonomische Anwendung, da die Bearbeitungszeit der SCL-90-R als zu lang empfunden wurde. Eine detaillierte Auswertung der gewonnenen Daten und Erkenntnisse aus der jahrelangen Anwendung der SCL-90-R zeigten, dass sechs Items pro Skala ausreichten, um die interessierenden Syndrome mit ausreichender Präzision darzustellen. Zu diesem Zweck wurden die Items mit den höchsten Faktorladungen ausgewählt und der Test so um 37 Items reduziert (vgl. **Tabelle 5**).

Tabelle 5: Übersicht über die Skalen und Itemanzahl der SCL-90 und ihrer Nachfolger

Skala	SCL-90-R	BSI	BSI-18
Somatisierung	12	7	6
Zwanghaftigkeit	10	6	-
Unsicherheit im Sozialkontakt	9	4	-
Depressivität	13	6	6
Ängstlichkeit	10	6	6
Aggressivität / Feindseligkeit	6	5	-
Phobische Angst	7	5	-
Paranoides Denken	6	5	-
Psychotizismus	10	5	-
Gesamtanzahl	90	53	18

Die Auswertung erfolgt analog zu der SCL-90. Das BSI wurde in Deutschland anhand einer Stichprobe aus 600 Erwachsenen aus der Bevölkerung und 589 Studenten normiert.

Trotz der ökonomischen Vorteile erreichte das Nachfolgeinstrument nicht ganz die Verbreitung der SCL-90. Diese ist international im Vergleich zu ihren Abkömmlingen in Wissenschaft und Diagnostik am weitesten verbreitet. So zeigte u.a. eine Umfrage von in der Praxis in Deutschland tätigen Psychologinnen und Psychologen einen hohen Stellenwert des Instruments im Alltag der psychologischen Diagnostik (Roth und Herzberg Yorck 2008).

Der primär intendierte Einsatzbereich beider Instrumente ist das Screening und die Abklärung psychischer Beschwerden und Belastungen bei somatisch Erkrankten, insbesondere in Bezug auf maligne Neoplasien. Mittlerweile hat sich der Indikationsbereich sehr erweitert und beide Verfahren sind dann einsetzbar, wenn eine deutliche psychische Belastung der Untersuchten zu erwarten ist. Ein wichtiger Einsatzbereich ist die Evaluation und Qualitätssicherung in der Psychotherapie (Geisheim 2002).

Die interne Konsistenz ist in den meisten Studien für beide Instrumente als gut bis sehr gut beschrieben. In den Normierungsstudien errechneten die Autoren für Cronbachs α Werte für den GSI zwischen 0,92 bzw. 0,95. Die Werte für Cronbachs α der einzelnen Subskalen waren weniger gut als die Werte des *GSI* und fielen sehr unterschiedlich aus (0,42 bis 0,82). Die Retest-Reliabilitäten innerhalb einer Woche lagen zwischen $r_{tt} = 0,90$ (GSI) und $r_{tt} = 0,70$ (Subskalen). Studien zur psychometrischen Evaluation erbrachten unterschiedlich gute Ergebnisse. Die einzelnen Symptomdimensionen zeigen häufig gute konvergente Validitäten mit anderen Instrumenten, z.B. mit dem *Minnesota Multiphasic Personality Inventory* (Dinning und Evans 1977).

Kritisch wird vor allem die Faktorstruktur der Fragebögen betrachtet (Boulet und Boss 1991, Ruipérez et al. 2001) und die teilweise weniger zufriedenstellenden Ergebnisse bezüglich der diskriminanten Validität der neun Skalen. Einige Autoren bezweifeln die Nützlichkeit der Subskalen bezüglich der formalen Diagnostik von psychischen Erkrankungen und sehen den Wert des Instruments in der Messung von allgemeiner psychischer Belastung (Boulet und Boss 1991, Derogatis und Fitzpatrick 2004, Piersma et al. 1994). Die Ergebnisse der einzelnen Subskalen werden vor allem als hilfreicher Ansatzpunkt in der weiteren klinischen Diagnostik betrachtet. Besonders im Kontext der somatischen Versorgung werden die Instrumente vor diesem Hintergrund und mit den bekannten Einschränkungen angewendet (Derogatis und Fitzpatrick 2004).

Im Rahmen der psychometrischen Evaluation und Erforschung der SCL-90-R und des BSI stellte sich heraus, dass die Skalen der Fragebögen für unterschiedliche Patientenkol-

lektive und Erkrankungen unterschiedlich gute Ergebnisse und eine unterschiedliche diagnostische Güte zeigen (Johnson et al. 1996).

2.3.2 Die Entwicklung des BSI-18

Durch die verhältnismäßig hohe Zahl der Items und durch die Vielzahl der erfassten psychopathologischen Syndrome messen die beiden Instrumente mitunter viele Aspekte, die im Alltag der Psychosomatik, Medizinpsychologie, Psychotherapie, Psychiatrie und Rehabilitation eine eher untergeordnete Rolle spielen. Plastische Beispiele hierfür sind die Skalen *paranoides Denken* oder *Psychotizismus*. Weitaus relevanter erscheinen in diesen Bereichen der medizinischen Versorgung die Dimensionen *Somatisierung* (SOMA), *Ängstlichkeit* (ANGS) und *Depressivität* (DEPR). So zeigen jüngere epidemiologische Untersuchungen, dass somatoforme, Angst- und depressive Störungen zu den häufigsten psychischen Erkrankungen in der Allgemeinbevölkerung zählen und häufige Komorbiditäten in der Primärversorgung und anderen organmedizinischen Disziplinen darstellen (Jacobi et al. 2004).

Vor diesem Hintergrund extrahierte Derogatis die drei korrespondierenden Subskalen des BSI und veröffentlichte diese als BSI-18 (vgl. **Tabelle 5**). Die folgenden Ausführungen zum BSI-18 beziehen sich fast ausschließlich, sofern nicht anders gekennzeichnet, auf das zu den Fragebogen gehörende Manual (Derogatis 2000, Derogatis und Fitzpatrick 2004).

Der Fragebogen enthält 18 Items aus dem BSI respektive der SCL-90-R, von denen jeweils 6 auf die Dimensionen SOMA, ANGS und DEPR fallen. Die Skalen und ihre Items werden in **Tabelle 6** vorgestellt.

Die Items wurden aufgrund der Itemanalyse, der Ladungen in der Faktorenanalyse der SCL-90-R sowie aus inhaltlichen Erwägungen, insbesondere in Bezug auf die Prävalenz der Symptome ausgewählt. Alle Items bis auf das Item *Gedanken, sich das Leben zu nehmen* weisen eine hohe Prävalenz in der Bevölkerung auf. Das Item *Hitzewallungen oder Kälteschauer* aus der Skala SOMA wurde aufgrund teststatistischer Erwägungen entfernt. Insgesamt ergibt sich eine im Vergleich zur SCL-90-R deutlich kürzere durchschnittliche Bearbeitungszeit von ca. 4 Minuten. Der Test eignet sich sowohl zur einmaligen Durchführung als auch zur Verlaufs- und Therapiekontrolle über einen längeren Zeitraum mit mehrmaliger Durchführung und kann aufgrund seiner Einfachheit nach kurzer Schulung von der gesamten Breite des medizinischen Personals eingesetzt werden. Das Instrument gilt als sensitiv und wird vor allem als Screeningverfahren verwendet.

Tabelle 6: Verteilung der Items des BSI-18 auf die Subskalen. „Wie sehr litten sie in den letzten 7 Tagen unter...?“

Somatisierung (SOMA)	Depression (DEPR)	Angst (ANGS)
1. Ohnmachts- oder Schwindelgefühlen	2. dem Gefühl, sich für nichts zu interessieren	3. Nervosität oder innerem Zittern
4. Herz- und Brustschmerzen	5. Einsamkeitsgefühlen	6. dem Gefühl, gespannt oder aufgeregt zu sein
7. Übelkeit oder Magenverstimmung	8. Schwermut	9. plötzlichem Erschrecken, ohne Grund
10. Schwierigkeiten beim Atmen	11. dem Gefühl, wertlos zu sein	12. Schreck- oder Panikanfällen
13. Taubheit oder Kribbeln in einzelnen Körperteilen	14. einem Gefühl der Hoffnungslosigkeit angesichts Ihrer Zukunft	15. so starker Ruhelosigkeit, dass Sie nicht mehr still sitzen können
16. Schwächegefühl in einzelnen Körperteilen	17. Gedanken, sich das Leben zu nehmen	18. Furchtsamkeit

Analog zur SCL-90-R und dem BSI kann ein *Global Severity Index* (GSI) als Maß für die psychische Belastung über alle Dimensionen berechnet werden. Die Berechnung der beiden anderen Indices der SCL-90-R (PSDI und PST) ist auf Grund der Kürze des Fragebogens nicht möglich. Die Auswertung des BSI-18 erfolgt analog zur SCL-90-R respektive des BSI durch Summierung der Itemrohwerter. Die Items des BSI-18 werden ebenfalls auf einer fünfstufigen Likert-Skala von ‚*überhaupt nicht*‘ (entspricht einem Punktwert von 0) bis ‚*sehr stark*‘ (entspricht einem Punktwert von 4) kodiert. Eine Umkodierung ist aufgrund von fehlenden invertierten Items nicht nötig. Im Unterschied zum BSI und zur SCL-90 werden keine Mittelwerte gebildet. Die Summation der Itemwerte ergibt die Rohwerte für die drei Skalen bzw. den GSI. Es findet keine Gewichtung der einzelnen Itemwerte statt, so dass ein maximaler Wert von 24 Punkten auf jeder Subskala erzielt werden kann. Die Addition der drei Subskalen ergibt den GSI mit einem maximalen Wert von 72 Punkten. Der Test kann ausgewertet werden, wenn nicht mehr als 2 Items pro Skala fehlen. Zuerst wird im Fall von fehlenden Items der GSI berechnet. Als Wert für die fehlenden Items wird der ganzzahlig gerundete Durchschnitt der Übrigen angenommen.

Die Auswertung der Skalenrohwerter erfolgt mit Hilfe einer Tabelle mit standardisierten T-Werten. Das Manual enthält geschlechtsspezifische T-Werte aus zwei Normstichproben aus den USA; einer Bevölkerungsstichprobe (N = 1134) und einer onkologischen Stichprobe (N = 1543). Die Cut-Off-Werte wurden im Manual analog zu den Vorgängerfragebögen gewählt, da zum Zeitpunkt der Veröffentlichung keine großen unabhängigen Studien eigene Cut-Off-Werte für den Fragebogen definierten. Ein Proband gilt somit als positiv getestet, wenn der GSI oder zwei der drei Dimensionen T-Werte größer/ gleich 63 erreichen

(Derogatis 2000). Eine Auswertung und Beurteilung auf drei Ebenen wird vorgeschlagen, somit sollten der GSI, die Subskalen und die einzelnen Items/ Symptome betrachtet werden.

Der durchführende Diagnostiker sollte auch Eindrücke abseits des Tests zulassen und sich nicht ausschließlich an den erreichten Punktwerten orientieren. Ist eine Versuchsperson über dem Cut-Off-Wert, besteht der hochgradige Verdacht einer psychischen Störung, diese sollte mit weiterführender Diagnostik abgeklärt und näher spezifiziert werden.

Besonders durch die Reduktion auf die o.g. Skalen ist in der Auswertung zu bedenken, dass das Instrument unter Umständen auch Teilaspekte von Erkrankungen misst, die in der theoretischen Skalenstruktur nicht mehr enthalten sind: Beispielfhaft sei hier die Schizophrenie genannt, da sie häufig mit den Symptomen der Angst bzw. Somatisierung und Depression einhergeht. Ebenso sind bei einer Depression unter Umständen erhöhte Werte in den beiden anderen Subskalen zu erwarten, sofern die Erkrankung mit Somatisierung oder einer Angstsymptomatik einhergeht. Mit Einschränkungen bezüglich der kategorialen Diagnostik ist somit zu rechnen.

2.3.3 Die psychometrischen Gütekriterien des BSI-18

Das BSI-18 wurde besonders im anglo-amerikanischen Sprachraum in den letzten Jahren vermehrt eingesetzt. Zahlreiche Studien unterschiedlicher Fachbereiche haben das Instrument genauer untersucht und evaluiert. Eine Medline-Recherche im Dezember 2014 ergab unter dem Suchbegriff „BSI-18“ 91 wissenschaftliche Publikationen. Es folgt ein Überblick über die Ergebnisse aktueller Forschung zum BSI-18, insbesondere im Hinblick auf die psychometrischen Eigenschaften des Fragebogens.

2.3.3.1 Reliabilität

Die Reliabilität des BSI-18 wurde in der überwiegenden Mehrheit der vorliegenden Studien mit Hilfe der internen Konsistenz (Cronbachs α) untersucht. Derogatis (2000) beschreibt in der im Manual publizierten Normstichprobe (N = 1134) interne Konsistenzen zwischen 0,74 und 0,89. Die Subskalen DEPR ($\alpha = 0,84$) und der GSI ($\alpha = 0,89$) waren reliabler als die Skalen SOMA ($\alpha = 0,74$) und ANGS ($\alpha = 0,79$). Zu ähnlichen Ergebnissen im Sinne einer verhältnismäßig geringeren Reliabilität der Skala SOMA sowie deutlich

besseren Ergebnissen des GSI kommt eine Vielzahl von internationalen Studien. Einen Überblick gibt Tabelle 7

Die Test-Retest-Reliabilitäten für die Skalen des Vorgängers BSI über einen Zeitraum von 6 Wochen liegen bei: SOMA $r_{tt} = 0,68$, DEPR $r_{tt} = 0,84$, ANGS $r_{tt} = 0,79$, GSI $r_{tt} = 0,90$. Für den BSI-18 wird von ähnlichen Werten ausgegangen. Diese Annahme konnte Andeu (2008) an 200 Patienten mit psychischen Erkrankungen im Wesentlichen bestätigen. Einzig die Re-Test-Reliabilität der Skala SOMA war deutlich höher ($\alpha = 0,76$). Die interne Konsistenz scheint sich zwischen den Geschlechtern kaum zu unterscheiden (Asner-Self et al. 2006). Die Werte der Männer waren jeweils leichtgradig besser (vgl. Tabelle 8).

Tabelle 7: Interne Konsistenz des BSI-18 in unterschiedlichen Stichproben

Autor(en)	Land	Stichprobe	SOMA α	DEPR α	ANGS α	GSI α
Andreu et al., 2008	E	$N = 200$ Patienten mit psychischen Erkrankungen/ Retest bei $n = 103$ ohne Intervention nach 15 Tagen	0,78 ($r_{tt}=0,76$)	0,88 ($r_{tt}=0,82$)	0,71 ($r_{tt}=0,68$)	0,89 ($r_{tt}=0,76$)
Asner-Self et al., 2006	USA	$N = 100$ Immigranten	0,82	0,81	0,81	0,91
Franke 2011	D	$N = 638$ Psychotherapiepatienten	0,79	0,84	0,84	0,91
Franke et al., 2010	D	$N = 409$ nach Nieren-Tx	0,76	0,87	0,83	0,91
Galdon et al., 2008	E	$N = 175$ Brustkrebspatientinnen	0,62	0,73	0,72	0,82
Meachen et al., 2008	USA	$N = 257$ Patienten nach Schädel-Hirn-Trauma (2 Teilstichproben)	0,61-0,75	0,64-0,84	0,74-0,83	0,84-0,91
Petkus et al., 2009	USA	$N = 142$ mobilitätseingeschränkte Senioren	0,69	0,87	0,81	0,89
Recklitis & Rodriguez, 2007	USA	$N = 221$ Erwachsene mit Krebs in der Kindheit	0,82	0,92	0,87	0,94
Recklitis et al., 2006	USA	$N = 8945$ Erwachsene mit Krebs in der Kindheit	0,75	0,88	0,81	0,90
Rusdorf et al., 2010	D	$N = 217$ Rehabilitationspatienten	0,58	0,81	0,79	0,87
Wang et al., 2010	USA	$N = 710$ Drogenkonsumenten (3 Teilstichproben)	0,83-0,84	0,85-0,87	0,79-0,88	0,91-0,93

Tabelle 8: Geschlechterunterschiede von Cronbachs α nach (Asner-Self et al. 2006).

N = 100 Immigranten (USA)	GSI	SOMA	ANGS	DEPR
männlich	0,937	0,808	0,884	0,832
weiblich	0,884	0,803	0,721	0,799
Total	0,912	0,817	0,810	0,812

2.3.3.2 Faktorielle Struktur

Wie schon bei den Vorgängerversionen wird besondere Kritik an der faktoriellen Validität des BSI-18 geübt. Dabei finden sich vornehmlich drei unterschiedliche Faktorenlösungen. Diskutiert wird neben einem latenten großen Faktor auch eine drei- oder vierfaktorielle Lösung. **Tabelle 10** am Ende des Abschnitts fasst den Stand der wissenschaftlichen Diskussion tabellarisch zusammen. Die exploratorische Faktorenanalyse der Bevölkerungsnormstichprobe (N = 1134) konnte die angenommene dimensionale Struktur nur bedingt wiedergeben (vgl. **Tabelle 9**).

Tabelle 9: Ladungen (> 0,40) der Faktorenanalyse der Bevölkerungsstichprobe des BSI-18 (Derogatis 2000).

Item „Wie sehr litten sie in den vergangenen 7 Tagen unter...“	I	II	III	IV
1. Ohnmachts- oder Schwindelgefühlen	-	0,64	-	-
2. dem Gefühl, sich für nichts zu interessieren	0,76	-	-	-
3. Nervosität oder innerem Zittern	-	-	0,60	0,42
4. Herz- und Brustschmerzen	-	0,73	-	-
5. Einsamkeitsgefühlen	0,71	-	-	-
6. dem Gefühl gespannt oder aufgeregt zu sein	0,41	-	0,67	-
7. Übelkeit oder Magenverstimmung	-	0,67	-	-
8. Schwermut	0,81	-	-	-
9. plötzlichem Erschrecken ohne Grund	-	-	-	0,85
10. Schwierigkeiten beim Atmen	-	0,73	-	-
11. dem Gefühl, wertlos zu sein	0,77	-	-	-
12. Schreck- oder Panikanfällen	0,43	-	-	0,78
13. Taubheit oder Kribbeln in einzelnen Körperteilen	-	0,65	-	-
14. einem Gefühl der Hoffnungslosigkeit angesichts Ihrer Zukunft	0,74	-	-	-
15. so starker Ruhelosigkeit, dass Sie nicht mehr still sitzen können	-	-	0,63	-
16. Schwächegefühl in einzelnen Körperteilen	0,49	0,67	-	-
17. Gedanken, sich das Leben zu nehmen	0,53	-	-	0,43
18. Furchtsamkeit	-	-	0,53	0,80

Vier Faktoren erreichten in der Berechnung das Eigenwertkriterium von „*größer eins*“. Obwohl auf Grund der drei Dimensionen des BSI-18 eine Drei-Faktorenlösung zu erwarten gewesen wäre, widersprechen die Ergebnisse nicht grundsätzlich der angenommenen Faktorenstruktur. Der erste Faktor entsprach der Skala DEPR. Alle sechs Items der Skala luden auf den ersten Faktor, zusätzlich zeigten die Items *Gefühl der Aufregung*, *Nervosität oder inneres Zittern* und *Schwächegefühl* leichte Ladungen auf den „*Depressionsfaktor*“. Diese waren jedoch deutlich geringer ausgeprägt (bis 0,49) als die Ladungen von 5 der 6 Items der Skala DEPR (0,74 bis 0,81). Faktor 2 enthielt hohe Ladungen für alle sechs Items der Skala SOMA und eine leichte Ladung für *Schreck und Panikanfälle*. Vier der sechs Items aus der Skala Ängstlichkeit laden hoch auf den dritten Faktor, der somit als „*Angstfaktor*“ bezeichnet werden könnte. Die Items *Nervosität oder inneres Zittern* und *Furchtsamkeit* der Skala ANGS laden gleichermaßen auf den 3. und den 4. Faktor. Der 4.

Faktor erreichte genau das Eigenwertkriterium und scheint die Panikstörungen abzubilden, da besonders die drei Items, die Symptome von Panik und plötzlichem Erschrecken messen, hohe Ladungen aufwiesen.

Die zweite im Manual publizierte Normstichprobe besteht aus einem onkologischen Patientenkollektiv (N = 1543); hier fanden Zabora et al. (2001), wie Derogatis zuvor, eine Vier-Faktorenlösung. Fünf Items aus der Skala ANGS luden hoch auf den ersten Faktor. Das verbleibende Item der Skala (*Ruhelosigkeit*) lud stattdessen, zusammen mit den meisten Items der Skala DEPR, auf den zweiten Faktor. Das Item *Einsamkeit* lud gleichermaßen auf den 1. und auf den 2. Faktor. Der dritte Faktor repräsentiert alle sechs Items der Skala SOMA. Ein vierter Faktor verpasste das Eigenwertkriterium knapp. Nur das Item *Gedanken sich das Leben zu nehmen* erreichte in diesem Faktor eine signifikante Ladung. Knapp unter dem Cut-Off-Wert ($\geq 0,40$) lagen die Faktorladungen von zwei Items der Skala ANGS. Für die Items *Schreck- oder Panikanfälle* und *Plötzliches Erschrecken ohne Grund* fand sich somit ein verstärkter Zusammenhang mit Suizidalität. Diese Beobachtungen decken sich weitestgehend mit den Erkenntnissen aus der Bevölkerungsstichprobe von Derogatis (2000). Auch hier ergab die Faktorenanalyse einen vierten Faktor, der einen Zusammenhang zwischen der Suizidalität und den genannten Items der Skala ANGS zeigte. Dies ist von besonderem Interesse, da andere Studien zeigen konnten, dass suizidales Grübeln und auch schwere Depression häufig mit Panik einhergehen (Vollrath et al. 1990).

Die Tatsache, dass der erste Faktor der Faktorenanalyse von Zabora und Kollegen (2001) die Items der Angstskala abbildete, während bei Derogatis (2000) der Depressionsfaktor die meiste Varianz aufklärte, scheint durch unterschiedliche Stichproben erklärt. Es ist wahrscheinlich, dass onkologische Patienten ängstlicher sind als die Normalbevölkerung und somit die unterschiedliche Reihenfolge der Faktoren zu erklären ist. Darüber hinaus mag auch der vierte Faktor in der Analyse von Zabora und Mitarbeitern (2001) mit der besonderen Situation onkologischer Patienten mit begleitender Angst und der besonderen Auseinandersetzung mit dem Tod zusammenhängen.

Eine drei-faktorielle Lösung wurde u.a. von Recklitis und Mitarbeitern (2006), Franke und Kollegen (2011) sowie der Arbeitsgruppe um Galdon (2008) präferiert. Recklitis und Kollegen (2006) untersuchten eine onkologische Stichprobe (N = 8945, Erwachsene mit Krebs in der Kindheit). Die Hauptkomponentenanalyse ergab eine Drei-Faktorenlösung in Übereinstimmung mit der intendierten Faktorenstruktur des Instruments. Auch die oben beschriebene vier-faktorielle Lösung wurde durch eine weitere Berechnung mit vier festen

Faktoren bestätigt. Der erste ermittelte Faktor bildete die Skala DEPR ab, der zweite die Skala SOMA. Die Faktoren drei und vier korrelierten mit der Skala ANGS, wobei Faktor drei die Paniksymptome (Items 9, 12, 18) und Faktor vier die Erregungssymptome (Items 3, 6, 15) abbildete. Auch die konfirmatorische Faktorenanalyse mit Strukturgleichungsmodell konnte beide Lösungen bestätigen. Es fanden sich in Bezug auf die Faktorstruktur keine Geschlechtsunterschiede.

Die psychometrischen Merkmale des BSI-18 bei 638 stationären Psychotherapiepatienten in Deutschland wurden von Franke und Mitarbeitern untersucht (2011). Eine exploratorische und eine konfirmatorische Faktorenanalyse konnten die postulierte drei-faktorielle Lösung jeweils bestätigen.

Galdon und Kollegen (2008) präferierten nach einer konfirmatorischen Faktorenanalyse in einer onkologischen Stichprobe (N = 175) das Drei-Faktorenmodell, wobei alle geprüften Modelle den statistischen Kriterien nicht genügten.

Im Gegensatz dazu konnten in den Stichproben aus den Arbeiten von Asner-Self und Kollegen (2006), Franke und Mitarbeitern (2010), der Arbeitsgruppe von Meijer (2011) sowie Rusdorf und Kollegen (2010) die intendierte Faktorenstruktur des BSI-18 nicht wiedergegeben werden. Asner-Self und Mitarbeiter (2006) untersuchten die Faktorstruktur anhand von Immigranten und Flüchtlingen aus Zentralamerika in den Vereinigten Staaten und extrahierten drei Faktoren, wobei der erste Faktor mittlere bis hohe Ladungen für alle Items der Skala DEPR aufwies und fünf der sechs Items der Skala ANGS und zwei Items aus der Skala SOMA enthielt. Der zweite Faktor enthielt vor allem Items der Skala SOMA sowie jeweils ein Item der beiden anderen Dimensionen. Der letzte Faktor zeigte eine ähnlich heterogene Mischung aus Items aller Dimensionen mit einem Schwerpunkt auf Items der Skala SOMA. Gefolgert wurde, dass der BSI-18 einen großen Faktor darstellt, der im weitesten Sinne psychische Belastungen misst.

Die exploratorische Faktorenanalyse von Rusdorf und Kollegen (2010) erfasste ebenfalls drei Faktoren, die nicht den Skalen entsprachen. In der konfirmatorischen Prüfung konnte jedoch die Item-Skalen-Zuordnung empirisch bestätigt werden.

Auch Franke und Mitarbeiter (2010) fanden in einer exploratorische Faktorenanalyse bei nierentransplantierten Patienten drei Faktoren. Alle Items der Skala DEPR sowie ein Item der Skala ANGS luden auf den ersten Faktor, dieser könnte somit als Darstellung der Skala DEPR gesehen werden. Bei den beiden anderen Faktoren war die Zuordnung weniger eindeutig. Vier der 18 Items zeigten keine Ladungen auf die theoretisch korrespondier-

ten Faktoren. Die konfirmatorische Faktorenanalyse ergab bei grenzwertigen Fit-Indices eine schwache Anpassung.

Meijer und Mitarbeiter (2011), die das Instrument mittels der Item-Response-Theorie untersuchten, fanden einen einzelnen latenten Faktor. Es wurden drei Stichproben untersucht: 487 ambulante psychiatrische Patienten, 266 Studenten und 207 Häftlinge. Alle untersuchten Parameter wurden als deutliche Hinweise auf die Unidimensionalität gewertet. Weiterhin fanden die Autoren, dass der GSI vornehmlich von den Items der Skalen ANGS und DEPR abhängig scheint. In diesem Sinne waren neun der zehn Items mit der höchsten diskriminativen Potenz aus den genannten Skalen. In der Studentenchprobe erlaubten die Items der Skala SOMA keinen Rückschluss auf die Ergebnisse des GSI.

Andreu und Kollegen (2008) untersuchten die psychometrischen Merkmale des Fragebogen anhand von 200 spanischen Patienten in psychiatrischer ambulanter Behandlung. Für die postulierten Lösungen mit drei bzw. vier Faktoren konnten keine zufriedenstellenden Ergebnisse gefunden werden. Das Vier-Faktorenmodell erreichte schlussendlich jedoch leicht bessere Testergebnisse in der konfirmatorischen Faktorenanalyse. Die Autoren diskutierten insbesondere einen Bias durch die Stichprobe und favorisierten bei mehrdeutigen statistischen Ergebnissen konträr dazu eine Drei-Faktoren Lösung.

Wang et al. (2013) untersuchten 303 Drogenkonsumenten aus China. Die Autoren dichotomisierten die Items (keine Symptome versus Symptome) und führten mit den dichotomen Items eine konfirmatorische Faktorenanalyse durch. Die drei-faktorielle Lösung konnte so bestätigt werden. Es fanden sich ebenfalls Hinweise für einen zu Grunde liegenden generellen Faktor. Die Analyse ohne Dichotomisierung führte nicht zu einer Bestätigung der postulierten Faktorenstruktur.

Von Torres und Mitarbeitern (2013) wurden kulturelle, sprachliche und geschlechtsspezifische Unterschiede in Bezug auf die Skalenstruktur des BSI-18 in einer Stichprobe aus Personen mexikanischer Herkunft mittels konfirmatorischer Faktorenanalysen untersucht. In der Gesamtstichprobe hatte sowohl das ein- als auch das drei-faktorielle Modell Bestand. Andere Modelle wurden nicht geprüft. Es ergaben sich nur leichte Unterschiede zwischen den getesteten Subgruppen. Es gab kaum Geschlechtsunterschiede. Die meisten Items wurden unabhängig von Geburtsland und Übersetzung interpretiert. Kleine Unterschiede der Faktorladungen ergaben sich zwischen dem spanischen und dem englischen Fragebogen bei den Items *Schreck- oder Panikanfälle*, *Gedanken sich das Leben zu nehmen* und *Furchtsamkeit*.

Tabelle 10: Übersicht über die Studien zur faktoriellen Struktur des BSI-18 in Anlehnung an Franke (2012)

Autor(en)	Land	Stichprobe	Ergebnis
Andreu et al., 2008	E	N = 200 Patienten mit psychischen Erkrankungen/ Retest bei n = 103 ohne Intervention nach 15 Tagen	CFA: 4 Faktoren: SOMA, DEPR, ANGS-generelle Angst (3, 6, 15), ANGS-Panik (9, 12, 18)
Asner-Self et al., 2006	USA	N = 100 Immigranten	EFA: 3 Faktoren, andere Zuordnung der Items: 3,7,8,10,18, 1-Faktor-Modell wird präferiert
Franke et al., 2010	D	N = 409 Patienten nach Nierentransplantation	EFA: DEPR, ANGS, SOMA, andere Zuordnung: 1, 4, 6, 15 CFA: 3 Faktoren + GSI best.
Franke et al., 2011	D	N= 638 Psychotherapiepatienten	EFA: 3 Faktoren, Abbildung der Skalen möglich CFA: 3 Faktoren Abbildung der Skalen möglich
Galdon et al., 2008	E	N = 175 Brustkrebspatientinnen	CFA: 3 Faktoren + GSI best.
Petkus et al., 2009	USA	N = 142 häuslich gebundene ältere Erwachsene	CFA: 3 Faktoren + GSI best.
Prelow, Weaver, Swenson & Bowman, 2005	USA	N = 1115 lateinamerikanische Mütter mit niedrigem Einkommen	EFA und CFA: 1-Faktor-Modell wird präferiert
Recklitis et al., 2006	USA	N = 8945 Erwachsene mit Krebs in der Kindheit	EFA: 3 Faktoren best. CFA: 3 Faktoren best./ auch 4 Faktoren-Modell möglich.
Rusdorf et. al. 2010	D	N = 217 (orthopädische und pulmonologische Reha)	EFA: 3 Faktoren, Abbildung der Skalen nicht möglich CFA: Skalenstruktur best.
Wang et al., 2010	USA	N = 710 Drogenkonsumenten (3 Teilstichproben)	CFA: 3 Faktoren + GSI best.
Wang et al., 2013		N = 303 Drogenkonsumenten	CFA mit dichotomisierten Items, Skalenstruktur bestätigt
Zabora et al., 2001	USA	N = 1643 Krebspatienten	EFA: 4 Faktoren: ANGS, DEPR, SOMA, Suizid/ mit Ausnahme der Items 5, 8, 15, 17,

2.3.3.3 Itemanalyse und Zusammenhänge mit soziodemographischen Merkmalen

Genauere Angaben zur Itemanalyse werden in den meisten Studien nicht erwähnt. Exemplarisch werden einige Ergebnisse dargestellt.

Die Trennschärfe der Items wird mit $r_i = 0,34$ bis $r_i = 0,70$ als mäßig bis hoch beschrieben (Zabora et al. 2001). Die Korrelationen der Items lagen zwischen $r = 0,43$ und $0,86$ zu der dazugehörigen Subskala und $0,42$ und $0,71$ zum GSI, wobei die Items der Skala DEPR generell höhere Korrelationen mit dem GSI aufwiesen (Franke et al. 2010).

