Eur Arch Psychiatry Clin Neurosci (1990) 240:96–102

European Archives of Psychiatry and Clinical Neuroscience

Consistencies and Discrepancies in Self- and Observer-Rated Anxiety Scales

A Comparison Between the Self- and Observer-Rated Marks-Sheehan Scales

Margot Albus[1,*], Wolfgang Maier[2], David Shera[3], and Per Bech[4]

[1]Psychiatric Hospital, University of Munich, Nussbaumstrasse 7, W-8000 Munich 2, Federal Republic of Germany
[2]Department of Psychiatry, University of Mainz, Untere Zahlbacherstrasse 8, W-6500 Mainz, Federal Republic of Germany
[3]Massachusetts General Hospital, Biostatistic Unit, Fruit Street, Boston, MA 02114, USA
[4]Frederiksborg Amts Centralsygehus, 48, Dyrehavevej, DK-3400 Hillerod, Denmark

Received January 4, 1990

Summary. The Marks-Sheehan anxiety scales are the only scales where self-ratings and observer ratings are perfectly matched by the number, the content and the scaling of the items. Therefore these scales are an excellent tool to investigate the compatibility and to study different structures in self- and observer ratings. This was done by using the data material on the Marks-Sheehan scales of the Cross National Collaborative Panic Study. In this study 1168 outpatients who met the DSM-III criteria for panic disorder were randomly allocated either to alprazolam, imipramine or placebo treatment. Our results show that the Marks-Sheehan scales are highly comparable to other established rating scales. Both scales have a similar stable and consistent factor pattern for somatic symptoms but not for psychic symptoms of anxiety. Our results provide empirical evidence that the low consistencies between self- and observer ratings reported so far can be improved by using a rating scale which matches up item by item for self- and observer ratings. However, other sources of disagreement can only be solved by more elaborated item descriptions and training of patients as well as raters to obtain a better compatibility between self- and observer rating scales.

Key words: Hamilton anxiety scale – Marks-Sheehan scales – SCL-90 – Self-ratings – Observer ratings

Introduction

Both, self-rating and observer rating scales are usually used in clinical trials to determine outcome as well as the time point of substantial improvement and response to treatment. Out of the variety of scales available the most established observer scales in anxiolytic drug trials are the Hamilton anxiety scale (HAS, Hamilton 1969), and the Hamilton depression scale (HAD, Hamilton 1967). Both Hamilton scales contain items reflecting the other scale (Williams 1986). The HAS consists of 14 items, each of them being a Likert scale (Likert 1932), a categorized scale with 5 categories for each item. In testing the HAS for homogeneity a general factor of anxiety as well as a "psychic" versus "somatic" factor was shown by factor analysis (Hamilton 1969). Recently, the adequacy of the HAS in drug trials with patients suffering from panic attacks has been questioned, since there is no distinction between attacks of anxiety and generalized or persistent anxiety provided in that scale.

The symptom checklist (SCL), one of the most comprehensive anxiety self-rating scales was developed out of the Hopkins Symptom Checklist (HSCL; Parloff et al. 1954) which was originally designed to assess psychoneurotic complaints in outpatients (Bech 1987). The later developments of the SCL (SCL-90, Derogatis et al. 1976) have included items for phobic anxiety and free floating anxiety. The SCL-90 consists of 90 individual items with five categories for each item ranging from 0 (not at all) to 4 (extremely). The SCL-90 quantifies current psychopathology in terms of nine different primary symptom dimensions: somatization, obsessive-compulsive behaviour, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation and psychoticism.

As can be seen from the description of those scales, self- and observer ratings used so far in clinical trials share one feature limiting their comparability: they are not matched by the number of items, the content and the scaling of the different items.

Therefore, there is a considerable risk of different constructs being measured by self- and observer rating scales. Owing to those limitations, it is not surprising

* *Present address:* Mental State Hospital Haar, Vockestrasse 72, W-8013 Haar, Federal Republic of Germany
Offprint requests to: M. Albus (present address)

that self- and observer rating scales are frequently providing quantitatively different results for the same study (Williams et al. 1972; Möller and von Zerssen 1985; Snaith and Taylor 1985; Lipman et al. 1986).

Discrepancies between self- and observer ratings can be reduced partially by training of observers and patients. Additionally, to reduce variance between self- and observer rating scales due to different item pools there is need for a rating scale that allows self-assessment and observer assessment on the same items. So far, only Zung (Zung 1971, 1976) has provided corresponding scales for self- and observer-rated anxiety containing the same items.

