

Fremdbeurteilungs-Skalen und Selbstbeurteilungs-Skalen messen unterschiedliche Konzepte der Depressivität

Aus einer großen Multicenterstudie an stationär behandelten depressiven Patienten

Hans-Jürgen Möller*, München, und Florian Seemüller, Garmisch-Partenkirchen

Im Rahmen einer großen Multicenterstudie an stationär behandelten depressiven Patienten (n = 1014) wurde die Übereinstimmung der mit der Hamilton-Depressions-skala (HAMD), der Montgomery-Åsberg-Depressionsskala (MADRS) und dem Beck-Depressions-Inventar (BDI) gemessenen Depressivitätswerte korrelationsstatistisch und faktorenanalytisch untersucht. Es zeigte sich eine nur mäßige Korrelation der Selbstbeurteilungswerte mit den Fremdbeurteilungswerten. In der explorativen Faktorenanalyse stellten sich drei Faktoren dar, von denen einer ausschließlich durch BDI-Items gebildet wurde. Die Selbstbeurteilung mit dem BDI misst offensichtlich ein anderes Depressionskonzept als die Fremdbeurteilung. Das sollte bei der Planung von Studiendesigns und im klinischen Alltag berücksichtigt werden.

Schlüsselwörter: Fremdbeurteilungs-Skalen, Selbstbeurteilungs-Skalen, HAMD, MADRS, BDI

Psychopharmakotherapie 2024;31:9–14.



Observer-rating scales and self-rating scales measure different concepts of depression: Results of HAMD, MADRS and BDI in a huge multicenter study on depressive in-patients

The concordance between the depression-scores of HAMD, MADRS and BDI were compared using correlation statistics and explorative factor-analysis in a huge multicenter-study on depressive inpatients (N = 1014). The correlation between self-rating and observer-ratings were only modest. The factor-analysis resulted in three factors, one of which was only composed by BDI-items. Self-ratings with the BDI apparently measure concept of depression different from observer-ratings. This should be taken into account when planning research studies or in clinical practice.

Key words: Observer-rating scales, self-rating scales, HAMD, MADRS, BDI

Standardisierte Beurteilungsskalen werden in der Psychiatrie angewendet, um psychopathologische Dimensionen wie das depressive Syndrom oder das ängstliche Syndrom in valider und reliabler Weise zu beurteilen/messen. Neben den vorrangig verwendeten Fremdbeurteilungs-Skalen stehen Selbstbeurteilungs-Skalen zur Verfügung, die wegen ihrer Praktikabilität und Zeitökonomie häufig eingesetzt werden, oft anstelle von Fremdbeurteilungs-Skalen oder im Sinne von die Fremdbeurteilung ergänzenden, häufigeren Messungen. Beide Beurteilungsverfahren haben ihre jeweils eigenen Vor- und Nachteile, die hier nicht weiter erörtert werden können. Es sei auf die diesbezügliche Darstellung von Möller 2009 [14] verwiesen.

In der deutschen Psychiatrie werden im Rahmen von Studien oder in der täglichen Praxis zur Fremdbeurteilung des depressiven Syndroms meistens die Hamilton-Skala (HAMD) oder die Montgomery-Åsberg-Depressionsskala (MADRS) eingesetzt, zur Selbstbeurteilung des depressiven Syndroms das Beck-Depressions-Inventar (BDI). Viele Nutzer gehen davon aus, dass die so erhaltenen Fremdbeurteilungs- oder Selbstbeurteilungswerte eine hohe Übereinstimmung aufweisen und mehr oder weniger austauschbar sind. Das trifft aber offenbar nicht zu, weder auf der Ebene einfacher Korrelationsstatistik noch auf der Ebene der faktoranalytischen Dimensions-Analyse [3, 6]. Dies sei im Folgenden am Datensatz einer großen naturalistischen Depressionsstudie dargestellt.

* Die Arbeit ist meinem verehrten akademischen Lehrer, Herrn Prof. Dr. Detlev von Zerssen (30.10.1926–14.08.2023), gewidmet.