Die Skalenwerte des BSI-18 scheinen relativ unabhängig von Alter und Bildungsstand der Testpersonen. In den Normstichproben wird nur das Geschlecht berücksichtigt. Franke und Kollegen (2010) sowie Rusdorf und Mitarbeiter (2010) fanden keine Zusammenhänge zu soziodemographischen Daten wie Alter und Geschlecht sowie klinischen Daten, etwa Krankheitsdauer und Komorbiditäten. Einzig ältere Menschen hatten tendenziell höhere Belastungen in Bezug auf die somatischen Symptome. Zusätzlich schienen die Skalen gegenüber sprachlichen Unterschieden stabil (Asner-Self et al. 2006). Carlson und Kollegen (2004) zeigten anhand von 3095 onkologisch erkrankten Patienten, dass die männlichen Probanden tendenziell erhöhte Testergebnisse in Bezug auf SOMA und die weiblichen auf die DEPR erzielten, während die Ergebnisse in Bezug auf die Skalen ANGS und GSI sich nicht unterschieden. Die Stärke der psychischen Belastung zeigt deutliche Unterschiede in Bezug auf die Art der Krebserkrankung und ist bei Minderheiten und sozialer Benachteiligung in der Gesellschaft tendenziell stärker ausgeprägt. Die US-amerikanischen Normstichproben unterscheiden nur nach Geschlecht und schließen alte Testpersonen (älter 70 Jahre) aus (Derogatis 2000, Zabora et al. 2001).

Eine deutsche Normierungsstichprobe aus der Allgemeinbevölkerung ist noch nicht publiziert (Franke et al. 2012). Die hier von Franke beschriebenen Belastungen bei allen vier Skalen im hohen Alter finden in internationalen Normstichproben keine Erwähnung. Internationale Studien zur altersspezifischen Bestimmung der T-Werte für das BSI-18 gibt es nicht. Für den Bereich der Psycho-Onkologie gibt es für den deutschen Sprachraum noch keine Norm- und Trennwerte.

2.3.3.4 Diskriminante und konvergente Validität

Die veröffentlichten Daten zur konvergenten und diskriminanten Validität zeigten überwiegend gute Ergebnisse. So konnte Derogatis (2000) deutliche Zusammenhänge mit dem besser untersuchten Vorgängerinstrument SCL-90-R nachweisen. Er fand hohe Korrelationen (jeweils $r > 0,90$) zwischen dem GSI und den korrespondierenden Skalen der beiden Schwesterinstrumente. Zabora (2001) beschrieb Korrelationen der beiden Global Severity Indices von 0,84, die Ergebnisse von Recklitis und Rodriguez (2007) zeigten ebenfalls hohe Korrelation der BSI-18 Skalen zu denen der SCL-90-R ($r = 0,89$ bis $0,94$). Dies wurde als Hinweis gewertet, dass der BSI-18 in der Lage ist, zuverlässig die Ergebnisse der Vorgänger zu reproduzieren.

Die Arbeitsgruppe um Prinz (2013) verglich die SCL-90-R mit ihren Kurzversionen BSI, SCL-27, BSI-18, SCL-14 und SCL-K-9 in einem Patientenkollektiv mit affektiven Störungen (N = 2727). Es fanden sich nur geringe Unterschiede hinsichtlich der internen Konsistenz in den Subskalen der SCL-90 und ihrer Abkömmlinge. Erwartungsgemäß fanden sich hohe Korrelationen der korrespondierenden Subskalen. Die divergente Validität der Instrumente untereinander war mäßig.

Tabelle 11: Korrelation zwischen GSI und Subskalen mit vergleichbaren Skalen der SCL-90-R anhand der Bevölkerungsstichprobe (N = 1134)

SCL-90-R – BSI-18	Korrelation
Depression	0,93
Angst	0,96
Somatisierung	0,91
GSI	0,93

Hohe Korrelationen ($r > 0,74$) zwischen den Gesamtscores des BSI-18 und den etablierten Instrumenten *Beck-Depressions-Inventar* (BDI) und *Beck-Angst-Inventar* (BAI), sowie dem *Minnesota Multiphasic Personality Inventory* (MMPI) wurden beschrieben (Andreu et al. 2008). Die BSI-18 Subskala SOMA korrelierte hoch mit dem BAI und seinen beiden Subskalen ($r > 0,77$), wobei die Korrelation mit den physischen Symptomen der Angst höher ausfiel ($r = 0,81$). Gute Korrelationen fanden sich auch mit der „*health concern*“ Skala des MMPI-2 ($r = 0,67$). Die BSI-18 Skala DEPR korrelierte hoch mit dem BDI ($r = 0,83$) und der entsprechenden Subskala des MMPI-2 ($r = 0,78$). Die Skala ANGS, getrennt in Panik- und Angststörungen, zeigte ebenfalls gute konvergente Zusammenhänge mit dem BAI, wobei die panikmessenden Items die höchsten Korrelationen ($r = 0,77$) mit dem BAI Faktor *subjektive Angst und Panik* aufwiesen. Die Korrelationen mit der *Ängstlichkeitsskala* des MMPI-2 fielen geringer aus ($r = 0,66$ [Angst] und $r = 0,47$ [Panik]). Als Hinweis auf die divergente Validität, fanden sich deutlich kleinere oder negative Korrelationen mit den MMPI-Skalen *antisoziales Verhalten* und *bizarres Denken* ($r = 0,35$ bis $-0,25$).

Es fanden sich weiterhin hohe Korrelationen der Skala DEPR ($r = 0,70$) und der Skala ANGS ($r = 0,78$) mit den korrespondierenden Skalen der *Hospital Anxiety and Depression Scale* (HADS) (Rusdorf et al. 2010). Der GSI zeigte enge Zusammenhänge mit der therapeutischen Fremdeinschätzung bezüglich der Schwere von berichteten Symptome (Franke et al. 2010).

Franke und Kollegen (2010) setzten das BSI-18 mit den Fragebögen SF-36 (Bullinger und Kirchberger 1998) und F-Sozu-K-22 (Fydrich et al. 2007) in Beziehung und beschrie-

ben signifikante Korrelation zwischen der *Körperlichen Summenskala* des SF-36 und der Skala SOMA ($r = -0,46$) sowie der *Psychischen Summenskala* und den übrigen drei Indices ($r = -0,55$ DEPR, $r = -0,56$ ANGS, $r = -0,59$ GSI). Die negativen Korrelationen der Indices entsprechen den Erwartungen und bestätigten die theoretisch zu vermutenden konvergen-ten Zusammenhänge der zwei Fragebögen. Der Gesamtwert des F-Sozu-K-22 korrelierte am deutlichsten mit der Skala DEPR ($r = -0,45$) und weniger deutlich mit der Skala ANGS ($r = -0,27$) und der Skala SOMA ($r = -0,16$). Bei den 32% der Versuchspersonen, die durch den BSI-18 als psychisch hoch belastet identifiziert wurden, ergaben sich mit hoher Wahr-scheinlichkeit bei den beiden anderen Instrumenten stimmige Ergebnisse im Sinne einer geringeren körperlichen und psychischen Lebensqualität und einer niedrigeren sozialen Unterstützung.

Hoffman und Mitarbeiter (2004) untersuchten das *Distress Management Screening Measure* (DMSM) und verwendeten zur Validitätsprüfung das BSI sowie das BSI-18. Die Korrelationen des GSI mit dem DMSM lagen bei $r = 0,59$ (BSI) und $r = 0,61$ (BSI-18). Die BSI Subskalen zeigten geringere Korrelationen mit dem DMSM.

Tabelle 12 gibt einen Überblick über Veröffentlichungen, in denen die konvergente und divergente Validität des BSI-18 geprüft wurde.

Tabelle 12: Überblick über konvergente und divergente Validität des BSI-18 in Anlehnung an Franke (2012)

Autor	Land	Stichprobe	Validitätsmaße
Andreu et al., 2008	E	<i>N</i> = 200 Patienten mit psychischen Erkrankungen/ Retest bei <i>n</i> = 103 ohne Intervention nach 15 Tagen	Korrelationen mit BDI, BAI & MMPI/ Änderungssensitivität bestätigt
Asner-Self et al., 2006	USA	<i>N</i> = 100 Immigranten	Geschlechts- und Subpopulationseffekte. Kein Änderung durch Übersetzung in Spanische
Franke 2011	D	<i>N</i> = 638 Psychotherapiepatienten	Niedrigere Belastungen auf allen 4 Skalen durch Suchtpatienten Hohe Korrelationen der Selbstbeurteilung mit dem Beeinträchtigungs-Schwere-Score hohe Korrelationen mit transplantations-assoziiertes Belastung
Franke et al., 2010	D	<i>N</i> = 409 Patienten nach Nierentransplantation	hohe Korrelationen mit transplantations-assoziiertes Belastung
Jacobsen et al., 2005	USA	<i>N</i> = 380 ambulante Krebspatienten	ROC-Analysen/ sehr gute Korrelation zu Distress-Thermometer und HADS
Meachen et al., 2008	USA	<i>N</i> = 257 Patienten nach Schädel-Hirn-Trauma	Prädiktor für funktionalen, psychischen & psychosozialen Status
Petkus et al., 2009	USA	<i>N</i> = 142 häuslich gebundene ältere Erwachsene	höhere Werte der Skalen DEPR und ANGS nach traumatischem Ereignis. Sensitivität und Spezifität nach strukturiertem Interview wurden untersucht. AUC zufriedenstellend. Cut-off für den GSI entsprechend einem T-Wert von 50 vorgeschlagen
Prelow, Weaver, Swenson & Bowman, 2005	USA	<i>N</i> = 1115 lateinamerikanische Mütter mit niedrigem Einkommen	Korrelative Zusammenhänge mit diversen psychosozialen Stressoren
Recklitis & Rodriguez, 2007	USA	<i>N</i> = 221 Erwachsene mit Krebs in der Kindheit	sehr gute Korrelationen mit der SCL-90-R Sensitivität und Spezifität wurden untersucht. Cut-off für den GSI entsprechend einem T-Wert von 50 vorgeschlagen
Rusdorf et. al. 2010	D	<i>N</i> = 217 (orthopädische und pulmonologische Reha)	Hohe Korrelationen mit den korrespondierenden Skalen der HADS
Wetherell et. al. 2010	USA	<i>N</i> = 97, geriatrische Patienten	Im Vergleich mit der HADS geringere diagnostische Güte im Vergleich zu einem strukturierten Interview für generalisierte Angststörungen AUC = .573
Zabora et al., 2001	USA	<i>N</i> = 1643 Krebspatienten	Sensitivität und Spezifität wurden untersucht. Cut-off für den GSI entsprechend einem T-Wert von 57 vorgeschlagen

2.3.4 Cut-Off-Werte, Sensitivität und Spezifität des BSI-18

In Bezug auf den GSI und die neun Subskalen wurde für den BSI und die SCL-90 ein Cut-Off-Wert von 63 (T-Wert) vorgeschlagen und in einigen Studien in Patientenkollektiven bestätigt (z.B. Todd et al. 1997). Diese Regel wurde für den BSI-18 übernommen. Im Manual findet sich keine eigenständige Überprüfung der Cut-Off-Werte für den BSI-18.

Bereits frühere Forschungen zu den Vorgängerinstrumenten führten zu unterschiedlichen Vorschlägen für die Cut-Off-Werte in verschiedenen Populationen (Johnson et al. 1996). Diese Einschränkung ist auch für den BSI-18 zu erwarten. Auf Grund der von Zabora und Kollegen (2001) erhobenen Daten wurden T-Werte geschlechterspezifisch für onkologische Patienten errechnet. Vorgeschlagen wurde die 25. Perzentile des GSI als Cut-Off-Wert (entsprechend eines T-Scores von 57) zu verwenden. Somit ergaben sich eine Sensitivität von 91,2% und eine Spezifität von 92,6%. Einschränkend muss erwähnt werden, dass das inhaltlich verwandte Instrument BSI als vergleichender Standard verwendet wurde und daher statistisch hohe Zusammenhänge zu erwarten sind.

Recklitis & Rodriguez (2007) untersuchten in einer onkologischen Stichprobe die SCL-90 und den BSI-18 mit einem T-Wert von größer 63 als Cut-Off und fanden nur eine geringe Übereinstimmung in Bezug auf die positivgetesteten Personen. Es fanden sich 79 positiv getestete Personen mit Hilfe der SCL-90-R und nur 33 Personen über dem Cut-Off mit dem BSI-18. Die Sensitivität (mit der SCL-90 als „Goldstandard“) lag somit bei nur 41,8% mit einer Spezifität von 100%. Unter Verwendung des alternativen Vorschlags zur Höhe des Cut-Off-Werts (T-Wert > 57) von Zabora und Mitarbeitern (2001) ergab sich eine Sensitivität von 83,6% und eine Spezifität von 97,9%. Auf Grund ihrer Ergebnisse formulierten die Autoren ihrerseits einen alternativen Vorschlag zum Cut-Off (T-Wert \geq 50). Somit erhöhe sich die Sensitivität (97,5%) auf Kosten einer noch akzeptablen Spezifität (85,2%).

Merport und Recklitis (2012) fanden in einer Stichprobe mit 193 Überlebenden einer Krebserkrankung in der Kindheit eine gute diagnostische Übereinstimmung mit dem Vorgängerinstrument als Standard. Eine *Receiver Operating Characteristic Analyse* (ROC) zeigte eine *area under the curve* (AUC) von 0,92. Jedoch ergab sich nach dem im Manual vorgeschlagenen Cut-Off-Wert nur eine unzureichende Sensitivität von 45,2%. Es wurden die oben dargestellten Vorschläge (T-Wert \geq 63, \geq 50 und \geq 57) verglichen. Ein Cut-Off

von ≥ 50 ergab Werte von 87,1% für die Sensitivität und 83,3% für die Spezifität. Dieser Cut-Off-Wert wurde als Alternative für dieses Patientenkollektiv vorgeschlagen.

Petkus und Mitarbeiter (2010) evaluierten das BSI-18 in einer geriatrischen Stichprobe mit 142 Patienten, die ambulante Pflegeleistungen in Anspruch nahmen; im Vergleich mit einem strukturierten Interview ergab sich eine AUC von 0,89 für depressiven Erkrankungen und ein Wert von .80 für Angsterkrankungen. Ein T-Wert von 50 wurde als Cut-Off vorgeschlagen. Auf diesem Niveau ließ sich eine Sensitivität von 88% erreichen. Die vorgeschlagenen Cut-Off-Werte von Derogatis (> 63) führten zu deutlich niedrigeren Werten, so dass von den Angsterkrankten nur 41% erkannt werden konnten.

Thekkumpurath und Kollegen (2009) verglichen das *Distress Thermometer* (DT), das BSI-18 und den *General Health Questionnaire-12* (GHQ-12) in der Palliativmedizin mit den Ergebnissen eines SKID-Interviews mittels ROC-Analysen. Bei 150 Patienten, von denen 34% eine psychiatrische Diagnose erhielten unterschied sich die AUC für die Instrumente DT, BSI-18 und GHQ-12 nicht wesentlich (jeweils 0,73, 0,73 und 0,76). Die Sensitivität und die Spezifität lagen für den BSI-18 bei 78% und 62%. Der Fragebogen war den beiden anderen Instrumenten minimal überlegen oder gleichwertig.

Von Wetherell und Mitarbeiter (2007) wurde die *Hospital Anxiety and Depression Scale* (HADS; Herrmann-Lingen et al., 1995) und das BSI-18 in Bezug auf die zuverlässige Reproduktion von zuvor mit einem strukturierten Interview diagnostizierten Angststörungen bei geriatrischen Patienten verglichen. Ein ROC-Analyse zeigte, dass die HADS in der Lage war, mit ihrer Angstskala die Patienten hinreichend genau zu erkennen. Die AUC betrug 0,80 und es ergab sich für die HADS eine Sensitivität von 97% und eine Spezifität von 67% bei einem Cut-Off von > 7 . Im Unterschied dazu fiel die AUC für die Skala ANGS des BSI-18 deutlich geringer aus (0,57). Das Instrument war somit nicht in der Lage, die betroffenen Personen verlässlich zu identifizieren. Die HADS ist in diesem Zusammenhang als das zuverlässigere Instrument zu betrachten. Betrachtet wurden jedoch nur die Skalen DEPR und ANGS des Instrumentes, der GSI wurde nicht berechnet.

2.3.5 Verbreitung und Anwendung des BSI-18

Zunehmend zu etablieren scheint sich der Fragebogen in der pädiatrischen und adulten Psychoonkologie. Eine Vielzahl unterschiedlicher Erhebungen wurde in den vergangenen Jahren durchgeführt (Biegler et al. 2012, Bober et al. 2013, Bornbaum et al. 2012, Brackett et al. 2012, Buchbinder et al. 2011, Clark et al. 2010, Gianinazzi et al. 2013, Grassi et al.

2013, Hjorleifsdottir und Oskarsson 2010, Kwak et al. 2013, Mertens et al. 2008, Nikolaichuk et al. 2011, Tremolada et al. 2012, Wenninger et al. 2013, Zebrack et al. 2007, Zeltzer et al. 2009).

Weitere Verwendungen fanden sich in der Ernährungs – und Sportmedizin (Imayama et al. 2011), der Suchtforschung (Greenfield et al. 2012), in den Gender-Studies (Mustanski et al. 2010) und zur Untersuchung der psychischen Gesundheit von Müttern mit Frühgeborenen (Voegtline et al. 2010). Mehrere Autoren nutzten das Instrument zur Erfassung psychischer Belastung in geriatrischen Bevölkerungsgruppen (Choi et al. 2013, Gum et al. 2009, Petkus et al. 2009, Petkus et al. 2012) und bei Patienten mit einer posttraumatischen Belastungsstörung (Kimron und Cohen 2012).

Weitere Studien finden sich zu Alkoholmissbrauch und psychischer Gesundheit nach Traumatisierung (Adams et al. 2006), zur Untersuchung von psychogenen Bewegungsstörungen (Anderson et al. 2007, Hopp et al. 2012) und psychischem Distress bei Afro-Amerikanern (Utsey und Hook 2007).

Auch zur Evaluation psychologischer Therapien, z.B. nach dem Erlernen von Coping-Strategien bei Patienten mit einem Schädel-Hirn-Trauma (Backhaus et al. 2010) sowie zur Therapiekontrolle von Biofeedbackverfahren (Alvarez et al. 2013) wurde das Instrument verwandt.

Die zunehmende Bedeutung der Skala äußert sich unter anderem in ihrer Verwendung bei der Evaluation anderer Selbstbeurteilungsinstrumente. Das BSI-18 wurde in unterschiedlichen Studien als Referenz und Maß für die konvergente Validität des *Distress Thermometers* (DT) verwandt (Bevans et al. 2011, Chambers et al. 2014, Cohen et al. 2012, Gessler et al. 2008, Jacobsen et al. 2005, Low et al. 2009, Martinez et al. 2013). Weitere Beispiele sind die *Functional Assessment of Chronic Illness Therapy - Spiritual Well-being Scale* (Canada et al. 2008) und die *Courtauld Emotional Control Scale* (Dura et al. 2010), der *Recent Physical Symptoms Questionnaire* (RPSQ) und der *Comorbid Medical Conditions Questionnaire* (CMCQ) (MacLean et al. 2012).

Auch in der somatischen Medizin liegen mögliche Verwendungsgebiete. Palapattu und Kollegen (2004) konnten zeigen, dass hohe Testwerte auf der Skala SOMA einen signifikanten und unabhängigen Prädiktor der Krankheitsprogression eines Urothelkarzinoms darstellen. Weitere Einsatzgebiete sind die Transplantationsmedizin (Franke et al. 2010, Schneekloth et al. 2012), die Schmerzforschung (Sieberg et al. 2011) und die Rehabilitationsmedizin (Serber et al. 2012).

2.4 Schlussfolgerungen für diese Untersuchung und Fragestellungen

Depressivität, Somatisierung und Angstsymptome gehören zu den häufigsten Merkmalen von Erkrankungen im allgemeinmedizinischen, psychiatrischen und psychotherapeutischen Bereich. Auch in der somatischen Versorgung zeigen epidemiologische Studien die große Relevanz für diese Erkrankungen als Komplikation oder Komorbidität auf. Die Störungen werden nicht häufig genug erkannt und behandelt.

Das BSI-18 kann als kurzes und anwendungsökonomisches Instrument eine sinnvolle Ergänzung zu den bestehenden diagnostischen Methoden sowohl zum Screening als auch zur Verlaufsdiagnostik sein.

Vor dem Hintergrund der zunehmenden Verbreitung in Wissenschaft und medizinischer Versorgung im deutschsprachigen Raum ist eine Untersuchung der psychometrischen Qualität der deutschen Version sinnvoll und nötig. Erfahrungen in der Entwicklung von Selbstbeurteilungsinstrumenten zeigten unterschiedlich gute Ergebnisse in verschiedenen Patientenkollektiven. Von besonderem Interesse für den nationalen Einsatz ist daher eine Untersuchung eines deutschsprachigen Patientenkollektivs aus dem psychosomatisch-psychotherapeutischen Bereich. Eine ausführlichere Validierung erfolgte bisher vor allem für den angloamerikanischen Sprachraum.

Aus diesen Überlegungen ergeben sich mehrere Fragenkomplexe für diese Arbeit:

Wie ist die Reliabilität einzuschätzen?

- Wie hoch ist die interne Konsistenz der drei Subskalen und des GSI?
- Können für den deutschen Sprachraum die in der Tendenz guten Ergebnisse in Bezug auf die Reliabilität der internationalen Studien bestätigt werden?
- Findet sich ebenfalls eine geringere Reliabilität für die Skala SOMA im Vergleich zu den anderen drei Skalen wie in vielen Vorgängeruntersuchungen?
- Wie reliabel ist das BSI-18 im Vergleich mit anderen verbreiteten Selbstbeurteilungsskalen?
- Welche Rückschlüsse lässt die Itemanalyse (Trennschärfe, Schwierigkeit, Homogenität, etc.) auf die Konstruktion des Instrumentes zu?

Welche Aussage lässt sich zur faktoriellen Validität treffen?

- Kann die Skalenstruktur in der exploratorischen Faktorenanalyse wiedergegeben werden?
- Falls nein: Ergeben sich Hinweise auf eine ein- bzw. vierfaktorielle Struktur wie in der Literatur beschrieben?
- Ist eine Auswertung auf syndromaler Ebene sinnvoll oder ist nur der GSI als Maß für die allgemeine psychische Belastung hinreichend valide?

Ist eine kriteriumsbezogene Validität gegeben?

- Finden sich die zu erwartenden Zusammenhänge zu verbreiteten und gut evaluierten Selbstbeurteilungsskalen (SF-12 und PHQ-D)?
- Können zu erwartende Korrelationen mit unterschiedlichen externen Kriterien wie z.B. „*Arbeitsunfähigkeit*“, „*Arztbesuche pro Monat*“ und „*Selbsteinschätzung der eigenen Depressivität*“ bestätigt werden?
- Welche Rolle spielen soziodemographische Merkmale und das Geschlecht?

Welche Güte hat die diskriminante Validität?:

- Erlauben die Ergebnisse der Skalen zuverlässige Rückschlüsse auf die im Erstgespräch gestellte klinische Diagnose und die vom Untersucher eingeschätzte Schwere der Störung?
- Wie stellt sich die Sensitivität und Spezifität des Instrumentes dar?

3 Material und Methoden

3.1 Studiendesign und Durchführung

Die zugrundeliegende Stichprobe für die Evaluationsstudie des BSI-18 hinsichtlich seiner teststatistischen Eigenschaften setzt sich aus 1086 Patienten zusammen, die sich zwischen Januar 2008 und Dezember 2009 in der Ambulanz des Instituts und der Poliklinik für Psychosomatische Medizin und Psychotherapie an den Standorten Universitätsklinikum Hamburg-Eppendorf (UKE) und in der Schön Klinik Eilbek vorstellten. Alle während des Studienzeitraums in der psychosomatischen Ambulanz vorstelligen Patienten erhielten vor ihrem Erstgespräch ein Set von Fragebögen. Enthalten war neben den Selbstbeurteilungsinstrumenten BSI-18, SF-12 und PHQ-D eine Einverständniserklärung und ein Fragebogen zu soziodemographischen Daten mit Angaben zur Person (Alter, Familienstand und Schulbildung, aktuelle berufliche Situation, etc.) und weiteren anamnestischen Fragen (z.B. Arbeitsfehltag, Gewicht, Medikation, Therapieerfahrungen, Therapieerwartungen, Wohlbefinden, Schmerzen, etc).

Die Eingabe der Daten erfolgte anonym durch studentische Hilfskräfte und den Autor in eine *Microsoft Access* Datenbank, in der alle Daten der Fragebögen und die im Erstgespräch gestellten Diagnosen zusammengetragen wurden.

Als Ausschlusskriterien galten: Kognitive Beeinträchtigungen (nach klinischem Urteil), psychotische Erkrankungen oder schwere formale Denkstörungen sowie unzureichende Kenntnisse der deutschen Sprache, die auf ein mangelndes Verständnis der Fragebögen schließen ließen. Darüber hinaus wurden Patientendaten mit grob unvollständig ausgefüllten Fragebögen oder offensichtlich falsch ausgefüllten Fragebögen aus der Datenbank gelöscht oder gar nicht erst aufgenommen.

3.2 Beschreibung der Untersuchungsstichprobe

Die Stichprobe umfasst insgesamt 1086 Patienten im Alter zwischen 17 und 88 Jahren, im Mittel waren die Patienten zum Zeitpunkt der Erhebung 41 Jahre alt (SD = 14,3 Jahre). Zwei Patienten machten keine Angabe zum Alter.

Der überwiegende Teil der Patienten ist weiblich (N = 742; 68,3%). Es nahmen nur knapp halb so viel männliche Patienten an der Studie teil (N = 335; 30,8%). Neun Patienten haben keine Aussage über ihr Geschlecht getroffen. Die deutliche Mehrheit der Patienten hatte die deutsche Staatsbürgerschaft (92%).

Die untersuchten Patienten waren überwiegend ledig (47%). 32 Prozent der Patienten waren verheiratet, zwölf Prozent waren geschieden und vier Prozent lebten getrennt, weitere drei Prozent waren verwitwet.

Die meisten der befragten Personen lebten zum Zeitpunkt der Untersuchung mit einem Partner bzw. einem Partner und Kindern (45%). 32,2 Prozent der Patienten lebte allein, sechs Prozent ohne Partner mit Kindern, neun Prozent lebten mit den Eltern zusammen.

Der überwiegende Teil der Stichprobe benannte den Realschulabschluss (33%) als höchsten Schulabschluss. Danach folgten: Abitur (21%), Haupt-/und Volksschule (23%) und abgeschlossenes (Fach-)Hochschulstudium (17%).

Die deutliche Mehrheit der Patienten arbeitet in einem Angestelltenverhältnis (40%), 15 Prozent der Teilnehmenden waren arbeitslos und 8,3 Prozent befanden sich noch in der Ausbildung. Fünf Prozent gaben an als Arbeiter ihr Geld zu verdienen, ebenso viele gaben jeweils an Hausfrau/ -mann zu sein oder selbstständig tätig zu sein. 31 Prozent der Patienten waren zum Zeitpunkt der Untersuchung nach eigener Aussage arbeitsunfähig. 15 Prozent waren auf Zeit oder dauerhaft berentet bzw. befanden sich in einem laufenden Rentenverfahren.

Ein Großteil der Stichprobe (41%) hat keine Psychotherapieerfahrungen. 42 Prozent der Patienten befanden sich und 16 Prozent sind aktuell in psychotherapeutischer Behandlung. Lediglich 26% der Befragten nahmen keine Medikamente ein bzw. machten keine Angaben bezüglich einer regelmäßigen Einnahme. 29 Prozent der Patienten nahmen zum Zeitpunkt der Befragung Antidepressiva oder Stimmungsaufheller und 16 Prozent regelmäßig Beruhigungsmittel. Schlafanregende Medikation wurde von 10 Prozent der Befragten eingenommen. 17 Prozent gaben an, regelmäßig Schmerzmedikamente zu benötigen. Ein Viertel der Patienten benötigte darüber hinaus andere Medikamente aus somatischen Indikationen (z.B. Antihypertensiva). Lediglich sieben Prozent der Befragten hatten in den letzten zwei Wochen keinen Arzt aufgesucht. Über 40% der Patienten hatten mehr als zwei Kontakte mit ärztlichem Personal. Eine ausführliche Auflistung findet sich in **Tabelle 41** im Anhang.

Die Patienten wurden weiterhin gebeten auf einer Skala von *1 = stark eingeschränkt* bis *10 = nicht eingeschränkt* anzugeben, wie sie ihre Leistungsfähigkeit der letzten 2 Wochen bewerten bzw. wie sehr sie innerhalb dieses Zeitraumes an Depressionen, Ängsten und Schmerzen litten (*1 = gar nicht* bis *10 = sehr stark*). Hierbei zeigte sich eine subjektive Einschränkung der Stichprobe in Bezug auf die Leistungsfähigkeit ($M = 4,14$; $SD = 2,37$).

Ein großer Teil der Stichprobe gab weiterhin an, in den letzten 2 Wochen Schmerzen gehabt zu haben (M = 4,60; SD = 3,22), an Ängsten (M = 5,14; SD = 3,26) und Depressionen (M = 4,97; SD = 3,17) gelitten zu haben. Weiterhin erfasste ein Item den Gesundheitszustand in den letzten 2 Wochen auf einer Skala von 0 = *nicht beeinträchtigt* bis 4 = *schwer beeinträchtigt*. Sowohl der körperliche als auch der seelische Gesundheitszustand wurde im Mittel als mittelmäßig bis schwer beeinträchtigt bezeichnet (körperlicher Gesundheitszustand: M = 2,37; SD = 1,06; seelischer Gesundheitszustand: M = 2,57; SD = 1,03).

Der körperliche und seelische Gesundheitszustand der letzten 14 Tage wurde von den Patienten überwiegend als subjektiv eingeschränkt angegeben. Lediglich fünf Prozent gaben an, körperlich nicht beeinträchtigt zu sein, 45 Prozent fühlen sich leicht bis mittelschwer beeinträchtigt und fast die Hälfte fühlte sich körperlich erheblich bis schwer eingeschränkt. Die Angaben zum seelischen Zustand zeigen ein ähnliches Bild. Hier fühlen sich nur vier Prozent nicht beeinträchtigt. 40 Prozent waren leicht bis mittelschwer eingeschränkt und 56 Prozent fühlen sich erheblich oder schwer beeinträchtigt.

Aus den vorliegenden Arztbriefen der psychosomatischen Ambulanz wurden im Zuge der Datenerhebung die im Erstgespräch gestellten Diagnosen (nach ICD-10) aufgenommen. Die Haupt- und Nebendiagnosen der eingeschlossenen Patienten verteilten sich wie folgt: Depressive Erkrankungen (F31-F34; n = 682), phobische und sonstige Angststörungen (F40-F41; n = 255), dissoziative und somatoforme Störungen (F44-F45; n = 315), psychische und Verhaltensstörungen durch psychotrope Substanzen (F1, n = 76).

Tabelle 13: Ausgewählte Diagnosen – Häufigkeiten in der Stichprobe (N=1086)

Ausgewählte Diagnosen (Mehrfachnennungen möglich)		n	Prozent
dissoziative und somatoforme Störungen (ICD10: F44-F45) n = 315	F44	19	1,7%
	F45	299	27,5%
Angststörungen (ICD-10: F40-F41) n = 241	F40	86	7,9%
	F41	169	15,6%
depressive Erkrankungen (ICD-10: F32 – F34) n = 682	F32	442	40,7%
	F33	198	18,2%
	F34	72	6,6%

Ein Großteil der Patienten hat neben der gestellten Hauptdiagnose weitere psychische oder somatische Diagnosen erhalten. So waren z.B. von 442 Patienten, die an einer depressiven Episode (F32) litten, auch an einer somatoformen Störung (F45) erkrankt, weitere 74 Patienten hatten zusätzlich eine Panikstörung (F41) und 31 eine phobische Störung (F40). 25% aller befragten Personen haben mindestens eine somatische Diagnose und bei 64,5%

wurde mehr als eine Diagnose gestellt. 361 Personen hatten nur eine Diagnose, 445 bekamen zwei und 189 Personen drei Diagnosen. 68 Personen hatten mehr als drei Diagnosen erhalten. Einen Überblick über die Komorbiditäten gibt **Tabelle 42** im Anhang.

3.3 Instrumente

Die beiden zur Berechnung der kriteriumsbezogenen Validität herangezogenen Selbstbeurteilungsinstrumente werden im Folgenden kurz vorgestellt.