However, item description as well as scoring of the items is not identical. Marks and Sheehan proposed two scales with 35 items each, an observer and a self-rating scale for the assessment of anxiety, which are based on a previously developed self-rating scale (Marks and Matthews 1979) and have already been validated in a previous study (Ballenger et al. 1988). The Marks-Sheehan scale is the only scale where self- and observer ratings are perfectly matched by the number, content, item description and the scoring of the items. Moreover, both scales are tapping all aspects of psychopathology from which anxiety patients are suffering. Each of the 35 items represents a particular symptom like the SCL-90 in contrast to an overall syndrome score as in the HAS. The item-by-item correspondence as supplied by the Marks-Sheehan scale allows evaluation of the compatibility of self- and observer rating inventories with special emphasis on different attitudes of patients and clinicians in scoring individual items as well as psychic and somatic aspects of anxiety. Additionally, different structures in self- and observer ratings were studied by means of factor analysis. Thus, the Marks-Sheehan scale is an excellent tool for investigating the relative advantages of self- and observer rating scales and to find out the sources of disagreement.

Methods

Study design. The data material for the validation of the Marks-Sheehan scales is derived from the cross-national collaborative panic study, phase II (Klerman et al. 1990). This study compared in a randomized double-blind placebo-controlled design the therapeutic efficacy of alprazolam and imipramine in patients who met the DSM-III criteria (APA 1980) for panic disorder. The diagnosis of panic disorder was made by means of a structured clinical interview (SCID, Spitzer and Williams 1988). Any degree of phobic avoidance was allowed. Patients with a current or past major depressive episode without melancholia were included when the features of anxiety were more prominent than those of depression during the present episode and panic disorder had an earlier onset than depression within the present episode. After a drug off-time for at least 1 week, the medication was randomly allocated over a total period of 8 weeks. The age range was 18 to 65 years. For further details, see Klerman et al. (1990).

Centres, sample size and completion rates. Twelve different centres all over the world participated in the study. A total of 1168 outpatients were randomized after having met the inclusion criteria. For baseline evaluation, data on the Marks-Sheehan scales were available for 1152 patients; 1090 patients completed at least 3 weeks of the trial, while 812 completed the 8-week treatment period with data available on the Marks-Sheehan scales from 685 patients. The drop-out rate was different for the three treatment groups: 17.4% in the alprazolam group, 30.2% in the imipramine group and 43.7% in the placebo group.

Ratings and reliability. The data collected at baseline, at the end of week 1, 3, 6 and 8 during the 8-week treatment period were assessed independently by patients and clinicians and included the 14-item HAS (Hamilton 1969), the HAD (Hamilton 1967), the SCL-90 (Derogatis et al. 1976), a global rating of the clinicians and patients for the overall improvement relative to baseline and the Marks-Sheehan scale.

The Marks-Sheehan scale consists of the clinician-rated anxiety scale (CRAS), and the patient-rated anxiety scale (PRAS). Both scales are identical in the number of items ($n = 35$), the content, description and scoring of the items. Each item has five levels ranging from 0 (not at all) to 4 (severely). The items cover all symptoms that are assumed to be related to anxiety and include panic attacks, phobia, anticipatory anxiety, depression and obsessive compulsive symptoms.

The reliability of the structured interview and of the ratings was ensured by an extensive 1-week training of all investigators of the 12 participating centres before the study began. Each of the centres had to supply at least three videotaped SCID interviews (Spitzer and Williams 1988). The tapes were evaluated blindly. The inter-rater reliability was sufficient with regard to diagnosis and global scores of the scales administered (kappa > 0.70). Data were scored on to computer-readable forms and checked and reviewed extensively for error and inconsistencies. Statistical analyses were done at the Biostatistical Unit of the Massachussets General Hospital under the direction of Dr. Lavori. Since this paper focuses on the comparison between self- and observer ratings of the Marks-Sheehan scales, only statistical analyses concerning the Marks-Sheehan scales will be reported here.

Statistical analysis. For testing the compatibility of self- and observer rating scales two general methods were applied:

1. Correlating the corresponding items in self- and observer rating scales: Spearman correlation coefficients between CRAS and PRAS for each of the 35 items were computed separately for week 0 and 8. Also, correlations between the factors obtained by the principal component analysis after varimax rotation of CRAS and PRAS were done. Moreover, Spearman correlation coefficients were calculated for the isolated two factors of the CRAS and PRAS as well as for CRAS and PRAS (35 items) total and Hamilton anxiety total scores and the SCL-90 subscale anxiety.

