M. Gárriz[1]
F. Gutiérrez[1,2]

Personality disorder screenings: a meta-analysis

[1] Psychology Department
Institut Clínic de Neurociències
Hospital Clínic Barcelona
Barcelona (Spain)

[2] IDIBAPS
Barcelona (Spain)

Personality disorders (PD) are highly prevalent and impairing stable patterns of maladaptive behavior that are associated to high health care costs. Although PD detection in clinical settings is a priority issue, it is still unknown which are the most reliable and valid screening instruments. For this purpose, 26 studies examining the diagnostic ability of 19 different screening tools that included structured interviews as the gold standard were meta-analyzed. The total median for the kappa was 0.40 (range 0.14 to 0.86), indicating poor to moderate agreement. Brief instruments created *ad hoc* showed the best predictive ability (Md_{kappa}=0.56) and are, because of their easy application, the first-line screening tools.

Key words:
Personality disorders. Personality assessment. Diagnosis.

Actas Esp Psiquiatr 2009;37(3):148-152

Cribado de trastornos de la personalidad: un metaanálisis

Los trastornos de la personalidad son patrones estables de comportamiento desadaptativo altamente prevalentes e incapacitantes, que se asocian a un elevado coste sanitario. Pese a ser su detección en la clínica una cuestión prioritaria, ignoramos qué instrumentos de cribaje son fiables y válidos. Este estudio meta analiza veintiséis estudios que examinaron la capacidad diagnóstica de 19 diferentes instrumentos de cribaje y utilizaron una entrevista estructurada como estándar diagnóstico. La mediana total de kappa fue de 0,40 (rango 0,14 a 0,86), indicando un acuerdo de pobre a moderado. Los instrumentos breves creados *ad hoc* mostraron la mejor capacidad predictiva (Md_{kappa}=0,56) y son, por su rapidez de administración, los instrumentos de primera elección.

Palabras clave:
Trastornos de la personalidad. Evaluación personalidad. Diagnóstico.

Correspondence:
F. Gutiérrez
Servei de Psicología
Institut Clínic Neurociències
Hospital Clínic Barcelona
Villarroel 170
08036 Barcelona (Spain)
E-mail: fguti@clinic.ub.es

INTRODUCTION

Human beings differ from one another in many stable psychological characteristics, such as anxiety, impulsivity, affiliation, dominance or persistence. These characteristics are partially heritable, are present since adolescence or early adult thood, and largely determine the biography of the individual.[1] When these traits are so extreme or inflexible as to cause significant disfress or impairment in some domine of the subject's functioning (work, family, social), a personality disorder (PD) is diagnosed.

Accurate detection of PD in clinical settings is essential for several reasons. Prevalence of maladapted personalities is very high in comparison with that of other mental disorders: 9%-13% in the general population and 24% in primary health care.[2,3] Functional impairment and suffering associated to these disorders are comparable to those caused by other severe and chronic disorders such as major depression.[4] They increase the risk of occurrence, and worsen the evolution, of affective, anxious, psychotic, eating and substance abuse disorders.[5] Finally, the presence of personality disorder also affects prognosis and adherence to treatment in medical diseases.[6] Thus, it is not surprising that PDs are associated to greater use and costs of health care services than the population mean,[7] that they have a greater likelihood of being perceived as difficult by their doctors, and that they are a source of *burnout*.[8]

Unfortunately, our capacity to detect PD does not correspond with its clinical and social relevance. The unstructured clinical interview, which is the most widespread method of diagnosis, lacks reliability. The general practitioner finds it difficult to discriminate between patients with PD and other patients who have poor compliance or unpleasant attitude, and refers the former to the specialist less often.[7] Even the specialists obtain unacceptable agreement indexes when they do not use structured interviews.[9] On the other hand, these interviews (SCID-II, SIDP, PDE, IPDE), that have become the gold standard for diagnosis, require previous training and 1 to 4 hours per patient, which makes them unsuitable in most health care centers.

In the 1980's, a search was boque for PD screening instruments which would reduce detection cost and which therefore results in more accurate referrals, more adequate treatments, reduction in health care costs and less burnout of the professionals. A wide range of personality interviews and questionnaires were examined for this purpose, with unequal results. However, these studies have still not been reviewed and we lack indications for the choice of most adequate screening instrument. The present study has performed a meta-analysis of the published literature on the diagnostic utility of PD screenings.

