

Patients' feedback after computer-assisted diagnostic interviews for mental disorders

JÜRGEN HOYER, UWE RUHL, DENIS SCHOLZ, & HANS-ULRICH WITTCHEN

Institute of Clinical Psychology and Psychotherapy, Technical University of Dresden, Dresden, Germany

(Received 29 September 2004; revised 11 May 2005; accepted 14 June 2005)

Abstract

Little is known about how psychotherapy patients perceive and evaluate computer-assisted diagnostic interviews for mental disorders. Using the World Health Organization's Composite International Diagnostic Interview in its computer-administered form, psychologists interviewed 236 psychotherapy patients, who evaluated the interview with regard to content, comprehensibility, and acceptance. More than 87% of patients evaluated the interview positively. Higher symptom severity and comorbidity, but not depression, were associated with a slightly lower but still favorable appraisal. The results indicate that the use of computerized clinical diagnostic interviews, previously usually restricted to research, seems to be a time-efficient, economical, and acceptable approach for the diagnostic phase of psychotherapy. The indications of diminished acceptability among multimorbid and severely disturbed patients warrant further study. Implications for quality assurance and practice research networks are discussed.

Zusammenfassung

Die Bewertung computer-gestützter diagnostischer Interviews für psychische Störungen durch die Patienten. Wir wissen wenig darüber, wie Psychotherapiepatienten computergestützte Interviews für psychische Störungen erleben und bewerten. Trainierte Psychologen untersuchten 236 konsekutive Patienten einer Psychotherapieambulanz mit dem Composite International Diagnostic Interview (CIDI) in seiner computergestützten Version. Danach beurteilten die befragten Patienten das Interview auf der Basis einer 15 Items umfassenden Liste hinsichtlich Inhalt, Verständlichkeit und Akzeptanz. 87% der Patienten bewerteten das Interview positiv, und gaben z.B. an froh zu sein, dass „die Befragung so genau und ausführlich war“. Größere Symptombelastung und höhere Komorbidität, nicht aber das Ausmaß an Depression, waren mit einer etwas schlechteren, absolut gesehen aber immer noch guten Bewertung des Interviews assoziiert. Die Ergebnisse zeigen, dass computergestützte Interviews, die bisher fast nur im Forschungsbereich eingesetzt wurden, eine zeit- und kosteneffiziente sowie für den Patienten akzeptable Möglichkeit für die Eingangsdagnostik psychotherapeutischer Behandlungen darstellen. Die Hinweise auf eine geringere Akzeptanz bei multimorbiden und stärker beeinträchtigten Patienten sollten weiter untersucht werden. Implikationen für die Qualitätssicherung und Praxisforschungsnetzwerke werden diskutiert.

A comprehensive, reliable, and valid diagnosis reflecting patients' problems and diagnostic status according to established diagnostic classificatory systems (International Classification of Diseases, 10th ed. [ICD-10]; World Health Organization [WHO], 1993, and Diagnostic and Statistical Manual of Mental Disorders, 4th ed. [DSM-IV]; American Psychiatric Association, 1994) is a mandatory prerequisite for research and treatment. However, beyond the pros of clinical or categorical methods, the cons, mainly their notoriously low interrater reliability and validity, have been discussed since Meehl's seminal analysis in 1954. With the introduction of

explicit descriptive diagnostic criteria in DSM, some of the reservations have been resolved, although routine diagnostic decisions may remain of doubtful reliability (Basco et al., 2000; Steiner, Tebes, Sledge, & Walker, 1995).

In response to this problem, two types of diagnostic interviews were developed for research purposes that revealed increased reliability, similar to that of routine laboratory measures: (a) clinical interviews (e.g., Schedules for Clinical Assessment in Neuropsychiatry [SCAN]; Wing et al., 1990; Structured Clinical Interview for DSM-IV-TR [SCID]; First, Spitzer, Gibbon, & Williams, 2002; Anxiety Disorders Interview Schedule for DSM-IV [ADIS]; Brown, DiNardo, & Barlow, 1994) requiring trained and experienced clinicians and (b) nonclinical interviews (e.g., Diagnostic Interview Schedule [DIS]; Robins, Helzer, Croughan, & Ratcliff, 1981) or the WHO's (1990) Composite International Diagnostic Interview (CIDI), which can be reliably administered even by laypersons for most diagnostic classes. The major procedural difference is that clinical interviews are only loosely standardized, giving clinical rating a considerable weight, whereas nonclinical interviews are fully standardized (i.e., all mandatory questions are explicitly spelled out along with specified rules of coding and analysis).

