

# The Chinese Bilingual SCID-I/P Project: Stage 3 — Multi-site Inter-rater Reliability

E So, I Kam, L Lam

## Abstract

**Objective:** The Chinese Bilingual Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders-IV is a recently introduced semi-structured psychiatric diagnostic instrument for use for Chinese patients. The aim of this study was to ascertain the multi-site inter-rater reliability.

**Patients and Methods:** Newly registered outpatients from 4 centres were consecutively recruited. Multi-site inter-rater reliability is expressed in percentage agreement.

**Results:** Forty one outpatients from 14 diagnostic categories were recruited. When principal diagnosis, comorbidity, and life-time diagnosis were considered together, the overall percentage agreement was 74% — 75% for the Li Ka Shing Psychiatric Centre, 74% for the Alice Ho Nethersole Hospital Psychiatric Clinic, 60% for the North District Hospital Psychiatric Clinic, and the 85% for the Alice Ho Nethersole Hospital Pain Clinic. When principal diagnosis alone was considered, the percentage agreement was 82% for the Li Ka Shing Psychiatric Centre, 83% for the Alice Ho Nethersole Hospital Psychiatric Clinic, 63% for the North District Hospital Psychiatric Clinic, and 90% for the Alice Ho Nethersole Hospital Pain Clinic. The overall percentage agreement for principal diagnoses was 80%, which improved to 87% when only the psychiatric patients were considered.

**Conclusions:** The Chinese Bilingual Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders-IV is a reliable diagnostic instrument with a satisfactory degree of agreement among different raters and across different sites for use in the outpatient setting. Specifically, the inter-rater reliability is superior for psychiatric than for non-psychiatric settings. At this stage, use for comorbid diagnosis is not recommended.

**Key words:** Asian continental ancestry group, Outpatients, Reproducibility of results

## Introduction

The Chinese Bilingual Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV; CB-SCID-I/P) is a semi-structured diagnostic instrument for making the Axis I DSM-IV diagnoses.<sup>1</sup> The CB-SCID-I/P is a semi-translated Chinese version of the SCID-I/P,<sup>2</sup> in which the DSM-IV diagnostic criteria remains in its original English format. Translation followed the standard back-translation procedure.<sup>3</sup> Research methodology paralleled that of the original authors.<sup>4,5</sup> The study is

divided into the following 5 stages:

- stage 1 — translation of the SCID-I/P. Test-retest reliability and clinician-rater reliability of the CB-SCID-I/P on mood disorders and schizophrenia and related psychotic disorders in an inpatient population
- stage 2 — clinician-rater reliability of the CB-SCID-I/P on anxiety disorders, adjustment disorders, and ‘no diagnoses’ in an outpatient population
- stage 3 — multi-site inter-rater reliability of the CB-SCID-I/P in an outpatient population
- stage 4 — reliability of the CB-SCID-I/P for low-prevalence disorders by enhanced sampling
- stage 5 — reliability of the SCID-I/P major psychiatric diagnoses against pooled data from stages 1, 2, and 4.

Stages 1 and 2 have already been completed.<sup>6,7</sup> In stage 1, test-retest reliability between the SCID raters achieved percentage agreement of 89.6% with  $\kappa$  of 0.84. When compared with the clinicians’ best-estimated diagnoses, the overall  $\kappa$  for clinician-rater reliability was 0.77.<sup>6,7</sup>

The results from stage 2 gave an overall  $\kappa$  of 0.71 for clinician-rater reliability.<sup>6</sup> Taken together, stages 1 and 2 support the CB-SCID-I/P as a sensitive and specific

*Dr Eddie So, FHKCPsych, FHKAM, Department of Psychiatry, Tai Po Hospital, Tai Po, Hong Kong, China.*

*Dr Irene Kam, MRCPsych, FHKAM, Department of Psychiatry, Shatin Hospital, Shatin, Hong Kong, China.*

*Professor Linda Lam, MRCPsych, FHKAM, Department of Psychiatry, Chinese University of Hong Kong, 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:** 27 October 2004; **Accepted:** 1 December 2004.

diagnostic instrument for ascertaining DSM diagnoses in Chinese people.

