

Dimensionen der Hamilton-Depressionsskala (HAMD)

Faktorenanalytische Untersuchungen

Wolfgang Maier, Michael Philipp und Armard Gerken

Psychiatrische Klinik der Johannes-Gutenberg-Universität (Direktor: Prof. Dr. O. Benkert),
Langenbeckstrasse 1, D-6500 Mainz, Bundesrepublik Deutschland

Dimensions of the Hamilton-Depression-Scale (HAMD)

A Factor Analytical Study

Summary. The results of factor analysis of HAMD are not unitary. Studying the stability in course and the independence of selection criteria for the population we rated two groups ($n = 107$, $n = 98$) of inpatients with HAMD; these groups differ in nosological selection criteria. We furthermore rated one group at different timepoints (before treatment and 3 weeks after). The one-factor-solution is the only stable one and the only one which is independent from selection criteria. In all groups the one-factor-solutions show a high similarity with other published solutions of the German version of HAMD. But there is no way to extract a general factor. Thus the ability of HAMD in judging the severity of depression is doubtful. Ways are discussed to circumvent this problem.

Key words: Depression scales – Depression – Severity of depression – Rating scales

Zusammenfassung. Die Resultate der faktorenanalytischen Untersuchungen der HAMD sind uneinheitlich. Um Verlaufsstabilität und Stichprobenunabhängigkeit der faktoriellen Zerlegungen zu prüfen, wurden zwei Gruppen depressiver Patienten ($n = 107$, $n = 97$), die nach operationalisierten Kriterien selektiert wurden, mit der HAMD beurteilt. Diese Gruppen unterscheiden sich nach ihrer nosologischen Zusammensetzung. In einer der Gruppen wurden die Ratings zu verschiedenen Zeitpunkten (vor Therapie bzw. 3 Wochen nach Beginn der Therapie) durchgeführt. Die Einfaktorenlösung erwies sich als einzige als stichprobenunabhängig und verlaufsstabil. In allen Gruppen zeigten die Einfaktorenlösungen eine hohe Ähnlichkeit mit entsprechenden Faktoren anderer publizierter Faktorenanalysen der deutschen Version der HAMD: es gelang jedoch nicht, einen Generalfaktor zu extrahieren. Das zieht die Fähigkeit der HAMD zur Beurteilung des Schweregrades der depressiven Symptomatik in Zweifel. Es werden Möglichkeiten diskutiert, diese Schwierigkeit zu umgehen.

Schlüsselwörter: Depressionsskala – Depression – Schweregrad der Depression – Ratingskalen

Einleitung

Zur Beurteilung von Therapieeffekten von Psychopharmaka sind psychologische Untersuchungsinstrumente notwendig. Bei der Evaluation von Antidepressiva stellen sich für psychologische Erhebungsverfahren dabei die folgenden Aufgaben:

1. Beurteilung des Schweregrades der depressiven Symptomatik zur Erfassung von Therapieeffekten;
2. Beurteilung des Ausprägungsgrades von Subtypen des depressiven Syndroms zur Erfassung der Zielsymptomatik oder von Prognosekriterien für die Antidepressiva-Therapie.

Die Hamilton-Depressions-Skala (HAMD) ist das gebräuchlichste Beurteilungsinstrument bei der Evaluation des Therapieeffektes von Antidepressiva. Neuerdings sind Zweifel an der Eignung der HAMD als Schweregradindikator entstanden [2, 3, 24]. Daher ist es notwendig, die Eignung der HAMD für die genannten Verwendungszwecke zu prüfen; eine Möglichkeit hierzu sind Faktorenanalysen [28], die eine Variablenmenge (Items) auf eine geringe Zahl von Variablen (Faktoren) reduzieren (Faktorenanalyse als Mittel der Datenreduktion).

Im folgenden werden Faktorenanalysen der HAMD dargestellt und bezüglich der drei Fragen diskutiert:

1. Sind Faktoren extrahierbar, deren Globalsumme als Indikator für den Schweregrad der depressiven Symptomatik interpretiert werden kann?
2. Sind Faktoren extrahierbar, deren Globalsumme als Indikator für den Schweregrad von Subtypen der depressiven Symptomatik interpretiert werden können?
3. Sind die jeweils extrahierten Faktoren stabil, d. h.
 - a) Sind sie von der Auswahl der Selektionskriterien für die untersuchte Patientenstichprobe abhängig?
 - b) Sind sie vom Erhebungszeitpunkt (z. B. vor Therapie bzw. 3 Wochen nach Beginn der Therapie) abhängig?

Zu diesen faktorenanalytischen Fragestellungen sagen die bisherigen Analysen [1, 7, 11–13, 15, 20, 25, 30] folgendes aus:

1. Ein Faktor, der auf allen Items bzw. auf fast allen Items mittlere oder hohe Ladungen trägt, konnte nicht in allen Faktorenanalysen extrahiert werden; ein Generalfaktor [28, Seite 58] ist aber wünschenswert, wenn die HAMD in

ihrem Globalsummskore den Schweregrad der depressiven Symptomatik repräsentieren soll.

