

This is an author-produced electronic version of an article accepted for publication in the *British Journal of Psychiatry*. The definitive publisher-authenticated version is available online at <http://bjp.rcpsych.org>.

## Cross-cultural Feasibility, Reliability and Sources of Variance of the Composite International Diagnostic Interview (CIDI)

H.-U. WITTCHEN, L. N. ROBINS, L. B. COTTLER, N. SARTORIUS, J. D. BURKE, D. REGIER  
and PARTICIPANTS IN THE MULTICENTRE WHO/ADAMHA FIELD TRIALS

The CIDI is a fully standardised diagnostic interview designed for assessing mental disorders based on the definitions and criteria of ICD-10 and DSM-III-R. Field trials with the CIDI have been conducted in 18 centres around the world, to test the feasibility and reliability of the CIDI in different cultures and settings, as well as to test the inter-rater agreement for the different types of questions used. Of 590 subjects interviewed across all sites and rated by an interviewer and observer, 575 were eligible for analysis. The CIDI was judged to be acceptable for most subjects and was appropriate for use in different kinds of settings. Many subjects fulfilled criteria for more than one diagnosis (lifetime and six-month). The most frequent lifetime disorders were generalised anxiety, major depression, tobacco use disorders, and agoraphobia. Percentage agreements for all diagnoses were above 90% and the kappa values were all highly significant. No significant numbers of diagnostic discordances were found with lifetime, six-month, and four-week time frames.

The Composite International Diagnostic Interview (CIDI) has been produced in the framework of a major project (the Joint Project on Diagnosis and Classification of Mental Disorders, and Alcohol- and Drug-Related Problems) undertaken by the World Health Organization (WHO) and the US Alcohol, Drug Abuse and Mental Health Administration (ADAMHA). It is a comprehensive, fully standardised diagnostic interview for the assessment of mental disorders according to the definitions and criteria of ICD-10 (World Health Organization, 1989a), and DSM-III-R (American Psychiatric Association, 1987). The CIDI has been designed for use in a variety of cultures and settings. Although it is primarily intended for use in epidemiological studies of mental disorders, the CIDI can also be used for clinical and research purposes, and can be supplemented by modules for diagnoses not covered in the core version. In the course of development (Robins et al, 1988), the instrument was subjected to a number of tests for feasibility, diagnostic coverage, test-retest reliability as well as procedural reliability (Semler et al, 1987, 1988; Farmer et al, 1987; Semler, 1990; Wittchen et al, 1990).

As part of the further development and test of the CIDI, this paper summarises the findings of the first major international WHO field trial of the 'pre-final' version of the CIDI. The study was initiated to explore the CIDI's cross-cultural acceptability and feasibility, and to determine its reliability in different settings and countries. Because several test-retest reliability studies of the CIDI approach have already been conducted (Burke, 1986; Semler et al, 1987, 1988; Wittchen et al, 1989) a complex inter-rater reliability design was chosen for the field trials, in order to allow for the identification of cross-cultural problems in the CIDI as well as to study reasons for disagreement between raters. The major objectives of the CIDI field trials were: (a) to report data on the feasibility of the CIDI in different cultures and settings, (b) to assess reliability between raters using the CIDI, (c) to report the CIDI's inter-rater agreement for the different types of questions used in the CIDI, and (d) to evaluate possible reasons for discrepancies in rating between an interviewer and an observer.

## Method

For the test of inter-rater reliability, each subject was examined by an interviewer/observer pair. The interviewer conducted the interview, while the observer independently - and without having access to the interviewer ratings - scored a separate booklet while sitting in the same room. While sitting by and making the CIDI codings, the observer additionally noted all the interviewer's mistakes and inconsistencies detected, by circling the respective items. Once the interviewer had finished the CIDI, the observer was then free to ask all these circled questions or even sections of the CIDI again, to verify that his/her original codings were correct. If necessary the observer then recoded the CIDI according to the second administration of the respective CIDI question or section. During this procedure the interviewer was not allowed to change any of his/her previous codings.

The data from the resulting two CIDI booklets – the original CIDI codings of the interviewer and the final CIDI codings of the observer – were used to calculate the inter rater agreement. After the respondent had left the room, the interviewer and the observer reviewed all questions circled by the observer. Using the Discrepancy Resolution Sheet (DRS) (see below), all discrepancies between the codings of the observer and those of the interviewer were discussed to identify the reason for the discrepancy. Multiple ratings could be made for each of the discrepant CIDI items. The following classes of reasons for discrepancies were used: (a) mistakes in the presentation of a CIDI question, (b) mistakes in the additional 'probe' questions for severity and exclusion criteria, (c) coding errors, (d) misunderstandings on either the interviewer's or the patient's side, (e) ambiguous questions and (f) changes in the respondent's answers to the same question. If possible, both the interviewer and the observer made final consensus ratings. These codings were then written down on the DRS. Only discrepancies in the codings of the CIDI symptom questions, thus no discrepancies for the so-called 'recency' and 'severity' questions, were discussed.