This strict standardization of nonclinical interviews such as the CIDI makes them also applicable in a computer-administered form and facilitates a cost-efficient and reliable diagnosis. Computer-assisted interviews have, therefore, become the standard procedure in many epidemiological and clinical studies. Within clinical practice, however, such instruments that have been originally developed for epidemiological surveys and mental health research still seem to be rarely used. One reason may be the considerable restriction for the clinician, who cannot apply the interview in a flexible and adaptive manner. Alternatively, this restriction may have important methodological and practical advantages like an enhanced quality and comparability of diagnoses and the fact that trained staff members are able to conduct the interview in addition to diagnosticians. The latter aspect can also help facilitate naturalistic psychotherapy studies.

Despite the extensive theoretical debate among experts concerning the most favorable diagnostic strategy in different practice and research fields, the respondents in epidemiological studies or the patients in clinical studies themselves have, at least to our knowledge, not systematically been asked about their subjective appraisal of such an interview situation. This information, however, seems crucial when it involves the transfer of diagnostic procedures from a research setting to clinical routine practice.

The subjective appraisal of a test and the demands involved for the recipients are often neglected aspects of a test, which have been summarized under the label of test acceptance (Hausknecht, Day & Thomas, 2004; Kubinger, 2006). Test acceptance is a pragmatic criterion mainly relevant for practical test use. Test acceptance may be high for the majority of self-report measures and must not necessarily be studied in this area. However, with regard to more time-consuming and demanding test applications as are necessary in structured clinical interviewing, test acceptance becomes more central.

In the present investigation, we explored how patients perceived and rated standardized diagnostic interviews as part of their diagnostic workup. Using a standardized feedback questionnaire, administered directly after the interview sessions with the CIDI, we aimed to explore test acceptance.

Test acceptance may be associated not only with characteristics of the test itself but also with characteristics of the respondents. For instance, lower education may influence the comprehensibility of a test instruction. Therefore, we analyzed the extent to which test acceptance was associated with sociodemographic variables (age, sex, and education). Psychopathological and clinical variables such as symptom severity and multimorbidity of mental disorders or depression may also be linked to test acceptance. We hypothesized that, generally, individuals being more burdened may exhibit a lower test acceptance because a relatively demanding test situation must be more strenuous for them.

Method

Sample

The sample consisted of 236 consecutive patients at the Psychotherapeutic Outpatient Clinic of the Technical University in Dresden, Germany. The 157 (66.5%) women and 79 (33.5%) men were a mean of 37.7 years old ($SD=13.7$); 28 (11.8%) had a lower level, 104 (44.1%) an average level, and 104 (44.1%) a higher level of education (e.g., the „Abitur“ [general qualification for university entrance] or academic degrees).

All patients were diagnosed using the CIDI (see later discussion). The majority of patients had at least one anxiety disorder (67.8%), an affective disorder (49.6%), or both (36.9%). Somatoform disorders (26.7%), eating disorders (4.7%), and substance disorders (including nicotine dependence; 16.9%) were less frequent. The most frequent specific diagnoses were major depression (40.7%), specific phobia (34.3%), panic disorder and agoraphobia (31.8%), dysthymia (24.2%), and social phobia (20.3%). Because of high comorbidity rates (65.6%) and heterogeneous combinations of comorbid disorders in our sample, we formed the following categories to isolate the effect of the diagnosis of depression on test acceptance: (a) those having an anxiety disorder but no depression ($N=64$; age: $M=35.10$ years, $SD=12.16$; 71% female); (b) those having any affective disorder but no anxiety disorder ($N=25$, $M=37.80$, $SD=13.17$; 64% female), (c) those having an anxiety disorder and an affective disorder ($N=61$; age: $M=35.36$ years, $SD=13.23$; 75% female), and (d) those having neither anxiety nor affective disorders ($N=85$; age: $M=41.42$ years, $SD=14.68$; 58% female).

Procedure

Subsequent to their online diagnostic assessment with the CIDI, participants were requested to complete an evaluative questionnaire anonymously.