Prof. Dr. med. Hans-Jürgen Möller, Klinik für Psychiatrie und Psychotherapie der Ludwig Maximilians-Universität München, Nussbaumstraße 7, 80336 München, E-Mail: Hans-Juergen.Moeller@med.uni-muenchen.de

Priv.-Doz. Dr. med. Florian Seemüller, kbo-Lech-Mangfall-Klinik Garmisch-Partenkirchen

Stichprobe und Methodik

Die HAMD-, MADRS- und BDI-Daten einer großen, prospektiven, naturalistischen Multicenterstudie [19] an stationär behandelten, depressiven Patienten (n = 1014) wurden ex post hinsichtlich der Übereinstimmung von Fremd- und Selbstbeurteilung der depressiven Symptomatik analysiert. Der psychopathologische Zustand der Patienten wurde alle zwei Wochen während des stationären Aufenthalts und in den folgenden vier Jahren bei jährlichen Nachuntersuchungen mithilfe dieser standardisierten Untersuchungsmethoden beurteilt.

Die naturalistische Studie war Teil der Forschungsaktivitäten des Deutschen Depressionsforschungs-Netzwerks, das vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wurde. An der Studie nahmen Patienten von 12 Universitätskliniken und nicht universitären Kliniken teil. Die Details zur Patientenstichprobe und zum Studiendesign finden sich in der Darstellung der Hauptstudie [19].

Die HAMD hat in der hier verwendeten Version (HAMD-17) 17 Items, die MADRS zehn Items, das BDI 21 Items. Die Items der MADRS und des BDI werden jeweils auf einer Skala von 0 bis 4 bewertet. Das gilt auch für neun Items der HAMD; die restlichen Items der HAMD werden auf einer Skala von 0 bis 2 beurteilt. Alle Fremdbeurteilungen wurden von erfahrenen Ärzten mit einem Minimum von vier Jahren Psychiatrie-Weiterbildung und Erfahrung in der Anwendung der Fremdbeurteilungsskalen durchgeführt. Alle Untersuchungen des jeweiligen Patienten wurden vom selben Untersucher durchgeführt.

Üblichen Standards für Analysen zur Skalen-Psychometrie folgend, wurde nur ein Zeitpunkt pro Patient einbezogen. Dabei wurde bedacht, dass Ratings bei Studienbeginn beziehungsweise bei stationärer Aufnahme geprägt sind durch höhere Schweregrade und geringere Variabilität, während Ratings am Studienende bzw. bei Entlassung geringere Schweregrade und höhere Variabilität zeigen. Um dadurch bedingte Bias zu vermeiden und um sicherzugehen, dass jeder Patient mit dem gleichen Gewicht in die Studie geht, wurde die von Uher et al. [23] vorgeschlagene Methode angewandt. Diese Methode basiert auf einem „random week dataset“, wobei nur ein Rating-Zeitpunkt nach Zufallsauswahl für jeden Patienten ausgewählt wird.

Für die kombinierte Analyse der Fremdbeurteilungsskalen (HAMD und MADRS) und der Selbstbeurteilungsskala BDI wurde eine explorative Hauptkomponenten-Faktorenanalyse unter Verwendung einer Pearson-Korrelationsmatrix durchgeführt. Zur Bestimmung einer sinnvollen Zahl von Faktoren wurden „parallel plots“ (parallele Koordinaten) genutzt, da diese Methode sich gegenüber anderen Methoden, zum Beispiel der verbreitet genutzten „eigenvalue-greater-than-one“-Regel (K1 Rule), als überlegen gezeigt hat. Um die Parallelanalysen durchzuführen, wurden die Eigenwerte des originalen Datensatzes mit den Eigenwerten von 500 Zufallspermutatio-

nen („random permutations“) der Daten verglichen. Eigenwerte größer als die der „random permutations“ legen die Existenz einer „internal data structure“ nahe und somit einen interpretierbaren Faktor. Es wurde eine oblique Promax-Rotation durchgeführt, weil zu erwarten ist, dass Dimensionen der Depressivität miteinander korrelieren. Um eine klar durchschaubare Darstellung der Daten zu gewährleisten, wurden nur Lösungen mit einem Absolutwert größer als 0,4 dargestellt. Alle statistischen Analysen wurden mit der Statistik-Software R, Version 4.0.4, durchgeführt.

Ergebnisse

Von 1014 Patienten lagen komplette HAMD-Ratings vor, von 919 Patienten komplette MADRS-Ratings und von 990 komplette BDI-Ratings. Daraus ergaben sich für die Analyse Daten von 3690 Visiten von 755 Patienten mit komplettem Datensatz in allen drei Skalen. Die mittlere Anzahl der stationären Behandlungstage war $53,6 \pm 47,5$ Tage. Die Patienten waren $45,5 \pm 11,9$ Jahre alt. 62 % der Patienten waren weiblich.