2. Correspondence of the factors obtained by principal component analysis: principal component analysis without rotation and with varimax rotation between observer and self-rating scales were carried out separately for the CRAS and the PRAS for baseline (week 0), week 3, 8 and end-point data for all patients where both ratings were available. According to the distribution of the eigenvalues a four-, three- and two-factor solution turned out to be appropriate. Then, consistency of factors over time was used as the curcial criterion for choosing the definite number of factors. The consistency of varimax rotated factors obtained for the self-rating scale and those obtained for the observer rating scale served as the indicator of the comparability of both methods of assessment. Since the two-factor solution turned out to be the most appropriate with regard to temporal stability, the two-factor solution was chosen for further analysis.

Results

Means and Standard Deviation

In all but one item (item 30, tension) as well as in total scores the PRAS showed higher means (P ranging from

Table 1. Means (*x*) and standard deviation (SD) of clinician-rated anxiety (CRAS) and patient-rated anxiety (PRAS) individual items for week 0 and week 8

	CRAS Week 0 ($n = 1158$)		PRAS Week 0 ($n = 1157$)		CRAS Week 8 ($n = 689$)		PRAS Week 8 ($n = 689$)	
	x	SD	*x*	SD	*x*	SD	*x*	SD
1. Light-headedness	1.6	1.1*	2.0	1.3	0.6	0.8***	0.7	1.0
2. Rubbery legs	1.2	1.0*	1.6	1.3	0.4	0.7*	0.6	1.0
3. Unsteady	1.1	1.1*	1.7	1.4	0.3	0.6*	0.6	0.9
4. Dyspnoea	1.5	1.1*	1.8	1.3	0.4	0.7*	0.6	0.9
5. Tachycardia	1.8	1.1*	2.1	1.3	0.6	0.8***	0.7	1.0
6. Chest pain or pressure	1.3	1.1*	1.6	1.3	0.5	0.7***	0.6	0.9
7. Smothering of choking	1.4	1.2*	1.7	1.4	0.4	0.8***	0.5	0.9
8. Paraesthesia	1.0	1.1*	1.4	1.3	0.3	0.6***	0.4	0.8
9. Hot or cold spells	1.4	1.0*	1.7	1.3	0.5	0.8*	0.7	1.0
10. Nausea	0.9	1.1*	1.5	1.4	0.3	0.6*	0.6	1.0
11. Diarrhoea	0.5	0.8*	0.7	1.1	0.1	0.4**	0.2	0.6
12. Headaches	1.4	1.1*	1.7	1.3	0.6	0.8*	0.8	1.0
13. Feeling tired or weak	1.5	1.1*	2.2	1.3	0.7	0.9*	1.1	1.2
14. Startled easily	1.1	1.1*	1.4	1.3	0.4	0.7***	0.5	0.9
15. Sweating	1.3	1.1***	1.5	1.4	0.6	0.8	0.7	1.2
16. Derealization	0.7	1.0*	1.2	1.3	0.2	0.5***	0.3	0.7
17. Depersonalization	0.5	0.9*	0.9	1.3	0.1	0.3*	0.2	0.6
18. Worrying about health	1.1	1.2*	2.0	1.4	0.5	0.8*	0.9	1.1
19. Feeling insane or out of control	0.9	1.1*	1.3	1.4	0.3	0.6	0.3	0.7
20. Fear of dying	1.3	1.3*	1.9	1.5	0.4	0.7*	0.6	1.0
21. Trembling	1.3	1.0*	1.7	1.3	0.4	0.7*	0.6	0.9
22. Waves of depression	1.1	1.0*	1.7	1.4	0.5	0.7*	0.7	1.0
23. Emotional liability	1.1	1.1*	1.7	1.4	0.5	0.8*	0.8	1.0
24. Dependent on others	1.4	1.3**	2.6	2.5	0.6	1.0	0.7	1.1
25. Compulsive rituals	0.2	0.5*	0.7	1.2	0.0	0.2*	0.3	0.7
26. Obsessive thoughts	0.2	0.6*	1.0	1.3	0.1	0.3*	0.3	0.8
27. Initial insomnia	1.1	1.2*	1.5	1.5	0.2	0.6*	0.4	0.9
28. Middle insomnia	1.0	1.1*	1.6	1.4	0.3	0.7*	0.6	1.0
29. Phobias	1.9	1.2*	2.3	1.4	0.9	1.0	1.0	1.2
30. Tension	2.3	0.9***	2.4	1.3	1.0	0.9	0.9	1.1
31. Anxiety	1.3	1.0*	2.8	1.1	0.5	0.7*	1.1	1.1
32. Spontaneous major panic attacks	2.0	1.1*	2.4	1.3	0.3	0.8*	0.5	0.9
33. Spontaneous minor panic attacks	1.5	1.2*	2.1	1.2	0.5	0.8*	0.7	0.9
34. Anticipatory anxiety episodes	2.0	1.1*	2.3	1.3	0.8	0.9	0.8	1.1
35. Situational panic attacks	1.6	1.3*	2.1	1.4	0.4	0.8*	0.7	1.0
Total	43.2	18.5*	59.7	25.7	14.9	3.2*	21.6	19.0

Significant differences between PRAS and CRAS individual item means are indicated with * $P < 0.0001$; ** $P < 0.001$; *** $P < 0.05$

< 0.0001 to < 0.05) and a wider variation compared to the CRAS at baseline (week 0) as well as at week 8.