3.3.1 Short-Form 12 Health Survey (SF-12)

Der SF-12 ist die zwölf Items umfassende Kurzversion des SF-36 Fragebogens (Bullinger und Kirchberger 1998) zur Erfassung der gesundheitsbezogenen Lebensqualität. Der SF-36 wurde als Folge der *Medical Outcomes Study* (Tarlov et al. 1989) entwickelt. Der Fragebogen ist als generisches Instrument konzipiert, d.h. er soll Vergleiche unterschiedlicher Nationalitäten und Erkrankungen ermöglichen.

Der SF-36 Fragebogen erfasst mit 36 Fragen aus unterschiedlichen Teilbereichen gesundheitsbezogener Lebensqualität und in 8 Dimensionen subjektive Gesundheit und enthält sowohl skalierte als auch dichotome Items.

Der SF-36 enthält acht Skalen: *Körperliche Funktionsfähigkeit, Körperliche Rollenfunktion, Körperliche Schmerzen, Allgemeine Gesundheitswahrnehmung, Vitalität, Soziale Funktionsfähigkeit, Emotionales Rollenverhalten* und *Psychisches Wohlbefinden*, sowie zwei globale Dimensionen zur subjektiven Gesundheit: *die Körperliche Summenskala (KSK)* und *die Psychische Summenskala (PSK)*. Ein einzelnes Item erfasst den Verlauf des Gesundheitszustandes im Vergleich zum Vorjahr.

Es gibt für die Fragebögen jeweils eine „Akutversion“, welche sich auf die Gesundheit der letzten Wochen bezieht, und eine Version mit Bezug auf die vergangenen vier Wochen.

Der Test ist international und in Deutschland weit verbreitet und an großen Patientengruppen und in unterschiedliche Altersstufen geschlechtsspezifisch normiert (Deutsche Normierungsstudie N = 2914). Der Fragebogen wird in der Regel ab einem Alter von 14 Jahren im klinischen Bereich und in der epidemiologischen Forschung eingesetzt. Die deutsche Version wurde bei Studien mit großen Patientenkollektiven angewendet und u.a. in der Bundesgesundheitsurvey eingesetzt (Bellach et al. 2000). Die Reliabilität des In-

struments (Cronbachs α) der Subskalen liegt zwischen $r = 0,57$ und $r = 0,94$, wobei die Mehrheit der Skalen Werte größer $r = 0,70$ aufweist.

Die Anwendung und Analyse der Ergebnisse des SF-36 zeigte, dass der körperliche und psychische Summenfaktor 80-85% der Varianz der acht Subskalen ausmachten, sowohl bei Stichproben aus einem Patientenkollektiv als auch bei gesunden Kontrollgruppen. Speziell die Entwicklung von Summenskalen erlaubte eine Reduktion des Fragebogens auf 12 Items. Durch statistische Regressionsmodelle gelang eine Itemselektion, die gewichtet, eine Reproduktion der psychischen und körperlichen U.S. Normstichprobe ermöglichte.

3.3.2 Patient Health Questionnaire (PHQ-D)

Der *Patient Health Questionnaire (PHQ-D)* – auch *Gesundheitsfragebogen für Patienten* (Löwe et al. 2002) genannt – ist ein Screeninginstrument zur Diagnostik von psychischen Störungen (Gräfe et al. 2004). Der Fragebogen gilt als valide und zeitökonomisch, seine guten diagnostischen Eigenschaften wurden in vielen Studien belegt (Kroenke et al. 2001, Kroenke et al. 2002, Spitzer et al. 1999, Spitzer et al. 2000). Der Fragebogen eignet sich sowohl zur Status- als auch zur Verlaufsdagnostik. Es handelt sich bei dem Instrument um die deutsche Version des *PRIME MD Patient Health Questionnaire (PHQ)* (Spitzer et al. 1999). Ein Modulsystem ermöglicht eine flexible Anpassung an unterschiedliche diagnostische Fragestellungen und Patientengruppen und eignet sich zur Diagnostik der häufigsten psychischen Störungen in der Primärmedizin. In der Komplettversion beinhaltet der Test Fragen zu Störungen des *Essverhaltens, Depression, Angst- und Panikstörungen, somatoformen Störungen und Alkoholmissbrauch*. Darüber hinaus fragt das Instrument nach *psychosozialer Funktionsfähigkeit, kritischen Lebensereignissen*, anderen Stressoren sowie *Menstruation, Schwangerschaft und Geburt*. Dabei ermöglicht der Test neben der kategorialen Diagnostik eine Aussage über den Schweregrad mittels der Skalen: *Depressivität (PHQ-9), somatische Symptome (PHQ-15) und Ängstlichkeit (GAD-7)*.

3.4 Testtheorie und statistische Analysen

3.4.1 Klassische Testtheorie und Testgütekriterien

Ein psychometrischer Test ist ein Verfahren zur Untersuchung empirisch abgrenzbarer Persönlichkeitsmerkmale und zur Quantifizierung der absoluten oder relativen Ausprägung eines Zustands, einer Fähigkeit oder einer Eigenschaft bzw. einer qualitativen Einordnung von Versuchspersonen in Personenklassen. Testverfahren werden in vielen unterschiedli-

chen Wissenschaften (z.B. Medizin, Psychologie, Wirtschafts- und Sozialwissenschaften) in der Forschung und in alltäglichen Anwendungen genutzt .

Zur Entwicklung von Testverfahren werden zwei unterschiedliche Testtheorien verwandt – die klassische und die probabilistische Testtheorie. Beide haben ein theoretisches Fundament und sie genügen klar definierten Gütekriterien (Lienert et al. 1994, Rost 2004).

Für spezifische Fragestellungen und Anforderungen werden laufend neue Verfahren entwickelt oder bewährte modifiziert und überprüft. Ein Großteil der genutzten Verfahren wurde nach der klassischen Testtheorie entwickelt. Aus der Vielzahl vorhandener Instrumente erfolgt die Auswahl und Interpretation eines Tests anhand von Testgütekriterien, die eine Aussage über Genauigkeit und das Vorhersagevermögen eines Verfahrens ermöglichen.

In der Psychologie unterscheidet man unterschiedliche Gruppen von Testverfahren: psychometrische Persönlichkeitstests, Persönlichkeits-Entfaltungs-Verfahren und Leistungstests sowie standardisierte Interviews und standardisierte Beobachtungen.

Die Einordnung der erhobenen Daten erfolgt mathematisch im Vergleich zur Norm und in Kenntnis der Befunde von Erkrankten bzw. anderen Merkmalsträgern. Diese Form der Diagnostik ist ökonomisch und inzwischen der Standard in unterschiedlichen Beratungs-, Betreuungs- und Behandlungssystemen (Bühner 2011).

Nach Möglichkeit sollten psychologische Tests eindimensionale Konstrukte messen, so dass ein bestimmtes Testergebnis möglichst eindeutig auf eine latente Variable bzw. ein Konstrukt zurückzuführen ist, insbesondere vor dem Hintergrund der Verwendung eines psychologischen Tests in der Diagnostik und möglicher spezifischer therapeutischer Konsequenzen.

Zu beachten ist, dass ein psychologischer Test zum einen Fähigkeiten und Eigenschaften eines Probanden misst und zum anderen eine Aussage über momentane Zustände treffen kann. Erstere sind theoretisch über die Zeit weitestgehend konstant, während die Testergebnisse je nach dem momentanen Zustand der Merkmalsausprägung differieren. Diagnosen oder anderweitig intendierte Aussagen können hierbei im Rahmen einer Querschnittuntersuchung (aktuell vorhandener Zustand) oder einer Längsschnittuntersuchung (Veränderungen über die Zeit) getroffen werden.

Die eben genannten Merkmale einer Person werden als *Konstrukte* oder auch als *latente Variablen* bezeichnet und sind in den wenigsten Fällen direkt zu messen, sondern werden aus dem Verhalten der Person bzw. den ermittelten Antworten indirekt abgeleitet. Im Ge-

gensatz dazu werden die einzelnen Items eines Tests (z.B. Fragen) als beobachtbare bzw. manifeste Variablen bezeichnet und sind ein Indikator auf das zugrunde liegende Konstrukt. Es wird angenommen, dass eine unterschiedliche Ausprägung einer latenten Variable zu unterschiedlichen Wertigkeiten der Items führt (Bühner 2011).

Die klassische Testtheorie nimmt an, dass Messfehler nicht miteinander korrelieren und setzt eine Eindimensionalität voraus. Messen mehrere Items dasselbe Konstrukt, d.h. liegen reflektive Indikatoren vor, wird angenommen, dass Korrelationen zwischen den Items auf Variationen der latenten Variable zurückzuführen sind. Bleibt diese konstant sollten die einzelnen Items nicht mehr korrelieren, es bestünden nur noch unkorrelierte Messfehler, sie wären *lokal unabhängig*. Dies ist die formale Voraussetzung für Eindimensionalität. Die Ergebnisse einzelner Personen in einem Test können z.B. auf Grund von systematischen Fehlern (wie Übungseffekten) oder unsystematischen Fehlern (wie Unkonzentriertheit) differieren. Die klassische Testtheorie ist eine reine Messfehlertheorie. Mit ihrer Hilfe lässt sich der Grad der Reliabilität definieren. Einschränkend wird aufgeführt, dass sie nur unsystematische Messfehler berücksichtigt und in der Praxis systematische Messfehler kaum zu vermeiden sind, so dass die Theorie teilweise nicht mehr korrekt anwendbar ist und hieraus abgeleitete Reliabilitätsmaße ihre Wertigkeit verlieren (Bühner 2011, Petermann et al. 2006). Alternativ können Testverfahren nach der probabilistischen Testtheorie entwickelt werden, dies ist in der Vergangenheit in Deutschland z.B. bei Lernstandserhebungen (PISA) erfolgt.

Zur Beurteilung der diagnostischen Güte eines Tests wurden Haupt- und Nebengütekriterien formuliert. Die Hauptgütekriterien sind Objektivität, Reliabilität und Validität sowie Skalierbarkeit. Diese lassen sich wiederum in unterschiedliche Untergruppen differenzieren und sollen in dieser Arbeit untersucht werden (Bühner 2011).

3.4.2 Objektivität

Die Objektivität ist ein Maß für die Unabhängigkeit der Ergebnisse eines Testes von dem jeweilig durchführenden Untersucher. Interpretation, Durchführung und Auswertung sollten standardisiert erfolgen. In diesem Sinne unterscheidet man:

- Durchführungsobjektivität
- Auswertungsobjektivität
- Interpretationsobjektivität

Die Handbücher der Fragbögen enthalten in der Regel klare Anweisungen zur korrekten Durchführung der Testungen, so dass Durchführung, Wertung und Interpretation insbesondere bei Selbstbeurteilungsinstrumenten streng standardisiert erfolgen können.

3.4.3 Reliabilität

Die Reliabilität beschreibt die Präzision, mit der ein Test ein bestimmtes Merkmal misst, egal ob intendiert oder nicht.

Mathematisch ist die Reliabilität definiert als der Anteil der Varianz der wahren Werte an der Varianz der beobachteten Werte, bzw. äquivalent zur quadrierten Korrelation zwischen den wahren und beobachteten Werten. Die Varianz der wahren Werte ist jedoch bei einer einmaligen Testung nicht zu ermitteln und lässt sich z.B. mit Hilfe von Korrelationen von parallelen Testungen schätzen (Bühner 2011).

Somit bedeutet eine geschätzte Reliabilität von $p_{tt} = 0,50$, dass eine Beeinflussung des Testwertes durch die Varianz der Messfehler und die systematische Varianz das ermittelte Testergebnis zu gleichen Teilen bestimmt. Die Reliabilitätsschätzungen nehmen ein gleichmäßiges Auftreten der Messfehler in allen Bereichen der Eigenschaftsausprägung an. Es werden drei Methoden zur Schätzung der Reliabilität unterschieden:

- Interne Konsistenz/ Halbierungsvalidität
- Retest-Reliabilität
- Paralleltest-Reliabilität

In der hier vorliegenden Studie wurde die Reliabilität des BSI-18 aufgrund des Studiendesigns und der besseren Vergleichbarkeit zu vorherigen Studien im Hinblick auf die interne Konsistenz untersucht. Der Cronbach- α -Test (Cronbach 1951) ist heutzutage die Standardmethode zur Berechnung der inneren Konsistenz, da er bei im wesentlichen oder essenziell tau-äquivalenten Messungen genaue Schätzungen der Reliabilität liefert. Essenziell tau-äquivalent bedeutet, dass die zu Grunde liegende Eigenschaft mit derselben Skala gemessen wird. Die ermittelten Werte müssen nicht gleich genau sein und können unterschiedliche Mittelwerte annehmen. Die Messfehler der Messungen sollten unkorreliert sein. Die wahren Werte der Messungen unterscheiden sich möglicherweise um eine additive Komponente (Bühner 2011).

Ist die Voraussetzung der Tau-Äquivalenz nicht erfüllt, gibt Cronbachs α die untere Grenze der Reliabilität an, sie wird demnach unterschätzt. Sind die Messfehler nicht wie angenommen unkorreliert, kann der Korrelationskoeffizient nicht genau geschätzt werden.

Korrelierte Messfehler können mit einer konfirmatorischen Faktorenanalyse identifiziert werden.

Zur Berechnung von Cronbachs α muss der Proband den Test nur einmalig durchführen, es erfolgt eine Aufteilung des Testes in viele *Untertests* entsprechend der Anzahl der Items oder in äquivalente Gruppen inhaltlich gleicher Items.

Berechnet wird die interne Konsistenz auf Basis der Korrelationen, Kovarianzen und Itemstreuungen. In dieser Arbeit wird die interne Konsistenz auf diese Weise für die Gesamtskala (GSI) und die drei Subskalen betrachtet. Zusätzlich wurde die Reliabilität der Skalen getrennt nach Schulbildung und Geschlecht bestimmt. Ein Wert größer $\alpha = 0,90$ wird in dem Zusammenhang als hoch betrachtet. Liegt der Koeffizient zwischen $\alpha = 0,80$ und $\alpha = 0,90$ wird von einer mittleren Messgenauigkeit gesprochen, Werte unter $\alpha = 0,80$ gelten als ungenau (Fisseni 2004). Auch wenn die interne Konsistenz einer Skala sehr hoch ist, verbirgt sich hinter dieser nicht zwangsläufig ein einzelner Faktor. Auch sehr hohe Werte von Cronbachs α können mit mehreren zugrunde liegenden Faktoren zustande kommen (Schmitt 1996).

Bei der Schätzung der Reliabilität mit Testhalbierungsmethoden können je nach Trennungsregel unterschiedliche Reliabilitäten geschätzt werden, dieser Fehler kann durch die Bestimmung der internen Konsistenz großteils vermieden werden, da der Korrelationskoeffizient als durchschnittliche Korrelation aller möglichen Halbierungen interpretiert werden kann. Cronbachs α nimmt mit steigender Itemanzahl und hoher Korrelation zwischen den Items zu, eine Zunahme ist jedoch auch bei zusätzlichen Items mit geringer Korrelation zur Skala, also bei einer geringen Trennschärfe, zu beobachten (Bühner 2011).

Es gibt unterschiedliche Einflussgrößen auf die Reliabilität, hier sind vornehmlich die Homogenität oder Heterogenität der Testitems, ihre Trennschärfe, die Varianz der Testwerte und Messfehler zu nennen (Bühner 2011). Die Beschreibung dieser Testkennwerte wird in der Itemanalyse durchgeführt.

3.4.4 Itemanalyse

Die Itemanalyse dient der Bestimmung der teststatistischen Güte der einzelnen Items. Die Analyse wird mit der Absicht durchgeführt, suboptimale Items und Itemeigenschaften zu identifizieren, insbesondere hinsichtlich der Güte der Items, die intendierte Testabsicht abzubilden. Sie beschäftigt sich somit mit der Zusammenstellung von einzelnen Items zu

einem Test und prüft nach verschiedenen Kriterien, welche Items im Sinne des gesamten Tests brauchbar oder weniger brauchbar sind.

Der BSI-18 Fragebogen enthält ausschließlich Likert-Skalen-Items. Zur Analyse wurden neben der deskriptiven Statistik *Mittelwert* (M), *Standardabweichung* (SD) und den Verteilungseigenschaften wie *Schiefte* und *Exzess*, zwei weitere Werte herangezogen: Die *Schwierigkeit* und die *Trennschärfe*.

Der Schwierigkeitsindex, auch Index der „kategorialen Häufigkeiten“, gibt den Anteil der Probanden wieder, die ein Item als „richtig“ beantwortet haben. Idealerweise trennt ein Test Probanden mit hoher von solchen mit niedriger Merkmalsausprägung. Zu dieser Trennung tragen nur Items bei, die nicht von allen Probanden auf die gleiche Art gelöst werden (Michel und Conrad 1982).

Die psychometrische Itemschwierigkeit ist als Zustimmung zu einem Item in Schlüsselrichtung der Skala definiert (Bühner 2011). Der Schwierigkeitsindex ist bei zweistufigen Antworten definiert als die Zahl der richtigen Lösungen durch die Zahl der Probanden und gibt somit die Wahrscheinlichkeit p wieder, dass eine Frage „richtig“ beantwortet wird. Für mehrstufige Antworten ist p nicht definiert. Alternativ lässt sich je nach Item der Mittelwert und die Streuung als Äquivalent zu p verwenden (Fisseni 2004).

Theoretisch sind Werte um $p = 0,5$ (entsprechend MW = 2 auf der Likertskala des BSI-18) als ideal anzusehen, in der Praxis jedoch sind Items unterschiedlicher Schwierigkeit von Vorteil, da sie eine bessere Abstufung der Testergebnisse erlauben. Auf diese Weise wird die Homogenität zu Gunsten der Differenzierungsvielfalt gelockert (Fisseni 2004).

Analog gelten Items mit einem Mittelwert deutlich größer 2 als *schwer*, da im Mittel die Testpersonen dieses Item in Richtung einer stärkeren Merkmalsausprägung beantwortet haben; somit ist eine rechtssteile Verteilung der Antworthäufigkeiten zu erwarten. In diesem Sinne wird ein Item als *leicht* bezeichnet, wenn weniger Personen bei dieser Frage hohe Werte auf der Likert-Skala angegeben haben. In diesem Fall liegt in der Regel eine linkssteile Verteilung vor. Ist ein Item sehr *schwer* bzw. *leicht* kommt es zu einer extrem schiefen Verteilung, es treten Decken- und Bodeneffekte auf, die Standardabweichung ist tendenziell erniedrigt, ebenso nimmt die Trennschärfe tendenziell niedrigere Werte an.

Die Trennschärfe (r_{it}) stellt den wichtigsten Itemkennwert dar und bezeichnet eine Korrelation zwischen einem einzelnen Item und einer zu messenden Eigenschaft, respektive der korrespondierenden Skala, d.h. die Trennschärfe für innere Kriterien (Eigentrennschärfe) ist definiert als die Korrelation zwischen Item- und Testscore mit dem Ziel, Items zu

finden, die hoch mit dem zugrunde liegenden Kriterium korrelieren. Hier besteht ein Zusammenhang mit der Reliabilität, da hohe Eigentrennschärfen tendenziell mit einer hohen Reliabilität einhergehen.

Unterschieden wird die Fremdtrennschärfe, hierbei wird die Korrelation mit einem externen Kriterium beschrieben, z.B. einem externen Testscore oder einer messbaren Variable (z.B. Fehltag). In diesem Fall besteht eine Ähnlichkeit zur kriteriumsbezogenen Validität - auch Itemvalidität. In der vorliegenden Untersuchung wird die Eigentrennschärfe beurteilt. Berechnet wurde die Trennschärfe nach der part-whole-Methode, d.h. das einzelne Item wurde aus der korrelierten Gesamtskala herausgerechnet, um fälschlich hohe Trennschärfen zu verhindern.

Trennschärfekoeffizienten größer $r_{it} = 0,50$ sind als hoch, zwischen 0,50 und 0,30 als mäßig und unter 0,30 als niedrig zu bewerten (Fisseni 2004). Die Höhe der Trennschärfe ist dabei abhängig von dem Schwierigkeitsindex und der Interkorrelation der Items untereinander. Mit einer Schwierigkeit von $p = 0,50$ und einer Trennschärfe von $r_{it} = 1$ ergeben sich die meisten Möglichkeiten, Probanden zu unterscheiden (Lienert et al. 1994). Auch empirisch zeigt sich ein Zusammenhang zwischen p und r_{it} , wobei niedrige und hohe Schwierigkeiten mit einer geringen Trennschärfe einhergehen und mittlere Schwierigkeiten eine höhere Trennschärfe aufweisen. Bezüglich der Interkorrelation von Items ist eine hohe Korrelation mit einer hohen Trennschärfe vergesellschaftet. Eine eingeschränkte Varianz ergibt sich aus der Untersuchung homogener Stichproben, in diesem Fall entstehen tendenziell geringere Trennschärfen

Die *Homogenität* r_{ii} beschreibt, wie hoch die einzelnen Items einer Skala im Durchschnitt miteinander korrelieren. Hohe Werte für die Homogenität bedeuten, dass die Items ähnliche Informationen erfassen (Bortz et al. 2006). Die Schwierigkeit der einzelnen Items bestimmt das Ausmaß der Homogenität, größere Schwierigkeitsunterschiede ergeben eine geringere Interkorrelation und eine geringere Reliabilität. Prinzipiell sind unkorrelierte (d.h. heterogene) Items gleicher Schwierigkeit oder positiv korrelierte (d.h. homogene) Items unterschiedlicher Schwierigkeit gewünscht. Die Homogenität gilt bei Interitemkorrelationen von $> 0,4$ als hoch (Lienert et al. 1994). Eine hohe Homogenität lässt redundante Items bzw. eine große Ähnlichkeit der Items vermuten.

3.4.5 Validitätsanalyse

Die Validität beschreibt das Maß, in dem ein Test misst, was er zu messen vorgibt. Man unterscheidet grundsätzlich drei Validitätsarten:

- Inhaltsvalidität,
- Kriteriumsvalidität
- Konstruktvalidität

In dieser Studie werden vornehmlich zwei Aspekte der Validität näher untersucht – die kriteriumsbezogene und die Konstruktvalidität im Sinne einer faktoriellen Validitätsanalyse.

3.4.5.1 Kriteriumsbezogene Validität

Zur Berechnung der kriteriumsbezogenen Validität werden die Ergebnisse des zu untersuchenden Instruments mit anderen Instrumenten und diagnostischen Methoden, die gleiche Merkmale erfassen sollen, verglichen und Korrelationen mit den Ergebnissen der anderen Verfahren hergestellt. Voraussetzung sind hierbei möglichst hohe Validitätskoeffizienten der zum Vergleich hinzugezogenen Instrumente gegenüber dem zu messenden Merkmal, d.h. die Vergleichsinstrumente sollten ihrerseits möglichst valide sein.

Handelt es sich bei den zu vergleichenden Variablen um quantitative Größen (z.B. Skalensummen Werte der Fragebögen BSI-18 und PHQ-D), wurde der Zusammenhang der beiden Größen mittels des Korrelationskoeffizienten ρ nach Spearman berechnet. In diesem Zusammenhang gelten Validitätskoeffizienten $\rho \geq 0,60$ als hoch, 0,40 bis 0,60 als mittelmäßig und unter 0,40 als niedrig (Fisseni 2004).

Für kategoriale Variablen der zum Vergleich hinzugezogenen Fragebögen und die im Erstgespräch ermittelten Diagnosen nach ICD-10, wurden nonparametrische Verfahren (z.B. Mann-Whitney U-Test) verwandt, um statistisch signifikante Unterschiede zwischen den Teilstichproben zu identifizieren.

3.4.5.2 Sensitivität, Spezifität und Cut-Off-Werte

Zur Berechnung der diagnostischen Güte des Instruments wurden exemplarisch anhand eines in der Literatur vorgeschlagenen Cut-Off-Wertes die *Sensitivität*, die *Spezifität* und der *positive bzw. negative prädiktive Wert* berechnet.

Die Sensitivität ist der Prozentsatz richtiger, positiver Ergebnisse eines Untersuchungsverfahrens in Bezug auf das Vorhandensein der gesuchten Störung. Die Spezifität beschreibt den Prozentsatz von richtig negativen Ergebnissen. Außer von der Trennschärfe des Verfahrens hängen die Maße auch von der Festlegung des Trennkriteriums ab. Eine Steigerung der Sensitivität geht z.B. durch Anheben des Cut-Off-Werts zu Lasten der Spezifität und vice versa.

Weiterhin lässt sich der positive prädiktive Wert auch als positiver Vorhersagewert beschreiben. Er gibt an, wie viele der getesteten Personen, die durch den Test als krank beschrieben wurden, auch wirklich krank sind. Analog dazu berechnet sich der negative prädiktive Wert, der den Anteil der Gesunden an den negativ getesteten Personen angibt.

Berechnet wurden die Sensitivität und Spezifität anhand der im Manual vorgeschlagenen Cut-Off-Werte und es erfolgte eine Berechnung einer *Receiver Operating Characteristic* (ROC)-Kurve der einzelnen Skalen zur Bestimmung. In einer ROC-Analyse können für jeden möglichen Cut-Off-Wert die entsprechenden relativen Häufigkeitsverteilungen in Form von Sensitivität und Spezifität berechnet werden. Die ermittelten Werte werden in ein Koordinatensystem aufgetragen. Das Diagramm ergibt eine aufsteigende gekrümmte Kurve, unterhalb derer eine Diagonale als Referenz aufgetragen ist. Diese veranschaulicht eine Zufallsverteilung. Eine Kurve, die nahe der Diagonalen liegt, deutet auf eine niedrige Trennschärfe des Tests hin. Weiterhin lässt sich die Fläche unterhalb der Kurve berechnen. Werte zwischen 0,5 und 1 sind möglich, wobei 0,5 einer Zufallsverteilung entspricht und der Test keine Vorhersage ermöglichen würde (Hilgers et al. 2013).

3.4.5.3 Faktorielle Validität

Eine Faktorenanalyse versucht, die Zusammenhänge der Testitems untereinander durch zugrunde liegende homogene Faktoren bzw. latente Variablen zu erklären. Die ermittelten Faktoren können dann zur Modellbildung verwendet werden. Grundsätzlich unterscheidet man zwischen *konfirmatorischer Faktorenanalyse* (CFA) und *exploratorischer Faktorenanalyse* (EFA).

Mit Hilfe der EFA kann von den Variablen einer Stichprobe auf eine kleinere Anzahl zugrundeliegender unabhängiger Einflussgrößen – Faktoren genannt – schließen. Variablen, die untereinander stark korrelieren, werden zu einem Faktor zusammengeschlossen. Variablen aus verschiedenen Faktoren weisen somit eine geringere Korrelation untereinander auf als Variablen, denen derselbe Faktor zugrunde liegt. Ziel der Faktorenanalyse ist

die Identifizierung und Darstellung von Faktoren, welche die Zusammenhänge der Items untereinander möglichst vollständig beschreiben können.

Die Faktorenanalyse verläuft in mehreren Schritten. Zu Beginn einer Faktorenanalyse werden die Itemwerte mittels einer z-Transformation standardisiert, um dann die Inter-Item-Korrelationen mit einer Pearson-Korrelation zu errechnen und eine Korrelationsmatrix zu erstellen. Mittels Matrizenrechnung werden die sogenannten Eigenwerte und Eigenvektoren bestimmt und die Kommunalitäten geschätzt (Bortz et al. 2006).

In der Regel werden dann so viele Faktoren extrahiert wie Eigenwerte über einen Wert von 1 vorliegen. Jeder Eigenwert hat einen zugehörigen Eigenvektor. Dieser besteht aus den einzelnen Faktorladungen. Eine Faktorladung ist die Korrelation zwischen der Variablen und dem errechneten Faktor. Zur Berechnung der Faktoren wurden vielfältige Verfahren entwickelt. Das gängigste ist die Hauptkomponentenanalyse (Bühner 2011). Diese findet auch in dieser Studie Verwendung.

Als Cut-Off-Wert für die einzelnen Faktorladungen wurde 0,40 zur besseren Vergleichbarkeit der vorhandenen Literatur und aufgrund von Empfehlungen zur besseren Differenzierung von starken und schwachen Faktorladungen gewählt (Comrey und Lee 1992).

Zur Abschätzung der optimalen Faktorenzahl kann ein Scree-Test verwandt werden (auch „Ellenbogenkriterium“); mit diesem ist eine einfache, graphische Darstellung möglich. Hier werden die Eigenwerte der exploratorische Faktorenanalyse in ein Koordinatensystem eingebracht. Mit der Anzahl der Faktoren nehmen der Eigenwert und die zusätzlich aufgeklärte Varianz ab.

3.4.5.4 Statistische Analyse

Die statistische Auswertung und Aufbereitung wurde mit Hilfe der statistischen Software PASW Statistics 18.0 (ehemals SPSS, Statistical Package for the Social Sciences) für Mac OS X durchgeführt.

Zunächst wurde geprüft, ob die 18 Items der BSI signifikant von einer Normalverteilung abweichen. Nach Durchführung eines Kolmogorov-Smirnov-Test konnte die Nullhypothese (Normalverteilung) mit einer Irrtumswahrscheinlichkeit von $\alpha < 0,001$ (korrigiert nach Lilliefort) nicht bestätigt werden. Auch die Rohwerte der 4 Skalen und die Items der zur Bestimmung der Validität herangezogenen Fragebögen SF-12 und PHQ-D waren nicht normalverteilt. Eine Normalverteilung ist generell bei diskreten Antwortkategorien und Items unterschiedlicher Schiefe nicht zu erwarten.

Wegen der Abweichung von der Normalverteilung wurden nonparametrische Verfahren verwandt. Korrelative Zusammenhänge wurden nach Spearman berechnet. Gruppenvergleiche wurden nach Kruskal-Wallis und Mann-Whitney berechnet. Das Signifikanzniveau lag, sofern nicht anders angegeben, bei $p < 0,05$.

An wenigen Stellen wurden parametrische Verfahren verwandt, die streng genommen eine Normalverteilung voraussetzen; jedoch gelten z.B. die Ergebnisse eines T-Tests als robust bei dem Nichtvorliegen einer Normalverteilung, sofern keine größeren Ausreißer in den statistischen Daten vorliegen (Rasch und Guiard 2004). Einige Autoren empfehlen bei heterogenen Varianzen, unabhängig von der Verteilung der abhängigen Variablen, die Verwendung eines T-Tests zur Bestimmung der Mittelwertsunterschiede (Kubinger et al. 2009). Sofern nichtparametrische Verfahren verwandt wurden, erfolgte eine vergleichende Rechnung mit nonparametrischen Verfahren. Diese ergab keine signifikanten Unterschiede der Ergebnisse im Vergleich zu den entsprechenden parametrischen Verfahren. Für die verwendeten nonparametrischen Verfahren kann die Effektstärke nur geschätzt werden.

Für den wird Kruskal- Wallis-Test wird die Effektstärke Omega nach folgender Formel geschätzt:

$$w = \sqrt{\frac{\chi^2}{n-1}}$$

Ähnlich kann Phi zu Approximierung der Effektstärke eines Mann-Whitney-U-Tests herangezogen werden (Bühner 2011):

$$\phi = \sqrt{\frac{z^2}{N}}$$

Hierbei gilt eine Effektstärke von 0,1 als klein, 0,3 als mittel und 0,5 als groß (Cohen 1988).

4 Ergebnisse

4.1 Deskriptive Kennwerte und Ergebnisse der Itemanalyse

Die Mittelwerte der 18 Items des Fragebogens liegen in der vorliegenden Stichprobe ($N = 1086$) zwischen $M = 0,51$ (Item 17 „*Gedanken, sich das Leben zu nehmen*“) und $M = 1,96$ (Item 14 „*Hoffnungslosigkeit angesichts der Zukunft*“) und variieren somit deutlich. Die Standardabweichungen bewegen sich zwischen $SD = 0,95$ (Item 17 „*Gedanken, sich das Leben zu nehmen*“) und $SD = 1,49$ (Item 5 „*Einsamkeit*“), wobei geringere Mittelwerte tendenziell mit einer geringeren Streuung einhergehen.

Alle Items zeigen, wie bei einer Likert-Skala zu erwarten, eine rechtsschiefe bzw. linkssteile Verteilung ($v > 0$). Die Schiefe unterscheidet sich jedoch deutlich und reicht von $v = 1,65$ (Item 14 „*Hoffnungslosigkeit angesichts der Zukunft*“) bis $v = 2,13$ (Item 17 „*Gedanken, sich das Leben zu nehmen*“). Den niedrigsten Exzess und somit den flachgipfligsten Verlauf zeigt Item 14 („*Hoffnungslosigkeit angesichts der Zukunft*“ $\gamma^2 = 1,41$). Im Gegensatz dazu zeigt Item 17 („*Gedanken, sich das Leben zu nehmen*“) eine deutlich leptokurtische Verteilung ($\gamma^2 = 3,99$). Das Antwortformat wurde in allen Fällen komplett ausgenutzt.