Die Parallelanalyse für die Kombination der drei Skalen legte eine 3- bis 5-Faktor-Struktur nahe. Jedoch nahm die Differenz/der Unterschied zwischen simulierten und beobachteten Eigenwerten vom vierten zum sechsten Faktor ab. Um eine interpretierbare Beschreibung zu erhalten und zu viele „cross-loadings“ zu vermeiden, wurde eine Begrenzung auf drei Faktoren festgelegt.

Mit Faktoranalysen kann abgeschätzt werden, welche Symptomdimensionen den untersuchten Ratingskalen zugrunde liegen. Zur Untersuchung der Frage, ob HAMD, MADRS und BDI ein einziges zugrunde liegendes Konstrukt der Depression/Depressivität erfassen, wurde eine explorative Faktorenanalyse aller 48 Items der drei Skalen durchgeführt (Tab. 1). Sie ergab einen Faktor, den man mit dem Oberbegriff „Stimmung und Angst“ bezeichnen könnte. Er erklärt 14 % der Varianz. Er enthält unter anderem die Items beobachtete und berichtete Stimmung, genitale und somatische Symptome sowie angstbezogene Items aus HAMD oder MADRS. Der zweite Faktor besteht ausschließlich aus allen depressionstypischen BDI-Items. Er erklärt 16 % der Varianz. Der dritte Faktor, hier als „neurovegetativer Faktor“ bezeichnet, enthält die Items, die Schlafstörungen, Appetit und Gewichtsveränderungen betreffen, aus allen drei Skalen und zwei Suizid-Items (mit schwacher Ladung unterhalb des Kriteriums 0,4) aus den beiden Fremdbeurteilungsskalen. Er erklärt 10 % der Varianz. Insgesamt erklären die drei Faktoren 40 % der Varianz.

Die Korrelation zwischen den beiden Fremdbeurteilungsskalen HAMD und MADRS war mit 0,88 – wie erwartet – sehr hoch. Die Korrelation dieser beiden Fremdbeurteilungsskalen mit der Selbstbeurteilung durch den BDI hingegen war nur mäßig: 0,58 bzw. 0,59.

Tab. 1. Kombinierte explorative Faktoranalyse von MADRS, BDI and HAMD: Einzelitem-Ladungen über 0,4 in einer 3-Faktor-Lösung, erklärte Varianz und interne Konsistenz (Cronbachs alpha)

Item	Faktor 1	Faktor 2	Faktor 3
MADRS 1-Sichtbare Traurigkeit	-0,64		
MADRS 2-Berichtete Traurigkeit	-0,63		
MADRS 3-Innere Spannung	-0,63		
MADRS 4-Schlaflosigkeit			0,73
MADRS 5-Appetitverlust			0,74
MADRS 6-Konzentrationsschwierigkeiten	-0,73		
MADRS 7-Untätigkeit	-0,77		
MADRS 8-Gefühllosigkeit	-0,72		
MADRS 9-Pessimistische Gedanken	-0,68		
MADRS 10-Suizidgedanken			
HAMD 1-Depressive Stimmung	-0,55		
HAMD 2-Schuldgefühle	-0,51		
HAMD 3-Suizid/Suizidgedanken			
HAMD 4-Einschlafstörung			0,60
HAMD 5-Durchschlafstörung			0,66
HAMD 6-Schlafstörungen am Morgen			0,65
HAMD 7-Arbeit und sonstige Tätigkeiten	-0,68		
HAMD 8-Depressive Hemmung	-0,68		
HAMD 9-Erregung	-0,44		
HAMD 10-Angst – psychisch	-0,50		
HAMD 11-Angst – somatisch	-0,51		
HAMD 12-Körperliche Symptome – gastrointestinal			0,60
HAMD 13-Körperliche Symptome – allgemein	-0,58		
HAMD 14-Genitalsymptome	-0,42		
HAMD 15-Hypochondrie	-0,46		
HAMD 16-Gewichtsverlust			0,73
HAMD 17-Krankheitseinsicht			
BDI A-Traurigkeit		-0,63	
BDI B-Zukunftspessimismus		-0,70	
BDI C-Versagensgefühle		-0,79	
BDI D-Freudlosigkeit		-0,70	
BDI E-Schuldgefühl		-0,74	
BDI F-Gefühl gestraft zu werden		-0,59	
BDI G-Enttäuschungsgefühl		-0,78	
BDI H-Selbstvorwürfe		-0,76	
BDI I-Suizidgedanken		-0,67	
BDI J-Weinen		-0,48	
BDI K-Reizbarkeit		-0,53	
BDI L-mangelndes Interesse an Menschen		-0,66	
BDI M-Entschlussmangel		-0,67	
BDI N-Gefühl schlecht auszusehen		-0,58	
BDI O-mangelnder Arbeitsantrieb		-0,62	
BDI P-Schlafstörungen			0,51
BDI Q-Müdigkeit		-0,55	
BDI R-Appetitmangel			0,53
BDI S-Gewichtsverlust			0,65
BDI T-Gesundheitssorgen			
BDI U-mangelndes sexuelles Interesse		-0,47	
SS loadings	6,68	7,81	4,79
Erklärte Varianz	0,14	0,16	0,10
Kumulativ erklärte Varianz	0,14	0,30	0,40
Cronbachs alpha	0,91	0,91	0,84

