

## Methodische Untersuchungen zur Hamilton-Depression-Skala

U. BAUMANN

Institut für Psychologie, Universität Kiel, Olshausenstraße 40-60,  
D-2300 Kiel, Bundesrepublik Deutschland

Eingegangen am 25. März 1976

### METHODOLOGIC STUDIES OF THE HAMILTON RATING SCALE FOR DEPRESSION

**SUMMARY.** Within the scope of psychopharmacologic therapies, the Hamilton Rating Scale for Depression was applied 3 times on each of 197 psychiatric patients. An extended version (24 symptoms) of the Hamilton Scale was employed, which included the original symptoms. Various formal analyses (frequency of symptoms, influence of sex, item analyses, estimation of reliability, factor analysis) allow the following conclusions: With the original version, the single factor solution is the most consistent; several factors are not acceptable. With the extended version (used in the U. S. A.) the single factor solution can be equally recommended, but the two factor solution could also be taken into consideration, whereby a distinction is made between retarded depression and somatic depression aspects. The total score of the Hamilton Scale is satisfactorily consistent and hardly dependent on sex.

**KEY WORDS:** Depression Scale - Hamilton Scale - Psychopathology - Frequency of Symptoms.

**ZUSAMMENFASSUNG.** Im Rahmen von Psychopharmaka-Therapien wurde die Hamilton-Depressions-Skala bei 197 psychiatrischen Patienten je dreimal verwendet; dabei wurde eine erweiterte Version (24 Symptome) der Hamilton-Skala benutzt, die die ursprüngliche miteinschließt. Verschiedene formale Analysen (Symptommhäufigkeiten, Einfluß von Geschlecht, Itemanalysen, Konsistenzschätzung, Faktorenanalysen) lassen folgende Schlußfolgerungen zu: Bei der ursprünglichen Fassung ist die einfaktorielle Auswertung am adäquatesten; mehrere Faktoren sind nicht anzunehmen. Bei der erweiterten Version (vor allem in USA gebraucht) ist zwar die Einfaktorenlösung ebenso zu empfehlen, doch könnte man auch die Zweifaktorenlösung in Betracht ziehen, bei der zwischen gehemmt-depressiven und somatisch-depressiven Aspekten unterschieden wird. Der Gesamtwert bei der Hamilton-Skala ist von befriedigender Konsistenz (Reliabilität) und kaum vom Geschlecht abhängig.

**SCHLÜSSELWÖRTER:** Depressions-Skala - Hamilton-Skala - Psychopathologie - Symptommhäufigkeit.

## 1. EINLEITUNG<sup>1</sup>

In der psychiatrischen Forschung und Praxis wird eine Vielzahl von Selbst- und Fremdbeurteilungsskalen zur Erfassung depressiver Symptomatik verwendet (Gebhardt et al., 1969; Kerekjarto et al., 1970; Pichot, 1974). Dabei ist die von Hamilton entwickelte Fremdbeurteilungs-Skala (Hamilton-Skala, 1960, 1967, 1974) besonders stark verbreitet und liegt in verschiedenen Sprachen vor. Die ursprüngliche Fassung der Hamilton-Skala enthält 21 Symptome mit jeweils 3 (0-2) oder 5 (0-4) Einstufungskategorien. Aus den ersten 17 Items wird durch Zusammenzählen der Gesamtwert gebildet. Die restlichen 4 Symptome sind zwar von klinischer Bedeutung, werden aber nicht weiter ausgewertet. Da im Normalfall das Rating von 2 Untersuchern durchgeführt werden sollte, werden die Ergebnisse beider Rater für den Gesamtwert addiert (nicht Durchschnittswert!). Da andererseits die Skala oft von nur 1 Rater angewendet wird, sollte man daher immer angeben, ob die Werte auf 1 oder 2 Ratern basieren.

Obwohl die Skala meistens eindimensional ausgewertet wird, diskutiert Hamilton (1967) auch mehrere Faktoren, die er in getrennten Faktorenanalysen von Patienten- und Patientinnen-Stichproben gewonnen hat. Der erste Faktor stellt bei Männern und Frauen einen allgemeinen Depressionsfaktor dar. Faktor 2 wird als bipolarer Faktor "Retarded vs. agitated depression" interpretiert. Die nach Hamilton postulierte Bipolarität ist aber vor allem bei der Männer-Stichprobe zweifelhaft, da zwar 3 Symptome mit  $a_{2j} \geq 0.67$  positiv laden, aber nur 1 Symptom negativ mit  $a_{2j} = -0.36$ . (übrige  $a_{2j} \leq -0.16$ ); bei der Frauen-Stichprobe ist die Bipolarität eher nachzuweisen. Eindeutig bipolare Faktoren müßten sowohl hohe positive als auch hohe negative Ladungen aufweisen; Nullladungen zeigen dagegen, daß Items mit dem Faktor nichts zu tun haben und für die Interpretation des Faktors nicht heranzuziehen sind.

Während Faktor 2 in beiden Stichproben inhaltlich in etwa gleich ist, kann dies von den Faktoren 3 und 4 nicht behauptet werden. Im weiteren ist noch kritisch anzumerken, daß Hamilton zwar 6 Faktoren extrahiert und orthogonal rotiert (Varimax-Kriterium); andererseits werden aber nur 4 Faktoren interpretiert. Es wäre daher angezeigt gewesen, nur 4 Faktoren zu rotieren.

Neben der ursprünglichen Version (21 Symptome; Gesamtwert: 17 S.) wird in USA eine 24 Symptom-Version vom National Institute of Mental Health verwendet; es fehlen aber Angaben, wie die 3 zusätzlichen Symptome in die Struktur der bisherigen einzuordnen sind.