At the conclusion of each session the interviewer and the observer completed a simple interview report in order to document problems with the interview, including suggestions for resolving them. This report emphasised the identification of points of uncertainty, based on cultural factors, translation problems, or other problems in the interview schedule. Each interview was audiotaped, or if feasible videotaped, to enable the sources of discrepancies to be rated subsequently. In each centre, the editor or field supervisor reviewed three tapes per interviewer, at random, to check on the administration of the interview and to identify possible uncertainties with the study procedure.

### *Participating centres and sampling of subjects*

Altogether 18 centres participated in the CIDI field trials. The trials were carried out in a variety of settings ranging from specialised psychiatric in-patient and out-patient units to general practice settings. At each a minimum of 25 (mostly 32) subjects were interviewed. The sampling process for patients differed slightly between centres depending on the size of the institution and the number of available interviewers (see below). All sites selected patients typically found in their settings. Exclusion criteria were: (a) evident severe cognitive impairment in language or communication, (b) no signed consent to the study procedure, which included video- or audiotaping, (c) age under 18 or over 65 years. Table 3 indicates the sites, the number of subjects selected, and classifies the sites into three groups – psychiatric in-patient setting, psychiatric out-patient setting, and other – characterises the type of institution. Some sites sampled a proportion of their subjects in an in-patient setting and

another proportion in an out-patient or 'other' setting. Although at each site samples if possible were to be selected randomly, there were some inconsistencies in the selection process (see footnote to Table 3).

### *CWI interviewers and training*

At each site, every interviewer completed at least four interviews as an interviewer and another four as observer, in alternate order. A random assignment was feasible at only six sites. The number of interviewers per site varied between three and eight.

In addition an attempt was made to study potential differences between clinician and non-clinician interviewers (Table 3). 'Non-clinicians' were defined as interviewers who were not able to make a diagnosis in the absence of a standardised interview (medical or psychology students, nurses and social workers working in the institution). 'Clinicians' were defined as mental health specialists, able to make diagnoses independently (psychiatrists or clinical psychologists). Altogether 108 interviewers (53 clinical, 55 non-clinical) participated in the study. Pairing of clinician and non-clinician interviewers and observers was randomised at five sites only, in order to allow for an analysis of differences. Eighty-seven patients were examined by clinician pairs, 49 by non-clinician pairs, and 454 by clinician-non clinician pairs.

The CIDI requires for proper administration a minimum of five days' continuous training. The key investigators of all participating sites were trained centrally in a one-week training session using the CIDI manual, the standardised training package, as well as the item-by-item specifications available for the CIDI. These key trainers were then provided with the training materials to prepare their on-site training seminars according to the guidelines in this training package. Before commencing the study, each interviewer completed at least two CIDI interviews under the close supervision of the key investigator.

### *Instruments*

Three instruments were used in this trial: the CIDI, the Evaluation Sheet (ES) and the Discrepancy Resolution Sheet (DRS).

The CIDI consists of a package that includes: (a) the interview booklet, (b) the user manual, (c) the training package, (d) a data-entry program and (e) the diagnostic computer program. In wave I of the field trials a 'pre-final' CIDI version was used (CIDI-Core; World Health Organization, 1987). The version allowed the derivation of all DSM-III diagnoses (American Psychiatric Association, 1980) listed in Table 1 as well as the derivation of Present State Examination (PSE) CATEGO classes (Wing et al, 1974), compatible with selected ICD-9 diagnostic classes. Only for substance use disorders were the revised criteria according to DSM-III-R and ICD-10 taken into account.

The interview consists of 288 symptom questions, not all of which are asked of all respondents because of skip rules. Many of the symptom questions have to be asked in conjunction with so-called 'probe' questions to assess severity and the psychiatric relevance of each respondent's answer. Besides symptom questions (a), the CIDI further incorporates the following types of questions (Table 2): (b) questions for the assessment of the last and first occurrence of a symptom ('recency'), as well as (c) selected items for duration and frequency of selected syndromes.

The version used in the field trials differs from the final CIDI version (World Health Organization, 1987b) in the following ways: (a) it does not contain 22 necessary new questions to meet ICD-10 criteria as well as DSM-III-R criteria, (b) it contains several items adapted from the PSE, (c) unlike to the final CIDI it incorporates additional questions to assess the recency of individual symptoms, and (d) an earlier draft of the final alcohol and drug section was used (the results for this substance section are reported by Cottler et al (following paper, this issue)).

The ES was used to allow judgements about the following areas: (a) the acceptance of the CIDI questions by the patient, (b) the overall appropriateness of the CIDI in clinical and 'epidemiological' settings (depending on the site), (c) the length of the interview, and (d) all items and sections for which problems were reported by the sites.

The DRS was used to record all observed discrepancies between the interviewer and the observer, rating the most likely reasons for the discrepancy.

### *Translation of the CIDI*

Translations of the English version of the CIDI and the other training materials were required in Greek, Portuguese, Kannada (India), Spanish, Chinese, Norwegian, Dutch, Italian, and French; the CIDI had been pilot tested in German. Because of possible cultural and language-specific problems, all participating centres sent their final translation and back-translation, together with a listing of translation problems, to the field trial co-ordinator before the study. This allowed for modifications to be made in either the instrument or the manual.