Diskussion

Eine 3-Faktoren-Lösung der Kombination der drei Skalen erscheint als angemessen und am besten interpretierbar. Den ersten Faktor kann man mit dem Oberbegriff „Stimmung und Angst“ bezeichnen. Er enthält Items aus HAMD oder MADRS. Der Zweite enthält alle depressionstypischen BDI-Items. Der dritte Faktor, hier als „psychovegetativer Faktor“ bezeichnet, enthält Items, die Schlafstörungen, Appetit und Gewichtsveränderungen betreffen, aus allen drei Skalen und zwei Suizid-Items (mit schwacher Ladung) aus den beiden Fremdbeurteilungsskalen. Diese Items bezeichnen Symptome, die nicht für die Depression spezifisch sind, sondern lediglich sensitive Charakteristika einer Depression im Rahmen eines korrekt diagnostizierten Samples Depressiver sind.

Alle depressionstypischen BDI-Items laden in einem gewichtigen Faktor, der 16% der Gesamtvarianz aller drei Skalen erklärt. Dieses Ergebnis unterstützt die Auffassung, dass die Selbstbeurteilung von Symptomen der Depression eine andere Krankheitsdimension erfasst als die Fremdbeurteilung. Bedenkt man, dass HAMD und MADRS eine 50%ige Überlappung der Symptome haben, so ist diese Trennung in zwei separate Faktoren bemerkenswert. Andererseits ist auch bemerkenswert, dass nicht alle Items des BDI auf den Faktor 2 laden, sondern einige wenige auch auf den Faktor 3, in dem es so zur Vermischung von Items aus Selbst- und Fremdbeurteilung kommt.

Ein deutlicher Anhaltspunkt für die nur bescheidene Übereinstimmung von Fremd- und Selbstbeurteilung der Depression ergibt sich auch aus den Korrelationskoeffizienten. Während die Gesamtscores der beiden Fremdbeurteilungsskalen HAMD und MADRS eine Korrelation von 0,88 zeigen, bestehen zwischen HAMD bzw. MADRS und BDI nur Korrelationswerte von 0,58 bzw. 0,59. Dies entspricht in der Größenordnung Werten, wie sie auch in der Literatur beschrieben wurden, zum Beispiel im Spektrum von relativ niedrigen Werten zwischen 0,30 und 0,40 bei Coles et al. 2001 [6] und etwas höher bei Beck et al. 1988 [3].

Interessant ist der Vergleich mit dem Ergebnis der Studie von Uher et al. 2008 [23], die ebenfalls eine explorative Faktorenanalyse der drei Skalen durchführten. Es ergab sich ebenfalls eine 3-Faktoren-Lösung. Bei dieser fand sich ein gewichtiger Selbstbeurteilungsfaktor mit fast allen BDI-Items plus Suizid-Item aus HAMD und MADRS und dem Item Schuld aus HAMD. Außerdem stellte sich ein Faktor Stimmung und Angst aus HAMD und MADRS dar sowie ein neurovegetativer Faktor mit Schlaf- und Appetit-Items aus allen drei Skalen.

