

A. Somma, S. Borroni, C. Maffei,
E. Besson, A. Garbini, S. Granozio,
B. Limuti, G. Perego, A. Pietrobon,
C. Rugi, E. Turano, A. Fossati

Vita-Salute San Raffaele University, Milan, Italy,
and San Raffaele Turro Hospital, Milan, Italy

Inter-rater reliability of the Italian Translation of the Structured Clinical Interview for *DSM-5* Personality Disorders (SCID-5-PD): a study on consecutively admitted clinical adult participants

Summary

Objectives

The aim of the present study was to evaluate the inter-rater reliability of the Italian translation of the Structured Clinical Interview for DSM-5 Personality Disorders (SCID-5-PD) in a sample of clinical adult participants.

Methods

A sample of 104 in- and outpatients who were consecutively admitted to the Clinical Psychology and Psychotherapy Unit of San Raffaele Turro Hospital, Milan, Italy, were administered the SCID-5-PD by trained graduate clinical psychologist using a pairwise interview design.

Results

In the present study, intraclass correlation coefficient (ICC) values ranged from .88 (Dependent PD and Histrionic PD) to .94 (Avoidant PD) for dimensional SCID-5-PD interview dimensional ratings (median ICC value = .94). Adequate Cohen κ values were observed for SCID-5-PD dichotomous ratings of presence of clinically significant subthreshold features (median κ value = .78, SD = .06), as well as for SCID-5-PD interview categorical PD diagnoses (median κ value = .89, SD = .11).

Conclusions

The present study findings suggest that the Italian translation of the SCID-5-PD is likely to yield reliable assessment of both dimensional and categorical PD diagnoses, at least in a sample of clinical adults who volunteered to ask for psychotherapy treatment.

Key words

Inter-rater reliability • SCID-5-PD • Clinical adult participants

Correspondence

Andrea Fossati
Clinical Psychology and Psychotherapy Unit,
San Raffaele Turro Hospital,
via Stamira d'Ancona 20, 20127, Milan, Italy •
Tel. +39 02 26433241 • Fax +39 02 26433408 •
E-mail: fossati.andrea@hsr

Introduction

Notwithstanding various concerns with the personality disorder (PD) categories in use since the third edition of Diagnostic and Statistical Manual of Mental Disorders (*DSM-5*)¹ – for example, lack of empirically validated cutoffs, extensive comorbidity, and temporal instability² – the *DSM-5*³ retained traditional PD symptom criteria in Section II, which reprints *DSM-IV*⁴ Axis II PD symptom criteria (although an hybrid, dimensional-categorical Alternative Model of PDs was provided in *DSM-5* Section III).

After the publication of *DSM-5*, work began on revising the Structured Clinical Interview for *DSM-IV* Axis II Personality Disorders (SCID-II)⁵. SCID-II has been renamed the Structured Clinical Interview for *DSM-5* Personality Disorders (SCID-5-PD)⁶ to reflect the fact that PDs are no longer listed on Axis II given the elimination of the multiaxial system in *DSM-5*. Over the years, the SCID-II has been widely used in PD research. For instance, some studies have used it to investigate patterns of PD co-

occurrence with other mental disorders⁷⁻¹⁰ or medical conditions^{11,12}. SCID-II has also been used¹³⁻¹⁵ to select groups of study subjects with a particular PD diagnosis (e.g., Antisocial PD, Borderline PD, etc.). Finally, other studies relied on the SCID-II to investigate the underlying structure of personality pathology^{e.g., 16} and for comparison with other assessment methods for PDs^{e.g., 17-19}. The Italian translation of the pre-publication edition of the SCID-II showed adequate inter-rater reliability, as it was indicated by Cohen κ coefficient values ranging from .48 (Mixed PD diagnosis) to .98 (Narcissistic PD diagnosis) for *DSM-IV* axis II PD categorical diagnoses, and by intraclass correlation coefficient (ICC) values ranging from .90 (Depressive PD) to .98 (Antisocial PD) for PD symptom counts²⁰.