2. Faktoren, die auf einigen Items mittlere bzw. hohe Ladungen tragen, auf den anderen Items dagegen niedrige Ladungen aufweisen, zeigen in den verschiedenen Analysen ungenügende Gemeinsamkeiten. So hat Hamilton [12] einen bipolaren Faktor isoliert, auf dem die Items „Hemmung“ bzw. „Erregung“ Koeffizienten mit hohen Absolutwerten und jeweils entgegengesetztem Vorzeichen aufweisen; solche bipolaren Faktoren konnten in anderen Untersuchungen [z. B. 1] nicht reproduziert werden.

Auf dem Hintergrund dieser widersprüchlichen Resultate haben Möller und v. Zersen [22] für den deutschsprachigen Raum verbindliche faktorenanalytische Zerlegungen der HAMD gefordert. Hierzu sind Analysen aus verschiedenen Zentren notwendig.

Die bisher einzige faktorenanalytische Zerlegung der deutschsprachigen Version der HAMD [1] wurde an einer nach der ICD klassifizierten Patientenstichprobe durchgeführt, in der sich Patienten mit neurotischer und endogener Depression sowie mit schizoaffektiver Psychose befanden; ein Fünftel der Patienten war älter als 60 Jahre. Dabei resultierte als einzige Lösung, die sowohl im Verlauf als auch bei Zufalls- teilung der Population hinreichende Stabilität aufwies, eine Einfaktorenlösung; eine mehrfaktorielle Auswertung der HAMD bzw. eine Beurteilung von Prägnanztypen der depressiven Symptomatik mit der HAMD ist damit in Frage gestellt. Außerdem stellt der extrahierte Faktor keinen Generalfaktor dar; mehr als ein Fünftel der Items trägt Ladungskoeffizienten kleiner als 0,4. Damit ist jedoch auch die Eignung der HAMD zur Beurteilung des Schweregrades der depressiven Symptomatik zweifelhaft. Angesichts der Verbreitung der HAMD in der Evaluationsforschung ist dieses Resultat beunruhigend. In [1] wird nicht weiter diskutiert, wodurch dieses Ergebnis bedingt ist; insbesondere wird nicht untersucht, ob die Zusammensetzung der Stichprobe hierfür verantwortlich ist. Es sind also Wiederholungsstudien notwendig, in denen untersucht wird, ob die Resultate in [1] durch die Selektion der Patienten bedingt sind.

Untersuchungsmethoden

Charakterisierung der Stichprobe (s. Tabelle 1)

Stichprobe 1 und Stichprobe 3 umfassen dieselbe Patientengruppe; die beiden Stichproben unterscheiden sich bezüglich verschiedener Erhebungszeitpunkte der HAMD. Stichprobe 2 umfaßt wie Stichprobe 1 stationäre Patienten, die innerhalb der ersten vier Tage des stationären Aufenthaltes nach der HAMD beurteilt wurden; Stichprobe 2 ist dabei heterogener zusammengesetzt (bezüglich der diagnostischen Zuordnung) als Stichprobe P1.

Stichprobe P1. Sämtliche Patienten erfüllen die Kriterien von Major Depressive Disorder (endogenous type—sicher) nach RDC [31]; es mußte mindestens die 2. abgesetzte depressive Phase vorliegen. Diese Stichprobe umfaßt sämtliche stationäre Patienten, die an einer von fünf Antidepressivastudien teilnahmen. Die Beurteilung auf der HAMD erfolgte nach einer mindestens 4tägigen antidepressivafreien Periode vor Ansetzen der antidepressiven Medikation.

Stichprobe P2. Diese Stichprobe besteht aus allen Patienten, die in einem festen Zeitraum in die Psychiatrische Universitätsklinik Mainz aufgenommen wurden und sich einer minde-

stens 14tägigen stationären Therapie unterzogen und die Kriterien A und B von Minor Depressive Disorder gemäß RDC genügten; Patienten, die nach RDC an einer Schizophrenie litten, wurden (ebenso wie in P1) ausgeschlossen.

Die Stichproben P1 und P2 sind disjunkt.

Stichprobe P3. Die Patienten dieser Gruppe sind sämtlich Patienten, die auch der Stichprobe P1 angehören; sie wurden am Ende der Antidepressivastudien (nach 21 Tagen) erneut mit der HAMD beurteilt. Bei 19 Patienten in P1 mußten die Antidepressivastudien abgebrochen werden; diese Patienten gehörten nicht der Stichprobe P3 an.

In keine der Stichproben wurden Patienten aufgenommen, die an einer ernsthaften körperlichen Erkrankung, einem Anfallsleiden oder einem Abusus von Medikamenten, Alkohol oder Drogen litten.

Durchführung der Untersuchung

Trainierte Beurteiler wendeten in sämtlichen drei Stichproben die deutsche Version der HAMD mit 21 Items an (Version: CIPS [6]). In der Stichprobe P2 wurde die Exploration durch ein strukturiertes Interview nach present state examination (PSE) [34] eingeleitet. Die Beurteiler der Stichproben P1 und P3 beteiligten sich nicht am Rating in der Stichprobe P2, die Ratings fanden jeweils vormittags statt.

