The reliability of the Brazilian version of the Composite International Diagnostic Interview (CIDI 2.1)

M.I. Quintana[1],
S.B. Andreoli[2],
M.R. Jorge[3],
F.L. Gastal[4] and
C.T. Miranda[5]

[1]Departamento de Psiquiatria, UNIFESP, São Paulo, SP, Brasil
[2]Núcleo de Estatística e Metodologia Aplicada, Departamento de Psiquiatria, UNIFESP, São Paulo, SP, Brasil and UNISANTOS, Santos, SP, Brasil
[3]Departamento de Psiquiatria, UNIFESP, São Paulo, SP, Brasil
[4]Escola de Medicina da Universidade Católica de Pelotas, Pelotas, RS, Brasil and Organização Nacional de Acreditação Hospitalar do Brasil, Brasília, DF, Brasil
[5]PAHO, Division of Mental Health, Washington DC, USA

Abstract

Correspondence

M.I. Quintana
Rua Dr. Bacelar, 334
04026-001 São Paulo, SP
Brasil
Fax: +55-11-5084-7060/7061
E-mail: quintana@psiquiatria.epm.br

The objective of the present study was to determine the reliability of the Brazilian version of the Composite International Diagnostic Interview 2.1 (CIDI 2.1) in clinical psychiatry. The CIDI 2.1 was translated into Portuguese using WHO guidelines and reliability was studied using the inter-rater reliability method. The study sample consisted of 186 subjects from psychiatric hospitals and clinics, primary care centers and community services. The interviewers consisted of a group of 13 lay and three non-lay interviewers submitted to the CIDI training. The average interview time was 2 h and 30 min. General reliability ranged from kappa 0.50 to 1. For lifetime diagnoses the reliability ranged from kappa 0.77 (Bipolar Affective Disorder) to 1 (Substance-Related Disorder, Alcohol-Related Disorder, Eating Disorders). Previous year reliability ranged from kappa 0.66 (Obsessive-Compulsive Disorder) to 1 (Dissociative Disorders, Maniac Disorders, Eating Disorders). The poorest reliability rate was found for Mild Depressive Episode (kappa = 0.50) during the previous year. Training proved to be a fundamental factor for maintaining good reliability. Technical knowledge of the questionnaire compensated for the lack of psychiatric knowledge of the lay personnel. Inter-rater reliability was good to excellent for persons in psychiatric practice.

Introduction

The Composite International Diagnostic Interview (CIDI) is a fully standardized, structured interview that provides a psychiatric diagnosis through computerized algorithms (1), according to the International Classification of Diseases, 10th revised edition (ICD 10) (2) and the Diagnostic and Statistical Manual of the American Psychiatric Association, 4th edition (DSM IV) (3). The CIDI was developed in 1980 by the World Health Organization (WHO) in collaboration with the former US Alcohol, Drug Abuse and Mental Health Administration (ADAMHA) as a joint project for the diagnosis and classification of mental disorders, and alcohol and drug-related problems. Its greatest appeal is

that it was designed to be applied by trained lay interviewers in epidemiologic studies and clinical trials and at research centers (4-8). The CIDI can be administered to individuals older than 18 years regardless of their social, economic and cultural status, and does not depend on whether or not such patients are illiterate (7-9).

The CIDI is available in paper-and-pencil and computer-administered forms (self-administered) with a diagnostic coverage for both lifetime and a 12-month period. The average time needed to administer the questionnaire is 75 min (8). The questions are explicit and positive answers are further explored by a specified probing system, the probe for chart (PFC), which determines the psychiatric significance of the symptom in terms of its relevance in the following situations: a) if it interferes significantly with life and activities; b) if the individual had to take medication more than once; c) if the symptom led the individual to consultation with a physician or another professional; d) if a psychiatric etiology was ever attributed to it by a doctor, and if the symptom was never associated with physical illness or injury or use of alcohol, drugs or other medication. The PFC leads the interviewer to a standardized decision tree (algorithm) that will deter-

mine if the symptom was present but was not important enough for the individual to seek assistance (code 2); if the symptom occurred, but was due to the use of medication, drug, alcohol, or caused by trauma or a physical disorder (code 3 and 4) or, finally, if it is a psychiatric symptom (code 5).

