

C. Kühner<sup>3</sup> · C. Bürger<sup>3</sup> · F. Keller<sup>1</sup> · M. Hautzinger<sup>2</sup>

<sup>1</sup> Abteilung Kinder- und Jugendpsychiatrie, Universitätsklinikum Ulm

<sup>2</sup> Abteilung Klinische und Entwicklungspsychologie am Psychologischen Institut der Universität Tübingen

<sup>3</sup> AG Verlaufs- und Interventionsforschung, Zentralinstitut für Seelische Gesundheit, Mannheim

# Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II)

## Befunde aus deutschsprachigen Stichproben

Das Beck-Depressionsinventar (BDI) [5, 13] ist das weltweit am weitesten verbreitete Selbstbeurteilungsinstrument zur Beurteilung der Depressionsschwere. Zahlreiche Studien bescheinigen dem Verfahren gute psychometrische Gütekriterien wie interne Konsistenz, Retestreliabilität, Konstruktvalidität und Änderungsensitivität unter Behandlung [6, 12]. Mit der 3. und weiteren Revisionen des Diagnostischen und Statistischen Manuals Psychischer Störungen stellte sich jedoch die Frage, inwieweit das BDI geeignet ist, die entsprechenden depressiven Symptomkriterien adäquat abzubilden. So erfasst das Instrument nur 6 der 9 DSM-IV-Kriterien-Symptome vollständig, 2 Kriterien (Schlaf, Appetit) sind nur teilweise, ein weiteres (Agitation) ist überhaupt nicht erfasst. Dagegen enthält das BDI auch Items, die für die Diagnosestellung irrelevant sind. Diese Überlegungen führten 1996 zur Entwicklung des BDI-II [7]. Hauptanliegen war eine Verbesserung der inhaltlichen Validität des Verfahrens durch Aufnahme neuer diagnoserelevanter und Ausschluss diagnoseirrelevanter Items.

Gegenüber dem BDI wurden nur 3 Items in ursprünglicher Form beibehal-

ten (Bestrafungsgefühle, Selbstmordimpulse, Libidoverlust), 3 diagnoseunspezifische Items wurden gestrichen (negatives Körperbild, Arbeitsunfähigkeit, Hypochondrie) und bei den restlichen Items mindestens eine Antwortalternative umformuliert. Der Beurteilungszeitraum der Items wurde in Anlehnung an DSM-IV auf 2 Wochen erweitert [7].

Beck et al. [7] berechneten anhand von ROC-Analysen für das BDI-II neue Cut-Offs zur Einschätzung des Schweregrads einer Depression. Danach zeigen 0–13 Punkte keine oder minimale, 14–19 Punkte milde, 20–28 Punkte moderate, und 29–63 Punkte schwere depressive Symptomatik an.

Publikationen zu Gütekriterien des BDI-II liegen derzeit hauptsächlich aus dem englisch- und spanischsprachigen Raum vor. Faktorenanalytische Untersuchungen identifizierten 2–3 hoch interkorrelierte Faktoren, die dafür sprechen, auch weiterhin den Summenwert des Fragebogens als Maß selbstbeurteilter Depressionsschwere zu verwenden [7]. Die interne Konsistenz des BDI-II liegt in klinischen und nichtklinischen Stichproben im Bereich von  $0,84 \leq \alpha \leq 0,94$  [2, 7, 8, 16, 18, 27, 28]. In einer Patientenstichprobe lag

die Retestreliabilität im Einwochenzeitraum bei  $r_{tt}=0,93$  [5]. Mit einer früheren BDI-Version korrelierte das BDI-II im Bereich zwischen  $r=0,84$  und  $r=0,93$  [7, 10], dabei lagen die durchschnittlichen BDI-II-Werte um 2–3 Punkte höher. Dies korrespondiert mit der Erhöhung der Cut-Offs zur Einschätzung der Depressionsschwere (s. oben).

Mit anderen Depressionsskalen korreliert das Verfahren im Bereich von  $r=0,68$ – $0,89$  [2, 7, 16, 27], etwas geringer ( $0,37 \leq r \leq 0,60$ ) mit selbst- und fremdbeurteilter Angst [7, 27, 28] sowie im ähnlichen Bereich mit konstrukt-nahen kognitiven Skalen [7, 8]. Bislang fehlen Arbeiten zu Zusammenhängen zwischen BDI-II und konstruktfernen Verfahren (z. B. Persönlichkeitstraits). Andere Untersuchungen zeigten, dass das BDI-II zwischen depressiven und Patienten mit anderen psychiatrischen Störungen [2, 17, 23] und zwischen verschiedenen Schweregraden einer Depression [3, 29] differenzieren kann. Damit bescheinigen die internationalen Studien dem BDI-II gute Reliabilitäts- und Validitätskennwerte. Entsprechende Studien im deutschen Sprachraum liegen bislang jedoch nicht vor.