Nimmt man den Mittelwert als Maß für die Schwierigkeit (di) der Items, ergeben sich deutlich unterschiedliche Schwierigkeiten, wobei jede der drei Subskalen Items mit unterschiedlichen Schwierigkeiten enthält und somit theoretisch eine gute Differenzierung zwischen Individuen mit hoher und niedriger Merkmalsausprägung gegeben ist. Die mittlere Itemschwierigkeit über alle Items beträgt 1,25. Die Gesamtschwierigkeit des Fragebogens ist folglich als relativ niedrig zu beurteilen, zumal es sich bei der Stichprobe um ein Patientenkollektiv handelt und tendenziell höhere Zustimmungen als in einer Bevölkerungsstichprobe erwartet werden. Von einer mittleren Schwierigkeit ist auszugehen, wenn ein Item von der Hälfte der Versuchspersonen positiv beantwortet wurde; in diesem Falle bei einem Mittelwert von 2. Bei einem Screeninginstrument werden tendenziell niedrigere Schwierigkeiten erwartet. Eine Übersicht über die deskriptiven und Verteilungskennwerte der einzelnen Items vermittelt **Tabelle 14**.

Tabelle 14: Deskriptive Statistik der BSI - 18 Items (N = 1086)

Item (Wie sehr litten Sie in den letzten 7 Tagen unter..?)	M	SD	Schiefe	Kurtosis
1. „Ohnmachts- oder Schwindelgefühlen“	0,88	1,151	1,304	0,812
2. „dem Gefühl, sich für nichts zu interessieren“	1,56	1,331	0,414	-1,057
3. „Nervosität oder innerem Zittern“	1,65	1,325	0,311	-1,081
4. „Herz- und Brustschmerzen“	0,79	1,088	1,317	0,856
5. „Einsamkeitsgefühlen“	1,75	1,494	0,246	-1,380
6. „dem Gefühl, gespannt oder aufgeregt zu sein“	1,86	1,328	0,156	-1,140
7. „Übelkeit oder Magenverstimmung“	1,24	1,336	0,779	-0,630
8. „Schwermut“	1,76	1,428	0,207	-1,308
9. „plötzlichem Erschrecken, ohne Grund“	0,72	1,151	1,536	1,256
10. „Schwierigkeiten beim Atmen“	0,88	1,140	1,257	0,682
11. „dem Gefühl, wertlos zu sein“	1,50	1,467	0,470	-1,223
12. „Schreck- oder Panikanfällen“	0,98	1,303	1,078	-0,146
13. „Taubheit oder Kribbeln in einzelnen Körperteilen“	1,07	1,280	0,922	-0,384
14. „einem Gefühl der Hoffnungslosigkeit angesichts Ihrer Zukunft“	1,96	1,478	0,065	-1,409
15. „so starker Ruhelosigkeit, dass Sie nicht mehr still sitzen können“	0,93	1,150	1,120	0,307
16. „Schwächegefühl in einzelnen Körperteilen“	1,38	1,369	0,568	-0,973
17. „Gedanken sich das Leben zu nehmen“	0,51	0,953	2,125	3,985
18. „Furchtsamkeit“	1,06	1,207	0,933	-0,166

Die mittlere Schwierigkeit der einzelnen Skalen unterscheidet sich mäßig, wobei die Skala DEPR ($d_i = 1,13$) schwieriger erscheint als die Skalen ANGS ($d_i = 1,9$) und SOMA ($d_i = 0,78$). Möglicherweise ist hier eine Verzerrung im Rahmen der unterschiedlichen Größe der Teilstichproben bzw. der Anzahl der psychischen Diagnosen möglich. Die beobachtete Schwierigkeitsrangfolge entspricht der Verteilung in der Stichprobe.

Die Trennschärfen (r_i) nach der part-whole Methode, d.h die korrigierte Item-Skala-Korrelation, liegen für den GSI zwischen $r_i = 0,39$ (Item 1 „Ohnmachts- oder Schwindelgefühle“) und $r_i = 0,66$ (Item 8 „Schwermut“). Somit zeigen alle Items mäßig bis hohe Trennschärfen. Bis auf die Items 1, 4, 7, 10 und 17 sind die Trennschärfen größer als 0,50 und somit als hoch zu bewerten, wobei die Items der Skala SOMA in Bezug auf den GSI tendenziell niedrigere Trennschärfen aufweisen. Dies zeigt sich auch bei der Betrachtung der Trennschärfen für die einzelnen Subskalen: Hier zeigt sich bei der Skala SOMA eine geringere Trennschärfe der Items ($r_i = 0,44$ bis $0,57$) als bei den Skalen DEPR ($r_i = 0,50$ bis $0,75$) und ANGS ($r_i = 0,53$ bis $0,67$). Die Trennschärfen der beiden zuletzt genannten Skalen sind ausnahmslos als hoch zu bewerten (vgl. zusammenfassend Tabelle 15).

Tabelle 15: Korrigierte Item-Skala-Korrelation bzw. Tensschärfe (r_i) des BSI-18 nach Skalen (N = 1086)

Item	GSI	SOMA	DEPR	ANGS
1	0,391	0,442	-	-
2	0,588	-	0,602	-
3	0,646	-	-	0,677
4	0,459	0,521	-	-
5	0,566	-	0,664	-
6	0,625	-	-	0,665
7	0,446	0,392	-	-
8	0,658	-	0,709	-
9	0,547	-	-	0,534
10	0,393	0,446	-	-
11	0,615	-	0,750	-
12	0,582	-	-	0,639
13	0,385	0,504	-	-
14	0,651	-	0,743	-
15	0,560	-	-	0,563
16	0,521	0,570	-	-
17	0,445	-	0,503	-
18	0,646	-	-	0,645

In **Tabelle 16** ist die durchschnittliche Inter-Item-Korrelation als Maß für die Homogenität dargestellt. Ein Wert zwischen 0,20 und 0,40 gilt als erstrebenswert (Bühner 2011). Die Varianz ist in allen Skalen gering. Es kommen keine negativen Korrelationen vor, wie bei gleichgerichtet gepolten Items zu erwarten. Die durchschnittliche Inter-Item-Korrelation liegt für die Skalen GSI und SOMA niedriger (0,33 und 0,32) als für die Skalen ANGS und DEPR (0,47 und 0,51). Die Skala GSI ist als globaler Index weniger homogen als die drei Subskalen.

Tabelle 16: Inter-Item-Korrelationen der BSI-18 Skalen (N = 1086)

Skala	Mittelwert	Min	Max
GSI	0,326	0,104	0,662
SOMA	0,324	0,198	0,499
DEPR	0,512	0,333	0,662
ANGS	0,469	0,365	0,634

4.2 Reliabilitätsanalyse

Zur Bestimmung der Reliabilität wurde eine Analyse der internen Konsistenz durchgeführt. **Tabelle 17** illustriert die Ergebnisse der Analyse. Es zeigten sich für die Skalen interne Konsistenzen zwischen $\alpha = 0,74$ und $\alpha = 0,89$. Nach Fisseni (2004) ist die Reliabilität des GSI, operationalisiert über die interne Konsistenz, somit als mittel bis hoch einzuschätzen ($\alpha = 0,90$). Die Werte der Skalen DEPR ($\alpha = 0,87$) und ANGS ($\alpha = 0,84$) zeigen eine mittlere Messgenauigkeit. Die Skala SOMA zeigt den geringsten Wert mit $\alpha = 0,74$; somit ist diese Dimension verhältnismäßig ungenau und die Reliabilität als niedrig zu bewerten.

Es zeigen sich keine großen Ausreißer der Teilstichproben nach Bildungsstand und Geschlecht (vgl. **Tabelle 18**), es lässt sich keine Gruppe ausmachen, bei der die Reliabilität für alle 4 Skalen substanziell besser bzw. schlechter ausfällt. Die Werte für die interne Konsistenz scheinen vielmehr um die Werte der Gesamtstichprobe zu streuen. Insgesamt zeigen sich keine Differenzen zur Gesamtstichprobe größer als $\alpha = 0,08$. Die Reliabilität ist für die Teilstichproben, mit Ausnahme der Skala SOMA, als mittel bis hoch einzuschätzen. Somit fanden sich in dieser Stichprobe keine Hinweise für geschlechts- oder bildungsspezifische Reliabilitätsunterschiede.

Tabelle 17: Skalenstatistiken und Cronbachs Alpha der 4 Subskalen des BSI-18 (N = 1086)

Skala	Mittelwert	SD	Cronbachs α
GSI	22,51	13,98	0,90
SOMA	6,25	4,88	0,74
ANGS	7,22	5,59	0,84
DEPR	9,04	6,36	0,87

Tabelle 18: Reliabilität nach Geschlecht und Bildungsstand (Cronbachs α)

Teilstichprobe	GSI	SOMA	DEPR	ANGS
Männlich (n = 335)	0,89	0,77	0,85	0,83
Weiblich (n = 742)	0,90	0,72	0,87	0,85
Volksschule /Hauptschule (n = 218)	0,91	0,76	0,86	0,89
Realschule (n = 356)	0,89	0,70	0,86	0,83
Abitur (n = 249)	0,90	0,78	0,86	0,81
Akademiker (n = 187)	0,90	0,73	0,88	0,84

4.3 Auswertung der Testergebnisse

Der Mittelwert für den GSI liegt bei 22,51 ($SD = 13,98$). Im Mittel erreichten die Testpersonen höhere Werte auf der Skala DEPR ($M = 9,04$; $SD = 6,36$) als auf den Skalen ANGS ($M = 7,22$; $SD = 5,69$) und SOMA ($M = 6,25$; $SD = 4,88$). Da es sich bei

der Stichprobe um ein Patientenkollektiv handelt, ist dies auch durch die unterschiedliche Verteilung der Diagnosen erklärt, wobei depressive Diagnosen (*F33* 18,2% und *F32* 40,7%) zahlenmäßig führen, gefolgt von Somatisierungsstörungen (*F45* 27,5%) und zuletzt Angsterkrankungen (*F40* 7,9% und *F41* 15,6%). Dies würde tendenziell höhere Werte auf der Skala DEPR erklären; die Ergebnisse bilden die zahlenmäßige Verteilung der restlichen Diagnosen nicht im Sinne einer entsprechenden Rangfolge der Mittelwerte ab (vgl. **Tabelle 17**).

4.3.1 Korrelation der Skalen

Die Skalen des BSI-18 korrelieren deutlich miteinander. Die höchsten Korrelationen, berechnet nach Spearman, zeigt der GSI zu den Skalen ANGS, SOMA und DEPR ($r = 0,87$; $0,74$ und $0,85$). Es kann somit angenommen werden, dass ANGS und DEPR zum globalen Index mehr beitragen als die Items der Skala SOMA. Etwas geringer sind die Korrelationen zwischen den Subskalen. Die Skala SOMA korreliert etwas höher mit der Skala ANGS ($r = 0,56$) als mit der Skala DEPR ($r = 0,41$). Die Skala ANGS zeigt die höchste Korrelation der Subskalen untereinander mit der Skala DEPR ($r = 0,62$). Somit lässt sich schließen, dass für Patienten, die hohe Testwerte auf einer Subskala erreichen, tendenziell auch erhöhte psychische Belastungen in den anderen zwei Dimensionen zu erwarten sind.

Tabelle 19: Korrelation nach Spearman der Skalen untereinander (N =1086)

	GSI	SOMA	DEPR
SOMA	0,74**	-	-
DEPR	0,85**	0,41**	-
ANGS	0,87**	0,556**	0,619**

**Alle Korrelationen nach Spearmans ρ sind auf dem 0,01 Niveau signifikant (zweiseitig).

4.3.2 Geschlechtsunterschiede und Alterseffekte

Untersucht man die vier Summenskalen des BSI-18 auf Unterschiede bezüglich des Geschlechts, zeigen sich für drei der vier Skalen unter Verwendung eines Mann-Whitney-U-Tests signifikante Unterschiede, die geschätzte Effektstärke ist jeweils sehr gering. In allen vier Dimensionen erreichten die weiblichen Versuchspersonen im Mittel höhere Testwerte und zeigten somit eine höhere Belastung (vgl. **Tabelle 20**). Die mittlere Differenz betrug 2,38 Punkte für den GSI ($p = 0,023$), 0,59 Punkte für die Skala SOMA ($p = 0,043$), 1,01 Punkte für die Skala DEPR ($p = 0,033$) und 0,77 Punkte für die Skala ANGS ($p = 0,094$). Unter der Annahme einer gerichteten Nullhypothese („weibliche Testpersonen

sind höher belastet“), die aufgrund der bekannten Tendenz in der Literatur gerechtfertigt scheint, ergibt sich bei einer einseitigen Testung auch für die Skala ANGS ein signifikanter Wert ($p = 0,047$).

Tabelle 20: Skalen nach Geschlecht mit Mann-Whitney-U-Test Signifikanzniveau (2-seitig)

Skala/Statistik	Geschlecht	n	Mittelwert	SD	Mittlerer Rang
SOMA					
p = 0,043	männlich	335	5,82	4,82	511
z = - 2,02					
φ = 0,061	weiblich	742	6,41	4,86	552
DEPR					
p = 0,033	männlich	335	8,343	5,88	509
z = -2,136					
φ = 0,065	weiblich	742	9,36	6,56	553
ANGS					
p = 0,094	männlich	335	6,67	5,15	515
z = - 1,674					
φ = 0,051	weiblich	742	7,45	5,78	550
GSI					
p = 0,023	männlich	335	20,76	12,76	507
z = 2,270					
φ = 0,069	weiblich	742	23,09	14,35	554

Tabelle 21: GSI und Subskalenwerte des BSI-18 differenziert nach Alter

Alter (N = 1084)		GSI	SOMA	DEPR	ANGS
Gruppe 1	M	21,52	5,50	9,42	6,60
(n = 304)	SD	13,55	4,83	6,27	5,11
<= 30 Jahre	Mttl. Rang	522	488	564	514
Gruppe 2	M	23,00	6,35	9,04	7,61
(n = 503)	SD	14,13	4,95	6,31	5,84
31 – 50 Jahre	Mttl. Rang	553	547	544	561
Gruppe 3	M	22,62	6,91	8,53	7,18
(n = 277)	SD	14,17	4,70	6,50	5,62
> 51 Jahre	Mttl. Rang	545	594	516	540
Statistik	p (Kruskal-Wallis-H)	0,391	0,000	0,187	0,121
	X ²	1,876	16,896	3,349	4,218
	Effektstärke (w)	-	0,125	-	-
	Mann-Whitney	-	3>2>1	-	2>1

Zur Prüfung von Unterschieden der Testergebnisse in Bezug auf das Alter der Patienten wurde ein Kruskal-Wallis-Test mit post-hoc-Analyse mittels Mann-Whitney-U-Test durchgeführt. Vorher wurden die Patienten in drei Altersgruppen eingeteilt. Eine vergleichende Varianzanalyse zeigte für die Skalen SOMA ($p = 0,002$) und ANGS ($p = 0,045$) signifikante Unterschiede bezüglich der drei Altersklassen. Im Kruskal-Willis-Test waren nur Unterschiede der Skala SOMA signifikant. Die daraufhin durchgeführte post-hoc-Analyse fand signifikante Rangunterschiede in Bezug auf die Skala SOMA mit aufsteigendem Alter. Die älteren Patienten hatten im Mittel einen Testwert, der 1,41 Punkte über dem

der Jungen lag. Die geschätzte Effektstärke ($w = 0,125$) deutet auf einen eher geringen Effekt hin. Auf der Skala ANGS zeigte sich die Gruppe zwischen 30 und 50 in der Post-hoc-Analyse signifikant belasteter als die jüngste Altersklasse. Eine synoptische Übersicht zu diesen Analysen bietet **Tabelle 21**.

4.3.3 Zusammenhänge mit dem Bildungsstand

Ein Kruskal-Wallis-Test zur Beschreibung des Einflusses des Bildungsstandes auf die Höhe der psychischen Belastung, unter Ausschluss von kleineren Teilstichproben (z.B. Schüler; $N = 8$), brachte keine signifikanten Ergebnisse (vgl. **Tabelle 22**).

Tabelle 22: BSI-18 Skalen nach Bildungsstand mit Kruskal-Wallis-Test ($N = 1028$)

Kruskal-Wallis	Ihr höchster Schulabschluss?	n	Mittlerer Rang
GSI (n.s. / p = .995)	Haupt-/Volksschulabschluss	218	518,77
	Realschulabschluss	374	511,99
	Fachabitur/Abitur	249	514,00
	abgeschlossenes (Fach-) Hochschulstudium	187	515,22
SOMA (n.s. / p = .147)	Haupt-/Volksschulabschluss	218	551,01
	Realschulabschluss	374	515,97
	Fachabitur/Abitur	249	490,23
	abgeschlossenes (Fach-) Hochschulstudium	187	501,31
DEPR (n.s. / p = .636)	Haupt-/Volksschulabschluss	218	506,63
	Realschulabschluss	374	521,25
	Fachabitur/Abitur	249	526,89
	abgeschlossenes (Fach-) Hochschulstudium	187	493,67
ANGS (n.s. / p = .207)	Haupt-/Volksschulabschluss	218	503,89
	Realschulabschluss	374	497,76
	Fachabitur/Abitur	249	520,84
	abgeschlossenes (Fach-) Hochschulstudium	187	551,90

4.3.4 Vergleich von Patienten mit und ohne Migrationshintergrund

Eine Betrachtung der Testergebnisse in Bezug auf unterschiedliche kulturelle Herkunft der Patienten ist nur eingeschränkt möglich. Es wurde nur nach der Staatsangehörigkeit und nicht nach einem möglichen Migrationshintergrund gefragt. Die Größen der Teilstichproben der übrigen Nationalitäten waren ausnahmslos im unteren einstelligen Bereich, so dass eine Auswertung nur zwischen Patienten mit deutscher Staatsbürgerschaft und der Gesamtheit der übrigen Nationalitäten möglich war. Es fanden sich, bis auf die Skala DEPR, signifikante Gruppenunterschiede im Mann-Whitney-U-Test mit einer stärkeren Belastung von Personen ohne deutsche Staatsbürgerschaft.

Tabelle 23: BSI-18 Skalenunterschiede differenziert nach Nationalität

Nationalität		GSI	SOMA	DEPR	ANGS
Deutsch (n = 991)	M	22,08	6,05	8,98	7,04
	SD	13,72	4,82	6,31	5,49
	Mttl. Rang	531	527	536	529
andere Nationalität (n = 86)	M	27,50	8,35	9,80	9,35
	SD	15,72	4,90	6,89	6,24
	Mttl. Rang	635	678	570	646
Mann-Whitney-Test	p	0,003	0,000	0,336	0,001
	Φ	0,091	0,132	-	0,102
	Z	-2,994	-4,335	-0,963	-3,324

4.3.5 Zusammenhänge mit Arbeitsunfähigkeit und Berentung

Als indirekten Hinweis auf die kriteriumsbezogene Validität wurden die Skalenwerte in Hinblick auf die aktuelle berufliche Situation untersucht. Erwartungsgemäß fanden sich erhöhte Belastungen bei gegenwärtig arbeitsunfähigen Personen in allen Dimensionen, wobei sich hier die höchste Effektstärke (leicht bis moderater Effekt) der Subskalen bei den somatischen Symptomen und dem GSI zeigte (vgl. **Tabelle 24**).

Tabelle 24: BSI-18 Skalenunterschiede differenziert nach Arbeitsunfähigkeit

Sind sie aktuell arbeitsunfähig?		GSI	SOMA	DEPR	ANGS
Nein (N = 565)	M	19,29	5,17	7,86	6,25
	SD	13,15	4,57	6,05	5,22
	Mttl. Rang	422	426	431	438
Ja (N = 397)	M	26,44	7,32	10,73	8,39
	SD	13,90	4,77	6,49	5,77
	Mttl. Rang	566	560	554	544
p (Mann-Whitney-U)		,000	,000	,000	,000
Z		-7,868	-7,384	-6,761	-5,859
ϕ		0,253	0,238	0,218	0,189

Patienten mit einer Altersrente oder einem laufendem Rentenverfahren bzw. Frührentner waren bei den ängstlichen Symptomen nicht signifikant höher belastet als Nichtberentete, wie aus **Tabelle 25** ersichtlich. Patienten mit einem laufenden Verfahren waren somatisch höher belastet und depressiver, während Altersrentner weniger hohe Werte auf der Skala *Depressivität* errangen als Nichtberentete. Die höchste Belastung in der Selbstbeurteilung der letzten 2 Wochen lag in der Gruppe der zeitlich befristet Berenteten bzw. bei Personen mit einem Rentenbegehren vor. Dies ist für die Skalen GSI, DEPR und SOMA signifikant. Gemessen an den Effektstärken lag der Schwerpunkt der Belastung bei den depressiven Symptomen. Die Effekte waren jeweils leichtgradig.

Tabelle 25: BSI-18 Skalenunterschiede differenziert nach Berentung (N=1015)

„Sind sie berentet?“		GSI	SOMA	DEPR	ANGS
Nein (1)	M	22,21	6,02	9,00	7,19
	SD	13,81	4,80	6,29	5,53
	Mttl. Rang	507	501	511	509
Altersrente (2)	M	19,70	6,31	6,99	6,40
	SD	13,98	5,00	5,88	5,46
	Mttl. Rang	450	516	416	464
Zeitrente /laufendes Verfahren (3)	M	26,56	7,41	10,83	8,32
	SD	15,26	4,75	6,96	6,58
	Mttl. Rang	590	593	587	549
Statistik	p (Kruskal-Wallis-H)	0,014	0,044	0,001	0,196
	X ²	8,579	6,233	13,338	3,260
	Effektstärke (w)	0,092	0,078	0,115	0,057
		1<3	1<3	3>1>2	1=2=3
	Mann-Whitney	2<3	2=3		
		1=2	1=2		

4.3.6 Zusammenhänge mit Familienstand und Wohnsituation

Tabelle 26 zeigt, dass sich signifikante Gruppenunterschiede in Abhängigkeit von Familienstand für die Skalen GSI, SOMA und DEPR fanden. Ledige Patienten waren auf der Skala SOMA gegenüber den anderen Gruppen niedriger belastet. Hier ist ein Bias auf Grund des niedrigeren Lebensalters anzunehmen. Insbesondere im Bezug auf depressive Symptome zeigen die Daten eine deutlich niedrigere Belastung mit einer leichten Effektstärke der verheirateten gegenüber den unverheirateten Personen. Bezüglich des GSI waren geschiedene oder verwitwete Personen höher belastet als verheiratete Personen. Alleinlebende Patienten erreichten höhere Werte auf dem GSI und waren depressiver als Patienten, die mit einem Lebenspartner zusammen wohnten. Wohnten die Personen noch bei ihren Eltern, waren sie bezüglich der depressiven Symptome ebenfalls höher belastet als mit einem Lebenspartner (vgl. *Tabelle 27*).

Tabelle 26: Skalen nach Familienstand (N = 1052)

Familienstand		GSI	SOMA	DEPR	ANGS
Ledig (Ø 32,9 J.) n = 503	M	22,50	5,78	9,68	7,04
	SD	13,79	4,82	6,39	5,42
	Mttl. Rang	530	495	557	520
Verheiratet (Ø 47,5 J.) n = 341	M	21,35	6,53	7,70	7,11
	SD	14,32	4,93	6,18	5,74
	Mttl. Rang	496	545	462	518
geschieden/verwitwet/in Trennung lebend (Ø 51,5 J.) n = 208	M	24,44	6,91	9,72	7,81
	SD	14,21	4,92	6,50	5,85
	Mttl. Rang	569	571	558	556
	p (Kruskal-Wallis-H)	,022	,004	,000	,289
	X ²	7,625	11,069	22,702	2,484
	Effektstärke (w)	0,085	0,103	0,147	-
	Mann-Whitney	2=1	2>1	1>2	2=1
		3=1	3>1	3=1	3=1
		3>2	2=3	3>2	2=3

Tabelle 27: Skalen nach Wohnsituation (N = 1012)

Wohnsituation		GSI	SOMA	DEPR	ANGS
allein (+Kinder) n = 411 (1)	M	23,46	6,04	10,18	7,25
	SD	13,52	4,59	6,16	5,58
	Mttl. Rang	537	502	566	514
mit Partner (+Kinder) n = 506 (2)	M	21,19	6,39	7,78	7,03
	SD	14,28	5,03	6,22	5,64
	Mttl. Rang	479	514	452	500
mit den Eltern n = 95 (3)	M	22,88	6,08	9,85	6,95
	SD	14,59	5,41	7,06	5,17
	Mttl. Rang	520	484	541	507
	p (Kruskal-Wallis-H)	0,011	0,591	0,000	0,784
	X ²	8,959	1,053	36,078	,488
	Effektstärke (w)	0,094	-	0,189	-
	Mann-Whitney	1>2	2=1	1>2	2=1
		3=1	3=1	3=1	3=1
		3=2	2=3	3>2	2=3

4.3.7 Einfluss von Arztkontakten und Psychotherapieerfahrung

Patienten, die in der Vergangenheit psychotherapeutisch behandelt wurden, waren in allen Dimensionen höher belastet als Patienten ohne Therapieerfahrungen. Befanden sich die Testpersonen zum Zeitpunkt der Erhebung in einer Psychotherapie, erreichten sie auf allen Skalen des BSI-18 im Mittel die höchsten Werte (vgl. **Tabelle 28**), wobei die Effektstärken des GSI und der Skala DEPR am größten waren (leicht bis mäßig).

Mit häufigeren Arztkontakten steigt auch der Grad der psychischen Belastung. Patienten, die mehr als zweimal in den letzten zwei Wochen einen Arzt aufgesucht hatten, waren in allen Dimensionen höher belastet, wie in **Tabelle 29** dargestellt. Betrachtet man die Effektstärken, führen vor allem somatische Beschwerden und Angsterkrankungen zu ver-

mehrten Arztkontakten ($w = 0,245$ und $0,190$). Patienten, die den Arzt nicht aufgesucht hatten, waren entsprechend weniger belastet.

Tabelle 28: Skalen nach Psychotherapieerfahrungen (N = 1058)

Psychotherapieerfahrung		GSI	SOMA	DEPR	ANGS
noch nie (1) (n = 435)	M	18,90	5,42	7,42	6,06
	SD	13,41	4,74	6,10	5,30
	Mttl. Rang	446	473	448	463
früher (2) (n = 447)	M	23,90	6,39	9,89	7,62
	SD	13,71	4,83	6,27	5,52
	Mttl. Rang	561	541	569	555
zur Zeit (3) (n = 176)	M	28,24	7,93	11,20	9,11
	SD	13,74	4,93	6,28	5,89
	Mttl. Rang	654	639	630	630
	p (Kruskal-Wallis-H)	,000	,000	,000	,000
	X ²	66,434	38,221	57,743	43,302
	Effektstärke (w)	0,251	0,190	0,234	0,202
	Mann-Whitney	3>2>1	3>2>1	3>2>1	3>2>1

Tabelle 29: Skalen nach Arztkontakten in den letzten 2 Wochen (N = 1045)

Anzahl der Arztbesuche in den letzten 14 Tagen?		GSI	SOMA	DEPR	ANGS
Kein Mal (1) (n = 76)	M	16,08	4,33	6,87	4,88
	SD	12,06	4,71	5,36	4,85
	Mttl. Rang	378	387	423	386
Einmal (2) (n = 243)	M	21,59	5,39	9,25	6,95
	SD	13,99	5,05	6,12	5,63
	Mttl. Rang	500	459	534	506
Zweimal (3) (n = 301)	M	20,60	5,65	8,52	6,43
	SD	13,23	4,57	6,36	5,00
	Mttl. Rang	484	491	496	487
Zweimal und mehr (4) (n = 425)	M	25,47	7,37	9,74	8,36
	SD	14,02	4,64	6,56	5,82
	Mttl. Rang	589	607	553	583
	p (Kruskal-Wallis-H)	,000	,000	,001	,000
	X ²	44,493	62,803	15,490	37,795
	Effektstärke (w)	0,206	0,245	0,122	0,190
	Mann-Whitney	4>3	4>3	4>3	4>3
		4>2	4>2	4=2	4>2
		4>1	4>1	4>1	4>1
		3=2	3=2	3=2	3=2
		3>1	3>1	3=1	3>1
		2>1	2=1	2>1	2>1

4.4 Kriteriumsbezogene Validität

4.4.1 Selbsteinschätzung

Das Set von Fragebögen enthielt unter anderem mehrere Fragen zum allgemeinen und psychischen Gesundheitszustand der Patienten (vgl. Abschnitt 3.2). Es wurden Korrelationen zu den BSI-18 Skalen berechnet, die zusammenfassend in **Tabelle 30** dargestellt sind. Die Einschränkung der Leistungsfähigkeit korrelierte erwartungsgemäß negativ und signifikant mit allen vier Skalen. Schmerzen korrelierten am höchsten mit der Skala SOMA ($r = 0,383$) und numerisch deutlich geringer mit den anderen BSI-18-Skalen. Die Selbsteinschätzung in Bezug auf Depression zeigte hohe Korrelationen zu den Skalen GSI ($r = 0,636$) und DEPR ($r = 0,663$), mittlere zur Skala ANGS ($r = 0,534$) und eine deutlich geringere zur Skala SOMA ($r = 0,327$). Zur Angst fanden sich hohe Korrelationen zur Skala ANGS ($r = 0,610$) und mittlere zu den Skalen GSI ($r = 0,577$) und DEPR ($r = 0,471$), sowie eine geringe Korrelation zur Skala SOMA ($r = 0,337$). Insgesamt stützen die Ergebnisse die kriteriumsbezogene Validität des Instruments. Die Einschätzung des körperlichen und seelischen Gesundheitszustands zeigte vergleichbare Ergebnisse zu den Korrelationen mit den Summenskalen des SF-12 (vgl. den nachfolgenden Abschnitt 4.4.3). Subjektive Einschränkungen der körperlichen Gesundheit korrelierten am deutlichsten mit der Skala SOMA ($r = 0,458$), während Einschränkungen der seelischen Gesundheit die höchsten Korrelationskoeffizienten mit dem GSI ($r = 0,640$) sowie den Skalen DEPR ($r = 0,605$) und ANGS ($r = 0,550$) zeigten (vgl. **Tabelle 31**).

Tabelle 30: Korrelation der Skalen mit der Selbstauskunft nach dem Befinden in den letzten 2 Wochen (auf einer Skala von 1-10, N = 1080)

	GSI	SOMA	DEPR	ANGS
Leistungsfähigkeit	-,450**	-,345**	-,414**	-,342**
Schmerzen	,224**	,383**	,101**	,136**
Depression	,636**	,327**	,663**	,534**
Ängste	,577**	,337**	,471**	,610**

**Alle Korrelationen nach Spearmans ρ sind auf dem 0,01 Niveau signifikant (zweiseitig).

Tabelle 31: Korrelation der Skalen mit „Wie beurteilen Sie Ihren Gesundheitszustand in den letzten 2 Wochen?“ (N = 1080)

	GSI	SOMA	DEPR	ANGS
Körperlicher Zustand	,310**	,452**	,150**	,208**
Seelischer Zustand	,640**	,397**	,605**	,550**

**Alle Korrelationen nach Spearmans-Rho sind auf dem 0,01 Niveau signifikant (zweiseitig).

4.4.2 BSI-18-Skalenwerte in Abhängigkeit der Diagnose nach ICD-10

Der Zusammenhang zwischen den im Erstgespräch gestellten klinischen Diagnosen nach ICD und den BSI-18-Werten wird nachfolgend dargestellt. Die Mittelwerte wurden für die Skalen in Abhängigkeit der einzelnen Diagnosen errechnet und der Unterschied mittels T-Tests für unabhängige Stichproben mit Bonferroni-Korrektur zur Verhinderung einer Alphafehler-Kumulierung auf Signifikanz untersucht. **Tabelle 32** stellt die Ergebnisse der Analyse dar. Aus Gründen der Übersichtlichkeit werden die Ergebnisse verkürzt dargestellt.