The CIDI comprises 288 symptom questions distributed throughout 14 sections, of which 10 are for diagnostic purposes and 4 are non-diagnostic (Table 1). Training for use of the CIDI 2.1 should follow the norms and regulations established by the WHO (8).

The goal of the present investigation was to study the inter-rater reliability of the paper-and-pencil CIDI 2.1 in mental health services. For this purpose, the questionnaire and manuals where translated into Portuguese, as recommended by the World Health Organization.

Material and Methods

The reliability of the CIDI 2.1 was studied by the inter-rater method, which is used by an interviewer and an observer to interview the subject at the same time but making independent codifications. We used the inter-rater method instead of the test-retest method to eliminate any clinical variation and to keep losses to a minimum. A total of 186 subjects were interviewed. The data were obtained from a variety of mental health services in order to improve symptom variability. These included: psychiatric hospitals (No. 82), psychiatric outpatient clinics (No. 54), community health centers (No. 6), and primary care units (No. 40). Data from primary care units were obtained to include non-psychiatric patients. The sample included individuals over 18 years of age, except for Eating Disorder patients, who were 16 years old, since this disorder is more prevalent among teenagers.

The CIDI version 2.1 was translated and submitted to the entire process of trans-cultural adjustment by a bilingual psychiatrist.

Table 1. Diagnostic and non-diagnostic sections of the Brazilian version of the CIDI 2.1.

A. Demographic Aspects
B. Disorders Caused by Tobacco
C. Dissociative and Somatoform Disorders
D. Phobic, Anxiety and Panic Disorders
E. Depressive Disorders
F. Maniac Disorders
G. Schizophrenia and Other Psychotic Disorders
H. Eating Disorders
J. Alcohol Abuse Disorders
K. Obsessive-Compulsive Disorders
L. Psychoactive Substance Abuse Disorders
M. Amnestic Disorders and Other Cognitive Disorders
O. Subject's Comments
P. Interviewer's Observations
X. Interviewer's Codifications

A total of ten research psychiatrists from two São Paulo universities (Federal University of São Paulo and University of São Paulo) analyzed the CIDI section. Rather than doing a back-translation, the team preferred to have each section checked by two specialists in the disorder who evaluated translation aspects, psychological phenomena and cultural adaptation, as recommended by Rubio-Stipec et al. (10). All comments were analyzed and changes were made to improve the questionnaire.

The interviewing team was composed of medical students (lay interviewers) and professional staff in the mental health field (non-lay interviewers). The lay interviewers were submitted to the entire training program as proposed by the WHO. The non-lay interviewers had already been trained in the previous CIDI versions and only required a review. The training followed the format of the WHO (8) except for the introduction of consensus interviews (meetings). Such meetings took place either immediately after the interviews or within a maximum period of two days and consisted of a discussion of any divergent opinions between the interviewer and the observer, with the objective of reaching an agreement. This method was used during the training, the pilot studies and the field trial. The interviewer and the observer were selected at random and alternated their roles (interviewer/observer) for each interview.

A different method was adopted in the selection of the subjects interviewed according to the data collection site. In the psychiatric hospitals and outpatient clinics, the physician was requested to provide the patient's main psychiatric diagnosis, comorbidities and clinical history. The subjects coming from drug abuse treatment facilities and non-governmental organizations were initially submitted to a psychiatric interview. The purpose of this interview was to verify the main diagnosis (using the ICD-10 diagnostic criteria), comorbidity and the subject's life history. At the primary

care unit, the sample was checked in two stages: the Self-Reporting Questionnaire (SRQ 30) (11) was applied to the subjects who were in the waiting room via a non-psychiatric consultation. The subjects were classified either as positive (8 or more positive questions in the SRQ) or negative (less than 7 positive questions). Two to three subjects were randomly chosen from each group and then referred to a psychiatrist for interview in order to confirm the primary diagnosis, comorbidity and the subject's life history. If the subject conformed to the inclusion criteria for the study and agreed to participate, the CIDI interview would then take place (with the interviewers performing a blind study for the psychiatric diagnosis).

