

The Chinese-bilingual SCID-I/P Project: Stage 1 — Reliability for Mood Disorders and Schizophrenia

E So, I Kam, CM Leung, D Chung, Z Liu, S Fong

Abstract

Objective: To report on the reliability of the Chinese-bilingual Structured Clinical Interview for DSM-IV (Axis I, Patient version) Project (CB-SCID-I/P), stage 1: 'Mood disorders' and 'Schizophrenia and other psychotic disorders'.

Patients and Methods: The CB-SCID-I/P project is a multi-staged translation-validation study of the SCID-I/P (patient research version 2.0, 8/98 revision). Inpatients from the Prince of Wales Hospital were consecutively recruited. Two DSM-IV diagnostic groups were chosen for reliability study for this stage: 'mood disorders' and 'schizophrenia and other psychotic disorders'. Test-retest and joint assessment methods were used to assess inter-rater reliability and clinicians' best-estimate diagnosis was used to assess rater-clinician reliability. Kappa value was used to represent the level of reliability.

Results: 144 inpatients were recruited during a 4-month period. The overall kappa for inter-rater reliability was 0.84, with percentage agreement at 89.6%. Rater-clinician reliability had a kappa of 0.84 for bipolar affective disorder, 0.76 for mood disorder, and 0.75 for schizophrenia. The overall kappa value was 0.77.

Conclusion: A good degree of agreement was achieved between the SCID raters as well as between the rater and clinician diagnoses. The Chinese-bilingual SCID-I/P is therefore a reliable instrument in generating DSM-IV schizophrenia and mood disorders diagnoses for inpatients.

Key words: Chinese, Diagnosis, Inpatients, Mental status schedule, Mood disorders, Schizophrenia

Introduction

A structured psychiatric interview is a procedure for making psychiatric diagnoses via an interviewing process. The term

Acknowledgement: *The authors wish to thank Professor Linda Lam for advising on the project's methodology and assistance in proof reading the original manuscript.*

Dr E So, MBBS, FRANZCP, FRACGP, Department of Psychiatry, Tai Po Hospital, Tai Po, Hong Kong, China.

Dr I Kam, MBChB, MRCPsych, FHKCPsych, FHKAMPsych, Department of Psychiatry, Prince of Wales Hospital, Shatin, Hong Kong, China.

Dr CM Leung, MBBS, MRCPsych, FHKCPsych, FHKAMPsych, Department of Psychiatry, Prince of Wales Hospital, Shatin, Hong Kong, China.

Dr D Chung, MBChB, MRCPsych, FHKCPsych, FHKAMPsych, Department of Psychiatry, Shatin Hospital, Shatin, Hong Kong, China.

Dr Z Liu, MD, Department of Psychiatry, The Second Affiliated Hospital, Hunan Medical University, Changsha, Hunan, China.

Dr S Fong, MBBS, MRCPsych, FHKCPsych, Department of Psychiatry, Shatin Hospital, Shatin, Hong Kong, China.

Address for correspondence: *Dr Eddie So, Department of Psychiatry, Tai Po Hospital, Tai Po, Hong Kong, China.
E-mail: somp@ha.org.hk*

Submitted: 20 August 2002; **Accepted:** 12 August 2003

'structured' implies the use of specific wording of questions that are being asked, a predetermined order of questions to go through, and systematically, a range of diagnoses to be covered. Theoretically, a structured interview gives the advantage of having a standardised assessment procedure, allows wide coverage, and reduces 'premature closure' to information access and verification, therefore leading to improved reliability, and hopefully improved validity of the diagnoses. There can be varying degrees of structure in a structured interview, ranging from fully structured, to semi-structured, to free form clinical interview, the latter essentially a diagnostic interview carried out by the clinician without any diagnostic aids.

The Structured Clinical Interview for DSM-IV (SCID) is a semi-structured face-to-face psychiatric research instrument for obtaining reliable DSM-IV diagnoses, developed by the Biometrics Research, New York State Psychiatric Institute, NY.¹ Since its inception 20 years ago, the SCID has undergone several revisions, from DSM-III, to DSM-III-R, to the present DSM-IV-based diagnostic instrument that incorporates SCID-I for Axis I disorders and the SCID-II for Axis II disorders. Various versions of the SCID-I have been produced to serve different patient groups and specific purposes.²⁻⁷

In addition to its original research designation, it has also been employed in clinical settings for complementing the diagnostic and the documentation process, as well as assisting in interview training. Versions of SCID (for DSM-IV) currently include the standard Patient (SCID-I/P) and the Non-patient (SCID-I/NP) research editions, a shorter Clinician Version (these are for Axis I disorders), and an SCID-II for Axis II disorders. In addition, there is a KID-SCID for children, plus a few specifically modified versions adopted by individual users. These SCID instruments are at present among the most widely used research tools internationally. The SCID has been translated into more than 10 other languages throughout the world, with the patient research version (SCID-I/P) being the most popular. Its utility, reliability and transportability has been well established and reported. The inter-rater reliability of the original SCID-I (DSM-III-R-based), with 592 subjects from 6 sites using test-retest design, gave kappa values above 0.60 for most categories,^{8,9} which is highly comparable to the Diagnostic Interview Schedule (DIS) and Schedule for Affective Disorders and Schizophrenia (SADS).^{10,11}

It is intended that SCID raters are either clinicians or mental health professionals trained to perform the diagnostic assessments. Subjects may be either psychiatric or general medical patients (using the patient version, SCID-I/P), or individuals who do not identify themselves as patients, such as subjects in a community survey of mental illness or family members of psychiatric patients (using the non-patient version, SCID-I/NP).

The language and diagnostic coverage are appropriate for use with adults and adolescents (aged 18 years or older). Individuals with severe cognitive impairment or severe psychotic symptoms are not suitable for SCID interviewing.¹² Intrinsic to the semi-structured design, clinical judgement is demanded when going through the order and questions on the SCID manual, the process being punctuated with follow-up questions for clarification. In addition to interviewing the subject, ancillary data from other sources is permitted in making a final SCID diagnosis, e.g. from family members, previous hospital records, referral notes, and observations of clinical staff.^{13,14}

The first Chinese version of the SCID was available in 1988, translated by Wilson and Young at the Hunan Medical University based on the DSM-III.¹⁵ It was further developed in Taiwan to assimilate the DSM-III-R version.¹⁶ This SCID-I/P (DSM-III-R) version has proved itself popular and presently still sees active service among Hong Kong researchers.¹⁷⁻¹⁹ A feature of this Chinese version is its distinctive Taiwanese colloquial flavour.

