CEU

- Trabajo realizado por la Biblioteca Digital de la Universidad CEU-San Pablo

- Me comprometo a utilizar esta copia privada sin finalidad lucrativa, para fines de investigación y docencia, de acuerdo con el art. 37 de la M.T.R.L.P.I. (Modificación del Texto Refundido de la Ley de Propiedad Intelectual del 7 julio del 2006)

13

JAD 00954

Concordance between clinical and self-report depression scales during the acute phase and after treatment

Carmen Senra [a] and Aquilino Polaino [b]

[a] Universidad de Santiago de Compostela, Santiago, Spain and [b] Universidad Complutense de Madrid, Madrid, Spain

Summary

The concordance between self-report and clinical rating scales of depression increases during progress from the acute depressive episode to recovery or improvement of symptoms. We investigated this convergence in a group of 52 outpatients with DSM-III major depression disorders using three widely employed depression scales and their parallel formats (i.e., alternative modes of administering the scales). The six instruments were applied at admission and after 12 and 24 weeks' treatment. The results indicate that the increase in the global concordance between scales may be a statistical effect deriving from broadening of the range of scores.

Introduction

Several studies have found that there is less concordance between self-rating and clinical rating depression scales during the acute depressive episode, prior to treatment, than upon repeated rating after improvement (Bailey and Coppen, 1976; Davies et al., 1975; Dratcu et al., 1987; Fava et al., 1986; Prusoff et al., 1972a,b; White et al., 1984). After examining possible sources of the discrepancy between patient and clinician before and after treatment, Paykel et al. (1970, 1973), Paykel and Prusoff (1973) and Prusoff et al. (1972a) reported that at the height of the acute illness the severe characteristics of the episode play an important role in such discrepancy. Specifically, at the time of the patient's admission the clinician's evaluation depends not only on the patient's verbal report but also on numerous clinical observations of facial expression, gestures, postural shifting, speech quality and rate, etc; however, his contact with the patient's internal world is limited, and his evaluation is affected by

---

Correspondence to: Carmen Senra Rivera, Departamento de Psicología Clínica, Universidad de Santiago de Compostela, 15705 Santiago, La Coruña, Spain.

his own biases. The patient himself, even though reasonably capable of evaluating his condition, is hardly in a position to place himself on a wide spectrum of depressions, and self-report scales involve systematic elements favouring discrepancy, such as age, personality traits and severity.

Concordance between clinical rating and self-rating increases as the patient's condition improves. Prusoff et al. (1972a) attribute this to the fact that during follow-up the clinician's evaluation depends largely on the patient's verbal report, observations playing only a minor role. Thus post-treatment evaluation does not reflect the severity of the episode but merely the presence or absence of symptoms (Prusoff et al., 1972a).

Biggs et al. (1978) have suggested that the increase in concordance between clinical rating and self-rating is basically a statistical effect related to score range; that is, at the time of admission all patients fall in a rather narrow severity range, but the range broadens during treatment because of differences in the rate at which patients improve.

After re-examining all the above-mentioned factors, Paykel and Norton (1986) concluded that it was not yet clear which of them was responsible for the increase in concordance.

The purpose of the present study was to ascertain whether the increase in concordance between self-evaluations and clinical assessments during treatment is indeed a statistical effect or whether it is due to clinical improvement removing the characteristics that caused the initial discrepancy. To rule out discrepancy due to the use of different scales by interviewer and patient, patients with DSM-III diagnoses of major depression disorders were administered the standard formats of the Beck Depression Inventory (BDI; Beck et al., 1961), the Self-Rating Depression Scale (SDS; Zung, 1965) and the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960), and also their parallel formats; i.e., clinical interview forms of the BDI and SDS and a self-report form of the HRSD. The scale used as the parallel format of the HRSD was the Carroll Rating Scale (CRS; Carroll et al., 1981), a direct self-rated adaptation of the original HRSD, to which it corresponds closely. For the parallel formats of the BDI and SDS, the symptom items

of the standard self-reports formats were used in scoring information elicited by interview in a way that maintained the item content, rating time frame and method of rating symptoms (by frequency or by severity) of the original self-report formats.