## 2. FRAGESTELLUNG

In unserer Arbeit wird die Hamilton-Skala (Symptome s. Tabelle 2) auf ihre formale Genauigkeit hin überprüft. Gleichzeitig sollen die

---

<sup>1</sup> Herrn Prof. Dr. med. J. Angst, der das Material für die vorliegenden Analysen zur Verfügung stellte, sei diese Arbeit zum Danke gewidmet für die vielen Forschungsmöglichkeiten, die der Verfasser während seiner Tätigkeit in der Forschungsdirektion der Psychiatrischen Universitätsklinik Zürich erhalten hat.

21- und 24-Symptomen-Versionen einander gegenübergestellt werden. Im einzelnen sind folgende Fragen von Interesse:

- Wie häufig und in welcher Ausprägung kommen die einzelnen Symptome bei hospitalisierten Patienten vor und wie verändert sich die Häufigkeit resp. Intensität im Laufe einer Psychopharmaka-Therapie? Zusätzlich ist der Einfluß vom Geschlecht auf die Ergebnisse zu diskutieren.
- Wie gut genügen die beiden Versionen den Kriterien der klassischen Testtheorie (Trennschärfe, Konsistenz)?
- Wie sieht die faktorielle Struktur der beiden Versionen (21, 24) aus?

Um die Ergebnisse breiter abzusichern, sollen jeweils mehr als eine Stichprobe herangezogen werden (Kreuzvalidierung).

Die Berechnungen wurden im Universitäts-Rechenzentrum der Universität Zürich (IBM 370-155) mit verschiedenen Programmen der Programmbibliothek PSYCHLIB (U. Baumann) durchgeführt.<sup>2</sup>

### 3. STICHPROBE

#### 3.1. Methodische Überlegungen

Die gesamte Stichprobe umfaßt 197 Patienten, die an den Tagen 0 (Ausgangsbefund), 10 und 20 im Rahmen von verschiedenen Pharmatherapien u. s. auch mit der Hamilton-Skala untersucht worden sind. Bei diesen Studien, die zwischen 1968 und 1974 stattfanden, haben verschiedene Rater die Interviews und Ratings durchgeführt. Für die Hamilton-Skala standen dabei Formulare mit den englischen Symptomenamen und Manuals mit Erläuterungen in deutscher Sprache zur Verfügung. Die Raters haben im Rahmen ihrer Ausbildung Trainingskurse zur Interviewtechnik und zur Erhebung des psychopathologischen Befundes erhalten.

Bei einzelnen der 197 Patienten mußte z. T. die Therapie wegen Unverträglichkeit abgebrochen werden (Drop-Outs), weshalb am 10. und 20. Tag nicht mehr bei allen Patienten eine Hamilton-Skala vorliegt.

Ausgangsbefund (Tag 0): 197 Patienten

Tag 10: 194 Patienten

Tag 20: 163 Patienten.

Aus methodischen Gründen wird bei den Drop-Outs der letzte verfügbare Befund auch für diejenigen Ratingtage verwendet, in denen der Patient nicht untersucht worden ist. Damit vermeidet man eine zu positive Bewertung von Psychopharmaka-Wirkungen, die bei Nichtberücksichtigung von Drop-Outs erfolgen würde. Ein Vergleich der Anzahl vorhandener Symptome am Tag 20 bei den 163 und den 197 (163 + 34 nachgeführte Drop-Outs) Patienten zeigt dies deutlich (Tabelle 2, Kap. 4), indem ohne Drop-Outs die Symptommhäufigkeiten

---

<sup>2</sup> Fräulein S. Zulauf und Fräulein G. Schneidewind, die bei der Aufbereitung der Daten und den Berechnungen mitgeholfen haben, sei dafür herzlich gedankt.

durchschnittlich geringer sind als unter Berücksichtigung der Drop-Outs. Bei 19 der 24 Symptomen trifft dies zu, was gem. Binominaltest mit  $P < 0.05$  (einseitig) signifikant ist. Wir werden daher bei den Ergebnissen am Tag 20 meistens die Stichprobe verwenden, bei der die Drop-Outs berücksichtigt sind ( $N = 197$ ). Da am Tag 10 nur 3 Drop-Outs vorliegen, haben wir auf eine getrennte Auszählung der Häufigkeiten verzichtet und nur mit der Gesamtstichprobe ( $N = 197$ , incl. 3 Drop-Outs) gerechnet.

Bei unserem Material wurde die Hamilton-Skala jeweils von 1 Rater erhoben; bei Vergleichsdaten, die von 2 Ratern stammen, werden wir daher die entsprechenden Werte durch 2 dividieren.

### 3. 2. Soziographische und klinische Daten

Wie schon in Kapitel 2 betont, werden die Ergebnisse jeweils an verschiedenen Unterstichproben überprüft. Wir haben folgende Stichproben gebildet:

- Gesamtstichprobe  $N = 197$  (HG)
- Unterteilung nach Geschlecht: männlich  $N = 63$  (HM)  
weiblich  $N = 134$  (HW)
- Unterteilung nach Zufall:  $N = 99$  (H1)  
 $N = 98$  (H2)

In Tabelle 1 sind die soziographischen und klinischen Daten der Gesamtstichprobe und der männlichen (HM) und weiblichen (HW) Stichprobe aufgeführt. Das Patientengut ist bezüglich der verschiedenen Daten breit gestreut. Die beiden Stichproben HM (männlich) und HW (weiblich) unterscheiden sich nur in der Dauer der Krankheit, gerechnet von der Erstmanifestation an ( $\chi^2 = 10.03$ ,  $df = 5$ ,  $P < 0.05$ ), indem die Frauen länger krank sind; dagegen ist die Zahl der früheren Manifestationen gleich. Bei den meisten Patienten und Patientinnen liegt eine eher gehemmt depressive Symptomatik im Vordergrund, was bei der Interpretation der Ergebnisse zu berücksichtigen ist.