Correlations Between CRAS Total and PRAS Total and Other Rating Scales.

Spearman correlation coefficients between CRAS total and PRAS total were $r = 0.69$ ($P < 0.001$) at week 0 and $r = 0.82$ ($P < 0.001$) at week 8.

Spearman correlation coefficients between CRAS total and HAS total were $r = 0.83$ ($P < 0.0001$) at week 0 and $r = 0.89$ ($P < 0.0001$) at week 8.

Spearman correlation coefficients between PRAS total and the SCL-90 subscale anxiety were $r = 0.84$ ($P < 0.0001$) at week 0 and $r = 0.90$ ($P < 0.0001$) at week 8.

Spearman Correlation Coefficients Between the Individual Items of CRAS and PRAS

Table 2 lists the Spearman correlation coefficients between each of the 35 items of the CRAS and the PRAS. The correlations range between 0.72 (initial insomnia) and 0.32 (obsessive-compulsive symptoms) at week 0

Table 2. Spearman correlation coefficients between each of the 35 items of the patient-rated anxiety scale (PRAS) and the clinician-rated anxiety scale (CRAS)

	PRAS/ CRAS	Baseline ($n = 1152$)	Week 8 ($n = 685$)
1. Light-headedness	1	$r = 0.65$	0.74
2. Rubbery legs	2	0.65	0.69
3. Unsteady	3	0.61	0.62
4. Dyspnoea	4	0.62	0.68
5. Tachycardia	5	0.61	0.73
6. Chest pain or pressure	6	0.67	0.71
7. Smothering of choking	7	0.65	0.72
8. Paraesthesia	8	0.66	0.66
9. Hot or cold spells	9	0.63	0.71
10. Nausea	10	0.62	0.51
11. Diarrhoea	11	0.71	0.70
12. Headaches	12	0.62	0.65
13. Feeling tired or weak	13	0.57	0.64
14. Startled easily	14	0.61	0.58
15. Sweating	15	0.64	0.75
16. Derealization	16	0.52	0.46
17. Depersonalization	17	0.44	0.38
18. Worrying about health	18	0.42	0.47
19. Feeling insane or out of control	19	0.57	0.56
20. Fear of dying	20	0.67	0.61
21. Trembling	21	0.62	0.64
22. Waves of depression	22	0.50	0.52
23. Emotional liability	23	0.39	0.47
24. Dependent on others	24	0.65	0.70
25. Compulsive rituals	25	0.32	0.34
26. Obsessive thoughts	26	0.32	0.30
27. Initial insomnia	27	0.72	0.74
28. Middle insomnia	28	0.68	0.66
29. Phobias	29	0.49	0.56
30. Tension	30	0.45	0.54
31. Anxiety	31	0.33	0.45
32. Spontaneous major panic attacks	32	0.41	0.62
33. Spontaneous minor panic attacks	33	0.39	0.52
34. Anticipatory anxiety episodes	34	0.42	0.58
35. Situational panic attacks	35	0.45	0.54

($P < 0.0001$ for all correlation coefficients)

with a median of 0.61 and between 0.75 (sweating) and 0.30 (obsessive thoughts) at week 8 with a median of 0.62. Self- and observer-rated anxiety scales are highly correlated in items that reflect somatic symptoms of anxiety. In contrast, the correlations between self- and observer ratings in items referring to derealization, depersonalization, obsessive-compulsive symptoms, psychic symptoms of anxiety and panic attacks are low. Overall, the correlations between the individual items of self- and observer-rated anxiety scales are higher at week 8 compared to baseline.