Analysenmethoden

Jeder der drei gewonnenen Datensätze wurde einer Itemanalyse (Mittelwerte, Varianzen, Trennschärfen, innere Konsistenz) einer Faktorenanalyse (Hauptfaktorenanalyse schiefwinkliger und orthogonaler Rotation) unterzogen. In den Faktorenanalysen wurden sämtliche Faktoren berücksichtigt, die Eigenwerte größer 1 aufweisen [27, S. 125].

Ergebnisse

1. Mittelwerte, Itemanalysen

- a) Der Mittelwert von P1 (22,26; $s = 6,70$) und P3 (12,46; $s = 7,73$) liegt unter den in [1] angegebenen Werten: 24,19

Tabelle 1. Zusammensetzung der untersuchten Patientengruppe

	P1	P2	P3
Anzahl der Patienten	($n = 107$)	($n = 98$)	($n = 88$)
Geschlecht:			
weiblich	($n = 69$)	($n = 71$)	($n = 62$)
männlich	($n = 38$)	($n = 27$)	($n = 26$)
Alter (Mittelwert) (jedenfalls 20–60 J)	42,1 J	40,2 J	41,0 J
Major Depressive Disorder (i. S. RDC)	Alle	($n = 43$)	Alle
Major Depressive Disorder endogenous type (sicher) (i. S. RDC)	Alle	($n = 22$)	Alle
Bipolar (bipolar I oder II)	($n = 33$)	($n = 7$)	($n = 26$)
Unipolar Major Depressive Disorder	($n = 74$)	($n = 36$)	($n = 62$)
Minor Depressive Disorder (i. S. RDC)	∅	35	∅
Schizoaffective Disorder (depressive type) (i. S. RDC)	∅	20	∅

Tabelle 2. Statistische Kennwerte

	P1	P2	P3
<i>HAMD-Version mit 17 Items:</i>			
Mittelwert	22,26	18,19	12,46
Standardabweichung	6,70	7,34	7,73
In Globalsumme repräsentierter Varianzanteil	22,1%	24,2%	33,1%
Konsistenz (α -Koeffizient v. Cronbach)	0,72	0,73	0,83
<i>HAMD-Version mit 21 Items:</i>			
Mittelwert	24,98	20,83	14,82
Standardabweichung	6,95	7,69	8,05
In Globalsumme repräsentierter Varianzanteil	21,5%	23,6%	32,5%
Konsistenz (α -Koeffizient v. Cronbach)	0,73	0,75	0,85

($s = 7,32$) für Tag 0 und $14,18$ ($s = 8,86$) für Tag 21 (jeweils 17-Items-Version); in der Stichprobe von Hamilton [11] wird ein Mittelwert von $23,03$ angegeben. Die Vergleichsstichproben sind dabei jeweils nach anderen als den hier verwendeten Kriterien erhoben.

b) Die Trennschärfe der Items 14 (Libidoverlust) (0,05), 15 (Hypochondrie) (0,07), 16 (Gewichtsverlust) (0,05) weisen

Werte auf, die nicht signifikant (Niveau 1%) von 0 abweichen (HAMD-Version mit 17 Items). Entsprechend sind in der Version mit 21 Items die Items 19 (Depersonalisation) (0,01) und 21 (Zwangssymptome) (0,03) in jedenfalls einer der untersuchten Stichproben nicht hinreichend trennscharf.

c) Die Konsistenzkoeffizienten (innere Konsistenz nach Cronbach) liegt im gewünschten Bereich [17, S.309]. Die Konsistenz liegt bei der Version mit 21 Items geringfügig höher als in der Version mit 17 Items. Wird der Konsistenzkoeffizient als Reliabilitätsschätzung verwendet, so zeigt er eine ausreichende Reliabilität der Skala an.

2. Extrahierbarkeit von Faktoren zur Beurteilung von Subtypen des depressiven Syndroms

Bei der Darstellung verschiedener Qualitäten in einer Skala bedient man sich in der Literatur vor allem zweier Rotationsprinzipien: der orthogonalen und der schiefwinkligen Rotation.

In sämtlichen untersuchten Stichproben ergeben sich für die Version mit 17 Items jeweils 4 Faktoren; diese Lösung ist mit dem Scree-Graph-Kriterium [27, S. 127] verträglich. Bei Verwendung beider Rotationsprinzipien können diese Faktoren nach Leitsymptomen charakterisiert werden (Hemmung, Erregung — psychische Angst, Hypochondrie — somatische Angst, Schlafstörungen).

