Journal of Personality Disorders, 14(2), 162-170, 2000

PERSONALITY DISORDERS IN A DUTCH FORENSIC PSYCHIATRIC SAMPLE: CONVERGENCE OF INTERVIEW AND SELF-REPORT MEASURES

Corine de Ruiter, PhD, and Peter G.J. Greeven, PhD

Convergence of PDQ-R- and SIDP-R-derived personality disorder diagnoses was studied in a sample of 85 forensic psychiatric patients. For categorical diagnoses, the mean kappa was .34, but on a dimensional level convergence was somewhat higher. Paranoid, antisocial and borderline personality disorders had prevalence rates around 40%; the other personality disorders occurred with much lower frequency. The PDQ-R yielded more diagnoses, except for antisocial, histrionic, narcissistic, and sadistic personality disorder. Because the latter disorders are among the most prevalent in forensic settings, and because they have important risk and treatment implications, the PDQ-R is not suitable as a screening device in forensic populations. Semistructured interviews that make use of collateral information are recommended for diagnosing personality disorders in forensic subjects.

In the absence of a gold standard, the most valid assessment method for Axis II disorders remains an issue for debate among researchers and clinicians (Gabbard, 1997; Westen, 1997). Semistructured interviews and self-report inventories are the most frequently used assessment methods. Of these, interviews are generally viewed as the most valid, while self-report inventories are considered to be possibly adequate as screening devices (Hyler, Skodol, Oldham, Kellman, & Doidge, 1992; Hunt & Andrews, 1992). However, Westen (1997) has recently criticized the use of interview techniques based on direct questions to the subject, because a lack of insight and defensive responding are characterictic of some personality disorders, thereby limiting the validity of Axis II interviews.

The issues of defensiveness, social desirability, and deception are especially relevant in forensic populations, where these traits are ubiquitous. A

From the Department of Psychology, University of Amsterdam, Utrecht, The Netherlands, and the Department of Research and Public Relations, Dr. Henri van der Hoeven Kliniek (C.R.); the Ministry of Justice, The Hague, The Netherlands (P.G.).

Address correspondence to Corine de Ruiter, PhD, Department of Research and Public Relations, Dr. Henri van der Hoeven Kliniek, P.O. Box 174, 3500 AD UTRECHT, The Netherlands; E-mail: hoklires@wxs.nl.

number of authors have advised against the use of self-report inventories with forensic subjects, unless they have robust measures of distortion (Hare, 1991; Gacono & Meloy, 1994). Ideally, ratings based on an interview with a forensic subject would need to be corroborated by information obtained from informants (e.g., significant others) and/or file information. Although Westen (1997) argued that diagnoses based on the interview methods assessing Axis II disorders, such as the Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II; Spitzer, Williams, Gibbon, & First, 1990), the International Personality Disorder Examination (IPDE; Loranger, 1993), and the Structured Interview for DSM-IV Personality Disorders (SIDP-IV; Pfohl, Blum, & Zimmerman, 1995) are based exclusively on information obtained through direct questioning of the subject, we note that the manual for the SIDP-IV explicitly advises the assessor also to use informant and/or file information to arrive at a final diagnosis.

The current study was designed to examine the relationship between the diagnoses made on the basis of the Structured Interview for DSM-III-R Personality Disorders (SIDP-R; Pfohl, Blum, Zimmerman, & Stangl, 1989) and those made on the basis of a self-report inventory, the Personality Diagnostic Questionnaire-Revised (PDQ-R; Hyler & Rieder, 1987) in a sample of forensic psychiatric patients. On the basis of earlier findings (Hunt & Andrews, 1992; Trull & Larson, 1994), we expected the PDQ-R to overdiagnose personality disorders when compared to the interview method. However, given the tendency of many forensic subjects to be reluctant in admitting pathology, we also expected the PDQ-R to be less sensitive than the interview in detecting Cluster B disorders, such as antisocial and narcissistic personality disorders.

METHOD

Subjects. Eighty-five patients who were involuntarily committed to a forensic psychiatric hospital under the Dutch TBS-order, a judicial measure which can be translated as "disposal to be treated on behalf of the state," participated in a 2-year prospective treatment outcome study. Exclusionary criteria were schizophrenia or other psychotic disorder and mental retardation. All patients had committed a serious and violent criminal act, for which the court had not held them entirely responsible due to their mental disorder (either an Axis I and/or Axis II disorder). Thirty-two percent of the sample had been convicted of (attempted) homicide, 17% arson, 18% rape, and the others included extortion, indecent assault, aggravated assault, threat and/or pedophilic offenses. Ninety-three percent of the sample was male. Mean age of the sample was 26 (SD = 6.6; range = 17–47).