Table 3. Factor analysis with varimax rotation for the clinician-rated anxiety scale (CRAS) at baseline (week 0) and week 8 (only factor loadings exceeding 0.40 are listed)

	Factor 1		Factor 2	
	Baseline	Week 8	Baseline	Week 8
1. Light-headedness	0.46	0.60		
2. Rubbery legs	0.54	0.50		
3. Unsteady	0.45	0.48		
4. Dyspnoea	0.58	0.70		
5. Tachycardia	0.48	0.68		
6. Chest pain or pressure	0.65	0.66		
7. Smothering or choking	0.56	0.63		
8. Paraesthesia	0.52	0.52		
9. Hot or cold spells	0.49	0.51		
10. Nausea	0.49	0.42		
11. Diarrhoea				
12. Headaches	0.47	0.50		
13. Feeling tired or weak	0.53	0.43		
14. Startled easily	0.46			0.49
15. Sweating				
16. Derealization			0.59	0.64
17. Depersonalization			0.57	0.66
18. Worrying about health		0.45		
19. Feeling insane or out of control			0.56	0.50
20. Fear of dying	0.46	0.61		
21. Trembling	0.46	0.56		
22. Waves of depression	0.42	0.43	0.41	0.53
23. Emotional liability	0.42	0.41	0.42	0.49
24. Dependent on others		0.43	0.64	
25. Compulsive rituals				0.54
26. Obsessive thoughts				0.54
27. Initial insomnia	0.52			
28. Middle insomnia	0.54	0.43		
29. Phobias		0.44	0.58	
30. Tension	0.51	0.63	0.48	
31. Anxiety		0.52		
32. Spontaneous major panic attacks	0.58	0.67		
33. Spontaneous minor panic attacks		0.53		
34. Anticipatory anxiety episodes		0.50	0.59	
35. Situational panic attacks		0.46	0.59	

Variance explained by each factor:
Factor 1: 6.179 = 25%; factor 2: 4.496 = 5.5% (baseline)
Factor 1: 7.942 = 28.5%; factor 2: 3.900 = 5.3% (week 8)

Principal Component Analysis

The eigenvalues show that factor 1 is by far the highest (PRAS – week 0: factor 1 = 10.88; week 8: factor 1 = 12.61; CRAS - week 0: factor 1 = 8.75; week 8: factor 1 = 10.00). The Kaiser criterion suggests a nine to seven

Table 4. Factor analysis with varimax rotation for the patient-rated anxiety scale (PRAS) at baseline (week 0) and week 8 (only factor loadings exceeding 0.40 are listed)

	Factor 1		Factor 2	
	Baseline	Week 8	Baseline	Week 8
1. Light-headedness			0.54	0.58
2. Rubbery legs			0.61	0.62
3. Unsteady			0.54	0.57
4. Dyspnoea			0.61	0.64
5. Tachycardia			0.47	0.58
6. Chest pain or pressure			0.65	0.63
7. Smothering or choking			0.61	0.53
8. Paraesthesia			0.58	0.60
9. Hot or cold spells			0.51	0.61
10. Nausea			0.49	0.41
11. Diarrhoea				
12. Headaches			0.50	0.58
13. Feeling tired or weak		0.43	0.57	0.55
14. Startled easily				0.42
15. Sweating				0.47
16. Derealization	0.56	0.60		
17. Depersonalization	0.54	0.55		
18. Worrying about health		0.49	0.42	0.45
19. Feeling insane or out of control	0.57	0.63		
20. Fear of dying		0.51	0.49	0.41
21. Trembling	0.46	0.47		0.49
22. Waves of depression	0.55	0.62		
23. Emotional liability	0.59	0.67		
24. Dependent on others	0.67	0.56		
25. Compulsive rituals	0.52	0.62		
26. Obsessive thoughts	0.53	0.63		
27. Initial insomnia			0.45	0.44
28. Middle insomnia			0.47	0.46
29. Phobias	0.72	0.65		
30. Tension	0.63	0.60		0.45
31. Anxiety	0.65	0.66		0.45
32. Spontaneous major panic attacks	0.42	0.48	0.46	0.51
33. Spontaneous minor panic attacks	0.43	0.44		0.43
34. Anticipatory anxiety episodes	0.72	0.72		
35. Situational panic attacks	0.72	0.64		

Variance explained by each factor:
Factor 1: 6.823 = 31.1%; factor 2: 6.177 = 6.1% (baseline)
Factor 1: 7.515 = 36%; factor 2: 6.891 = 5.1% (week 8)

Table 5. Spearman correlation coefficients between total score (tot.), factor 1 (F1) and factor 2 (F2) of the clinician-rated anxiety scale (CRAS) and the patient-rated anxiety scale (PRAS) at baseline (week 0), week 3 and week 8 (* $P < 0.0001$)