**Tab. 1** Verteilungsparameter des BDI-II-Summenwerts in verschiedenen Mannheimer Stichproben

	Akut Depressive (G1) (n=36)	Remittierte Depressive (G2) (T1, n=39)	(T2, n=65)	Nichtklinische Stichprobe (G3) (n=315)
Parameter				
Arithmetischer Mittelwert	33,06	15,02	10,45	7,69
Standardabweichung	9,43	8,53	8,76	7,52
Konfidenzintervall M (95%)	29,9–36,3	12,3–17,8	8,3–12,6	6,9–8,5
Median	31,50	15,00	9,00	6,00
Schiefe	0,07	–0,09	0,75	1,52 <sup>a</sup>
Kurtosis	–0,76	–0,87	–0,16	2,91
Streubreite	35	32	35	41

<sup>a</sup>Signifikante Abweichung von der Normalverteilung ( $p < 0,001$ ).

Ziel des vorliegenden Beitrags ist es, erste Ergebnisse zu den psychometrischen Gütekriterien des deutschsprachigen BDI-II [14] vorzustellen, die an verschiedenen klinischen und nichtklinischen Stichproben am Zentralinstitut für Seelische Gesundheit, Mannheim, gewonnen wurden.

## Methode

### Stichproben

**TS 1: Stationär Depressive (n=13).** Diese Patienten (25–54 Jahre,  $M=43,2$  J [SD 9,9], 69% weiblich) füllten während ihres stationären Aufenthalts in der Psychiatrischen Klinik des Zentralinstituts für Seelische Gesundheit, Mannheim, einmalig das BDI-II aus. Behandlungsdiagnosen nach ICD-10 waren F32 und F33, die Patienten waren zum Zeitpunkt der Untersuchung klinisch depressiv im Ausmaß einer major-depressiven Episode (DSM-IV).

**TS 2: Nicht remittierte Depressive (n=23).** Diese Patienten (18–70 Jahre,  $M=40,7$  J [SD 10,8], 57% weiblich) sind Teilstichprobe einer DFG-Längsschnittstudie<sup>1</sup>, die während ihres stationären Aufenthalts (T<sub>0</sub>) am ZI rekrutiert wurde (ICD-10 Behandlungsdiagnosen: F32, F33 und F34.1) und einen (T<sub>1</sub>) und 6 Monate nach Entlassung (T<sub>2</sub>) nachuntersucht wurden. TS 2 enthält diejenigen Patienten, die zum Zeitpunkt der BDI-II-Bearbeitung (T<sub>0</sub>, T<sub>1</sub> oder T<sub>2</sub>) im Strukturierten

Klinischen Interview für DSM-IV SKID [30] depressiv im Ausmaß einer major-depressiven Episode (DSM-IV) waren.

**TS 3: Entlassene Depressive T1 (n=52).** Diese Patienten (18–70 Jahre,  $M=43,4$  J [SD 12,4], 48% weiblich) repräsentieren die Gesamtstichprobe unipolar depressiver Patienten der DFG-Studie, die zu T<sub>1</sub> den BDI-II ausgefüllt hatten. Etwa 25% waren zu diesem Zeitpunkt teil- oder nicht remittiert (SKID).

**TS 4: Gemeindestichprobe I (n=89).** Diese Stichprobe ist eine gesunde Gemeindestichprobe aus der Mannheimer Bevölkerung (18–70 Jahre,  $M=45,9$  J [SD 13,0], 49% weiblich), die parallel zur DFG-Patientenstichprobe (TS 3) untersucht wurde.

**TS 5: Gemeindestichprobe II (n=118).** Diese unausgelesene Bevölkerungsstichprobe (18–70 Jahre,  $M=46,6$  J [SD 14,8], 61% weiblich) wurde im Rahmen einer genetisch-epidemiologischen Studie einmalig untersucht.<sup>2</sup>

**TS 6: Studenten (n=108).** Diese Stichprobe besteht aus Studenten verschiedener Fachrichtungen der Universität Mannheim (18–30 Jahre,  $M=21,8$  J [SD 2,0], 61% weiblich).

Für bestimmte Teilauswertungen wurden TS 1 und TS 2 zur Gruppe der „akut Depressiven“ (G<sub>1</sub>) zusammengefasst. Pati-

enten aus TS 3, welche zu T<sub>1</sub> im SKID keine oder nur geringe Residualsymptomatik (0–2 Depressionssymptome) aufwiesen, repräsentieren die Gruppe der remittierten Patienten (G<sub>2</sub>), TS 4 bis TS 6 die nichtklinische Stichprobe (G<sub>3</sub>).