## 4. ERGEBNIS DER SYMPTOMAUSZÄHLUNG

### 4. 1. Häufigkeit der Symptome (Gesamtstichprobe HG)

In Tabelle 2 sind für die einzelnen Symptome die Mittelwerte und der Prozentsatz an Patienten mit vorhandenen Symptomen aufgeführt (Ratingtag 0, 10, 20 incl. Drop-Outs). Von den 21 (24 Symptomen) kommen am Ausgangsbefund 15 (18) bei über 50% der Patienten vor; am Tag 10 sind es noch 12 (15), am Tag 20 ( $N = 197$ ) 9 (12). Die drei zusätzlichen Symptome der 24-Version treten also an allen drei Ratingtagen relativ häufig auf. Folgende 5 Symptome sind durchgehend sehr häufig. (Ausgangsbefund  $P \geq 92\%$ , Ratingtag 20:  $P \geq 66\%$ )

1 Depressive Stimmung	100%	(0)	95	(10)	85	(20)
7 Beeinträchtigung in Arbeit und Interesse	95%	(0)	85	(10)	78	(20)
10 Angst (psych.)	92%	(0)	85	(10)	75	(20)
22 Hilflosigkeit	92%	(0)	78	(10)	66	(20)
23 Hoffnungslosigkeit	93%	(0)	79	(10)	69	(20)

Tabelle 1. Soziographische und klinische Daten der Gesamtstichprobe (HG), Stichprobe HM (männlich) und HW (weiblich)

Stichprobe	HG	HM	HW
N	197	63	134
<u>Alter:</u>			
16-40	38	12	26
41-60	108	33	75
61-81	51	18	33
M (s)	51.9 (13.2)	51.3 (13.1)	52.2 (13.3)
Median, Range	54 (16-18)	55 (16-69)	55 (18-81)
<u>Krankheitsdauer seit Erstmanifestation</u>			
unbekannt	10	4	6
< 1 Jahr	27	13	14
1-5	57	20	37
6-10	29	11	18
11-20	33	9	24
> 20	41	6	35
Median, Range	7 (0-51)	4 (0-45)	8 (0-51)
<u>Zahl der früheren Manifestationen</u>			
unbekannt	11	4	7
unsicher, aber mehrere	28	10	18
unsicher, aber mind. 10	11	5	6
0	33	12	21
1, 2	58	18	40
3-5	35	9	26
6-10	12	1	11
11-20	9	4	5
Range	0-20	0-20	0-20
<u>Diagnosen (ICD)</u>			
2957 Schizoaff. Psychos.	32	5	27
2960 Involutionenidepress.	43	19	24
2962 Monopolare Depress.	65	20	45
2963 Bipolare Depress.	20	7	13
2980 Reakt. depr. Psychose	6	3	3
3004 Depr. Neurose	12	4	8
andere Diagnosen	19	5	14

Tabelle 2. Mittelwerte und Prozentsatz der Patienten mit vorhandenem Symptom für Ratingtag 0, 10, 20 bei Gesamtschichprobe (HG) = 197

Symptom	Theoret. Max.	Mittelwert M			Prozent "Symptom vorhanden"			
		0	10	20	0	10	20 <sup>a</sup>	20 <sup>b</sup>
1. Depressive Stimmung	4	3.09	2.28	1.94	100	93	85.3	84.7
2. Schuldgefühle	4	1.62	1.16	0.95	66	55	54.2	53.6
3. Suicid	4	1.48	0.54	0.44	73	35	27.4	23.3
4. Einschlafstörungen	2	1.31	0.78	0.65	78	57	48.2	46.6
5. Durchschlafstörungen	2	1.34	0.94	0.71	81	63	52.3	50.3
6. Schlafstörungen am Morgen	2	1.31	0.87	0.64	79	60	46.7	48.5
7. Beeinträchtigung in Arbeit und Interesse	4	2.89	1.93	1.68	95	85	77.7	76.1
8. Hemmung	4	1.78	1.23	1.09	85	69	58.9	56.4
9. Erregung	4	1.00	0.83	0.75	58	51	51.3	48.5
10. Angst (psych.)	4	2.39	1.71	1.48	92	85	75.1	75.5
11. Angst (somat.)	4	1.12	0.88	0.81	58	50	47.7	48.5
12. Gastro-intestinale Symptome	2	1.10	0.88	0.76	73	61	54.8	54.6
13. Allgem. somat. Symptome	2	1.18	0.92	0.76	80	66	55.8	53.4
14. Genitalsymptome (Libidoverlust)	2	0.52	0.43	0.29	31	27	20.3	20.9
15. Hypochondrie	4	0.59	0.48	0.51	30	27	28.9	27.6
16. Mangel an Krankheitseinsicht	2	0.82	0.68	0.60	53	49	43.1	38.7
17. Gewichtsverlust	2	0.65	0.18	0.12	46	13	8.6	7.4
18. Tagesschwankungen	2	0.82	0.66	0.52	53	45	39.1	39.9
19. Derealisation und Depersonalisat.	4	0.34	0.26	0.21	16	13	11.2	9.8
20. Wahnsymptome	4	0.99	0.69	0.65	38	26	24.9	22.1
21. Zwangssymptome	2	0.22	0.16	0.16	14	12	12.2	11.7
22. Hilflosigkeit	4	2.52	1.69	1.44	92	78	66.0	64.4
23. Hoffnungslosigkeit	4	2.47	1.76	1.53	93	79	69.0	67.5
24. Gefühl der Wertlosigkeit	4	1.91	1.38	1.11	78	63	55.3	54.0

<sup>a</sup> N = 197 (incl. Drop-Outs)<sup>b</sup> N = 163 (excl. Drop-Outs)

- Symptom ohne Berücksichtigung der Drop-Outs häufiger als mit Berücksichtigung

Diese 5 Symptome stellen, da sie beim Ausgangsbefund bei fast allen Patienten vorkommen ( $P \geq 92\%$ ) notwendige Bedingungen für das depressive Zustandsbild dar. Im Gegensatz dazu sind folgende 5 Symptome relativ selten ( $P \leq 38\%$ ):

14 Genitalsymptome	31% (0)	27% (10)	20% (20)
15 Hypochondrie	30% (0)	27% (10)	29% (20)
19 Depers. u. Derealisation	16% (0)	13% (10)	11% (20)
20 Wahnsymptome	38% (0)	26% (10)	25% (20)
21 Zwangssymptome	14% (0)	12% (10)	12% (20)

Die Seltenheit der drei letzten Symptome wird auch von Hamilton (1967) betont, 12 weitere Symptome treten relativ häufig auf, indem sie beim Ausgangsbefund  $P \geq 50\%$ , am 20. Tag  $P \geq 39\%$  haben: Schuldgefühle (2), Schlafstörungen (4-6), Hemmung (8), Erregung (9), Angst somat. (11), allg. somat. Symptome (13), Mangel an Krankheitseinsicht (16), Tagesschwankungen (18), Gefühl der Wertlosigkeit (24).