	Week 0 ($n = 1152$)	Week 3 ($n = 954$)	Week 8 ($n = 689$)
	|R|	|R|	|R|
PRAS F1/PRAS F2	−0.00	0.04	0.05
CRAS F1/CRAS F2	−0.03	−0.00	0.04
PRAS F1/CRAS F1	0.47*	0.66*	0.74*
PRAS F1/CRAS F2	0.50*	0.46*	0.33*
PRAS F2/CRAS F1	0.53*	0.49*	0.26*
PRAS F2/CRAS F2	−0.42*	−0.40*	−0.24*
PRAS tot./CRAS tot.	0.69*	0.83*	0.82*

factor solution for both scales for the baseline as well as for the week 8 data (PRAS - week 0: 2–9 range from 1.93 to 1.04; week 8: factors 2–8 range from 1.85 to 1.13; CRAS – week 0: factors 2–7 range from 2.12 to 1.06; week 8: factors 2–8 range from 1.79 to 1.05). The eigenvalues suggest the existence of a general factor, thus proposing a one-factor solution. To test whether besides this general factor of anxiety a "psychic" and "somatic" factor like in the HAS is present in PRAS and CRAS, additional criteria were applied. The screegraph by Cattell (Überla 1971) proposes either a four-, three- or two-factor solution. Since this criterion is not very decisive in suggesting a definite number of factors, the consistency over time of the varimax rotated factor solutions was used as an additional criterion for the decision as to which factorial solution would be the most appropriate one. Thus, a factor solution seemed appropriate when for all factors the requirement that the corresponding factors for baseline and week 8 were correlated with at least 0.50 (Burt coefficient; Schneewind and Catell 1970) was fulfilled. Neither the four- nor the three-factor solution of the CRAS and PRAS met this requirement. The two-factor solution was the first that met the criterion.

Table 3 shows the results for the two-factor solution after varimax rotation for the 35-item CRAS at baseline and week 8. Only loadings exceeding 0.40 are reported. Items in bold type have consistent and satisfactory factor loadings at both time points. Factor 1 can be interpreted as an overall panic attack factor including somatic symptoms as well as fear of dying, which are frequently present during major panic attacks.

Only three items can be identified as loading higher than 0.40 at each of the two time points exclusively on factor 2; those items describe derealization, depersonalization and feeling out of control. Consequently, the factor pattern for somatic symptoms is consistent, the pattern for psychic symptoms of anxiety is inconsistent. The variance explained by both factors is satisfactory at baseline (= 30.5%) as well as at week 8 (= 33.8%).

Table 4 lists the results of factor analysis with varimax rotation for PRAS at baseline and week 8. Factor 2 corresponds to factor 1 in CRAS and can be considered as a factor exclusively reflecting somatic symptoms during panic attacks. Factor 1 shows a consistent pattern in items reflecting derealization, depersonalization, loss of control, depressed mood, obsessive-compulsive symptoms as well as phobic and anticipatory anxiety. Again, the variance explained by both factors is high (37.2% at

baseline, 41.1% at week 8). Comparing the factor pattern of PRAS and CRAS, the two-factor solution shows consistent results for somatic symptoms of anxiety, but inconsistencies in psychic symptoms of anxiety.

Table 5 lists the Spearman correlation coefficients between PRAS total score, the exact factor scores of factor 1, factor 2 and CRAS total score, factor 1 and factor 2 and between CRAS factor 1 and 2 and PRAS factor 1 and 2 at baseline, week 3 and week 8. PRAS and CRAS total scores are obtained by addition of the 35 individual item scores.

Because of the inconsistencies in the factorial structure of both rating scales, the correlations between the corresponding factors of PRAS and CRAS are quite low. In contrast, the correlation between PRAS and CRAS total scores are satisfactory and again increase during the treatment period. Correlations between the two factors in each scale show that the two factors measure different constructs.

Discussion

The correspondence between CRAS and PRAS reflects the general relationship between self- and observer anxiety scales. This relationship can be illustrated most appropriately by scales covering all aspects of anxiety and by scales carefully matched by the number, content and scaling of all items that are included in self- and observer rating scales. Since the Marks-Sheehan scales correspond to these requirements, a part of the general relationship between self- and observer rating scales can be obtained by comparing the total scores of CRAS and PRAS. The high correlations between the CRAS and the Hamilton total score as well as between the PRAS and the SCL-90 total score show that both Marks-Seehan scales are highly comparable to the most established concurrent rating scales. The correlations between the total scores show that matching by number, content and scaling of the items increases the compatibility of self- and observer ratings compared to self and observer rating scales with different items (Snaith and Taylor 1985; Lipman et al. 1986).

The correlations between the total scores of individual items of patient- and observer-rated anxiety scales increased from baseline to week 8. This finding is in line with other psychopharmacological treatment studies (Möller 1990). Since the variances were not higher at week 8 compared with baseline, which would suggest that that finding be interpreted as caused by a wider variation due to different outcomes at week 8, most likely training effects (in the patient group who are less familiar with ratings) explain that finding.