Auf den Unterschied der mit Fremd- oder Selbstbeurteilung erfassten Dimensionen der Depression wurde auch schon in früheren Arbeiten hingewiesen. Insbesondere die diesbezüglichen Publikationen von von Zerssen sind hier zu erwähnen [24–26], der obendrein mit seinen psychometrischen Untersuchungen darstellte, dass die Fremdbeurteilung psychopatho-

logischer Symptome eine erheblich größere Differenzierung psychopathologischer Phänomene ermöglicht als die Selbstbeurteilung.

Es gibt verschiedene Gründe für die Unterschiede zwischen Selbst- und Fremdbeurteilung [14]:

- Die Selbstbeurteilung ist stärker Bias-abhängig; zum Beispiel neigen schwer depressive Patienten zur Unterschätzung ihrer Symptome, während leicht depressive eher zur Überschätzung ihrer Symptome neigen [11, 12].
- Einige Aspekte der Depression können nicht adäquat durch Selbstbeurteilung festgestellt werden, da sie vorrangig durch einen Beobachter erkannt werden. Das gilt beispielsweise für Symptome wie mimische Verarmung, psychomotorische Verlangsamung [15] oder Hypochondrie.
- Selbstbeurteilung kann beeinflusst sein durch bestimmte Antwort-Bias, wie etwa Antworten im Sinne sozialer Erwünschtheit oder Antworten im Sinne der Übertreibung zum Beispiel in der Hoffnung, mehr Hilfe zu bekommen [13].
- Die adäquate Selbstbeurteilung ist abhängig vom Bildungsgrad und der Fähigkeit zur Introspektion [7].

Berücksichtigend, dass auch die Fremdbeurteilung nicht frei von Bias ist – sie kann beispielsweise stark beeinflusst werden durch die Erwartungshaltung des Kliniklers in Bezug auf die Wirksamkeit einer Therapie –, kann bzw. sollte die Selbstbeurteilung als eine eigene sinnvolle Beurteilungsdimension angesehen werden [8, 14].

Hinsichtlich der Frage der Übereinstimmung von Fremd- und Selbstbeurteilung darf der Aspekt der Skalenauswahl und der mit den Skalen erfassten Items nicht unberücksichtigt bleiben. Während HAMD und MADRS die klassischen Symptome der Depression erfassen, wurde der BDI im Rahmen der kognitiven Depressionstheorie entwickelt [1]. Der Kern dieser Theorie ist, dass Depression sich aus negativen Gedanken über sich selbst, die Welt und die Zukunft entwickelt. Daraus folgend fokussiert der BDI unter anderem sehr stark auf kognitive Items, die sich auf negative Selbstwahrnehmungen beziehen. Eine neuere Version, der BDI-II, wurde unter anderem an Depressionskriterien der DSM-IV-Diagnose Depression angepasst [2, 9, 10] und erreicht gegebenenfalls durch diese Annäherung an klinische Depressionssymptomatik eine höhere Übereinstimmung mit der MADRS: 0,68 bis 0,70.

Bei den erwähnten Untersuchungen von Möller und von Zerssen zur Übereinstimmung von Fremd- und Selbstbeurteilung wurden die von von Zerssen entwickelten Selbstbeurteilungsskalen [26] eingesetzt. Deren Konstruktion ging davon aus, dass ein Patient nur wenige Dimensionen der Psychopathologie selbst erfassen kann, weshalb Selbstbeurteilungsskalen entsprechend einfach gestaltet sein müssen, auf dem Patienten unbekannte Symptom-Termini verzichten müssen (da er diese nicht versteht oder sogar missversteht) und in einer einfachen Alltagssprache formuliert sein müssen. So beschränkte sich

von Zerssen in der Konstruktion der Skalen weitgehend auf Adjektiv- und Beschwerden-Listen.

Auch andere Untersuchungen zur Übereinstimmung von Selbstbeurteilung und Fremdbeurteilung der Depression, die auf separat entwickelte Skalen zurückgreifen, zeigten die nur bescheidene Korrelation zwischen Selbst- und Fremdbeurteilung, wie etwa die Studie von Carroll et al. 1973 [4], die für die Selfrating Depression Scale von Zung [28] und die HAMD nur eine Korrelation von 0,41 fand.