Besides a number of smaller adjustments in some CIDI items, the following problems were acknowledged. (a) Parts of the section for organic brain syndrome were difficult to adapt to some rural settings and in less developed countries (where there was lesser ability to write or draw, and fewer equivalents for sentences and words used) which required modifications based on advice from local neurological and neuropsychological research centres. (b) The questions for sexual problems as well as symptoms related to pregnancy in unmarried subjects were difficult to use in some countries like India. (c) There were difficulties in translating some symptom meanings into Chinese and Kannada (e.g. headaches, weakness, indicators for trouble in concentrating) (in these cases examples were added in parenthesis to the words in the question to clarify the underlying concepts). (d) With regard to the specific probe questions, a set of rules was developed for developing countries on how to code certain explanations of symptoms by 'barefoot' doctors or healers. (e) Because of difficulties determining weight and height criteria in some rural sites, interviewer judgement was used to assess these criteria.

The translated versions of the CIDI-Core were produced in the same format as the English original version, page by page, to avoid mistakes and difficulties in data analysis.

### *Study procedure and data analysis*

The translations and back-translations of the CIDI were finalised by December 1987. In early 1988 the on-site training courses and the preparation of the study requirements took place. Data collection started in February and ended in May 1988. The CIDI interviews were locally and centrally edited and checked for mistakes and inconsistencies and were then entered on the computer for the final analysis in Munich. Data analysis was completed in July 1988 and results were circulated to all centres for comment. Final analyses were made in the autumn

and winter of 1988. For the reliability analyses of the CIDI/DSM-III diagnoses, a modified version of the computer program to analyse information collected with the Diagnostic Interview Schedule (DIS) Version III was used (Semler et al, 1987). This program allows the derivation of the diagnostic categories with and without DSM-III exclusion rules. These were not applied in the current analysis. The reliability and acceptability of the substance use disorder sections are reported in the following paper.

## **Results**

Of 590 subjects interviewed across all sites and rated by the interviewer and the observer, 575 were eligible for analysis; 15 were excluded because of violations of the design or incomplete codings not allowing the application of the diagnostic computer program. The mean age of subjects was 39.3 years (s.d. 14.3), 41.4% were male, 58.6% female, 34.3% were never married, 44.9% were married, 9.9% were divorced, 4.7% were separated and 5.7% were widowed. Types of setting and numbers of patients at each site are shown in Table 3.

### *Feasibility and acceptance of the CIDI across centres*

The majority of raters in the 18 centres judged the CIDI as being acceptable (49.3% as good, 41.5% as moderate, 9.2% as poor), with no country-by-country variation. Concerns were expressed by only seven interviewers in centres that included in-patients, specifically with regard to the rather lengthy section for alcohol and drugs, the format of the depression section (being too long and complicated for severely disturbed subjects) and the repetitiveness of some probe and recency questions.

The CIDI was also judged as being appropriate for most of the settings, especially out-patient and primary-care settings (25.9% as very appropriate, 59.3% as appropriate). It should be acknowledged, however, that in clinical settings with more seriously ill subjects, 30.7% of the interviewers rated parts of the CIDI as inappropriate. Less favourable ratings were especially given for the sections for schizophrenia and depression. At two of the sites interviewing almost exclusively acute psychiatric in-patients – the majority of whom had symptoms of psychosis (Amsterdam and Milan) – doubts were expressed regarding the validity of some of the codings of delusions and hallucinations as well as the assessment of feelings of worthlessness and sleep disturbance, and here the CIDI was criticised for relying exclusively on the subject's self-report.

One general criticism from many sites concerned the length of the CIDI. Only a few (11.2%) of the interviews lasted less than one hour, while 35.4% lasted one to two hours and 36.4% lasted two to three hours. In 52 subjects the administration of the CIDI had to be split in two sections, in six cases even in three; all of these were conducted in in-patient settings. Independent of type of setting or site, the majority rated the CIDI as being too long (65%) or even much too long (15%), particularly the sections for depression, alcohol and drugs. The centre in India reported that the illiterate respondents took substantially longer than literate subjects.

### *Number of lifetime and six-month diagnoses across centres*

Table 4 indicates the number and percentage of subjects fulfilling diagnostic criteria according to the CIDI/DSM-III program without the optional DSM-III exclusion rules. Diagnoses for alcohol and drug abuse and dependence were not included in this analysis (see following paper). A total of 62 subjects did not meet DSM-III criteria for any diagnosis. In

both the lifetime and six-month time frames, many subjects fulfilled criteria for more than one diagnosis. The most frequent lifetime disorders were generalised anxiety (50.3%), major depression (31.0%), tobacco use disorders (30.3%), and agoraphobia (30.1%). Lowest frequencies were found for eating disorders, pathological gambling and somatisation. There was some variability between centres depending on the kind of setting and the severity of the patients' conditions. In-patient institutions had the highest mean number of lifetime diagnoses per patient (3.2 per patient), followed by out-patient settings (2.6 diagnoses). The lowest number of diagnoses were found – as expected – in the general practice settings (1.8) and in the health maintenance organisation in St Louis, USA.