Die beiden Symptome Suicid (3) und Gewichtsverlust (17) nehmen von 73% auf 27% ab (Suicid), resp. von 46% auf 9% (Gewichtsverlust) und lassen sich daher nicht zu den übrigen Symptomen einreihen. Bei diesen beiden Symptomen ist auch die Abnahme im Verlaufe der Therapie am stärksten, während andere Symptome z. T. in ihrer Häufigkeit nicht abnehmen (Hypochondrie, Zwangssymptome). Wie der nächste Abschnitt zeigt, ist aber neben der Häufigkeit eines Symptoms auch dessen Ausprägung zu berücksichtigen.

#### 4.2. Symptomausprägung (Gesamtstichprobe HG)

Ein Vergleich der Ausprägungsgrade verschiedener Symptome untereinander ist insofern problematisch, als die Einstufungsbreite zum einen 0-2, zum anderen 0-4 beträgt. Dazu kommt, daß z. T. die logischen Bedeutungen einander entsprechender Werte verschieden sind (Beisp. Symptom 2 und 8) und die Skalierung nicht immer eindimensional erfolgt (Beisp. Symptom 18, aber auch Symptom 8) (vgl. Baumann et al., 1975).

Beispiel (Graduierung nach Hamilton, 1967):

Symptom Hemmung (Nr. 8; 0-4): Wert 4 bedeutet, daß Interview wegen zu großer Hemmung nicht möglich ist, Symptome nicht eingestuft werden können (formaler Aspekt eines Symptoms).

Symptom Schuldgefühle (Nr. 2; 0-4): Wert 4 bedeutet Schuldwahn mit oder ohne Halluzinationen (inhaltlicher Aspekt: maximale Ausprägung).

Symptom Tagesschwankungen (Nr. 18; 0-2): Wert 2 bedeutet, daß Tagesschwankungen eindeutig vorhanden sind, während Wert 1 sie als fraglich vorhanden charakterisiert. Werte 0 und 2 betreffen das Vorhandensein, Wert 1 die Entscheidungssicherheit.

Analysiert man trotz der geäußerten Bedenken die Mittelwerte der Symptome, so findet man - wie zu erwarten - einen engen Zusammenhang zwischen Häufigkeit eines Symptoms und der durchschnittlichen Ausprägung; die im vorherigen Kapitel als häufig aufgeführte Symptome weisen auch hohe Mittelwerte auf, die selteneren niedrige Mittelwerte (Symptome mit 0-2 Skalierung sind in ihrem Wert zu verdoppeln, um Vergleiche durchzuführen). Bei mittleren oder niedrigen Mittelwerten ist daher zu prüfen, ob es sich um ein seltenes, aber im Auftretensfall stark ausgeprägtes Symptom handelt oder um ein häufiges, das meistens schwach ausgeprägt ist. So haben Erregung (Nr. 9;  $M = 1.00$ ,  $P = 58\%$ ) und Paranoide Symptome

(Nr. 20;  $M = 0.99$ ,  $P = 38\%$ ) gleich hohe Mittelwerte, unterscheiden sich aber in der Häufigkeit. Beim ersteren handelt es sich um ein häufigeres, aber nicht so stark ausgeprägtes Symptom, beim letzteren um ein selteneres, aber im Auftretensfall stärker ausgeprägtes Symptom.

Betrachtet man bei den einzelnen Symptomen den Mittelwertsverlauf über die drei Ratingtage, so ist in allen Symptomen (Ausnahme: Genitalsymptome (Nr. 14)) die Abnahme zwischen dem Ausgangsbefund und Tag 10 größer als zwischen Tag 10 und 20. Für eine differenziertere Interpretation müßten aber die Häufigkeiten bei den einzelnen Symptomen mitberücksichtigt werden.

#### 4. 3. Einfluß von Geschlecht

Nach Hamilton (1967) sind die einzelnen Symptome bei Männern und bei Frauen verschieden stark ausgeprägt (Vergleich der Mittelwerte). Insbesondere haben Männer höhere Mittelwerte in den Symptomen "Beeinträchtigung in Arbeit und Interesse (7)", Einschlafstörungen (4), Hemmung (8) und niedrigere Werte bei den allgemeinen somatischen Symptomen (13), Angst (Psych., 10), depressive Stimmung (1) und Erregung (9). Numerisch sind 10 der 17 Mittelwerte (Symptome 1-17) bei Männern höher. In unserem Material lassen sich ähnliche Tendenzen feststellen, indem beim Ausgangsbefund sogar 13 von 16 Symptomen (1 Symptom ohne Unterschied) bei den Männern stärker ausgeprägt sind (Binomialtest  $P < 0.05$ ). Am Tag 10 und 20 verwischen sich dagegen die Unterschiede.

In der Häufigkeit der Symptome sind dagegen kaum Unterschiede zwischen Männern und Frauen zu finden (Chiquadratstest,  $P > 0.01$ ). Ausnahmen bildet das Genitalsymptom, das bei Männern häufiger vorhanden ist (Ausgangsbefund  $P < 0.01$ , Tag 10  $P < 0.10$ , Tag 20  $P < 0.05$ ) und das Symptom "Derealisation und Depersonalisation", das bei Frauen häufiger ist (Tag 10 und 20  $P < 0.10$ ).