Method

Subjects

The study was carried out in the Neuropsychiatric Out-patients Service of the La Robleda Psychiatric Hospital, Santiago de Compostela, Spain. The sample comprised 52 patients with DSM-III major depression disorders; 34 (65.4%) were women and 18 (34.6%) men. Their mean age was 34.7 years (SD = 8.32 years). Twenty-seven patients were married, nineteen were single, five were widowed and one was divorced. No patients had a diagnosis of DSM-III bipolar or melancholia disorders, none had a history of addiction other than to tobacco, all were drug-free for at least 2 weeks before being interviewed, and all were willing and able to complete the inventories.

Instruments

Semistructured Diagnostic Interview (SDI). Patients were selected by means of an ad hoc semistructured interview (Senra, 1990) that both fulfilled DSM-III guidelines for diagnosis of major depression disorders and included items of the interview forms of the scales used. The SDI consisted of five sections recording 1) personal data; 2) essential signs and symptoms; 3) concomitant signs and symptoms; 4) disqualifying symptoms; and 5) the profile of the episode.

The Beck Depression Inventory (BDI).
The BDI quantifies the severity of 21 depressive symptoms. Though originally designed to be administered by trained interviewers, it is now generally used as a self-administered scale (Steer et al., 1986). Each of the 21 items is composed of four alternative statements rating the severity of a symptom from 0 to 3. The Spanish version of the BDI was developed and validated by Conde and Useros (1974) and Conde et al. (1976). In this study we used both the standard 21-item self-rated

version of the Inventory (B-SR) and a clinical interview format (B-CI). Administration of the latter did not consist in the interviewer simply asking the patient the same questions as are asked on the self-report form, but rather in the interviewer using the alternative assertions listed in the BDI items as a guide for the translation of elicited information into scores for the corresponding symptoms. Mood, for example, was scored 0–3 depending upon which of the mood assertions constituting the first DBI item was most closely matched, in the interviewer's opinion, by the patient's account of his or her symptoms.

The Zung Self-Rating Depression Scale (SDS).

The SDS comprises 20 items scored 1–4 according to symptom frequency. The Spanish version has been developed and validated by Conde and coworkers (Conde et al., 1970a,b; Conde and Esteban, 1973; Conde and Sanchez, 1969). In this study we used both the standard self-rating format of the scale (S-SR) and a clinical interview format of identical content and scoring structure (S-CI; Conde and Franch, 1984). The first question of the S-CI, for example, is 'How often have you felt sad or depressed in the past week, including today'?, and is scored 1–4 depending on whether the patient's answer comes closest, in the interviewer's opinion, to 'None or a little of the time' (1), 'Some of the time' (2), 'Good part of the time' (3), 'Most of the time' (4).

The Hamilton Rating Scale for Depression (HRSD).

The HRSD is designed to quantify the results of clinical interviews of patients with diagnoses of depressive affective disorders (Hamilton, 1960, 1967). The original version contained 21 items, but only 17 were considered necessary for evaluation of the overall severity of depression, the omitted items being diurnal variation, derealization, paranoid symptoms and obsessional symptoms. Its administration has been standardized by the introduction of the Structured Interview Guide for the HRSD (SIGH-D; Williams, 1988). In this study we used the standard 17-item version (H-CI).

The Carroll Rating Scale for Depression (CRS).

The CRS was designed 'as a direct self-rated adaptation of the original 17-item HRSD' (Carroll et al., 1981). The HRSD items scored 0–4 are represented in the CRS by four yes/no questions of increasing severity, the HRSD items scored 0–2 by two yes/no questions, so that the CRS consists of a total of 52 questions. Although it does not retain the same method of rating severity as the HRSD, it closely matches the latter in item specificity and content (Feinberg et al., 1981; Nasr et al., 1984; Smouse et al., 1981). In this study we used the standard format of the CRS (C-SR) as the self-rated format parallel to the HRSD.