Ein ganz anderer Weg wurde in den letzten drei Dekaden versucht: die Entwicklung von Selbstbeurteilungs-Skalen, die eine direkte sprachliche Umsetzung einer Fremdbeurteilungs-Skala sind. Dies wurde von Carroll et al 1981 [5] für die von ihm entwickelte Selbstbeurteilungs-Version der HAMD mit der üblichen Fremdbeurteilungs-Version der HAMD erprobt. Es ergab sich eine Korrelation von 0,80, die aber in einer Replikationsstudie von Robbins et al 1985 [16] auf 0,46 schrumpfte. Trotz anfänglichen Interesses fand dieser Ansatz im weiteren Verlauf, selbst in den USA, keine dauernde Resonanz, genauso wenig wie eine Selbstbeurteilungs-Version der MADRS [27], und wurde durch neuere Ansätze von Rush ersetzt. Er entwickelte auf der Basis von DSM-IV-Symptomkriterien der Depression und weiteren klinischen Depressionssymptomen das Inventory of Depressive Symptomatology (IDS) [18] und daraus – unter Weglassung der ergänzenden klinischen Depressionssymptome – das als gekürzte Version abgeleitete 16-Item Quick Inventory of Depressive Symptomatology (QIDS). Beide Skalen wurden von Anfang an neben den Fremdbeurteilungs-Versionen (IDS-C und QIDS-C) auch als sprachlich analoge Selbstbeurteilungs-Versionen (IDS-SR und QIDS-SR) vorgelegt. Sowohl für die originären Langversionen als auch für die Kurzversionen wurden sehr gute Korrelationen von über 0,80 zwischen der jeweils analogen Selbst- und Fremdbeurteilungsskala gefunden, ebenso zwischen den beiden Selbstbeurteilungsskalen und der HAMD-17 sowie der HAMD-24 [17]. Diese hohen Werte könnten unter anderem damit zusammenhängen, dass die Korrelationen nur bei Therapieende berechnet wurden.

Angesichts dieser engen Korrelation ist es umso erstaunlicher, dass sich bei kategorialer Auswertung im Sinne der Definition von Remission und Response zwischen Selbst- und Fremdbeurteilung erhebliche Unterschiede zeigten. Während sich bei der Festlegung von Remission oder Response durch die Selbstbeurteilungsskalen IDS-SR und QIDS-SR recht ähnliche Ergebnisse zeigten, ergaben sich erhebliche Diskrepanzen (mehr als 20%) zwischen den jeweiligen Response- und Remitter-Definitionen der beiden Selbstbeurteilungsskalen und denen der HAMD (in der Studie HAMD-24) [17].

In der auf die STAR-D-Studie bezogenen Auswertung von Trivedi et al. 2006 [22] waren die diesbezüglichen Differenzen nicht so extrem, betrug aber zwischen den QIDS- und HAMD-24-definierten Remittern am Ende der ersten Behand-

lungsphase immerhin 5% (32% bei Selbstbeurteilung, 27% bei Fremdbeurteilung) – eine Differenz, die für ein Therapieeffizienz-Kriterium durchaus relevant ist. Im Gegensatz zu der in der Publikation von 2003 zum Ausdruck gebrachten Auffassung der Autoren scheint eine solche Diskrepanz sowie eine solche Überschätzung des Therapieergebnisses durch Selbstbeurteilung zumindest für Therapiestudien nicht hinnehmbar. Bei der Interpretation der Ergebnisse unserer Studie müssen unter anderem folgende Einschränkungen berücksichtigt werden:

- Es handelt sich um eine Stichprobe stationär behandelter depressiver Patienten im Schweregradbereich von mittel bis schwer.
- Wegen Missing-Data-Problemen konnten nicht alle Patienten in die Auswertung einbezogen werden.
- Die Fremdbeurteilungsskalen wurden pro Patient jeweils vom selben Arzt bearbeitet, der auch die Exploration durchführte. Eine alternative Strategie mit jeweils einem weiteren Untersucher wäre zu aufwendig gewesen und hätte obendrein die Varianz erhöht.
- Die verwendeten Skalen bezogen nicht die Symptomatik der atypischen Depression mit ein, die spezifischere Skalen erfordert [21].

Weitergehende, differenziertere statistische Analysen zur selben Stichprobe mit skalenbezogenen Fragestellungen, unter anderem mit konfirmatorischen Faktorenanalysen zu HAMD, MADRS und BDI, wurden in einer umfangreichen Publikation von Seemüller et al. 2023 [20] vorgelegt.