## Untersuchungsinstrumente

**Deutsche Version des Beck-Depressionsinventars II [14].** Das BDI-II besteht aus 21 vierstufigen Items mit inhaltlichen Aussagen, die nach aufsteigendem Schweregrad geordnet sind. Beurteilungszeitraum sind die letzten 2 Wochen. Die Bearbeitungsdauer liegt zwischen 5 und 10 min. Durchführungs-, Auswertungs- und Interpretationsobjektivität sind durch die standardisierte Vorgabe und Anweisung zur Bestimmung und Interpretation des Summenwerts gegeben. Die deutschsprachige Übersetzung des Fragebogens erfolgte durch die Autoren dieser Arbeit. Dabei wurden zunächst unabhängige Übersetzungen angefertigt. Die gemeinsam abgestimmte deutsche Fassung wurde dann von einer unabhängigen Person ins Amerikanische zurück übersetzt.

**Weitere Instrumente.** Zur Untersuchung von Validitätsaspekten des BDI-II wurden für verschiedene Stichproben Verfahren eingesetzt, die entweder depressive Symptomatik in der Selbstbeurteilung (Fragebogen zur Depressionsdiagnostik nach DSM-IV FDD-DSM-IV [19], Gesundheitsfragebogen für Patienten PHQ-D [24], WHO-5 Wellbeing Index [4]) oder Fremdbeurteilung (SKID [30], Montgomery-Asperg-Depressionsskala MADRS [26]) erfassen. Zudem wurden Verfahren zu depressionsnahen kognitiven Konstrukten (Response Styles Questionnaire, Skala Rumination RSQ-R [21], Fragebogen zur Erfassung der Handlungs- und Lageorientierung HAKEMP [22], Fragebogen zur Dysfunktionalen und Funktionalen Selbstaufmerksamkeit DFS [15], Selbstwertskala nach Rosenberg SES [11]) sowie der WHO-Fragebogen zur gesundheitsbezogenen Lebensqualität WHO-QOL-BREF [1] eingesetzt. In einer Teilstichprobe (TS 5) kam das NEO-5-Faktoren-Inventar NEO-FFI [9] als Persönlichkeitsinventar zur Anwendung. In TS

<sup>1</sup> DFG-Projekt „Symptombezogene Reaktionsstile und Depression“ (KU1464–1/1,3)

<sup>2</sup> Wir danken Frau Prof. Dr. M. Rietschel, ZI Mannheim, für die Überlassung dieser Daten

5 und TS 6 wurde zusätzlich das Beck-Angstinventar BAI [25] eingesetzt.

## Ergebnisse

**Teststatistiken.** Die Verteilungsparameter des BDI-II-Summenwerts für akut Depressive ( $G_1$ ), remittierte Depressive ( $G_2$ ) und die nichtklinische Stichprobe ( $G_3$ ) sind in **Tab. 1** dargestellt. Die Konfidenzintervalle der jeweiligen Mittelwerte überschneiden sich nicht. Der Mittelwert der akut Depressiven liegt im Bereich „schwere Depression“ ( $\geq 29$ ), der der remittierten Patienten im Bereich „milde Depression“ (14–19) und der der nichtklinischen Stichprobe im Bereich „keine/minimale Depression“ ( $\leq 13$ ).

**Interne Konsistenz.** Die Cronbachs  $\alpha$ -Koeffizienten der internen Konsistenz des BDI-II liegen für sämtliche Teilstichproben im Bereich von  $\alpha \geq 0,84$  [akut Depressive ( $n=36$ ):  $\alpha=0,84$ , entlassene Depressive ( $n=52$ )  $\alpha=0,90$ , nichtklinische Stichprobe ( $n=315$ )  $\alpha=0,89$ ].

**Retestreliaibilität.** Bei 23 Studenten der TS 6 wurde das BDI-II im Abstand von 3 Wochen wiederholt. Die Retestkorrelation betrug  $r=0,78$  ( $p<0,001$ ). In der DFG-Gemeindestichprobe (TS 4,  $n=86$ ) resultierte ein Retestkoeffizient (5 Monate) von  $r=0,78$  ( $p<0,001$ ). Für die Patientenstichprobe (TS 3,  $n=50$ ) betrug dieser Koeffizient  $r=0,47$  ( $p<0,001$ ), hier sind jedoch a priori keine vergleichbar hohen Retestkoeffizienten zu erwarten, da von interindividuell variablen nachstationären Verläufen auszugehen ist.

**Inhaltsvalidität.** Das BDI-II wurde entwickelt, um depressive Symptome nach DSM-IV abzubilden. Alle relevanten Depressionskriterien nach DSM-IV sind nun in Itemform aufgeführt, so dass die inhaltliche Validität als gegeben betrachtet werden kann.