A limitation of both stage 1 and stage 2 was that the same SCID raters were used. The reliability of the instrument when used by less experienced clinicians remains unknown. This study aimed at establishing the inter-rater reliability among novice raters across various sites. The same transcript of the CB-SCID-I/P was used. Because the level of agreement is lower when the CB-SCID-I/P is administered to outpatients than when it is administered to inpatients,<sup>7</sup> inter-rater reliability between inexperienced and experienced raters would therefore become more sensitive when used for an outpatient cohort. In addition, the use of 2 new recruitment sites, the North District Hospital Psychiatric Clinic (NDH) and the Alice Ho Nethersole Hospital Pain Clinic (PC), a non-psychiatric facility that also acted as a control towards over-diagnosis, ascertained the broader applicability of the instrument.

## Patients and Methods

Between 2002 and 2004, patients were independently and consecutively recruited from the Li Ka Shing Psychiatric Centre (LKS), Alice Ho Nethersole Hospital Psychiatric Clinic (AHNH), NDH, and PC. There was no prescreening and the same recruitment criteria were used. Test-retest design was employed. Data from each clinic were pooled together. Population characteristics approximated each other among the sites. The degree of inter-rater reliability was expressed in the form of percentage agreement. The recruitment criteria were as follows:

- fluent in the Cantonese dialect
- age from 16 to 65 years
- ability to give written informed consent
- absence of history of major head injury, serious neurological

or medical problems, severe pervasive developmental disorder (autism), or significant cognitive deficits (mental retardation, dementia).

Rater 1 consisted of 2 experienced SCID raters, both of whom had been extensively involved in the development of the instrument and stages 1 and 2 of the reliability study. There were 6 novice raters (rater 2) from the 4 sites. Two raters completed the SCID workshop in 2000 (from LKS and AHNH). The others were individually trained to administer the instrument (from NDH and PC). All raters were clinicians. They were allowed to use whatever information was available to them at the time of the assessment. Diagnoses were generated according to the CB-SCID-I/P manual. The number of diagnoses generated was limited to a maximum of 3 — namely, principal diagnosis, comorbid diagnosis, and lifetime diagnosis. The novice conducted the first SCID interview. The retest interviews took place immediately after the initial interview, with the exception of 3 patients for whom the interviews took place 2 weeks apart. The second rater was blinded to the first interviewer's findings. A principal diagnosis according to DSM-IV was generated for each patient. When indicated, a comorbid diagnosis and/or lifetime diagnosis was also entered. Only Axis I diagnoses were recorded.

## Results

It took between 35 and 80 minutes to conduct an SCID interview. All data were processed by the Statistical Package for the Social Sciences version 11.

Forty one patients were recruited, as follows: 11 from LKS, 12 from AHNH, 8 from NDH, and 10 from PC. All patients gave written informed consent and completed the study. Table 1 shows the patients' characteristics according to the study sites. The demographic characteristics of the

*Table 1. Patients' characteristics.*

	Li Ka Shing Psychiatric Centre	Alice Ho Nethersole Hospital Psychiatric Clinic	North District Hospital Psychiatric Clinic	Alice Ho Nethersole Hospital Pain Clinic
Sex				
Male	4	5	3	3
Female	7	7	5	7
Mean age ± standard deviation (range) [years]	41.1 ± 8.4 (21.0 - 47.0)	35.7 ± 11.4 (20.0 - 55.0)	39.4 ± 8.9 (21.0 - 54.0)	42.8 ± 5.3 (37.0 - 51.0)
Mean education ± standard deviation (range) [years]	9.5 ± 3.6 (5.0 - 16.0)	7.6 ± 2.8 (4.0 - 11.0)	8.3 ± 3.7 (4.0 - 16.0)	8.0 ± 3.0 (4.0 - 14.0)
Marital status				
Married	6	8	4	9
Widowed	0	0	1	0
Divorced	2	2	1	1
Separated	0	0	0	0
Single	3	2	2	0
Total	11	12	8	10

4 groups of patients were comparable. Women were over-represented.