Die Symptome kommen also bei Männern und Frauen gleich häufig vor, bei Männern sind aber die Durchschnittswerte der Ausprägung - insbesondere am Ausgangsbefund - höher als bei Frauen.

Betrachtet man die Symptomkonfigurationen (Profil) der beiden Stichproben HM und HW, so sind zwischen Männern und Frauen nur geringe Unterschiede feststellbar. Zur Überprüfung wurde die Rangkorrelation zwischen dem Durchschnittsprofil (Symptommittelwerte) der Männer und dem der Frauen gerechnet (17 Symptome wegen Vergleich mit Hamilton; Wert von Symptomen mit Range 0-2 wurde verdoppelt):

Rangkorrelation M/W bei Hamilton (1967)	$r = 0.88$
" " " eigenes Material	$r = 0.90$ (Tag 0)
" " " " "	$r = 0.93$ (Tag 10)
" " " " "	$r = 0.97$ (Tag 20).

Vergleicht man die Symptomprofile von Hamilton (1967) und unserem Material, so ist die Übereinstimmung geringer:

Männer: Hamilton/eigenes Material Tag 0	$r = 0.62$
" " " Tag 10	$r = 0.54$
" " " Tag 20	$r = 0.45$
Frauen: Hamilton/eigenes Material Tag 0	$r = 0.63$
" " " Tag 10	$r = 0.56$
" " " Tag 20	$r = 0.53$



Wie weit die geringe Übereinstimmung der Profilkonfigurationen durch verschiedene Patientenstichproben bedingt ist, läßt sich nicht klären, da bei Hamilton (1967) keine weiteren Angaben zu den Patienten vorliegen. Eine Analyse der Diskrepanzen zwischen den beiden Rangfolgen (Ausgangsbefund) der Symptommittelwerte (hoher Mittelwert: niedriger Rang) gibt aber Anhaltspunkte zu den Profilunterschieden. So stehen bei den englischen Patientinnen das Symptom Gewichtsverlust und Genitalsymptome stärker im Vordergrund (niedrigere Rangplätze) als bei unserem Material, während es sich beim Symptom "Mangel an Krankheitseinsicht" umgekehrt verhält. Einschlafstörungen und Gewichtsverlust sind bei den englischen Patienten von größerer Bedeutung, Durchschlafstörungen dagegen von geringerer als in unserem Material.

Die folgenden Korrelationen zeigen, daß bei den Zürcher-Patienten die Symptomkonfiguration über alle 3 Ratingtage hinweg ziemlich konstant bleibt:

Vergleich Tag 0/10	$r = 0.94$ (Männer),	$r = 0.86$ (Frauen)
" " 0/20	$r = 0.88$ "	$r = 0.86$ "
" " 10/20	$r = 0.95$ "	$R = 0.97$ "

## 5. MITTELWERT UND VERTEILUNG DER GESAMTSKALA

Hamilton (1967) gibt für die ursprüngliche Version (21 Symptome, aber nur 17 verrechnet) folgende Mittelwerte für den Gesamtwert bei psychiatrischen Patienten an:

Männer:  $M = 23.17$   $s = 11.72$   $N = 152$

Frauen:  $M = 22.83$   $s = 11.72$   $N = 120$

Der Unterschied ist mit  $P > 0.05$  nicht signifikant (t-Test).

Die Höhe der Mittelwerte ist auch in anderen Untersuchungen ähnlich (auf 1 Rater bezogener Gesamtwert):

$M = 21.0$   $s = 14.8$   $N = 51$  (Fahy, 1974)

$M = 21.8$   $N = 53$  (Waldron et al., 1965)

$M = 21.3$   $N = 50$  (Rose et al., 1965)

Auch in unserem Material ist der Wert im Ausgangsbefund bei der Gesamtstichprobe in der gleichen Größenordnung (ursprüngliche Version: 17 Symptome verrechnet):

Tag 0  $M = 24.19$   $s = 7.32$  24-Symptome:  $M = 33.46$   $s = 9.94$

Tag 10  $M = 16.72$   $s = 8.89$  "  $M = 23.32$   $s = 12.76$

Tag 20  $M = 14.18$   $s = 8.86$  "  $M = 19.80$   $s = 12.97$

Männer haben zwar - wie bei Hamilton - an allen drei Ratingtagen numerisch einen höheren Mittelwert als Frauen, doch ist der Unterschied nur beim Ausgangsbefund mit  $P < 0.10$  bedeutsam (t-Test).

Analysiert man Schiefe und Exzess, so haben wir an allen drei Ratingtagen bei den Stichproben HG, HM, HW, H1 und H2 keine signifikante Abweichung von der Normalverteilung ( $P > 0.01$ ). Dies ist für statistische Berechnungen natürlich besonders vorteilhaft, da damit parametrische Verfahren zugänglich sind.

## 6. ERGEBNIS DER ITEMANALYSEN

Die Ergebnisse der Itemanalysen (Trennschärfen: Produkt-Moment-Korrelation mit part-whole-Korrektor, Konsistenzkoeffizienten nach Cronbach; vgl. Lienert, 1969) für die Stichproben H1 und H2 (2 Zufallsstichproben aus der

Tabelle 3. Ergebnis der Itemanalysen bei den Stichproben H1, H2 (Trennschärfe  $r_{it}$ , Konsistenz  $\alpha$ )

Ursprüngliche Version (21 Symptome, Auswertung von 17 Symptomen)

	Tag 0		Tag 10		Tag 20	
	H1	H2	H1	H2	H1	H2
Anzahl Items mit						
$r_{it} \quad P > 0.01$	6	4	2	1	2	1
$\bar{r}_{it} /$	0.33	0.33	0.48	0.46	0.46	0.51
$\alpha$	0.73	0.73	0.86	0.86	0.85	0.88

Version 24 (Analyse sämtlicher Symptome)

Anzahl Items mit						
$r_{it} \quad P > 0.01$	4	7	3	2	2	3
$\bar{r}_{it} /$	0.33	0.31	0.48	0.45	0.48	0.49
$\alpha$	0.78	0.77	0.90	0.88	0.90	0.91

Gesamtstichprobe) sind in Tabelle 3 aufgeführt, und zwar für die ursprüngliche Version (21 Symptome, Auswertung der ersten 17) und die erweiterte Fassung (24 Symptome).