There are other instruments available in Chinese such as the translated DIS developed by Robins in the 1970s,²⁰ which has been revised by the Department of Psychiatry, Chinese University of Hong Kong, and used in The Shatin Community Mental Health Survey.^{21,22} Of the other popular research diagnostic instruments, e.g. SADS by Endicott and Spitzer¹¹ and Composite International Diagnostic Interview (CIDI, a fully computer-based instrument, also using the

DSM diagnostic system), none have so far been made available in a Chinese language. Another structured psychiatric interview for the Chinese population is one based upon the Chinese Classification of Mental Disorder, Second Edition, 1989 (CCMD-2), the national system of psychiatric classification officiated in mainland China, which is not used in Hong Kong.²³

As Spitzer and others have succinctly argued, there is not and there may never be a diagnostic gold standard for psychopathology.²⁴⁻²⁶ We concur with their views and consider that diagnostic instruments are not a replacement for properly conducted psychiatric interviews, albeit that his process also has intrinsic limitations.^{24,27} However, large-scale research and epidemiological studies demand collection of data that are reliable as well as valid, in a cost-effective manner, that could also be carried out by staff with less than specialist training. The SCID purported to be systematic and objective in obtaining data through a structured interview format without losing valuable clinical expertise.^{8,28}

We therefore sought to produce a Chinese version of the SCID that would reliably access DSM-IV diagnoses, so that future psychiatric research could be kept compatible with the current DSM system. This will allow local findings to be communicable internationally, as well as facilitating trans-cultural research. The SCID-I/P was chosen for translation. Approval for the translation was obtained from the Biometrics Research, New York State Psychiatric Institute.

Chinese Translation of SCID-I/P

The translation of the SCID-I/P was conducted in collaboration with staff from the Hunan Medical University, PRC. In line with the Spanish-bilingual version of the SCID-I/P, we took advantage of the 3-column format of the SCID-I/P and retained the DSM-IV criteria in English, which is the middle column. The translation is confined to the first column (i.e. the stem and the follow-up questions), leaving the rest of the material in its original language, consequentially the designated Chinese-bilingual version. There are several advantages to a partial translation of the SCID. First and foremost, it avoids the requirement to translate the DSM-IV en-bloc. Second, by preserving a mechanism that permits interchangeability between the DSM-IV criteria and matched CCMD-II criteria, the translated bilingual SCID-I/P can be flexibly modified into an instrument that generates CCMD-II diagnoses, thereby allowing direct comparison between the 2 systems. Furthermore, it is anticipated that when the future DSM-V comes into action, this in-built adaptation will readily provide scope for updating the Chinese-bilingual SCID-I/P (CB-SCID-I/P), by allowing full incorporation of DSM-IV successors. Given the above considerations, a bilingual version is preferable to a comprehensively translated SCID-I. The translation process adopted the standard translation back-translation method,²⁹ the same methodology as used in the translation of the DIS by Chen et al.²¹

The completed CB-SCID-I/P consisted of an 'Overview' section, a 'Screening' section, 9 'Diagnostic Group' modules, 1 'Optional' diagnostic module, and the 'Summary Score Sheet'.

Overview of the Project

The project adopted a multi-stage, multi-site approach. Several factors were influential to the project's design. First, the base rate for different diagnostic entities varies with patient populations and treatment settings. For example, the prevalence of schizophrenia and bipolar disorders in an inpatient cohort is expected to be much higher compared with the outpatient population, and will be different again from that of a community sample. Subjects taken from an inpatient source will therefore be over-represented with schizophrenia and bipolar disorder patients, and under-represented in the anxiety disorders, etc. Second, the low prevalence disorders, such as post-traumatic stress disorder, eating disorder, delusional disorder, or schizoaffective disorder, will naturally represent a much smaller sample unless recruitment is done from highly specialised units. Third, patients with certain disorders with unique psychopathological characteristics — such as hypochondriasis or somatisation disorder — frequently populate medical outpatient clinics or present to GP's rooms, rather than to psychiatric services. Fourth, the level of cooperation from patients and the veracity of their information, could vary significantly depending on the treatment setting.³⁰⁻³² Lastly, there is the effect of individual clinical settings, with their differing areas of expertise, culture, and training emphasis. It is therefore mandatory that an instrument be assessed for its reliability not just within a homogenous group of subjects or clinicians, but across different treatment settings and clinicians of different training backgrounds.^{13,33,14}

The CB-SCID-I/P Project was divided into the following stages:

- stage 1 — single site inter-rater and rater-clinician reliability of CB-SCID-I/P for mood disorders, and Schizophrenia and related psychotic disorders, for adult inpatients
- stage 2 — reliability of CB-SCID-I/P for anxiety disorders, adjustment disorders and 'no diagnoses' in outpatients
- stage 3 — multi-site inter-rater reliability of CB-SCID-I/P
- stage 4 — reliability of CB-SCID-I/P for low prevalence disorders using enhanced sampling technique
- stage 5 — reliability of CB-SCID-I/P on pooled data from stages 1, 2, 3, and 4.

At present, subjects from primary care settings are excluded. The gaps between psychiatric manifestations at the primary care level and a referral psychiatry service are well acknowledged. It is hazardous to apply the DSM-IV criteria in search of psychiatric diagnoses from a primary care sample.³⁴⁻³⁶ Moreover, given the descriptions and diagnostic criteria of the DSM primary care version (DSM-PC), it is likely that the middle column of the SCID-I/P has

to be highly modified to include the DSM-PC diagnostic decision tree if the integrity and efficiency of the SCID-I/P is to be preserved. The SCID-II for personality disorders will not be translated; empirical evidence suggested that the current categorical approach to personality disorder will give way to a dimensional model, such that future assessment would be concerned less with formal 'diagnoses' and more with the measurement of individual differences on relevant trait dimensions.³⁷

According to the original intention, all SCID raters were qualified clinicians and had completed the SCID training workshop.

Patients and Methods

Patients

Inpatients fulfilling the following criteria were consecutively recruited from the Department of Psychiatry, Prince of Wales Hospital between February and May 2000: Cantonese speaking; age from 16 to younger than 70 years; able to give informed and written consent; absence of previous history of major head injury, serious neurological or physical problems, severe pervasive developmental disorder (e.g. autism), or significant cognitive deficits (e.g. mental retardation, dementia).

Subjects for rater-clinician reliability were consecutively recruited using the patient list obtained from the admission office. Subjects for inter-rater reliability were randomly recruited, and further divided into 2 groups: the first group was assessed by the joint interview method, the second group by the test-retest method. No prescreening was carried out.

Raters, Procedures, and Order of Assessment

SCID raters consisted of 2 junior psychiatrists (ES and IK), both of whom completed a 2-day structured SCID training workshop prior to the commencement of the study. The SCID raters were blinded to the admission diagnosis and subsequent events. Diagnoses generated were according to the Chinese-bilingual SCID-I/P manual. No limits were placed on the number of diagnoses generated, although only the 2 most relevant diagnoses were to be assessed.

