Eur Arch Psychiatry Clin Neurosci (2011) 261:357–367
DOI 10.1007/s00406-010-0175-1

ORIGINAL PAPER

The Inventory of Depressive Symptomatology (IDS-C_{28}) is more sensitive to changes in depressive symptomatology than the Hamilton Depression Rating Scale ($HAMD_{17}$) in patients with mild major, minor or subsyndromal depression

Isabella Helmreich · Stefanie Wagner · Roland Mergl · Antje-Kathrin Allgaier · Martin Hautzinger · Verena Henkel · Ulrich Hegerl · André Tadić

Received: 10 June 2010 / Accepted: 22 November 2010 / Published online: 4 December 2010

Abstract Depression rating scales play a decisive role in the assessment of the severity of depression and the evaluation of the efficacy of antidepressant treatments. The Hamilton Depression Rating Scale (HAMD) is regarded as the 'gold standard'; nevertheless, studies suggest that the Inventory of Depressive Symptomatology (IDS) is more sensitive to detect symptom changes. The aim of the present study was to investigate whether the IDS is more sensitive in detecting changes in depression symptoms in patients with mild major, minor or subsyndromal depression (MIND). Biweekly IDS-C_{28} and $HAMD_{17}$ data from 340 patients of a 10-week randomized, placebo-controlled trial comparing the effectiveness of sertraline and cognitive-behavioural therapy in patients with MIND were analysed. We investigated sensitivity to change for both scales (1) from assessment-to-assessment, (2) in relation to depression severity level, and (3) in relation to DSM-IV depression criterion symptoms. The IDS-C_{28} was more sensitive in detecting changes in depression symptomatology over the treatment course as well as for different severity levels, especially in patients with a low depression severity. It assesses the DSM-IV criteria more thoroughly, is better able to track the change of cognitive symptoms and to identify residual symptoms. Both scales are well able to assess depressive symptomatology. However, the IDS-C_{28} surpasses the $HAMD_{17}$ in detecting small changes especially in the core symptoms of depression. This is important for an optimal treatment by capturing early improvements, enabling prompt reactions and detecting residual symptoms.

Keywords Minor depression · Hamilton Depression Rating Scale ($HAMD_{17}$) · Inventory of Depressive Symptomatology (IDS-C_{28}) · Sensitivity to change · Depression severity

U. Hegerl and A. Tadić contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00406-010-0175-1) contains supplementary material, which is available to authorized users.

I. Helmreich (✉) · S. Wagner · A. Tadić
Department of Psychiatry and Psychotherapy,
University Medical Centre Mainz, Untere Zahlbacher Strasse 8,
5131 Mainz, Germany
e-mail: helmreich_i@psychiatrie.klinik.uni-mainz.de

R. Mergl · U. Hegerl
Department of Psychiatry,
University of Leipzig, Leipzig, Germany

A.-K. Allgaier
Department of Child and Adolescent Psychiatry,
Ludwig-Maximilians-University, Munich, Germany

M. Hautzinger
Institute of Psychology, Department of Clinical
and Developmental Psychology, University of Tuebingen,
Tuebingen, Germany

V. Henkel
Department of Psychiatry and Psychotherapy,
Ludwig-Maximilians-University, Munich, Germany

Introduction

Clinician rating scales play a decisive role in the assessment of depression severity as well as in the evaluation of antidepressant treatment. Meaningful and time-efficient measurements are of great importance in order to accurately capture depression symptoms, make effective treatment decisions and to be able to conduct cost-efficient clinical trials.

The *Hamilton Depression Rating Scale* (HAMD, [15]) is a worldwide used clinician-administered rating scale for the evaluation of depression severity. It is still considered to be the 'gold standard'. Though it exists in different versions with up to 36 items (for review see [44]), the 17-item version is the most frequently used one [3] and meets established criteria for reliability and validity. Nevertheless, it has been criticized, because of its conceptual shortcomings [3]. Developed 50 years ago, the $HAMD_{17}$ reflects only in parts the current DSM-IV operationalization of the diagnosis 'Major Depression' [1]. Behavioural and somatic symptoms are overrepresented, cognitive symptoms are de-emphasised and atypical symptoms are missing. Most DSM-IV depression criteria are assessed only by one or two items (e.g., depressive mood). Therefore, one can expect that a slight improvement in the main area of depressive symptomatology would not be reflected adequately by a change of the total score of the $HAMD_{17}$. Previous studies have shown that the $HAMD_{17}$ is not able to detect small improvements in classic depressive symptoms like depressed mood, guilt, and suicidal ideation [5–7], resulting in a limited sensitivity for changes in depressive symptomatology [3, 19, 33]. Moreover, the scale was not constructed according to classical test theory, i.e., the psychometric and conceptual quality of each item is different, generating a total score of which the meaning is multidimensional and unclear [3, 19, 33].

Many studies postulate modern and better psychometrically constructed rating scales as a replacement of the $HAMD_{17}$ (i.e., [3, 43]). The *Inventory of Depressive Symptomatology* (IDS, [34, 35]) is one of the favoured options in this discussion. The IDS was developed as a 28-item clinician rating scale [34] and was later expanded to 30 items [35] to capture all DSM-IV atypical symptom features [2]. Its development was guided by the DSM-IV diagnostic criteria for major depression allowing to capture all nine criterion symptom domains as well as commonly associated symptoms (e.g., anxiety, irritability and subtype symptoms such as melancholic or atypical). It has very good psychometric properties (e.g., a high internal consistency) and has been proven to be sensitive to depression severity changes [8, 36–38, 40].