Instruments. At the start of the study, all subjects completed a battery of self-report questionnaires, including the Personality Diagnostic Questionnaire-Revised (PDQ-R; Hyler & Rieder, 1987). The PDQ-R consists of 133 true/false items that are consistent with DSM-III-R Axis II criteria. Each subject was also administered the Dutch version of the Structured Interview for DSM-III-R Personality Disorders (SIDP-R; Pfohl, Blum, Zimmerman, & Stangl, 1989; van den Brink & de Jong, 1992). The SIDP-R consists of 160 questions organized into 17 areas such as interpersonal functioning, emotional expression, and perception of threat. The interviewer did not rate items while interviewing but was expected to take detailed

notes. The interviewer also examined available chart materials. In the current study these consisted of all the available judicial and psychological/psychiatric reports on the offender-patient.

Two interviewers, who were unaware of the PDQ-R scores of the patients, conducted the SIDP-R interviews. Interviewers had received extensive training in the administration and scoring of the SIDP-R by the authors of the Dutch translation. The interviews were taped and final diagnoses were arrived at via consensus between the two interviewers.

RESULTS

The prevalence of categorical personality disorders and dimensional personality disorder traits (i.e., the number of Axis II criteria met) was determined on the basis of the SIDP-R and the PDQ-R. Categorical diagnoses are determined by the required number of Axis II criteria for each specific disorder, as specified in the DSM-III-R (American Psychiatric Association, 1987). Concurrent validity between SIDP-R and PDQ-R diagnoses was determined for categorical diagnoses and for the number of Axis II criteria met for each disorder.

Table 1 presents frequencies and percentages of categorical DSM-III-R personality disorder diagnoses. The McNemar-test was used to examine whether the categorical diagnoses derived from the SIDP-R and the PDQ-R differed significantly.

As shown in Table 1, only three disorders showed a significant difference between the SIDP-R and the PDQ-R results: paranoid, schizotypal, and self-defeating personality disorders. In all three cases, the PDQ-R revealed a greater number of diagnoses. Also shown in Table 1, antisocial and borderline personality disorders are the most prevalent disorders in this forensic psychiatric sample, with percentages of 52 and 35, respectively, as measured by the SIDP-R.

Table 1 also demonstrates that the diagnostic overlap between the two instruments is actually quite low. For instance, for schizoid personality disorder, the PDQ-R yielded 13 diagnoses and the SIDP-R 10, but only 4 patients received a diagnosis of schizoid personality disorder on the basis of both instruments. This modest diagnostic overlap is reflected in the outcome of the subsequent analyses. Diagnostic agreement between the SIDP-R and the PDQ-R on categorical diagnoses was determined by means of three coefficients: percentage observed agreement, kappa, and Yule's Y (Spitznagel & Helzer, 1985). Table 2 presents the results of these analyses.

A value of 70% to 80% observed agreement is generally considered acceptable (Zegers, 1991). When we used this criterion, only paranoid and antisocial personality disorders did not meet this agreement. However, the kappa value, which corrects for chance agreement, is generally much lower. Only four personality disorders (borderline [kappa = .54], histrionic [kappa = .49], self-defeating [kappa = .45], and dependent [kappa = .40]) meet the criterion of acceptable agreement (kappa between .40 and .74; Shrout, Spitzer, & Fleiss, 1987). The kappa coefficient is markedly attenuated by low prevalence rates, for which coefficient Yule's Y provides a correction. On the basis

TABLE 1. Frequency (and Percentage) of DSM-III-R Personality Disorder Categories as Measured by PDQ-R and SIDP-R (N = 85)

	SIDP-R	PDQ-R	Diagnosis on SIDP-R and PDQ-R	McNemar test
Personality disorder	N (%)	N (%)	N (%)	$p <$
Paranoid	21 (24.7)	39 (45.9)	15 (17.6)	.01
Schizoid	10 (11.8)	13 (15.3)	4 (4.7)	NS
Schizotypal	12 (14.1)	21 (24.7)	7 (8.2)	.05
Histrionic	16 (18.8)	15 (17.6)	9 (10.1)	NS
Antisocial	44 (51.8)	39 (45.9)	27 (31.8)	NS
Narcissistic	20 (23.5)	17 (20.0)	6 (7.1)	NS
Borderline	30 (35.3)	37 (43.5)	24 (28.2)	NS
Obsessive-compulsive	13 (15.3)	14 (16.5)	6 (7.1)	NS
Dependent	14 (16.5)	14 (16.5)	7 (8.2)	NS
Avoidant	7 (8.2)	20 (23.5)	6 (7.1)	NS
Passive-aggressive	16 (18.8)	21 (24.7)	7 (8.2)	NS
Self-defeating	12 (14.1)	20 (23.5)	9 (10.1)	.05
Sadistic	14 (16.5)	11 (12.9)	5 (5.9)	NS
Average number of disorders	2.7	3.3		
One disorder minimum	68 (80.0)	71 (83.5)	NS	
No personality disorder	17 (20.0)	14 (16.5)	NS	