Although the design of the SCID states that ancillary data should be used for reaching the final diagnosis whenever available, the raters in the current study were not permitted to obtain further information on the patient except the name and date of admission from the admission office. The rationale for this restriction was that by prohibiting such access, the final diagnostic assessment would be based only on the result of the SCID interview and thus be independent of the adequacy of review of the collateral material. This restriction policy paralleled the vigorous testing of SCID as in the multi-site test-retest reliability study reported by Williams et al.⁹

Two senior psychiatrists (CL and DC) provided the best-estimate diagnosis, the 'gold standard' that the SCID diagnoses were to be compared with. They were informed of the names of the subjects who had been successfully

interviewed by the SCID raters, and then approached the subjects individually to arrange for a clinical interview. The clinicians assessed each patient separately; there was no stipulated policy as to the order in which subjects were seen by the clinicians. The clinical interview took place within 2 weeks after the SCID interview, so as to minimise the mental state variation. The clinicians were blinded to the SCID findings. They were allowed free access to the referral notes, case notes, old records, and laboratory results, as well as further information from friends and relatives, if desired, as in a normal diagnostic assessment.

Each clinician generated a primary diagnosis according to the DSM-IV, and a second one for comorbidity when appropriate. In cases where their diagnoses differed, the clinicians reviewed the case together in order to arrive at the best-estimate diagnosis (diagnoses), which was then entered as the “Primary Clinical Diagnosis”, plus a comorbid diagnosis when indicated. Axis II diagnosis was discarded.

Recruitment periods for inter-rater reliability and rater-clinician reliability were conducted separately, with SCID raters always assessing subjects before the clinicians. Two different methods, test-retest and joint interviewing, were employed to assess inter-rater reliability. Subjects for inter-rater reliability testing were randomly recruited and randomly assigned to the SCID raters. Subjects recruited between February and the end of March were jointly interviewed by the SCID raters on the same occasion, within 2 weeks of their admission. This formed the first part of the inter-rater reliability exercise. Subjects recruited from April onward were assessed using the test-retest method, forming the second part for inter-rater reliability evaluation. In line with the joint assessment requirement, the initial rater’s assessment took place within 2 weeks of the admission, and the second rater’s assessment within the subsequent 2 weeks, in order that changes in the mental state were minimised.^{4,9,38} There were no requirements as to the order in which subjects were seen by the SCID raters. Recruitment for rater-clinician reliability assessment occurred between 1 March and 30 May, 2000.

The results for inter-rater reliability and rater-clinician reliability were determined by the kappa index. Unlike percentage agreement, kappa corrects for chance levels of agreement.^{13,39} Kappa values greater than 0.70 are considered to reflect good agreement. Values from 0.50 to 0.70 suggest fair agreement, whereas values less than 0.50 indicate poor agreement. Values less than 0.0 reflect less than chance agreement.⁴⁰

Rationale and Validity of Inter-rater Reliability Methodology

The reliability of a diagnostic instrument is generally evaluated by comparing the degree of agreement between independent evaluations by 2 or more interviewers across a group of subjects, which is a clinical as much as a statistical judgement.⁴¹ As noted elsewhere, both the joint interview method and the test-retest method have their individual merits.^{8,13} The ideal would be to combine the 2 methods within 1 instrument, an approach that would improve the quality of the data obtained, albeit requiring extra effort and expense.

This combined approach was adopted in stage 1 of the CB-SCID-I/P project in the assessment of inter-rater reliability.

During ‘joint interviewing’ (also known as simultaneous rating), 1 rater observed the interview while the other was conducting the assessment. Clarifications with the subject were allowed but there was to be no discussion or exchange of information between the 2 raters during the session. The subject’s response was scored separately, leading to independent diagnoses. The joint interview method also offered an opportunity for the SCID raters to monitor each other’s interviewing style, tracking of the subject’s psychopathology, and in particular encouraged immediate discussion of events at the completion of scoring. Desirably, the raters closely approximated each other in proficiency with the SCID manual by the end of the joint assessment. The attainment of a high level of methodological uniformity allowed the SCID diagnoses from the 2 raters to be legitimately pooled for the subsequent rater-clinician reliability stage.

The test-retest method acted as an ongoing mechanism for monitoring the SCID raters’ level of diagnostic reliability throughout the rater-clinician assessment period, in addition to its role for generating inter-reliability data.

Results

It took between 30 and 75 minutes to conduct a CB-SCID interview, depending on the complexity of the case involved. All data were processed by the Statistical Package for the Social Sciences (SPSS) software, version 9.0.

Analysis of Subjects

257 patients were admitted during the recruitment period for rater-clinician reliability assessment admitted over this period, with an age range of 16 to 92 years. There were 122 males (47.5%) and 135 females (52.5%). 227 (88.3%) fulfilled the selection criteria (age range, 16 to 66 years), including 110 males (48.5%), with a mean age of 36.3 ± 12.4 years. 103 patients (44.7%) were admitted through the emergency department, 72 (33.2%) were admitted from the outpatient clinic, 29 (13%) were transferred from other departments of the hospital, 13 (5.8%) were transferred from another hospital (Shatin Infirmary), and 10 were direct admissions.

130 patients (66 males; 50.7%) gave informed and written consent and completed all assessments. Of these 130 patients, 71 (49.3%) were admitted through the emergency department, 49 (34%) were admitted from the outpatient clinic, 19 (13.2%) were transferred from other departments, and 5 (3.5%) were transferred from the Shatin Infirmary. None were direct admissions. The breakdown for source of admission is closely comparable to the identified population. The SCID raters successfully interviewed all 130 subjects, and the clinicians assessed all subjects within 2 weeks after the SCID interview.

Fourteen subjects (the jointly interviewed group) were also assessed by the clinicians. Among these 14 subjects, the SCID diagnoses of the raters were concordant in 12. The differences were resolved by adopting the SCID

diagnosis that bears the highest hierarchical order. These 14 subjects were added to the pool for rater-clinician reliability, increasing the number of subjects to 144.

The pooled rater-clinician reliability subjects had a mean age of 36.0 ± 12.4 years (range, 16 to 66 years) and included 77 male (53.5%) and 67 female (46.5%) patients. Mean duration of education was 8.7 ± 4.0 years (27% had primary level education or less; 11.8% had secondary level education or higher). 45.8% were married, 43.1% were single, 9.8% were separated or divorced, and 1.4% were widowed. Of the 144 SCID interviewed subjects, 138 were assessed by CL, 140 were assessed by DC, and 134 were assessed by both clinicians, giving a total of 144 subjects and 157 diagnoses for rater-clinician reliability analysis.