Measures

CIDI. Psychopathological and diagnostic assessments were based on the computer-assisted version of the Munich Composite International Diagnostic Interview (DIA-X/M-CIDI; Brown et al., 1994), a modified version of the WHO-CIDI (Wittchen & Pfister, 1997) supplemented by questions to cover a fuller range of DSM-IV diagnoses and subtypes. The DIA-X/M-CIDI is a fully structured interview that allows for the assessment of symptoms, syndromes, and 4-week, 12-month, and (partly) lifetime diagnoses of DSM-IV mental disorders along with information about onset, duration, and severity of threshold and subthreshold definitions. The following groups of DSM-IV disorders are covered: mental disorders resulting from a general medical condition, schizophrenia and possible psychotic disorders (screening without further differential diagnosis), substance-related disorders (nicotine, alcohol, and drug), depressive

disorders and bipolar disorders, anxiety disorders, posttraumatic stress disorder, obsessive-compulsive disorder, somatoform disorders, and eating disorders. Subthreshold diagnoses were computed as well. Diagnoses were assessed mainly within a 12-month time frame along with information about age at onset, course, duration, and persistence. For some disorders (affective, psychotic, and eating disorders and some anxiety disorders), the lifetime history was assessed as well.

The DIA-X/M-CIDI was supplemented by a respondent's booklet with cognitive aids to assist respondents in dating symptom onset and recency, answering complicated symptom questions, and identifying course patterns. The average time to complete the computerized DIA-X/M-CIDI, including additional questionnaires, was 63 min (Wittchen & Pfister, 1997). The test-retest reliability of the DIA-X/M-CIDI DSM-IV diagnoses (mean interval = 38 days) was found to be acceptable to good ($k = 0.56-0.81$ for the main diagnostic categories). The procedural validity of the M-CIDI diagnoses compared with attending physicians' independent clinical consensus diagnoses in a sample of 68 randomly chosen patients was acceptable to very good ($k = 0.50-0.96$, excluding psychotic disorders [0.21]). Further details of the CIDI's psychometric properties have been presented elsewhere (see Reed et al., 1998; Wittchen, Lachner, Wunderlich, & Pfister, 1998; Wittchen & Pfister, 1997).

Brief Symptom Inventory. The Brief Symptom Inventory (BSI; Derogatis, 1993; German version: Franke, 2000), a well-validated self-report short form of the Symptom Checklist-90-R (Derogatis, 1977), was also used. The BSI has 53 items, each rated on a 5-point scale ranging from 0 (not at all) to 4 (extremely). The inventory measures nine symptom dimensions: somatization, obsessive-compulsivity, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Three global indexes of distress can also be obtained, of which we used only the Global Severity Index (GSI), an indicator of the current overall level of distress.

Feedback scale for the CIDI. The feedback scale was developed specifically for this research to explore the test acceptance of the CIDI. We formulated 15 items, which covered the interview's evaluative aspects, mainly with respect to its cognitive load and the emotional response to being interviewed in a fully standardized format (see Table I for the complete list of items). All items were answered on a 4-point Likert scale. Seven items were positively formulated (Items 2, 5, 6, 8-10, 15) and eight negatively formulated (Items 1, 3, 4, 7, 11-14). The items were chosen to cover a broad range of relevant evaluative aspects and content domains rather than to reflect underlying psychological dimensions. Item-total correlations ranged from .28 to .62. Because the internal consistency (Cronbach's alpha) of the feedback scale was good ($\alpha = .85$), we calculated a mean score of the feedback form to reduce information and to allow for an easier overview of correlation analyses.

Interviewers. All CIDI interviews were conducted by extensively trained graduate psychology students and postgraduate psychotherapy students.

Statistical analyses. Because of the scales' non-normality, we analyzed Spearman's rho correlations in addition to descriptive statistics.

Results

The results of the interview feedback showed that most patients agreed with positive and disagreed with negative feedback statements (see Table I). This was true for different content

aspects such as comprehensibility of the interview and its items (Items 1, 3, 6, 12), procedure (Items 4, 11, 13), concentration (Items 5 and 9), general well-being during the interview (Items 2, 7, 8, 10), and global feedback (Items 14 and 15). An interesting finding was that, although a number of respondents (27.2%) obviously had expected a more therapeutic rather than a strictly diagnostic procedure (i.e., they „would have preferred a normal face-to-face interview from patient to therapist“), they could flexibly convert to what was offered and indicated they were satisfied with the evidently high quality of the diagnostic procedure (Item 15; 87.2%).

To analyze how feedback was associated with age, sex (1=female, 2=male), and level of education (1=low, 2=medium, 3=high), Pearson and Spearman's rho correlations were computed. None of these correlations were significant (feedback and age [N=236]: $r=.06$, ns; feedback and sex [N=233]: $r=-.03$, ns; feedback and education [N=227]: $r=.03$, ns).

As shown in Table II, test acceptance in diagnostic subgroups differed significantly in the overall test, $F(3, 231)=5.88$, $p<.01$, a finding attributable to one significant between-groups contrast: Test acceptance in the group with anxiety disorders and affective disorders was lower than in the group with none of these disorders.