of Yule's Y, 9 of the 13 disorders show acceptable to good agreement. The exceptions are narcissistic (Y = .18), passive-aggressive (Y = .28), antisocial (Y = .33) and paranoid (Y = .35) personality disorders. Finally, specificity and sensitivity for the PDQ-R were determined. If we assume that the SIDP-R provides the most valid diagnosis (Zimmerman & Coryell, 1990; Duijsens, Bruinsma, Jansen, Eurelings-Bontekoe, & Diekstra, 1996), the sensitivity of the PDQ-R gives the frequency with which the PDQ-R gives a positive diagnosis, in case the SIDP-R also gives a positive diagnosis. The specificity of the PDQ-R is the frequency with which it gives no diagnosis when the SIDP-R also gives no diagnosis. The sensitivity of the PDQ-R for the diagnosis of any personality disorder is good (.87). Sensitivity is low for schizoid, narcissistic, dependent, obsessive-compulsive, passive-aggressive, and sadistic personality disorders (.50). For 11 of the 13 personality disorders, specificity is .80. Specificity of the PDQ-R for the absence of a personality disorder is much lower (.30). This is caused by the low base rate.

The mean numbers of Axis II criteria met (dimensional scores) on the PDQ-R and the SIDP-R are 2.9 (SD = 1.2) and 2.4 (SD = 1.2), respectively. The mean number of criteria met is significantly greater for the PDQ-R than for the SIDP-R (t test, $p < .001$). For 8 of the 13 disorders, the mean number

TABLE 2. Agreement Between SIDP-R and PDQ-R for Categorical Diagnoses ($N = 85$)

	P_o (%)	Kappa	Yule's Y	Sensitivity	Specificity
Paranoid	65	.27	.35	.72	.63
Schizoid	82	.25	.46	.40	.88
Schizotypal	78	.28	.42	.58	.81
Histrionic	85	.49	.57	.56	.91
Antisocial	66	.32	.33	.62	.71
Narcissistic	71	.14	.18	.30	.83
Borderline	78	.54	.57	.80	.88
Obsessive-compulsive	82	.34	.45	.46	.89
Dependent	84	.40	.51	.50	.90
Avoidant	82	.37	.68	.86	.82
Passive-aggressive	73	.21	.28	.44	.80
Self-defeating	71	.45	.61	.75	.85
Sadistic	82	.30	.42	.36	.92
Average	77	.34	.45	.57	.82
One personality disorder	75	.25	.30	.87	.30

Note. P_o = percentage observed agreement.

of criteria met is significantly greater for the PDQ-R. For narcissistic, obsessive-compulsive, and passive-aggressive personality disorders, the PDQ-R also yields a greater number of criteria met, but the differences are not significant. For antisocial and sadistic personality disorders, the result is reversed, but not significant.

Table 3 presents Pearson product–moment correlation coefficients for the number of Axis II criteria met for the PDQ-R and SIDP-R. For all disorders, the dimensional scores on the SIDP-R and PDQ-R were significantly correlated (all $ps < .01$). Antisocial ($r = .60$), selfdefeating ($r = .55$), and borderline ($r = .50$) personality disorder showed the highest correlations. With the exception of passive-aggressive, dependent, and paranoid personality disorders, all other disorders showed the largest correlation with the same disorder diagnosed using the other instrument. On a dimensional level, there appeared to be more diagnostic agreement between the two instruments than on a categorical level.

DISCUSSION

Eighty percent of our forensic psychiatric sample fulfilled diagnostic criteria for at least one personality disorder with paranoid, antisocial, and borderline personality disorders the most prevalent. These findings are in line with a number of recent studies that showed prevalence rates for personality disorders to be between 60% and 80% in various forensic psychiatric samples from England and Sweden (Blackburn, Crellin, Morgan & Tulloch, 1990;

TABLE 3. Significant Pearson Product-Moment Correlations Between SIDP-R and PDQ-R for Number of Personality Disorder Criteria Met ($N = 85$)