Subjects fulfilling the inclusion criteria, who were not successfully recruited or dropped out [97 of 227; 44 males (45.4%)], had a mean age of 37.5 ± 12.8 years. The breakdown of reasons for unsuccessful recruitment/dropout is as follows: 41 declined to participate in the interviews; 23 were discharged before being approached by the SCID rater; 12 remained markedly psychotic by the end of the second week and were unable to give consent; 8 gave consent but were mentally unsuitable for the SCID interviewing; 7 completed the SCID interviews but subsequently withdrew consent for further participation; and 5 had multiple admissions (3 were admitted for weekly maintenance ECT; and 2 were admitted more than once throughout the study period).

Of the 23 patients who were discharged before being approached by the SCID raters, the reasons for early discharge were as follows: crisis admission, with discharge within a few days ($n = 9$); uncontrollable behaviour requiring transfer to gazetted hospital ($n = 5$); discharge against medical advice category (reasons included over-representation of anxiety disorders and personality disorders, and patients with substance use problems who withdrew consent to detoxification; $n = 6$); and patients with chronic schizophrenia who presented themselves to the emergency department after normal working hours were admitted to the acute care wards; on transit to the rehabilitation unit is located in Shatin Infirmary ($n = 3$).

Inter-rater Reliability

Forty seven subjects were interviewed for establishing inter-rater reliability. Twenty seven were assessed by the joint interview method; in the remainder, the test-retest method was used. Both principal and comorbid diagnoses were registered.

Table 1 shows the distribution of diagnoses made by the 2 SCID raters and their diagnostic concordance. Fifty two of 58 diagnoses (89.6%) were in agreement.

Table 2 shows the base rate and kappa value of various diagnoses generated by the 2 methods. The kappa values for mood disorders and schizophrenia from both reliability assessment methods were compatible. Their respective kappa values for mood disorders ($\kappa = 1.00$ and 0.91) and schizophrenia and related psychotic disorders ($\kappa = 0.94$ and 1.00) were high, indicating excellent agreement between the raters and the 2 inter-rater methods.

Alcohol and sub-stance use disorders ($\kappa = 1.00$, minimal qualifying base rate of 5) and anxiety disorders group ($\kappa = 0.79$, minimum qualifying base rate of 3) were also reported. However, the effect of a very low base rate upon the kappa value is apparent, such that these kappa values are very unstable and should be interpreted with caution, if at all.

Table 3 shows the extent of agreement between the SCID raters for the severity of major depressive episodes. No cases were identified as 'mild'. The agreement between the raters in determining cases with moderate severity was marginal, as illustrated by the low kappa value of 0.49. The kappa value improved with cases of severe depression, reflecting better agreement between the raters for severe depressive episode without psychotic features ($\kappa = 0.75$) and very good agreement for subjects with psychotic features ($\kappa = 1.00$).

The raters also had excellent agreement in diagnosing manic episodes (Table 4). Table 5 shows the agreement on the subtypes of schizophrenia. There were no cases of catatonic or residual subtypes. Rater agreement was very good for diagnosis of the 3 subtypes of schizophrenia present (paranoid, disorganised, and undifferentiated).

Table 1. Comparison of diagnoses made by the SCID raters.

Rater: ES (no. of diagnoses)	Rater: IK (no. of diagnoses)						Total
	Mood disorders	Schizophrenia and related psychotic disorders	Alcohol and substance use disorders	Anxiety disorders	Other diagnoses	No diagnosis	
Mood disorders	26	1	—	—	—	1	28
Schizophrenia and related psychotic disorders	—	18	—	—	—	—	18
Alcohol and substance use disorders	—	—	5	—	—	—	5
Anxiety disorders	—	—	—	2	1	—	3
Other diagnoses	—	—	—	—	1	—	1
No diagnosis	2	—	—	1	—	—	3
Total	28	19	5	3	2	1	58

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Table 2. Base rate and kappa values of Axis I disorders in 47 subjects produced by 2 inter-rater reliability assessment methods.

Diagnostic group	No. of diagnoses										κ (Combined)
	Joint interview method					Test-retest method					
	Rater: ES		Rater: IK		κ	Rater: ES		Rater: IK		κ	
	n	%	n	%		n	%	n	%		
Mood disorders	15	42.8	15	44.1	1.00	13	56.5	13	54.2	0.91	0.91
Schizophrenia and related psychotic disorders	10	28.5	11	32.5	0.93	8	34.8	8	33.3	1.00	1.00
Alcohol and substance use disorders	3	8.6	3	8.8	1.00	2	8.7	2	8.3	1.00	1.00
Anxiety disorders	3	8.6	2	5.9	0.79	0	0	1	4.2	—	—
No Axis I diagnosis	1	3.2	1	3.0	—	—	—	—	—	—	—
Unknown/deferred	0	0	1	3.0	—	—	—	—	—	—	—
Total	32	100	33	100		22	100	25	100		Overall k: 0.84

Table 3. Extent of agreement between SCID raters for severity of major depressive episodes.

Severity of major depressive episode (n = 17)	Agreement (no. of cases)		Disagreement (no. of cases)	κ Value
	Positive	Negative		
Mild	0	17	0	—
Moderate	3	9	5	0.49
Severe				
Without psychotic features	8	5	4	0.75
With psychotic features	1	16	0	1.00

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Table 4. Agreement and disagreement between SCID raters on subtypes of schizophrenia.

Type of current episode of bipolar affective disorder (n = 7)	Agreement (no. of cases)		Disagreement (no. of cases)	κ Value
	Positive	Negative		
BAD (all subtypes)	7	40	0	1.000
Mania	5	2	0	1.000
Mixed	0	7	0	—
Hypomania	0	7	0	—
Depressed	1	6	0	—
Unspecified	1	6	0	—

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Table 5. Agreement and disagreement between SCID raters on subtypes of schizophrenia.

Subtype of schizophrenia (n = 11)	Agreement (no. of cases)		Disagreement (no. of cases)	κ Value
	Positive	Negative		
Schizophrenia (all subtypes)	11	35	1	0.94
Paranoid	4	7	0	1.00
Catatonic	0	11	0	—
Disorganised	4	7	0	1.00
Undifferentiated	3	8	0	1.00
Residual	0	11	0	—

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Rater-clinician Reliability

Table 6 contrasts the frequencies of diagnoses (both primary and comorbid) made by the raters and the clinicians. Of the 144 subjects, the SCID raters made 29 comorbid diagnoses along the principal diagnoses, and the clinicians 13. The most common diagnoses made were schizophrenia (21% to 26%), major depressive episode (20% to 25%), and bipolar affective disorder (12% to 15%).