Tabelle 3. Faktorenanalysen für Stichproben P1, P2, P3 für HAMD (17 Items). Schiefwinklige Rotation, Items mit Ladungen kleiner 0,40 sind nicht aufgeführt; Ladungen der Items auf den Faktoren sind in Klammern angegeben

Stichproben	Faktoren			
	Faktor „Hemmung“	Faktor „Erregung – psychische Angst“	Faktor „Hypochondrie – somatische Angst“	Faktor „Schlafstörungen“
P1 Major depressive disorder (endogenous) akut Wash-out Phase ($n = 107$)	Depressive Stimmung (0,44) Suicidalität (0,79) Hemmung (0,69) Körperl. Sympt. – gastroint. (0,44) 1	Depressive Stimmung (0,40) Erregung (0,61) Angst psych. (0,61) Krankheitsuneinsichtigkeit (0,53) 2	Arbeit (0,83) Körperl. Symptome – allgem. (0,46) Hypochondrie (0,40) 3	Einschlafstörungen (0,61) Durchschlafstörungen (0,80) Schlafstörungen am Morgen (0,60) 4
P2 Major und Minor depressive disorder akut teilweise unter Medikation ($n = 98$)	Depressive Stimmung (0,51) Arbeit (0,57) Hemmung (0,81) Körperl. Symptome – allgem. (0,46) Genitalstörungen (0,46) 1	Suizidalität (0,52) Erregung (0,52) Angst psych. (0,72) 2	Som. Angst (0,55) Körperl. Symptome – allgem. (0,55) Hypochondrie (0,45) 4	Einschlafstörungen (0,69) Durchschlafstörungen (0,90) Schlafstörungen am Morgen (0,63) 3
P3 Major depressive disorder (endogenous) nach 3 Wochen Therapie unter Standard. Medikation ($n = 88$)	Arbeit (0,68) Hemmung (0,72) Körperl. Symptome – allgem. (0,40) Gewichtsverlust (0,45) 2	Depressive Stimmung (0,49) Schuldgefühle (0,45) Suizidalität (0,50) Erregung (0,59) Angst psych. (0,56) 1	Som. Angst (0,49) Hypochondrie (0,70) 4	Einschlafstörungen (0,57) Durchschlafstörungen (0,77) Schlafstörungen am Morgen (0,85) 3

Die Ziffern bezeichnen die Rangfolge der Faktoren bezüglich der erklärten Varianz

Die in Tabelle 3 gegebene Zuordnung der Items zu den Faktoren ist unabhängig von der verwendeten Rotationsmethode. Allerdings zeigt sich eine mangelnde Kongruenz der einzelnen Faktoren zwischen den verschiedenen Stichproben: Das Item „Suizidalität“ lädt in der Stichprobe P1 auf dem durch das Leitsymptom „Hemmung“ charakterisierten Faktor relevant (größer 0,4); in den beiden anderen Stichproben lädt „Suizidalität“ auf dem Faktor, der durch „Erregung — psychische Angst“ als Leitsymptom gekennzeichnet ist. Es entspricht der klinischen Erfahrung, daß im Therapieverlauf eine Erhöhung des Antriebspotentials häufig mit einem Weiterbestehen der depressiven Stimmung der kognitiven Fehleinstellung (z. B. Suizidalität) einhergehen kann. Allein die „Schlafstörungen“ weisen eine weitgehende Kongruenz auf.

Entsprechend zeigen die Ähnlichkeitskoeffizienten für den Globalvergleich zwischen den Faktorenlösungen verschiedenen Umfangs (Koeffizienten nach Fischer und Roppert [9]) und die Ähnlichkeitskoeffizienten für den Vergleich der einzelnen Faktoren in der 4-Faktoren-Lösung (Koeffizienten nach Burt [29]) relativ niedrige Werte. Die Fischer-Roppert-Ähnlichkeitskoeffizienten zeigen für die resultierenden 4-Faktoren-Lösungen beim Vergleich zwischen den verschiedenen Stichproben P1 und P2 bzw. P2 und P3 unzureichende Werte (jeweils kleiner 0,80), wenn als Normwerte die in [10] angegebenen Werte zugrunde gelegt werden (Tabelle 4). Die Faktorenkongruenzkoeffizienten nach Burt sind nur beim Vergleich der Faktoren, die durch das Leitsymptom „Schlafstörungen“ charakterisiert sind, ausreichend hoch (zugrundelegen der in [28] angegebene Normwert = 0,90) (Tabelle 5).

Bei der Analyse der Version mit 21 Items ergeben sich in den verschiedenen Stichproben Lösungen verschiedenen Umfangs: Eine 5-Faktoren-Lösung in P3, eine 6-Faktoren-Lösung in P2 und eine 7-Faktoren-Lösung in P1. Die Lösungen sind jeweils mit dem Scree-Graph-Kriterium verglichen. Die Ähnlichkeitskoeffizienten für die 4-, 5-, 6- und 7-Faktoren-Lösungen sind nicht ausreichend (gemäß den in [10] angegebenen Normwerten), allerdings resultiert auch bei dieser Version in allen Stichproben ein Faktor mit dem Leitsymptom

Tabelle 4. Ähnlichkeitskoeffizienten von Hauptkomponentenlösungen (Fischer-Poppert)