### *Diagnostic agreement*

The kappa values across centres were excellent in all diagnostic sections (Table 5). Percentage agreements for all diagnoses were above 90%, kappa values were highly significant, and with three exceptions above 0.90. Although the CIDI requires complex ratings for many of the symptom items - ranging from simple 'yes/no' responses, to complex probe questions with five different coding options, to judgements of onset and recency of selected symptoms and all syndromes - no significant diagnostic discordance was found. The lowest kappa values were found for somatisation, schizophreniform disorders and anorexia, possibly due to low base rates.

Given the base-rate problem with kappa and the relatively small number of subjects per site, it is not possible to report separate concordance statistics for each site. Specific intersite differences however were examined by analysing the average number of discrepant ratings between interviewer and observer. On average 1.68% (s.d. 1.64%) of all symptom ratings made were discrepant. The number of discrepancies correlated ( $r=0.38$ ,  $P<0.01$ ) with the overall number of ratings, thus indicating that in-patient settings with multisymptomatic subjects revealed slightly higher numbers of discrepancies (2.04) than out-patient (1.52) settings.

### *Agreement on the item level*

The concordance of raters for the symptom questions in all diagnostic sections was good to excellent. Figure 1 displays the box-plots of kappa values for symptom questions for each of the diagnostic sections. There are only a few items indicating some disagreement among raters which deserve closer attention: for example, in the depression section the item "Did you tell a doctor about this spell?"; in the schizophrenia section a few questions assessing the psychosocial impact of hallucinations and delusions; and for obsessive-compulsive disorders the question about had a doctor been told about the syndrome. The section for organic brain dysfunction has a wide score distribution because a number of items (drawing, naming objects, handling of objects) require close observation of the patient's behaviour, which was not possible within our research design, thus causing some disagreements, and because of problems arising in translation, as discussed under 'Method'.

Because the CIDI is specifically designed to assess the occurrence of disorders over the whole life-span, the analysis of all time-related questions was of specific interest. For the 'recency' questions, which are asked in some sections for individual symptoms as well as for syndromes, almost perfect agreement was found when dichotomising the rating options into 'present in the last four weeks' and 'not present in the last four weeks'. The only section with a noticeable variation was the one for obsessive-compulsive disorder. This discrepancy may be due to the rather lengthy symptom question for obsessive thoughts that was frequently readministered by the observer who felt that this question was not asked appropriately by the

interviewer. In these cases the recency code for the last occurrence of the symptom was sometimes different in the interviewers' and observers' codings. Concordance for items for which the subjects were asked for the age of first occurrence of a symptom (lifetime) (all sections) as well as the frequency of episodes and their length (depression, mania, schizophrenia sections) was determined by using the intraclass correlation coefficient (ICC). The ICC values revealed excellent concordance, with scores ranging between 0.70 and 0.98.

### *Discrepancies*

The analysis of the DRS revealed an average of almost 4.3 discrepancies (1.68% of all CIDI questions) per subject. The highest proportions of discrepancies (number of discrepant judgements/number of overall judgements per section) were found in the sections for somatisation (33%), phobias (24%), obsessive-compulsive disorder (24%), mood disorders (11%), and psychotic disorders (10%). The most frequent types of disagreement were mistakes by an interviewer in either the presentation of the question (23%) or the administration of the complicated probe flowchart questions (39%), which need to be followed rigidly. Another relatively frequent reason for discrepancies was that the respondent had simply changed his/her mind (24%) when the question was readministered by the observer. Implicit problems with the formulation of the question (8%) or coding errors (1%) account for only a small proportion of the discrepancies.

Because the vast majority ( $n = 454$ ) of examinations were conducted by pairing a clinical with a non-clinical interviewer, it is difficult to study systematically the effect of type of interviewer. In comparing, however, the mean number of discrepancies found for clinician pairs ( $n = 77$ ; mean 1.13% discrepancies) with the findings for non-clinician pairs ( $n = 49$ ; mean 1.65% discrepancies) or the clinician-non clinician pairs (mean 1.62% there is an indication that the clinician pairs made fewer discrepant decisions. This interpretation must be cautious, however, because data from the clinician-clinician comparisons come from three non-psychiatric out-patient sites with relatively asymptomatic subjects only, and thus could not be regarded as being representative. No differences were found for the comparison of 'type of interviewers' versus 'type of discrepancies'.

### **Discussion**

Our findings emphasise the generally good acceptance and appropriateness of the CIDI for the assessment of symptoms and syndromes of mental disorders and its classification of them along the criteria of established diagnostic systems in different settings, countries, and cultures. Only few CIDI questions were identified as candidates for modification because of national or cultural reasons. The only general criticism referred to the length of the interview. Factors mainly responsible for this proved to be the many individual 'recency' questions for specific symptoms, the sections for somatoform disorders (somatisation, hypochondrias and persistent pain disorder), depressive disorders and, especially, the two new sections for alcohol and drug abuse and dependence. The solution to this problem lies in the deletion of most symptom recency questions, together with the implementation of summary recency questions.