Procedure and data analysis

Before patients were selected for the study, inter-rater reliability was evaluated by having two raters jointly interview 20 patients not included in the study and (independently) rate them on the three interview forms. The inter-rater product-moment correlations were 0.80 for the SDS, 0.89 for the HRSD, and 0.93 for the BDI.

The patients were tested in the course of the routine evaluation procedure of the Outpatients Service. They were told the purpose of the research and the importance of reporting their symptoms as accurately as possible. Patient selection began with the SDI. The patients fulfilling initial criteria for inclusion in the study then rated on the B-CI, S-CI and H-CI by a psychologist (the same for all patients) who was experienced in conducting interviews for this purpose. Immediately afterwards, the patients filled in the B-SR, S-SR and C-SR in a quiet room under the supervision of another psychologist (the same for all patients) who was also experienced in the use of these instruments and who had no knowledge of the patients' interview ratings. Both the self-report and interview formats of the three scales were first administered at the time of admission, and all referred to the patient's state in 'the past week, including today'. All scales were administered in the morning to minimize differences associated with circadian variations. After scale administration, an experienced psychiatrist ignorant of the scale ratings had a routine diagnostic interview with the patients and prescribed medi-

cation. The final requisite for a patient's inclusion in the study was that the psychiatrist should diagnose DSM-III major depression; only two patients were excluded by this criterion. The six-scale procedure was again applied after 12 and 24 weeks of treatment with tricyclic antidepressants. Of the 52 patients with whom the study began, 45 continued to completion.

The hypothesis that the concordance between self-report and interview increases during treatment was evaluated in three ways.

1) The Pearson correlations among the total scores of the three standard formats were calculated for each time point, and for each scale pair the significance of differences among the correlations at the different time points was evaluated.

2) The same procedure as above was applied to Pearson correlations between the total scores of corresponding SR and CI formats, i.e. B-SR/B-CI, S-SR/S-CI, C-SR/H-CI.

3) For each scale (BDI, SDS or HRSD/CRS), the following procedure was carried out to determine the evolution of SR-CI concordance at item level. First, for each item of the scale and each time point, the correlation between the SR and CI scores for that item was calculated (45 data points, one per subject, at each time point). Next, the three sets of itemwise SR-CI correlations so obtained (one set for each time point) were treated as distinct correlation samples, and

TABLE 1

Means ± SD's and ranges (in parentheses) of the total scores for all scale formats and evaluation times

Scale	Pre-treatment ($N = 52$)	12 week ($N = 49$)	24 week ($N = 45$)
B-SR	25.18 ± 3.78 (19–35)	16.53 ± 6.91 (5–30)	11.84 ± 6.90 (0–27)
B-CI	22.94 ± 2.88 (18–32)	14.26 ± 6.43 (5–29)	9.68 ± 5.94 (0–26)
S-SR	51.11 ± 2.52 (46–59)	43.69 ± 6.60 (31–55)	38.02 ± 8.75 (20–53)
S-CI	48.63 ± 2.30 (44–55)	40.89 ± 6.48 (30–52)	35.44 ± 7.47 (20–50)
C-SR	23.57 ± 2.95 (20–31)	14.18 ± 5.83 (6–26)	9.80 ± 6.72 (0–24)
H-CI	20.79 ± 2.07 (17–25)	11.46 ± 5.03 (5–23)	7.06 ± 6.11 (0–20)

TABLE 2

Pearson correlations among the standard Beck, Zung, and Hamilton scales on admission and after 12 and 24 weeks of treatment

Week	N	BDI-SDS	BDI-HRSD	SDS-HRSD
0	52	0.61 *	0.51 *	0.48 *
12	49	0.85 *	0.76 *	0.62 *
24	45	0.92 *	0.89 *	0.82 *

* $P < 0.001$.

the null hypothesis that the samples for the three different time points belonged to the same population of correlations was tested a) with the Wilcoxon matched pairs test, and b) by one-way analysis of variance.