Alcohol and substance use disorder was the most frequent comorbid diagnosis. Two subjects had no Axis I diagnoses. In 2 other cases, the SCID raters deferred from making a diagnosis because of significant reservations as to the veracity of the subjects' information.

The clinicians used 6 fewer diagnostic categories than the raters. The distribution of the principal diagnoses made by the SCID raters and the clinicians (Table 7) shows that, of the 144 principal diagnoses, 24 were in disagreement

(overall $\kappa = 0.77$). Most of the disagreement was clustered within the mood disorder category. The SCID and the clinician best-estimate diagnoses (Table 8) demonstrated good agreement for bipolar affective disorder ($\kappa = 0.84$), major depressive disorder ($\kappa = 0.76$) and schizophrenia ($\kappa = 0.75$).

Alcohol and substance use disorders had the highest kappa, at 0.926; it is included to illustrate the fact that patients admitted voluntarily to detoxification units are highly unlikely to be assessed as not having a substance related problem, whether by a research or clinical diagnostic process.^{6,30}

The criteria of clinician best-estimate diagnoses, the false-negative (best-estimate diagnoses not identified by the SCID) and false-positive (SCID diagnoses not confirmed by the best-estimate diagnoses) rates for the major diagnostic categories were calculated, and the results are shown in Table 9.

Table 6. Frequency of SCID and clinician diagnoses.

Diagnosis	SCID		Clinician best estimate	
	Frequency	%	Frequency	%
Bipolar affective disorder	21	12.1	23	14.7
Major depressive disorder	35	20.2	39	24.9
Dysthymic disorder	10	5.8	2	1.3
Depressive disorder NOS	4	2.3	1	0.6
Substance-induced mood disorder	1	0.6	—	—
Schizophrenia	36	20.8	40	25.6
Schizophreniform disorder	4	2.3	1	0.6
Schizoaffective disorder	2	1.2	—	—
Delusional disorder	3	1.7	3	1.9
Brief psychotic disorder	2	1.2	—	—
Psychotic disorder due to a GMC	1	0.6	1	0.6
Substance-induced psychotic disorder	7	4.0	8	5.1
Psychotic disorder NOS	4	2.3	6	3.8
Alcohol abuse or dependence	4	2.3	4	2.5
Other substance abuse or dependence	11	6.4	11	7.0
Panic disorder	3	1.7	2	1.3
Agoraphobia without history of panic disorder	1	0.6	1	0.6
Obsessive-compulsive disorder	4	2.3	3	1.9
PTSD	2	1.2	1	0.6
Anxiety disorder due to a GMC	1	0.6	—	—
Anxiety disorder NOS	1	0.6	—	—
Pain disorder	1	0.6	—	—
Bulimia nervosa	3	1.7	3	1.9
Adjustment disorder	8	4.5	8	5.1
No Axis I diagnosis	2	1.2	0	0
Diagnosis deferred/unknown	2	1.2	—	—
Total	173	100	157	100

Abbreviations: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version); NOS = not otherwise specified; GMC = general medical condition; PTSD = post-traumatic stress disorder.

Table 7. Comparison of SCID and clinician best-estimate principal diagnoses and overall kappa value.

Clinician (no. of cases)	SCID (no. of cases)					Total*
	Bipolar affective disorder	Major depressive disorder other depressive disorders	Schizophrenia and related psychotic disorders	Alcohol and substance use disorders	Other Axis I diagnoses	
Bipolar affective disorders	19	1	2	—	1	23
Major depressive disorder and other depressive disorders	2	30	2	—	6	40
Schizophrenia and related psychotic disorders	—	3	53	1	2	59
Alcohol and substance use disorders	—	—	—	4	1	5
Other Axis I diagnosis	—	2	—	1	14	17
Total*	21	36	57	6	24	144

* Overall $\kappa = 0.77$.

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Table 8. Base rates and kappa values for SCID raters for SCID rater and clinician DSM-IV principal and comorbid diagnoses (combined).

Diagnoses	SCID diagnoses (n = 173)		Clinician DSM-IV diagnoses (n = 157)		κ Value
	n	%	n	%	
Bipolar affective disorder	21	12.1	23	14.6	0.84
Major depressive disorder	35	21.4	39	25.1	0.76
Schizophrenia	36	20.8	40	25.8	0.75
Alcohol and substance use disorders	15	8.7	15	9.6	0.93

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Table 9. False-negative and false-positive rates for Axis I diagnoses comparing SCID with clinician best-estimate procedure.

Diagnosis	False-negative rate			False-positive rate		
	Diagnosed as present by best estimate	Not identified by SCID		Diagnosed as absent by best estimate	Identified by SCID only	
		n	n		%	n
Bipolar affective disorder	23	4	17.4	121	2	1.7
Major depressive disorder	39	9	23.1	105	4	3.8
Schizophrenia	40	9	22.5	104	5	4.8
Alcohol and substance use disorders	15	1	6.7	129	1	0.8

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

The agreement, sensitivity, and specificity of the specific diagnoses made by SCID are further summarised in Table 10. For bipolar affective disorder, both the specificity (0.98) and the sensitivity (0.83) were high, and the kappa value (0.84) indicated very good agreement. The specificity values for major depressive disorder (0.96) and schizophrenia (0.95) were high but the sensitivities were lower (0.77 and 0.78, respectively), indicating a higher diagnostic threshold from the SCID raters. Nevertheless, the kappa values for these diagnostic categories (0.76 and 0.75, respectively) indicated good agreement between SCID and clinician diagnoses.

The modules B (psychotic symptoms) and C (psychotic differential) of the SCID are designed for detecting psychotic symptoms and categorising them into different diagnostic groups. There was good agreement between SCID rater and clinicians' assessment of psychotic features in the current episode (Table 11).

Analysis of reliability for comorbid diagnoses was calculated by taking all subjects into consideration, irrespective of the principal diagnostic group, the only exception to the study's limitation on diagnostic categories. In all, there were 33 subjects with comorbid diagnosis as

per the SCID raters, the clinicians or both. Only 9 of the 33, however, had comorbid diagnoses entered by both the SCID raters and the clinicians (Table 12). SCID raters made disproportionately more comorbid diagnoses than the clinicians. Among the 9 subjects identified with comorbidity by both SCID raters and the clinicians, 8 were suffering from active alcohol or substance misuse or dependency. The reliability for this subgroup of dual diagnosed subjects was therefore satisfactory, but this was not the case for other categories, in particular those subjects with past history of substance misuse. Overall agreement was poor.

Discussion

Using the Chinese-bilingual SCID-I/P, this study demonstrated a significant degree of diagnostic agreement between the 2 SCID raters and between the SCID raters and the clinicians, when measured against 2 major DSM diagnostic groups (mood disorders and schizophrenia and related psychotic disorders) in a Chinese adult inpatient population. The demographics of the recruited subjects (144 out of 227) were comparable to the sampling population in terms of gender ratio, age range, and source of admission.