### Diagnostic Profile

For the 'principle diagnosis', rater 1 and rater 2 each generated 31 DSM diagnoses and 10 'no diagnosis' for the 41 patients; there was disagreement for the principal diagnosis for 8 patients. Six patients were given a 'comorbid diagnosis', of which half were in agreement. Rater 1 diagnosed 5 patients, whereas rater 2 diagnosed 4 patients with comorbid conditions. Four of these patients were from the AHNH site. Seven patients were given a 'lifetime diagnosis', of which 4 were in agreement. Rater 1 entered 7 lifetime diagnoses and the novices entered 4. Overall, there were 14 DSM categories. Rater 1 and rater 2 disagreed with each other for 11 patients, with 14 discrepant diagnostic pairs. The diagnostic profiles from each site are represented in Tables 2 to 5.

Table 2 shows the diagnostic profile of the 11 patients from LKS. Twelve SCID diagnostic pairs were generated. The raters differed in diagnoses for patients 2, 6, and 11. Both raters agreed that patients 4 and 8 had no psychiatric diagnoses.

Table 3 shows the profile of diagnoses at the AHNH site. There were 18 DSM diagnostic pairs among the 12 patients. Rater 1 entered 16 diagnoses and rater 2 entered 14 diagnoses. The raters differed for patients 12, 14, and 18, amounting to 6 discrepant diagnostic pairs from 3 patients.

Table 4 shows the profile of the 8 patients from NDH. There were 10 diagnostic pairs and 4 were discrepant. Rater 1 made 10 SCID diagnoses versus 9 for rater 2. The raters differed for patients 24, 30, and 31.

Table 5 shows the profile of patients from PC. Rater 1 and rater 2 generated equal amounts of SCID diagnoses. Both

raters agreed that 6 of the patients had no principle psychiatric diagnoses, but disagreed on the diagnosis of 2 patients.

### Multi-site Inter-rater reliability

Table 6 compares the principal diagnoses of rater 1 with those of rater 2 at all sites. The diagnosis of patient 18 is considered to be equivalent to major depressive disorder. The diagnosis of 'substance-induced psychotic disorder' is incorporated into 'schizophrenia and related disorders'. 'Panic disorder with agoraphobia' is included in the 'panic disorder' category. 'Pain disorder' and 'undifferentiated somatoform disorder' are included in 'somatisation disorders'. The number of diagnoses made by the experienced raters was the same as for the novices. There was disagreement about the diagnosis for 8 patients. The percentage agreement between raters 1 and 2 was 80%.

Table 7 shows the correlation between the 3 psychiatric sites. When the PC was excluded, there were 31 patients and those with no diagnosis decreased to 4, resulting in a more specific inter-rater reliability evaluation for the psychiatric cohort. Rater 1 diagnosed 1 more patient than rater 2. The percentage agreement was 87%.

Table 8 shows the data for all sites and all diagnoses. The number of diagnostic pairs increased to 54 and non-concordance increased to 14. The rate of agreement was 74%.

## Discussion

The patient's demographics and the range of DSM diagnoses resembled the characteristics expected from an average outpatient clinic. The profile for principle diagnosis between rater 1 and rater 2 were identical, with no observable bias towards any particular diagnostic pattern. Experienced

Table 2. Diagnostic profile of patients at the Li Ka Shing Psychiatric Centre (n = 11).