Es fanden sich in den parametrischen Verfahren signifikante ($p < 0.05$) Mittelwertsunterschiede für eine Vielzahl der untersuchten Diagnosen. Patienten mit einer *Alkoholabhängigkeit (F10)* und *depressiven Erkrankungen (F32-F24)* hatten signifikant höhere Testergebnisse auf allen vier Skalen. Patienten mit einer *Angststörung (F40-F41)* hatten signifikant höhere Belastungen auf den Skalen ANGS, GSI und SOMA. Die Diagnose *somatoforme Störung (F45)* ging mit signifikant erniedrigten Testwerten in der Skala DEPR und erhöhten Testwerten in der Dimension SOMA einher. *Essstörungen* waren mit höheren Scores bei DEPR vergesellschaftet. Personen mit einer *Persönlichkeitsstörung* hatten erhöhte Testwerte auf den Skalen GSI, DEPR und ANGS. Patienten mit einer *Adipositas (E66)* waren bis auf die Skala SOMA überall signifikant niedriger belastet. Keine Gruppenunterschiede zeigten sich für die Diagnosen: *Reaktionen auf schwere Belastungen und Anpassungsstörungen (F43)* und *psychische Faktoren und Verhaltenseinflüsse bei andernorts klassifizierten Krankheiten (F54)*.

Für die drei Diagnosengruppen (depressive Erkrankungen, Angststörungen und Somatisierungsstörungen), die der intendierten dimensional Struktur des BSI-18 entsprechen, erfolgten die Berechnungen zusätzlich mit nonparametrischen Verfahren.

Es wurden jeweils die an dieser Störung Nichterkrankten und die Erkrankten verglichen. Zu beachten ist in diesem Fall, dass die Vergleichsgruppe der Nichterkrankten aus einem Patientenkollektiv besteht, somit gibt es keine Personen ohne psychiatrische oder somatische Diagnose. Es liegen daher wahrscheinliche Belastungen in anderen Bereichen vor, somit können Unterschiede durch das Vorhandensein der untersuchten Störung oder durch das Nichtvorhandensein einer anderen erklärt werden. Patienten mit einer *Angststörung (F40-F41)* erreichten signifikant höhere Werte auf den Skalen GSI, SOMA und ANGS. Übereinstimmend mit der postulierten Skalenstruktur fand sich der größte Unterschied für die Skala ANGS ($\varphi = 0,275$), entsprechend einer mittleren Effektstärke. Hö-

here Werte auf allen vier Skalen, die im Vergleich zwischen Patienten mit *depressiven Erkrankungen* und den Patienten ohne diese Diagnose gefunden wurden, sind statistisch signifikant. Auch hier punkten die Patienten skalenkonform mit dem eindeutigsten Unterschied auf der Skala DEPR ($\varphi = 0,404$). Die Effektstärke ist hier als mittel bis hoch zu bewerten. Patienten mit einer *somatoformen Störung (F45)* hatten signifikant höhere Werte der Skala SOMA und signifikant niedrigere auf der Skala DEPR.

Tabelle 32: BSI-18 Mittlere Testergebnisse in Abhängigkeit der im Erstgespräch gestellten Diagnose

Diagnose			GSI	SOMA	DEPR	ANGS
F1x	Nein (n = 1010)	M	21,97 _a	6,10 _a	8,84 _a	7,03 _a
		SD	13,83	4,83	6,32	5,53
	Ja (n = 76)	M	29,67 _b	8,33 _b	11,68 _b	9,66 _b
		SD	14,08	5,07	6,33	5,86
F32-F34	Nein (n = 405)	M	16,32 _a	5,10 _a	5,80 _a	5,42 _a
		SD	12,16	4,37	5,40	4,99
	Ja (n = 681)	M	26,19 _b	6,94 _b	10,96 _b	8,28 _b
		SD	13,69	5,03	6,10	5,66
F40-F41	Nein (n = 845)	M	21,12 _a	5,80 _a	8,90 _a	6,43 _a
		SD	13,91	4,73	6,46	5,38
	Ja (n = 241)	M	27,36 _b	7,85 _b	9,53 _a	9,99 _b
		SD	13,12	5,05	5,96	5,42
F43	Nein (n = 972)	M	22,51 _a	6,28 _a	9,03 _a	7,19 _a
		SD	13,95	4,87	6,34	5,60
	Ja (n = 114)	M	22,48 _a	6,00 _a	9,06 _a	7,42 _a
		SD	14,23	4,94	6,55	5,55
F45	Nein (n = 787)	M	22,62 _a	5,86 _a	9,57 _a	7,19 _a
		SD	13,97	4,78	6,44	5,59
	Ja (n = 299)	M	22,22 _a	7,28 _b	7,64 _b	7,30 _a
		SD	14,00	4,99	5,91	5,59
F54	Nein (n = 1020)	M	22,63 _a	6,27 _a	9,08 _a	7,27 _a
		SD	14,03	4,86	6,38	5,61
	Ja (n = 66)	M	20,67 _a	5,97 _a	8,30 _a	6,39 _a
		SD	13,02	5,08	6,03	5,32
F6x	Nein (n = 981)	M	22,02 _a	6,23 _a	8,69 _a	7,11 _a
		SD	13,85	4,85	6,33	5,51
	Ja (n = 105)	M	27,07 _b	6,50 _a	12,31 _b	8,25 _b
		SD	15,42	5,20	6,16	6,66
E66	Nein (n = 955)	M	23,08 _a	6,29 _a	9,27 _a	7,53 _a
		SD	13,85	4,89	6,38	5,57
	Ja (n = 131)	M	18,31 _b	6,02 _a	7,35 _b	4,94 _b
		SD	14,25	4,78	5,96	5,18

Anmerkung: Werte in der gleichen Zeile und Untertabelle, bei denen das Subskript nicht identisch ist, unterscheiden sich stark bei $p < 0.05$ im zweiseitigen Test auf Gleichheit für Spaltenmittel.

Tabelle 33: BSI-18 Skalenunterschiede bei Patienten mit einer depressiven Störung (F32/F33/F34)

depressive Störung		GSI	SOMA	DEPR	ANGS
Nein (n = 405)	M	16,33	5,10	5,80	5,42
	SD	12,16	4,37	5,40	4,99
	Mttl. Rang	400	470	380	438
Ja (n = 681)	M	26,19	6,94	10,96	8,28
	SD	13,69	5,03	6,10	5,66
	Mttl. Rang	629	587	641	607
p (Mann-Whitney-U)		,000	,000	,000	,000
Z		-11,62	-5,98	-13,30	-8,58
φ		0,353	0,181	0,404	0,260

Tabelle 34: BSI-18 Skalenunterschiede bei Patienten mit einer somatoformen Störung

somatoforme Störung		GSI	SOMA	DEPR	ANGS
Nein (n = 787)	M	22,62	5,86	9,57	7,19
	SD	13,97	4,78	6,44	5,59
	Mttl. Rang	546	518	569	541
Ja (n = 299)	M	22,22	7,28	7,64	7,30
	SD	14,00	4,99	5,91	5,59
	Mttl. Rang	537	612	476	549
p (Mann-Whitney-U)		,680	,000	,000	,720
Z		-0,41	-4,44	-4,40	-0,36
φ		0,012	0,135	0,134	0,011

Tabelle 35: BSI-18 Skalenunterschiede bei Patienten mit einer Angststörung (F40/F41)

Angststörung		GSI	SOMA	DEPR	ANGS
Nein (n = 845)	M	21,12	5,80	8,90	6,43
	SD	13,91	4,73	6,46	5,38
	Mttl. Rang	510	514	535	497
Ja (n = 241)	M	27,26	7,85	9,53	9,99
	SD	13,12	5,05	5,96	5,42
	Mttl. Rang	660	647	573	705
p (Mann-Whitney-U)		,000	,000	,095	,000
Z		-6,52	-5,80	-1,70	-9,07
φ		0,198	0,176	0,051	0,275

Zusammenfassend fanden sich signifikant höhere Werte und die deutlichsten Effektstärken in den entsprechenden Skalen, dies kann als Hinweis für die kriteriumsbezogene Validität gewertet werden. Bei zwei der drei Teilstichproben fanden sich zusätzlich signifikant erhöhte Punktwerte für den GSI. Es fanden sich weiterhin (außer bei der *somatoformen Störung*) auch signifikant erhöhte Belastungen in den anderen Dimensionen, die jedoch geringer ausfielen (vgl. **Tabelle 33**, **Tabelle 34** und **Tabelle 35**)

4.4.3 Vergleich zwischen BSI-18, PHQ-D und SF-12

Zur Abschätzung der kriteriumsbezogenen Validität erfolgte zudem der Vergleich der Ergebnisse zwischen den Instrumenten BSI-18, PHQ-D und SF-12. In einigen Fällen wurde der PHQ-D unvollständig ausgefüllt, so dass nicht alle Skalen auswertbar waren; daher wird in der Tabelle die Fallzahl für die jeweiligen Berechnungen angegeben. Korrelationskoeffizienten von $> 0,60$ gelten als hoch, $0,40$ bis $0,60$ als mittelmäßig und unter $0,40$ als niedrig (Fisseni 2004).

Vergleicht man die dimensionale Auswertung des PHQ-D mit den Subskalen des BSI-18 zeigen sich teils deutliche Korrelationen. Die Korrelationskoeffizienten, berechnet nach Spearman, liegen zwischen $r = 0,796$ (Depression) und $r = 0,722$ (Somatisierung). Die Ängstlichkeitsskalen korrelieren mit $r = 0,799$. Somit sind die statistischen Zusammenhänge zwischen den korrespondierenden Skalen als Hinweis auf die kriteriumsbezogene Validität des Fragebogens ausnahmslos als hoch zu bewerten. Die berechneten Zusammenhänge zwischen den nicht korrespondierenden Konstrukten fielen inhaltlich nachvollziehbar geringer aus, es bestehen jedoch auch hier messbare Zusammenhänge als Hinweis für Belastungen in den anderen Dimensionen.

Es fanden sich weiterhin deutliche Korrelationen der drei Skalen des PHQ-D mit dem GSI des BSI-18. Besonders die Skalen PHQ-9 (*Depressivität* $\rho = 0,798$) und GAD-7 (*Ängstlichkeit* $\rho = 0,78$) zeigen hohe Zusammenhänge mit dem GSI, während die Skala PHQ-15 (*Somatisierung* $\rho = 0,571$) einen mittelmäßigen Zusammenhang zeigt. Der PHQ-D Stressscore korrelierte am deutlichsten mit dem GSI ($\rho = 0,615$). Die geringste Korrelation ($\rho = 0,340$) zeigt die Skala *Somatisierung* des BSI-18 mit der Skala PHQ-9 (*Depressivität*) des PHQ-D (vgl. **Tabelle 36**).

Tabelle 36: Korrelation der BSI-18-Skalen mit den dimensional Skalen des PHQ-D

	GSI	SOMA	DEPR	ANGS
Depressivität PHQ-9 (n = 1031)	,798**	,491**	,796**	,651**
Somatisierung PHQ-15 (n = 932)	,571**	,722**	,341**	,440**
Ängstlichkeit GAD-7 (n = 1034)	,780**	,452**	,670**	,799**
PHQ Stress Score (n = 936)	,615**	,406**	,611**	,500**

**Alle Korrelation nach Spearmans-Rho sind auf dem 0,01 Niveau signifikant (zweiseitig).

Den gültigen BSI-18 Fragebögen (N = 1086) standen 716 korrekt und vollständig ausgefüllte SF-12-Fragebögen gegenüber. Die hier vorliegende Stichprobe aus Patienten der psychosomatischen Ambulanz hatte erwartungsgemäß deutliche Einschränkungen der gesundheitsbezogenen Lebensqualität, wobei die *körperliche Summenskala* in etwa eine, die

psychische Summenskala in etwa zwei Standardabweichungen unter den Werten der Normstichprobe lagen, augenscheinlich kongruent mit der Zusammensetzung des hier untersuchten Patientenkollektivs, in dem Patienten mit somatoformen Beschwerden einen geringeren Anteil haben (vgl. **Tabelle 37**).

Als Maß für die kriteriumsbezogene Validität wurden die Korrelationen der *körperlichen* und der *psychischen Summenskala* mit den Skalen des BSI-18 berechnet. Es fanden sich mittelmäßig bis hohe Korrelationen zwischen der psychischen Summenskala und dem GSI ($\rho = 0,59$) sowie den Skalen ANGS ($\rho = 0,52$) und DEPR ($\rho = 0,64$). Die beiden somatischen Skalen zeigten einen mittelmäßigen Zusammenhang ($\rho = 0,43$). Die Ergebnisse sind inhaltlich stimmig.

Tabelle 37: Korrelationen der Skalen des SF-12 mit dem BSI-18 (N = 716)

	GSI	SOMA	DEPR	ANGS
SF-12 psychische Summenskala	-,609**	-,266**	-,654**	-,542**
SF-12 körperliche Summenskala	-,221**	-,448**	-,102**	-,076*

** . Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* . Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

4.5 Faktorielle Validität

Zur Bestimmung der faktoriellen Validität wurde eine exploratorische Faktorenanalyse mittels varimaxrotierter Hauptkomponentenanalyse mit Kaiser-Normalisierung durchgeführt. Es wurden ausschließlich Faktoren mit einem Eigenwert größer 1 (Kaiser-Guttman-Kriterium) extrahiert und Faktorladungen mit einem Absolutwert unter 0,40 unterdrückt.

Unter diesen Bedingungen konnten für diese Stichprobe 3 Faktoren extrahiert werden, die zusammen eine Varianzaufklärung von 54,1% erzielen, wobei der erste 21,1%, der zweite 17,9% und der dritte Faktor 15,2% der Gesamtvarianz aufklärt (vgl. **Tabelle 38**). Ein vierter Faktor verpasste das Eigenwertkriterium knapp (0,974).

Tabelle 38: BSI-18 exploratorische Faktorenanalyse - Erklärte Gesamtvarianz (N = 1086)

Faktor	Anfängliche Eigenwerte		Rotierte Summe der quadrierten Ladungen	
	Gesamt	% der Varianz	Gesamt	% der Varianz
1	6,684	37,134	3,792	21,069
2	1,931	10,729	3,226	17,924
3	1,138	6,322	2,735	15,193
4	,974	5,411		

Auch die Interpretation des Scree-Tests lässt eine dreifaktorielle Struktur erwarten, da der Knick, an dem die Eigenwerte sich asymptotisch der Abszisse annähern, auf Höhe des dritten Faktors zu liegen kommt. Der Scree-Plot ist in **Abbildung 1** wiedergegeben.

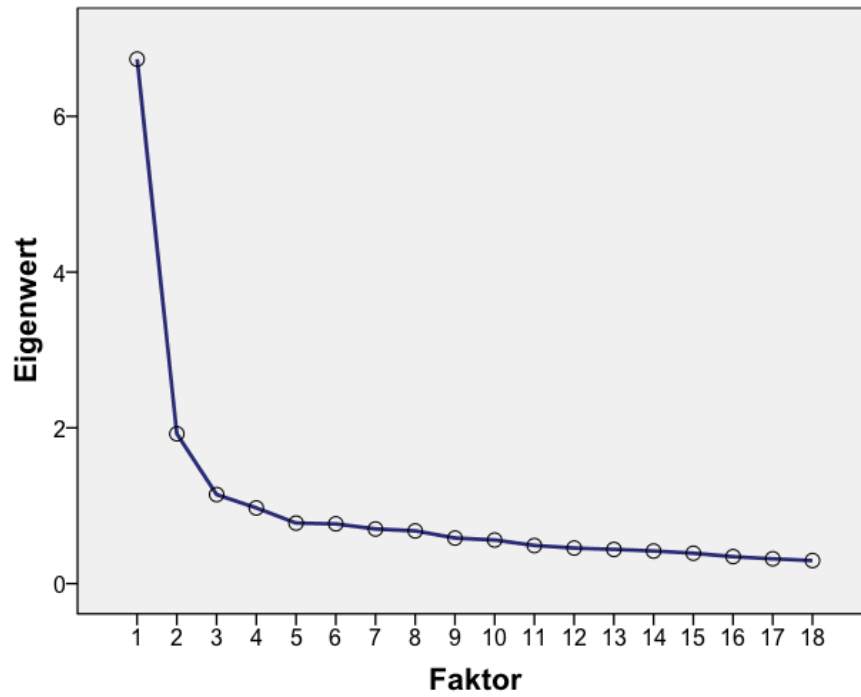


Abbildung 1: Scree-Test der Faktorenanalyse des BSI-18

Die Faktorstruktur entspricht weitestgehend der dimensionalen Struktur des BSI-18 (vgl. **Tabelle 39**). Der erste Faktor enthält hohe Ladungen von 0,622 (Item 17 „*Gedanken, sich das Leben zu nehmen*“) bis 0,830 (Item 11 „*dem Gefühl, wertlos zu sein*“) für alle Items der Skala DEPR. Darüber hinaus lädt das Item 18 („*Furchtsamkeit*“) ebenfalls auf den „*Depressionsfaktor*“, jedoch numerisch geringer (0,416). Der zweite Faktor enthält alle Items der Skala ANGS mit Ladungen zwischen 0,496 (Item 9 „*plötzlichem Erschrecken ohne Grund*“) und 0,753 (Item 3 „*Nervosität und innerem Zittern*“). Der dritte Faktor schließlich gibt die Skala SOMA wieder. Die Ladungen der entsprechenden 6 Items liegen zwischen 0,401 (Item 7 „*Übelkeit oder Magenverstimmung*“) und 0,743 (Item 13 „*Taubheit oder Kribbeln in einzelnen Körperteilen*“). Zusammenfassend lässt sich festhalten, dass die hypothetische dimensionale Struktur des BSI-18 durch die Faktorenanalyse gestützt wird.

Tabelle 39: BSI-18 exploratorische Faktorenanalyse – Komponentenmatrix mit Ladungen größer 0,40 (N = 1086)

BSI-18 Item	Faktor 1	Faktor 2	Faktor 3
1. Ohnmachts- oder Schwindelgefühlen	-	-	,488
2. dem Gefühl, sich für nichts zu interessieren	,633	-	-
3. Nervosität oder innerem Zittern	-	,753	-
4. Herz- und Brustschmerzen	-	-	,657
5. Einsamkeitsgefühlen	,741	-	-
6. dem Gefühl, gespannt oder aufgeregt zu sein	-	,738	-
7. Übelkeit oder Magenverstimmung	-	-	,401
8. Schwermut	,732	-	-
9. plötzlichem Erschrecken ohne Grund	-	,496	-
10. Schwierigkeiten beim Atmen	-	-	,664
11. dem Gefühl, wertlos zu sein	,830	-	-
12. Schreck- oder Panikanfällen	-	,713	-
13. Taubheit oder Kribbeln in einzelnen Körperteilen	-	-	,743
14. einem Gefühl der Hoffnungslosigkeit angesichts Ihrer Zukunft	,796	-	-
15. so starker Ruhelosigkeit, dass Sie nicht mehr still sitzen können	-	,588	-
16. Schwächegefühl in einzelnen Körperteilen	-	-	,693
17. Gedanken sich das Leben zu nehmen	,622	-	-
18. Furchtsamkeit	,416	,646	-

4.6 Diskriminante Validität, Sensitivität und Spezifität des BSI-18

Zur Beurteilung der diskriminanten Validität des BSI-18 und seiner Subskalen wurden die Sensitivität und Spezifität für die vorgeschlagenen Cut-Off-Werte (Franke et al. 2012) der deutschen Normierungsstichprobe berechnet. Weiterhin wurden ROC-Analysen für die vier Skalen mit den im Erstgespräch gestellten Diagnosen als dichotome Variablen berechnet und die ermittelten Cut-Off-Werte mit den vorgeschlagenen Werten aus der Literatur verglichen.

Unter Verwendung der geschlechtsspezifischen T-Werte der Skalen des BSI-18 mit dem vorgeschlagenen Cut-Off (≥ 60) fand sich für die *depressiven Störungen* eine Sensitivität von 86,3% und eine Spezifität von 44,9%. Der BSI-18 war dem PHQ-D minimal überlegen (85,6% und 44,9%). Der GSI (90,7%) zeigte eine noch etwas höhere Sensitivität als die Skala DEPR. Auch die zwei übrigen Skalen konnten Erkrankte trotz fehlender Items bezüglich einer depressiven Symptomatik mit geringerer Sensitivität erkennen (SOMA = 69%; ANGS = 82,5%). Hier scheinen in diesem Patientenkollektiv vor allem ängstliche Symptome komorbide aufzutreten. Die Spezifität war für alle Skalen gering (33,3% bis 44,9%).

Die Auswertung von *Angst- und Panikstörungen* zeigte ein ähnliches Bild. Das BSI-18 erreichte eine deutlich höhere Sensitivität (92,1%) als der PHQ-D (66,2%); erneut war der GSI sensitiver bei geringerer Spezifität als die syndromal passende Skala. Die Skalen waren allesamt wenig spezifisch. Die Skala ANGS des BSI-18 war nur zu 29% spezifisch, während der PHQ-D etwas spezifischer war (36,7%).

Patienten mit *somatoformen Störungen* konnten von der Skala SOMA nur zu 73% erkannt werden. Die Spezifität lag bei 38,6%. Besser war der GSI (81,3%) zu Lasten der Spezifität (18%). Der PHQ-D konnte nur knapp die Hälfte richtig identifizieren und war am spezifischsten (57,6%). Die Skala ANGS erkannte wider Erwarten mehr Erkrankte als die inhaltlich passende Skala SOMA.

Zusammenfassend war das BSI-18 für alle Diagnosen in dieser Stichprobe für die vorgeschlagenen Cut-Off-Werte sensitiver als der PHQ-D, während es sich für die Werte der Spezifität erwartungsgemäß reziprok verhielt. Der GSI scheint den kategorialen Skalen bezüglich der Sensitivität ebenbürtig oder überlegen zu sein, ist jedoch sehr wenig spezifisch. Dies wird auch bei der Betrachtung von inhaltlich nicht kongruenten „Skalen-Diagnosen-Paaren“ deutlich. Auch diese sind häufig zu einem gewissen Grad sensitiv. Es

liegen häufige komorbide Symptome aus anderen Symptomkomplexen vor, dies bildet auch die Tatsache ab, dass 64,5% der Probanden mehr als eine Diagnose nach ICD-10 im Erstgespräch erhielten.

Tabelle 40: Klassifikationsbewertung des BSI-18

Diagnose	Skala	Sensitivität	Spezifität	Pos. prädiktiver Wert	Neg. prädiktiver Wert
Depressive Erkrankungen (F32-34)	ANGS	82,5	36	68,4	54,9
	DEPR	86,3	44,9	72,6	66,1
	SOMA	69	42,2	66,8	44,8
	GSI	90,7	33,3	69,5	68,1
	PHQ-9	85,6	44,9	71,4	66,2
Angsterkrankungen (F40-F41)	ANGS	92,1	29	27,7	92,9
	DEPR	80	26,9	23,8	82,5
	SOMA	75,9	38,3	26	84,8
	GSI	92,1	21,1	25	90,4
	GAD-7	66,2	36,7	23,2	78,9
Somatoforme Störung (F45)	ANGS	76,3	24,8	27,8	73,3
	DEPR	68,5	22,9	25,2	65,8
	SOMA	73,9	38,6	31,4	79,6
	GSI	81,3	18	27,4	71,7
	PHQ-15	54,9	57,6	31,7	78,1

In der ROC-Analyse des BSI-18 wurden die Summenskalen des PHQ-D mit den entsprechenden Subskalen des BSI-18 verglichen. Zusätzlich wurden drei Fragen aus dem „allgemeinen soziodemographisch-sozialanamnestischen“ Fragebogen hinzugezogen, z.B.: „Wie sehr litten Sie in den letzten 2 Wochen an Depressionen?“ (Likert-Skala von 1-10). Die Stichprobengröße verringerte sich aufgrund fehlender Antworten jeweils um ca. 10%, daher entstehen leichte Differenzen der Sensitivität und Spezifität im Vergleich zu der oben aufgeführten Berechnung.

Für die depressiven Erkrankungen (F32-F34) zeigt sich für die vier unterschiedlichen Skalen ein ähnlicher Verlauf. Die Flächen unter der Kurve unterschieden sich kaum, den höchsten Wert erreichte die Skala DEPR (0,746) gefolgt vom der PHQ-9 (0,739) und der direkten Frage nach den Depressionen (0,724), zuletzt der GSI (0,71). Die Flächen unter der Kurve der entsprechenden Skalen für die Diagnose F45 nach ICD-10 fielen geringer aus als bei den depressiven Erkrankungen. Die Skalen konnten Erkrankte und Gesunde kaum trennen. Auffällig war, dass die Frage „Wie sehr litten Sie in den letzten 2 Wochen unter Schmerzen“ (1-10 auf einer Likert-Skala) einen höheren Wert (0,66) erreichte als die vergleichbaren syndromalen Skalen des BSI-18 und PHQ-D (0,583 und 0,593). Mittels des GSI scheint eine Trennung nicht möglich zu sein, sein AUC-Wert lag bei 0,485 und entspricht somit der Zufallsverteilung.

Die Analyse der Grenzwertoptimierungskurve mit den Diagnosen F40 und F41 als Zustandsvariable zeigte, dass die Skala ANGS (0,697), gemessen an der Fläche unter der Kurve, dem Ängstlichkeitsindex des PHQ-D (0,618) überlegen scheint. Der GSI erreichte einen vergleichbaren Wert von 0,641. Erneut schien eine diagnostische Aussage mit dem Item „*Wie sehr litten Sie in den letzten 2 Wochen unter Ängsten?*“ am besten möglich zu sein.

Zusammenfassend zeigt sich deutlich, dass die diskriminativen Fähigkeiten des Fragebogens auf syndromaler Ebene gering sind, hohe Sensitivitäten gehen mit einer geringen Spezifität einher, wobei Depressionen und Angsterkrankungen deutlich besser abgebildet werden als die Somatisierungsstörungen. Der GSI ist in allen Dimension am sensitivsten, lässt jedoch keine Aussage über die Art der Störung zu.

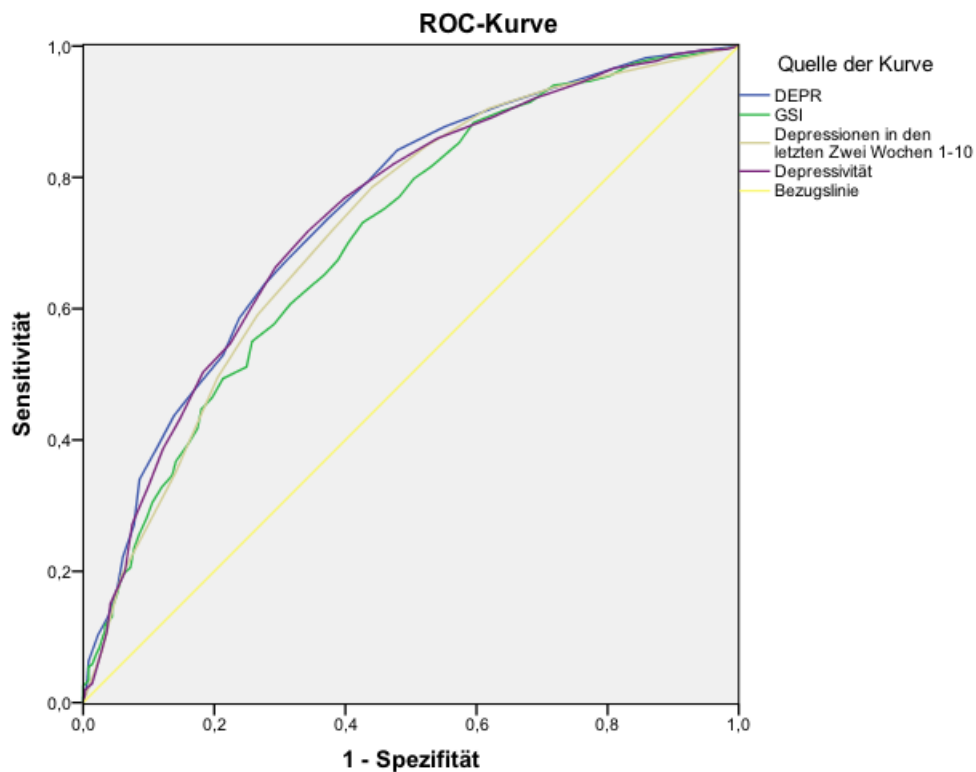


Abbildung 2: ROC Kurve der Diagnosen F32-F34 (N = 1020) der Skalen GSI, DEPR, PHQ-9 („Depressivität“), Selbsteinschätzung der Depressivität (auf einer Skala von 1-10)

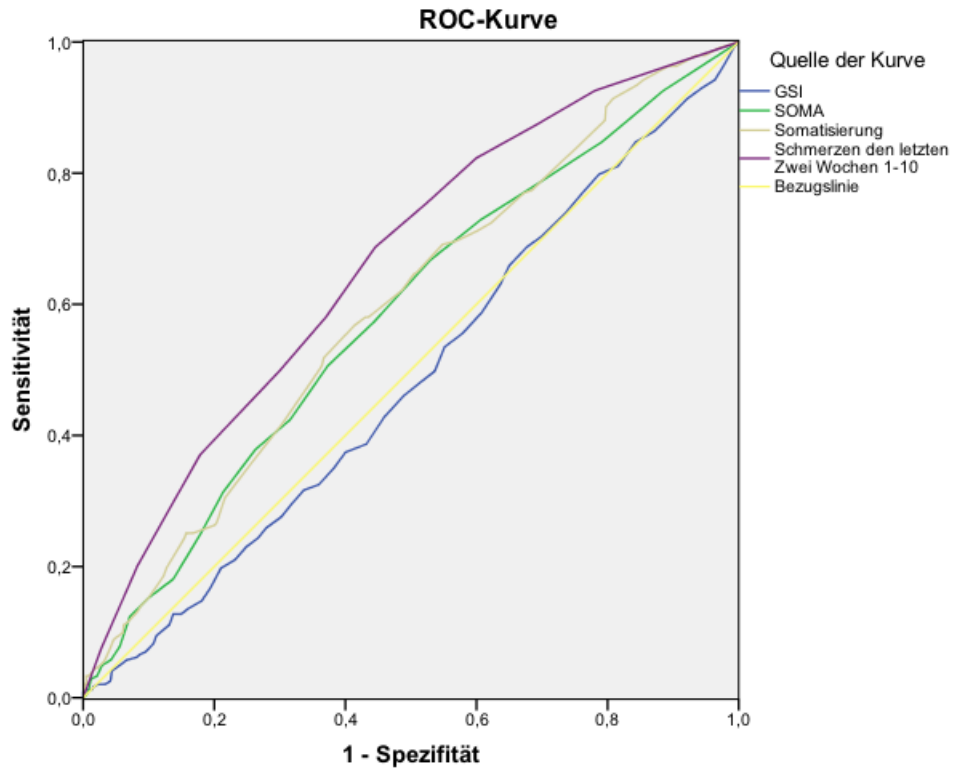


Abbildung 3: ROC Kurve der Diagnosen F45 (N = 905) für die Skalen GSI, SOMA, PHQ-15 („Somatisierung“), Selbsteinschätzung von Schmerzen (auf einer Skala von 1-10)

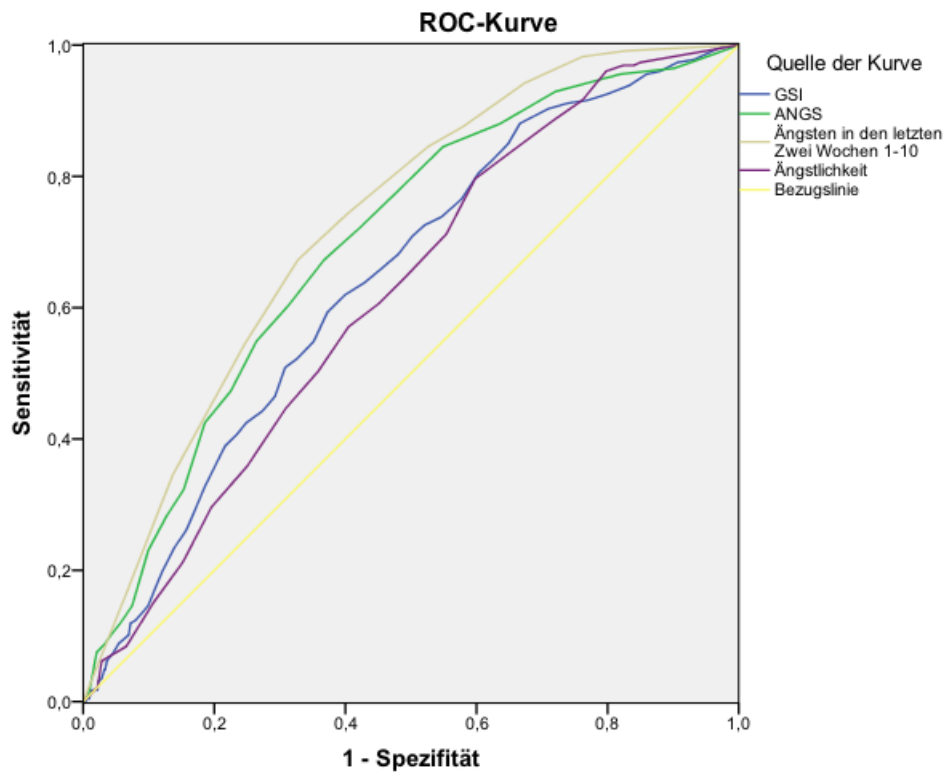


Abbildung 4: ROC Kurve der Diagnosen F40/F41 (N = 1002) für die Skalen GSI, ANGS, GAD-7 („Ängstlichkeit“), Selbsteinschätzung der Ängstlichkeit (auf einer Skala von 1-10)

5 Diskussion

Die Resultate der vorliegenden Arbeit, einschließlich der Implikationen für den klinischen Alltag, sollen im Folgenden dargestellt, in den wissenschaftlichen Kontext eingeordnet, und diskutiert werden. Ziel der Dissertation war die Prüfung der psychometrischen Eigenschaften des BSI-18 für ein deutsches Patientenkollektiv mit vornehmlich psychosomatischen Erkrankungen.