Although the *DSM-IV* PD criteria are unchanged in *DSM-5*, the SCID-5-PD is the updated version of the former SCID-II. The SCID-5-PD interview questions have been thoroughly reviewed and revised to optimally capture the construct embodied in the diagnostic criteria. First and colleagues thoroughly revised each SCID-5-PD question in order to ensure that question wording accurately captures the corresponding *DSM-5* PD criterion construct; this revision process ended in a number of major wording changes (i.e., although *DSM-5* Section II PD criteria are identical to *DSM-IV* axis II PD criteria, several questions in SCID-5-PD are markedly different from SCID-II questions assessing the same PD criterion)⁶. In addition, a dimensional scoring component has been added to the SCID-5-PD. Finally, the assessment of the *DSM-IV* research categories Passive-Aggressive (Negativistic) PD and Depressive PD were removed from the SCID-5-PD, given their elimination as research categories in *DSM-5*.

The SCID-5-PD provides at least one question for each diagnostic criterion of the 10 PD diagnoses that are listed in the *DSM-5* (i.e., Avoidant PD, Dependent PD, Obsessive-Compulsive PD, Paranoid PD, Schizotypal PD, Schizoid PD, Histrionic PD, Narcissistic PD, Borderline PD, and Antisocial PD); the SCID-5-PD allows also for Other Specified (i.e., Mixed) PD categorical diagnosis. Ordinarily the entire SCID-5-PD is administered; however, it is also possible to evaluate only those PD that are of particular interest to the clinician or researcher. Different from the SCID-II, after having administered the SCID-5-PD, for each PD, the interviewer first indicates whether the categorical threshold has been met (e.g., at least 5 out of 9 criteria for Borderline PD). If the categorical threshold has not been met for a particular PD, the interviewer is asked to indicate the presence of clinically significant subthreshold features of that PD.

Finally, the SCID-5-PD allows the interviewer to make a dimensional rating for each of the *DSM-5* PDs by summing up the individual scores for the ratings and circling

the appropriate number. Although this is not an official feature of the *DSM-5* Section II PD classification, the idea of dimensionalizing the PD categories in this way has been proposed by researchers as a potentially useful addition to the categorical classification²¹. For each disorder, the interviewer sums up all of the ratings (i.e., “0” = *Absent*, “1” = *Subthreshold*, and “2” = *Threshold*), producing a dimensional score for that disorder that reflects both threshold and subthreshold ratings for the criteria. It should be observed that the highest possible dimensional score for each PD is twice the total number of possible criteria. Different from SCID-II, the SCID-5-PD does not explicitly allow the interviewer to report the symptom count for each PD on the scoring sheet.

Consistent with the SCID-II, SCID-5-PD is also provided with a true/false, self-report personality questionnaire (SCID-5-SPQ) as a screening tool to shorten the time that it takes the clinician to administer the instrument. The SCID-5-SPQ acts as a low-threshold screening device with intentionally high rates of false positives⁶. Each of the 106 questions in the SCID-5-SPQ corresponds to an initial interview question in the SCID-5-PD (identified by numbers in the left-hand column of both instruments), with the exception of adult Antisocial PD criteria that are not listed in the SCID-5-SPQ. Subjects usually need about 20 minutes to complete the SCID-5-SPQ. Afterward, the SCID-5-PD interview is administered, with the clinician needing to inquire only about the items screened positive (i.e., “YES” answers) on the SCID-5-SPQ. To minimize the risk for false negatives, the interviewer is encouraged to explore items for which any evidence emerges during the SCID-5-PD interview, regardless of the subject's response on the SCID-5-SPQ (e.g., asking about suspiciousness if the subject acts suspicious during the interview even though the subject may have denied it on the SCID-5-SPQ). Recently, the SCID-5-PD has been translated into Italian²².

Thus, starting from these considerations, we aimed at testing the inter-rater reliability of the Italian translation of the SCID-5-PD in a sample of consecutively admitted clinical adult participants using a pairwise interview design. Although other methods are available for assessing inter-rater reliability of psychiatric diagnoses that may yield accurate estimates of actual diagnostic agreement (e.g., independent interview designs)²³, in the present study we relied on pairwise interview design because it allows comparability of our findings with previous data on SCID-II inter-rater reliability e.g.,²⁰, while being akin to typical training to diagnostic assessment (i.e., being provided with ecological validity). In order to extend previous data on the psychometric properties of the SCID-II, and to put the SCID-5-PD to a “risky test” of its psychometric characteristics, we included in

the present study only interviewers with limited clinical experience in the assessment of PDs.

Method

Participants

The sample was composed of 104 adult participants who were consecutively admitted to the Clinical Psychology and Psychotherapy Unit of San Raffaele Turro Hospital from September 2016 to March 2017.