Reliability was checked by calculating the kappa coefficient. Further information was obtained for this analysis by evaluating the reports of the interviewers and the consensus meetings, in order to identify any idiosyncrasies observed when administering the questionnaire. Kappa is defined as a randomly adjusted measurement of the agreement between two interviewers that occurs by chance (12). It ranges from 1 (perfect agreement) to -1 (complete disagreement). A kappa score of zero does not indicate a poor level of agreement, but indicates that agreement is no better than the randomly expected level. The method proposed by Landis and Koch (13) was adopted for the interpretation of the kappa values. These investigators suggest that scores higher than 0.75 correspond to a "very good" level of agreement, those between 0.75 and 0.40 correspond to "good agreement" or "satisfactory" levels and those below 0.40, to a "poor" level of agreement.

Results

A total of 186 subjects were interviewed, of whom 54% were women with an average age of 37 years (16-73 years) and 64% were unmarried, divorced, separated or widowed. The average educational level was 7 years of

schooling, and 68% of the subjects were unemployed at the time of the interview. The average duration of the interview was 2 h and 30 min, ranging from 50 min (subject with no psychiatric diagnosis) to 3 h and 40 min (subject with an Eating Disorder diagnosis). Most interviews (80%) were completed in one session.

The diagnostic section that presented the highest number of discrepancies between the interviewer and the observer during the consensus meeting was Depressive Disorders (20.4%), followed by the Somatoform and Dissociative Disorders (19.6%) and by Schizophrenia and Other Psychotic Disorders (17.2%). When each question was evaluated separately, the codification discrepancies were most frequent for questions E12 (During one of those periods did you feel worthless/guilty nearly every day?), E12C (Was the R worthless/guilty only about being impaired by depression?), E29 (In your lifetime, how many different periods have you had that lasted two weeks or more when you felt depressed/lost interest in things/felt a lack of energy and had some of the problems we have been talking about?), and G2 (Was there ever a time when you believed people were following you? Is example im-

plausible?). Questions E12 and G2 guide the interviewer to judge if the symptom that the respondent is referring to could be due to his depressive state (E12) or if it is "implausible" or not (G2), introducing personal judgment into the interview. In E29 we systematically observed two types of error, one in carrying out the questionnaire instructions and the other regarding the understanding of the question by the interviewer. In the analysis of the CIDI 2.1 question-by-question reliability, these same questions presented the lowest kappa values of the questionnaire.

The overall reliability of the CIDI 2.1 was very good, with kappa values equal to 0.94 (SE: 0.035) for lifetime diagnoses and 0.84 (SE: 0.042) for 12-month diagnoses. Diagnostic agreement tended to be closely similar for both periods of time, but slightly higher for lifetime diagnoses (Table 2).

Most of the questions related to the "first time" and "last time" of the occurrence of a symptom showed a very good level of agreement (kappa more than 0.80). The lowest kappa values were found for the questions referring to the "first time" in the following diagnoses: Bipolar Affective Disorder ($\kappa = 0.77$), Mild Depressive Episode ($\kappa = 0.66$), Mania with Psychotic Symptoms ($\kappa = 0.56$), and mixed Obsessive Compulsive Disorder ($\kappa = 0.50$).

Discussion

The Brazilian version of the CIDI 2.1 showed good operational performance in Brazilian mental health services and was also well accepted within different social levels and settings.

The interviewers had difficulties in understanding some of the CIDI questions, especially in the depression, anxiety and schizophrenia sections. Special attention should be given to questions in which the lay examiner uses his/her personal judgment by making allowances for the inclusion of personal concepts, the reliability of the questionnaire is

Table 2. Diagnostic reliability of the Brazilian version of the CIDI 2.1.