The sensitivity to change in depressive symptomatology is an important criterion for the evaluation of efficacy of any antidepressant treatment. There are only a few studies comparing the sensitivity to change of the IDS-C_{30} and the $HAMD_{17}$ directly. Additionally, most of these studies have been carried out in patients with major depressive disorder (MDD). Rush and colleagues [35] analysed 59 out-patients with MDD in an open-label trial of fluoxetine. They observed that the IDS was more sensitive to detect small changes in symptom severity in the lower scoring range than the $HAMD_{17}$, while Lauge et al. [24] found that the IDS and the HAMD are equal in their ability to detect changes in depressive symptoms during treatment in patients with MDD. By analysing data from 993 outpatients with non-psychotic MDD in a 12-week open trial of nefazodone [13], Rush and colleagues [36] demonstrated that the IDS-C_{30} declared less patients as responder (defined as a total score reduction $\geq$50% compared to baseline) than the $HAMD_{17}$ (IDS-C_{30}: 46.2% vs. $HAMD_{17}$: 64.4%), indicating that the $HAMD_{17}$ is less sensitive to residual symptoms than the IDS-C_{30}. To the best of our knowledge, there is only one study comparing the IDS-C_{30} with the $HAMD_{17}$ in a sample of patients with minor depression. In a randomized, double-blind, placebo-controlled trial with fluoxetine in 162 patients, Judd and colleagues [23] found that the IDS-C_{30} was more sensitive than the $HAMD_{17}$ in detecting response to treatment. However, the published data of Judd and colleagues [23] were restricted to analyses of mean total scores. The relationship between changes in depression severity and subscores of the scales, e.g., changes in DSM-IV depression criteria symptoms, were not examined. Minor depression is more associated with mood and cognitive symptoms than with vegetative symptoms [18, 32], which are better covered by the items of the IDS-C_{30}. Therefore, analyses of subscales are of high interest in the comparison of both scales. In order to further explore this, we analysed a sample of patients with MIND (mild major, minor or subsyndromal depression). The aim of this study was to investigate (1) whether the IDS-C_{28} is more sensitive in detecting changes in depression severity than the $HAMD_{17}$ during the course of an antidepressant treatment; (2) whether the sensitivity for change is dependent on the depression severity level; and (3) how sensitivity for change is dependent on DSM-IV criteria of depression.

Methods

Sample and data collection

The sample consisted of 368 patients with mild major, minor and subsyndromal depression. The data were collected during a prospective, randomized, single-centre, placebo-controlled, parallel-group, 10-week clinical trial with five treatment arms: sertraline, placebo, manual-guided cognitive-behavioural group therapy, guided self-help group, and patients' choice condition. The study compared the effectiveness of sertraline and cognitive-behavioural therapy. It was performed within the framework of the German Research Network on Depression and Suicidality. All participants gave their written informed consent to

participate in the study. The study was approved by an independent Ethics Review Committee (Medical Faculty, Ludwig-Maximilians-University Munich, Munich, Germany).

Details of the study design have been described elsewhere [17]. In brief, patients were referred by primary-care providers to the study centre. Inclusion criteria required that participants had to be a minimum of age 18 years, and have a diagnosis of subthreshold (minor) depression, dysthymia or major depressive disorder (according to DSM-IV criteria) with mild to moderate severity, i.e., a $HAMD_{17}$ total score ≥ 8 and ≤ 22. In order to create a sample as close as possible the population found in general practice, comorbidities such as somatoform disorders and anxiety disorders were allowed. Patients were excluded from the study if they had acute suicidality, a diagnosis of brief recurrent depression, bipolar affective disorder, addiction (alcohol, benzodiazepines, illicit drugs), schizophrenia, schizotypal personality disorder or delusional disorder, obsessive–compulsive disorder, severe somatic diseases or current psychotherapeutic or antidepressant treatment. The diagnoses according to DSM-IV criteria were confirmed using a German computer-administered structured clinical interview for DSM-IV (DIA-X, [45]), which is based on the Composite International Diagnostic Interview [46]. As primary efficacy measure for the severity of depressive symptomatology the IDS-C_{28} [34] and the $HAMD_{17}$ [16] were assessed biweekly by trained and blinded clinicians.

Statistical analysis

The intent-to-treat sample (i.e., all randomized patients) was analysed. Patients were excluded if baseline assessments for the IDS-C_{28} and the $HAMD_{17}$ were not available (IDS-C_{28}: $n = 17$; $HAMD_{17}$: $n = 1$) and the $HAMD_{17}$ baseline total score was <8 ($n = 10$), resulting in a total number of 340 patients. Baseline clinical and demographical data were computed.

The first hypothesis (*Is the IDS-C_{28} more sensitive in detecting small changes in depression severity than the $HAMD_{17}$*) was analysed by calculating mean sum scores ($\pm$standard deviation, SD) by assessment for each scale. Agreement in total scores between the scales was tested with Pearson correlations. In order to characterise changes in depressive symptomatology (1) between baseline and each of the five subsequent assessments and (2) from assessment-to-assessment, we analysed the mean sum score changes (student's *t*-test for dependent samples, $P \leq 0.01$) and determined effect sizes (d_i). In order to take between-patient variability into account, we computed an extension of d to individual cases (d_i) according to the formula from Vittengl et al. [41], i.e., we calculated the standardized difference between dependent means. We tested d_i for differences between the scales with student's *t*-test for dependent samples ($P \leq 0.05$). In order to investigate if both scales indicate the same direction of score change, three groups were formed according to the direction of change: decrease, increase or no change in total scores. The direction of change was calculated for both scales from assessment-to-assessment (baseline up to week 10) and the data were restructured into cases, resulting in 1,182 data points (subjects $\times$ visits) for which ratings for both scales were available. Percentage of change in the total group as well as change agreement was analysed by using kappa (κ) statistics and McNemar-Bowker's tests ($P \leq 0.05$).

In order to test the second hypothesis (*The higher sensitivity of the IDS-C_{28} is dependent from the depression severity level—lower severity levels should go along with a higher sensitivity of the IDS-C_{28}*), three depression severity categories were formed applying the cut-off values established by Paykel [30], which differentiate between treatment effects in general practice. In order to facilitate the understanding, we named the categories of Paykel according to the guideline of the National Institute for Health and Clinical Excellence (NICE) [29] as follows: $HAMD_{17}$ score 0–12 = subthreshold, 13–15 = mild, >15 = moderate. Again, the three change directions increase, decrease and no change were characterised in percent for each severity level for both scales from assessment-to-assessment (baseline up to week 10) by restructuring the data (see above) and change agreement between the scales was analysed by using kappa (κ) statistics and McNemar-Bowker's tests ($P \leq 0.05$).