Stichprobenpaare	Zahl der vorkommenden Faktoren			
	1	2	3	4
P1/P2	0,91	0,79	0,67	0,75
P1/P3	0,87	0,82	0,88	0,92
P2/P3	0,87	0,76	0,68	0,78

Tabelle 5. Faktorkongruenz-Koeffizient nach Burt für den Vergleich von Faktoren mit gleichen Leitsymptomen von jeweils zwei Stichproben

Stichproben	Faktoren			
	Faktor „Hemmung“	Faktor „Erregung“	Faktor „Schlafstörungen“	Faktor „Hypochondrie“
P1/P2	0,68	0,64	0,93	0,20
P1/P3	0,45	0,76	0,87	0,47
P2/P3	0,75	0,67	0,88	0,72

Werte unter 0,8 zeigen eine unzureichende Vergleichbarkeit der einzelnen Faktoren in den verglichenen Populationen an

„Schlafstörungen“, der eine ausreichende Stichprobenunabhängigkeit aufweist.

3. Extrahierbarkeit eines Generalfaktors zur Beurteilung des Schweregrades der depressiven Symptomatik

Der Schweregrad kann sich sowohl in Symptomen (Items) als auch in Symptomverbänden bzw. Syndromen (Faktoren) niederschlagen; entsprechend können bei der Faktorenanalyse, die zur Extraktion eines als Schweregradindikator interpretierbaren Faktors durchgeführt wird, in jeder Stichprobe zwei verschiedene Korrelationsmatrizen zugrunde gelegt werden:

- a) die Korrelationsmatrix der einzelnen Items der HAMD (Schweregrad als Faktor erster Ordnung);
- b) die Korrelationsmatrix für jene Faktoren, die aus den Items extrahiert werden konnten und auf eine nicht-orthogonale Struktur rotiert wurden (Schweregrad als Faktor 2. Ordnung).

Für die Version mit 17 Items ergeben sich bei der Analyse der Korrelationsmatrix der einzelnen Items die folgenden Resultate (siehe Tabelle 6):

a) Erwartungsgemäß hat in allen untersuchten Stichproben der unrotierte und varianzstärkste Faktor die meisten Items mit mittlerer oder höherer Ladung (Koeffizienten 0,4). Diese Items kennzeichnen keinen Subtyp des depressiven Syndroms, so daß der varianzstärkste Faktor am ehesten als Schweregradindikator interpretiert werden kann. Aber die Zahl der Items mit relevanter Ladung ist unbefriedigend: in den Stichproben P1 und P2 sind 6 Items und in der Stichprobe P3 sind 5 Items mit Werten unter 0,4 geladen.

Ein Generalfaktor kann in der Faktorenanalyse eines Datensatzes unter Umständen aus methodischen Gründen nicht extrahiert werden: bei zu geringer Heterogenität der Stichprobe [27, Seite 295] bzw. bei Fehlen eines heterogenen Kontextes [15] kann ein tatsächlich vorhandener Generalfaktor unter Umständen nicht extrahiert werden. Doch weder eine Heterogenisierung der Stichprobe (Vereinigung der beiden Stichproben P1 und P2 zu einer neuen Stichprobe, deren HAMD-Daten einer Faktorenanalyse unterworfen werden)

Tabelle 6. Unrotierte varianzstärkste Faktoren in verschiedenen Stichproben

Itemnr.	P1	P2	P3
1	0,80	0,65	0,89
2	0,51	0,26	0,65
3	0,57	0,21	0,72
4	0,32	0,51	0,51
5	0,30	0,53	0,30
6	0,44	0,54	0,46
7	0,61	0,62	0,69
8	0,57	0,38	0,67
9	0,45	0,21	0,65
10	0,76	0,59	0,53
11	0,21	0,59	0,53
12	0,50	0,69	0,60
13	0,47	0,64	0,70
14	0,10	0,54	0,37
15	0,13	0,25	0,13
16	0,10	0,59	0,31
17	0,41	0,15	0,27

noch die Einbettung der 17-Items-Version der HAMD in einen heterogenen Kontext (Analyse der 21-Items-Version anstelle der Version mit 17 Items) führt zu einer Erhöhung der Anzahl der auf dem varianzstärksten Faktor relevant ladenden Items.

Für die unrotierten varianzstärksten Faktoren ist die Stichprobenunabhängigkeit gegeben (Tabelle 4): der niedrigste Wert für den Fischer-Roppert-Koeffizienten liegt bei 0,87 (Vergleich von P2 und P3). Derselbe Koeffizient ergibt beim Vergleich der Einfaktorenlösungen, die von Baumann [1] für verschiedene Ratingtage angegeben werden, stets Werte größer als 0,85 und mittlere Werte von 0,91.