The correlation between the BSI GSI and the total feedback score was $-.33$ ($p<.001$). Correlations of the BSI's subscales with the feedback score were in the same range ($-.16$ for anxiety and $-.34$ for interpersonal sensitivity; all $ps<.05$).

The number of comorbid diagnoses and the total score of the feedback form correlated significantly negatively ($r=-.24$, $p<.01$).

Discussion

This study, conducted in an unselected consecutive sample of 236 patients undergoing the routine diagnostic session in a psychotherapeutical outpatient clinic, revealed that standardized diagnostic interviews are highly accepted. Detailed results showed that, despite the substantial and evident effort of going through all of the DSM-IV Axis I diagnostic categories (only 55% of the patients „enjoyed“ the questions), patients understood that the questions were related to their problems, and they were glad to have been interviewed in such a straightforward and comprehensive manner.

These overall positive results were found to be unrelated to age, sex, and education. This fact indicates that, at least in this sample, the simple yes-no format of most questions in the interview is suitable for older and less well-educated people, and it thus confirms the broad applicability of the CIDI.

Further analyses showed that psychopathological symptom severity and psychiatric multimorbidity were associated with a slightly less favorable, although on average still positive (cf. Table II, Group 3), feedback rating of the interview. These effects ranged from small to moderate in size.

The specific factors explaining this situation need to be explored. Although it is not surprising that the more impaired patients face more difficulties undergoing the interview, it is not clear whether the slightly less favorable appraisal is a direct effect of patients' impairment or

simply an effect of the longer interview duration because of more disorders. The fact that having a depression alone did not influence test acceptance favors the latter interpretation: If there is depression and no other disorder, the interview will be, except for the stem questions, focused on depression. If there are numerous comorbid disorders (including depression or not), there will not be such a clear focus in the interview, which will also require more time. Probably this effect will have a negative impact on test acceptance. This view is also in accordance with the fact that, when comparing differing diagnostic groups, only the group with a comorbid combination of at least two diagnoses exhibited slightly lower test acceptance (for one of the comparisons, see Table II).

Limitations of this research include the fact that our research strategy aimed at an economic investigation only of the specific comments after the application of a computerized interview, and thus no comparison group could be meaningfully examined. Further research should aim at the pros and cons of comparative interview strategies from a more global patient perspective. Also the specificity of our sample, which comprised only psychotherapy outpatients, mostly with anxiety and depression, must be taken into account. The psychotherapy outpatients whom we see in our university setting may, to a certain degree, expect advanced diagnostic procedures and hence may accept a standardized interview better than other patients in routine care. The results may, therefore, not be applicable to more severely disturbed patients (e.g., those with substance use disorders, psychotic symptoms) or inpatients. Because personality disorders were not diagnosed in this sample, a further unanswered question is whether the diagnosis of personality disorders may have an impact on the test acceptance. A final critical point is that the tendency to answer in a socially desirable manner could not be controlled for appropriately, so that a positive bias in the data cannot be definitely ruled out.

Keeping these limitations in mind, we believe that the generally favorable results have important implications concerning the use of standardized interviews such as the CIDI: Psychotherapy outpatients who are interviewed with the CIDI appraise this experience as consistent with their expectation of professional practice and, accordingly, give a generally positive feedback. Thus, a broader and routine use of the CIDI in practice as well as in research about practice seems possible.

The lack of psychotherapy studies in naturalistic settings is now often recognized (Borkovec, Echemendia, Raguesa, & Ruiz, 2001; Hartmann & Zepf, 2003; Leichsenring, 2004), but research efforts are often expensive and not economical. Clinical diagnoses that are solely based on therapists' subjective impressions are clearly insufficient. The CIDI can be recommended as an alternative tool that must not necessarily be used by psychotherapists or researchers but can also be used by trained personnel. Thus, it can promote collection of data also in naturalistic settings without compromising the diagnostic and methodological standards and the patients' acceptance of the study.

Table I. Items of the Evaluation Sheet for the Interview and Frequencies (%) of Answers (N = 236)