To test the third hypothesis (*The sensitivity for change is dependent on DSM-IV criteria of depression*), we analysed which symptoms of a major depressive episode according to DSM-IV contribute to the total score change. We used the alignment of the items of the IDS-C_{28} with the DSM-IV as well as with the $HAMD_{17}$ [8, 35, 43] in order to create 12 depression criteria (see Table 1), which reflect all nine DSM-IV criteria as well as additional symptoms like anxiety or atypical symptoms. A subject was counted as having a symptom if it is present to any degree (i.e., having a response value of 1, 2, 3, or 4). This resulted in a total number of 1,506 data points (subjects $\times$ visits) for which ratings for both scales were available.

For each scale, the total percentage as well as the percentage of change for the three categories (decrease, increase, no change in total score) was calculated from assessment-to-assessment for each of the 12 domains, both for the total group and for the different severity levels (according to $HAMD_{17}$ cut-off values).

Table 1 DSM-IV criterion symptoms for depression and the items of the IDS-C_{28} and the $HAMD_{17}$

DSM-IV criterion	IDS-C_{28} (item number)	$HAMD_{17}$ (item number)
1. Depressive mood	Mood (sad) (5)	Depressive mood (1)
	Mood (irritable) (6)	
	Mood (anxious) (7)	
	Reactivity of mood (8)	
	Mood variation (9)	
	Quality of mood (10)	
	Outlook (Future) (17)	
2. Loss of interest/pleasure	Involvement (19)	Work and activities (7)
	Pleasure/enjoyment (excludes sexual activities) (21)	Genital symptoms (14)
	Sexual interest (22)	
3. Appetite/weight	Appetite (decreased/increased) (11, 12)	Somatic symptoms gastrointestinal (12)
	Weight (decrease/increase) within the last 2 weeks (13, 14)	Loss of weight within the last week (16)
4. Sleep disturbance	Sleep onset insomnia (1)	Insomnia early (initial insomnia) (4)
	Mid-nocturnal insomnia (2)	Insomnia middle (5)
	Early morning insomnia (3)	Insomnia late (terminal insomnia) (6)
5. Energy/fatigue	Energy/fatigability (20)	Somatic symptoms general (13)
	Somatic complaints (25)	
6. Psychomotoric agitation/slowing	Psychomotor agitation (24)	Agitation (9)
	Psychomotor slowing (23)	Retardation (8)
7. Self-esteem/self-blame	Outlook (self) (16)	Feelings of guilt (2)
8. Concentration/problems in decision-making	Concentration/decision making (15)	
9. Suicidal ideation	Suicidal Ideation (18)	Suicide (3)
Additional symptoms		
10. Anxiety (psychic and somatic)	Mood (anxious) (7)	Anxiety, psychic (10)
	Mood (irritable) (6)	Anxiety, somatic (11)
	Sympathetic arousal (26)	
	Panic/Phobic symptoms (27)	
	Gastrointestinal (28)	
11. Atypical symptoms	Hypersomnia (4)	
	Reactivity of mood (8)	
	Appetite (increased) (12)	
	Weight (increase) within the last 2 weeks (14)	
12. Other symptoms		Hypochondriasis (15)
		Insight (17)

IDS-C_{28} 28-item version of the Inventory of Depressive Symptomatology-Clinician-Rated, *$HAMD_{17}$* 17-item version of the Hamilton Depression Rating Scale

For the remitter of the total sample ($HAMD_{17}$ score ≤ 7, $n = 290$), differences in residual symptoms (defined as any symptom which is still present to any degree (i.e., having a response value of 1, 2, 3, or 4)) were calculated and evaluated with chi-square tests and $df = 1$ ($P \leq 0.05$). Agreement between the scales was evaluated again by κ statistics. All analyses were done using SPSS 17.0.

Results

Patients' characteristics

A total of 340 patients (32.4% men, 67.6% women; mean age ($\pm$SD) = 46.17 $\pm$ 14.82 years) were included in the analysis. The main diagnoses were Double Depression (42.4%), Major Depression (32.1%), Depressive Disorder

NOS (20.6%), Dysthymic Disorder (2.9%), and Subsyndromal Depressive Disorder (2.1%). Almost half of the patients (41.2%) had a psychiatric comorbid diagnosis. The mean baseline sum scores were 16.48 ± 4.25 points on the $HAMD_{17}$ and 27.29 ± 7.52 points on the IDS-C_{28}, indicating a moderate depressive symptomatology at study entry.

Hypothesis 1 Sensitivity to change during the course of antidepressant treatment

The strength of agreement between sum scores of the IDS-C_{28} and the $HAMD_{17}$ was excellent for each assessment (baseline: 0.85, week 2: 0.92, week 4: 0.93, week 6: 0.94, week 8: 0.94, week 10: 0.96). Both scales had a similar pattern of mean raw score changes during the study period (see Fig. 1).

In order to characterise the changes between baseline and each of the five subsequent assessments, we analysed the mean sum score changes and determined effect sizes (see Table 2).

Compared to baseline, the mean sum scores decreased significantly at each measurement occasion in both scales ($P < 0.01$, *t*-tests). The order of effect sizes was consistent for both scales with medium effect sizes between baseline and week 2 (IDS-C_{28}: $d = 0.42$; $HAMD_{17}$: $d = 0.44$) and large effect sizes between baseline and week 10 (IDS-C_{28}: $d = 0.80$; $HAMD_{17}$: $d = 0.79$). Between the scales there was no significant difference for any assessment ($P \geq 0.05$ for each analysis). The analysis of mean sum score changes from assessment-to-assessment revealed that the IDS-C_{28} detected a significant decrease between each of the assessments, while the $HAMD_{17}$ sum scores didn't decline significantly between weeks 4 and 6 as well as between weeks 6 and 8 ($P \leq 0.01$, *t*-tests; see Table 2). Effect sizes were rather small (see Table 2) and comparable for both scales ($P \geq 0.05$ for each analysis).