Patient number	Principal diagnosis		Comorbid diagnosis		Lifetime diagnosis	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
1	Adjustment disorders	Adjustment disorders	—	—	—	—
2	Dysthymia*	Post-traumatic stress disorder*	—	—	—	—
3	Major depressive disorder	Major depressive disorder	—	—	—	—
4	No diagnosis	No diagnosis	—	—	—	—
5	Adjustment disorders	Adjustment disorders	—	—	—	—
6	Adjustment disorders*	Major depressive disorder*	—	—	—	—
7	Obsessive compulsive disorder	Obsessive compulsive disorder	—	—	—	—
8	No diagnosis	No diagnosis	—	—	—	—
9	Generalised anxiety disorder	Generalised anxiety disorder	—	—	—	—
10	Undifferentiated somatoform disorder	Undifferentiated somatoform disorder	—	—	—	—
11	Panic disorder	Panic disorder	—	—	Major depressive disorder*	—*

\* Discrepant data.

**Table 3. Diagnostic profile of patients at the Alice Ho Nethersole Hospital Psychiatric Clinic (n = 12).**

Patient number	Principal diagnosis		Comorbid diagnosis		Lifetime diagnosis	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
12	Adjustment disorders*	Dysthymia*	Generalised anxiety disorder*	—*	Dysthymia*	—*
13	No diagnosis	No diagnosis	—	—	Substance misuse: hallucinogen	Substance misuse: hallucinogen
14	Agoraphobia without panic*	No diagnosis*	—*	Specific phobia*	—	—
15	Panic disorder	Panic disorder	—	—	—	—
16	Schizophrenia	Schizophrenia	—	—	—	—
17	Panic disorder	Panic disorder	—	—	Major depressive disorder	Major depressive disorder
18	Mood disorder due to general medical condition*	Major depressive disorder*	Agoraphobia without panic	Agoraphobia without panic	—	—
19	Generalised anxiety disorder	Generalised anxiety disorder	—	—	—	—
20	Substance-induced psychotic disorder	Substance-induced psychotic disorder	—	—	—	—
21	Panic disorder	Panic disorder	Social phobia	Social phobia	—	—
22	Panic disorder	Panic disorder	—	—	—	—
23	No diagnosis	No diagnosis	—	—	—	—

\* Discrepant data.

**Table 4. Diagnostic profile of patients at the North District Hospital Psychiatric Clinic (n = 8).**

Patient number	Principal diagnosis		Comorbid diagnosis		Lifetime diagnosis	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
24	Major depressive disorder*	Adjustment disorders*	—	—	—	—
25	Panic disorder with agoraphobia	Panic disorder with agoraphobia	—	—	—	—
26	Substance misuse: hallucinogen	Substance misuse: hallucinogen	—	—	—	—
27	Major depressive disorder	Major depressive disorder	—	—	—	—
28	Generalised anxiety disorder	Generalised anxiety disorder	—	—	—	—
29	Schizophrenia	Schizophrenia	—	—	—	—
30	Generalised anxiety disorder*	Adjustment disorders*	—	—	—	—
31	Major depressive disorder*	Adjustment disorders*	Substance misuse: alcohol and benzodiazepines	Substance misuse: alcohol and benzodiazepines	Dysthymia*	—*

\* Discrepant data.

**Table 5. Diagnostic profile of patients at the Alice Ho Nethersole Hospital Pain Clinic (n = 10).**

Patient number	Principal diagnosis		Comorbid diagnosis		Lifetime diagnosis	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
32	Pain disorder	Pain disorder	—	—	—	—
33	Pain disorder	Pain disorder	Post-traumatic stress disorder*	—*	—	—
34	No diagnosis	No diagnosis	—	—	Major depressive disorder	Major depressive disorder
35	No diagnosis	No diagnosis	—	—	—	—
36	Adjustment disorders	Adjustment disorders	—	—	—	—
37	No diagnosis	No diagnosis	—	—	—	—
38	No diagnosis*	Major depressive disorder*	—	—	—	—
39	No diagnosis	No diagnosis	—	—	—	—
40	No diagnosis	No diagnosis	—	—	—	—
41	No diagnosis	No diagnosis	—	—	Major depressive disorder	Major depressive disorder

\* Discrepant data.