The finding of the good inter-rater reliability of the CIDI in the present study was not unexpected because of (a) results of earlier studies in the context of the development of the DIS (compare overview by Burke, 1986), (b) previous reliability and validity studies of the CIDI (Farmer et al, 1987; Semler et al, 1988; Wittchen et al, 1989; Semler, 1990), and (c) subsequent attempts to improve this instrument further. Nevertheless, the fact that



concordance for almost all diagnoses was above a kappa value of 0.9 needs explanation. Regarding test-retest reliability, in a study of the CIDI with two different interviewers examining the patient within two to three days, Semler et al (1987) found kappa values that are with one exception almost 0.2 kappa values lower than those we report (see Fig. 2). There are several explanations for our higher agreement coefficients, all related to our specific design characteristics. (a) It might be easier to get better concordance when both raters are in the same room. With an observer present, the interviewer might have felt obliged to keep more closely than otherwise to the rigid rules for administering the CIDI. (b) The two ratings in the inter-rater design were obtained within the same session, thus reducing the possible sources of variance occurring because of a change of the patient's psychopathological state. Change in the patient's condition might be regarded as an important source of variance in a test-retest design with a two- to three-day interval. (c) This study design is of course more vulnerable to design violations than a more strict test-retest design with two independent investigations. In some cases observers might not have strictly coded the patient response, but rather what he/she suspected would be the final coding of the interviewer, thus violating the design intent. Although this assumption is difficult to test, it is relatively unlikely that this has had a major effect on our data. (d) A more likely explanation might be seen in differences of the sample selection. Semler's study was conducted in severely and acutely disturbed in-patients, including a relatively high proportion of psychotic patients, whereas the present study included many out-patient settings, with some monosymptomatic cases.

The analysis of the reasons for discrepancies demonstrated that the major source of discrepancies lies in subtle modifications of the original CIDI question and deviations or modifications of the complex set of rules for the probe questions. Complying with the CIDI rules seems to be even more important if the instrument is going to be used by such a heterogeneous group of users as in our field trials. Although we used clinician and non-clinician interviewers, we could not demonstrate any important differences between these two groups. In this respect it seems important to emphasise that the CIDI requires, for clinician as well as non-clinician interviewers, a comprehensive one-week course of structured training before its administration.

Thus, we can conclude that the CIDI, as an almost completely standardised diagnostic instrument, reduces effectively one major source of disagreement, that is inter-rater variance, typically found to be a frequent source of disagreement in all less standardised psychiatric interviews. Whether the high inter-rater reliability of the CIDI also results in an improved content validity and procedural validity needs to be determined in further studies. Studies of this kind are currently under way with a slightly revised and shortened final version of the CIDI (World Health Organization, 1989b) in a second wave of the WHO/ADAMHA CIDI field trials.

## FEASIBILITY OF THE CIDI

Table 1  
The CIDI diagnostic coverage (DSM-III and PSE items)

CIDI diagnostic coverage of DSM-III disorders	CIDI diagnostic coverage of 63 PSE items <sup>1</sup>
<i>Eating disorders</i>	(1) Physical health
(1) Anorexia nervosa	(2) Physical illness
(2) Bulimia	(3) Psychosomatic symptoms
<i>Organic mental disorders</i>	(4) Worrying <sup>3</sup>
(3) Dementia (unspecified)	(5) Tension pains
<i>Substance use disorders<sup>2</sup></i>	(6) Tiredness <sup>3</sup>
(4) Alcohol abuse	(7) Muscular tension <sup>3</sup>
(5) Alcohol dependence	(8) Restlessness <sup>3</sup>
(6) Barbiturate, sedative abuse	(9) Hypochondriasis <sup>3</sup>
(7) Barbiturate, sedative dependence	(10) Nervous tension <sup>3</sup>
(8) Opiate abuse	(11) Free-floating anxiety <sup>3</sup>
(9) Opiate dependence	(12) Anxious foreboding <sup>3</sup>
(10) Cocaine abuse	(14) Panic attacks <sup>3</sup>
(11) Amphetamine abuse	(15) Situational anxiety <sup>3</sup>
(12) Amphetamine dependence	(16) Anxiety on meeting people <sup>3</sup>
(13) Hallucinogen abuse	(17) Specific phobias <sup>3</sup>
(14) Cannabis abuse	(18) Avoidance of anxiety <sup>3</sup>
(15) Cannabis dependence	(19) Inefficient thinking <sup>3</sup>
(16) Tobacco dependence	(20) Poor concentration <sup>3</sup>
<i>Schizophrenic disorders</i>	(21) Neglect due to brooding <sup>3</sup>
(17) Schizophrenia	(22) Depressed mood
<i>Psychotic disorders not elsewhere classified</i>	(23) Hopelessness
(18) Schizophreniform	(24) Suicidal
<i>Affective disorders</i>	(25) Morning depression <sup>3</sup>
<i>Major affective disorders</i>	(26) Social withdrawal <sup>3</sup>
(19) Manic-depressive episode	(27) Self-depreciation <sup>3</sup>
(20) Manic episode	(28) Lack of self-confidence <sup>3</sup>
(21) Bipolar disorder	(29) Simple ideas of reference <sup>3</sup>
(22) Major depression, single episode	(30) Guilty ideas of reference <sup>3</sup>
(23) Major depression, recurrent	(31) Pathological guilt <sup>3</sup>
<i>Other specific affective disorders</i>	(32) Loss of weight
(24) Dysthymic disorder	(33) Delayed sleep <sup>3</sup>
<i>Atypical affective disorders</i>	(34) Anergia and retardation <sup>3</sup>
(25) Atypical bipolar disorder	(35) Early waking <sup>3</sup>
<i>Anxiety disorders</i>	(36) Loss of libido <sup>3</sup>
<i>Phobic disorders</i>	(37) Irritability <sup>3</sup>
(26) Agoraphobia with panic attacks	(38) Expansive mood
(27) Agoraphobia without panic attacks	(39) Subjective ideomotor pressure
(28) Social phobia	(40) Grandiose ideas
(29) Simple phobia	(41) Obsessional checking <sup>3</sup>
<i>Anxiety states</i>	(42) Obsessional cleanliness
(30) Panic disorder	(43) Thought insertion
(31) Generalised anxiety disorder	(44) Thought broadcast
(32) Obsessive-compulsive disorder	(45) Thought block
<i>Somatisation disorders</i>	(46) Delusion of thoughts read
(33) Somatisation disorder	(47) Verbal hallucinations
(34) Psychosexual dysfunction	(48) Dissociative hallucinations
(35) Pathological gambling	(49) Olfactory hallucinations
	(50) Other hallucinations
	(51) Delusions of control
	(52) Delusions of reference
	(53) Delusions of misinterpretation <sup>3</sup>
	(54) Delusions of persecution
	(55) Delusions of grandiose ability <sup>3</sup>
	(56) Delusions of grandiose identity <sup>3</sup>
	(57) Delusional explanations <sup>3</sup>
	(58) Slowness and underactivity <sup>3</sup>
	(59) Behaves as if hallucinated
	(60) Blunted affect
	(61) Slowness of speech <sup>3</sup>
	(62) Neologisms
	(63) Incoherence of speech
	(64) Adequacy of interview