Die Kürze des BSI-18 soll eine effizientere und anwendungsökonomischere Diagnostik ohne allzu große Einschränkungen in der diagnostischen Güte gewährleisten. Die Nützlichkeit und Güte der Subskalen bezüglich der formalen Diagnostik von psychischen Erkrankungen wird von vielen Autoren kritisch gesehen. Das Instrument wird vielfach als Maß für eine allgemeine psychische Belastung verstanden (Asner-Self et al. 2006).

Im Rahmen der Forschung an den Vorgängerinstrumenten SCL-90-R und BSI konnte gezeigt werden, dass die Skalen der Vorgängerfragebögen in unterschiedlichen Patientenkollektiven unterschiedliche Ergebnisse und eine unterschiedliche diagnostische Güte zeigen (Johnson et al. 1996). Eine genauere Untersuchung der Psychometrie in einem deutschen psychosomatischen Patientenkollektiv ist, unter der Annahme einer zunehmenden Verbreitung, von großer praktischer Bedeutung.

5.1 Inhaltliche Validität

Betrachtet man inhaltlich die im BSI-18 enthaltenen Items und vergleicht diese mit empirisch erhobenen Beschwerden bei Depressionen, Angsterkrankungen und Somatisierungsstörungen, findet sich vor dem Hintergrund des geringen Umfangs des Instruments, insbesondere in Bezug auf die vielfältigen somatoformen und funktionellen Beschwerden, nur eine geringe Anzahl der möglichen Beschwerden im Fragebogen wieder. Die Verringerung der abgefragten Symptome wurde, im Hinblick auf die gewünschte Kürzung des Instruments, bewusst in Kauf genommen. Die Items wurden aufgrund der Itemanalyse, der Faktorenanalyse sowie aus inhaltlichen Erwägungen, insbesondere in Bezug auf die Prävalenz der Symptome ausgewählt.

Daher fehlen häufige *somatoforme* Symptome wie z.B. *muskuloskelettale Beschwerden*. Auch *Kopfschmerzen* und *Palpitationen* (Franz und Schepank 1996) werden nicht erfasst. Bei den möglichen Symptomen einer depressiven Störung (Dilling et al. 2011) werden z.B. die *Konzentrations- und Aufmerksamkeitsstörung*, *Schuldgefühle*, *Schlafstörungen* und der *verminderte Appetit* nicht erfragt. Auch die Kennzeichen des somatischen Syndroms wer-

den nur zum Teil abgebildet. Es fehlen unter anderem: die *Unfähigkeit, auf freudige Situationen emotional zu reagieren*, das *Morgentief*, die *psychomotorische Hemmung* und ein möglicher *Gewichts- und Appetitverlust*.

Die somatischen Aspekte der Angst- und Panikerkrankungen werden teilweise in der Skala SOMA abgebildet. Insgesamt konzentrieren sich die Items der Skala ANGS auf Symptome der Angst und die physiologische Reaktion. Nach Ätiologie oder Bezug der Ängste wird nicht gefragt. Die Items der SCL-90 Skala *phobische Angst* wurden nicht berücksichtigt. Hier könnten Patienten mit vornehmlich phobischen Ängsten, da sie in Abwesenheit des Stimulus nicht zwangsläufig häufiger Ängste als Nicht-Erkrankte erleben, übersehen werden. Aus rein inhaltlichen Überlegungen ist der Fragebogen eher dazu in der Lage, die *anderen Angststörungen (F41)* zu erkennen.

Die deutliche Reduktion der Itemanzahl geht mit einer verminderten Breite an erhobenen Symptomen einher. Die Kürze des Fragebogens impliziert bereits Schwächen in Bezug auf die Diagnostik auf syndromaler Ebene und legt eine Verwendung zum Screening allgemeiner Psychopathologie nahe. Insbesondere im Hinblick auf die vielfältigen Symptome einer *somatoformen Störung (F45)*, kann eine verminderte Sensitivität und Spezifität auch mit der Reduktion der abgefragten Symptomkomplexe zusammenhängen. Diese Einschränkung liegt in der Kürzung begründet und müsste somit auch auf andere Instrumente wie z.B. die PHQ-2 (Depressivität) oder GAD-2 (Ängstlichkeit) zutreffen, die noch kürzer sind. Dennoch sind trotz deutlicher Kürzung für beide Instrumente hinreichende Sensitivitäten und Spezifitäten berichtet (Kroenke et al. 2007, Lowe et al. 2005). Wahrscheinlich liegt dies in der Vielfältigkeit somatoformer Symptome begründet, die sich nicht ohne weiteres auf „Kernsymptome“ herunterbrechen lassen.

5.2 Teststatistische Kennwerte

Die statistische Itemanalyse des Fragebogens erbrachte weitestgehend positive Ergebnisse, die den gängigen Gütekriterien der Fragebogenkonstruktion genügen.

Die Mittelwerte der 18 Items des Fragebogens liegen in der vorliegenden Stichprobe zwischen $M = 0,51$ (Item 17 „*Gedanken sich das Leben zu nehmen*“) und $M = 1,96$ (Item 14 „*Hoffnungslosigkeit angesichts der Zukunft*“). Vergleicht man die ermittelten Mittelwerte dieser Studie mit anderen Publikationen, insbesondere mit Normstichproben, zeigt sich erwartungsgemäß eine erhöhte psychische Belastung. Ein Vergleich mit den T-Werten aus einer deutschen bevölkerungsrepräsentativen Stichprobe (Franke et al. 2012) ergibt für

die vorliegende Stichprobe im Mittel T-Werte von 67 für den GSI, 64 für die Skala SOMA und jeweils 66 für die Skalen ANGS und DEPR.

Alle Items zeigten eine rechtsschiefe bzw. linkssteile Verteilung. Betrachtet man die Schwierigkeit (d_i) der Items, ergeben sich deutlich unterschiedliche Schwierigkeiten, wobei jede der drei Subskalen Items mit unterschiedlichen Schwierigkeiten enthält und somit theoretisch eine gute Differenzierung zwischen Individuen mit hoher und niedriger Merkmalsausprägung gegeben ist (Fisseni 2004). Die mittlere Itemschwierigkeit über alle Items beträgt 1,25. Die Gesamtschwierigkeit ist folglich als relativ niedrig zu beurteilen, insbesondere vor dem Hintergrund der aus Patienten bestehenden Stichprobe.

Die Schwierigkeit der einzelnen Skalen unterscheidet sich mäßig, wobei die Skala Depression ($d_i = 1,13$) schwieriger erscheint als die Skalen Ängstlichkeit ($d_i = 0,90$) und Somatisierung ($d_i = 0,78$). Mit einer Schwierigkeit von $p = 0,50$ (entspricht einem Wert von 2 auf einer 4-stufigen Likert-Skala) und einer Trennschärfe von $r_{it} = 1$ ergeben sich theoretisch die meisten Möglichkeiten, Probanden zu unterscheiden (Lienert et al. 1994).

Betrachtet man die Mittelwerte der Skalen, findet sich für den GSI ein Wert von $M = 22,51$ ($SD = 13,98$). Entsprechend der mittleren Itemsschwierigkeiten finden sich im Mittel höhere Werte auf der Skala DEPR ($M = 9,04 / SD = 6,36$) als auf den Skalen ANGS ($M = 7,22 / SD = 5,69$) und SOMA ($M = 6,25 / SD = 4,88$). Da es sich bei der Stichprobe um ein Patientenkollektiv handelt, sind Verzerrungen durch die unterschiedliche Häufigkeit der Diagnosen möglich, wobei depressive Störungen ($F33 = 18,2\% / F32 = 40,7\%$) zahlenmäßig führten, gefolgt von Somatisierungsstörungen ($F45 = 27,5\%$) und Angsterkrankungen ($F40 = 7,9\% / F41 = 15,6\%$). Dies könnte die tendenziell höheren Werte auf der Skala DEPR erklären.

Die Trennschärfe (r_i) nach der part-whole Methode liegt für den GSI zwischen $r_i = 0,39$ und $0,66$. Trennschärfekoeffizienten größer $r_{it} = 0,50$ gelten als hoch, zwischen $0,50$ und $0,30$ als mäßig und unter $0,30$ als niedrig (Fisseni 2004). Bis auf 5 Items sind die Trennschärfen daher als hoch zu bewerten. Die Items der Skala SOMA weisen bezüglich des GSI tendenziell niedrigere Trennschärfen auf. Bei der Betrachtung der Trennschärfen in den Subskalen findet sich in der Skala SOMA eine geringere Trennschärfe der Items ($r_i = 0,44$ bis $0,57$), als bei den Skalen DEPR ($r_i = 0,50$ bis $0,75$) und ANGS ($r_i = 0,53$ bis $0,67$). Die Trennschärfen der Items der beiden zuletzt genannten Skalen sind ausnahmslos als hoch zu werten.

Eine Interitemkorrelation zwischen 0,20 und 0,40 gilt als erstrebenswert (Bühner 2011). Die Skala GSI ist als globaler Index weniger homogen als die drei Subskalen. Die durchschnittliche Itemkorrelation lag für die Skalen GSI und SOMA bei 0,326 und 0,324, während die Items der Skalen ANGS und DEPR (0,469 und 0,512) homogener beantwortet wurden.

Theoretisch ist die Höhe der Trennschärfe abhängig von dem Schwierigkeitsindex und der Interkorrelation der Items. Diese statistische Regel der klassischen Testtheorie lässt sich empirisch auch an der vorliegenden Stichprobe beobachten.

So zeigt z.B. die Skala SOMA mit der niedrigsten Itemkorrelation und der niedrigsten Schwierigkeit auch die geringste Trennschärfe. Diese Beobachtung findet sich auch in der tendenziell geringeren Reliabilität und der geringeren diagnostischen Güte (im Sinne geringerer Werte für die Sensitivität und Spezifität) wieder. Für die restlichen Skalen finden sich in oben genannter Reihenfolge ähnliche Zusammenhänge im Sinne günstiger Kennwerte der Itemanalyse. Gemäß diesen Überlegungen lässt sich spekulieren, ob eine andere Auswahl an „somatoformen“ Items oder eine Umformulierung mit dem Ziel einer höheren Schwierigkeit sinnvoll wäre.

Eine Analyse der soziodemographischen Merkmale „Alter“ und „Geschlecht“ ergab bei drei der vier Skalen geschlechts- und bei zwei Skalen altersspezifische Unterschiede. In allen vier Dimensionen erreichten die weiblichen Versuchspersonen im Mittel höhere Testwerte und zeigten somit eine höhere Belastung. Die Gruppenunterschiede sind unter der Annahme einer gerichteten Nullhypothese signifikant. Höhere Werte in der weiblichen Bevölkerung wurden sowohl in den deutschen als auch in den englischen Normierungstichproben beschrieben und finden Eingang in die Cut-Off-Werte (Derogatis 2000, Franke et al. 2012, Zabora et al. 2001).

Für die Skalen SOMA ($p = 0,002$) und ANGS ($p = 0,045$) fanden sich altersspezifische signifikante Unterschiede. Es fanden sich signifikante Mittelwertdifferenzen in Bezug auf die Skala SOMA zwischen der jungen und der älteren Teilstichprobe ($p = 0,002$). In Punkto Ängstlichkeit zeigten sich die 30 bis 50-Jährigen signifikant belasteter als die jungen Patienten ($p = 0,046$). Kongruent zu den hier erhobenen Daten fanden sich Altersunterschiede in der deutschen Normstichprobe (Franke et al. 2012).

Darüber hinaus wurden von Franke et. al. für die Skalen SOMA, DEPR und GSI besonders hohe Werte bei sehr alten Personen (75 - 94 Jahre) beschrieben. Diese Ergebnisse konnten (möglicherweise auf Grund der geringen Teilstichprobe dieser Altersklasse) in

dieser Studie nicht reproduziert werden. Die US-amerikanischen Normstichproben unterscheiden nur nach Geschlecht und schließen alte Testpersonen (> 70 Jahre) aus. Die von Franke beschriebenen Belastungen bei allen vier Skalen im hohen Alter finden in internationalen Normstichproben keine Erwähnung. Wünschenswert wären internationale Studien zur altersspezifischen Bestimmung der T-Werte für das BSI-18 und ggf. eine Berücksichtigung in der Auswertung des Instruments.

Es zeigten sich (bis auf die Skala DEPR) in den hier erhobenen Daten stärkere Belastungen von Personen, ohne deutsche Staatsbürgerschaft. Die Größe der Teilstichproben der übrigen Nationalitäten waren ausnahmslos im unteren einstelligen Bereich, so dass eine Auswertung nur zwischen Patienten mit deutscher Staatsbürgerschaft und der Gesamtheit der übrigen Nationalitäten verglichen werden konnte. Diese Beobachtung lässt keine Rückschlüsse auf eine Beeinflussung der Testwerte durch eine real erhöhte Belastung der Migranten oder durch interkulturelle Unterschiede im Antwortverhalten zu. Untersuchungen zur Psychometrie von Selbstbeurteilungsverfahren bei Personen mit Migrationshintergrund wären wünschenswert.

Für den Bereich der Psychoonkologie gibt es für den deutschen Sprachraum noch keine Norm- und Trennwerte. Eine Erhebung und ein Vergleich mit den von Zabora (2001) publizierten Daten ist sinnvoll, da frühe Studien interkulturelle Unterschiede der SCL-90-R zeigen konnten (Franke 1994).

Hinweise für eine Beeinflussung der Testergebnisse durch unterschiedliche Bildungsstände konnten nicht gefunden werden. Asner-Self et al. (2006) konnten ebenfalls keine statistisch signifikanten Haupt- oder Interaktionseffekte mit den drei Subskalen als abhängige, sowie Sprache und Geschlecht als unabhängige Variablen finden. Auch Franke et al. (2010) fanden bei Nierentransplantierten keine statistisch signifikanten Zusammenhänge zu soziodemographischen Daten.

5.3 Reliabilität

Die Reliabilität wurde in dieser Studie mit Hilfe von Cronbachs Alpha geschätzt. Dies erscheint vor dem Hintergrund der großen Verbreitung dieses Standardverfahren sinnvoll und hinreichend.

In der untersuchten Stichprobe fanden sich interne Konsistenzen zwischen $\alpha = 0,74$ und $\alpha = 0,90$. Nach Fisseni (2004) ist die Reliabilität des GSI als mittel bis hoch einzuschätzen

($\alpha = 0,898$). In den vorhandenen Studien zum BSI-18 (vgl. Kapitel 2.3.3.1) fanden sich ähnliche Ergebnisse (GSI $\alpha = 0,82$ bis $0,94$).

Die Werte der Skalen DEPR ($\alpha = 0,89$) und ANGS ($\alpha = 0,84$) zeigten eine mittelmäßige Messgenauigkeit. In der Literatur finden sich sehr unterschiedliche Ergebnisse ($\alpha = 0,73$ bis $0,91$ und $\alpha = 0,71$ bis $0,88$). Die Skala SOMA zeigte den geringsten Wert für Cronbachs Alpha ($\alpha = 0,74$). Die Reliabilität ist als niedrig zu bewerten. Auch hier liegt das Ergebnis im Bereich der publizierten Ergebnisse ($\alpha = 0,58$ bis $0,83$). Arbeiten mit kleinen Stichproben führten eher zu geringeren Werten für Cronbachs α (vgl. **Tabelle 7**: Interne Konsistenz des BSI-18 in unterschiedlichen Stichproben, Seite 22). Die hier errechneten Werte bestätigen die guten Ergebnisse früherer Studien in einem deutschsprachigen, psychosomatischen Patientenkollektiv. Zum wiederholten Male zeigte sich, dass der GSI besonders reliabel ist und die Skala SOMA gegenüber den anderen Skalen deutlich zurücksteht. Eine Beeinflussung der Reliabilität durch Bildung und Geschlecht konnte nicht gefunden werden.

Das Vorgängerinstrument SCL-90 zeigte in der Eichstichprobe aus der Normalbevölkerung mit 2141 Probanden (Hessel et al. 2001) höhere Werte für Cronbachs Alpha des GSI und die Skala SOMA ($\alpha = 0,97$ und $\alpha = 0,83$) und vergleichbar hohe Werte für die Skalen DEPR und ANGS ($\alpha = 0,87$ und $\alpha = 0,84$). Die im Manual der SCL-90 aufgeführte Stichprobe aus 5057 Psychotherapieklienten ergab nur minimale Abweichungen zur Eichstichprobe (Franke und Derogatis 2000) in Bezug auf die Reliabilität der SCL-90. Im Vergleich mit dem BSI-18 erscheint vornehmlich die geringere Reliabilität der Skala SOMA von Bedeutung, da die Reliabilität des GSI des BSI-18 als hoch zu bewerten ist. Eine Verminderung der Reliabilität durch die verminderte Anzahl an Items der Skala SOMA ist daher zu vermuten.

Prinz et al. verglichen die SCL-90 und ihre Kurzversionen in Bezug auf ihre psychometrischen Eigenschaften bei Patienten mit affektiven Störungen (2013). Die Autoren konnten ebenfalls eine Abnahme der Reliabilität in Zusammenhang mit der Kürzung der Instrumente beobachten. Die Skala SOMA zeigte die deutlichste Abnahme (SCL-90 $\alpha = 0,86$, BSI $\alpha = 0,80$, BSI-18 $\alpha = 0,79$), die Reliabilität der übrigen Skalen nahm ebenfalls mit geringer werdender Itemanzahl ab. Die Unterschiede waren jedoch gering und statistisch nicht signifikant.

Die Werte für Cronbachs α für den GSI erhöhen sich nicht wesentlich durch das Weglassen einzelner Items. Die Reliabilität der Skala DEPR würde davon profitieren, das Item

17 „Gedanken sich das Leben zu nehmen“ aus der Skala zu entfernen. Auf Grund klinischer und inhaltlicher Überlegungen erscheint dies nicht sinnvoll. Ob ein Zusammenhang bezüglich der geringeren Reliabilität der Skalen ANGS und SOMA und den tendenziell leichteren Items dieser Skalen besteht, lässt sich aus den vorliegenden Daten nicht ableiten. In den Normstichproben für die Allgemeinbevölkerung findet sich diese Konstellation nicht wieder (Derogatis 2000, Franke et al. 2012, Zabora et al. 2001).

Re-Test Erhebungen wurden in dieser Studie nicht erhoben. Die Test-Retest Reliabilitäten für den Vorgänger BSI über einen Zeitraum von 6 Wochen liegen zwischen 0,68 und 0,90 (SOMA $r_{tt} = 0,68$; DEPR $r_{tt} = 0,84$; ANGS $r_{tt} = 0,79$; GSI $r_{tt} = 0,90$). Für den BSI-18 wird von ähnlichen Werten ausgegangen. Zwei Autoren (Andreu et al. 2008, Spitzer et al. 2011) überprüften die Retest-Reliabilität für den BSI-18 in einer spanischsprachigen (SOMA $r_{tt} = 0,76$; DEPR $r_{tt} = 0,82$; ANGS $r_{tt} = 0,68$; GSI $r_{tt} = 0,76$) und einer deutschsprachigen (SOMA $r_{tt} = 0,52$; DEPR $r_{tt} = 0,80$; ANGS $r_{tt} = 0,65$; GSI $r_{tt} = 0,77$) Studie. In der Literatur finden sich unterschiedliche Empfehlungen zur Bewertung der Re-Test-Reliabilität. Einige Autoren erachten Werte von $r_{tt} > 0,50$ als ausreichend (Streiner und Norman 2003), während andere erst höhere Werte ($r_{tt} > 0,80$) als hinreichend reliabel erachten (Bortz et al. 2006). Einschränkend ist zu erwähnen, dass das BSI-18 als „State-Skala“ psychische Belastungen explizit in den vergangenen 7 Tagen erfasst. Daher ist eine hohe Änderungssensitivität wahrscheinlich.

5.4 Kriteriumsbezogene Validität

Die kriteriumsbezogene Validität wurde auf Grund von verschiedenen externen Kriterien, wie den Fragebögen SF-12, PHQ-D, psychiatrischen Diagnosen und soziodemographischen Daten auf Plausibilität geprüft.

Die Skalenwerte wurden weiterhin in Hinblick auf private und berufliche Lebensumstände untersucht. Die Ergebnisse entsprechen den in vielfacher Hinsicht erwarteten Belastungen der unterschiedlichen Teilstichproben.

Konkret fanden sich erhöhte Belastungen bei gegenwärtig arbeitsunfähigen Personen in allen Skalen, wobei sich (inhaltlich nachvollziehbar) die höchste Effektstärke bei den somatischen Symptomen zeigte. Patienten mit einer Zeitrente oder einem laufenden Rentenverfahren waren vor allem bezüglich der depressiven Symptome höher belastet. Stimmig erscheint auch die Beobachtung, dass Patienten, die in der Vergangenheit oder aktuell psychotherapeutisch behandelt wurden, auf allen Skalen höher belastet waren. Befanden sich

die Testpersonen zum Zeitpunkt der Erhebung in einer Psychotherapie, waren die Unterschiede am deutlichsten, wobei die Effektstärken des GSI und der Skala DEPR am größten waren. Arztkontakte als Indikator für eine körperliche oder psychische Belastung führten bei Patienten, die mehr als zweimal in den letzten zwei Wochen einen Arzt aufgesucht hatten, zu signifikant höheren Werten auf allen vier Skalen. Nach Effektstärken führten vor allem somatische Beschwerden und Angsterkrankungen zu vermehrten Arztkontakten. Patienten, die keinen Arzt aufgesucht hatten, waren erwartungsgemäß weniger belastet.

Auch in Hinblick auf private Lebensumstände und die Wohnsituation konnten erwartete Gruppenunterschiede nachgewiesen werden. Es fand sich z.B. in Bezug auf depressive Symptome eine deutlich niedrigere Belastung von Personen mit Partner gegenüber denen ohne Lebenspartner. Bezüglich des GSI waren Geschiedene oder Verwitwete höher belastet als verheiratete Personen. Alleinlebende waren depressiver und erreichten höhere Werte für den GSI als Personen mit einem Lebenspartner.

Das Set von Fragebögen enthielt einzelne Items zum allgemeinen und psychischen Gesundheitszustand der Patienten. Auch hier fanden sich teilweise hochkorrelative positive und negative Zusammenhänge, die eine konvergente bzw. diskriminante Validität des Instruments stützen. Die subjektive Einschränkung der Leistungsfähigkeit korrelierte erwartungsgemäß positiv mit allen 4 Skalen. Schmerzen korrelierten am höchsten mit der Skala SOMA ($r = 0,38$) und kaum mit den restlichen Dimensionen. Die Selbsteinschätzung in Bezug auf eine Depression zeigte hohe Korrelationen zu den Skalen GSI ($r = 0,64$) und DEPR ($r = 0,66$). Die Selbsteinschätzung in Bezug auf Ängste ergab hohe Korrelationen zur Skala ANGS ($r = 0,61$). Die Korrelationen mit den anderen Subskalen waren jeweils geringer ausgeprägt. Patienten mit der Hauptdiagnose Adipositas (E66) waren bis auf die Skala SOMA überall signifikant niedriger belastet.

Bei Patienten mit einer Angststörung (F40-F41) gegenüber Patienten ohne diese Diagnose fand sich übereinstimmend mit der postulierten Skalenstruktur die höchste Effektstärke auf der Skala ANGS. Die Patienten waren auf den Skalen ANGS, GSI und SOMA signifikant höher belastet. Patienten mit einer Depression (F23-F34) erreichten gegenüber denjenigen ohne diese Diagnose auf allen 4 Skalen höhere Testwerte. Auch hier punkten die Patienten skalenkonform mit der höchsten Effektstärke auf der Skala DEPR. Bei Vorliegen einer somatoformen Störung (F45) errechneten sich im Vergleich zu denen ohne Diagnose signifikant höhere Werte auf der Skala SOMA und signifikant niedrigere auf der Skala DEPR.

Die erhobenen Ergebnisse zur kriteriumsbezogenen Validität stimmen im Wesentlichen mit anderen Erhebungen überein. So fanden z.B. Franke et al. (2011) bei Patienten mit einer Depression höhere Summenwerte auf der Skala Depressivität. Angstpatienten waren auf den Skalen SOMA und ANGS höher belastet. Suchtpatienten hatten signifikant niedrigere Belastungen als die Übrigen auf allen 4 Skalen. Insgesamt stützen die Ergebnisse die Validität des Fragebogens, es finden sich für alle Skalen im Wesentlichen erwartete Zusammenhänge mit den untersuchten Kriterien.

Die Interkorrelation der Skalen war in der untersuchten Stichprobe relativ hoch. Es fanden sich erwartungsgemäß die höchsten Korrelationen mit dem GSI zu den Skalen ANGS, SOMA und DEPR ($r = 0,87, 0,74$ und $0,85$). Eine leicht höhere Korrelation der Skala SOMA mit der Skala ANGS ($r = 0,56$) als mit der Skala DEPR ($r = 0,41$) und eine hohe Korrelation der Skala ANGS mit der Skala DEPR ($r = 0,62$) lassen sich inhaltlich gut nachvollziehen. Die Skala SOMA scheint weniger zum GSI beizutragen als die anderen beiden Skalen. Ähnliche Zusammenhänge wurden bereits in anderen Stichproben beobachtet (z.B. Meijer et al. 2011).

Die recht hohen Interkorrelationen der Skalen wurden bereits für die SCL-90-R kritisiert (z.B. Cyr et al. 1985). Allerdings bestehen für die untersuchten Symptome und Syndrome bekanntermaßen hohe Komorbiditätsraten (Fawcett und Kravitz 1983). Daher erscheint es logisch, dass sich die teilweise überlappenden Symptome von Angsterkrankungen, Somatisierungsstörungen und Depressionen auch in einer hohen Interkorrelation der Skalen zeigen.

Die statistischen Zusammenhänge zwischen den korrespondierenden Skalen des BSI-18 und des PHQ-D sind ausnahmslos als hoch zu bewerten ($r = 0,72$ bis $0,80$). Die berechneten Zusammenhänge zwischen den nicht korrespondierenden Konstrukten fielen nachvollziehbar geringer aus. Die *körperliche Summenskala* des SF-12 korrelierte mit der Subskala SOMA ($r = 0,43$), die *psychische Summenskala* zeigte mittelmäßig bis hohe Zusammenhänge mit dem GSI ($r = 0,59$) sowie den Skalen ANGS ($r = 0,52$) und DEPR ($r = 0,64$), die restlichen Korrelationen waren ausnahmslos niedrig. Die Korrelationen der PHQ-D mit dem SF-12 unterschieden sich nur marginal von den Zusammenhängen mit dem BSI-18.

Auch Franke (2010) fand Korrelationen in vergleichbarer Ausprägung mit der *Körperlichen Summenskala* des umfangreicheren Instrument SF-36 und der *Skala Somatisierung* ($r = 0,46$) sowie der *Psychischen Summenskala* und den übrigen drei Indices ($r = 0,55$ bis $0,59$). Die Korrelationen der Indices entsprechen den Erwartungen und bestätigen die theo-

retisch zu vermutenden konvergenten (und divergenten) Zusammenhänge der vier Fragebögen. Die erhobenen Daten scheinen stabil in unterschiedlichen Patientenpopulationen.

In anderen Arbeiten verschiedener Autoren wurden bereits gute Ergebnisse des BSI-18 im Sinne konvergenter Validität publiziert. Es fanden sich unter anderem hohe Korrelationen der Skalen DEPR und ANGS mit den korrespondierenden Skalen der HADS-D (Rusdorf et al. 2010). Ebenfalls wurden hohe Übereinstimmungen des BSI-18 mit dem BDI (Prinz et al. 2013) und dem BAI sowie dem MMPI beschrieben (Andreu et al. 2008).

5.5 Faktorielle Validität

Wie schon die Vorgängerfragebögen (z.B. Cyr et al. 1985) wird der BSI-18 vor allem bezüglich seiner Faktorenstruktur kritisiert. Frühere Arbeiten zur Faktorenstruktur des BSI-18 kamen zu unterschiedlichen Ergebnissen, die möglicherweise durch unterschiedliche Patientenkollektive zu erklären sind. In der Literatur werden im Wesentlichen eine drei- oder vierfaktorielle Lösung diskutiert oder ein einziger latenter Faktor angenommen.

In dieser Arbeit wurde eine exploratorische Faktorenanalyse mittels varimaxrotierter Hauptkomponentenanalyse mit Kaiser-Normalisierung durchgeführt. Es konnten drei Faktoren extrahiert werden, die zusammen eine Varianzaufklärung von 54,1% erreichten. Die extrahierten Faktoren entsprechen weitestgehend der dimensional Struktur des BSI-18. Der erste Faktor enthält hohe Ladungen für alle Items der Skala DEPR. Darüber hinaus zeigt das Item 18 „*Furchtsamkeit*“ ebenfalls eine geringe Ladung auf den „*Depressionsfaktor*“. Der zweite Faktor repräsentiert die Skala ANGS und enthält die entsprechenden sechs Items, während der dritte Faktor schließlich die Skala SOMA wiedergibt, so dass die hypothetische Faktorenstruktur ohne Einschränkungen bestätigt werden konnte.

Eine ähnliche gewählte Stichprobe (638 Psychotherapiepatienten in Deutschland) wurde von Franke et al. (2011) untersucht. Die exploratorische und konfirmatorische Faktorenanalyse konnten die postulierte 3-faktorielle Lösung bestätigen. In weiteren internationalen Publikationen (Recklitis et al., 2006, Galdon et al., 2006; Petkus et al., 2009, Wang et al., 2010) und in einer deutschen klinischen Studie (Franke et al., 2010) fanden sich ebenfalls Hinweise auf die 3-Faktoren-Lösung. In der US-amerikanischen Normstichprobe (Derogatis 2000) erreichten vier Faktoren das Eigenwertkriterium. Während die ersten beiden Faktoren weitestgehend die Skalen DEPR und SOMA abbildeten, waren die Items der Subskala ANGS auf zwei Faktoren verteilt, einen „Angstfaktor“ und einen „Panikfaktor“.

Zabora et. al. (2001) errechneten ebenfalls vier Faktoren, entsprechend der drei Skalen ANGS (5 der 6 Items), DEPR und SOMA. Der vierte Faktor bestand nur aus Item 17 „*Gedanken sich das Leben zu nehmen*“. Bei Galdon et. al. (2008) genügten alle geprüften Modelle den statistischen Kriterien nicht. Asner-Self et al. (2006) und Meijer, de Vries, & van Bruggen (2011) fanden Hinweise, dass der BSI-18 einen großen Faktor darstellt, der im weitesten Sinne psychische Belastungen mit einer robusten Skala misst.

Insgesamt ist die faktorielle Validität in verschiedenen Stichproben unterschiedlich bewertet worden, wobei in niedriger belasteten Stichproben die Skalenstruktur tendenziell weniger robust erscheint. Die uneinheitlichen Ergebnisse wecken Zweifel an der Skalenstruktur des Fragebogens und eine formale und kategoriale Diagnostik sollte mit dem BSI-18 nicht allein und unkritisch erfolgen. Eine Trennung der Skala ANGS in zwei Gruppen zu je drei Items erscheint auf Grund der geringen Skalengröße mit geringer Reliabilität nicht geboten, zumal in dieser Untersuchung kein Hinweis auf eine sinnvolle Differenzierung gefunden wurde.