Fifty-five (52.9%) participants were female and 49 (47.1%) were male; participants' mean age was 44.30 years, $SD = 14.60$ years. Fifty-one (49.0%) participants were single, 37 (35.6%) participants were married, 13 (12.5%) participants were divorced, and two (1.9%) participants were widows/widowers. Nineteen (18.3%) participants had junior high school degree, 53 (51.0%) participants had high school degree, and 32 (30.7%) participants had university degree. Nineteen (18.3%) participants were unemployed; white collar ($n = 33$, 31.7%), free-lance professional ($n = 20$, 19.2%), and university student ($n = 12$, 11.5%) were the most frequently reported occupations. Fifty-nine (56.7%) participants were inpatients, and 45 (43.3%) were outpatients. Fifty eight (55.8%) participants received at least one *DSM-IV* axis I diagnosis; in this sample, mood disorders ($n = 44$, 42.3%) were the most frequently diagnosed *DSM-IV* axis I diagnosis. *DSM-IV* psychiatric disorder diagnoses were assessed by the clinicians who were following the participants in treatment or by trained clinical psychologists during their initial assessment interviews; since axis I diagnoses were not assessed using standardized interviews and were not the primary focus of this research, they were used mainly for descriptive purposes in the current study.

Psychiatric disorder diagnoses were assessed by the clinicians who were following the participants in treatment or by trained clinical psychologists during their initial assessment interviews; since psychiatric disorder diagnoses were not assessed using standardized interviews and were not the focus of this research, they were used only for descriptive purposes in the current study. All participants were admitted to the Clinical Psychology and Psychotherapy Unit in order to receive psychotherapy treatment for interpersonal difficulties and/or problems with behavior and emotional regulation on a strictly voluntary basis; inpatient participants were referred to the Unit by the clinicians who were following them in treatment.

Potential participants were screened for the following exclusionary criteria: (1) age less than 18 years; (2) IQ less than 80; (3) diagnosis of schizophrenia, schizoaffective disorder, schizophreniform disorder, or delusional disorder according to *DSM-IV* diagnostic criteria; (4) diagnosis of dementia or organic mental disorder ac-

cording to *DSM-IV* diagnostic criteria; and (5) education level lower than elementary school. All participants in the current research passed this screen.

Participants with psychiatric disorder diagnoses were administered the SCID-5-PD after acute symptom remission according to the judgment of the clinicians who were following them in treatment to avoid confounding effects of psychiatric disorders on these measures²⁴. The absence of acute symptom remission was considered an exclusion criterion from the study.

All participants volunteered to take part in the study after being presented with a detailed description and all were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct; none of the participants received an incentive, either directly or indirectly for participating, and were administered all measures as part of their routine clinical assessment.

Participants were administered the SCID-5-PD as part of routine clinical assessment and blind to the aim of the present study; interviewers were also kept blind to the aim of the study (they were required to perform pairwise interviews with independent rating as part of their routine training).

Measures

All participants were administered the Italian translation of the SCID-5-PD²². In the translation process, the authors closely followed Denissen and colleagues' indications²⁵. First, the SCID-5-PD was translated into Italian by one of the authors (A.S.) after obtaining official permission for this translation. The guiding principle was to respect the items' original meaning; then, two of the coauthors (A.F., S.B.) reviewed the translation independently. After reaching a consensus, a native English professional translator translated the Italian version back into English. When the latest version differed from the English original, the first author, the second author and the professional translator come to an agreement on the definitive Italian translation. Then, the consensus translation was sent to the authors for their comments; this process was iteratively carried until final approval of the official Italian translation of the scale.

*Structured Clinical Interview for DSM-5 Personality Disorders (SCID-5-PD)*⁶. The SCID-5-PD is a 119-item semi-structured interview designed to assess the 10 *DSM-5* PDs in Clusters A, B, and C. The SCID-5-PD can be used to make PD diagnoses, either categorically (present or absent) or dimensionally (summing the ratings [0, 1, or 2] for each diagnosis and treating these sums as dimensions). In the present study, the SCID-5-PD was preceded by the administration of its self-report screening questionnaire (SCID-5-SPQ). The validity of the personality questionnaire as a measure for screening SCID-II PD psychopathology has been previously reported²⁶, and SCID-5-PD enables direct probing of

negative SCID-5-SPQ answers when this is considered clinically rerating assessment for the 10 *DSM-5* PDs. In the present study, the number of positively answered items on the SCID-5 SPQ for each PD showed moderate, albeit positive and significant correlations (i.e., Pearson r values) with the SCID-5 dimensional rating for the corresponding PD; in particular, these convergent validity values ranged from .30 (Schizotypal PD) to .67 (Histrionic PD), median r value = .51, 25th percentile = .35, 75th percentile = .56, all $ps < .01$. On average, discriminant validity coefficients of the SCID-5 SPQ (i.e., median r value between the number of positively answered items on the SCID-5 SPQ for a given PD and a SCID-5 dimensional rating for the other PDs) were all small, ranging from -.11 (Histrionic PD) to .23 (Avoidant PD), with a median r value of .01.