Results

Total scores

Table 1 lists the mean total score, with its standard deviation and range, of each format used at each time point. Note that for each scale, without exception, the range of scores increases progressively with increasing treatment time: from 16–27 for the B-SR, from 14–26 for the B-CI, from 13–33 for the S-SR, from 11–30 for the S-CI, from 11–24 for the C-SR, and from 8–20 for the H-CI.

Tables 2 and 3 list, for each time point, Pearson correlations among the standard formats of the scales (Table 2) and between the SR and CI formats of each pair of formats, the pairwise differences among the correlations for the three time points were all statistically significant, with the sole exception of the difference between the SDS/HRSD correlations at times 0 (admission) and 12 weeks ($P = 1.9$, n.s., Table 4).

TABLE 3

Pearson correlations between the parallel formats of the scales on admission and after 12 and 24 weeks of treatment

Week	N	B-SR/B-CI	S-SR/S-CI	C-SR/H-CI
0	52	0.71 *	0.65 *	0.70 *
12	49	0.85 *	0.82 *	0.82 *
24	45	0.93 *	0.93 *	0.92 *

* $P < 0.001$.

TABLE 4

Inter-time comparisons of inter-format (total score) correlations

Inter/ time	BDI/ SDS	BDI/ HRSD	SDS/ HRSD	B-SR/ B-CI	S-SR/ S-CI	C-SR/ H-CI
0–12	$t = 10.0$ *	$t = 5.99$ *	$t = 1.9$ n.s.	$t = 4.45$ *	$t = 5.41$ *	$t = 3.40$ *
12–24	$t = 2.88$ [a]	$t = 2.78$ [a]	$t = 2.49$ [b]	$t = 2.39$ [b]	$t = 3.24$ *	$t = 2.67$ [b]
0–24	$t = 12.7$ *	$t = 8.79$ *	$t = 4.43$ *	$t = 6.85$ *	$t = 8.65$ *	$t = 6.08$ *

* $P < 0.001$; [a] $P < 0.01$; [b] $P < 0.05$.

TABLE 5

Wilcoxon matched pairs test t values for inter-time differences in itemwise SR-CI correlation, for each scale. ($N = 45$)

	Pre-treat- ment– 12 week	12 week– 24 week	Pre-treat- ment– 24 week
B-SR/B-CI	$t = 46.5$	$t = 93$	$t = 55$
	$t_{0.05} = 46$	$t_{0.05} = 59$	$t_{0.05} = 52$
S-SR/S-CI	$t = 55.5$	$t = 94.5$	$t = 68.5$
	$t_{0.05} = 52$	$t_{0.05} = 52$	$t_{0.05} = 40$
C-SR/H-CI	$t = 52$	$t = 44.5$	$t = 27.5$
	$t_{0.05} = 30$	$t_{0.05} = 25$	$t_{0.05} = 21$

Analysis at item level

Table 5 presents, for each scale, the Wilcoxon's t values obtained as described in Methods for the inter-time differences between SR-CI item score correlations. None of the inter-time differences proved significant. The same conclusion, i.e., that for each scale the differences among the item score correlation samples afforded by the SR-CI pairs at the three time points are not statistically significant, was also reached using Fisher's transformation of the correlation coefficients and a one-way analysis of variance with the time points as factor levels: for S-SR/S-CI correlations $F(38/2) = 1.34$ ($P = 0.27$, $\epsilon = 0.73$); B-SR/B-CI $F(40/2) = 2.97$ ($P = 0.06$); and for C-SR/H-CI $F(32/2) = 0.91$ ($P = 0.41$).