Inter-rater Reliability of the CB-SCID-I/P

Overall kappa values of the inter-rater reliability for the mood disorder and the schizophrenia groups were both above 0.9. Compared with the reliability obtained in Williams et al's multi-site study,⁹ our results implied a very high diagnostic reliability. It is acknowledged that a high degree of reliability could naturally be due to the function of the structured interview itself, better diagnostic criteria, or some combination of both. We propose, in addition to the above considerations, that a further explanation might be found in the degree of homogeneity between the SCID raters. A high degree of approximation in administering and scoring SCID diagnoses by the raters is predictable because, as indicated earlier (see Patients and Methods), the joint interviewing method carried the additional purpose of cross monitoring and fine-tuning the raters' performance. The benefit is that, first, in the subsequent rater-clinician reliability assessment, the SCID diagnoses by the 2 raters can be pooled together as 1 group. Second, the 2 raters will set the benchmark for training and evaluation in the project's subsequent multi-site inter-rater reliability assessment.

Of the 2 inter-rater reliability assessment methods, the 'joint interview' method had a higher kappa value for the mood disorders group when compared with the 'test-retest' method, but this result was reversed in the schizophrenia and related psychotic disorders group. This discrepancy is marginal overall, although the finding that the kappa for the test-retest method was not inferior to that of the joint assessment method is contrary to Grove's prediction.⁴⁰ All 4 kappa values were above 0.90, indicating a very satisfactory level of reliability.

For individual diagnostic groups, kappa for a SCID diagnosis of mania was 1.00, i.e. 100% agreement. Kappa

Table 10. Sensitivity, specificity, and kappa values of SCID diagnoses in relation to clinician best-estimate diagnoses.

Diagnosis	Sensitivity	Specificity	κ Value
Bipolar affective disorder	0.83	0.98	0.84
Major depressive disorder	0.77	0.96	0.76
Schizophrenia	0.78	0.95	0.75
Alcohol and substance use disorders	0.93	0.99	0.93

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Table 11. Agreement between SCID diagnoses and clinician best-estimate diagnosis on detection of psychotic symptoms.

	Psychotic features as detected by SCID rater (n)			
	Yes	No	Total*	
Psychotic features as detected by clinician (n)	Yes	63	7	70
	No	5	69	74
Total*	68	76	144	

* $\kappa = 0.83$.

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

Table 12. Agreement between SCID and clinician comorbid diagnosis in 144 subjects.

	Comorbid diagnosis detected by SCID (n)			
	Yes	No	Total*	
Comorbid diagnosis detected by clinician (n)	Yes	9	4	13
	No	20	111	131
Total*	29	115	144	

* $\kappa = 0.347$.

Abbreviation: SCID = Structured Clinical Interview for DSM-IV (Axis I, Patient version).

values for SCID diagnoses across the measured subtypes of schizophrenia were also 1.00 in the paranoid, disorganised, and undifferentiated subtypes.

Amidst the high overall kappa values for mood disorders, differences in the kappa for degree of severity of major depressive episodes (MDE) varied remarkably. Kappa for diagnosing a MDE of moderate severity was less than chance agreement, and for severe MDE without psychotic features was good but substantially below the overall kappa value. This incongruous finding reflected the fact that the diagnosis of a major depressive episode is objectively structured upon the DSM-IV criteria, while severity estimation could only be arrived at with further inference from Axis IV and Axis V information, and is in turn substantively reliant on subjective clinical judgement. The observed discrepancy is therefore most probably intrinsic to the DSM system rather than a weakness of the Chinese-bilingual SCID-I/P instrument or a methodological flaw.

From a statistical point of view, the issue of reliability in severity assessment is worthy of further scrutiny. No subjects were being diagnosed as suffering from a mild form of MDE, resulting in a voided kappa for mild MDE. This was not surprising, as a patient with mild MDE would rarely be admitted. In our study, none of 17 MDE diagnoses were identified as mild by the raters, while 9 cases caused disagreement between the raters as to whether they were moderate or severe in nature. The conclusion is that severity assessment using the CB-SCID-I/P at this stage remains speculative. An outpatient population sample, with higher prevalence of milder forms of psychiatric illnesses, would be more likely to provide a definitive answer on the reliability assessment for severity.

Rater-clinician Reliability of the CB-SCID-I/P

The base rates of major mood disorders and schizophrenia were sufficiently representative to allow for more refined estimation of rater-clinician reliability. Kappa values for these diagnostic groups were all above 0.70, indicating of a good degree of reliability of the Chinese-bilingual SCID-I/P.

Across the board, SCID raters made more types of diagnosis, and were more likely to make diagnoses from the minor diagnostic categories, or disorders that are regarded as nosologically unstable. This phenomenon is believed to be a true reflection of actual clinical practice. Agreement among clinicians for diagnoses such as dysthymia, schizoaffective disorder, brief psychotic disorder, and psychotic disorder NOS were notably poor and uncommon.

Despite the availability of criteria-based diagnostic systems such as the DSM, a good proportion of diagnosticians remain skeptical regarding the instrument's precision and also concerning the vexed issue of diagnostic instability.^{42,43} The controversy of sub-threshold disorders has added further complexity to the issue.³⁶ Where the situation allows, clinicians invariably deflected from committing on such groups of diagnoses.

Results for comorbidity were also interesting and reflected the prevailing pattern in clinical practice. Among the 144 subjects, SCID raters committed 29 comorbid diagnoses in comparison to clinicians' 13. That is, on average 1 in 5 subjects were diagnosed to have SCID comorbidity, versus 1 in 11 by clinicians. It is uncertain whether the SCID instrument leads to over-diagnosis, or the clinicians were in general refraining from making comorbid diagnoses (other than alcohol or substance misuse).

The concept of comorbidity is still relatively nascent, coming to prominence with the development of the DSM-III-R. Authorities struggled with the concept that a collection of DSM symptoms necessarily establishes a separate disorder, as opposed to the traditional wisdom that a variety of psychiatric symptoms are common to a lot of conditions. Adding to the controversy has been the fact that comorbid diagnoses are much more unstable and do not always contribute extra information about the disease and treatment process.^{26,42}

In the present study, there are insufficient data to define the suitability of the Chinese-bilingual SCID-I/P for assessing comorbid diagnoses. In fact, it is highly likely that even with an adequately large sample size, the reliability regarding comorbid diagnosis would be dictated by the style of practice rather than the instrument per se.