**Table 6. Correlation of principal diagnoses for all patients (n = 41).**

Principal diagnoses by rater 1	Principal diagnoses by rater 2												
	MDD	DT	SRD	SMD	PD	AWOP	PTSD	GAD	SD	OCD	AD	No diagnosis	
Major depressive disorder	3										1		
Dysthymia							1						
Schizophrenia and related disorders			3										
Substance misuse disorders				1									
Panic disorder					6								
Agoraphobia without panic											1		
Post-traumatic stress disorder													
Generalised anxiety disorder								3			1		
Somatisation disorders									3				
Obsessive compulsive disorder										1			
Adjustment disorder	2	1									3		
No diagnosis	1											10	

The number of patients with a principal diagnosis by either rater was 31 and no diagnosis was 10.

Abbreviations: MDD = major depressive disorder; DT = dysthymia; SRD = schizophrenia and related disorders; SMD = substance misuse disorders; PD = panic disorder; AWOP = agoraphobia without panic; PTSD = post-traumatic stress disorder; GAD = generalised anxiety disorder; SD = somatic disorder; OCD = obsessive compulsive disorder; AD = adjustment disorder.

raters are more likely to commit on comorbid and lifetime diagnoses than the less experienced raters (in stage 1, SCID raters made twice the amount of comorbid diagnoses compared with the clinicians).<sup>6,7</sup> However, this trend was not borne out when only non-psychiatric patients were considered. In line with the results from stage 2, this finding refutes the argument that the instrument has a propensity towards over-diagnosis. A sample size of 41 patients is too small to ascertain individual  $\kappa$  scores for each type of disorder, but is sufficient to demonstrate

the overall degree of inter-rater reliability between the 2 groups of raters. When restricted to principle diagnosis, the percentage agreement was 80%, which improved to 87% when only psychiatric patients were considered, closely approximating the 90% percentage agreement in stage 1 of the study. It should be emphasised that stage 1 involved inpatients while stage 3 involved outpatients.

As argued by Robins<sup>8,9</sup> and demonstrated in stage 2,<sup>7</sup> a lower degree of inter-rater reliability is therefore expected due to the reduced prevalence of psychopathology among out-

**Table 7. Correlation of principal diagnoses for all patients from 3 psychiatric sites (n = 31).**

Principal diagnoses by rater 1	Principal diagnoses by rater 2												
	MDD	DT	SRD	SMD	PD	AWOP	PTSD	GAD	SD	OCD	AD	No diagnosis	
Major depressive disorder	3											1	
Dysthymia							1						
Schizophrenia and related disorders			3										
Substance misuse disorders				1									
Panic disorder					6								
Agoraphobia without panic													1
Post-traumatic stress disorder													
Generalised anxiety disorder								3				1	
Somatisation disorders									1				
Obsessive compulsive disorder										1			
Adjustment disorder	2	1										2	
No diagnosis													4

The number of patients with a principal diagnosis by either rater was 31 and no diagnosis was 4.

Abbreviations: MDD = major depressive disorder; DT = dysthymia; SRD = schizophrenia and related disorders; SMD = substance misuse disorders; PD = panic disorder; AWOP = agoraphobia without panic; PTSD = post-traumatic stress disorder; GAD = generalised anxiety disorder; SD = somatic disorder; OCD = obsessive compulsive disorder; AD = adjustment disorder.

**Table 8. Correlation of principal diagnoses for all patients.**

Principal diagnoses by rater 1	Principal diagnoses by rater 2													
	MDD	DT	SRD	SMD	PD	AWOP	SSP	PTSD	GAD	SD	OCD	AD	No diagnosis	
Major depressive disorder	6											1	1	
Dysthymia								1					2	
Schizophrenia and related disorders			3											
Substance misuse disorders				3										
Panic disorder					6									
Agoraphobia without panic						1							1	
Social and specific phobias							1							
Post-traumatic stress disorder													1	
Generalised anxiety disorder									3			1	1	
Somatisation disorders										3				
Obsessive compulsive disorder											1			
Adjustment disorder	2	1										3		
No diagnosis	1						1						10	

The number of patients with a principal diagnosis by either rater was 31 and no diagnosis was 10.