Der Zusammenhang zwischen dem Globalsummscore der HAMD und den unrotierten varianzstärksten Faktoren in den untersuchten Stichproben ist eng: die Spearman-Korrelation zwischen den beiden Variablen (Globalsummscore der 17-Items-Version bzw. mit Ladungskoeffizienten gewichteter Summscore des unrotierten varianzstärksten Faktors) beträgt in P1 0,96, in P2 0,95 und in P3 0,97. Der Ähnlichkeitskoeffizient nach Burt zwischen den beiden Variablen zeigt in sämtlichen 3 Stichproben Werte größer 0,91. Entsprechend können der Globalsummscore der HAMD und die unrotierten varianzstärksten Faktoren als weitgehend identisch angesehen werden.

Die Version mit 21 Items zeigt eine noch geringere Tendenz zur Generalfaktorenlösung als die kürzere Version: die Anzahl der Items mit irrelevanten Ladungen (Koeffizienten kleiner 0,4) beträgt in P1 9, in P2 11 und in P3 6. Die Ähnlichkeitskoeffizienten zwischen diesen Faktoren erbringen niedrigere Werte (der niedrigste Wert ist 0,76 bei Vergleich des Faktors von P1 mit dem entsprechenden Faktor von P2).

b) Beim Versuch einen Faktor zweiter Ordnung als Index für den Schweregrad der depressiven Symptomatik zu gewinnen, wird eine schiefwinklige Mehrfaktorenlösung (Faktoren erster Ordnung) zugrunde gelegt (in allen 3 Stichproben die in Tabelle 3 dargestellten Lösungen). Bei dieser Faktorenanalyse zweiter Ordnung (17-Items-Version) ergibt sich in den Stichproben P1 und P3 eine 1-Faktoren-Lösung und in der Stichprobe P2 eine 2-Faktoren-Lösung; bipolare Faktoren resultieren nicht. Ein Generalfaktor kann aber in den Stichproben P1 und P2 weder unter den unrotierten noch unter den rotierten Faktoren ausfindig gemacht werden; lediglich in P3 resultiert ein Generalfaktor.

Für die HAMD (Version 21 Items) resultiert bei der Faktorenanalyse zweiter Ordnung in der Stichprobe P3 ein Generalfaktor. Auch bei Zugrundelegen der mehrfaktoriellen Zerlegung der HAMD (Version mit 21 Items), die von Rhoades [26] vorgeschlagen wurde, kann in keiner der untersuchten Stichproben ein Faktor zweiter Ordnung isoliert werden, der auf sämtlichen Faktoren erster Ordnung (diese sind die Subskalen von Rhoades) relevante Ladungen aufweist.

Diskussion

Die Itemanalysen zeigen, daß die Items der HAMD in den Mittelwerten und Trennschärfen breit streuen. Die Trennschärfen für die Items „Hypochondrie“, „Genitalstörungen“, „Gewichtsverlust“ genügen bei relativ geringen Mittelwerten nicht den in [17, Seite 126ff.] angegebenen Kriterien. Die Koeffizienten für die innere Konsistenz — die auch als Reliabilitätskoeffizienten interpretiert werden können — sind für die beiden Versionen der Skala ausreichend [17, S.242]. Die Erweiterung von 17 auf 21 Items bringt nur einen beschei-

denen Zuwachs an innerer Konsistenz. Sieht man in der Erweiterung vor allem ein Mittel zur Reliabilitätssteigerung, so kann auf die Erweiterung verzichtet werden. Diese Konsequenz wird auch durch die relativ geringe Stichprobenunabhängigkeit der Faktorenanalyse der Skala mit 21 Items nahegelegt. Dieses Resümee entspricht dem von Baumann [1].

In jeder Stichprobe ergaben sich für die 17-Items-Version jeweils 4-Faktoren-Lösungen. Für die Version mit 21 Items ergaben sich Lösungen, deren Faktorenzahlen stichprobenabhängig sind. Die meisten bisher vorgelegten Analysen geben vergleichbare Faktorenzahlen an. Insbesondere zur bisher einzigen Analyse mit der deutschsprachigen Version [1] bestehen keine wesentlichen Widersprüche: dort werden 1-bis 4-, 5-, 6-Faktoren-Lösungen als mit dem Scree-Graph-Kriterium verträglich angesehen, wobei die Anzahl der Eigenwerte, die größer als 1 sind, vier bis sechs beträgt. Die 4 Faktoren, die für die Version mit 17 Items in allen untersuchten Stichproben extrahiert werden konnten, zeigen keine bipolaren Strukturen; sie lassen sich mit Leitsymptomen charakterisieren (Hemmung; Erregung — psychische Angst, Hypochondrie — somatische Angst, Schlafstörungen), die gebräuchlichen klinischen Differenzierungen entsprechen.

Allerdings sind die gewonnenen Faktoren nicht zur Erfassung des Schweregrades von Subtypen des depressiven Syndroms geeignet; denn sie sind nicht hinreichend verlaufsstabil und stichprobenunabhängig. Diese Unstabilität kann die inhaltliche Heterogenität der Mehrfaktorenlösungen, die bisher in der Literatur angegeben wurden, erklären.