1. Item 13 is not covered. 2. According to criteria of ICD-10 and DSM-III-R.  
3. Questions were added to cover these; other items were adequately covered by questions from the Diagnostic Interview Schedule.

Table 2  
Types of question in the CIDI

Diagnostic system	Code for criteria	CIDI question	Codes
<i>Example of a standard question with probes (1 2 3 4 5)</i>			
SOM3RB8	C2	Have you ever had a lot of trouble with back pain?	PRB: 1 2 3 4 5
SMPN3RP1			
PPICD			
<i>Example of a simplified no/yes question</i>			
SMPN3RA	C12	Have there ever been 6 months or more in your life when most of the time you were bothered a great deal by (LIST PAINS CODED '5'Q.C1-C11)	NO ..... 1 YES ..... 5
PPICDA			
<i>Example of an age-of-onset and recency question</i>			
	C14	ONS/REC: When was the (first/last) time you were bothered a lot by these pain(s)?	ONS: 1 2 3 4 5 6 AGE ONS: _/_ REC: 1 2 3 4 5 6 AGE REC: _/_
<i>Example of a time/episode-related question</i>			
DEP3RSE	E41	Now I'd like to ask about spells when you felt both (depressed/OWN EQUIVALENT) and had some of these other problems like (LIST ANY ITEM CODED PRB5* FROM QS. E7-E31). In your lifetime how many spells like that have you had that lasted two weeks or more?	_/_ # SPELLS
DEPICDRT			
DEP3RRT			

The labels in the left-hand column are a unique feature of the CIDI-core, and reflect which diagnostic criteria that particular question is used to assess.

Table 3  
Sites and types of settings of CIDI field trials

Site	Type of setting/patients	No. of subjects	No. of interviewers	
			Clinical	Non-clinical
<i>In-patient</i>				
Cardiff, UK	Hospital/psychiatric ward	32	3	5
Jena, Germany	Neuropsychiatric university hospital and unit	25	3	2
Beijing, China	Medical university/psychiatric ward	32	2	2
Luxembourg, Luxembourg	Central hospital	32	4	0
Minneapolis, USA	University hospital chemical dependency treatment programme	21	1	2
Amsterdam, the Netherlands	University hospital/psychiatric ward	40	3	3
<i>Out-patient</i>				
Sydney, Australia	Hospital anxiety clinic	32	2	5
Bangalore, India	Rural and general out-patient clinic	32	4	4
Athens, Greece	General and mental hospital out-patient departments	37	3	2
San Juan, Puerto Rico	General and psychiatric out-patient department/university	40	2	2
Sao Paulo, Brazil	General and psychiatric out-patient clinic	29	4	4
Minneapolis, USA	University hospital affective disorders unit	16	2	3
Paris, France	Psychiatric out-patient consultations/hospital	25	2	1
Mainz, Germany	Psychiatric out-patient clinic/university	32	3	5
Milan, Italy	Psychiatric out-patient service/university hospital	32	2	2
Porto, Portugal	Community intervention unit	32	5	3
Oslo, Norway	Psychiatric outdoor clinic and general hospital	32	3	3
<i>Other</i>				
St Louis, USA	Health maintenance organisation	25	3	3
Munich, Germany	General practitioner attenders	29	2	4
Totals		575	53	55

In Paris five of the subjects examined were interviewed in the internal medicine ward of the hospital. These five cases were left in the analysis, because the reason for consultation as well as their characteristics did not differ from the other subjects selected in this site. In Athens - as part of an ongoing study in that unit - 12 relatives of psychiatric cases were examined. In Mainz - for the same reason as Athens - 16 subjects were included. In Milan four of the selected psychiatric out-patients were admitted as in-patients after they had been selected for inclusion in the study.