Procedures

Since 12 graduate psychologists in their first year of training as clinical psychologists trained in administering the SCID-5-PD participated in the present study, we used a pairwise interview design in order to assess the inter-rater reliability of the SCID-5-PD diagnoses. Raters were paired randomly. Each rater served approximately equally as interviewer and observer. The participant attribution to interviewer-observer pairs was randomized by consecutive admission.

Data analyses

SCID-5-PD dimensional scores inter-rater reliability was assessed using Shrout and Fleiss's one-way ANOVA intraclass correlation coefficient (1,1) (ICC 1,1) ²⁷. Repeated measure ANOVA was used to assess differences in mean scores on dimensional assessed personality disorder ratings. Cohen simple κ coefficient ²⁸ was used to measure interrater reliability of SCID-5-PD categorical judgements; the marginal homogeneity test was used to assess differences in marginal frequencies. κ coefficient was computed only for PDs that were diagnosed at least two times by either interviewer or observer. Cohen κ coefficient was used also to evaluate the inter-rater reliability of the presence of clinically significant subthreshold features of a given PD.

Results

In the present study, the average administration time for SCID-5-PD interview was 90 minutes, $SD = 17.90$ minutes (min. = 60 minutes, max. = 180 minutes). In the present study, the average number of SCID-5-PD interview PD diagnoses was 0.70 (min. = 0 PD diagnosis, max. = 4 PD diagnoses), $SD = 0.73$, for Rater 1 and 0.64, $SD = 0.75$ for Rater 2, ICC(1,1) = .81, 95% confidence interval = .73-.87.

Descriptive and inter-rater reliability statistics for SCID-5-PD dimensional ratings are listed in Table I.

TABLE I. *The Italian translation of the Structured Clinical Interview for DSM-5 Personality Disorders: descriptive and inter-rater reliability statistics for dimensional personality disorder ratings in a sample of consecutively admitted clinical adult participants (N = 104).*

	Rater 1 (Interviewer)		Rater 2 (Observer)		ICC	95% CI
	M	SD	M	SD		
APD	1.81 ^a	2.41	1.88 ^a	2.40	.94	.91 - .96
DPD	2.51 ^a	2.91	2.50 ^a	2.83	.88	.83 - .92
OCPD	2.17 ^a	2.80	2.21 ^a	2.84	.93	.89 - .95
PPD	1.49 ^a	2.37	1.56 ^a	2.58	.94	.91 - .96
SZPD	0.69 ^a	2.19	0.79 ^a	2.25	.96	.94 - .97
SPD	0.19 ^a	0.76	0.22 ^a	0.79	.96	.94 - .97
HPD	2.39 ^a	3.39	2.20 ^a	3.14	.88	.83 - .92
NPD	4.36 ^a	4.93	4.22 ^a	4.77	.92	.89 - .95
BPD	2.88 ^a	3.98	2.63 ^a	3.68	.89	.85 - .93
ASPD	0.50 ^a	1.95	0.56 ^a	2.18	.73	.69 - .86

SCID-5-PD: Structured Clinical Interview for DSM-5 Personality Disorders; ICC: Intraclass correlation coefficient (1,1); 95% CI: 95% confidence interval for ICC (1,1); PD: Personality disorder; APD: Avoidant PD; DPD: Dependent PD; OCPD: Obsessive-compulsive PD; PPD: Paranoid PD; SZPD: Schizotypal PD; SPD: Schizoid PD; HPD: Histrionic PD; NPD: Narcissistic PD; BPD: Borderline PD; ASPD: Antisocial PD. Means with different superscripts are different according to repeated measure ANOVA.