5.6 Sensitivität und Spezifität des Fragebogens

Unter Verwendung der geschlechtsspezifischen T-Werte für die Skalen des BSI-18 von Franke et. al. (2012) mit Cut-Off-Wert von ≥ 60 ergaben sich für die depressiven Störungen eine Sensitivität von 86,3% (Spezifität von 44,9%). Die Angst- und Panikstörungen wurden (mit einer Spezifität von 29%) zu 92,1% erkannt. Patienten mit somatoformen Störungen konnten von der Skala Somatisierung nur zu 73% identifiziert werden. Auch die Spezifität lag nur bei 38%. Besonders auffällig war hier, dass die Skala ANGS mehr somatoforme Störungen erkannte als die inhaltlich passende Skala SOMA. Die Sensitivität des GSI für alle 3 Krankheitsgruppen lag bei 90,7%, 92,1% und 81,3%. Der GSI ist den kategorialen Skalen bezüglich der Sensitivität (bei deutlich geringerer Spezifität) überlegen.

In dieser Stichprobe war der BSI-18 für die gewählten Cut-Off-Werte sensitiver als der PHQ-D, während sich die Werte für die Spezifität reziprok verhielten. Diese Beobachtung wird mit dem gewählten Cut-Off-Wert zusammenhängen.

Abermals finden sich Hinweise auf eine geringere diagnostische Güte und schlechtere psychometrische Eigenschaften der Subskala SOMA. Dies sollte bei der Interpretation der Ergebnisse besondere Beachtung finden. Bereits zu dem BSI und der SCL-90-R wurden unterschiedliche Vorschläge für Cut-Off-Werte in verschiedenen Populationen vorgeschlagen (Johnson et al. 1996). Laut Manual besteht eine deutliche psychische Belastung, ana-

log zu den Vorgängerfragebögen BSI und SCL-90, bei T-Werten ≥ 63 in mindestens zwei Subskalen oder dem GSI. Die in der Auswertung von SCL-90-R und BSI postulierte alternative Cut-Off-Schwelle ab einem T-Wert von 60, von auffälliger psychischer Belastung zu sprechen (Franke und Derogatis 2002), wurde auch für den BSI-18 vorgeschlagen und in dieser Studie zur Berechnung herangezogen.

Zabora et al. schlugen einen T-Wert von 57 als Cut-Off-Wert vor. Recklitis und Rodriguez (2007) fanden nur eine geringe Übereinstimmung in Bezug auf die positiv getesteten Personen der SCL-90-R und dem BSI-18 unter Verwendung der klassischen Cut-Off-Werte. Es wurde daher vorgeschlagen, den T-Wert des BSI-18 abermals zu senken (> 49). Auch Merport & Recklitis (2012) empfehlen nach einem Vergleich zwischen BSI-18 und der SCL-90 einen Cut-Off-Wert von ≥ 50 . Eine Normierung anhand der Schwesterinstrumente SCL-90 und BSI ist auf Grund der großen Ähnlichkeit nicht sehr valide.

Petkus et. al. (2010) verglichen das BSI-18 mit einem strukturierten Interview. Mit einem T-Wert von 50 wurde eine Sensitivität von 88% beschrieben. Die vorgeschlagenen Cut-Off-Werte von Derogatis (>63) führten nur zu einer Sensitivität von 41%.

Die ROC-Analyse der Skalen des BSI-18, im Vergleich zum PHQ-D und direkten Fragen nach der entsprechenden Störung (z.B. „*Wie sehr litten Sie in den letzten 2 Wochen an Depressionen?*“) wurde ebenfalls durchgeführt.

Für die depressiven Erkrankungen (ICD-10 F32-F34) unterschieden sich die AUC-Werte (0,71 bis 0,74) kaum. Die AUC der Skalen für die Diagnose somatoformer Störungen (F45) fiel deutlich geringer aus. Alle Instrumente konnten Erkrankte und Nichterkrankte kaum trennen. Auffällig war, dass die Frage „*Wie sehr litten Sie in den letzten 2 Wochen unter Schmerzen?*“ einen höheren Wert (0,66) erreichte als die vergleichbaren syndromalen Skalen des BSI-18 und des PHQ-D (0,58 und 0,59). Für die Diagnosen F40 und F41 zeigten sich bessere Ergebnisse. Die Skala ANGS (0,70) scheint dem Ängstlichkeitsindex des PHQ-D (0,62) in punkto AUC überlegen zu sein. Der GSI erreichte einen vergleichbaren Wert von 0,64. Erneut schien eine diagnostische Aussage mit dem externen Item „*Wie sehr litten Sie in den letzten 2 Wochen unter Ängsten?*“ am besten möglich zu sein.

Wetherell, Birchler et al. (2007) verglichen die HADS und das BSI-18 in Bezug auf ihre Differenzierungsfähigkeit von zuvor mit einem strukturierten Interview diagnostizierter Angststörungen bei geriatrischen Patienten. Ein ROC-Analyse zeigte, dass die HADS in der Lage war mit ihrer Angstskala die Patienten hinreichend genau zu erkennen. Die AUC

betrug 0,80. Im Unterschied dazu fiel die AUC für das BSI-18 deutlich geringer aus (0,57). Die HADS ist in diesem Zusammenhang als das zuverlässigere Instrument eingeschätzt worden. Thekkumpurath et al. (2009) verglichen das Distress Thermometer (DT), das Brief Symptom Inventory-18 und den General Health Questionnaire-12 (GHQ-12) in der Palliativmedizin mit den Ergebnissen eines SKID-Interviews. Die ROC-Analyse unterschied sich für die Instrumente DT, BSI-18, und GHQ-12 nicht wesentlich. Die Sensitivität und Spezifität lagen für den BSI-18 bei 78% und 62% und war den beiden anderen Instrumenten minimal überlegen oder gleichwertig.

Insgesamt verdeutlichen die hier erhobenen Daten und die Erkenntnisse aus früheren Studien, dass die diskriminativen Fähigkeiten des Fragebogens auf syndromaler Ebene, insbesondere für die somatoformen Erkrankungen, gering sind. Hohe Sensitivitäten können nur durch Senkung der Cut-Off-Werte und mit einer geringen Spezifität errungen werden. Die Depressionen und Angsterkrankungen werden deutlich besser abgebildet. Es zeigte sich in den zitierten Erhebungen, dass die Spezifität der kategorialen Diagnostik gering ist. Diese Beobachtung kann durch die vorliegende Arbeit gestützt werden. Die Studienlage unterstreicht die Bedeutung des BSI-18 als Screeninginstrument.

5.7 Methodenkritik und Ausblicke hinsichtlich möglicher zukünftiger Forschung, Implikationen und Schlussfolgerungen für die Praxis

Die Stichprobe umfasst insgesamt 1086 ambulante Patienten im Alter zwischen 17 und 88 Jahren. Es liegt somit eine vergleichsweise große Stichprobe vor. Der überwiegende Teil der Patienten ist weiblich (n = 742). Es nahmen deutlich weniger männliche Patienten an der Studie teil (n = 335). Statistische Schwierigkeiten und Einschränkungen ergeben sich unter anderem aus einer ungleichmäßigen Verteilung der Haupt- und Nebendiagnosen und einer fehlenden Kontrollgruppe ohne psychische Erkrankungen. Eine nach Diagnosen oder soziodemographischen Gesichtspunkten gewichtete Auswahl der eingeschlossenen Patienten erfolgte nicht. Wahrscheinlich ist die Stichprobe repräsentativ für Patienten einer psychosomatischen Hochschulambulanz. Auf Grund unterschiedlicher Betriebsorganisation, Schwerpunkte und möglichen Spezialsprechstunden ist ein Vergleich der Ambulanzen nicht ohne weiteres möglich. Zumindest gehören die in dieser Studie am meisten vertretenen Störungsbilder zu den häufigsten Gründen einer Konsultation in einer psychosomatischen Hochschulambulanz. Die Schwere der vorliegenden Störungen dürfte vor dem Hin-

tergrund von häufig vorkommenden stationären Aufnahmen als Folge des Erstgesprächs, schwerer als in der rein ambulanten Versorgung sein.

Als eher schwerwiegendere Einschränkung der erhobenen Daten ist zu beachten, dass die Vergleichsgruppe der „Nichterkrankten“ aus einem Patientenkollektiv besteht und somit tendenziell Belastungen auf Grund weiterer psychiatrischer Erkrankungen bestehen. An dieser Stelle zeigt sich besonders deutlich eine Einschränkung der vorhandenen Erhebung, das Fehlen einer gesunden Vergleichsstichprobe. Unterschiede der Skalenwerte könnten durch das Vorhandensein der untersuchten Störung, weiteren Störungen oder dem Nichtvorhandensein einer anderen erklärt werden. Eine Unterscheidung in Haupt- und Nebendiagnosen erfolgte nicht. Exemplarisch genannt sei hier, dass Patienten mit einer somatoformen Störung (F45) auf der Skala DEPR weniger belastet sind als die Übrigen. Hier ist im Speziellen ein Bias durch Mehrfachdiagnosen wahrscheinlich, denn eine signifikant niedrigere Rate an depressiven Symptomen bei Patienten mit einer somatoformen Störung erscheint inhaltlich nicht stimmig. Kritisch muss daher hinterfragt werden, inwieweit die Beurteilung der kriteriumsbezogenen Validität und der Sensitivität und Spezifität des Fragebogens durch die Stichprobe verfälscht ist.

In Bezug auf sinnvolle Cut-Off-Werte besteht beim BSI-18 bisher kein Konsens. Bei den erhobenen Daten handelt es sich nicht um eine repräsentative Stichprobe der (Sub-) Bevölkerung, die eine Normierung erlaubt. Eine Festlegung auf Cut-Off-Werte (ggf. auch für unterschiedliche Patientenkollektive) ist jedoch von großer praktischer Bedeutung. In diesem Rahmen erscheinen weitere Untersuchungen sinnvoll. Im Rahmen der psychometrischen Evaluation und Erforschung der SCL-90-R und des BSI stellte sich bereits heraus, dass die Fragebögen für unterschiedliche Patientenkollektive und Erkrankungen unterschiedlich gute Ergebnisse und eine unterschiedliche diagnostische Güte zeigen (Johnson et al. 1996).

Aussagen zur Re-Test-Reliabilität können an Hand der vorliegenden Daten nicht beantwortet werden. Für eine deutsche Stichprobe liegen jedoch Daten von Spitzer et al. (2011) vor. Die postulierte faktorielle Struktur wurde in dem vorliegenden Patientenkollektiv durch eine exploratorische Faktorenanalyse gestützt. Vor dem Hintergrund eines der teils kontroversen Ergebnisse aus anderen Erhebungen (s.o.) erscheint dies relevant. Eine konfirmatorische Faktorenanalyse zur Bestätigung der Ergebnisse mit einer weiteren statistischen Methode erfolgte nicht.

Auf Grund einer großen Anzahl von Patienten mit Migrationshintergrund wären Untersuchung zur Verständlichkeit und möglichen Beeinflussung durch Sprachbarrieren und die Frage nach interkulturellen Unterschieden zu klären.

Bei der inzwischen außerordentlichen Vielzahl an unterschiedlichen gut evaluierten Selbstbeurteilungsskalen erscheint eine weitere sinnvolle Zielsetzung der breite Vergleich unterschiedlicher Selbstbeurteilungsinstrumente innerhalb unterschiedlicher Stichproben. An inhaltlich vergleichbaren Selbstbeurteilungsskalen wurde nur der PHQ-D in dieser Studie untersucht. Insbesondere weitere weit verbreitete und gut validierte zeitökonomische Verfahren wie z.B.: HADS, BDI, BAI, SOMS und ADS wären im direkten Vergleich interessant. Auf diesem Wege ließen sich die sonst nur indirekt vergleichbaren psychometrischen Vor- und Nachteile erfassen. Die Frage, welches Selbstbeurteilungsinstrument zur Erfassung von Depressivität, Ängstlichkeit und Somatisierung ausgewählt werden sollte, ist kaum allgemeingültig zu beantworten. Es gibt eine Vielzahl von Instrumenten, die für unterschiedlichste Verwendungszwecke und Patientenkollektive untersucht sind. Der Zweck der Anwendung (Therapiekontrolle, Screening, etc.), die Anlehnung an Diagnosesysteme (DSM, ICD) sowie Anwendungsökonomie oder die Vergleichbarkeit mit anderen wissenschaftlichen Studien sind Beispiele für Entscheidungskriterien. Es gibt für die o.g. Störungsfelder eine Vielzahl an Instrumenten, die teilweise nur in Subpopulationen, im Rahmen der Kreuzvalidierung oder indirekt miteinander verglichen wurden.

Zur Messung von Depressivität werden (neben der SCL-90 und dem BSI-18) beispielsweise häufig die Instrumente PHQ-9, BDI-II und HADS herangezogen. Es gibt für alle genannten Instrumente gute Daten bezüglich ihrer Validität und Reliabilität in Bezug auf die Erfassung von Depressivität. Einen eindeutigen Konsens, welches der Instrumente, nun das geeignetere sei, gibt es nicht. In einem direkten Vergleich waren z.B. die PHQ-9 und die HADS ähnlich valide, während sie sich vor allem in den Grenzwerten bzw. der Einschätzung des Schweregrades einer Störung unterschieden (Cameron et al. 2008, Kendrick et al. 2009). Ähnlich verhält es sich mit den Ängstlichkeit messenden Skalen BAI und GAD-7.

Die Verwendung von Screeninginstrumenten verbessert die Behandlung jedoch nicht zwangsläufig. Eine Übersichtarbeit kam zu dem Ergebnis, dass bei Depressionen durch Screening keine Verbesserung der Therapie erreicht wurde (Thombs u. Ziegelstein, 2014).

6 Zusammenfassung

Das Ziel der vorliegenden Dissertation war die Evaluation der psychometrischen Merkmale der deutschen Version des Brief Symptom Inventory (BSI-18) in einem psychosomatischen Patientenkollektiv. Die untersuchte Stichprobe bestand aus 1086 Patienten aus der Hochschulambulanz des Zentrum für Innere Medizin, Institut und Poliklinik für Psychosomatische Medizin und Psychotherapie in Hamburg an den Standorten Universitätsklinikum Hamburg-Eppendorf (UKE) und in der Schön Klinik Eilbek. Führend waren depressive Erkrankungen (ICD-10: F31–F34; $n = 682$), gefolgt von phobischen und sonstigen Angststörungen (ICD-10: F40/F41; $n = 255$), dissoziativen und somatoformen Störungen (ICD10: F44/F45; $n = 315$).

Es ergaben sich in der Tendenz positive Befunde in der Item- und Reliabilitätsanalyse. Die Reliabilität des GSI (operationalisiert über die interne Konsistenz) ist als mittel bis hoch einzuschätzen ($\alpha = 0,90$). Die Skalen DEPR ($\alpha = 0,87$) und ANGS ($\alpha = 0,84$) zeigen eine mittelmäßige Messgenauigkeit, während die Reliabilität der Skala SOMA ($\alpha = 0,74$) als niedrig zu bewerten ist. Es fanden sich gute konvergente Befunde zu den etablierten Fragebögen SF-12 und PHQ-D und plausible statistische Zusammenhänge zu soziodemographischen Kriterien. Die Subskala SOMA ist den restlichen Skalen nach psychometrischen Merkmalen unterlegen. Die Ergebnisse der exploratorischen Faktorenanalyse stützen die Skalenstruktur des BSI-18. Es wurden drei skalenkonforme Faktoren extrahiert. Untersuchungen der Sensitivität und Spezifität der einzelnen Subskalen zeigen die Schwächen des Instruments in Bezug auf die kategoriale Diagnostik.

Das Instrument sollte daher vornehmlich als Screeningfragebogen eingesetzt werden. Eine Normierung war mit der erhobenen Stichprobe nicht möglich. Es konnte gezeigt werden, dass die deutsche Version des BSI-18 ein reliables, valides und ökonomisches Selbstbeurteilungsinstrument zum Screening von Depression, Somatisierung und Ängstlichkeit ist.

7 Abkürzungsverzeichnis

AUC	Area under the curve
Abb.	Abbildung
ANGS	Skala Ängstlichkeit des BSI-18
BAI	Beck-Angst-Inventar
BDI	Beck-Depressions-Inventar
BSI	Brief Symptom Inventory
BSI-18	Brief Symptom Inventory 18
bzw.	beziehungsweise
CFA	konfirmatorische Faktorenanalyse
d.h.	das heißt
DEPR	Skala Depressivität des BSI-18
d_i	Itemschwierigkeit
EFA	exploratorische Faktorenanalyse
et al.	et alii
ff.	und folgende
GAD-7	Ängstlichkeitsskala des PHQ-D
GSI	Global Severity Index
ICD-10	International Classification of Diseases
KSK	Körperliche Summenskala des SF-12
M	Mittelwert
MMPI	Minnesota Multiphasic Personality Inventory
N	Stichprobengröße
n	Größe der Teilstichprobe
p	Wahrscheinlichkeit
P.	Punkte
PHQ-15	Somatisierungsskala des PHQ-D
PHQ-7	Depressivitätsskala des PHQ-D
PHQ-D	Patient Health Questionnaire
r	Korrelationskoeffizient
R	Spannweite
r_i	korrigierte Trennschärfe
ROC	Receiver-Operated-Curves
r_{tt}	Re-Test-Reliabilität
S.	Seite
s.	siehe
SCL-90-R	Symptom-Checklist-90-R(evised Version)
SD	Standardabweichung
SF-12	Short-Form 12 Health Survey
SF-36	Short-Form 36 Health Survey
SOMA	Skala Somatisierung des BSI-18
Tab.	Tabelle
u.a.	unter anderem
v	Schiefe
vgl.	vergleiche
vs.	Versus
α	Cronbachs Alpha (interne Konsistenz)
γ^2	Exzess

8 Literaturverzeichnis

Teile der Ergebnisse dieser Dissertation wurden im Rahmen einer Ko-Autorenschaft unter dem Titel: „*Die Kurzform des Brief Symptom Inventory (BSI-18): Erste Befunde zu den psychometrischen Kennwerten der deutschen Version*“ im Fachblatt „*Fortschritte der Neurologie – Psychiatrie*“, dem wissenschaftlichen Organ des Berufsverbandes Deutscher Nervenärzte, publiziert (Spitzer et al. 2011).

Adams RE, Boscarino JA, Galea S. *Alcohol use, mental health status and psychological well-being 2 years after the World Trade Center attacks in New York City*. Am J Drug Alcohol Abuse 2006; 32: 203-224.

Aigner M, Paulitsch K, Berg D, Lenz G. *Psychopathologie: Anleitung zur psychiatrischen Exploration*. 1. Aufl. Stuttgart: UTB GmbH; 2014: 62ff.

Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, de Girolamo G, Graaf R, Demeyttenaere K, Gasquet I, Haro JM, Katz SJ, Kessler RC, Kovess V, Lepine JP, Ormel J, Polidori G, Russo LJ, Vilagut G, Almansa J, Arbabzadeh-Bouchez S, Autonell J, Bernal M, Buist-Bouwman MA, Codony M, Domingo-Salvany A, Ferrer M, Joo SS, Martinez-Alonso M, Matschinger H, Mazzi F, Morgan Z, Morosini P, Palacin C, Romera B, Taub N, Vollebergh WA, Eseméd/Mhedeá Investigators ESotEoMDP. *Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project*. Acta Psychiatr Scand Suppl 2004: 21-27.

Alvarez J, Meyer FL, Granoff DL, Lundy A. *The effect of EEG biofeedback on reducing postcancer cognitive impairment*. Integ Cancer Ther 2013; 12: 475-487.

Anderson KE, Gruber-Baldini AL, Vaughan CG, Reich SG, Fishman PS, Weiner WJ, Shulman LM. *Impact of psychogenic movement disorders versus Parkinson's on disability, quality of life, and psychopathology*. Mov Disord 2007; 22: 2204-2209.

Andrade L, Caraveo-Anduaga JJ, Berglund P, Bijl RV, De Graaf R, Vollebergh W, Dragomirecka E, Kohn R, Keller M, Kessler RC, Kawakami N, Kilic C, Offord D, Ustun TB, Wittchen HU. *The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys*. Int J Methods Psychiatr Res 2003; 12: 3-21.

Andreu Y, Galdon MJ, Dura E, Ferrando M, Murgui S, Garcia A, Ibanez E. *Psychometric properties of the Brief Symptoms Inventory-18 (Bsi-18) in a Spanish sample of outpatients with psychiatric disorders*. Psicothema 2008; 20: 844-850.

Asner-Self KK, Schreiber JB, Marotta SA. *A cross-cultural analysis of the Brief Symptom Inventory-18*. Cultur Divers Ethnic Minor Psychol 2006; 12: 367-375.

- Backhaus SL, Ibarra SL, Klyce D, Trexler LE, Malec JF. *Brain injury coping skills group: a preventative intervention for patients with brain injury and their caregivers*. Arch Phys Med Rehabil 2010; 91: 840-848.
- Bellach BM, Ellert U, Radoschewski M. *Der SF-36 im Bundes-Gesundheitssurvey*. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 2000; 43: 210-216.
- Berking M, Rief W. *Klinische Psychologie und Psychotherapie für Bachelor Band I: Grundlagen und Störungswissen Lesen, Hören, Lernen im Web*. In, Springer-Lehrbuch. Berlin, Heidelberg: Springer; 2012: 10ff..
- Bevans M, Wehrlen L, Prachenko O, Soeken K, Zabora J, Wallen GR. *Distress screening in allogeneic hematopoietic stem cell (HSCT) caregivers and patients*. Psychooncology 2011; 20: 615-622.
- Biegler K, Cohen L, Scott S, Hitzhusen K, Parker P, Gilts CD, Canada A, Pisters L. *The role of religion and spirituality in psychological distress prior to surgery for urologic cancer*. Integr Cancer Ther 2012; 11: 212-220.
- Bober SL, Zhou ES, Chen B, Manley PE, Kenney LB, Recklitis CJ. *Sexual function in childhood cancer survivors: a report from Project REACH*. J Sex Med 2013; 10: 2084-2093.
- Bornbaum CC, Fung K, Franklin JH, Nichols A, Yoo J, Doyle PC. *A descriptive analysis of the relationship between quality of life and distress in individuals with head and neck cancer*. Support Care Cancer 2012; 20: 2157-2165.
- Bortz J, Döring N, Bortz D. *Forschungsmethoden und Evaluation : für Human- und Sozialwissenschaftler*. 4., überarb. Aufl. Heidelberg: Springer-Medizin-Verl.; 2006: 220ff, 378ff.
- Boulet J, Boss MW. *Reliability and validity of the Brief Symptom Inventory*. Psychological Assessment 1991; 3: 433-437.
- Brackett J, Krull KR, Scheurer ME, Liu W, Srivastava DK, Stovall M, Merchant TE, Packer RJ, Robison LL, Okcu MF. *Antioxidant enzyme polymorphisms and neuropsychological outcomes in medulloblastoma survivors: a report from the Childhood Cancer Survivor Study*. Neuro Oncol 2012; 14: 1018-1025.
- Bräutigam W, Christian P. *Allgemeines psychosomatisches Syndrom*. Psychosomatische Medizin. Stuttgart: Thieme; 1981: S 320–324.
- Brawman-Mintzer O, Lydiard RB, Emmanuel N, Payeur R, Johnson M, Roberts J, Jarrell MP, Ballenger JC. *Psychiatric comorbidity in patients with generalized anxiety disorder*. Am J Psychiatry 1993; 150: 1216-1218.
- Buchbinder D, Casillas J, Krull KR, Goodman P, Leisenring W, Recklitis C, Alderfer MA, Robison LL, Armstrong GT, Kunin-Batson A, Stuber M, Zeltzer LK. *Psychological outcomes of siblings of cancer survivors: a report from the Childhood Cancer Survivor Study*. Psychooncology 2011; 20: 1259-1268.

- Bühner M. *Einführung in die Test- und Fragebogenkonstruktion*. 3., aktualisierte und erw. Aufl. München u.a.: Pearson Studium; 2011: 30f, 39f, 58f, 141f, 147f, 153f, 161f, 171f, 178f, 217f, 235f, 257f, 295f, 299f, 320f, 340f, 342f.
- Buller R, Maier W, Benkert O. *Das Paniksyndrom: Symptome, Verlauf, Prädiktoren*. Berlin Heidelberg New York Tokio: Springer; 1988.
- Bullinger M, Kirchberger I. *Fragebogen zum Gesundheitszustand : SF-36 ; Handanweisung*. Göttingen u.a.: Hogrefe Verl. für Psychologie; 1998.
- Cameron IM, Crawford JR, Lawton K, Reid IC. *Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care*. Br J Gen Pract 2008; 58: 32-36.
- Cameron OG, Thyer BA, Nesse RM, Curtis GC. *Symptom profiles of patients with DSM-III anxiety disorders*. Am J Psychiatry 1986; 143: 1132-1137.
- Canada AL, Murphy PE, Fitchett G, Peterman AH, Schover LR. *A 3-factor model for the FACIT-Sp*. Psychooncology 2008; 17: 908-916.
- Carlson LE, Angen M, Cullum J, Goodey E, Koopmans J, Lamont L, MacRae JH, Martin M, Pelletier G, Robinson J, Simpson JS, Specca M, Tillotson L, Bultz BD. *High levels of untreated distress and fatigue in cancer patients*. Br J Cancer 2004; 90: 2297-2304.
- Chambers SK, Zajdlewicz L, Youlden DR, Holland JC, Dunn J. *The validity of the distress thermometer in prostate cancer populations*. Psychooncology 2014; 23: 195-203.
- Choi NG, Sirey JA, Bruce ML. *Depression in Homebound Older Adults: Recent Advances in Screening and Psychosocial Interventions*. Curr Transl Geriatr Exp Gerontol Rep 2013; 2: 16-23.
- Clark KL, Loscalzo M, Trask PC, Zabora J, Philip EJ. *Psychological distress in patients with pancreatic cancer--an understudied group*. Psychooncology 2010; 19: 1313-1320.
- Cohen J. *Statistical power analysis for the behavioural sciences* Hillsdale: Lawrence Erlbaum; 1988.
- Cohen M, Gagin R, Cinamon T, Stein T, Moscovitz M, Kuten A. *Translating 'distress' and screening for emotional distress in multicultural cancer patients in Israel*. Qual Life Res 2012; 21: 555-562.
- Comrey AL, Lee HB. *A First Course in Factor Analysis*. 2. Aufl. Hillsdale, NJ: Erlbaum Publishing; 1992.
- Cronbach LJ. *Coefficient alpha and the internal structure of tests*. Psychometrika 1951: 297-334.
- Cyr JJ, McKenna-Foley JM, Peacock E. *Factor structure of the SCL-90-R: is there one?* J Pers Assess 1985; 49: 571-578.

- Derogatis LR. *SCL-90-R: Administration, scoring & procedures manual-I for the R(vised) version*. Johns Hopkins University: School of Medicine: Eigendruck; 1977.
- Derogatis LR. *BSI-18. Brief Symptom Inventory-18. Administration, Scoring, and Procedures Manual*. Minneapolis, MN: NCS Pearson, INC; 2000.
- Derogatis LR, Fitzpatrick M. *The SCL-90-R, the Brief Symptom Inventory (BSI) and the BSI-18*. Maruish ME, Hrsg. The use of psychological testing for treatment planning and outcomes assessment. 3. Ed. Aufl. Mahwah, NJ: Erlbaum; 2004: 1-41.
- Dilling H, Mombour W, Schmidt M. *Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F). Klinisch-diagnostische Leitlinien*. Bern: Huber; 2011: 156f., 190f., 224f..
- Dinning WD, Evans RG. *Discriminant and convergent validity of the SCL-90 in psychiatric inpatients*. J Pers Assess 1977; 41: 304-310.
- Dura E, Andreu Y, Galdon MJ, Ibanez E, Perez S, Ferrando M, Murgui S, Martinez P. *Emotional suppression and breast cancer: validation research on the Spanish Adaptation of the Courtauld Emotional Control Scale (CECS)*. Span J Psychol 2010; 13: 406-417.
- Fawcett J, Kravitz HM. *Anxiety syndromes and their relationship to depressive illness*. J Clin Psychiatry 1983; 44: 8-11.
- Fisseni H-J. *Lehrbuch der psychologischen Diagnostik: mit Hinweisen zur Intervention*. 3., überarb. und erw. Aufl. Göttingen u.a.: Hogrefe; 2004: 32f, 49f, 62f, 80f.
- Franke G, Jäger S, Gläser H, Barkmann C, Brähler E. *Psychometrische Analyse und Normierung des Brief Symptom Inventory 18 (BSI-18) in einer deutschen bevölkerungsrepräsentativen Stichprobe*. In, noch nicht publiziert; 2012: .
- Franke GH, Derogatis LR. *BSI : Brief symptom inventory von L. R. Derogatis ; (Kurzform der SCL-90-R) ; deutsche Version*. Göttingen: Beltz-Test; 2000.
- Franke GH, Derogatis LR. *Symptom-Checkliste von L. R. Derogatis : SCL-90-R ; deutsche Version ; Manual*. 2., vollst. überarb. und neu normierte. Aufl. Göttingen: Beltz-Test-GmbH; 2002.
- Franke GH, Jäger S, Morfeld M, Salewski C, Reimer J, Rensing A, Witzke O, Türk T. *Eignet sich das BSI-18 zur Erfassung der psychischen Belastung von nierentransplantierten Patienten? Is the BSI-18 a useful screening tool for psychological distress in kidney transplanted patients? Zeitschrift für Medizinische Psychologie* 2010; 19: 30-37.
- Franke GH. *Der Einsatz des Brief Symptom Inventory 18(BSI-1 8) bei Psychotherapiepatienten*. Psychotherapie, Psychosomatik, medizinische Psychologie 2011; 61: 82.
- Franke H. *Die Symptom-Checkliste von Derogatis. Deutsche Version*. Weinheim: Beltz; 1994.

- Franz M, Schepank H. *Epidemiologie funktioneller Erkrankungen*. Herrmann JM, Lisker H, Dietze GJ, Hrsg. Funktionelle Erkrankungen: Diagnostische Konzepte und therapeutische Strategien. München, Wien, Baltimore: Urban & Schwarzenberg; 1996: 37–52.
- Fydrich T, Sommer G, Brähler E. *Fragebogen zur Sozialen Unterstützung : F-SozU ; Manual*. Göttingen u.a.: Hogrefe; 2007.
- Galdon MJ, Dura E, Andreu Y, Ferrando M, Murgui S, Perez S, Ibanez E. *Psychometric properties of the Brief Symptom Inventory-18 in a Spanish breast cancer sample*. J Psychosom Res 2008; 65: 533-539.
- Geisheim C. *Das Brief Symptom Inventory (BSI) als Instrument zur Qualitätssicherung in der Psychotherapie*. Diagnostica 2002; 48: 28.
- Gessler S, Low J, Daniells E, Williams R, Brough V, Tookman A, Jones L. *Screening for distress in cancer patients: is the distress thermometer a valid measure in the UK and does it measure change over time? A prospective validation study*. Psychooncology 2008; 17: 538-547.
- Gianinazzi ME, Rueegg CS, Wengenroth L, Bergstraesser E, Rischewski J, Ammann RA, Kuehni CE, Michel G, for Swiss Pediatric Oncology G. *Adolescent survivors of childhood cancer: are they vulnerable for psychological distress?* Psychooncology 2013; 22: 2051-2058.
- Gräfe K, Zipfel S, Herzog W, Löwe B. *Screening psychischer Störungen mit dem "Gesundheitsfragebogen für Patienten (PHQ-D)"*. Diagnostica 2004; 50: 171-181.
- Grassi L, Johansen C, Annunziata MA, Capovilla E, Costantini A, Gritti P, Torta R, Bellani M, Italian Society of Psycho-Oncology Distress Thermometer Study G. *Screening for distress in cancer patients: a multicenter, nationwide study in Italy*. Cancer 2013; 119: 1714-1721.
- Green BL, Gleser GC, Stone WN, Seifert RF. *Relationships among diverse measures of psychotherapy outcome*. J Consult Clin Psychol 1975; 43: 689-699.
- Greenfield BL, Venner KL, Kelly JF, Slaymaker V, Bryan AD. *The impact of depression on abstinence self-efficacy and substance use outcomes among emerging adults in residential treatment*. Psychol Addict Behav 2012; 26: 246-254.
- Gum AM, Petkus A, McDougal SJ, Present M, King-Kallimanis B, Schonfeld L. *Behavioral health needs and problem recognition by older adults receiving home-based aging services*. Int J Geriatr Psychiatry 2009; 24: 400-408.
- Hanel G, Henningsen P, Herzog W, Sauer N, Schaefer R, Szecsenyi J, Lowe B. *Depression, anxiety, and somatoform disorders: vague or distinct categories in primary care? Results from a large cross-sectional study*. J Psychosom Res 2009; 67: 189-197.
- Hautzinger M, Bailer M. *ADS - Allgemeine Depressionsskala*. Göttingen: Beltz; 2003.