**Konvergente und diskriminante Validität.** In **Tab. 2** sind die Korrelationen des BDI-II mit konstrukt-nahen und konstrukt-fernen Verfahren aufgelistet. Hohe Übereinstimmungen resultierten in allen Stichproben mit der selbstbeurteilten Depressivität anhand des FDD-

## Zusammenfassung · Summary

Nervenarzt 2007 · 78:651–656 DOI 10.1007/s00115-006-2098-7  
© Springer Medizin Verlag 2006

C. Kühner · C. Bürger · F. Keller · M. Hautzinger

### Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II). Befunde aus deutschsprachigen Stichproben

#### Zusammenfassung

**Hintergrund.** Das Beck-Depressionsinventar (BDI) wurde 1996 einer Revision (BDI-II) unterzogen mit dem Ziel, ein Selbstbeurteilungsinstrument zur Beurteilung der Depressionsschwere, orientiert an den Depressionskriterien nach DSM-IV, verfügbar zu haben. Im vorliegenden Beitrag werden psychometrische Gütekriterien zum deutschsprachigen BDI-II berichtet.

**Patienten und Methoden.** Das BDI-II wurde ins Deutsche übersetzt und in Studien an depressiven Patienten und Gesunden eingesetzt.

**Ergebnisse.** Die Inhaltsvalidität des BDI-II wurde durch Anlehnung an DSM-IV verbessert. Die interne Konsistenz der Skala ist zufrieden stellend ( $\alpha \geq 0,84$ ), die Retestreliaibilität liegt bei  $r \geq 0,75$  in nichtklinischen Stich-

proben. Die Zusammenhänge zwischen BDI-II und konstrukt-nahen Skalen sind hoch, solche mit symptom-fernen Persönlichkeitsskalen niedrig. Das BDI-II diskriminiert gut zwischen unterschiedlichen Schweregraden der Depression und ist änderungssensitiv.

**Schlussfolgerung.** Das deutsche BDI-II weist gute psychometrische Kennwerte in klinischen und nichtklinischen Stichproben auf. Die revidierte Fassung des BDI kann nun zur Erfassung der selbstbeurteilten Schwere einer Depression und zur Dokumentation des Verlaufs depressiver Symptomatik unter Behandlung die alte Version ablösen.

#### Schlüsselwörter

Beck-Depressionsinventar II · Depression · Fragebogen · Reliabilität · Validität

### Reliability and validity of the Revised Beck Depression Inventory (BDI-II). Results from German samples

#### Summary

**Background.** The Beck Depression Inventory (BDI) underwent revision in 1996 (BDI-II) with the goal of addressing DSM-IV depression criteria. The present study assessed psychometric properties of the German version of the BDI-II.

**Patients and methods.** The BDI-II was translated into German and evaluated in a series of studies with clinical and nonclinical samples.

**Results.** The content validity of the BDI-II has improved by following DSM-IV symptom criteria. Internal consistency was satisfactorily high ( $\alpha \geq 0,84$ ), and retest reliability exceeded  $r \geq 0,75$  in nonclinical samples. Associations with construct-related scales (de-

pression, dysfunctional cognitive constructs) were high, while those with nonsymptomatic personality assessment (NEO-FFI) were lower. The BDI-II differentiated well between different grades of depression and was sensitive to change.

**Conclusion.** The German BDI-II demonstrates good reliability and validity in clinical and nonclinical samples. It may now replace the older version of the BDI for assessing self-rated severity of depression and course of depressed symptoms under treatment.

#### Keywords

Beck Depression Inventory II · Depression · Questionnaire · Reliability · Validity

**Tab. 2** Korrelationen zwischen BDI-II und ausgewählten Skalen (Mannheimer Stichproben, Messzeitpunkt T1)

Konstrukt	Instrument <sup>a</sup>	Depressive	Gemeinde I	Gemeinde II	Studenten
		(TS 3, n=52)	(TS 4, n=89)	(TS 5, n=118)	(TS 6, n=108)
Selbstbeurteilte Depression	FDD-DSM-IV	0,72***	0,81***	–	0,89***
	PHQ-9	0,74***	–	–	–
Fremdbeurteilte Depression	MADRS	0,68***	0,70***	–	–
Psychisches Wohlbefinden	WHO-5	–0,49***	–0,63***	–0,73***	–0,69***
Selbstbeurteilte Angst	BAI	–	–	0,65***	0,69***
Rumination	RSQ-R	0,52***	0,47***	–	–
Präokkupation	HAKEMP-P	0,19	0,24*	–	–
Zögern	HAKEMP-Z	0,37**	0,21*	–	–
Dysfunktionale Selbstaufmerksamkeit	DFS-D	0,54***	0,48***	–	–
Funktionale Selbstaufmerksamkeit	DFS-F	–0,37**	–0,43***	–	–
Selbstwertgefühl	SES	–0,53***	–0,53***	–0,60***	–0,64***
Lebensqualität physisch	WHOQOL-BREF	–0,57***	–0,59***	–	–0,67***
Lebensqualität psychisch	WHOQOL-BREF	–0,78***	–0,66***	–	–0,66***
Lebensqualität soziale Beziehungen	WHOQOL-BREF	–0,47***	–0,30**	–	–0,36***
Lebensqualität Umwelt	WHOQOL-BREF	–0,53***	–0,53***	–	–0,37***
Lebensqualität insgesamt	WHOQOL-BREF	–0,64***	–0,60***	–	–0,50***
Neurotizismus	NEO-FFI	–	–	0,70***	–
Extraversion	NEO-FFI	–	–	–0,36***	–
Offenheit	NEO-FFI	–	–	0,09	–
Verträglichkeit	NEO-FFI	–	–	–0,25**	–
Gewissenhaftigkeit	NEO-FFI	–	–	–0,25**	–