Schlussfolgerungen

Die Beurteilung depressiver Symptomatik sollte nicht allein auf der Basis des BDI erfolgen, da die so erfassten Dimensionen der Depression nicht ausreichend kongruent sind mit den durch Fremdbeurteilung (HAMD, MADRS) erfassten Dimensionen der Depression. Die Selbstbeurteilung mit dem BDI kann aber die mit der Fremdbeurteilung gesammelten Information über die depressive Symptomatik sinnvoll ergänzen und gibt durch den Fokus auf die Selbstbeurteilung eine komplementäre Sicht. Die revidierte Version BDI-II bringt möglicherweise Vorteile, was aber weiter überprüft werden muss.

Im Therapieverlauf ist eine kombinierte Form von Selbst- und Fremdbeurteilung in zeitlicher Versetzung möglicherweise aus pragmatischen Gründen sinnvoll, wenn man nicht so häufig Fremdbeurteilungen durchführen will: zum Beispiel bei Beginn der Behandlung MADRS und BDI, dann über ein bis zwei Wochen nur BDI, dann wieder MADRS und BDI etc. Alternativ zum BDI können in einem solchen Messkonzept zum Beispiel auch die Befindlichkeitsskala von von Zerssen [26] oder eine visuelle Analogskala für die repetitive Messung eingesetzt werden.

Die für das BDI dargestellte beschränkte Kongruenz mit HAMD und MADRS gilt auch für andere Selbstbeurteilungs-

Skalen, zum Beispiel für die in Deutschland über lange Zeit sehr häufig angewandte Befindlichkeitsskala von von Zerssen [26]. Für das im US-amerikanischen Sprachraum entwickelte Inventory of Depressive Symptoms (IDS) und seine Kurzform QIDS wurden, möglicherweise bedingt durch die analoge Item-Formulierung, bessere Kongruenzen mit deren Selbstbeurteilungs-Versionen (IDS-SR bzw. QIDS) in angloamerikanischen Studien erreicht. Das bedarf aber in Hinblick auf den Transfer in die deutsche Psychiatrie weiterer Überprüfung und Erprobung in deutschsprachigen Ländern.

Interessenkonflikterklärung

Die Autoren geben an, dass keine Interessenskonflikte vorliegen. Die Originalstudie wurde vom BMBF gefördert.

Am Deutschen Forschungsnetzwerk Depression beteiligte Untersucher und Zentren

Psychiatrische Universitätsklinik Berlin Charité, Campus Mitte (Andreas Heinz, Mazda Adli, Katja Wiethoff), Psychiatrische Universitätsklinik Berlin Charité Campus Benjamin Franklin (Isabella Heuser, Gerd Bischof), Berlin Auguste Viktoria Klinik (Joachim Zeiler, Robert Fisher, Cornelia Fähser), Berlin St. Hedwig (Florian Standfest), Berlin St. Joseph (Dorothea Schloth), Psychiatrische Universitätsklinik Düsseldorf (Wolfgang Gaebel, Joachim Cordes, Arian Mobascher), Inn-Salzach-Klinikum Wasserburg (Gerd Laux, Sissi Artmann), Isar-Amper-Klinikum München-Ost Haar (Wolfram Bender, Nicole Theyson), Psychiatrische Universitätsklinik Halle (Andreas Marneros, Dörthe Strube, Yvonne Reinelt, Peter Brieger), Psychiatrische Universitätsklinik Heidelberg (Christoph Mundt, Klaus Kronmüller, Daniela Victor), Psychiatrische Universitätsklinik München LMU (Hans-Jürgen Möller, Ulrich Hegerl, Roland Mergel, Michael Riedel, Florian Seemüller, Florian Wickelmaier, Markus Jäger, Thomas Baghai, Ingrid Borski, Constanze Schorr, Roland Bottlender), Max-Planck-Institut für Psychiatrie (MPIP) München (Florian Holsboer, Matthias Majer, Marcus Ising).