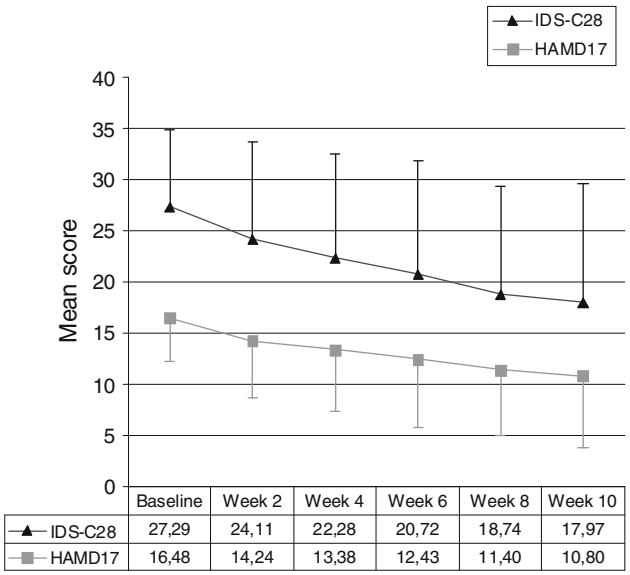

Fig. 1 Mean (±SD (standard deviation)) raw scores for IDS-C_{28} and $HAMD_{17}$ by assessment. *IDS-C_{28}* 28-item version of the Inventory of Depressive Symptomatology-Clinician-Rated, *$HAMD_{17}$* 17-item version of the Hamilton Depression Rating Scale

Table 2 *P*-values and effect sizes for IDS-C_{28} and $HAMD_{17}$ during the course of treatment

	P-value		Effect size (*d*)	
	IDS-C_{28}	$HAMD_{17}$	IDS-C_{28}	$HAMD_{17}$
BL-week 2	<0.01	<0.01	0.42	0.44
BL-week 4	<0.01	<0.01	0.53	0.55
BL-week 6	<0.01	<0.01	0.60	0.61
BL-week 8	<0.01	<0.01	0.83	0.79
BL-week 10	<0.01	<0.01	0.80	0.79
Week 2–4	0.01	0.01	0.19	0.17
Week 4–6	0.01	0.02	0.17	0.15
Week 6–8	0.01	0.03	0.19	0.14
Week 8–10	<0.01	<0.01	0.25	0.23

IDS-C_{28} 28-item version of the Inventory of Depressive Symptomatology-Clinician-Rated, *$HAMD_{17}$* 17-item version of the Hamilton Depression Rating Scale, *BL* baseline

Regarding the direction of change (increase, decrease, no change in total score), the IDS-C_{28} was better able to characterise any change in symptomatology (no change: IDS-C_{28} = 5.4% vs. $HAMD_{17}$ = 10.8%; see Table 3).

The agreement between both scales was moderate with $\kappa = 0.56$. Both scales differed significantly in their ratings (McNemar-Bowker's test; $P \leq 0.05$). Disagreement was highest with 73% in the category 'no change' (see Table 4).

Hypothesis 2 Sensitivity to change in relation to depression severity

Overall the direction of change ratings (increase, decrease, no change in total score) was also consistent for both scales for the different severity levels (see Table 3). However, the IDS-C_{28} detected more changes in symptomatology, especially in the subthreshold level (no change: IDS-C_{28} = 5.2% vs. $HAMD_{17}$ = 12.7%). Both scales differed significantly in their ratings (McNemar-Bowker's test; $P \leq 0.05$) in all categories except the mild ($P = 0.13$). In all severity levels, disagreement between scales was highest for the category 'no change' ranging from 58% up to 88% (see Table 4). Overall, agreement in change in total scores between scales was best for the subthreshold depression level ($\kappa = 0.58$) and medium for the mild ($\kappa = 0.54$) and moderate level ($\kappa = 0.50$). Results were similar when using the depression severity cut-off values for the IDS-C_{28} (data not shown).

Table 3 Percentage of change (increase, decrease and no change in total score) in the total group and the different depression severity levels (according to $HAMD_{17}$ cut-off values) for the IDS-C_{28} and the $HAMD_{17}$

	Total (n = 1,182)		Moderate (n = 444)		Mild (n = 235)		Subthreshold (n = 503)	
	IDS-C_{28}	$HAMD_{17}$	IDS-C_{28}	$HAMD_{17}$	IDS-C_{28}	$HAMD_{17}$	IDS-C_{28}	$HAMD_{17}$
Increase	36.6	35.2	28.8	25.5	34.9	31.1	44.3	45.7
No change	5.4	10.8	5.9	9.2	5.1	9.8	5.2	12.7
Decrease	58.0	54.0	65.3	65.3	60.0	59.1	50.5	41.6

IDS-C_{28} 28-item version of the Inventory of Depressive Symptomatology-Clinician-Rated, *$HAMD_{17}$* 17-item version of the Hamilton Depression Rating Scale

Table 4 Change agreement (κ) between $HAMD_{17}$ and IDS-C_{28} for the total group and the different depression severity levels (according $HAMD_{17}$ cut-off values)

IDS-C_{28}	Agreement with $HAMD_{17}$			
	Disagreement (%)	Increase (%) (n = 416)	No change (%) (n = 128)	Decrease (%) (n = 638)
Total group				
Increase (n = 433)	24	76	9	16
No change (n = 64)	73	39	27	34
Decrease (n = 685)	20	9	11	80
Subgroups according to depression severity				
		(n = 230)	(n = 64)	(n = 209)
Subthreshold				
Increase (n = 101)	15	85	7	8
No change (n = 13)	65	35	35	30
Decrease (n = 81)	28	13	15	72
		(n = 73)	(n = 23)	(n = 139)
Mild				
Increase (n = 133)	32	68	9	23
No change (n = 23)	58	25	42	33
Decrease (n = 243)	18	10	8	82
		(n = 113)	(n = 41)	(n = 290)
Moderate				
Increase (n = 48)	36	64	12	24
No change (n = 11)	88	50	12	38
Decrease (n = 136)	14	6	8	86

IDS-C_{28} 28-item version of the Inventory of Depressive Symptomatology-Clinician-Rated, *$HAMD_{17}$* 17-item version of the Hamilton Depression Rating Scale

Hypothesis 3 Sensitivity to change of DSM-IV criteria of depression

First, the DSM-IV criteria coverage was investigated for the total group (see Table 5).

