

Screening for personality disorder: a comparison of personality disorder assessment by patients and informants.

PAUL WALTERS,¹ PAUL MORAN,¹ PARTHA CHOUDHURY,² TENNYSON LEE,¹ ANTHONY MANN³

¹Health Services Research Department, Institute of Psychiatry, London, UK

²National Institute of Mental Health and Neuroscience, Bangalore, India

³Section of Epidemiology, Institute of Psychiatry, London, UK

ABSTRACT *The Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II Version 2.0) is becoming the most favoured instrument to measure personality disorder but takes up to an hour to complete. The Standardized Assessment of Personality (SAP), an informant-based measure, takes 10 to 15 minutes to complete. Both instruments have been validated independently. This study aimed to determine whether the SAP is a suitable screening instrument for personality disorder as measured by the SCID-II.*

Fifty-seven psychiatric patients were assessed for personality disorder using both the SAP and the SCID-II. The SAP assessments were conducted blind to the results of the SCID-II assessments.

Agreement between the two instruments in this population was low ($\kappa = 0.3$). The level of agreement differed between personality disorder categories, ranging from $\kappa = 0.4$ (antisocial) to -0.1 (narcissistic).

In this population of patients, the SAP proved to be a poor screen for the SCID-II. The study highlights the discrepancy between informant and self-report assessments for personality disorder.

Key words: personality disorder, screening

Introduction

The American Psychiatric Association defines a personality disorder as 'an enduring pattern of inner experience and behaviour that deviates markedly from the expectations of the individual's culture' (American Psychiatric Association, 1994). There is growing evidence that these disorders are associated with significant burden to the individuals affected, those around them and society as a whole. Personality disorders are associated with chronicity of mental illness (Mulder, 2002), suicidal behaviour (Harris and Barraclough, 1997), substance abuse (Brooner, King, Kidorf, Schmidt and Bigelow, 1997), crime (Hodgins, Mednick, Brennan, Schulsinger and Engberg, 1996), and increased health service costs (Rendu, Moran, Patel, Knapp and Mann, 2002). Nevertheless, there is still no consensus as to how to assess personality disorders (Zimmerman, 1994).

Three methods are available: unstandardized clinical interview, standardized clinical interview (semi-structured) and self-report questionnaire. Clinical diagnoses of personality disorder demonstrate poor reliability (Mellsop, Varghese, Joshua and Hicks, 1982) and although the introduction of standardized assessments has resulted in improvements in inter-rater reliability, the test-retest reliability of many instruments remains less satisfactory (Zimmerman, 1994). In addition, the diagnostic agreement between structured interviews and self-report questionnaires is poor, raising questions about the validity of personality disorder assessments (Perry, 1992; Tyrer, 1995).

The following interviews are commonly used to assess personality disorder for research purposes – the Structured Clinical Interview for DSM IV Personality Disorder (SCID-II Version 2.0) (First, Gibbon, Spitzer Williams and Benjamin, 1997), the

Standardized Assessment of Personality (SAP) (Mann, Jenkins, Cutting and Cowen, 1981), the International Personality Disorder Examination (IPDE) (Loranger, Sartorius, Andreoli, Berger, Buchheim, Channabasavanna, Coid, Dahl, Diekstra and Ferguson, 1994), The Diagnostic Interview for Personality Disorders (Zanarini, Frankenburg, Chauncey and Gunderson, 1987), and the Personality Assessment Schedule (Tyrer, Cicchetti, Casey, Fitzpatrick, Oliver, Balter, Giller and Harkness, 1984). The SAP is entirely informant based; the remainder depend on patient information.

One of the main difficulties in assessing personality is that the patient's mental state can bias the assessment. In addition, self-description relies on patients' ability to reflect upon their impact on others and to report honestly about their behaviour. These are abilities that may be specifically absent in some personality-disordered individuals: impairment of the ability to self-reflect is a defining feature of borderline personality disorder; deception and lying are features of antisocial personality disorder. The use of informants for information on personality status theoretically overcomes these problems (Mann, Jenkins et al., 1981), although informants may themselves have disturbed mental states or their own reasons for deceiving the interviewer.