In beiden Versionen erweist sich Item 14 (Genitalsymptome) als trennschwach ( $r_{it} P > 0.01$ ). Tagesschwankungen (Nr. 18; Rating 0, 10, 20) und Zwangssymptome (Nr. 21; Rating 0, 20) sind in der 24-Symptom-Version zusätzlich noch von geringer Trennschärfe.

Bei Selbsteinstufungen und dichtomen Items gelten mittlere Trennschärfen von  $0.32 \leq \bar{r}_{it} \leq 0.44$  als befriedigend (Skalen Extraversion und Neurotizismus; Baumann & Dittrich, 1976). Diese Größenordnung wird auch bei unseren Daten in der Fremdeinstufung erreicht (dabei ist aber zu berücksichtigen, daß mehrstufige Items eher zu höheren Trennschärfen führen sollten als dichtome (vgl. auch Baumann et al., 1974). Auch die Konsistenzkoeffizienten (Reliabilitätsschätzung) können - unter Berücksichtigung der Skalenlängen (17, 24 Symptome) als befriedigend bezeichnet werden; erwartungsgemäß ist die längere Skala etwas konsistenter:  $0.73 \leq \alpha_{17} \leq 0.88$ ,  $0.77 \leq \alpha_{24} \leq 0.91$ .

Die verschiedenen Item- und Skalenkennwerte ändern sich zwar von Tag 0 zu Tag 20 im Sinne einer Verbesserung, doch sind die Unterschiede - vor allem in der Konsistenz - nicht so gravierend, daß aus meßtheoretischer Sicht Einwände begründet wären. Von den formalen Gesichtspunkten her ist die 17-Version der 24-Version ebenbürtig und daher als kürzere und ökonomischere Form vorzuziehen.

## 7. FAKTORENANALYSEN

### 7.1. Ursprüngliche Version (Auswertung von 17 Symptomen)

Die Frage nach der Faktorenzahl ist bei der Hamilton-Skala insofern von Bedeutung, als die Skala eindimensional ausgewertet wird, Hamilton (1967) aber mehrfaktorielle Lösungen diskutiert. Die Zahl der Faktoren läßt sich im allgemeinen nicht eindeutig bestimmen (Pawlik, 1968; Überla, 1968), doch kann man Argumente anführen, welche Lösung welche Vorteile hat. Zur Klärung dieses Problems haben wir bei zwei Stichproben (H1 und H2) Hauptkomponentenanalysen über die ersten 17 Symptome gerechnet (Diagonale = 1, Produkt-Moment-Korrelation). Dies erlaubt uns, die Stichprobenabhängigkeit der Lösungen abzuschätzen. In Tabelle 4 haben wir die Eigenwerte für die drei Ratingtage aufgeführt.

Die Ähnlichkeit der verschiedenen, nach Varimax rotierten Lösungen des Ausgangsbefundes läßt sich nach Fischer & Roppert (1964) berechnen:

Faktorenzahl	1	2	3	4	5	6
Ähnlichkeit r	0.96	0.66	0.85	0.75	0.81	0.85

Es zeigt sich dabei eine stärkere Stichprobenabhängigkeit der Zweierlösung.

Nach dem Scree-Test (Cattell, 1966) sind 1-2, evtl. 4-6 Faktoren anzunehmen (je nach Untersuchungstag). Bei den höheren Lösungen (4-6 Faktoren) haben wir aber meistens pro Faktor zu wenig substantiell ladende Items ( $a_{ij} > 0.4$ ); z. T. müßten die Faktoren durch 2 Symptome charakterisiert werden, was nicht ausreicht. Da die Zweierlösung stark stichprobenabhängig ist, erscheint uns die Einfaktorenlösung am sinnvollsten. Diese hängt auch kaum vom Geschlecht ab (Faktorenstrukturvergleich Ausgangsbefund HM/HW:  $r = 0.94$ ).

Tabelle 4. Eigenwerte der ursprünglichen Version (Auswertung von 17 Symptomen): PC-Analyse

Nr.	Rating 0		Rating 10		Rating 20	
	H1	H2	H1	H2	H1	H2
	$\lambda$	$\lambda\%$	$\lambda$	$\lambda\%$	$\lambda$	$\lambda\%$
1	3.80	22.4	3.76	22.1	5.49	32.3
2	1.92	11.3	2.13	12.5	1.67	9.8
3	1.85	10.9	1.58	9.3	1.46	8.6
4	1.68	9.9	1.38	8.1	1.33	7.8
5	1.15	6.8	1.16	6.8	1.11	6.5
6	0.99	5.8	1.02	6.0	0.97	5.7
$\lambda_1 > 1$	5	6	5	5	5	4

Tabelle 5. Ladungen der rotierten Faktormatrix (Varimax)

Symp- tom	Rating 0		Rating 10		Rating 20	
	H1	H2	H1	H2	H1	H2
1	0.55	0.40	0.81	0.76	0.82	0.81
2	0.21	0.10	0.51	0.39	0.45	0.50
3	0.29	0.19	0.52	0.57	0.46	0.55
4	0.60	0.63	0.49	0.54	0.39	0.58
5	0.56	0.68	0.59	0.60	0.46	0.59
6	0.56	0.57	0.51	0.46	0.40	0.48
7	0.63	0.59	0.77	0.73	0.80	0.80
8	0.40	0.46	0.72	0.68	0.66	0.67
9	0.39	0.38	0.40	0.47	0.50	0.50
10	0.66	0.51	0.67	0.78	0.76	0.80
11	0.58	0.66	0.56	0.61	0.60	0.64
12	0.61	0.52	0.69	0.49	0.67	0.70
13	0.43	0.57	0.66	0.62	0.75	0.59
14	0.13	0.19	0.20	0.21	0.28	0.31
15	0.17	0.47	0.29	0.36	0.32	0.41
16	0.11	0.24	0.46	0.54	0.43	0.55
17	0.54	0.27	0.42	0.32	0.36	0.31

Die Ladungsmatrizen für die Stichproben H1 und H2 und die drei Ratingtage sind in Tabelle 5 abgedruckt.