The incidence was highest for the domains *(1) depressive mood*, *(2) loss of interest/pleasure*, *(4) sleep disturbance*, *(5) energy/fatigue*, *(8) concentration*, *(10) anxiety* and *(11) atypical symptoms* (occurring in at least 70%). The appetite/weight, suicidal ideation, hypochondriasis/insight items were not very common, occurring in less than half of subjects. Most change in symptomatology (increase or decrease, see Online Resource 1) occurred in the domains *(1) depressive mood*, *(2) loss of interest/pleasure*, *(4) sleep disturbance*, *(10) anxiety* and *(11) atypical symptoms*, while *(9) suicidal ideation* and *(12) other symptoms* were relatively stable. Agreement in change between the scales (see Online Resource 1) was highest for the DSM-IV criteria which were measured by an equal

Table 5 DSM-IV depression criteria coverage (in percent) in the total sample and the different depression severity levels for the IDS-C_{28} and the $HAMD_{17}$

DSM-IV criteria	Total (n = 1,506)		Moderate (n = 558)		Mild (n = 268)		Subthreshold (n = 680)	
	IDS-C_{28}	$HAMD_{17}$	IDS-C_{28}	$HAMD_{17}$	IDS-C_{28}	$HAMD_{17}$	IDS-C_{28}	$HAMD_{17}$
1. Depressive mood	94.4	86.0	99.8	99.3	99.6	99.3	87.6	70.0
2. Loss of interest/pleasure	83.1	89.1	98.2	99.8	92.9	97.8	66.9	79.1
3. Appetite/weight	47.7	26.2	66.7	44.8	52.2	25.7	30.3	11.0
4. Sleep disturbance	76.0	76.6	95.0	95.2	83.2	84.0	57.6	58.4
5. Energy/fatigue	91.2	60.0	99.3	84.2	98.1	72.0	81.8	35.4
6. Psychomotoric agitation/slowing	60.8	60.1	83.3	82.3	71.3	70.9	38.1	37.6
7. Self-esteem/self-blame	67.3	64.7	88.4	84.9	73.9	72.0	47.4	45.3
8. Concentration/problems in decision-making	72.0	–	91.3	–	80.8	–	52.7	–
9. Suicidal ideation	37.6	37.4	66.7	66.5	36.9	36.9	14.1	13.7
10. Anxiety (psychic and somatic)	93.9	92.9	99.8	99.5	98.9	98.5	87.1	84.9
11. Atypical symptoms	69.3	–	83.1	–	75.6	–	56.3	–
12. Other symptoms (hypochondriasis, insight)	–	35.9	–	54.7	–	36.9	–	19.9

IDS-C_{28} 28-item version of the Inventory of Depressive Symptomatology-Clinician-Rated, *$HAMD_{17}$* 17-item version of the Hamilton Depression Rating Scale

number of items, i.e., *(9) suicidal ideation* with $\kappa = 0.95$, *(6) psychomotoric agitation/slowing* ($\kappa = 0.75$), *(4) sleep disturbance* ($\kappa = 0.74$) and *(7) self-esteem/self-blame* ($\kappa = 0.63$). Disagreement between both scales was most pronounced in the domain *(5) energy/fatigue* ($\kappa = 0.19$) and the core depressive domain *(1) depressive mood* ($\kappa = 0.32$) as well as in *(2) loss of interest/pleasure* and *(10) anxiety* (both with $\kappa = 0.36$). In the domain *(12) other symptoms*, which was only included in the $HAMD_{17}$, not much change occurred (66% have 'no change').

Next, the data were analysed for the different severity levels separately (see Table 5). In the moderately depressed sample, the coverage of most criteria was quite high with 82–100%. Just the domains *(3) appetite/weight*, *(9) suicidal ideation* and *(12) other symptoms (hypochondriasis, insight)* had a lower incidence (45–67%). In the mildly depressed sample criteria which were also less present (compared to the moderately depressed sample) were *(3) appetite/weight*, *(9) suicidal ideation* and *(12) other symptoms*. In the subthreshold depressed sample criteria which were still common were *(1) depressive mood*, *(2) loss of interest/pleasure*, *(4) sleep disturbance*, *(5) energy/fatigue*, *(8) concentration/problems in decision-making*, *(10) anxiety* and *(11) atypical symptoms*.

Residual symptom coverage was similar for both scales (data not shown). Most common symptoms were anxiety (IDS-C_{28}: 73% vs. $HAMD_{17}$: 69%), depressed mood (IDS-C_{28}: 72% vs. $HAMD_{17}$: 45%), energy/fatigue (IDS-C_{28}: 66% vs. $HAMD_{17}$: 14%) and loss of interest/pleasure (IDS-C_{28}: 46% vs. $HAMD_{17}$: 59%). The IDS-C_{28} detected significantly more residual symptoms than the $HAMD_{17}$ (IDS-C_{28}: 96% (n = 278) vs. $HAMD_{17}$: 94% (n = 272); $\chi^2 = 189.161$; $P < 0.01$).

Differences in symptom change pattern were present in the course of depression. Most change (de- or increase) took place in the moderately depressed sample (see Online Resource 1). Symptom decreases occurred in all depression levels mainly in the domains *(1) depressive mood*, *(2) loss of interest/pleasure*, *(4) sleep disturbance* and *(10) anxiety*. In the mild (both scales) and the subthreshold (only IDS-C_{28}) depression level decrease also took place in the domain *(5) energy/fatigue*. In the IDS specific domain *(8) concentration* more than one-third showed improvement in the mildly and moderately depressed level, while in the subthreshold depressed level, 'no change' was more prominent (54%). Domains which were most stable were *(6) psychomotoric agitation/slowing*, *(9) suicidal ideation*, and *(12) other symptoms*. Agreement between the scales (see Online Resource 1) was worst for the moderate depression level with most κ values in the low to moderate range ($0.26 \leq \kappa \leq 0.66$) and best for the level 'mild' with most κ values in the moderate to excellent range ($0.28 \leq \kappa \leq 0.73$). The highest κ values with up to 0.97 had the domain *(9) suicidal ideation*. For this domain, ratings between scales were quite consistent among depression severity levels. Disagreement was again highest for the domains *(1) depressive mood*, *(2) loss of interest/pleasure*, *(5) energy/fatigue*, and *(10) anxiety* with κ values between 0.11 and 0.41. Results were similar when other depression severity cut-off values (i.e., the cut-off values established by Rush et al. [37]; $HAMD_{17}$ score 0–7 = subthreshold, 8–13 = mild, 14–19 = moderate, >19 = severe) were used (data not shown).