Discussion

In keeping with the general pattern observed in the literature, correlation between total self-report and clinical rating scores was poorer during the acute phase of the depressive episode than after treatment, when patients had improved symptomatically. This held both for correlations between standard self-report and standard clinical interview scales (Tables 2 and 4) and for correlations between the parallel SR and CI formats of the same scale (Tables 3 and 4), i.e., for SR and CI formats with no difference in content. The range of total scores on each scale increased considerably during treatment (Table 1), which was to be expected because patients differ in the speed and degree of their response to treatment.

When the scales were treated as samples of items, SR-CI concordance did not increase with treatment time (Table 5). This strongly suggests that the statistically significant improvements in total score correlations (Table 4) are not in fact due to improved concordance between clinician and patient, but are a statistical artifact due to the increase in total score ranges. Our results thus support the conclusions of Biggs et al. (1978). Likewise in keeping with this notion are the results of Fava et al. (1986), who in evaluating the sensitivity of three scales for detecting changes in depression found that two out of three correlations between observer-rated and self-rated measurements increased after a month's treatment in a group of patients suffering from depression but not in a group of normal controls whose score range presumably varied little between evaluations.

In spite of the above findings, it is important to point out that, the diagnostic characteristics of our subjects limit the extent to which this study tests the hypothesis put forward by Paykel and coworkers, since we were unable to examine discrepancies in major depression cases with symptoms of psychosis or melancholia. It seems likely

that the self-reports of patients suffering from these 'more severe' forms of depression are distorted by the patients' altered perception and testing of reality, and that in such cases a significant improvement in itemwise SR-CI correlation might occur during treatment.

In conclusion, the results reported here, in keeping with our previous findings (Polaino and Senra, 1991), corroborate the utility of self-reports for quantitating the severity of the symptoms and monitoring the progress of patients being treated for major depression with no psychotic or melancholic disorders. Note, however, that our conclusions strictly apply only to the Spanish versions of the scales used. Future research should extend the kind of comparison carried out in this work to depressive patients with other diagnoses.

Acknowledgements

References

Bailey, J. and Coppen, A. (1976) A comparison between the Hamilton Rating Scale and the Beck Inventory in the measurement of depression. Br. J. Psychiatry 128, 486–489.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. and Erbaugh, J. (1961) An inventory for measuring depression. Arch. Gen. Psychiatry 4, 561–571.

Biggs, J.T., Wylie, L.T. and Ziegler, V.E. (1978) Validity of the Zung Self-rating Depression Scale. Br. J. Psychiatry 132, 381–385.

Carroll, B.J., Feinberg, M., Smouse, P.E., Rawson, S.G. and Greden, J.F. (1981) The Carroll rating scale for depression. I. Development, reliability and validation. Br. J. Psychiatry 138, 194–200.

Conde, V., Escriba, J. and Izquierdo, J. (1970a) Estudio crítico de la escala autoaplicada para la medida cuantitativa de la depresión de Zung. Rev. Clin. Esp. 117, 71–78.

Conde, V., Escriba, J. and Izquierdo, J. (1970b) Evaluación estadística y adaptación castellana de la escala autoaplicada para la depresión de Zung. Arch. Neurobiol. 33, 281–303.

Conde, V. and Esteban, T. (1973) Revisión crítica de dos adaptaciones castellanas de la "Self-Rating Depression Scale (SDS)" de Zung. Arch. Neurobiol. 34, 375–392.

Conde, V., Esteban, T. and Useros, E. (1976) Revisón crítica de la adaptación castellana del cuestionario de Beck. Rev. Psic. Gral. Aplic. 31, 469–497.

Conde, V. and Franch, J.I. (1984) Escalas de evaluación comportamental para la cuantificación de la sintomatología psicopatológica en los trastornos angustiosos y depresivos. Upjohn, Madrid.