The comparison between CRAS and PRAS leads to a number of findings, some proving similarities, other discrepancies between self- and observer-rated Marks-Sheehan anxiety scales.

The higher means and standard deviations of the individual items of the patient rated anxiety scale reflect the fact that patients have a tendency toward evaluating themselves as being more severely ill. It is also reasonable to assume that some of the discrepancies in standard deviations between self- and observer rating are due to a different familiarity with rating scales in patients and observers as well as due to the considerably higher number of patients compared to raters.

The item-by-item correspondence shows that both scales have a high compatibility in items reflecting somatic symptoms, which is demonstrated by correlations as well as by factorial structure. However, two items, diarrhoea and sweating show neither consistent nor satisfying factor loadings in both scales. Since the factor analytical results are considered as the index of validity, it is recommended that these items are dropped.

The similar stable and consistent factor pattern for somatic symptoms in both scales contrasts with the inconsistent and instable factor pattern for psychic symptoms. As shown by correlational and factor analyses, items reflecting psychic symptoms of anxiety, depression and panic attacks are discrepant and not consistent between the two scales. According to the results of the two-factor solution for CRAS and PRAS the low correlation between the individual items of CRAS and PRAS (Table 1) reflecting psychic anxiety is at least partially explained by inconsistencies in the factorial structure of the two rating scales. Taking into account that the Marks-Sheehan scale does not provide an elaborated item description, the individual items are still likely to measure different constructs. Owing to the more precise connotation of somatic items in everyday language it is assumed that a more precise item definition would enhance the correlations. According to our results, the consistency of the Marks-Sheehan scale could be improved by dropping the item "startling easily", an item with inconsistent factorial structure between CRAS and PRAS and giving up the distinction between major and minor spontaneous panic attacks.

A two-factor solution is equally appropriate and stable for both scales. However, the variance explained is higher for the patient-rated anxiety scale. For both scales factorial analysis showed a somatic and psychic factor for anxiety as proposed by Hamilton (1969).

However, this overall similarity for the two factors obtained in the Marks-Sheehan scales as well as in the HAS, namely psychic and somatic anxiety, turns out to be stable only for the somatic factor when comparing the CRAS and PRAS. In both Marks-Sheehan scales, the inclusion of items specifically referring to panic attacks, which was claimed to be an advantage to the HAS, neither increased the consistency nor the distinction between constructs of persistent and panic anxiety. This is at least partially due to the inconsistencies in the factorial structure of both Marks-Sheehan scales. The consequences of these discrepancies for measuring anxiolytic drug effects will be discussed in a subsequent paper (Maier et al.; this issue).

Our data provide empirical evidence that the low consistencies between self- and observer ratings reported so far (Williams et al. 1972; Möller and von Zerssen 1985; Snaith and Taylor 1985; Lipman et al. 1986) can be improved by using a rating scale which matches up item by item for self- and observer rating. Although the correlations between the individual items of the Marks-

Sheehan scale are higher than those reported for different self- and observer scales (Snaith and Taylor 1985; Lipman et al. 1986) the Marks-Sheehan scales only partially fulfil the requirements for consistency. Other sources of disagreement may contribute to the diverging results between self- and observer rating scales: first, the tendency of patients to rate themselves as more severely ill and the tendency of observers to overemphasize an improvement in therapeutic studies; secondly, the different understandings of raters and patients of the meaning of the items, consequently resulting in different constructs measured. These crucial points cannot be solved by supplying new rating scales, but by more elaborate item descriptions and training of patients as well as raters to obtain a better compatibility between self and observer rating scales.

Acknowledgement. This is a report from the Cross National Collaborative Panic Study, Second Phase. This study consists of a Steering Committee: James H. Coleman, Chairman (Kalamazoo, Mich., USA); Gerald L. Klerman, Co-Chairman (New York, NY, USA). Project Directors: Jose Luis Ayuso (Madrid, Spain); Per Bech (Hillerod, Denmark); Otto Benkert (Mainz, FRG); Sydney Brandon (Leicester, England); Giovanni B. Cassano (Pisa, Italy); Jorge A. Costa e Silva (Rio de Janeiro, Brazil); George C. Curtis (Ann Arbor, Mich., USA); Juan R. de la Fuente (Mexico City, Mexico); Jose Guimon (Bilbao, Spain); Hanns Hippius (Munich, FRG); Yves Lecrubier (Paris, France); Carlos A. Leon (Cali, Colombia); Juan J. Lopez-Ibor Jr. (Madrid, Spain); Heinz Katschnig (Vienna, Austria); Juan Massana (Barcelona, Spain); Mogens Mellergaard (Glostrup, Denmark); Jan-Otto Ottoson (Gotenborg, Sweden); Ole J. Rafaelsen (deceased); Raben Rosenberg (Copenhagen, Denmark); Martin Roth (Cambridge, England); Javier Sepulveda (Mexico City, Mexico); Leslie Solyom (Vancouver, Canada); Marco Versiani (Rio de Janeiro, Brazil); Jean Wilmotte (Marchienne-au-Pont, Belgium).