Discussion

This report compares the IDS-C_{28} and the $HAMD_{17}$ regarding its sensitivity to change both from treatment-to-treatment and for different depression severity levels among patients with mild major, minor or subsyndromal depression.

Overall, correspondence in total scores between the IDS-C_{28} and the $HAMD_{17}$ was high ($r \geq 0.84$) and comparable to values found in other studies among depressive patients (e.g., [34–36]). The course of mean symptom change was comparable for both scales: depressive symptoms decreased rapidly in both scales in the early course of treatment (i.e., in the first 2 weeks) and were more stable afterwards. However, the IDS-C_{28} was slightly more sensitive in detecting a significant change between weeks 4, 6 and 8 than the $HAMD_{17}$. This indicates that the IDS-C_{28} is better able to identify small changes in symptomatology in this stage of the treatment. Regarding effect sizes, both scales were equally sensitive to change over time. Effect sizes increased over time and were moderate to large for both scales. Compared to other studies among patients with moderate to severe MDD (e.g., [38, 41]) effect sizes were lower, likely due to the lower depression severity scores of the MIND population, but comparable to the study of patients with MIND [23]. The IDS-C_{28} indicated more changes like increase or decrease in symptomatology than the $HAMD_{17}$. Agreement in change between scales was moderate. In conclusion and in accordance with our hypothesis, the IDS-C_{28} was found to be slightly more sensitive in indicating small changes in symptomatology over the treatment course than the $HAMD_{17}$.

In the subgroups with severity levels (subthreshold, mild, moderate), agreement between both scales was moderate. This was mainly a result of the disagreement in the 'no change' category (up to 88%). Congruent with our second hypothesis, the IDS-C_{28} detected more changes in symptomatology, especially in the lower depression severity level, due to its wider range of items.

Lastly, we investigated, if mood and cognitive symptoms, which are commonly associated with mild depression [18, 32], were more present among subthreshold to mildly depressed patients than among moderately depressed patients. The results of our study support this assumption. Overall, core symptoms for depression in our sample were disturbance in mood, motivation, capacity for pleasure and anxiety symptoms. In these domains, the most change in symptomatology occurred and disagreement between scales was highest. Agreement in change between the scales was highest for the DSM-IV criteria which were measured by an equal number of items, like e.g., *suicidal ideation*.

Regarding the symptom coverage, in the subthreshold to mild depressed sample, mood and cognitive symptoms (domains 1, 2, 5, 7 and 8) came to the fore (beside anxiety symptoms), while in the moderately depressed sample all DSM-IV criteria were highly present (only appetite/weight, suicide, insight and hypochondriasis have a lower incidence). The transition from mild to moderate depression severity was marked by an increase in suicidal ideation and more somatic symptoms like appetite/weight and psychomotor complaints, which are not commonly associated with MIND. Residual symptoms were present in the subsample of the remitters and consisted mainly of anxiety symptoms as well as impairment of mood, work and activities. Overall, the IDS-C_{28} identified more residual symptoms. In summary, through its wider item range (especially for the domains mood, energy/fatigue and concentration), the IDS-C_{28} revealed more differences in symptom coverage than the $HAMD_{17}$.

We also demonstrated that changes in symptoms varied over depression severity levels. Overall, changes mainly occurred in items for which a number of studies have demonstrated high reliability and sensitivity. In all depression severity levels, improvement took place mainly in core symptoms like depressed mood, loss of interest/pleasure, sleep disturbance and anxiety. Items which are more common among inpatients like suicide, psychomotoric agitation/retardation, loss of weight, and somatic symptoms [10] changed most in the moderately depressed sample and remained relatively stable in the other depression severity levels. In agreement with other studies (e.g., [9–11, 25]), the items *hypochondriasis* and *insight* of the $HAMD_{17}$ (criteria *other symptoms*) did not contribute much additional information to the total sum score, because they did not change much. These findings also support the use of uni-dimensional subscales derived from the $HAMD_{17}$ (e.g., [4, 11, 27, 28]).

Study limitations include that the 28-item version instead of the 30-item version of the IDS was used in our sample. Replications using the 30-item version are warranted in order to investigate how the missing items 'interpersonal sensitivity' and 'leaden paralysis/physical energy' contribute to the change pattern. A second limitation is that the same rater completed both interviews (IDS-C_{28} and $HAMD_{17}$); thus, the interviewer was not blinded to the results of the other scale, which could have enhanced agreement between scales. However, by having the same rater judging the symptomatology of one patient, existing differences between the scales are not affected by the inter-rater variability. Rater agreement in our study was high with 95% [17], but it still has an influence on the score ratings with 5% rater disagreement. Because independent ratings also have a negative impact on study time and costs, the advantages of having one rater for both scales outweighed the disadvantages. Another limitation of the study is that the real depression severity change cannot be

validated by an external criterion. However, concurrent validity was demonstrated by high correlation between the scales (over 0.84), indicating that both scales accurately measure depression. The lack of an external validation criterion also means that it cannot be fully excluded that the higher rates of changes detected by the IDS-C_{28} could also be a result of a higher false- positive rate, i.e., the IDS-C_{28} could detect changes in symptomatology where in fact there is no change, while the $HAMD_{17}$ would produce less false positives, e.g., that it does not change when there is no change. An additional source of a potentially higher variance of the IDS-C_{28} is the fact that it has more items than the $HAMD_{17}$; on the other hand, the higher number of items, the more detailed assessment of depressive symptoms and the identical weight of individual items for the total score of the IDS-C_{28} are generally regarded as its advantages over the $HAMD_{17}$ [e.g., 36]. Nevertheless, our results are in line with previous studies providing evidence that the IDS-C_{28} is more sensitive to change than the $HAMD_{17}$, especially in the detection of small improvements [23, 35, 36]. Further studies should focus on the validation of our results by using an external validation criterion like an independent global expert severity assessment (LEAD approach) involving all available data (e.g., clinical, biological and physiological assessments) [25].