<sup>a</sup>Information zu den Abkürzungen siehe Text. \* $p < 0,05$ , \*\* $p < 0,01$ , \*\*\* $p < 0,001$ .

DSM-IV (0,72–0,89), ähnlich mit der Depressionsskala PHQ-9 (0,74). Etwas niedrigere Koeffizienten resultierten mit der fremdbeurteilten Depressivität (MADRS,  $r = 0,68$ –0,70). Der WHO-5 Wellbeing Index psychischen Wohlbefindens korreliert mit dem BDI-II im Bereich von  $-0,49 \leq r \leq -0,63$ . Die Korrelation mit dem BAI (TS 5, TS 6) betrug  $r = 0,65$ –0,69.

Die Korrelationen mit konstrukt-nahen kognitiven Skalen liegen im mittleren bis hohen Bereich (■ Tab. 2). Die höchsten Zusammenhänge fanden sich mit den Skalen „Selbstwirksamkeit“ und „Selbstwertgefühl“, insbesondere in den nichtklinischen Stichproben. Im mittleren Bereich liegen die Korrelationen mit Skalen zur funktionalen und dysfunktionalen Selbstaufmerksamkeit (RSQ-R, DFS, HAKEMP). Die Korrelationskoeffizienten von BDI-II und den WHOQOL-BREF-Skalen zur gesundheitsbezogenen Lebensqualität variieren im mittleren bis hohen Bereich, die höchsten Koeffizienten resultierten für die körperliche und die psychische Lebensqualitätsdomäne und die allgemeine Lebensqualität.

Mit Ausnahme der NEO-FFI-Subskala „Neurotizismus“ (TS 5,  $r = 0,70$ ) liegen die Zusammenhangskoeffizienten der weiteren Skalen „Extraversion“, „Offenheit“, „Verträglichkeit“ und „Gewissenhaftigkeit“ mit dem BDI-II-Summenwert im unteren bis mittleren Bereich und damit für die Mehrzahl der Subskalen deutlich niedriger als mit konstruktverwandten Skalen (s. oben).

**Diagnostische Diskriminationsfähigkeit.** Hinweise auf eine gute Differenzierungsfähigkeit des BDI-II im Hinblick auf die Schwere der depressiven Symptomatik zeigte bereits der Vergleich der Verteilungen der Summenwerte der Gruppen G1, G2 und G3 (s. oben). Auch für den nachstationären Zeitraum gelingt eine gute Differenzierung von Gruppen nach Schweregrad. Vergleicht man die remittierten Patienten zu T1 (G2,  $n = 39$ ) mit teil- und nichtremittierten Patienten aus TS 3 (mindestens 3 klinisch relevante SKID-Symptome oder SKID-Diagnose einer majoren Depression zu T1 sind erfüllt,  $n = 13$ ), so resultieren hochsignifikante Differenzen (Vollremittierte:

$M = 15,0$  [ $SD = 8,5$ ], Teil- und Nichtremittierte:  $M = 25,7$  [ $SD = 11,6$ ],  $t(1, 51) = -3,58$ ,  $p < 0,001$ ).

Ähnlich wurde anhand der Verteilung der MADRS-Summenwerte die Stichprobe TS 3 in die Subgruppen leichte (0–2 Punkte,  $n = 18$ ), mittlere (3–10,  $n = 19$ ) und ausgeprägte fremdbeurteilte Depressivität ( $\geq 11$ ,  $n = 15$ ) unterteilt. Eine Varianzanalyse mit dem Faktor Gruppenzugehörigkeit lieferte einen hochsignifikanten Globalwert bezüglich der BDI-II-Summenscores (Gruppe 1:10,1 [ $SD$  7,7]; Gruppe 2:17,1 [ $SD$  6,5]; Gruppe 3:27,4 [ $SD$  9,6];  $F[2,49] = 19,8$ ,  $p < 0,001$ ). Paarweise durchgeführte Post-hoc-Vergleiche (Bonferroni-adjustiert) zeigten für alle Gruppenvergleiche signifikante Unterschiede in den durchschnittlichen Summenwerten des BDI-II ( $ps < 0,03$ ).