Literatur

1. Beck AT, et al. An inventory for measuring depression. *Arch Gen Psychiatry* 1961;4:561–71.
2. Beck AT, et al. Comparison of the Beck Depression Inventories-IA and -II in psychiatric outpatients. *J Pers Assess* 1996;67:588–97.
3. Beck AT, et al. Psychometric properties of the Beck Depression Inventory Twenty-five years of evaluation. *ClinPsychol Rev* 1988;8:77–100.
4. Carroll BJ, et al. Depression rating scales. A critical review. *Arch Gen Psychiatry* 1973;28:61–6.
5. Carroll BJ, et al. The Carroll rating scale for depression. I. Development, reliability and validation. *Br J Psychiatry* 1981;138:194–200.
6. Coles ME, et al. Psychometric evaluation of the Beck Depression Inventory in adults with social anxiety disorder. *Depress Anxiety* 2001;14:145–8.
7. Demyttenaere K, et al. Patient-assessed versus physician-assessed disease severity and outcome in patients with nonspecific pain associated with major depressive disorder. *Prim Care Companion J Clin Psychiatry* 2009;11:8–15.
8. Demyttenaere K, Jaspers L. Trends in (not) using scales in major depression: A categorization and clinical orientation. *Eur Psychiatry* 2020;63(1):e91.
9. Hautzinger M, et al. Beck Depressions-Inventar. Revision – BDI-II. 2. Aufl. Frankfurt: Hogrefe Verlag, 2009.
10. Kühner C, et al. Reliabilität und Validität des revidierten Beck Depressions-Inventar (BDI-II). Befunde aus deutschsprachigen Stichproben. *Nervenarzt* 2007;78:651–6.
11. Moeller HJ, von Zerssen D. Self-rating procedures in the evaluation of antidepressants. *Psychopathology* 1995;28(6):291–306.
12. Moeller HJ. Outcome criteria in antidepressant drug trials: Self-rating versus observer-rating scales. *Pharmacopsychiatry* 1991;24:71–5.
13. Moeller HJ. Rating depressed patients: Observer- vs self-assessment. *Eur Psychiatry* 2000;15:160–72.
14. Moeller HJ. Standardized rating scales in Psychiatry: Methodological basis, their possibilities and limitations and description of important scales. *World J Biol Psychiatry* 2009;10:257.
15. Parker G, et al. Psychomotor disturbance in depression: Defining the constructs. *J Affect Disord* 1993;27:255–65.
16. Robbins DR, et al. Use of the Hamilton rating scale for depression and the Carroll Self-Rating Scale in adolescents. *Psychiatry Res* 1985;14:123–9.
17. Rush AJ, et al. The 16 item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C) and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003;54:573–83.
18. Rush AJ, et al. The inventory of depressive symptomatology (IDS): psychometric properties. *Psychol Med* 1996;26:477–86.
19. Seemüller F, et al. Outcomes of 1014 naturalistically treated inpatients with major depressive episode. *Eur Neuropsychopharmacol* 2010;20:346–55.
20. Seemüller F, et al. A factor analytic comparison of three commonly used depression scales (HAMD, MADRS, BDI) in a large sample of depressed inpatients. *BMC Psychiatry* 2023;23:548.
21. Seemüller F, et al. Atypical symptoms in hospitalised patients with major depressive episode: Frequency, clinical characteristics, and internal validity *J Affect Disord* 2008;108:271–8.
22. Trivedi MH, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR-D: implications for clinical practice. *Am J Psychiatry* 2006;163:28–40.
23. Uher R, et al. Comparison and integration of three scales in the gendep study. *Psychol Med* 2008;38:289–300.
24. von Zerssen D, Cording C. The measurement of change in endogenous affective disorders. *Arch Psychiatr Nervenkr* 1978;226:95–112.
25. von Zerssen D. Clinical Self-Rating Scales (CSRS) of the Munich Psychiatric Information System (PSYCHIS). In: Sartorius et al. *Assessment of Depression*. Berlin: Springer Verlag, 1986:270–303.
26. von Zerssen D. Klinische Selbstbeurteilungsskalen KSbS aus dem Münchner Psychiatrischen Informationssystem (PSYCHYS München). *Befindlichkeits-Skala*. Weinheim: Beltz Verlag, 1976.
27. Wikberg C, et al. Comparison between the Montgomery-Asberg Depression Rating Scale- Self and the Beck Depression Inventory II in Primary Care. *Prim Care Companion CNS Disord* 2015;17:10.4088/PCC.14m01758.
28. Zung WW, et al. Self rating depression scale in an outpatient clinic. Further validation of the SDS. *Arch Gen Psychiatry* 1965;13:508–15.