In summary, IDS-C_{28} and the $HAMD_{17}$ have been both well able to measure depressive symptomatology in a quite representative sample of depressed primary-care patients with mild major, minor or subsyndromal depression. The IDS-C_{28} has some advantages over the $HAMD_{17}$, because it assesses the symptomatology more thoroughly through its wider range of items and can better detect small changes as well as residual symptoms. Detecting residual symptoms is of high importance because in the clinical praxis, many depressed patients do not fully remit and still have residual symptoms. These patients are at a higher risk of experiencing a relapse [20–22, 39] and show a poorer functioning [26], especially when phobic anxiety symptoms are present [47]. By treating residual symptoms successfully, patients have a better prognosis [12, 31]. Detecting changes in symptomatology during the early course of treatment helps to treat patients better by enabling a prompt reaction. Thus, as already stated [14], the use of the IDS-C is especially useful in research contexts in order to capture symptom profiles, identify subtypes, make decisions over the most effective treatment and predict long-term outcomes or study biological correlates and phenotypes.

So far, the IDS has rarely been used in clinical trials, possibly because of its longer execution time ($HAMD_{17}$ approx. 10–15 min vs. IDS-C_{30} 20–25 min) and its lower popularity and awareness. However, results from a standardized rater training in psychiatric raters for the $HAMD_{17}$ and the IDS-C_{30} [42] has shown that raters, who were unfamiliar with the IDS-C_{30} but familiar with the $HAMD_{17}$, are quickly (i.e., in three rater training sessions) able to learn the appropriate use of the IDS-C_{30}.

In conclusion, the results of our study provide further evidence for the higher sensitivity of the IDS-C than the $HAMD_{17}$ in detecting small changes in depression symptomatology (especially in patients with a rather low depression severity), in tracking the change of cognitive symptoms and in identifying residual symptoms, and thus support the use of the IDS-C in clinical practice as well as in research settings, e.g., in clinical trials evaluating the efficacy of antidepressant treatment.

Acknowledgments We are grateful to all patients who participated in the study, the general practitioners who screened the patients within the network of medical practices, Nuremberg North, especially Veit Wambach, MD, Vanadis Kamm-Kohl, MD, and the practice staff. We are indebted to Patrick Bussfeld, MD, Ute Hägele, MD, Winfried Scheunemann†, Michael Schütze, MD, the physicians who investigated the patients at the specialized study center, Michaela König, MSc, Stephanie Lösch, MSc, Michael Stürmer, MSc who monitored the progress of the study and Professor Hans-Jürgen Möller, MD for his contributions to design and methods of the RCT reported here (trial registration: clinicaltrials.gov; URL: http://www.clinicaltrials.gov; registration number: NCT00226642). We are also thankful to Suzanne Snead, PhD, for the fruitful discussion of the manuscript. Funding for the study was provided by the German Ministry for Education and Research (BMBF; grant: 01 GI 9922/0222/0452) within the promotional emphasis 'German Research Network on Depression and Suicidality'. The BMBF had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Press, Washington DC
2. American Psychiatric Association (APA) (2000) Practice guideline for the treatment of patients with major depressive disorder (revision). Am J Psychiatry 157(suppl. 4):1–45
3. Bagby M, Ryder AG, Schuller DR, Marshall MB (2004) The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry 161:2163–2177
4. Bech P, Rafaelsen OJ (1980) The use of rating scales exemplified by a comparison of the Hamilton and the Bech Rafaelsen Melancholia scale. Acta Psychiatr Scand 62:128–131
5. Bech P, Allerup P, Gram LF, Reisby N, Rosenberg R, Jacobsen O, Nagy A (1981) The Hamilton depression scale. Evaluation of objectivity using logistic models. Acta Psychiatr Scand 63:290–299
6. Bech P, Allerup P, Reisby N, Gram LF (1984) Assessment of symptom change from improvement curves on the Hamilton depression scale in trials with antidepressants. Psychopharmacology 84:276–281