**Table 4**  
Lifetime and six-month CIDI/DSM-III diagnoses (*n* = 575)  
(excluding alcohol and drug-related disorders)

	Lifetime		Six-month	
	<i>n</i>	%	<i>n</i>	%
Organic brain syndrome				
definite	18	3.1	18	3.1
probable	34	5.9	34	5.9
Any depressive disorder	224	39.0	173	30.1
dysthymia	102	17.7	64	11.1
major depressive - single	43	7.5	29	5.0
major depressive - recurrent	135	23.5	111	19.3
Bipolar disorder I	22	3.8	16	2.8
Bipolar disorder II	25	4.3	20	3.5
Obsessive-compulsive	74	12.9	54	9.4
Panic disorder	85	14.8	61	10.6
Generalised anxiety	289	50.3	226	39.3
Any phobic disorder	260	45.2	204	35.5
agoraphobia	173	30.1	149	25.9
agoraphobia with panic attacks	109	19.0	100	17.4
simple phobia	165	28.7	131	22.8
social phobia	116	20.2	98	17.0
Somatisation	4	0.7	4	0.7
Schizophrenic disorder	44	7.7	32	5.6
Schizophreniform disorder	8	1.4	8	1.4
Anorexia nervosa	2	0.3	1	0.2
Tobacco dependence/abuse	174	30.3	124	21.6
Pathological gambling	9	1.6	0	0

**Table 5**  
Inter-rater reliability (*n* = 575) across all centres

Diagnosis	Kappa	Percent- age agreement	Interviewer/ observer findings <sup>1</sup>
Organic brain syndrome	0.90	98.3	517 4 6 60
Any depressive disorder	0.95	97.6	345 8 6 216
dysthymia	0.96	99.0	472 5 1 97
major depressive - single	0.97	99.7	531 1 1 42
major depressive - recurrent	0.93	97.4	433 8 7 127
Bipolar disorder I	0.92	99.5	533 3 0 19
Bipolar disorder II	0.94	99.5	549 2 1 23
Obsessive-compulsive	0.94	98.6	497 4 4 70
Panic disorder	0.94	98.4	485 4 5 81
Generalised anxiety	0.96	97.9	280 6 6 283
Any phobic disorder	0.98	98.8	312 4 3 256
agoraphobia	0.99	99.5	398 0 3 173
agoraphobia with panic attacks	0.94	98.3	461 5 5 104
simple phobia	0.95	97.9	405 7 5 158
social phobia	0.97	99.1	457 3 2 113
Somatisation	0.67	99.7	571 2 0 2
Schizophrenic disorder	0.91	98.8	529 5 2 39
Schizophreniform disorder	0.89	99.7	565 0 2 8
Anorexia nervosa	0.80	99.8	572 0 1 2
Tobacco dependence/abuse	0.98	99.3	398 1 3 173

1. Number of agreements on subject having no disorder is in the upper left cell; number of agreements on subject having the disorder is in the lower right cell. The remaining cells show the disagreements. The full sample could not be analysed in the case of organic brain syndrome, bipolar I disorder, or agoraphobia.

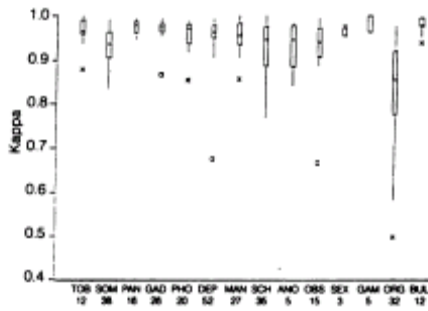


Fig. 1 Kappa score distribution (box-plot) for CIDI symptom questions for each diagnostic category (including PSE items). Box-plots are used to show the data distribution. The box shows the middle 50% of values (the 'spread') and the line inside the box shows the median value ('location'), and the length of line above and below the box ('tail') indicates the remaining 50% of the values. The 'x' indicates 'outliers' in terms of the score distribution; the 'o's are 'extreme outliers'. The bottom line indicates the diagnostic sections and the number of symptom questions for each diagnostic category: TOB = tobacco dependence, SOM = somatisation, PAN = panic disorder, GAD = generalised anxiety, PHO = phobic disorders, DEP = depression, MAN = mania, SCH = schizophrenic disorders, ANO = anorexia nervosa, OBS = obsession, SEX = psychosexual dysfunction, GAM = pathological gambling, ORG = organic brain syndrome, BUL = bulimia.