Vergleicht man die Ladungshöhen an den einzelnen Ratingtagen miteinander, so tragen die Symptome depressive Stimmung (1), Beeinträchtigung in Arbeit und Interesse (7), psychische Angst (10) am meisten zur Faktorenidentifikation bei. Weiterhin sind bedeutsam Durchschlafstörungen (5), Hemmung (8), somatische Angst (11), gastrointestinale Symptome (12) und allgemein somatische Symptome (13). Wenig zu tun mit dem Faktor haben Erregung (9), Mangel an Krankheitseinsicht (16), Schuldgefühle (2), Gewichtsverlust (17), Hypochondrie (15), Genitalsymptome (14). Bei Hamilton (1967) finden wir im ersten unrotierten Faktor gegenüber unserer Lösung eine leichte Akzentverschiebung, indem bei ihm die Symptome Schuldgefühle (2), Suicid (3) und Mangel an Krankheitseinsicht (16) mit relativ hohen Ladungen hinzukommen. Dagegen treten bei ihm psychische (10) und somatische (11) Angst und allgemeine somatische Symptome (13) eher in den Hintergrund. In beiden Fällen hat aber das Symptom Erregung (9) niedrige Ladungen. Während in unserer Lösung der zu interpretierende Faktor gehemmt-depressive und somatisch-depressive Aspekte umfaßt, liegt bei Hamilton's Ergebnissen der Akzent eher auf der gehemmt-depressiven Symptomatik.

## 7.2. Erweiterte Version (24 ausgewertete Symptome)

Nach ähnlichen Gesichtspunkten wie in Kap. 7.1. wurde auch die 24-Version analysiert. Aus Platzgründen sind nur die Eigenwerte des Ausgangsbefundes aufgeführt (Tabelle 6).

Ähnlich wie bei der ursprünglichen Version erklären die entsprechenden Eigenwerte am 10. Tag mehr Varianz als beim Ausgangsbefund und am Tag 20 Mehr als am Tag 10, was auf eine Zunahme der Homogenität hinweist (vgl. auch Zunahme der Konsistenzkoeffizienten Kap. 6). Beim Ausgangsbefund ist die Ähnlichkeit der Faktorenlösungen nach Fischer, Roppert (1964) in der Einer- und Zweierlösung am höchsten (Varimax-Rotation):

Faktorenzahl	1	2	3	4	5	6	7
Ähnlichkeit r	0.95	0.92	0.77	0.77	0.77	0.76	0.78

Tabelle 6. Eigenwerte der 24-Version (PC-Analyse)

Eigenwert	1	2	3	4	5	6	7	8	9
H1 $\lambda$	4.80	2.60	2.14	1.84	1.39	1.33	1.15	1.07	0.92
$\lambda\%$	20.0	10.8	8.9	7.7	5.8	5.5	4.8	4.4	3.8
H2 $\lambda$	4.45	3.16	1.67	1.53	1.45	1.22	1.14	1.07	1.00
$\lambda\%$	18.6	13.2	7.0	6.4	6.0	5.1	4.8	4.4	4.2

Auch bei der 24 Symptome umfassenden Version scheint uns die Ein-faktorenlösung am sinnvollsten zu sein, eventuell wäre aber auch die Zwei-faktorenlösung in Betracht zu ziehen. Bei der Einerlösung stehen - wie bei der 17-Version - im Vordergrund die depressive Stimmung (1), Beeinträchtigung in Arbeit und Interesse (7) und psychische Angst (10). Dazu kommen noch die neuen Symptome Hilflosigkeit (22) und Hoffnungslosigkeit (23). Zur Interpretation des Faktors sind noch die Symptome Verlangsamung (8), Gefühl der Wertlosigkeit (24) und gastrointestinale Symptome heranzuziehen. Im Gegensatz zur ursprünglichen Version (17 Symptome) tritt bei der 24-Lösung der somatische Aspekt eher in den Hintergrund.

Die Zweierlösung umfaßt einen mehr gehemmt depressiven und einen somatisch-depressiven Aspekt:

Gehemmt-depressiver Faktor: Hoffnungslosigkeit (23), Gefühl der Wertlosigkeit (24), Hilflosigkeit (22), depressive Stimmung (1), Schuldgefühle (2), Hemmung (8), psychische Angst (10), Gewichtsverlust (16), Wahnsymptome (20).

Somatisch-depressiver Faktor: Schlafstörungen (4-6), gastro-intestinale Symptome (12), allgemein-somatische Symptome (13), somatische Angst (11), Hypochondrie (15).

Der von Hamilton (1967) postulierte Faktor der gehemmt-agitierten Depression ließ sich nicht nachweisen; auch Binitie (1975) führt diesen Faktor nicht auf. Ob es sich daher um ein Zufallsprodukt (methodisch schwach abgesichert, vgl. Kap. 1) handelt, das einer Kreuzvalidierung nicht standhält, oder ob der betreffende Faktor nur in bestimmten, diagnostisch spezifisch zusammengesetzten Stichproben auftritt, läßt sich nicht entscheiden. Dazu müßten weitere Analysen an genau beschriebenen Stichproben vorliegen, die methodisch exakt durchgeführt worden sind.