From a review of studies comparing patient and informant accounts, Zimmerman (1994) reported that informants tend to provide more symptoms of personality disorder than patients. He suggested that the poor agreement between self-report and informant-based measures could be due to the following: social desirability biasing the subject's responses, denial of negative attributes due to subject's lack of insight, or the informant's confusion of pre-morbid personality with symptoms caused by episodes of illness (Zimmerman, 1994).

More recently three studies have compared patient and informant accounts of personality disorder. In a study of 90 psychiatric outpatients from Bangalore, India, Mann et al. (1999) compared the International Personality Disorder Examination (IPDE) (a semi-structured interview with the subject) and the Standardised Assessment of Personality (SAP) (an informant-based semi-structured interview) (Mann, Raven, Pilgrim, Khanna, Velayudham, Suresh, Channabasavanna, Janca and Sartorius, 1999). Overall agreement between the two instruments in

the detection of personality disorder was low ($\kappa = 0.4$). The level of agreement varied according to personality disorder category ranging from dependent personality disorder ($\kappa = 0.66$) to dissocial personality disorder ($\kappa = 0.09$).

Modestin and Puhan (2000) compared patients' self-response with informant accounts' of personality pathology (Modestin and Puhan, 2000). Although patient and informant accounts yielded the same number of diagnoses, agreement between the three sets of data was generally poor. Dreessen et al. (1998) examined the concordance between 42 psychotherapy outpatients and their informants using the SCID-II (Dreessen, Hildebrand, and Arntz, 1998). Again, only low or modest levels of agreement were found.

The SCID-II is a semi-structured interview with the patient. It is rapidly becoming the favoured personality disorder assessment tool by international researchers. However its usefulness as a case detector for use in epidemiological surveys and routine clinical work is limited by the fact that it takes a long time to administer. Mann et al. (1999) have already indicated that the SAP may be useful as a screen for IPDE personality disorder diagnoses in samples where the expected prevalence of personality disorder is low (Mann, Raven et al., 1999). They proposed a two-stage approach for epidemiological studies, with the longer instrument only being applied to those that screen positive for a SAP diagnosis of personality disorder (Mann, Raven et al., 1999). The aim of the current study was to determine the level of agreement between the SAP and the SCID-II. In particular we were interested in determining whether the SAP could act in the same way if the SCID-II is the 'gold standard'.

The Standardized Assessment of Personality (SAP) (Mann, Jenkins et al., 1981)

This is a semi-structured questionnaire conducted with an informant either face-to-face or by telephone. It takes about 10 to 15 minutes to complete. It has good inter-rater ($\kappa = 0.76$), inter-temporal ($\kappa = 0.54-0.75$) and inter-informant reliability ($\kappa = 0.93-0.96$) (McKeon, Roa, and Mann, 1984; Pilgrim, Mellers, Boothby and Mann, 1993; Ormel, Oldehinkel, Brilman and Brink, 1993).

The SAP opens with a screening set of questions used to identify key words that are then used to explore categories of personality disorder. Only those

categories with an identified key word are explored in more detail. In order to score positively for a particular personality disorder, a threshold number of criteria have to be reached and there has to be evidence of personal, social or occupational disability.

The Structured Clinical Interview for DSM IV Personality Disorder (SCID-II Version 2.0) (First, Gibbon et al., 1997)

This is a semi-structured diagnostic interview with the participant to assess DSM-IV personality disorder. Questions may necessitate further exploratory questions by the examiner in order to score a particular item. An item is scored as 1 (absent), 2 (sub-threshold), 3 (threshold). Multiple examples of specific items are frequently needed. It was designed to generate DSM-IV personality disorder diagnoses.

The SCID-II takes up to 1 hour to administer. Items are grouped according to the particular personality disorder. If a threshold is reached on a sufficient number of items in each personality disorder category, the disorder is deemed to be present. The instrument demonstrates acceptable internal consistency and inter-rater reliability (Maffei, Fossati, Agostoni, Barraco, Bagnato, Deborah, Namia, Novella and Petrachi, 1997). Good rates of diagnostic agreement between the DSM-III-R version and the DSM-IV version of the SCID-II have been reported except for the histrionic and dependent diagnostic categories (Poling, Rounsaville, Ball, Tennen, Kranzler and Triffleman, 1999).