	PDQ-R												
	SZD	STY	PAR	NAR	ASP	BPD	HST	AVD	DEP	OBS	PAG	SEL	SAD
SZD	**.44**	.28	.27	.26				.32				.27	
STY		**.45**	.31			.29		.39			.42	.43	
PAR		.28	**.32**		.31	.29		.33			.29	.41	
NAR			.26	**.30**							.30		
ASP		.28	.32		**.60**	.29					.34	.36	.36
BPD		.34	.34		.29	**.50**	.30		.30		.35	.50	
HST							**.43**						
AV								**.49**	.38			.33	
DEP								.33	**.44**			.27	-.25
OBS								.34	.45	**.26**			
PAG			.35	.30	.27	.37		.33	.37		**.39**	.42	
SEL									.36			**.55**	
SAD		.32			.38						.27		**.38**

Note. SZD = schizoid; STY = schizotypal; PAR = paranoid; NAR = narcissistic; ASP = antisocial; BPD = borderline; HST = histrionic; AVD = avoidant; DEP = dependent; OBS = obsessive-compulsive; PAG = passive-aggressive; SEL = self-defeating; SAD = sadistic. Correlations for the same disorder, diagnosed with each of the two instruments, are in bold.

Dietz, 1992; Coid, 1992; Dolan & Mitchell, 1994; Kullgren, Grann & Holmberg, 1996). Antisocial and borderline personality disorders were shown to be the most prevalent of the Axis II disorders (Dolan & Coid, 1993).

Diagnostic agreement between two instruments for the assessment of DSM-III-R personality disorders, one self-report questionnaire, the PDQ-R, and one semistructured interview, the SIDP-R, was studied on a categorical and a dimensional level. Hyler, Skodol, Oldham, Kellman, and Doidge (1992) advised investigators to compare their findings with those of colleagues who use the same instruments. We will thus compare our findings with those of Hyler et al. (1989), Zimmerman and Coryell (1990), and Trull and Larson (1994). On a categorical level, diagnostic agreement between the SIDP-R and the PDQ-R was generally higher in our study than in the other three studies. The mean kappa value in our study was .34, whereas the other authors found mean kappa values of .15 (Hyler et al., 1989), .14 (Zimmerman & Coryell, 1990) and .19 (Trull and Larson, 1994). A possible explanation for this divergence is the fact that some personality disorders had very low base rates in the three studies, and kappa values are attenuated by extreme base rates. Compared to the categorical approach, the dimensional approach led to somewhat higher diagnostic agreement between the two instruments, a finding that is in line with those of others (van den Brink, 1989; Zimmerman & Coryell, 1990). In general, however, convergent validity of the two instruments is not very high, a finding that several authors have attributed to the deficient construct validity of a number of the personality disorders (Perry, 1992; Zimmerman, 1994; Zimmerman & Coryell, 1990). The use of empirical/psychometric methodologies to improve the construct and discriminant validity of the personality disorder criteria sets is an important goal for future studies (cf. Blais & Norman, 1997).

The PDQ-R tended to overdiagnose personality disorders (cf. Hunt & Andrews, 1992) compared to the SIDP-R. However, this was not the case for histrionic, antisocial, narcissistic, and sadistic personality disorders where the SIDP-R yielded a greater number of diagnoses. This finding is in line with our hypothesis that Cluster B disorders are more difficult to detect by means of self-report instruments because of the lack of self-insight and defensiveness that are inherent to these disorders. Zimmerman and Coryell (1990) also found that histrionic and antisocial personality disorders were more often diagnosed on the basis of the SIDP-R than the PDQ-R.

A number of authors have suggested employing the PDQ-R as a screening device, to identify possible "true cases" of personality disorder or to rule out a personality disorder diagnosis (Hyler et al., 1992; Trull & Larson, 1994). We would like to caution against the use of the PDQ-R in this sense, particularly in forensic settings. Although the PDQ's average sensitivity for a personality disorder was .87, its specificity for a number of specific disorders/traits that are especially relevant in forensic samples in terms of risk assessment and treatment planning, such as narcissistic, sadistic, and schizoid personality disorders, is unacceptably low. This means that if one were to use the PDQ-R as a screening device, there would be a high probability of not recognizing these disorders. The use of a semi-structured interview in combination with collateral information is indispensable for the diagnosis of personality disorders in forensic settings.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Blackburn, R., Crellin, M. C., Morgan, E. M., & Tulloch, R. M. B. (1990). Prevalence of personality disorders in a special hospital population. *Journal of Forensic Psychiatry, 1*, 43–52.

Blais, M. A., & Norman, D. K. (1997). A psychometric evaluation of the DSM-IV personality disorder criteria. *Journal of Personality Disorders, 11*, 168–176.