- Hautzinger M, Keller F, Kühner C. *Beck-Depressions-Inventar : BDI-II ; Manual*. Rev. Aufl. Frankfurt am Main: Harcourt Test Services; 2006.
- Henningsen P. *Somatisierung und Affektregulation: Elemente eines interpersonellen Modells*. Rudolf G, Henningsen P, Hrsg. Somatoforme Störungen. Stuttgart: Schattauer; 1998: 185-198.
- Herrmann-Lingen C, Buss U, Snaith RP. *HADS-D: Hospital Anxiety and Depression Scale - Deutsche Version. Ein Fragebogen zur Erfassung von Angst und Depressivität in der somatischen Medizin*. Bern: Huber; 1995.
- Hessel A, Schumacher J, Geyer M, Brähler E. *Symptom-Checkliste SCL-90-R: Testtheoretische Überprüfung und Normierung*. Diagnostica 2001; 47: 27-39.
- Hilgers RD, Heitmann KU, Bauer P, Scheiber V. *Einführung in die Medizinische Statistik*. Springer Berlin Heidelberg; 2013: 81ff.
- Hiller W, Rief W, Braehler E. *Somatization in the population: from mild bodily misperceptions to disabling symptoms*. Soc Psychiatry Psychiatr Epidemiol 2006; 41: 704-712.
- Hjorleifsdottir E, Oskarsson GK. *Psychological distress in Icelandic patients with repeated recurrences of cancer*. Int J Palliat Nurs 2010; 16: 586-592.
- Hoffman BM, Zevon MA, D'Arrigo MC, Cecchini TB. *Screening for distress in cancer patients: the NCCN rapid-screening measure*. Psychooncology 2004; 13: 792-799.
- Hopp JL, Anderson KE, Krumholz A, Gruber-Baldini AL, Shulman LM. *Psychogenic seizures and psychogenic movement disorders: are they the same patients?* Epilepsy Behav 2012; 25: 666-669.
- Imayama I, Alfano CM, Kong A, Foster-Schubert KE, Bain CE, Xiao L, Duggan C, Wang CY, Campbell KL, Blackburn GL, McTiernan A. *Dietary weight loss and exercise interventions effects on quality of life in overweight/obese postmenopausal women: a randomized controlled trial*. Int J Behav Nutr Phys Act 2011; 8: 118.
- Jacobi F, Wittchen HU, Holting C, Hofler M, Pfister H, Muller N, Lieb R. *Prevalence, comorbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS)*. Psychol Med 2004; 34: 597-611.
- Jacobsen PB, Donovan KA, Trask PC, Fleishman SB, Zabora J, Baker F, Holland JC. *Screening for psychologic distress in ambulatory cancer patients*. Cancer 2005; 103: 1494-1502.
- Johnson LC, Murphy SA, Dimond M. *Reliability, construct validity, and subscale norms of the Brief Symptom Inventory when administered to bereaved parents*. J Nurs Meas 1996; 4: 117-127.
- Kendrick T, Dowrick C, McBride A, Howe A, Clarke P, Maisey S, Moore M, Smith PW. *Management of depression in UK general practice in relation to scores on*

- depression severity questionnaires: analysis of medical record data. *BMJ* 2009; 338: b750.
- Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. *Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication*. *Arch Gen Psychiatry* 2005; 62: 617-627.
- Khan AA, Khan A, Harezlak J, Tu W, Kroenke K. *Somatic symptoms in primary care: etiology and outcome*. *Psychosomatics* 2003; 44: 471-478.
- Kimron L, Cohen M. *Coping and emotional distress during acute hospitalization in older persons with earlier trauma: the case of Holocaust survivors*. *Qual Life Res* 2012; 21: 783-794.
- Kirmayer LJ, Robbins JM. *Three forms of somatization in primary care: prevalence, co-occurrence, and sociodemographic characteristics*. *J Nerv Ment Dis* 1991; 179: 647-655.
- Kroenke K, Spitzer RL, Williams JB. *The PHQ-9: validity of a brief depression severity measure*. *J Gen Intern Med* 2001; 16: 606-613.
- Kroenke K, Spitzer RL, Williams JB. *The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms*. *Psychosom Med* 2002; 64: 258-266.
- Kroenke K, Spitzer RL, Williams JB, Monahan PO, Lowe B. *Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection*. *Ann Intern Med* 2007; 146: 317-325.
- Kubinger K, Rasch D, Moder K. *Zur Legende der Voraussetzungen des t-Tests für unabhängige Stichproben*. *Psychologische Rundschau* 2009; 60: 26-27.
- Kwak M, Zebrack BJ, Meeske KA, Embry L, Aguilar C, Block R, Hayes-Lattin B, Li Y, Butler M, Cole S. *Trajectories of psychological distress in adolescent and young adult patients with cancer: a 1-year longitudinal study*. *J Clin Oncol* 2013; 31: 2160-2166.
- Leichsenring F. *Komplexe psychische Störungen - Komorbiditätsstrukturen bei stationären Psychotherapiepatienten*. *Complex mental disorders - comorbidity structures in in-patients*. *Nervenheilkunde* 2010; 29: 843.
- Lienert GA, Ratz U, Lienert R. *Testaufbau und Testanalyse*. 5. Aufl. Weinheim: Beltz Psychologie-Verl.-Union; 1994: 58f..
- Lipowski ZJ. *Somatization: the concept and its clinical application*. *Am J Psychiatry* 1988; 145: 1358-1368.
- Low J, Gessler S, Williams R, Daniells E, Brough V, Tookman A, Jones L. *Screening for distress and depression in cancer patients: is ultrashort depression screening a valid measure in the UK? A prospective validation study*. *J Pain Symptom Manage* 2009; 38: 234-243.

- Lowe B, Kroenke K, Grafe K. *Detecting and monitoring depression with a two-item questionnaire (PHQ-2)*. J Psychosom Res 2005; 58: 163-171.
- Löwe B, Spitzer RL, Zipfel S, Herzog W. *PHQ-D. Der Gesundheitsfragebogen für Patienten*. Karlsruhe: Pfizer; 2002.
- MacLean EW, Palsson OS, Turner MJ, Whitehead WE. *Development and validation of new disease-specific measures of somatization and comorbidity in IBS*. J Psychosom Res 2012; 73: 351-355.
- Margraf J, Ehlers A, Beck AT. *Beck-Angst-Inventar: BAI; Manual*. Frankfurt am Main: Harcourt Test Services; 2007.
- Martinez P, Galdon MJ, Andreu Y, Ibanez E. *The Distress Thermometer in Spanish cancer patients: convergent validity and diagnostic accuracy*. Support Care Cancer 2013; 21: 3095-3102.
- Meijer RR, de Vries RM, van Bruggen V. *An evaluation of the Brief Symptom Inventory-18 using item response theory: which items are most strongly related to psychological distress?* Psychol Assess 2011; 23: 193-202.
- Meller I, Fichter MM, Schroppel H. *Incidence of depression in octo- and nonagenarians: results of an epidemiological follow-up community study*. Eur Arch Psychiatry Clin Neurosci 1996; 246: 93-99.
- Merport A, Recklitis CJ. *Does the Brief Symptom Inventory-18 case rule apply in adult survivors of childhood cancer? Comparison with the Symptom Checklist-90*. J Pediatr Psychol 2012; 37: 650-659.
- Mertens AC, Sencer S, Myers CD, Recklitis C, Kadan-Lottick N, Whitton J, Marina N, Robison LL, Zeltzer L. *Complementary and alternative therapy use in adult survivors of childhood cancer: a report from the Childhood Cancer Survivor Study*. Pediatr Blood Cancer 2008; 50: 90-97.
- Michel L, Conrad W. *Theoretische Grundlagen psychometrischer Tests*. Groffmann K-J, Michel LH, Hrsg. Enzyklopädie der Psychologie, Themenbereich B Methodologie und Methoden. Göttingen: Hogrefe; 1982: 11-29.
- Möller H-J, Laux G, Deister A. *Psychiatrie und Psychotherapie*. Duale Reihe. 4. Aufl. Stuttgart: Thieme; 2009: 76ff, 110ff, .
- Mulrow CD, Williams JW, Jr., Gerety MB, Ramirez G, Montiel OM, Kerber C. *Case-finding instruments for depression in primary care settings*. Ann Intern Med 1995; 122: 913-921.
- Mustanski BS, Garofalo R, Emerson EM. *Mental health disorders, psychological distress, and suicidality in a diverse sample of lesbian, gay, bisexual, and transgender youths*. Am J Public Health 2010; 100: 2426-2432.

- Nekolaichuk CL, Cumming C, Turner J, Yushchyshyn A, Sela R. *Referral patterns and psychosocial distress in cancer patients accessing a psycho-oncology counseling service*. *Psychooncology* 2011; 20: 326-332.
- Noyes R, Jr., Woodman C, Garvey MJ, Cook BL, Suelzer M, Clancy J, Anderson DJ. *Generalized anxiety disorder vs. panic disorder. Distinguishing characteristics and patterns of comorbidity*. *J Nerv Ment Dis* 1992; 180: 369-379.
- Palapattu GS, Bastian PJ, Slavney PR, Haisfield-Wolfe ME, Walker JM, Brintzenhofesoc K, Trock B, Zabora J, Schoenberg MP. *Preoperative somatic symptoms are associated with disease progression in patients with bladder carcinoma after cystectomy*. *Cancer* 2004; 101: 2209-2213.
- Paykel ES, Brugha T, Fryers T. *Size and burden of depressive disorders in Europe*. *Eur Neuropsychopharmacol* 2005; 15: 411-423.
- Petermann F, Eid M, Bengel J. *Handbuch der psychologischen Diagnostik*. Göttingen u.a.: Hogrefe; 2006: 288ff.
- Petkus AJ, Gum AM, King-Kallimanis B, Wetherell JL. *Trauma history is associated with psychological distress and somatic symptoms in homebound older adults*. *Am J Geriatr Psychiatry* 2009; 17: 810-818.
- Petkus AJ, Gum AM, Small B, Malcarne VL, Stein MB, Wetherell JL. *Evaluation of the factor structure and psychometric properties of the Brief Symptom Inventory-18 with homebound older adults*. *Int J Geriatr Psychiatry* 2010; 25: 578-587.
- Petkus AJ, Gum A, Wetherell JL. *Thought suppression is associated with psychological distress in homebound older adults*. *Depress Anxiety* 2012; 29: 219-225.
- Piersma HL, Boes JL, Reaume WM. *Unidimensionality of the Brief Symptom Inventory (BSI) in adult and adolescent inpatients*. *J Pers Assess* 1994; 63: 338-344.
- Prinz U, Nutzinger DO, Schulz H, Petermann F, Braukhaus C, Andreas S. *Comparative psychometric analyses of the SCL-90-R and its short versions in patients with affective disorders*. *BMC Psychiatry* 2013; 13: 104.
- Pschyrembel W. *Pschyrembel klinisches Wörterbuch*. 259., neu bearb. Aufl. Berlin: de Gruyter; 2002: S. 904, 1617 und 1619
- Rasch D, Guiard V. *The robustness of parametric statistical methods*. *Psychology Science* 2004; 46: 175-208.
- Recklitis CJ, Parsons SK, Shih MC, Mertens A, Robison LL, Zeltzer L. *Factor structure of the brief symptom inventory--18 in adult survivors of childhood cancer: results from the childhood cancer survivor study*. *Psychol Assess* 2006; 18: 22-32.
- Recklitis CJ, Rodriguez P. *Screening childhood cancer survivors with the brief symptom inventory-18: classification agreement with the symptom checklist-90-revised*. *Psychooncology* 2007; 16: 429-436.

- Rief W, Hiller W, Heuser J. *SOMS. Das Screening für Somatoforme Störungen. Manual zum Fragebogen*. Bern: Huber; 1997.
- Rost J. *Lehrbuch Testtheorie - Testkonstruktion*. 2., vollst. überarb. und erw. Aufl. Bern u.a.: Huber; 2004: 17f, 32f.
- Roth M, Herzberg Yorck P. *Psychodiagnostik in der Praxis: State of the Art? Klinische Diagnostik und Evaluation* 2008; 1: 5-18.
- Ruipérez MA, Ibáñez MI, Lorente E, Moro M, Ortet G. *Psychometric properties of the Spanish version of the BSI: Contributions to the relationship between personality and psychopathology*. European Journal of Psychological Assessment 2001; 17: 241-250.
- Rusdorf S, Franke GH, Jäger S, Küch D, Rossband H, Morfeld M. *Psychometrische Prüfung des BSI-18*. 11. Nachwuchswissenschaftlerkonferenz der Fachhochschulen Sachsens, Sachsen-Anhalts und Thüringens; 2010; Fachhochschule Schmalkalden.
- Scharfetter C. *Allgemeine Psychopathologie: eine Einführung ; 26 Tabellen*. Thieme; 2002: 163ff.
- Schmitt N. *Uses and abuses of coefficient alpha*. Psychological Assessment 1996: 350-353.
- Schmitz N, Kruse J, Heckrath C, Alberti L, Tress W. *Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments*. Soc Psychiatry Psychiatr Epidemiol 1999; 34: 360-366.
- Schneekloth TD, Jowsey SG, Biernacka JM, Burton MC, Vasquez AR, Bergquist T, Drews MS, Gores GJ. *Pretransplant psychiatric and substance use comorbidity in patients with cholangiocarcinoma who received a liver transplant*. Psychosomatics 2012; 53: 116-122.
- Schoepf D, Heun R, Weiffenbach O, Herrmann S, Maier W. *4-Wochen-Pravalenz somatoformer Storungen und assoziierte psychosoziale Beeinträchtigung*. Nervenarzt 2003; 74: 245-251.
- Serber ER, Edwards-Hampton SA, Yeager B, Clair M, Taylor M, Galloway SK, Balliet WE, Madan A, Borckardt JJ. *Prevalence of Chest Pain, Depression, Somatization, Anxiety, Global Distress, and Substance Use among Cardiac and Pulmonary Rehabilitation Patients*. Pain Res Treat 2012; 2012: 138680.
- Sieberg CB, Williams S, Simons LE. *Do parent protective responses mediate the relation between parent distress and child functional disability among children with chronic pain?* J Pediatr Psychol 2011; 36: 1043-1051.
- Spitzer C, Hammer S, Lowe B, Grabe HJ, Barnow S, Rose M, Wingenfeld K, Freyberger HJ, Franke GH. *Die Kurzform des Brief Symptom Inventory (BSI-18): erste Befunde zu den psychometrischen Kennwerten der deutschen Version*. Fortschr Neurol Psychiatr 2011; 79: 517-523.

- Spitzer RL, Kroenke K, Williams JB. *Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire.* JAMA 1999; 282: 1737-1744.
- Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J. *Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study.* Am J Obstet Gynecol 2000; 183: 759-769.
- Streiner DL, Norman GR. *Health measurement scales : a practical guide to their development and use.* Oxford: Oxford University Press; 2003.
- Tarlov AR, Ware JE, Jr., Greenfield S, Nelson EC, Perrin E, Zubkoff M. *The Medical Outcomes Study. An application of methods for monitoring the results of medical care.* JAMA 1989; 262: 925-930.
- Thekkumpurath P, Venkateswaran C, Kumar M, Newsham A, Bennett MI. *Screening for psychological distress in palliative care: performance of touch screen questionnaires compared with semistructured psychiatric interview.* J Pain Symptom Manage 2009; 38: 597-605.
- Todd DM, Deane FP, McKenna PA. *Appropriateness of SCL-90-R Adolescent and Adults Norms for Outpatient and Nonpatient College Students.* Journal of Counseling Psychology 1997; 44: 294-301.
- Torres L, Miller MJ, Moore KM. *Factorial invariance of the Brief Symptom Inventory-18 (BSI-18) for adults of Mexican descent across nativity status, language format, and gender.* Psychol Assess 2013; 25: 300-305.
- Tremolada M, Bonichini S, Schiavo S, Pillon M. *Post-traumatic stress symptoms in mothers of children with leukaemia undergoing the first 12 months of therapy: predictive models.* Psychol Health 2012; 27: 1448-1462.
- Utsey SO, Hook JN. *Heart rate variability as a physiological moderator of the relationship between race-related stress and psychological distress in African Americans.* Cultur Divers Ethnic Minor Psychol 2007; 13: 250-253.
- Voegtline KM, Stifter CA, Family Life Project I. *Late-preterm birth, maternal symptomatology, and infant negativity.* Infant Behav Dev 2010; 33: 545-554.
- Vollrath M, Koch R, Angst J. *The Zurich Study. IX. Panic disorder and sporadic panic: symptoms, diagnosis, prevalence, and overlap with depression.* Eur Arch Psychiatry Neurol Sci 1990; 239: 221-230.
- Wang J, Kelly BC, Liu T, Zhang G, Hao W. *Factorial structure of the Brief Symptom Inventory (BSI)-18 among Chinese drug users.* Drug Alcohol Depend 2013; 133: 368-375.
- Weisman AD, Worden JW. *The existential plight in cancer: significance of the first 100 days.* Int J Psychiatry Med 1976; 7: 1-15.

- Wenninger K, Helmes A, Bengel J, Lauten M, Volkel S, Niemeyer CM. *Coping in long-term survivors of childhood cancer: relations to psychological distress.* Psychooncology 2013; 22: 854-861.
- Wetherell JL, Birchler GD, Ramsdell J, Unutzer J. *Screening for generalized anxiety disorder in geriatric primary care patients.* Int J Geriatr Psychiatry 2007; 22: 115-123.
- Zabora J, BrintzenhofeSzoc K, Jacobsen P, Curbow B, Piantadosi S, Hooker C, Owens A, Derogatis L. *A new psychosocial screening instrument for use with cancer patients.* Psychosomatics 2001; 42: 241-246.
- Zebrack BJ, Zevon MA, Turk N, Nagarajan R, Whitton J, Robison LL, Zeltzer LK. *Psychological distress in long-term survivors of solid tumors diagnosed in childhood: a report from the childhood cancer survivor study.* Pediatr Blood Cancer 2007; 49: 47-51.
- Zeltzer LK, Recklitis C, Buchbinder D, Zebrack B, Casillas J, Tsao JC, Lu Q, Krull K. *Psychological status in childhood cancer survivors: a report from the Childhood Cancer Survivor Study.* J Clin Oncol 2009; 27: 2396-2404.

9 Tabellenverzeichnis

Tabelle 1: Einjahres-Prävalenz funktioneller Beschwerden in der erwachsenen Allgemeinbevölkerung Mannheims (Franz und Schepank 1996)	11
Tabelle 2: Eine Auswahl an phobischen und anderen Angststörungen nach ICD-10	13
Tabelle 3: Die häufigsten Symptome einer Panikattacke nach Buller et. al. (1998)	14
Tabelle 4: Häufig genutzte Selbstbeurteilungsverfahren für Depressivität, Ängstlichkeit und Somatisierung	15
Tabelle 5: Übersicht über die Skalen und Itemanzahl der SCL-90 und ihrer Nachfolger	17
Tabelle 6: Verteilung der Items des BSI-18 auf die Subskalen. „Wie sehr litten sie in den letzten 7 Tagen unter...?“	20
Tabelle 7: Interne Konsistenz des BSI-18 in unterschiedlichen Stichproben	22
Tabelle 8: Geschlechterunterschiede von Cronbachs α nach (Asner-Self et al. 2006).	22
Tabelle 9: Ladungen ($> 0,40$) der Faktorenanalyse der Bevölkerungsstichprobe des BSI-18 (Derogatis 2000).	23
Tabelle 10: Übersicht über die Studien zur faktoriellen Struktur des BSI-18 in Anlehnung an Franke (2012)	27
Tabelle 11: Korrelation zwischen GSI und Subskalen mit vergleichbaren Skalen der SCL-90-R anhand der Bevölkerungsstichprobe (N = 1134)	29
Tabelle 12: Überblick über konvergente und divergente Validität des BSI-18 in Anlehnung an Franke (2012)	31
Tabelle 13: Ausgewählte Diagnosen – Häufigkeiten in der Stichprobe (N=1086)	39
Tabelle 14: Deskriptive Statistik der BSI - 18 Items (N = 1086)	53
Tabelle 15: Korrigierte Item-Skala-Korrelation bzw. Tensschärfe (r_i) des BSI-18 nach Skalen (N = 1086)	54
Tabelle 16: Inter-Item-Korrelationen der BSI-18 Skalen (N = 1086)	54
Tabelle 17: Skalenstatistiken und Cronbachs Alpha der 4 Subskalen des BSI-18 (N = 1086)	55
Tabelle 18: Reliabilität nach Geschlecht und Bildungsstand (Cronbachs α)	55
Tabelle 19: Korrelation nach Spearman der Skalen untereinander (N =1086)	56
Tabelle 20: Skalen nach Geschlecht mit Mann-Whitney-U-Test Signifikanzniveau (2-seitig)	57
Tabelle 21: GSI und Subskalenwerte des BSI-18 differenziert nach Alter	57
Tabelle 22: BSI-18 Skalen nach Bildungsstand mit Kruskal-Wallis-Test (N = 1028)	58

Tabelle 23: BSI-18 Skalenunterschiede differenziert nach Nationalität	59
Tabelle 24: BSI-18 Skalenunterschiede differenziert nach Arbeitsunfähigkeit	59
Tabelle 25: BSI-18 Skalenunterschiede differenziert nach Berentung (N=1015)	60
Tabelle 26: Skalen nach Familienstand (N = 1052)	61
Tabelle 27: Skalen nach Wohnsituation (N = 1012)	61
Tabelle 28: Skalen nach Psychotherapieerfahrungen (N = 1058)	62
Tabelle 29: Skalen nach Arztkontakten in den letzten 2 Wochen (N = 1045)	62
Tabelle 30: Korrelation der Skalen mit der Selbstauskunft nach dem Befinden in den letzten 2 Wochen (auf einer Skala von 1-10, N = 1080)	63
Tabelle 31: Korrelation der Skalen mit „Wie beurteilen Sie Ihren Gesundheitszustand in den letzten 2 Wochen?“ (N = 1080)	63
Tabelle 32: BSI-18 Mittlere Testergebnisse in Abhängigkeit der im Erstgespräch gestellten Diagnose	65
Tabelle 33: BSI-18 Skalenunterschiede bei Patienten mit einer depressiven Störung (F32/F33/F34)	66
Tabelle 34: BSI-18 Skalenunterschiede bei Patienten mit einer somatoformen Störung	66
Tabelle 35: BSI-18 Skalenunterschiede bei Patienten mit einer Angststörung (F40/F41)	66
Tabelle 36: Korrelation der BSI-18-Skalen mit den dimensional Skalen des PHQ-D	67
Tabelle 37: Korrelationen der Skalen des SF-12 mit dem BSI-18 (N = 716)	68
Tabelle 38: BSI-18 exploratorische Faktorenanalyse - Erklärte Gesamtvarianz (N = 1086)	68
Tabelle 39: BSI-18 exploratorische Faktorenanalyse – Komponentenmatrix mit Ladungen größer 0,40 (N = 1086)	70
Tabelle 40: Klassifikationsbewertung des BSI-18	72
Tabelle 41: Soziodemographische Merkmale der Stichprobe	107
Tabelle 42: Verteilung der Diagnosen und Komorbiditäten N = 1086	109

10 Abbildungsverzeichnis

- Abbildung 1:** Scree-Test der Faktorenanalyse des BSI-18 69
- Abbildung 2:** ROC Kurve der Diagnosen F32-F34 (N = 1020) der Skalen GSI, DEPR, PHQ-9 („Depressivität“), Selbsteinschätzung der Depressivität (auf einer Skala von 1-10) 73
- Abbildung 3:** ROC Kurve der Diagnosen F45 (N = 905) für die Skalen GSI, SOMA, PHQ-15 („Somatisierung“), Selbsteinschätzung von Schmerzen (auf einer Skala von 1-10) 74
- Abbildung 4:** ROC Kurve der Diagnosen F40/F41 (N = 1002) für die Skalen GSI, ANGS, GAD-7 („Ängstlichkeit“), Selbsteinschätzung der Ängstlichkeit (auf einer Skala von 1-10) 74

11 Anhang

Tabelle 41: Soziodemographische Merkmale der Stichprobe

Geschlecht (N = 1077)	männlich	335	31,1%
	weiblich	742	68,9%
Familienstand (N = 1080)	ledig	503	46,6%
	verheiratet	341	31,6%
	getrennt lebend	42	3,9%
	geschieden	131	12,1%
	verwitwet	35	3,2%
	Sonstiges	28	2,6%
Lebens- und Wohnsituation (N = 1078)	allein	347	32,2%
	mit Partner	324	30,1%
	allein mit Kind(ern)	64	5,9%
	mit Partner und Kind(ern)	182	16,9%
	mit Eltern	95	8,8%
	in Institutionen	7	0,6%
	Sonstiges	59	5,5%
Ihr höchster Schulabschluss? (N = 1069)	noch in der Schule	8	0,7%
	Haupt-/Volksschulabschluss	218	20,4%
	Abschluss - Polytechnische Oberschule	18	1,7%
	Realschulabschluss	356	33,3%
	Fachabitur/ Abitur	249	23,3%
	abgeschlossenes (Fach-) Hochschulstudium	187	17,5%
	Sonstiges	33	3,1%
aktueller beruflicher Status (N = 1069)	selbstständig	52	4,9%
	Beamter/in	15	1,4%
	Angestellter	433	40,7%
	Arbeiter	51	4,8%
	Sonstiges	36	3,4%
	Hausfrau/mann	58	5,4%
	arbeitslos	167	15,7%
	Rente (Früh-, Alters-, Witwenrente)	94	8,8%
	Erwerbs/Berufsunfähigkeits-rente	35	3,3%
	Studium/Schule	88	8,3%
	Sonstiges	32	3,0%
Sind sie zur Zeit arbeitsunfähig? (N = 838)	nein	565	67,4%
	ja, seit längerem	273	32,6%
Sind sie berentet? (N = 1015)	Nein	863	85,0%
	Altersrente	86	8,5%
	Zeitrente	36	3,5%
	laufendes Rentenverfahren	30	3,0%
Wie beurteilen Sie Ihren körperlichen Gesundheitszustand in den letzten 2 Wochen? (N = 1065)	nicht beeinträchtigt	54	5,1%
	leicht beeinträchtigt	163	15,3%
	mittelmäßig beeinträchtigt	326	30,6%
	erheblich beeinträchtigt	376	35,3%
	schwer beeinträchtigt	146	13,7%
Wie beurteilen Sie Ihren seelischen Gesundheitszustand in den letzten 2 Wochen? (N = 1058)	nicht beeinträchtigt	40	3,8%
	leicht beeinträchtigt	112	10,6%
	mittelmäßig beeinträchtigt	309	29,2%
	erheblich beeinträchtigt	403	38,1%
	schwer beeinträchtigt	194	18,3%
Welche Medikamente nehmen Sie regelmäßig? (Mehrfachnennungen möglich)	Antidepressiva/ Stimmungsaufheller	316	29%
	Beruhigungsmittel	174	16%
	Schlafmittel	105	10%
	Schmerzmittel	188	17%

	Abführmittel	15	1%
	andere	279	26%
	Keine Angabe / Keine Medikamente	280	26%
Sind Sie früher oder sind Sie zur Zeit in psychotherapeutischer Behandlung? (N = 1058)	noch nie	435	41,1%
	früher	447	42,2%
	zurzeit in Behandlung	176	16,6%
Wie oft haben Sie in den letzten zwei Wochen einen Arzt aufgesucht? (N = 1045)	kein mal	76	7,3%
	1-mal	243	23,3%
	2-mal	301	28,8%
	mehrmals	425	40,7%

Tabelle 42: Verteilung der Diagnosen und Komorbiditäten N = 1086

Diagnosen		F32	F33	F34	F40	F41	F43	F44	F45	F50	F54	F1	F60	E66	Som.
F32	n	442	-	19	31	74	19	4	121	46	22	30	26	45	95
	%	100,0	-	4,3	7,0	16,7	4,3	0,9	27,4	10,4	5,0	6,8	5,9	10,2	21,5
F33	n	-	198	12	20	35	9	1	53	32	8	22	20	21	42
	%	-	100,0	6,1	10,1	17,7	4,5	0,5	26,8	16,2	4,0	11,1	10,1	10,6	21,2
F34	n	19	12	72	6	5	4	-	17	5	9	4	5	3	15
	%	26,4	16,7	100,0	8,3	6,9	5,6	-	23,6	6,9	12,5	5,6	6,9	4,2	20,8
F40	n	31	20	6	86	14	7	-	21	11	4	8	7	8	14
	%	36,0	23,3	7,0	100,0	16,3	8,1	-	24,4	12,8	4,7	9,3	8,1	9,3	16,3
F41	n	74	35	5	14	169	11	1	39	9	5	16	7	9	22
	%	43,8	20,7	3,0	8,3	100,0	6,5	0,6	23,1	5,3	3,0	9,5	4,1	5,3	13,0
F43	n	19	9	4	7	11	114	4	20	6	6	9	8	5	15
	%	16,7	7,9	3,5	6,1	9,6	100,0	3,5	17,5	5,3	5,3	7,9	7,0	4,4	13,2
F44	n	4	1	-	-	1	4	19	3	2	1	2	1	-	1
	%	21,1	5,3	-	-	5,3	21,1	100,0	15,8	10,5	5,3	10,5	5,3	-	5,3
F45	n	121	53	17	21	39	20	3	299	8	16	17	16	11	42
	%	40,5	17,7	5,7	7,0	13,0	6,7	1,0	100,0	2,7	5,4	5,7	5,4	3,7	14,0
F50	n	46	32	5	11	9	6	2	8	130	1	11	15	15	20
	%	35,4	24,6	3,8	8,5	6,9	4,6	1,5	6,2	100,0	0,8	8,5	11,5	11,5	15,4
F54	n	22	8	9	4	5	6	1	16	1	66	2	2	5	66
	%	33,3	12,1	13,6	6,1	7,6	9,1	1,5	24,2	1,5	100,0	3,0	3,0	7,6	100,0
F1	n	30	22	4	8	16	9	2	17	11	2	76	9	3	7
	%	39,5	28,9	5,3	10,5	21,1	11,8	2,6	22,4	14,5	2,6	100,0	11,8	3,9	9,2
F60	n	26	20	5	7	7	8	1	16	15	2	9	70	6	9
	%	37,1	28,6	7,1	10,0	10,0	11,4	1,4	22,9	21,4	2,9	12,9	100,0	8,6	12,9
E66	n	45	21	3	8	9	5	-	11	15	5	3	6	131	131
	%	34,4	16,0	2,3	6,1	6,9	3,8	-	8,4	11,5	3,8	2,3	4,6	100,0	100,0
Som.	n	95	42	15	14	22	15	1	42	20	66	7	9	131	271
	%	35,1	15,5	5,5	5,2	8,1	5,5	0,4	15,5	7,4	24,4	2,6	3,3	48,3	100,0

12 Danksagung

Ich möchte zunächst den Patienten, deren Auskünfte ich für diese Untersuchung verwenden durfte, für Ihr Vertrauen und Ihre ehrlichen Antworten danken.

Ich danke weiterhin Herrn Prof. Dr. Carsten Spitzer und Herrn Prof. Dr. Bernd Löwe für ihre kompetente und freundliche Betreuung, Motivation und Unterstützung.

Darüber hinaus möchte ich allen Mitarbeitern des Zentrum für Innere Medizin, Institut und Poliklinik für Psychosomatische Medizin und Psychotherapie an den Standorten Universitätsklinikum Hamburg-Eppendorf (UKE) und in der Schön Klinik Eilbek für ihre Hilfsbereitschaft und die freundliche und angenehme Atmosphäre danken.

Ich danke meiner Lebenspartnerin für ihre Geduld, ihre Motivation und orthografische Unterstützung. Zuletzt danke ich meinen Eltern für die liebevolle Begleitung und großzügige Unterstützung meiner Ausbildung.

13 Eidesstattliche Erklärung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Sebastian Hammer