Data processing and preparation of the tapes were performed by Dennis B. Gillings, Quintiles (Chapel Hill, NC) and by Philip W. Lavori and the staff of the Biostatistical Unit, Massachusetts General Hospital (Boston, Mass.). Statistical analyses reported in this paper were performed mainly by Philip W. Lavori and the staff of the Biostatistical Unit, Massachusetts General Hospital (Boston, Mass.). Dennis B. Gillings and the staff of Quintiles (Chapel Hill, NC) also contributed statistical advice and analyses.

This study has been sponsored and supported by The Upjohn Company, Kalamazoo, Mich. through its Psychopharmacology Research Unit, Division of Medical Affairs, Robert P. Purpura, director; Carl P. Lewis; Cathy A. White.

References

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. American Psychiatric Association, Washington, DC

Ballenger JC, Burrows GD, DuPont RL, Lesser IM, Noyes R Jr, Pecknold JC, Rifkin A, Swinson RP (1988) Alprazolam in panic disorder and agoraphobia: results from a multicenter trial. I. Efficacy in short-term treatment. Arch Gen Psychiat 45:413–422

Bech P (1987) Quality of life in psychosomatic research. Psychopathology 20:169–179

Derogatis LR, Lipman RS, Covi L (1976) Self-report symptom inventory. In: Guy E (ed) ECDEU assessment manual for psychopharmacology. Rev Ed, Rockville, Maryland, pp 313–331

Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278–296

Hamilton M (1969) Diagnosis and rating of anxiety. In: Lader MH (ed) Studies of anxiety. Br J Psychiatry Spec Publ 3:76–79

Klerman GL (1990) Drug treatment of panic disorder: comparative efficacy of alprazolam, imipramine and placebo. Br J Psychiatry (accepted for publication)

Likert R (1932) A technique for the measurement of attitudes. Arch Psychol 40:1–55

Lipman RS, Covi L, Rickels K, McNair DM, Downing R, Kahn RJ, Lasseter VK, Faden V (1986) Imipramine and chlordiazepoxide in depressive and anxiety disorders. I. Efficacy in depressed outpatients. Arch Gen Psychiatry 43:79–85

Maier W, Albus M, Shera D, Buller R, Bech P (1990) Self and observer assessment in anxiolytic drug trials: a comparison of their validity. Eur Arch Psychiatry Clin Neurosci 240:103–108

Marks IM, Matthews AM (1979) Brief standard self-rating for phobic patients. Behav Res Ther 17:263–267

Möller HJ (1990) Outcome criteria in antidepressant drug trials: self rating versus observer rating scales. Eur J Psychopharmacol (*in press*)

Möller HJ, Zerssen D von (1985) Psychopathometrische Untersuchung zur prognostischen Bedeutung des psychopathologischen Befundes und der prämorbiden Persönlichkeit für den Entlassungszustand neuroleptisch behandelter Patienten mit schizophrenen Psychosen. Nervenarzt 56:498–505

Parloff MB, Kelman HC, Frank JD (1954) Comfort, effectiveness and self-awareness as criteria of improvement in psychotherapy. Am J Psychiatry 5:343–351

Schneewind KA, Catell RB (1970) Zum Problem der Faktorenidentifikation. Psychol Beitr 12:214–226

Snaith RP, Taylor CM (1985) Rating scales for depression and anxiety: a current perspective. Br J Clin Pharmacol 19:17S–20S

Spitzer RL, Williams JBW (1988) Revised diagnostic criteria and a new structured interview for diagnosing anxiety disorders. J Psychiatr Res 22:55–85

Überla K (1971) Faktorenanalyse. Springer, Berlin Heidelberg New York

Williams JG, Barlow DH, Agras WS, Jackson J (1972) Behavioral measurement of severe depression. Arch Gen Psychiatry 27: 330–333

Williams P (1986) The problem of diagnosis in psychiatric research. J Soc Med 79:1–2

Zung WWK (1971) A rating instrument for anxiety disorders. Psychosomatics 12:371–379

Zung WWK (1976) SAS. Self rating anxiety scale. In: Guy E (ed) ECDEU assessment manual for psychopharmacology. Rev Ed, Rockville, Maryland, pp 337–340