Cohen κ coefficient values for SCID-5-PD presence of clinically significant subthreshold features and categorical PD diagnoses are listed in Table II. The number of participants with individual *DSM-5* PD diagnoses exceeded the number of participants who received any PD diagnoses because of multiple PD diagnoses. Cohen κ coefficient was computed only for conditions that were rated as present for at least two participants by either Rater 1 (i.e., interviewer) or Rater 2 (i.e., observer). Confidence intervals were computed only for ICC(1,1) and Cohen κ coefficient values that were greater than 0.00 and lower than 1.00.

Unfortunately, only κ coefficient values $0.00 < \kappa < 1.00$ (i.e., only κ coefficient values for which we were able to compute the corresponding SE estimates) could be compared with those that were reported in Maffei and colleagues' study ²⁰ on the psychometric properties of the Italian translation of the pre-publication edition of the SCID-II. When we computed the appropriate chi-square statistic for comparing two independent κ coefficients (Fleiss, 1981), we observed no significant difference for Avoidant PD, $\chi^2(1) = 1.90$, $p > .10$, Dependent PD, $\chi^2(1) = 1.89$, $p > .10$, Obsessive-Compulsive PD, $\chi^2(1) = 0.38$, $p > .50$, Paranoid PD, $\chi^2(1) = 0.26$, $p > .60$, Narcissistic PD, $\chi^2(1) = 2.74$, $p > .05$, and Borderline

TABLE II. *The Italian Translation of the Structured Clinical Interview for DSM-5 Personality Disorders: descriptive and inter-rater reliability statistics for the presence of clinically significant subthreshold features and categorical PD diagnoses in a sample of consecutively admitted clinical adult participants (N = 104).*

	SCID-5-PD Presence of Clinically Significant Subthreshold Features						SCID-5-PD Categorical PD Diagnoses					
	Rater 1 (Interviewer)		Rater 2 (Observer)				Rater 1 (Interviewer)		Rater 2 (Observer)			
	N	%	N	%	κ	95% CI	N	%	N	%	κ	95% CI
APD	38	36.5 ^a	45	43.3 ^b	.82	.71 - .93	3	2.9 ^a	3	2.9 ^a	.66	.21 - 1.00
DPD	47	45.2 ^a	55	52.9 ^b	.81	.70 - .92	5	4.8 ^a	5	4.8 ^a	.58	.21 - .95
OCPD	43	41.3 ^a	51	49.0 ^b	.77	.65 - .89	6	5.8 ^a	5	4.8 ^a	.90	.72 - 1.00
PPD	35	33.7 ^a	37	35.6 ^a	.75	.61 - .88	4	3.8 ^a	3	2.9 ^a	.85	.57 - 1.00
SZPD	18	17.3 ^a	19	18.3 ^a	.77	.61 - .93	2	1.9 ^a	2	1.9 ^a	1.00	--
SPD	7	6.7 ^a	8	7.7 ^a	.93	.69 - 1.00	0	0.0	0	0.0	--	--
HPD	40	38.5 ^a	44	42.3 ^a	.78	.63 - .89	6	5.8 ^a	6	5.8 ^a	1.00	--
NPD	57	54.8 ^a	63	60.6 ^a	.73	.59 - .86	21	20.2 ^a	19	18.3 ^a	.88	.76 - .99
BPD	46	44.2 ^a	49	47.1 ^a	.79	.67 - .91	11	10.6 ^a	9	8.7 ^a	.89	.74 - 1.00
ASPD	8	7.7 ^a	9	8.7 ^a	.68	.42 - .94	3	2.9	3	2.9	1.00	--
OSPD	--	--	--	--	--	--	19	18.3 ^a	18	17.3 ^a	.84	.70 - .98
Any PD	--	--	--	--	--	--	63	60.6 ^a	61	58.7 ^a	.82	.70 - .93

SCID-5-PD: Structured Clinical Interview for DSM-5 Personality Disorders; κ : Cohen κ coefficient (1,1); 95% CI: 95% confidence interval for κ coefficient; PD: Personality disorder; APD: Avoidant PD; DPD: Dependent PD; OCPD: Obsessive-compulsive PD; PPD: Paranoid PD; SZPD: Schizotypal PD; SPD: Schizoid PD; HPD: Histrionic PD; NPD: Narcissistic PD; BPD: Borderline PD; ASPD: Antisocial PD; OSPD: Other Specified PD; --: Statistic not computed. The number of participants with individual DSM-5 PD diagnoses exceeded the number of participants who received any PD diagnoses because of multiple PD diagnoses. Proportions with different superscripts are significantly (i.e., $p < .05$) different according to marginal homogeneity test.