Abbreviations: MDD = major depressive disorder; DT = dysthymia; SRD = schizophrenia and related disorders; SMD = substance misuse disorders; PD = panic disorder; AWOP = agoraphobia without panic; PTSD = post-traumatic stress disorder; GAD = generalised anxiety disorder; SD = somatic disorder; OCD = obsessive compulsive disorder; AD = adjustment disorder.

patients. However, the fact that the percentage agreement becomes less promising when comparing 'psychiatric patients' with 'all patients' suggests that the CB-SCID instrument has less reliability and therefore less credibility for non-psychiatric application. Similarly, when all diagnoses are taken into consideration, the percentage agreement decreases from

87% to 74%, as found in stages 1 and 2, so the authors cannot recommend the use of the instrument for studying psychiatric comorbidity. Table 9 shows the different percentage agreements from each site and with different cohorts.

The strength of the present investigation is its multi-site multi-rater design and its consecutive recruitment process,

**Table 9. Inter-rater reliability for percentage agreement.**

Sites and patients	Principal diagnosis (%)	Combined diagnosis (%)
Li Ka Shing Psychiatric Centre (n = 11)	82	75
Alice Ho Nethersole Hospital Psychiatric Clinic (n = 12)	83	74
North District Hospital Psychiatric Clinic (n = 8)	63	60
Alice Ho Nethersole Hospital Pain Clinic (n = 10)	90	85
Psychiatric patients (n = 31)	87	71
All patients (n = 41)	80	74

which are crucial for ensuring general application of the instrument. The study demonstrated a satisfactory degree of test-retest inter-rater reliability of the CB-SCID-I/P when used by trained novice interviewers from 4 sites. Novices who had attended the training workshop showed better inter-rater reliability than raters who were trained individually. When standardised training is given, the degree of diagnostic accuracy among the novice raters closely approaches that among the experienced raters.<sup>10</sup> The results attest that the administration of the CB-SCID-I/P can be reliably conducted by trained clinicians and does not lead to over-diagnosis.

## Acknowledgements

The authors would like to thank Drs Sandra Chan, Teddy Chan, PT Ho, Simon Lui, Arthur Mak, and TS Wong for their contribution towards the study. Material support from the Prince of Wales Hospital, Alice Ho Nethersole Hospital, and North District Hospital is gratefully acknowledged.

## References

1. American Psychiatric Association. DSM-IV: diagnostic and statistical manual of mental disorders. 4th ed. Washington: American Psychiatric Association; 1994.

2. First MB, Gibbon M, Spitzer RL, Williams JBW. Structured clinical interview for axis I DSM-IV disorders — patient edition. New York: Biometrics Research Department, New York State Psychiatric Institute; 1966.
3. Chen CN, Wong J, Lee N. 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: Department of Health; 1985: 247-251.
4. Williams JB, Gibbon M, First MB, Spitzer RL. The Structured Clinical Interview for DSM-III-R (SCID), II: Multisite test-retest reliability. *Arch Gen Psychiatry* 1992;49:630-636.
5. Williams JB, Spitzer RL, Gibbon M. International reliability of a diagnostic intake procedure for panic disorder. *Am J Psychiatry* 1992;149: 560-562.
6. So E, Kam I, Liu Z, Fong S, Chung D, Leung CM. The Chinese bilingual SCID-I/P project: stage 1 — reliability for mood disorders and schizophrenia. *HK J Psychiatry* 2003;13:7-18.
7. So E, Kam I, Leung CM, Pang A, Lam L. The Chinese bilingual SCID-I/P project: stage 2 — reliability for anxiety disorders, adjustment disorders and 'no diagnosis'. *HK J Psychiatry* 2003;13(1):19-25.
8. Robins LN. Epidemiology: reflections on testing the validity of psychiatric interviews. *Arch Gen Psychiatry* 1985;42:918-924.
9. 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.
10. Ventura J, Liberman RP, Green MF. Training and quality assurance with the structured clinical interview for DSM-IV (SCID-I/P). *Psychiatr Res* 1998;79:163-173.