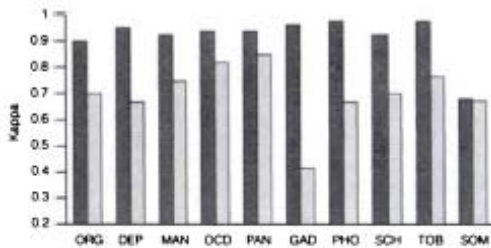


Fig. 2 Reliability of the CIDI: ■ inter-rater ( $n = 575$ ), □ test-retest ( $n = 60$ ; from Semler, 1990) (abbreviations as for Fig. 1).

## References

- AMERICAN PSYCHIATRIC ASSOCIATION (1980) *Diagnostic and Statistical Manual of Mental Disorders (3rd edn) (DSM-III)*. Washington, DC: APA.
- (1987) *Diagnostic and Statistical Manual of Mental Disorders (3rd edn, revised) (DSM-III-R)*. Washington, DC: APA.
- BURKE, J. D. (1986) Diagnostic categorization by the Diagnostic Interview Schedule (DIS): a comparison with other methods of assessment. In *Mental Disorders in the Community. Findings from Psychiatric Epidemiology* (eds 3. Barret & R. M. Rose), pp. 255-285. New York: Guilford Press.
- COTTLER, L. B., ROBINS, L. N. & HELZER, J. E. (1989) The reliability of the CIDI-SAM: a comprehensive substance abuse interview. *British Journal of Addiction*, 84, 810-814.
- & KEATING, S. (1990) Operationalization of alcohol and drug dependence criteria by means of a structured interview. In *Recent Developments in Alcoholism* (ed. M. Gallantes). New York: Plenum.
- , HELZER, J. E. & MAGER, D., et al (1991) Diagnostic agreement between DSM-III and DSM-III-R substance abuse disorders. *Alcohol Abuse and Dependence* (in press).
- EDWARDS, G. & GROSS, M. M. (1976) Alcohol dependence: provisional description of a clinical syndrome. *British Medical Journal*, i, 1058-1061.
- FARMER, A. E., KATZ, R., MCGIFFIN, P., et al (1987) A comparison between the Present State Examination and the Composite International Diagnostic Interviews. *Archives of General Psychiatry*, 44, 1064-1068.
- FEIGHNER, J. P., ROBINS, E., GUZE, S. B., et al (1972) Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26, 57-63.

- HELZER, J. E., ROBINS, L. N., CROUGHAN, J. L., et al (1981) The Renard Diagnostic Interview: its reliability and procedural validity with physicians and lay interviewers. *Archives of General Psychiatry*, 38, 393-398.
- ROBINS, L. N. (1989) Diagnostic grammar and assessment: translating criteria into questions. *The Validity of Diagnosis* (eds L. Robins & J. Barrett). New York: Raven Press.
- , HELZER, J. E., CROUGHAN, J., et al (1981) The National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. *Archives of General Psychiatry*, 38, 381-389.
- , WING, J., WITTCHEN, H.-U., et al (1988) The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry*, 45, 1069-1077.
- , COTTLER, L. B. & BABOR, T. (1990) *The CIDI Substance Abuse Module*. St Louis: Washington University.
- SEMLER, G. (1990) *Reliabilität und Validität des Composite International Diagnostic Interview (CIDI)*. Regensburg: Roderer Verlag.
- , WITTCHEN, H.-U., JOSCHKE, K., et al (1987) Test-retest reliability of a standardized psychiatric interview (DIS/CIDI). *European Archives of Psychiatry and the Neurological Sciences*, 236, 214-222.
- , -- & ZAUDIG, M. (1988) The test-retest reliability of the German version of the Composite International Diagnostic Interview on RDC diagnoses and symptom level. *International Classification in Psychiatry, Unity and Diversity* (eds J. E. Meziah & M. von Cranach), pp. 221-234. London: Cambridge University Press.
- WING, J. K., COOPER, J. E. & SARTORIUS, N. (1974) *The Description and Classification of Psychiatric Symptoms. An Instruction Manual for the PSE and CATEGO System*. London: Cambridge University Press.
- WITTCHEN, H.-U., BURKE, J. D., SEMLER, G., et al (1989) Recall and dating reliability of psychiatric interview (CIDI/DIS). *Archives of General Psychiatry*, 46, 437-443.
- , ROBINS, L. N., COTTLER, L., et al (1990) Interrater reliability of the Composite International Diagnostic Interview (CIDI). Results from the multicenter WHO/ADAMHA field trials (wave I). In *Psychiatry - A World Perspective*, vol. 1 (ed. C. N. Stefanis), pp. 125-132. International Congress Series Vol. 900. Amsterdam: Elsevier.
- WORLD HEALTH ORGANIZATION (1987) *Composite International Diagnostic Interview, Core Version 0.0*. Geneva: WHO.
- (1989a) ICD-10, Chapter V, Mental and Behavioral Disorders (including disorders of psychological development) – Diagnostic Criteria for Research. Draft for field trials. Geneva: WHO.
- (1989b) *Composite International Diagnostic Interview*. Geneva: WHO.