## 8. DISKUSSION

Betrachten wir zuerst die Ergebnisse der ursprünglichen Version, die 21 Symptome umfaßt, von denen die ersten 17 den Gesamtwert bilden:

Die Faktorenanalysen an verschiedenen Stichproben und an verschiedenen Ratingtagen bestätigen die Richtigkeit einer eindimensionalen Auswertung, wie sie von Hamilton von Anfang an vorgeschlagen wird. Es ließen sich aber keine ausreichenden Anhaltspunkte für eine Aufsplitterung der Symptomatik in mehrere Faktoren finden. Von daher, aber auch aus methodischen Gründen (vgl. Kap. 1) ist der Faktor der gehemmt-agitierten Depression als empirisch wenig gesichert zu bezeichnen.

Die Konsistenzschätzungen weisen auf eine gute formale Genauigkeit der Hamilton-Skala hin; dabei fällt aber das Genitalsymptom wegen ungenügender Trennschärfe und starker Abhängigkeit vom Geschlecht auf. Von den 4 zusätzlichen Symptomen (18-21) sind vor allem Depersonalisation und Derealisation und die Zwangssymptome im Vergleich zu den anderen Symptomen selten, so daß es gerechtfertigt erscheint, sie beim Gesamtwert unberücksichtigt zu lassen. Bei den beiden anderen Symptomen (Wahnsymptome und Tagesschwankungen) wäre aber eine Berücksichtigung im Gesamtwert zu diskutieren.

Bei der erweiterten Fassung (24 Symptome) ist die eindimensionale Auswertung ebenso sinnvoll wie bei der ursprünglichen Version. Dabei haben aber die Items Genitalsymptome, Zwangssymptome und Tages-

schwankungen ungenügende Trennschärfen. Eine Aufgliederung der Symptomatik in einen gehemmt-depressiven und einen somatisch-depressiven Teil könnte von der Faktorenanalyse her vertreten werden; ob den beiden verschiedenen Aspekten aber auch verschiedene inhaltliche Aussagen zugeordnet werden können, wäre zu klären.

Grundsätzlich ist noch zu betonen, daß bei der Skalierung die logische Bedeutung der verschiedenen Skalenwerte (0-2, 0-4) eindimensional sein muß, damit die vorgesehene Auswertung adäquat ist. Eine Vermischung von verschiedenen Ebenen der Entscheidungslogik - wie es in der bisherigen Version geschieht - ist daher zu vermeiden.

Bei unserer Untersuchung haben wir es absichtlich vermieden, neue Skalen, resp. Auswertungsformen vorzuschlagen. Wir halten es nicht für sinnvoll, breit verwendete Skalen leicht zu modifizieren, da dies nachher meistens ein Durcheinander zwischen ursprünglicher Version und modifizierten Versionen zur Folge hat. Unsere Analysen sollen vielmehr Anregung sein zu breiteren Kontrolluntersuchungen von bestehenden Untersuchungsinstrumenten, denn erst diese erlauben eine empirisch breit abgestützte Revision von Ratingverfahren.

## LITERATUR

- Baumann, U., Dittrich, A.: Überprüfung der Fragebogendimension P (Psychotizismus) im Vergleich zu Extraversion und Neurotizismus. Z. klin. Psychol. 5, 1-23 (1976)
- Baumann, U., Meyer, B., Suter, U.: Formale und inhaltliche Überprüfung der PND-Skalen. Diagnostica 20, 3-21 (1974)
- Baumann, U., Schneidewind, G., Angst, J., Helmchen, H., Hippus, H.: Zur Entscheidungslogik bei Symptombeurteilungen. Arch. Psychiat. Nervenkr. 220, 225-235 (1975)
- Binitie, A.: A factor-analytical study of depression across cultures (African and European). Brit. J. Psychiat. 127, 559-563 (1975)
- Cattell, R. B.: The Scree test for the number of factors. Mult. Beh. Res. 1, 112-125 (1966)
- Fahy, T. J.: Depression in hospital and in general practice: a direct clinical comparison. Brit. J. Psychiat. 124, 240-242 (1974)
- Fischer, G., Roppert, J.: Bemerkungen zu einem Verfahren der Transformationsanalyse. Arch. ges. Psychol. 116, 98-100 (1964)
- Gebhardt, R., Helmchen, H., Hippus, H., Kerekjarto, M. v., Lienert, G. A., Renfordt, E.: Beurteilungsskalen und Merkmalslisten für depressive Syndrome. In: Hippus, H., Selbach, H. (Ed.): Das depressive Syndrom. München: Urban & Schwarzenberg 1969
- Hamilton, M.: A rating scale for depression. J. Neurol. Neurosurg. Psychiat. 23, 56-62 (1960)
- Hamilton, M.: Development of a rating scale for primary depressive illness. Brit. J. soc. clin. Psychol. 6, 278-296 (1967)
- Hamilton, M.: General problems of psychiatric rating scales (especially for depression). p. 125-138, In: Pichot, P. (Ed.) Basel-New York: Karger 1974
- ICD: Diagnoseschlüssel u. Glossar psychiatrischer Krankheiten. 2. Aufl. Berlin, Heidelberg, New York: Springer 1971



- Kerekjarto, M. v., Lienert, G. A. : Depressionsskalen als Forschungsmittel in der Psychopathologie. *Pharmakopsychiatrie Neuro-Psychopharmakologie* 3, 1-21 (1970)
- Lienert, G. A. : Testaufbau und Testanalyse. 3. Aufl. Weinheim: Beltz 1969
- Pawlik, K. : Dimensionen des Verhaltens. Bern: Huber 1968
- Pichot, P. (Ed.) : *Psychological Measurements in Psychopharmacology*. Basel-New York: Karger 1974
- Rose, J. T., Leaky, M. R., Martin, I. C. A., Westhead, T. T. : A comparison of nortriptyline and amitriptyline in depression. *Brit. J. Psychiat.* 111, 1101-1103 (1965)
- Überla, K. : *Faktorenanalyse*. Berlin, Heidelberg, New York: Springer 1968
- Waldron, J., Bates, T. J. N. : The management of depression in hospital. *Brit. J. Psychiat.* 111, 511-516 (1965)