Strengths, Limitations, and Implications

Apart from the benefits of the project's multi-stage multi-site design, there are a number of methodological strengths within the current stage 1 study. The SCID ratings were completed by trained psychiatrists who conducted the interviews. All subjects were assessed within 2 weeks of admission. Mental states and diagnostic stability were therefore optimally preserved. Although the original SCID encourages the use of data from sources other than the patient, the raters were provided with minimal information other than demographics when administering the CB-SCID. This gap in ancillary information between the clinical and research interviews provided a reasonable and practical test of the CB-SCID-I/P's comparability to a standard clinical interview.

Another strength of the study is the randomisation process, a crucial aspect for inter-rater reliability study that ensured an even distribution of diagnoses between the 2 SCID raters. The consecutive recruitment process for rater-clinician reliability provided for maximal representation of the various diagnostic entities that characterised the service, such that the chance of a skewed sample was minimal.

A third strength is the use of 2 inter-rater reliability assessment tools. The 'joint assessment' method, as intended, resulted in a highly uniform approach of the raters to SCID-I/P administration, leading to the observed excellent agreement. The joint interview method is also more patient-friendly, as the subject is required to be interviewed only once and is therefore more likely to consent to participate. The drawback, however, is that the SCID raters may find coordinating their schedules throughout the reliability assessment phase an arduous task. The test-retest interview's merit lies in its flexibility to allow for reliable diagnosis to be made for the same subject by different interviewers at different time points, usually within 1 week, albeit extended up to 14 days in the present study.⁴

Lastly, the use of 2 senior psychiatrists with all the information at their disposal in order to arrive at each subject's diagnosis, is perhaps the cornerstone to the instrument's reliability. It is a more rigorous process than diagnosis in the average clinical practice and was used as the 'gold standard'.

The limitations of the present study are that only 2 categories of disorders are included, the paucity of subtypes within each diagnostic group, the lack of reliability of severity assessment (as previously noted), and the relatively small sample size.

Theoretically, a small sample with low base rates will inevitably lead to unstable kappa values, focusing on those diagnoses with higher base rates therefore served to increase the reliability coefficients and compensates for sample size

limitation. Low prevalence diagnostic subtypes, such as delusional disorder, schizoaffective disorders, and those that have low propensity for admissions, e.g. dysthymia, are legitimately under-represented, as demonstrated in our subjects. As planned, these weaknesses shall be compensated in the subsequent stages of the Chinese-bilingual SCID-I/P project, where enhanced sampling techniques will be targeted at these particular subject populations. It is anticipated that by stage 5, subjects from various sites shall be pooled together to build up the total sample size and the range of Axis I disorders.

Apart from the low base rate diagnostic groups, there are other groups of patients where the validity of the Chinese-bilingual SCID-I/P is unknown. This applied to elderly patients, who were therefore excluded, those who for various reasons declined to participate in the project, and in particular, those that were discharged shortly after their admission, before the SCID raters could approach them. The implication is that for the 2 ends of the disease spectrum, reflecting minor or short-term disturbance and severe psychotic illnesses, respectively, the reliability of the CB-SCID-I/P awaits further evaluation.

At an academic level, there are additional concerns about the SCID-I/P and its applications that warrant further consideration. Theoretically, the structure of the SCID interview, accompanied by its diagnostic algorithms, can lead the rater through a different clinical process than the 'usual' psychiatric assessment. Though the SCID procedure is more orderly by nature, it remains controversial as to which method is the more comprehensive. It may be that clinicians obtain information other than that elicited by the SCID, or possibly apply different notions of diagnostic characterisation to similar data.⁴⁴ Furthermore, given the context that SCID is a not a fully structured interview and requires the clinical judgement of the SCID rater, the reliability of the SCID should also be considered as a function of the particular interviewing circumstances in which it is being used.¹² It is unlikely that there will ever be a perfect diagnostic instrument, and possibly the most valid diagnoses can be obtained by adding the data from standardised protocols to extensive data derived from all other clinical and historical sources, which hopefully is achievable in optimum conditions.

Conclusion

The results from this study establish that the CB-SCID-I/P yields highly reliable DSM-IV diagnoses. The CB-SCID-I/P is recommended as an instrument for research of adult inpatients with mood disorders and schizophrenia. Excellent diagnostic agreement was obtained for major depressive episodes, bipolar disorder, and schizophrenia. For identification of psychotic symptoms, and subtypes of bipolar disorder and schizophrenia, reliability is excellent. Severity assessment for non-psychotic major depressive episode was unfortunately disappointing. Our observations also indicate that the CB-SCID-I/P made more comorbid

diagnoses than the unstructured clinical interviews. However, whether this is due to the SCID-I/P instrument's tendency to over-diagnose or the clinicians' higher threshold for comorbidity is as yet unclear. At this stage, use of the CB-SCID-I/P is not recommended in studies on comorbidity of mental disorders. Overall, the compatibility between our results and results those from other studies indicate that the Chinese translated version of SCID-I/P, at least for the A, B, C, and D modules, is a satisfactory reproduction of the original English version.

The present CB-SCID-I/P Project has completed its first stage of reliability assessment. However, there are still gaps in the research data. The applicability of the present data will be strengthened with the completion of the project's subsequent stages, as different diagnostic groups, patient populations, and treatment settings become involved via the multi-stage multi-site approach. In addition, when the CB-SCID-I/P can be administered in a uniform manner to patients from different clinical sites, allowing variations in diagnostic practices to be compared, subtle differences in perspectives and biases within the SCID-I/P may then be identified.

References

1. First MB, Gibbon M, Spitzer RL, Williams JBW. Structured Clinical Interview for Axis I DSM-IV Disorders – Patient Edition – (SCID-I/P, Version 2.0, February 1996 Final Version). Biometrics Research Department, New York State Psychiatric Institute.
2. Skre I, Onstad S, Torgersen S, Kringlen E. High interrater reliability for the Structured Clinical Interview for DSM-III-R Axis I (SCID-I). *Acta Psychiatr Scand* 1991;84:167-173.
3. Arntz A, van Beijsterveldt B, Hoekstra R, et al. The inter-rater reliability of a Dutch version of the Structured Clinical Interview for DSM-III-R Personality Disorders. *Acta Psychiatr Scand* 1992;85:394-400.
4. Williams JB, Spitzer RL, Gibbon M. International reliability of a diagnostic intake procedure for panic disorder. *Am J Psychiatry* 1992; 149:560-562.
5. Dunner DL, Tay LK. Diagnostic reliability of the history of hypomania in bipolar II patients and patients with major depression. *Compr Psychiatry* 1993;4:303-307.
6. Albanese MJ, Bartel RL, Bruno RF, et al. Comparison of measures used to determine substance abuse in an inpatient psychiatric population. *Am J Psychiatry* 1994;151:1077-1078.
7. Fennig S, Craig T, Lavelle J, et al. Best-estimate versus structured interview-based diagnosis in first-admission psychosis. *Compr Psychiatry* 1994;35:341-348.
8. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: history, rationale, and description. *Arch Gen Psychiatry* 1992;49:624-629.
9. Williams JB, Gibbon M, First MB, et al. The Structured Clinical Interview for DSM-III-R (SCID). II: multisite test-retest reliability. *Arch Gen Psychiatry* 1992;49:630-636.
10. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. *Arch Gen Psychiatry* 1981;38:381-389.
11. Endicott J, Spitzer RL. A diagnostic interview: the Schedule for Affective Disorder and Schizophrenia. *Arch Gen Psychiatry* 1978;5:837-844.
12. First MB, Gibbon M, Spitzer RL, Williams JBW. User's Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders – Research Version – (SCID, Version 2.0, February 1996 Final Version). Biometrics Research Department, New York State Psychiatric Institute.
13. Segal DL, Hersen M, Van Hasselt VB. Reliability of the Structured