7. Bech P, Stage KB, Nair NP, Larsen JK, Kragh-Sorensen P, Gjerris A (1997) The Major Depression Rating Scale (MDS). Inter-rater reliability and validity across different settings in randomized moclobemide trials. J Affect Disord 42:39–48
8. Drieling T, Schärer LO, Langosch JM (2007) The Inventory of Depressive Symptomatology: German translation and psychometric validation. Int J Methods Psychiatr Res 16:230–236
9. Entsuah R, Shaffer M, Zhang J (2002) A critical examination of the sensitivity of unidimensional subscales derived from the Hamilton Depression Rating Scale to antidepressant drug effects. J Psychiatr Res 36:437–448
10. Evans KR, Sills T, DeBrota DJ, Gelwicks S, Engelhardt N, Santor D (2004) An item response analysis of the Hamilton Depression Rating Scale using shared data from two pharmaceutical companies. J Psychiatr Res 38:275–284
11. Faries D, Herrera J, Rayamajhi J, DeBrota D, Demitrack M, Potter WZ (2000) The responsiveness of the Hamilton Depression Rating Scale. J Psychiatr Res 34:3–10
12. Fava GA, Rafanelli C, Grandi S, Canestrari R, Morphy M (1998) A six-year outcome for cognitive behavioral treatment of residual symptoms in major depression. Am J Psychiatry 155:1443–1445
13. Feiger AD, Bielski RJ, Bremner J, Heiser JF, Trivedi M, Wilcox CS, Roberts DL, Kensler TT, McQuade RD, Kaplita SB, Archibald DG (1999) Double-blind, placebo-substitution study of nefazodone in the prevention of relapse during continuation treatment of outpatients with major depression. Int Clin Psychopharmacol 14:19–28
14. Guillon CM, Rush AJ (1998) Toward a generalizable model of symptoms in major depressive disorder. Biol Psychiatry 44:959–972
15. Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry 23:56–62
16. Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278–296
17. Hegerl U, Hautzinger M, Mergl R, Kohnen R, Schütze M, Scheunemann W, Allgaier A-K, Coyne J, Henkel V (2010) Effects of pharmaco- and psychotherapy in depressed primary care patients. A randomized, controlled trial including a patient choice arm. Int J Neuropsychopharmacol 13:31–44
18. Howland RH, Schettler PJ, Rapaport MH, Mischoulon D, Schneider T, Fasiczka A, Delrahiem K, Maddux R, Lightfoot M, Nierenberg AA (2008) Clinical features and functioning of patients with minor depression. Psychother Psychosom 77:384–389
19. Iannuzzo R, Jaeger J, Goldberg J, Kafantaris V, Subletteet M (2006) Development and reliability of the HAM-D/MADRS interview: an integrated depression symptom rating scale. Psychiat Res 145:21–37
20. Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W, Paulus MP, Kunovac JL, Leon AC, Mueller TI, Rice JA, Keller MB (1998) A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. Arch Gen Psychiatry 55:694–700
21. Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W, Paulus MP, Kunovac JL, Leon AC, Mueller TI, Rice JA, Keller MB (1998) Major depressive disorder: a prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. J Affect Disord 50:97–108
22. Judd LL, Paulus MJ, Schettler PJ, Akiskal HS, Endicott J, Leon AC, Maser JD, Mueller T, Solomon DA, Keller MB (2000) Does incomplete recovery from first lifetime major depressive episode herald a chronic course of illness? Am J Psychiatry 157:1501–1504
23. Judd LL, Rapaport MH, Yonkers KA, Rush AJ, Frank E, Thase ME, Kupfer DJ, Plewes JM, Schettler PJ, Tollefson G (2004) Randomized, placebo-controlled trial of fluoxetine for acute treatment of minor depressive disorder. Am J Psychiatry 161:1864–1871
24. Lauge N, Behnke K, Søgaard J, Bahr B, Bech P (1998) Responsiveness of observer rating scales by analysis of number of days until improvement in patients with major depression. Eur Psychiatry 13:143–145
25. Maier W (1990) The Hamilton depression scale and its alternatives: a comparison of their reliability and validity. In: Bech P, Coppen A (eds) The Hamilton scales. Springer, Berlin, pp 64–71
26. Miller IW, Keitner GI, Schatzberg AF, Klein DN, Thase ME, Rush AJ, Markowitz J, Schlager DS, Kornstein SG, Davis SM, Harrison WM, Keller MB (1998) The treatment of chronic depression, part 3: psychosocial functioning before and after treatment with sertraline or imipramine. J Clin Psychiatry 59:608–619
27. Möller H-J (2001) Methodological aspects in the assessment of severity of depression by the Hamilton Depression Scale. Eur Arch Psychiatry Clin Neurosci 251(suppl. 2):13–20
28. Montgomery SA, Åsberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatry 134:382–389
29. National Institute for Health, Clinical Excellence (2009) Depression: the treatment and management of depression in adults. National Clinical Practice Guideline 90. National Institute for Health and Clinical Excellence, London
30. Paykel ES (1990) The use of the Hamilton Depression Scale in general practice. In: Bech P, Coppen A (eds) The Hamilton Scales. Springer, Berlin, pp 40–48
31. Paykel ES, Scott J, Teasdale JD, Johnson AL, Garland A, Moore R, Jenaway A, Cornwall PL, Hayhurst H, Abbott R, Pope M (1999) Prevention of relapse in residual depression by cognitive therapy: a controlled trial. Arch Gen Psychiatry 56:829–835
32. Rapaport MH, Judd LL, Schettler PJ, Yonders KA, Thase ME, Kupfer DJ, Frank E, Plewes JM, Tollefson GD, Rush AJ (2002) A descriptive analysis of minor depression. Am J Psychiatry 159:637–643
33. Rehm LP, O′Hara MW (1985) Item characteristics of the Hamilton rating scale for depression. J Psychiat Res 19:31–41
34. Rush AJ, Gilles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C (1986) The Inventory for Depressive Symptomatology (IDS): preliminary findings. Psychiatry Res 18:65–87
35. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi HM (1996) The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 26:477–486
36. Rush AJ, Carmody TJ, Reimitz PE (2000) The Inventory of Depressive Symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. Int J Methods Psychiatr Res 9:45–59
37. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB (2003) The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 54:573–583
38. Rush AJ, Carmody TJ, Ibrahim HM, Trivedi HM, Biggs MM, Shores-Wilson K, Crismon ML, Toprac MG, Kashner TM (2006) Comparison of self-report and clinician ratings on two inventories of depressive symptomatology. Psychiatr Serv 57:826–837
39. Thase ME, Simons AD, McGeary J, Cahalane JF, Hughes C, Harden T, Friedman E (1992) Relapse after cognitive behaviour therapy of depression: potential implications for longer courses of treatment. Am J Psychiatry 149:1046–1052
40. Trivedi MH, Rush AJ, Ibrahim HM, Carmody TJ, Biggs MM, Suppes T, Crismon ML, Shores-Wilson K, Toprac MG, Dennehy EB, Witte B, Kashner TM (2004) The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self Report

(IDS-SR) and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. Psychol Med 34:73–82
41. Vittengl JR, Clark LA, Kraft D, Jarrett R (2005) Multiple measures, methods, and moments: a factor-analytic investigation of change in depressive symptoms during acute-phase cognitive therapy for depression. Psychol Med 35:693–704
42. Wagner S, Helmreich I, Lieb K, Tadić A (2011) Standardized rater training for the Hamilton Depression Rating Scale ($HAMD_{17}$) and the Inventory of Depressive Symptoms (IDS_{C30}). Psychopathology 44:68–70. doi:10.1159/000318162)
43. Williams JBW (1988) A structured interview guide for the Hamilton Depression Rating Scale. Arch Gen Psychiatry 45:742–747
44. Williams JBW (2001) Standardizing the Hamilton Depression Rating Scale: past, present, and future. Eur Arch Psychiatry Clin Neurosci 251(suppl 2):6–12
45. Wittchen H-U, Pfister H (eds) (1997) DIA-X-interview. Instruktionsmanual zur Durchführung von DIA-X Interviews. Swets & Zeitlinger, Frankfurt am Main
46. World Health Organization (WHO) (1993) Composite international diagnostic interview, version 1.1. World Health Organization, Geneva
47. Yang H, Chuzi S, Sinicropi-Yao L, Johnson D, Chen Y, Clain A, Baer L, McGrath PJ, Stewart JW, Fava M, Papakostas GI (2010) Type of residual symptom and risk of relapse during the continuation/maintenance phase treatment of major depressive disorder with the selective serotonin reuptake inhibitor fluoxetine. Eur Arch Psychiatry Clin Neurosci 260:145–150