Coid, J. W. (1992). DSM-III diagnosis in criminal psychopaths: A way forward. *Criminal Behavior and Mental Health, 2*, 78–94.

Dietz, P. E. (1992). Mentally disordered offenders: Patterns in the relationship between mental disorder and crime. *Psychiatric Clinics of North America, 15*, 539–551.

Dolan, B. M., & Coid, J. W. (1993). *Psychopathic and antisocial personality disorders: Treatment and research issues.* London: Gaskell.

Dolan, B. M., & Mitchell, E. (1994). Personality disorder and psychological disturbance of female prisoners: A comparison with women referred for NHS treatment of personality disorder. *Criminal Behaviour and Mental Health, 4*, 130–143.

Duijsens, I. J., Bruinsma, M., Jansen, S. J. T., Eurelings-Bontekoe, E. H. M., & Diekstra, R. F. W. (1996). Agreement between self-report and semi-structured interviewing in the assessment of personality disorders. *Personality and Individual Differences, 21*, 261–270.

Gabbard, G. O. (1997). Finding the "Person" in personality disorders. *American Journal of Psychiatry, 154*, 891–893.

Gacono, C. B., & Meloy, J. R. (1994). *The Rorschach assessment of aggressive and psychopathic personalities.* Hillsdale, NJ: Erlbaum.

Hare, R. D. (1991). *Psychopathy checklist revised.* Toronto: Multi-Health Systems.

Hunt, C., & Andrews, G. (1992). Measuring personality disorders: The use of self-report questionnaires. *Journal of Personality Disorders, 6*, 125–133.

Hyler, S. E., & Rieder, R. O. (1987). *Personality Disorder Questionnaire-Revised.* New York: New York State Psychiatric Institute.

Hyler, S. E., Rieder, R. O., Williams, J. B. W., Spitzer, R. L., Lyons, M., & Hendler, J. (1989). A comparison of clinical and self-report diagnoses of DSM-III personality disorders in 552 patients. *Comprehensive Psychiatry, 30*, 170–178.

Hyler, S. E., Skodol, A. E., Oldham, J. M., Kellman, H. D., & Doidge, N. (1992). Validity of the Personality Diagnostic Questionnaire-Revised: A replication in an outpatient sample. *Comprehensive Psychiatry, 33*, 73–77.

Kullgren, G., Grann, M., & Holmberg, G. (1996). The Swedish forensic concept of severe mental disorder as related to personality disorders. *International Journal of Law and Psychiatry, 19*, 191–200.

Loranger, A. W. (1993). *IPDE user manual.* Geneve: World Health Organization.

Perry, J. C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry, 149*, 1645–1653.

Pfohl, B., Blum, N., & Zimmerman, M. (1995). *Structured Interview for DSM-III-IV Personality* (SIDP-IV). Iowa City: Department of Psychiatry, University of Iowa.

Pfohl, B., Blum, N., Zimmerman, M., & Stangl, D. (1989). *Structured Interview for DSM-III-R Personality* (SIDP-R). Iowa City: Department of Psychiatry, University of Iowa.

Shrout, P. E., Spitzer, R. L., & Fleiss, J. L. (1987). Quantification in psychiatric diagnosis revisited. *Archives of General Psychiatry, 44*, 172–177.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. (1990). *Structured Clinical Interview for the DSM-III-R Personality Disorders (SCID-II).* Washington, DC: American Psychiatric Press.

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

Trull, T. J., & Larson, S. L. (1994). External validity of two personality disorder inventories. *Journal of Personality Disorders, 8*, 96–103.

van den Brink, W. (1989). *Meting van DSM-III persoonlijkheidspathologie: Betrouwbaarheid en validiteit van de*

SIDP-R en as II van de DSM-III. [Assessment of DSM-III personality pathology: Reliability and validity of the SIDP-R and axis II of the DSM-III]. (Dissertation Rijksuniversiteit Groningen).

van den Brink, W., & Jong, C. A. J. de (1992). *The Dutch version of the Structured Interview for DSM-III-R Personality Disorders* (SIDP-R). Eindhoven: Instituut voor Verslavings zorg Brabant.

Westen, D. (1997). Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. *American Journal of Psychiatry, 154,* 895–903.

Zegers, F. E. (1991). Coefficients for interrater agreement. *Applied Psychological Measurement, 15,* 321–333.

Zimmerman, M. (1994). Diagnosing personality disorders: A review of issues and research methods. *Archives of General Psychiatry, 51,* 225–245.

Zimmerman, M., & Coryell, W. H. (1990). Diagnosing personality disorders in the community. *Archives of General Psychiatry, 47,* 527–531.