**Änderungssensitivität.** In TS 3 ( $n = 50$ ) betrug die Korrelation der Veränderungswerte aus BDI-II und FDD-DSM-IV  $r = 0,73$  ( $p < 0,001$ ), die der Veränderungswerte aus BDI-II und MADRS  $r = 0,69$  ( $p < 0,001$ ).

## Zusammenhänge mit soziodemographischen Merkmalen.

Der BDI-II-Summenwert war weder bei den akut Depressiven [G1 (n=36),  $r=-0,05$ ] noch bei den remittierten Patienten nach Entlassung [G2 (n=39),  $r=-0,20$ ] noch in der nichtklinischen Stichprobe [G3 (n=315),  $r=-0,01$ ] signifikant mit dem Alter assoziiert. Auch wurden keine Geschlechtsunterschiede im BDI-Summenwert identifiziert (alle  $ps>0,20$ ). Ähnliches gilt für das Schulbildungsniveau, das nur für die Teilstichproben TS 3 und TS 4 vorlag (TS 3:  $r=-0,01$ , TS 4:  $r=-0,04$ ).

## Diskussion

Mit der vorliegenden Arbeit liegen erstmals Ergebnisse zu psychometrischen Gütekriterien der deutschen Version des BDI-II [14] vor. Der ursprüngliche Fragebogen [5] wurde im Rahmen der amerikanischen Revision [7] zahlreichen Veränderungen unterworfen.

## Inhaltsvalidität

Hauptintention von Beck [7] war es, die inhaltliche Validität und damit die diagnostische Güte des Verfahrens zu verbessern. Die Aufnahme neuer, DSM-IV-relevanter Items erhöht die diagnostische Sensitivität des Verfahrens, während die Streichung diagnoseirrelevanter Items dessen diagnostische Spezifität verbessert. Nach dem Fragebogen zu Depressionsdiagnostik nach DSM-IV [19] und dem Depressionsmodul des Gesundheitsfragebogens für Patienten [24] handelt es sich beim BDI-II nun um das 3. Selbstbeurteilungsinstrument im deutschen Sprachraum mit dem Anspruch, den Schweregrad depressiver Symptomatik in direkter Anlehnung an die Symptomkriterien einer Major-Depression nach DSM-IV abzubilden. Allerdings sind die einzelnen DSM-IV-Kriterien im BDI-II, im Gegensatz zu den beiden anderen Verfahren [19, 24], nicht gleichwertig berücksichtigt. So geht das depressive Begleitsymptom „Wertlosigkeit/Schuldgefühle“ mit insgesamt 6 Items mit dem höchsten Gewicht in die Berechnung des Summenwerts ein. Damit legt das BDI-II – ähnlich wie seine Vorläuferversion – im Vergleich zu DSM-IV ein starkes Gewicht auf kognitive De-

pressionssymptome und spiegelt damit den kognitionstheoretischen Ansatz von Beck wider. Möglicherweise ist das Verfahren damit besonders sensitiv zur Messung von Effekten kognitiver Therapien, während andere – bezüglich der DSM-IV-Kriterien ausgewogenere – Verfahren besser für Pharmastudien einsetzbar sind oder mit Fremdratings übereinstimmen. Dies wäre jedoch in zukünftigen Studien zu prüfen.

Die Ausweitung des Beurteilungszeitraums auf 2 Wochen dient ähnlich wie die Streichung diagnoseirrelevanter Items dazu, die diagnostische Spezifität des Verfahrens zu erhöhen. Bislang ist allerdings ungeklärt, inwieweit die Validität der Selbsteinschätzung durch eine solche Ausweitung tangiert wird. Auch ist es möglich, dass mit der Ausweitung des Zeitrahmens eine höhere Spezifität mit einer geringeren Sensitivität des Verfahrens erkauft wird [19]. Auch dieser Aspekt könnte anhand geeigneter Studien geprüft werden.

Trotz dieser Einschränkungen lässt sich anhand der berichteten psychometrischen Gütekriterien feststellen, dass mit dem deutschen BDI-II ein Verfahren vorliegt, das gute Reliabilitäts- und Validitätskennwerte aufweist, die mit denen der amerikanischen Originalversion vergleichbar sind. Der Einsatz des Fragebogens in verschiedenen Quer- und Längsschnittstudien an depressiven und nichtklinischen Stichproben im Rahmen der vorliegenden Arbeit ermöglichte es darüber hinaus, Befunde zu psychometrischen Gütekriterien des BDI-II darzustellen, die über die bisher international publizierten Evaluationen des Verfahrens hinausgehen.