Method

Site and study participants

A non-random sample of 57 participants was recruited from inpatient and outpatient clinics at The Maudsley Hospital, London. The sample was chosen to represent patients with a range of psychiatric diagnoses and likely to have a high prevalence of personality disorder to test agreement levels efficiently in a clinical sample. Sixty-seven per cent ($n = 38$) of the participants were male. The mean age of the participants was 41 years ($SD = 10$). The population was predominantly white (80%), unemployed (62%), 65% were educated to the age of 16 years only. The case-note diagnoses of these patients were: personality disorder ($n = 15$); depression ($n = 12$); anxiety disorder ($n = 9$); psychosis (schizophrenia or bipolar

affective disorder) ($n = 8$); substance dependence ($n = 5$); unspecified ($n = 4$).

Forty-nine per cent of SAP informants ($n = 28$) were immediate family members and 51% ($n = 29$) were either close friends, other relatives or key workers. Sixty-five per cent of informants ($n = 35$) had known the patient for more than 5 years, and 35% had known the patient from 1 to 5 years.

Rating of the participants' personalities

All participants were examined using DSM-IV versions of both the SAP and the SCID-II (v2.0). Three psychiatrists (PW, PM and PC) were trained in these instruments and administered them to each subject. One conducted the SAP with a named informant, and the other interviewed the patient with the SCID-II. Rater and interview were randomized for each participant. Each interviewer was blind to the results of the previous interview.

The level of agreement between instruments was assessed using Cohen's weighted kappa (Cohen, 1960), using the SCID-II as the 'gold-standard' (Mann, Raven et al., 1999).

Results

The SAP and the SCID-II identified a similar proportion of the population as personality disordered (37/57 or 65% of the sample using the SCID-II; 39/57 or 68% using the SAP) (see Table 1). Sixty-five per cent of people had two or more co-morbid SCID-II personality disorders, and 41% had two or more co-morbid SAP personality disorders. The overall level of agreement between the two instruments was low ($kappa = 0.3$). Using the SCID-II as the 'gold standard', the positive predictive value (PPV) of a SAP diagnosis was 74% with a sensitivity of 78%. The negative predictive value (NPV) of a SAP diagnosis was 55% with a specificity of 50%.

Predictive values were also calculated for populations with a lower prevalence of personality disorder. For a population with a personality disorder prevalence of 10% (community sample) the negative predictive value (NPV) was 96% and the positive predictive value (PPV) was 15%. For a population with a 28% prevalence of personality disorder (cf. the Bangalore outpatient sample (Mann, Raven et al., 1999)) the NPV would be 87% and PPV would be 37%.

Table 1. Cross tabulation of the agreement between SAP and SCID-II ratings of personality disorder (n = 57)

SAP	SCID-II		Total
	PD absent	PD present	
PD absent	10	8	18
PD present	10	29	39
Total	20	37	57

PD – Personality disorder

SAP– Standardized Assessment of Personality

SCID-II – Structured Clinical Interview for DSM-IV Personality Disorders (Version 2.0)

Table 2 shows the level of agreement between the SAP and the SCID-II for the individual personality disorders. There was wide variation in the levels of agreement from kappa = 0.4 (antisocial personality disorder) to kappa = -0.08 (narcissistic personality disorder). Clusters A and C showed a greater level of agreement between the two instruments (kappa = 0.4). However there was a much lower level of agreement for Cluster B personality disorders (kappa = 0.2).

Discussion

There are important limitations to this study that need to be considered. The number of subjects was small and therefore the kappa values for individual

personality disorders must be treated cautiously. In addition, the study was based on a highly selected population with a prevalence of personality disorder of 65%. Also the choice of a non-random sample of psychiatric patients limits the generalizability of our findings. However, the two measures were used independently and neither depended on just one rater.