Conde, V. and Sanchez, J. (1969) La escala autoaplicada para la depresión de Zung. Arch. Neurobiol. 33, 535–558.

Conde, V. and Useros, E. (1974) El inventario para la medida de la depresión de Beck. Rev. Psiq. Psic. Med. Eur. Am. Lat. 12, 153–167.

Davies, B., Burrows, G. and Poynton, C. (1975) A comparative study of depression rating scales. Aust. NZ J. Psychiatry 9, 21–24.

Dratcu, L., Costa Ribeiro, L. and Calil, H.M. (1987) Depression assessment in Brazil. The first application of the Montgomery-Asberg Depression Rating Scale. Br. J. Psychiatry 150, 797–800.

Fava, G.A., Kellner, R., Lisansky, J., Park, S., Perini, G. and Zielezny, M. (1986) Rating depression in normals and depressives: Observer versus self-rating scales. J. Affect. Disord. 11, 29–33.

Feinberg, M., Carroll, B.J., Smouse, P. and Rawson, S.G. (1981) The Carroll rating scale for depression. III. Comparison with other rating instruments. Br. J. Psychiatry 138, 205–209.

Hamilton, M. (1960) A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.

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

Nasr, S.J., Altman, E., Rodin, M.B., Jobe, T.H. and Burg, B. (1984) Correlation of the Hamilton and Carroll depression rating scales: A replication study among psychiatric outpatients. J. Clin. Psychiatry 45, 167–168.

Paykel, E.S., Klerman, G.L. and Prusoff, B.A. (1970) Treatment setting and clinical depression. Arch. Gen. Psychiatry 22, 11–21.

Paykel, E.S. and Norton, K.R.W. (1986) Self-report and clinical interview in the assessment of depression. In: N. Sartorius and T.A. Ban (Eds.), Assessment of depression. Springer-Verlag, Berlin, pp. 356–366.

Paykel, E.S. and Prusoff, B.A. (1973) Response set and observer set in the assessment of depressed patients. Psychol. Med. 3, 209–216.

Paykel, E.S., Prusoff, B.A., Klerman, G.L. and DiMascio, A. (1973) Self-report and clinical interview ratings in depression. J. Nerv. Ment. Dis. 156, 166–182.

Polaino, A. and Senra, C. (1991) Measurement of depression: Comparison between self-reports and clinical assessments of depressed outpatients. J. Psychopathol. Behav. Assess. 13, 313–324.

Prusoff, B.A., Klerman, G.L. and Paykel, E.S. (1972a) Concordance between clinical assessment and patients self-report in depression. Arch. Gen. Psychiatry 26, 546–552.

Prusoff, B.A., Klerman, G.L. and Paykel, E.S. (1972b) Pitfalls in the self-report assessment of depression. Can. Psychiatr. Asssoc. J. 17, 101–107.

Senra, C. (1990) Semistructured diagnostic interview for DSM-III major depression. (Available from Dr. C. Senra,

Dpto. Psicología Clínica, 15705 Santiago, La Coruña, Spain).

Smouse, P.E., Feinberg, M., Carroll, B.J., Park, M.H. and Rawson, S.G. (1981) The Carroll rating scale for depression. II. Factor analyses of the feature profiles. Br. J. Psychiatry 138, 201–204.

Steer, R.A., Beck, A.T. and Garrison, B. (1986) Applications of the Beck Depression Inventory. In: N. Sartorius and T.A. Ban (Eds.), Assessment of depression. Springer-Verlag, Berlin, 123–142.

White, J., White, K. and Razani, J. (1984) Effects of endogenicity and severity on consistency and standard depression rating scales. J. Clin. Psychiatry 45, 260–261.

Williams, J.B.W. (1988) A Structured Interview Guide For the Hamilton Depression Rating Scale. Arch. Gen. Psychiatry 45, 742–747.

Zung, W.W.K. (1965) A self-rating depression scale. Arch. Gen. Psychiatry 12, 63–70.