In [32] konnten ebenfalls 4 Faktoren extrahiert werden, die durch dieselben Leitsymptome charakterisiert werden konnten; dagegen wurden in [4] 4 Faktoren verwendet, bei denen „Erregung“ und „Angst“ je einen eigenen Faktor bezeichnen. Auch die von Rhoades et al. [26] vorgeschlagene Subskalierung der HAMD (Version mit 21 Items) konnte nicht reproduziert werden.

Die Einfaktorenlösung ist dagegen verlaufsstabil und stichprobenunabhängig; sie zeigt außerdem in allen untersuchten Stichproben eine hinreichende Ähnlichkeit mit den in [1] isolierten Einfaktorenlösungen. Eine einfaktorielle Auswertung der Skala mit 17 Items erscheint unter diesem Gesichtspunkt adäquat; trotzdem genügt die Einfaktorenlösung den hier verwendeten Kriterien zur Extraktion der Faktoren nicht. Der Faktor der Einfaktorenlösung kann nur bedingt als Schweregradindex der depressiven Symptomatik interpretiert werden: in allen untersuchten Stichproben zeigen mindestens ein Viertel der Items auf diesen Faktoren irrelevante Ladungen. Die Summenbildung beim Faktor der Einfaktorenlösung und die Globalsummenbildung der HAMD repräsentieren inhaltlich dasselbe, denn sie sind hoch miteinander korreliert (größer 0,95). Kann also der Faktor der Einfaktorenlösung bedingt als Maßstab für den Schweregrad einer depressiven Symptomatik gelten, so ist dies auch für den Globalsummscore der HAMD der Fall. Andere Möglichkeiten zur faktorenanalytischen Isolierung eines Faktors, dessen Summscore als Schweregrad der depressiven Symptomatik interpretiert werden könnte, versagen ebenso (Konstruktion von Faktoren zweiter Ordnung, Faktorenanalyse von Summscores vorgegebener HAMD-Subskalen, Heterogenisierung der Stichprobe); so kann zusammenfassend der Schluß gezogen werden:

1. Unter den geprüften Bedingungen kann in der HAMD nur ein einziger Faktor isoliert werden, der keinen Subtyp der depressiven Symptomatik beschreibt; dieser Faktor ist

jedoch nur bedingt als Indikator für den Schweregrad der depressiven Symptomatik interpretierbar.

2. In der HAMD sind keine weiteren Informationen zu extrahieren, die eine hinreichende faktorielle Validität aufweisen.
3. Die HAMD ist eine recht unökonomische Skala, da jener Faktor, auf den allein eine Auswertung gestützt werden kann, maximal ein Drittel der Varianz erklärt; in anderen Skalen zur Beurteilung des Schweregrades der depressiven Symptomatik erklärt der Summenscore jedenfalls die Hälfte der Varianz [19].

Die HAMD wurde wegen der hohen inhaltlichen Heterogenität, dem Überwiegen der Items, die körperlich betonte Symptome abdecken, und der ungenügenden Beurteilung von Veränderungen des Schweregrades kritisiert [3, 24]. Um die damit gegebenen Nachteile für die Therapieevaluationsforschung zu vermeiden, wurden zur Fremdbeurteilung des Schweregrades der depressiven Symptomatik andere Skalen vorgeschlagen [3, 24] (Itemumfänge 11 bzw. 10). Die hier vorgetragene Kritik an der HAMD unterstreicht die Notwendigkeit zur Konstruktion und zur parallelen Verwendung anderer Skalen zur Beurteilung des Schweregrades des depressiven Syndroms.

Andererseits ist die HAMD die breit akzeptierte Grundlage für den Wirkungsnachweis von Antidepressiva; sie stellt die Vergleichsbasis für verschiedene Studien dar. Zur Sicherung der Vergleichbarkeit der Antidepressivastudien erscheint eine Ersetzung der HAMD durch bessere Konkurrenzskalen nicht als sinnvoll. Obwohl die Beschreibung von Therapieeffekten mit dem Globalsummscore der HAMD zur Sicherung der Vergleichbarkeit unabdingbar ist, sollte versucht werden, aus der HAMD einen Subscore zu extrahieren, der den Schweregrad der depressiven Symptomatik besser zu beurteilen vermag; Subscores wären auch aus den bereits abgeschlossenen Studien extrahierbar. Von solchen Subskalen wäre eine Erhöhung der Homogenität (z. B. Generalfaktorenlösung) ohne Reduktion der Reliabilität und der Validität zu fordern.

Bech et al. [2] haben bereits aufgrund anderer Erwägungen (Homogenitätsanalyse) eine Verkürzung der HAMD vorgeschlagen und durchgeführt. In [18] wurde eine Reduktion der HAMD (Version mit 17 Items) mit anderen als den von Bech verwendeten Methoden aufgrund des hier analysierten Datenmaterials durchgeführt, wobei eine ähnliche Skala resultierte; die Items dieser Subskalen repräsentieren den Kernbereich depressiver Symptomatik. Diese Resultate weisen ebenso wie die Analysen von Heimann [14] darauf hin, daß die Beurteilung von Symptomen aus dem Kernbereich depressiver Symptomatik besser zur Erfassung des Schweregrades geeignet ist als die Beurteilung breit gestreuter Symptome, die von den Items der HAMD (Version mit 17 bzw. 21 Items) dargestellt werden.