## Interne Konsistenz und Retestrelia-bilität

Die internen Konsistenzwerte ( $\alpha \geq 0,84$ ) lassen auf eine ausreichende Homogenität des Verfahrens schließen, die es rechtfertigt, den Summenwert des BDI-II als Maß der Depressionsschwere zu verwenden. Mit Koeffizienten von  $r > 0,75$  (3 Wochen, 5 Monate) in den beiden nichtklinischen Stichproben kann die Retestrelia-bilität als zufrieden stellend bezeichnet werden.

## Konvergente und diskriminante Validität

Das BDI-II weist hohe Übereinstimmung mit anderen Maßen depressiver Selbstbeurteilung auf, die ebenfalls auf DSM-IV-Basis konstruiert sind (FDD-DSM-IV, PHQ-9). Ähnliche Übereinstimmungen mit depressiven Selbstbeurteilungen werden auch in der Literatur berichtet [8, 16, 17]. Nur geringfügig niedriger liegt die Übereinstimmung des BDI-II mit der fremdbeurteilten Depression (MADRS).

Die Übereinstimmung mit selbstbeurteilter Angst in zwei nichtklinischen Stichproben ( $r=0,65-0,69$ ) liegt geringfügig über den in der Literatur berichteten Koeffizienten [7, 27] und sprechen für eine eher schlechte Differenzierbarkeit der Konstrukte Angst und Depression, zumindest auf Selbstbeurteilungsebene. Für die depressionsnahen kognitiven Skalen resultierten erwartungsgemäß signifikante, dem Betrag nach mittlere Korrelationen (vgl. [15]). Im Bereich subjektiver Lebensqualität fanden sich die engsten Zusammenhänge mit den WHO-QOL-BREF-Skalen „psychische Gesundheit“ und „körperliche Gesundheit“. Hier sind partielle Konstruktüberlappungen zu berücksichtigen, die durch ähnliche Items bedingt sind (z. B. Fragen zu Schlaf, Energie und Konzentration). Auch die hohen Korrelationen mit der allgemeinen Lebensqualität passen zu Literaturbefunden, die zeigen, dass globale Ratings zur Lebensqualität hauptsächlich die Schwere der Depressivität widerspiegeln [20].

Das BDI-II wies niedrige bis mäßige Korrelationen mit den Subskalen Extraversion, Offenheit, Verträglichkeit und Gewissenhaftigkeit des NEO-FFI auf. Dies kann als Hinweis auf die diskriminante Validität des BDI-II gewertet werden, da der NEO-FFI mit Ausnahme der Subskala Neurotizismus relativ symptomferne Persönlichkeitsdimensionen erfasst. Die Skala Neurotizismus korreliert aufgrund inhaltlicher Überlappungen dagegen sehr hoch mit dem BDI-II, sie erfasst u. a. emotionale und kognitive Aspekte von Depression [9].

## Diagnostische Diskriminationsfähigkeit und Änderungssensitivität

Der Summenwert des BDI-II differenzierte gut zwischen depressiven Patienten und Gesunden und zwischen verschiedenen Schweregraden depressiver Symptomatik. Hohe Korrelationen zwischen den Änderungswerten des BDI-II mit denen anderer Depressionskalen bescheinigen dem Verfahren auch eine gute Änderungssensitivität.

### Fazit für die Praxis

Mit den vorliegenden Ergebnissen zur psychometrischen Güte des BDI-II kann dieses unseres Erachtens nun für den Einsatz im deutschen Sprachraum empfohlen werden. Das BDI-II wird auch zukünftig international eine führende Rolle als Selbstrating depressiver Beschwerden spielen und breite Anwendung in der Therapie- und Verlaufsforschung sowie in der klinischen Praxis finden. Durch die Publikation der deutschen Version des Fragebogens [14] werden nun auch direkte Vergleiche mit entsprechenden internationalen Studien möglich sein. Wichtig erscheint nun, weitere Studien zum BDI-II in einem breiten Anwendungsfeld durchzuführen. Forschungsbedarf besteht insbesondere für die Einbeziehung von Patienten mit anderen psychischen und somatischen Primärerkrankungen bzw. komorbiden Störungen und für Untersuchungen zu Kennwerten des BDI-II unter anderen Rahmenbedingungen, wie im primären Versorgungsbereich und in der Rehabilitation.

### Korrespondierender Autor

PD Dr. C. Kühner

AG Verlaufs- und Interventionsforschung, Zentralinstitut für Seelische Gesundheit  
68072 Mannheim  
kuehner@zi-mannheim.de

**Interessenkonflikt.** Es besteht kein Interessenkonflikt. Der korrespondierende Autor versichert, dass keine Verbindungen mit einer Firma, deren Produkt in dem Artikel genannt ist, oder einer Firma, die ein Konkurrenzprodukt vertreibt, bestehen. Die Präsentation des Themas ist unabhängig und die Darstellung der Inhalte produktneutral.