- Clinical Interview for DSM-III-R: an evaluative review. *Compr Psychiatry* 1994;35:316-327.
14. Ventura J, Liberman RP, Green MF, et al. Training and quality assurance with the Structured Clinical Interview for DSM-IV (SCID-IP). *Psychiatry Res* 1998;79:163-173
 15. Wilson LG, Young DH. Diagnosis of severely ill inpatients in China: a collaborative project using the Structured Clinical Interview for DSM-III (SCID). *J Nerv Ment Disease* 1988;176:585-592.
 16. Lee Y, Chong MY, Hsu CC, Wen JK. Psychiatric diagnoses in patients with chronic fatigue symptoms. [In Chinese]. *Chinese Psychiatry* 1996; 10:234-242.
 17. Lee DT, Wong CK, Cheung LP, et al. Psychiatric morbidity following miscarriage: a prevalence study of Chinese women in HK. *J Affect Disord* 1997;43:63-68.
 18. Lee DT, Yip SK, Chiu HF, et al. Detecting postnatal depression in Chinese women: validation of the Chinese version of the Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 1998;172:433-437.
 19. Lee S, Yu H, Wing YK, et al. Psychiatric morbidity and illness experience of primary care patients with chronic fatigue in Hong Kong. *Am J Psychiatry* 2000;157:380-384.
 20. Bio-psychological Health Survey Inventory for Shatin Residents (a Cantonese Version for the NIMH/DIS-III/84). Hong Kong: The Chinese University of Hong Kong, Department of Psychiatry;1984.
 21. Chen CN, Wong J, Lee N, et al. A two stage screening in community survey: report of a pilot study. In: Yeh EK, Rin H, Yeh CC, Hwu HG, editors. *Prevalence of Mental Disorders*. Taipei, Taiwan: Department of Health;1985:247-251.
 22. Chen CN, Wong J, Lee N, et al. The Shatin community mental health survey in Hong Kong, II. Major Findings. *Arch Gen Psychiatry* 1993;50:125-133.
 23. Chinese Association of Neurology and Psychiatry. *Chinese Classification of Mental Disorders*. [In Chinese]. 2nd ed. Changsa: Publishing House of Hunan University 1989.
 24. Robins LN. Epidemiology: Reflections on testing the validity of psychiatric interviews. *Arch Gen Psychiatry* 1985;42:918-924.
 25. Robins LN, Helzer JE, Ratcliff KS, Seyfried W. Validity of the Diagnostic Interview Schedule, Version II DSM-III diagnoses. *Psychol Med* 1982;12:855-870.
 26. Spitzer RL. Psychiatric diagnosis: are clinicians still necessary? *Compr Psychiatry* 1983;24:399-411.
 27. Kosten TA, Rounsaville BJ. Sensitivity of psychiatric diagnosis based on the best estimate procedure. *Am J Psychiatry* 1992;149:1225-1227.
 28. Spitzer RL, Williams JB, Miriam G, First MB. *Structured Clinical Interview for DSM-III-R-Patient Edition (SCID-P, Version 1.0)*. Washington, DC: American Psychiatric Press;1990.
 29. Behling O, Law KS. *Translating questionnaires and other research instruments: problems and solutions*. London: SAGE University Paper; 2000.
 30. Goldfinger SM, Schutt RK, Seidman LJ, et al. Self-report and observer measures of substance abuse among homeless mentally ill persons in the cross-section and over time. *J Nerv Ment Dis* 1996;184: 667-672.
 31. Kranzler HR, Kadden RM, Bureson JA, et al. Validity of psychiatric diagnoses in patients with substance use disorders – is the interview more important than the interviewer? *Compr Psychiatry* 1995;36:278-288.
 32. Kranzler HR, Kadden RM, Babor TF, et al. Validity of the SCID in substance abuse patients. *Addiction* 1996;91:859-868.
 33. Strakowski SM, Hawkins JM, Keck PE, et al. The effects of race and information variance on disagreement between psychiatric emergency service and research diagnoses in first-episode psychosis. *J Clin Psychiatry* 1997;58:457-463.
 34. Frances A, First MB, Pincus HA, Widiger T. An introduction to DSM-IV. *Hosp Community Psychiatry* 1990;41:493-494.
 35. Pincus HA, Vettorello M, McQueen LE, et al. Bridging the gap between Psychiatry and Primary Care: The DSM-IV-PC. *Psychosomatics* 1995;36:328-335.
 36. Pincus HA, Wakefield Davis W, McQueen LE. Subthreshold mental disorders. *Br J Psychiatry* 1999;174:288-296.
 37. Parker G, Barrett, E. Editorial: personality and personality disorder: current issues and directions. *Psychol Med* 2000;30:1-9.
 38. Malow RM, West JA, Williams JL, Sutker PB. Personality disorders classification and symptoms in cocaine and opioid addicts. *J Consult Clin Psychol* 1989;57:765-767
 39. Fleiss JL. *Statistical methods for rates and proportions*. 2nd ed. New York, NY: Wiley;1981.
 40. Grove WM. The reliability of psychiatry diagnosis. In: Last CG, Hersen M, editors. *Issues in Diagnostic Research*. New York: Plenum; 1987:99-119.
 41. Dunn G. *Design and Analysis of Reliability Studies*. New York: Oxford University Press;1989.
 42. Blacker D, Tsuang MT. Contested boundaries of bipolar disorder and the limits of categorical diagnosis in psychiatry. *Am J Psychiatry* 1992;149:1473-1483.
 43. Hickie IB, Scott EM, Davenport TA. Somatic distress: developing more integrated concepts. *Curr Opin Psychiatry* 1998;11:153-158.
 44. Steiner JI, Tebes JK, Sledge WH, et al. A comparison of the Structured Clinical Interview for DSM-III-R and clinical diagnoses. *J Nerv Ment Dis* 1995;183:365-369.