In this study population we found that using the SCID-II as the 'gold-standard', the PPV and sensitivity of the SAP were acceptable (74% and 78% respectively), however, the NPV and specificity were lower (55% and 50% respectively). Sensitivity and specificity are two measures of the validity of a screening test and it is desirable for a screening test to be both highly sensitive and highly specific. The

Table 2. Number of diagnoses identified by SAP and SCID-II with each DSM-IV category of personality disorder and level of agreement

Diagnosis	Number of diagnoses		Kappa
	SAP	SCID-II	
Any personality disorder	39	37	0.3
Cluster A	19	20	0.4
Cluster B	21	27	0.2
Cluster C	22	21	0.4
Paranoid	14	16	0.4
Schizoid	7	6	0.2
Schizotypal	2	2	-0.04
Antisocial	9	17	0.4
Borderline	10	17	0.1
Histrionic	3	2	-0.04
Narcissistic	5	4	-0.08
Dependent	8	4	0.1
Avoidant	15	11	0.1
Obsessive	9	14	0.1

SAP– Standardized Assessment of Personality

SCID – II – Structured Clinical Interview for DSM-IV Personality Disorders (Version 2.0)

high sensitivity of the SAP compared to the SCID-II indicates a low false negative rate, but the lower specificity of the SAP means an accompanying high false positive rate. In this study, 50% of SAP cases were false positives on further testing with the SCID-II. This therefore suggests that the SAP was not an efficient means of identifying personality disorder in this population of psychiatric patients.

Sensitivity and specificity of a screen are independent of the prevalence of a disorder. However, the predictive value of a test is also of importance to clinicians and researchers and this is sensitive to changes in population prevalence. If the prevalence of personality disorder is 10% (the estimated prevalence of personality disorder in the community) (Samuels, Eaton, Bienvenu, Brown, Costa and Nestadt, 2002), the negative predictive value of the SAP would be 96% but the positive predictive value would be 15%. In Mann et al.'s (1999) study, comparing the SAP against the IPDE in a population with a prevalence of personality disorder of 28%, the negative and positive predictive values of a SAP diagnosis versus an IPDE diagnosis were 97% and 47% respectively. The predictive values obtained in our study using a prevalence of personality disorder of 28%, are similar (NPV 87%, PPV 37%). These figures suggest that the SAP may be a potentially useful screening instrument for the SCID II (as judged by the high NPV) for use in populations with a low prevalence of personality disorder, but less useful in populations with a higher prevalence of personality disorder (where the NPV falls). This finding lends some support to a two-stage process for identifying personality disorders in populations with a low prevalence of personality disorder. This, however, needs to be confirmed in a representative community sample.

This study has practical implications for both clinicians and researchers. As might be anticipated, a discrepancy was found between informant and self-report assessments of personality disorder. Although both assessments examine for the presence of DSM-IV personality disorder criteria, we find ourselves agreeing with other researchers who have concluded that the accounts obtained from patients and informants present different viewpoints: respectively one more experiential and the other more observational (Modestin and Puhon, 2000; Zimmerman, 1994). Good agreement between the two accounts is

therefore unlikely for the personality disorder categories as some depend on the former type of information, others the latter. Further research is needed to explore this interesting distinction. Clinicians and researchers should also bear this in mind when selecting a personality disorder assessment tool. One solution might be to use both approaches until further research is forthcoming (Zimmerman, 1994). The current study has shown that SAP is not an efficient screen for a SCID-II examination in a population with high prevalence of personality disorder. It may, however, provide a different perspective.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington DC: American Psychiatric Association, 1994.
- Bronner RK, King VL, Kidorf M, Schmidt CW, Jr., Bigelow GE. Psychiatric and substance use comorbidity among treatment-seeking opioid abusers. *Archives of General Psychiatry* 1997; 54: 71–80.
- Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; 20: 37–46.
- Dreessen L, Hildebrand M, Arntz A. Patient-informant concordance on the Structured Clinical Interview for DSM-III-R personality disorders (SCID-II). *Journal of Personality Disorders* 1998; 12: 149–61.
- First MB, Gibbon M, Spitzer RL, Williams JBW, Benjamin LS. Structured Clinical Interview for DSM-IV Axis II Personality Disorders. Washington: American Psychiatric Press, 1997.
- Harris EC, Barraclough B. Suicide as an outcome for mental disorders. A meta-analysis. *British Journal of Psychiatry* 1997; 170: 205–28.
- Hodgins S, Mednick SA, Brennan PA, Schulsinger F, Engberg M. Mental disorder and crime. Evidence from a Danish birth cohort. *Archives of General Psychiatry* 1996; 53: 489–96.
- Loranger AW, Sartorius N, Andreoli A, Berger P, Buchheim P, Channabasavanna SM, Coid B, Dahl A, Diekstra RF, Ferguson B. The International Personality Disorder Examination. The World Health Organization/Alcohol, Drug Abuse, and Mental Health Administration international pilot study of personality disorders. *Archives of General Psychiatry* 1994; 51: 215–24.
- Maffei C, Fossati A, Agostoni I, Barraco A, Bagnato M, Deborah D, Namia C, Novella L, Petrachi M. Interrater reliability and internal consistency of the structured clinical interview for DSM-IV axis II personality disorders (SCID-II), version 2.0. *Journal of Personality Disorders* 1997; 11: 279–84.