Item	Answer			
	1	2	3	4
1. Sometimes I could not follow the instructions	49.6	32.6	17.0	0.9
2. I enjoyed answering the questions	6.5	37.9	46.1	9.5
3. Sometimes I did not understand what the interviewer did	56.6	33.6	8.8	0.9
4. While I was answering the questions, I found the computer annoying	68.8	26.3	4.9	0
5. Throughout the interview, my concentration was good	3.4	21.8	50.9	23.9
6. The questions were clear to me	1.3	11.9	64.3	22.6
7. Now and then I wanted to quit the interview	54.6	31.4	13.1	0.9
8. The interview was a pleasant experience for me	3.4	20.5	61.5	14.5
9. During the interview, my endurance was steady	6.0	18.3	52.3	23.4
10. I'm satisfied with how I answered the questions	2.5	12.7	70.8	14.0
11. I found it annoying that the interviewer again and again did something with his computer	62.6	32.7	4.4	0.4
12. I did not understand how the questions were related to my problems	51.8	38.5	9.7	0
13. Anything related to computers makes me feel uncomfortable or anxious	73.5	23.9	1.8	0.9
14. I would have preferred a normal face-to-face interview from patient to doctor/therapist	32.0	40.8	23.7	3.5
15. I think it was good that the survey was done in such an exact and detailed manner	3.0	9.8	58.3	28.9

Note. 1 = completely disagree, 2 = disagree, 3 = agree, 4 = completely agree.

Table II. Total Score (Mean) of the Feedback Questionnaire in Different Diagnostic Subgroups (N = 236)

Diagnostic group	N	M	SD	Contrasts ^a	p
1. Anxiety, no depression	62	3.28	0.46		ns
2. Depression, no anxiety	29	3.13	0.44		ns
3. Anxiety and depression	109	3.09	0.41	3 < 4	< .01
4. Other disorders (no anxiety, no depression)	36	3.32	0.48	4 > 3	< .01

^aScheffé test.

References

- American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: American Psychiatric Association.
- Basco, M. R., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry*, 157, 1599-1605.
- Borkovec, T. D., Echemendia, R. J., Raguesa, S. A., & Ruiz, M. (2001). The Pennsylvania Research Network and future possibilities for clinically meaningful and scientifically rigorous psychotherapy effectiveness research. *Clinical Psychology: Science & Practice*, 8, 155-167.
- Brown, T. A., DiNardo, P. A., & Barlow, D. A. (1994). *Anxiety Disorders Interview Schedule for DSM-IV*. San Antonio, TX: Psychological Corporation.
- Derogatis, C. R. (1977). *SCL-90-R: Administration, scoring & procedures*. Manual I for the R(evisited) Version and other instruments of the Psychopathology Rating Scale Series. Baltimore, MD: Johns Hopkins University School of Medicine.
- Derogatis, L. R. (1993). *Brief Symptom Inventory (BSI): Administration, scoring and procedures manual (Vol. 3)*. Minneapolis, MN: National Computer Services.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I disorders, research version, patient edition*. New York: New York State Psychiatric Institute.
- Franke, G. (2000). *Brief Symptom Inventory*. Göttingen: Hogrefe. Hartmann, S., & Zepf, S. (2003). Effectiveness of psychotherapy in Germany: A replication of the Consumer Report study. *Psychotherapy Research*, 13, 235-242.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and metaanalysis. *Personnel Psychology*, 57, 639-683.
- Kublinger, K. D. (2006). *Psychologische Diagnostik*. Göttingen: Hogrefe.

- Leichsenring, F. (2004). Randomized controlled versus naturalistic studies: A new research agenda. *Bulletin of the Menninger Clinic* , 68, 137-151.
- Meehl, P. E. (1954). *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence* . Minneapolis: University of Minnesota Press.
- Reed, V., Gander, F., Pfister, H., Steiger, A., Sonntag, H., Trenkwalder, C., et al. (1998). To what degree the Composite International Diagnostic Interview (CIDI) correctly identifies DSM-IV disorders? Testing validity issues in a clinical sample. *International Journal of Methods in Psychiatric Research* , 7, 142-155.
- Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule (DIS): Its history, characteristics and validity. *Archives of General Psychiatry*, 38, 381-389.
- Steiner, J. L., Tebes, J. K., Sledge, W. H., & Walker, M. L. (1995). A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. *Journal of Nervous and Mental Disease* , 183, 365-369.
- Wing, J. K., Babor, T., Brugha, T., Burke, J., Cooper, J. E., Giel, R., et al. (1990). SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, 47, 589-593.
- Wittchen, H.-U., Lachner, G., Wunderlich, U., & Pfister, H. (1998). Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Social Psychiatry and Psychiatric Epidemiology*, 33, 568-578.
- Wittchen, H.-U., & Pfister, H. (1997). *Diagnostisches Expertensystem für psychische Störungen (DIA- X)*. Frankfurt: Swets & Zeitlinger.
- World Health Organization. (1990). *Composite International Diagnostic Interview (CIDI), Version 1.0*. Geneva: Author. World Health Organization. (1993). *Tenth revision of the International Classification of Diseases: Chapter V(F). Mental and behavioral disorders* . Geneva: Author.