PD, $\chi^2(1) = 0.05$, $p > .80$; rather, the Cohen κ value that we observed in the present study for SCID-5-PD Other Specified PD diagnosis ($\kappa = .84$) was significantly larger than the κ value that was reported by Maffei and colleagues²⁰ for SCID-II Not Otherwise Specified (Mixed) PD diagnosis ($\kappa = .46$), $\chi^2(1) = 62.14$, $p < .001$.

Discussion

Consistent with previous findings concerning the Italian translation of the pre-publication edition of the SCID-II²⁰, the Italian translation of the SCID-5-PD seemed to be provided with adequate inter-rater reliability, at least in a clinical sample of adults who voluntarily asked for psychotherapy.

The SCID-5-PD dimensional ratings of *DSM-5* Section II PDs seemed to yield reliable scores in clinical adult subjects, as it was evidenced by *ICC*(1,1) values ranging from .88 (Histrionic PD and Dependent PD) to .96 (Schizotypal PD and Schizoid PD), with a median *ICC*(1,1) value of .94. These *ICC*(1,1) values are suggestive of excellent inter-rater reliability by conventional standards^{29,30}, although care should be used in relying

on arbitrary cut-offs to evaluate effect size³⁰. Interestingly, in our study no significant differences on SCID-5-PD dimensional scores between interviewers and observers. As a whole, our findings indicate that SCID-5-PD may help clinicians in assessing the continuities between adaptive personality features and pathological personality traits. Since SCID-5-PD dimensional ratings do not yield merely symptom counts for the individual *DSM-5* Section II PDs, our *ICC*(1,1) values could not be compared with those that were reported in Maffei and colleagues' study²⁰, which were based on the number of criteria met by a given subjects on the SCID-II.

Although marginal homogeneity assumption was not met for Avoidant PD, $z_{MH} = 2.33$, $p < .05$, Dependent PD, $z_{MH} = 2.53$, $p < .05$, and Obsessive-Compulsive PD, $z_{MH} = 2.31$, $p < .05$, all Cohen κ values suggested that SCID-5-PD dichotomous ratings of presence of clinically significant impairment for subthreshold features were provided with at least adequate inter-rater reliability (median κ value = .78, $SD = .06$). These findings suggested that the Italian translation of the SCID-5-PD may allow clinicians to reliably assess subthreshold personality pathology, which is deemed clinically relevant by

the interviewer. Of course, this result does not mean that subthreshold features that are considered as clinically relevant by SCID-5-PD interviewers actually predict subject's level of impairment. In other terms, our study was designed to provide data on the inter-rater reliability of SCID-5-PD ratings; future studies should give evidence on the clinical validity of SCID-5-PD dimensional ratings, as well as of SCID-5-PD dichotomous ratings of presence of clinically relevant subthreshold features. As it was expected on previous reliability studies based on the SCID-II^{6,20}, the Italian translation of the SCID-5-PD yielded categorical diagnoses of *DSM-5* Section II PDs that were provided with good-to-excellent inter-rater reliability by conventional standards^{29,30}. Although in our sample the base rate was excessively small to yield stable agreement rate estimates for Schizotypal PD, Avoidant PD, and Antisocial PD, whereas Schizoid PD was never rated as present by either interviewers and observers, our data suggest that the Italian translation of the SCID-5-PD may help clinicians to reliably diagnose *DSM-5* Section II PD categories (median κ value = .89, SD = .11), including the controversial category of Other Specified PD diagnosis (which corresponds to *DSM-IV* Mixed PD diagnosis). Of course, we are aware that κ coefficient values may be influenced by a number of issues, including marginal homogeneity and disorder base rate³¹⁻³³; these issues stress the need for further studies on the inter-rater reliability of SCID-5-PD categorical diagnoses based on different samples of clinical participants. The sensitivity of Cohen κ statistic to PD base rate indicates that comparisons between κ coefficient values that were obtained in different samples should be considered with extreme caution³¹⁻³³. Even considering this limitation, our data suggested that the Italian translation of the SCID-5-PD yielded inter-rater reliability indices for the 6 *DSM-5* Section II PD categories that could be compared that were not significantly different from those that were reported in Maffei and colleagues' SCID-II study²⁰. Rather, we observed a κ coefficient value for the SCID-5-PD Other Specified PD diagnosis that was significantly larger than the corresponding κ coefficient value that was reported in Maffei and colleagues' study²⁰ for SCID-II Mixed PD diagnosis.