Literatur

1. Baumann U (1976) Methodische Untersuchungen zur Hamilton-Depressions-Skala. Arch Psychiatr Nervenkr 222:359-375
2. Bech P, Allerup P, Gram F, Reisby N, Rosenberg R, Jacobson O, Nagy A (1981) The Hamilton depressions scale. Acta Psychiatr Scand 63:290-299
3. Bech P (1984) The instrumental use of rating scales of depression. Pharmacopsychiatry 17:22-29
4. Beckmann H, Schmauss N (1983) Clinical investigations into anti-depressive mechanisms. I. Antihistaminic and cholinolytic effects: Amitriptyline versus promethazine. Arch Psychiatr Nervenkr 233:59-71
5. Cicchetti V, Prusoff BA (1983) Reliability of depression and associated clinical symptoms. Arch Gen Psychiatry 40:987-990
6. CIPS-Collegium Internationale Psychiatriae Salarum (1977) Internationale Skalen für Psychiatrie, Berlin
7. Cleary P, Guy W (1977) Factor analysis of the Hamilton depression scale. Drug Exp Clin Res 1:115-120
8. Endicott J, Cohen J, Nee J, Fleiss J, Sarantakos S (1981) Hamilton depression scale. Arch Gen Psychiatr 38:98-103
9. Fischer G, Roppert J (1964) Bemerkungen zu einem Verfahren der Transformationsanalyse. Arch Ges Psychol 116:98-100
10. Gebhardt F (1968) Die Ähnlichkeit von Faktorenmatrizen. Psych Beiträge 11:591-599
11. Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry 23:56-62
12. Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278-296
13. Hedlund IL, Vieweg BW (1979) The Hamilton rating scale for depression. A comprehensive review. J Operational Psychiatr 10:149-165
14. Heimann H (1975) Methodologische Probleme bei der Effizienzprüfung von Psychopharmaka. Arch Psychiatr Nervenkr 220:225-268
15. Hodder A, Burt CG, Holt NF, Gade JFJ (1965) Depressive state. Harper & Rowe, Springfield, Ill
16. Jager AG, Hormann HJ (1981) Determination of "g" (allgemeine Intelligenz) und die Bedeutung des Variablenkontextes bei explorativen Faktorenanalysen. Psych Beiträge 23:403-420
17. Lienert GA (1972) Testaufbau und Testanalyse. Beltz-Verlag, Weinheim
18. Maier W, Philipp M (1985) Improving the assessment of the severity of depression. A reduction of the Hamilton depression scale. Pharmakopsychiatry (in press)
19. Maier W, Philipp M (1985) Comparative analysis of rating scales for depression. Acta Psychiatr Scand (in press)
20. Michaux MH, Suideles A, Garmize C, Rossi AJ (1969) Depression factors in depressed and in heterogenous in-patient sample. J Neurol Neurosurg Psychiatry 32:609-613
21. Michaux MH, Zwaag LV, Kurland AA (1967) Dimensions of manifest depression in newly hospitalized patients. Br J Psychiatry 113:981-986
22. Möller HJ, Zerssen D v (1983) Psychometrische Verfahren. II. Standardisierte Beurteilungsverfahren. Nervenarzt 54:1-16
23. Mokken P (1972) Scale analysis. University Press, Den Haag
24. Montgomery P, Asberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatr 134:382-389
25. Mowbray RM (1972) The Hamilton scale for depression. A factor analysis. Psychol Med 2:272-280
26. Rhoades M, Overall JE (1983) The Hamilton-depression-scale: factor scoring and probit-scoring. Psychopharmacology Bull 19:121-125
27. Pawlik K (1968) Dimensionen des Verhaltens. Huber-Verlag, Bern
28. Revenstorff D (1976) Lehrbuch der Faktorenanalyse. Kohlhammer-Verlag, Stuttgart
29. Schneewind KA, Cattell RB (1970) Zum Problem der Faktorentifizierung. Psychol Beiträge 12:214-226
30. Smousse PE, Feinberg M, Carroll BJ, Park MH, Rawson SG (1981) The Carroll rating scale für depression. II. Factor-analysis of feature profiles. Br J Psychiatry 138:201-204
31. Spitzer RL, Endicott J, Robins E (1978) Research diagnostic criteria: Rational and reliability. Arch Gen Psychiatry 35:773-785
32. Weissmann MM, Klerman L, Prusoff BA, Sholomskas D, Padian N (1981) Depressed outpatients. Arch Gen Psychiatry 38:51-55
33. Welner J (1972) Eine internationale multizentrische Doppelblindprüfung eines neuen Antidepressivums: In: Kielholz P (Hrsg) Depressive Zustände. Huber-Verlag, Bern
34. Wing JK, Cooper JE, Sartorius N (1978) Present state examination (Deutsche Bearbeitung). Beltz, Weinheim