## Literatur

- Angermeyer MC, Kilian R, Matschinger H (2000) WHOQOL-100 und WHOQOL-BREF. Handbuch. Hogrefe, Göttingen
- Arnau RC, Meagher MW, Norris MP et al. (2001) Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychol* 20:112–119
- Ball R, Steer RA (2003) Mean Beck Depression Inventory-II scores of outpatients with dysthymic or recurrent-episode major depressive disorders. *Psychol Rep* 93:507–512
- Bech P (2004) Measuring the dimensions of psychological general well-being by the WHO-5. *QoL Newsletter* 32:15–16
- Beck AT, Ward CH, Mendelson M et al. (1961) An inventory for measuring depression. *Arch Gen Psychiatry* 4:561–571
- Beck AT, Steer RA, Garbin MG (1988) Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev* 8:77–100
- Beck AT, Steer RA, Brown GK (1996) Beck Depression Inventory, 2nd edn. Manual. The Psychological Corporation, San Antonio
- Bonilla J, Bernal G, Santos A et al. (2004) A revised Spanish version of the Beck Depression Inventory: psychometric properties with a Puer Rican sample of college students. *J Clin Psychol* 60:119–130
- Borkenau P, Ostendorf F (1993) NEO-Fünf-Faktoren-Inventar. Hogrefe, Göttingen
- Dozois DJA, Dobson KS, Ahnberg JL (1998) A psychometric evaluation of the Beck Depression Inventory-II. *Psychol Assess* 10:83–89
- Ferring D, Filipp SH (1996) Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala. *Diagnostica* 42:284–292
- Hautzinger M (1991) Das Beck-Depressioninventar (BDI) in der Klinik. *Nervenarzt* 62:689–696
- Hautzinger M, Bailer M, Worall H et al. (1994) Beck-Depressions-Inventar (BDI). Bearbeitung der deutschen Ausgabe. Testhandbuch. Huber, Bern
- Hautzinger M, Keller F, Kühner C et al. (2006) Das Beck-Depressionsinventar II. Deutsche Bearbeitung und Handbuch zum BDI II. Harcourt Test Services, Frankfurt a. M.
- Hoyer J (2000) Der Fragebogen zur Dysfunktionalen und Funktionalen Selbstaufmerksamkeit (DFS). *Diagnostica* 46:140–148
- Kojima M, Furukawa TA, Takahashi H et al. (2002) Cross-cultural validation of the Beck Depression Inventory-II in Japan. *Psychiatry Res* 110:291–299
- Krefetz DG, Steer RA, Gulab NA et al. (2002) Convergent validity of the Beck Depression Inventory-II with the Reynolds Adolescent Depression Scale in psychiatric inpatients. *J Pers Assess* 78:451–460
- Krefetz DG, Steer RA, Kumar G (2003) Lack of age differences in the Beck Depression Inventory-II scores of clinically depressed adolescent outpatients. *Psychol Rep* 92:489–497
- Kühner C (1997) Fragebogen zur Depressionsdiagnostik nach DSM-IV (FDD-DSM-IV). Hogrefe, Göttingen
- Kühner C (2002) Subjective quality of life: validity issues with depressed patients. *Acta Psychiatr Scand* 106:62–70
- Kühner C, Weber I (1999) Responses to depressed mood in depressed patients: an investigation of Nolen-Hoeksema's Response Styles Theory. *Psychol Med* 29:1323–1333
- Kuhl J (1994) Action versus state orientation: psychometric properties of the Action Control Scale (ASC-90). In: Kuhl J, Beckmann J (eds) Volition and personality: action versus state orientation. Hogrefe, Seattle, pp 47–60
- Kumar G, Steer RA, Teitelman KB et al. (2002) Effectiveness of Beck Depression Inventory-II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Assessment* 9:164–170
- Löwe B, Spitzer RL, Zipfel S et al. (2001). Gesundheitsfragebogen für Patienten (PHQ-D). Komplettversion und Kurzform. Pfizer, Karlsruhe
- Margraf J, Ehlers A (2005) Beck Angst Inventar. Hogrefe, Göttingen
- Montgomery SA, Asberg M (1979) A new depression rating scale designed to be sensitive to change. *Br J Psychiatry* 134:382–389
- Steer RA, Clark DA (1997) Psychometric characteristics of the Beck Depression Inventory-II with college students. *Meas Eval Counsel Dev* 30:128–136
- Steer RA, Kumar G, Ranieri WF et al. (1998) Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *J Psychopathol Behav Assess* 20:127–137
- Steer RA, Brown GK, Beck AT et al. (2001) Mean Beck Depression Inventory-II scores by severity of major depressive episode. *Psychol Rep* 88:1075–1076
- Wittchen HU, Wunderlich U, Gruschwitz S et al. (1997) SKID-I. Strukturiertes klinisches Interview für DSM-IV. Achse I: Psychische Störungen. Hogrefe, Göttingen