- Mann AH, Jenkins R, Cutting JC, Cowen PJ. The development and use of a standardized assessment of abnormal personality. *Psychol Med* 1981; 11: 839–47.
- Mann AH, Raven P, Pilgrim J, Khanna S, Velayudham A, Suresh KP, Channabasavanna SM, Janca A, Sartorius N. An assessment of the Standardized Assessment of Personality as a screening instrument for the International Personality Disorder Examination: a comparison of informant and patient assessment for personality disorder. *Psychol Med* 1999; 29: 985–89.
- McKeon J, Roa B, Mann A. Life events and personality traits in obsessive-compulsive neurosis. *Br J Psychiatry* 1984; 144: 185–9.
- Mellsop G, Varghese F, Joshua S, Hicks A. The reliability of axis II of DSM-III. *Am J Psychiatry* 1982; 139: 1360–1.
- Modestin J, Puhan A. Comparison of assessment of personality disorder by patients and informants. *Psychopathology* 2000; 33: 265–70.
- Mulder RT. Personality pathology and treatment outcome in major depression: a review. *American Journal of Psychiatry* 2002; 159: 359–71.
- Ormel J, Oldehinkel T, Brilman E, Brink WV. Outcome of Depression and Anxiety in Primary Care: A Three-Wave 3 1/2-Year Study of Psychopathology and Disability. *Arch Gen Psychiatry* 1993; 50: 759–66.
- Perry JC. Problems and considerations in the valid assessment of personality disorders. *Am J Psychiatry* 1992; 149: 1645–53.
- Pilgrim J, Mellers JD, Boothby H, Mann AH. Inter-rater and temporal reliability of the Standardized Assessment of Personality and the influence of informant characteristics. *Psychol Med* 1993; 23: 779–86.
- Poling J, Rounsaville BJ, Ball S, Tennen H, Kranzler HR, Triffleman E. Rates of personality disorders in substance abusers: a comparison between DSM-III-R and DSM-IV. *Journal of Personality Disorders* 1999; 13: 375–84.
- Rendu A, Moran P, Patel A, Knapp M, Mann A. Economic impact of personality disorders in UK primary care attenders. *British Journal of Psychiatry* 2002; 181: 62–6.
- Samuels J, Eaton WW, Bienvenu OJ, Brown C, Costa PTJ, Nestadt G. Prevalence and correlates of personality disorders in a community sample. *Br J Psychiatry* 2002; 180: 536–42.
- Tyrer P. Are personality disorders well classified in DSM-IV? In Livesley WJ (ed.) *The DSM-IV Personality Disorders*. New York: Guilford Press, 1995
- Tyrer P, Cicchetti DV, Casey PR, Fitzpatrick K, Oliver R, Balter A, Giller E, Harkness L. Cross-national reliability study of a schedule for assessing personality disorders. *Journal of Nervous and Mental Disease* 1984; 172: 718–21.
- Zanarini MC, Frankenburg FR, Chauncey DL, Gunderson JG. The diagnostic Interview for Personality Disorders: interrater and test-retest reliability. *Comprehensive Psychiatry* 1987; 28: 467–80.
- Zimmerman M. Diagnosing personality disorders. A review of issues and research methods. *Arch Gen Psychiatry* 1994; 51: 225–45.

*Correspondence: Dr Paul Walters, Health Services Research Department, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK.
Telephone (+44)(0)20 7848 0568.
Fax (+44)(0)20 7848 0333.
Email spoapaw@iop.kcl.ac.uk.*