Of course, our data should be considered in the light of several limitations. Pairwise interview designs are known to yield excessively optimistic estimates of the actual measurement reliability²³; however, it should be observed that pairwise interview designs represent the most commonly used approach to inter-rater reliability assessment because of their simplicity and ecological validity (they are closely akin to the typical training to clinical diagnosis). Our sample was of moderate size, and it was largely composed of participants with Narcissistic PD, Borderline PD, and Other Specified PD diagnosis. Sample with different clinical

and demographic characteristics may yield different results. Moreover, we relied on adult participants; this limits the generalizability of our findings on the inter-rater reliability of the Italian translation of the SCID-5-PD to clinical adolescent populations, as well as to elderly populations. In our study, we focused only on inter-rater reliability of the SCID-5-PD dimensional and categorical ratings, because clinician agreement remains a major aim of *DSM-5* Section II diagnostic criteria. Further studies are badly needed to yield data on test-retest reliability of SCID-5-PD ratings, as well as on the validity of SCID-5-PD dimensional/subthreshold ratings. In our study, the low base rate for selected PD diagnoses prevented us from formally assessing the inter-rater reliability of Schizoid PD; thus, further studies are needed before drawing definitive conclusions on the Italian translation of the SCID-5-PD.

Even keeping these limitations in mind, we feel that our data support the hypothesis that the Italian translation of SCID-5-PD is provided with adequate inter-rater reliability, at least among clinical adult participants who voluntarily asked for psychotherapy treatment.

Appendix

Interrater Reliability

Before any measurement instruments or interviews can be used for research or clinical applications, their reliability must be established. When the measurement method requires raters, as in the case of the SCID-5-PD, the reliability coefficient should take measurement error due to raters into account. Measurement of the extent to which raters assign the same score to the same variable is called inter-rater reliability. In other words, interrater reliability reflects the variation between two or more raters who assess the same group of subjects.

For continuous data, the intraclass correlation coefficient (ICC) is often used to assess interrater reliability. It is the correlation between two measurements made on same subject. Each subject assessed by multiple raters, and ICC helps to answer the question: "To what extent are the ratings within a subject homogeneous?". Ideally, we want raters to be interchangeable. The higher the ICC values the higher the interrater reliability. When two raters are responsible for measuring a variable on a categorical scale (e.g., presence/absence of a disorder, as in the case of categorical PD diagnoses), Cohen's κ represent a useful measure of inter-rater agreement. The higher Cohen's κ value, the better the agreement.

Conflict of interest

None.