	Lifetime diagnoses (kappa value)	12-month diagnoses (kappa value)
Substance-Related Disorder	1.00	–
Alcohol-Related Disorder	1.00	0.93 ± 0.038
Tobacco-Related Disorder	0.99 ± 0.012	0.98 ± 0.019
Schizophrenia	0.87 ± 0.50	0.94 ± 0.041
Maniac Disorders	0.92 ± 0.079	1.00
Bipolar Affective Disorder	0.77 ± 0.011	0.74 ± 0.125
Depressive Disorder	0.95 ± 0.022	0.94 ± 0.027
Dysthymia	0.91 ± 0.49	0.87 ± 0.073
Phobic-Anxiety Disorder	0.92 ± 0.033	0.92 ± 0.033
Obsessive-Compulsive Disorder	0.76 ± 0.133	0.66 ± 0.224
Post-Traumatic Stress	0.99 ± 0.014	0.95 ± 0.034
Dissociative Disorders	0.93 ± 0.042	1.00
Somatoform Disorders	0.93 ± 0.030	0.89 ± 0.053
Eating Disorders	1.00	1.00

Data are reported as means ± SD for 186 subjects.

impaired. A possible solution of these difficulties would be to change the format of the questions, to clarify specific rules, to minimize the influence of clinical judgment in the questionnaire and, finally, to exclude questions that lay interviewers are not able to answer themselves. The presence of severe psychotic symptoms or of intellectual deficits makes it impossible to understand the content in question.

The sections showing the greatest difficulties in administration, with the highest number of discrepancies between the examiner and the observer, involved: Depressive Disorders, Somatoform and Dissociative Disorders, Schizophrenia and Other Psychotic Disorders. Wittchen et al. (7) reported similar problems. In Wittchen's study, 575 subjects were included in 18 centers around the world (media of 25 individuals per center, including Brazil), testing the feasibility, cultural aspects and inter-rater reliability of the CIDI 1.0 in different cultures and settings. The overall acceptance was good (49.3%), agreements for all diagnoses were above 90% and the kappa values were all highly significant. The problems associated with the Somatoform and Dissociative Disorder sections are related to the constant use of the PFC. It would appear that medical knowledge is sometimes required in order to differentiate a psychiatric condition from a secondary disorder due to trauma or use of drugs, alcohol or medication. This perhaps explains the significant number of diagnoses of Somatoform and Dissociative Disorders in the present study. Another hypothesis which merits discussion is that in the event of being unable to identify a physical illness, the examiner often ends up with a Code 5 (positive for a psychiatric symptom) and in this way several false-positive cases are established.

Adequate training on the CIDI has proven to be one of the main factors affecting the reliability of the questionnaire. No statistically significant differences were found in

the performance of the lay interviewers compared to the non-lay interviewers. It can therefore be concluded that with adequate training and learning the rules and methods for administering the questionnaire, there is no need for previous training in psychiatry. Wittchen et al. (7) and Lopes (14) emphasized that for achieving good performance by examiners, it is more important to receive the proper training than to have medical knowledge.

The introduction of the consensus meetings during the training period has been helpful in establishing rules for administering the questionnaire, and in clarifying any doubts that may arise when applying the CIDI 2.1. The meetings were also helpful for controlling the quality of questionnaire administration, for calculating the inter-rater reliability, for verifying "problem questions" and for checking question errors in general.

The authors believe that some comments are required for an appropriate comparison of the results in the present study with those previously published in the relevant literature: a) the studies referred to in the present paper used previous versions of the CIDI; b) except for the multi-centric study coordinated by Wittchen et al. (7), most of the CIDI reliability studies were performed using the test-retest method. This difference in methodology (which reduces the clinical variability) could explain why the kappa values were higher. The overall reliability of the instrument for "lifetime" diagnoses ($\kappa = 0.94$) was higher than the average values reported in the literature, which ranged from 0.60 to 0.93 (15-17). No studies were found in the available literature describing the overall reliability value for the 12-month diagnoses in order to make a comparison with that of the present study ($\kappa = 0.84$). In all of the CIDI 2.1 diagnoses the kappa values were over 0.80, with the exception of the diagnoses of Bipolar Disorder ($\kappa = 0.77$ for lifetime diagnoses and $\kappa = 0.74$ for 12-month diagnoses) and Obsessive-Compulsive Disorder ($\kappa =$

0.76 for lifetime diagnoses and κ = 0.66 for 12-month diagnoses).

The duration of the interview still constitutes a problem and is viewed by the patients as their principal concern. The main complaint about the questionnaire lies in the interviewee's judgment as described by Wittchen et al. (7). In their study they found that 65% of the interviews lasted two or more hours for illiterate subjects or those who had depression symptoms or were alcohol or drug users. In the present format, the average interview time remained unchanged regardless of the subject's educational level. These results require careful evaluation since in the present sample only 5% of the subjects were illiterate whereas 41% had received formal education (more than 5 to 8 years of schooling). However, strangely enough, the interviews of illiterate subjects lasted on average less than 2 h when compared to the group with a higher educational level (2 h and 30 min).

On analyzing the questions related to time, i.e., those referring to the "last time symptoms", it would appear that agreement was lower than that obtained for the "first time". Wittchen et al. (17) showed results that were quite different for the same diagnoses, but also confirmed the fact that the reliability values for "first time" symptoms were lower than those for the appearance of "last time" ones. In the current literature we could not find any explanation for this situation. We believe that the first occurrence of a symptom is dramatic and easily remembered.

In general, the reliability of the Brazilian version of the CIDI 2.1 proved to be high when used under different settings and with subjects having a variety of psychiatric diagnoses. It is also a questionnaire that can be administered by lay interviewers who are well trained.

Acknowledgments

References

1. 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.

2. World Health Organization (1993). *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. World Health Organization, Geneva, Switzerland.

3. American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*. 4th edn. American Psychiatric Association, Washington, DC, USA.

4. Cottler LB, Robins LN, Grant BF et al. (1991). The CIDI-core substance abuse and dependence questions: cross-cultural and nosological issues. The WHO/ADAMHA field trial. *British Journal of Psychiatry*, 159: 653-658.

5. Tiemens BG, VonKorff M & Lin EH (1999). Diagnosis of depression by primary care physicians versus a structured diagnostic interview. Understanding discordance. *General Hospital Psychiatry*, 21: 87-96.

6. Wittchen H-U & Essau CA (1993). Comorbidity and mixed anxiety-depressive disorders: is there epidemiologic evidence? *Journal of Clinical Psychiatry*, 54 (Suppl): 9-15.

7. Wittchen H-U, Robins LN, Cottler LB, Sartorius N, Burke JD & Regier D (1991). Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials. *British Journal of Psychiatry*, 159: 645-653, 658.

8. World Health Organization (1997). *The Composite International Diagnostic Interview* (*CIDI*). World Health Organization, Genebra, Switzerland.

9. Robins LN, Helzer JE, Croughan J & Ratcliff KS (1981). National Institute for Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Archives of General Psychiatry*, 38: 381-389.

10. Rubio-Stipec M, Bravo M & Canino G (1991). The Composite International Diagnostic Interview (CIDI): an epidemiologic instrument suitable for using in conjunction with different diagnostic systems in different cultures. *Acta Psiquiatrica y Psicologica de America*

Latina, 37: 191-204.

11. World Health Organization (1994*). A User's Guide to the Self Reporting Questionnaire*. World Health Organization, Geneva, Switzerland.

12. Spitznagel E & Helzer J (1985). A proposed solution to the base rate problem in the kappa statistic. *Archives of General Psychiatry*, 42: 725-728.

13. Landis JR & Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33: 159-174.

14. Lopes C (1994). Reliability of the Brazilian version of the CIDI in a case-control study of risk factors for drugs abuse among adults in Rio de Janeiro. *Bulletin of PAHO*, 28: 34-41.

15. Farmer AE, Katz R, McGriffin P & Bebbington P (1987). A comparison between the present state exam and the CIDI. *Archives of General Psychiatry*, 44: 1064-1068.

16. Semler G, Wittchen HU, Joschke K, Zaudig M, von Geiso T, Kaiser S, von Cranach M & Pfister H (1987). Test-retest reliability of a standardized psychiatric interview (DIS/CIDI). *European Archives of Psychiatry and Neurological Sciences*, 236: 214-222.

17. Wittchen H-U, Robins LN & Cottler LB (1984). Inter-rater reliability of the Composite International Diagnostic Interview (CIDI). Results from the multi-center WHO/ADAMHA field trials (wave I). In: Stefanis CN (Editor), *Psychiatry - A World Perspective*. Vol. 1. International Congress Series. Vol. 900. Elsevier, Amsterdam, The Netherlands, 125-132.