References

- ¹ American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 3rd ed. Washington, DC: American Psychiatric Association 1980.
- ² Widiger TA, Trull TJ. *Plate tectonics in the classification of personality disorder*. *Am Psychol* 2007;62:71-83.
- ³ American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington, DC: American Psychiatric Publishing 2013.
- ⁴ American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. Washington, DC: American Psychiatric Association 1994.
- ⁵ First MB, Spitzer RL, Gibbon M, et al. *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID II)*. Washington, DC: American Psychiatric Publishing 1997.
- ⁶ First MB, Williams JBW, Benjamin L, et al. *Structured Clinical Interview for DSM-5 Personality Disorders (SCID-5-PD)*. Washington, DC: American Psychiatric Publishing 2016.
- ⁷ Casadio P, Olivoni D, Ferrari B, et al. *Personality disorders in addiction outpatients: prevalence and effects on psychosocial functioning*. *Subst Abuse* 2014;8:17-24.
- ⁸ Mulder RT, Joyce PR, Frampton CM. *Personality disorders improve in patients treated for major depression*. *Acta Psychiatr Scand* 2010;122:219-25.
- ⁹ Odlaug BL, Schreiber LR, Grant JE. *Personality disorders and dimensions in pathological gambling*. *J Pers Disord* 2012;26:381-92.
- ¹⁰ Williams ED, Reimherr FW, Marchant BK, et al. *Personality disorder in ADHD Part 1: assessment of personality disorder in adult ADHD using data from a clinical trial of OROS methylphenidate*. *Ann Clin Psychiatry* 2010;22:84-93.
- ¹¹ Calderone A, Mauri M, Calabrò PF, et al. *Exploring the concept of eating dyscontrol in severely obese patients candidate to bariatric surgery*. *Clin Obes* 2015;5:22-30.
- ¹² Uguz F, Engin B, Yilmaz E. *Quality of life in patients with chronic idiopathic urticaria: the impact of axis I and axis II psychiatric disorders*. *Gen Hosp Psychiatry* 2008;30:453-7.
- ¹³ Edens JF, Kelley SE, Lilienfeld SO, et al. *DSM-5 antisocial personality disorder: predictive validity in a prison sample*. *Law Hum Behav* 2015;39:123-9.
- ¹⁴ Gremaud-Heitz D, Riemenschneider A, Walter M, et al. *Comorbid atypical depression in borderline personality disorder is common and correlated with anxiety-related psychopathology*. *Compr Psychiatry* 2014;55:650-6.
- ¹⁵ Martín-Blanco A, Ferrer M, Soler J, et al. *Association between methylation of the glucocorticoid receptor gene, childhood maltreatment, and clinical severity in borderline personality disorder*. *J Psychiatr Res* 2014;57:34-40.
- ¹⁶ Sharp C, Wright AG, Fowler JC, et al. *The structure of personality pathology: both general ('g') and specific ('s') factors?* *J Abnorm Psychol* 2015;124:387-98.
- ¹⁷ Huprich SK, Paggeot AV, Samuel DB. *Comparing the Personality Disorder Interview for DSM-IV (PDI-IV) and SCID-II borderline personality disorder scales: an item-response theory analysis*. *J Pers Assess* 2015;97:13-21.
- ¹⁸ Rojas EC, Cummings JR, Bornoalova MA, et al. *A further validation of the Minnesota Borderline Personality Disorder Scale*. *Personal Disord* 2014;5:146-53.
- ¹⁹ Fossati A, Somma A, Borroni S, et al. *Borderline Personality Disorder and Narcissistic Personality Disorder Diagnoses from the Perspective of the DSM-5 Personality Traits: a Study on Italian Clinical Participants*. *J Nerv Ment Dis* 2016. doi: 10.1097/NMD.0000000000000587
- ²⁰ Maffei C, Fossati A, Agostoni I, et al. *Interrater reliability and internal consistency of the Structured Clinical Interview for DSM-IV Axis II personality disorders (SCID-II), version 2.0*. *J Pers Disord* 1997;11:279-84.
- ²¹ Oldham JM, Skodol AE. *Charting the future of Axis II*. *J Pers Disord* 2000;14:17-29.
- ²² First MB, Williams JBW, Benjamin LS, et al. *Structured Clinical Interview for DSM-5*. Arlington, VA, American Psychiatric Association, 2013. Tr. it. Milano: Raffaello Cortina Editore, 2016.
- ²³ Chmielewski M, Clark LA, Bagby RM, et al. *Method matters: understanding diagnostic reliability in DSM-IV and DSM-5*. *J Abnorm Psychol* 2015;124:764-9.
- ²⁴ Zimmerman M. *Diagnosing personality disorders. A review of issues and research methods*. *Arch Gen Psychiatry* 1994;51:225-45.
- ²⁵ Denissen JJ, Geenen R, van Aken MA et al. *Development and validation of a Dutch translation of the Big Five Inventory (BFI)*. *J Pers Assess* 2008;90:152-7.
- ²⁶ Richman H, Nelson-Gray R. *Nonclinical panicker personality: profile and discriminative ability*. *J Anx Dis* 1994;8:33-47.
- ²⁷ Shrout PE, Fleiss JL. *Intraclass correlations: uses in assessing rater reliability*. *Psychol Bull* 1979;86:420-8.
- ²⁸ Fleiss JL. *Statistical methods for rates and proportions*. New York: Wiley, 1981.
- ²⁹ Landis JR, Koch GG. *The measurement of observer agreement for categorical data*. *Biometrics* 1977;33:159-74.
- ³⁰ Cicchetti DV. *Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology*. *Psychol Assess* 1994;6:284-90.
- ³¹ Feinstein AR, Cicchetti DV. *High agreement but low kappa: I. The problems of two paradoxes*. *J Clin Epidemiol* 1990;43:543-9.
- ³² Cicchetti DV, Feinstein AR. *High agreement but low kappa: II. Resolving the paradoxes*. *J Clin Epidemiol* 1990;43:551-8.
- ³³ Viera AJ, Garrett JM. *Understanding inter-observer agreement: the kappa statistic*. *Fam Med* 2005;37:360-3.