METHOD

A bibliographic search, up-dated until December 2007, was made, using the PsycINFO, Medline and Academic Search Premier databases with the strategy ([personality disorder*] or [axis II] or [personality pathology] and [screen*]), and the references of the selected articles were reviewed. All the works were induded that 1) used any screening instrument of PD, 2) confirmed the presence/absence of the diagnosis by structured interview based on the DSM classification,[10] and 3) provided the information necessary for the calculation of the following diagnostic indexes: sensitivity (percentage of PD with positive result in the screening), specificity (percentage of non-PD with negative result in the screening), percentage of accurate results (PD and non-PD accurately classified), and Cohen's kappa, that take the agreements expected by chance into account. Kappa statistic was used as the principal outcome variable given that depends both on the sample prevalence of PD and on the predictive capacity of the screening, and that sensitivity and specificity can only be interprete together.

In the studies that used more than one screening instrument or more than one sample, independent analyses were conducted and for those that analyzed different cut offs the one with the highest Kappa value was used. Mean (M) and median (Md) were calculated, both weighted and unweighted by sample size, of the four diagnostic indexes for the total of the studies and for six screening instrument subgroups: *a)* brief instruments (<30 items) created *ad hoc* (n = 10), *b)* PDQ (Personality Diagnostic Questionnaire) (n = 6), *c)* TCI (Temperament and Character Inventory) (n = 5), *d)* SAP interview (Standardized Assessment of Personality) (n = 2), *e)* large instruments (>100 items) DSM-based (n = 5), and *f)* MCMI (Millon Clinical Multiaxial Inventory) (n = 5). The MMPI (Minnesota Multiphasic Personality Inventory) and SCL-90R (Symptoms Checklist) only had a single study and were not included in the analysis by subgroups.

RESULTS

Twenty-six studies fulfilled the inclusion criteria. These studies examined the properties of 19 different screening instruments in 29 samples, with a total of 35 results and 5,432 subjects (Table 1). The mediane Kappa for the total of the studies was 0.40 (SD 0.20; range 0.14 to 0.86). In the analysis by subgroups, the brief instruments created *ad hoc* obtained the highest kappa values (Md = 0.56), followed by PDQ (Md = 0.42), TCI (Md = 0.36), SAP (Md = 0.34), large DSM-based instruments (Md = 0.29) and MCMI (Md = 0.26). This order remained unchanged when the means were calculated instead of the medians and only changed slightly when weighted by the number of subjects: the brief instruments increased their advantage (Md_p = 0.73), and PDQ, TCI and SAP inverted their respective positions, although within the same score range (Md_p = 0.33 to 0.40) (Table 1).

The remaining diagnostic indexes were less informative. The hit tate was high (Md = 0.76), but there was little difference between instrument groups (0.68 to 0.78). Sensitivity results were inverse to those of Kappa, suggesting large DSM-based instruments (Md = 0.96) and MCMI (Md = 0.87) as being of first choice. However, this high sensitivity was reached at the expense of the lowest specificity indexes (Md = 0.61 and 0.48), indicating that half of the non-cases are erroneously classified as cases.

DISCUSSION

The available screening instruments show a poor or moderate predictive ability of DSM diagnosis. As expected, the best indexes on agreements are obtained for the brief instruments developed with this purpose (IOWA, SAP-AS, IIP-PD). These instruments are also suitable in clinical settings due to their reduced number of items (from 11 to 28), they can be easily administered as part of the general interview, and thus they are of first choice if the objective is simply to detect the presence of PD. However, they have two disadvantages. The first one is that the most instruments studied, IOWA, an 11-item interview, and IIP-PD, a questionnaire on interpersonal problems, show as a whole moderate indexes, while the instruments with better indexes, SAP-AS, a modification of SAP, and Nurberg's DSM criteria selection, only have one study, suggesting the possibility of bias. The second disadvantage is that they do not provide information on the type of disorder, although this point may be irrelevant if the objective is referral to a specialist who will make a more decurate diagnosis.

The self-administered questionnaires PDQ and TCI are acceptable alternatives. The PDQ assesses the presence of all the DSM PD criteria, so that it makes it possible to determine both the presence and type of disorder. However, it has a moderate-to-low agreement with structured interviews. Similar results are obtained with the TCI (previously TPQ), based on Cloninger's personality model and with roots in the dimensional models of personality rather than in psychiatric classification. This instrument evaluates four temperamental dimensions that reflect the type of personality, and three scales character that determine the existence of maladaptation or disorder. The TCI is more specific than sensitive and thus is a

Table 1 **Diagnostic ability (Kappa agreement, hit rate, sensitivity and specificity) of 19 different screening instruments in the literature.**

Study	Sample		Screening			Diagnostic standard	Diagnostic indexes			
	n	% PD	Screening	No. items	I/Q[1]		Kappa	Hit rate	Sensitivity	Specificity
Brief Inst. created										
Langbehn 1999[11]	433	57.7	IOWA	11	I	SIDP-R	0.52	0.76	0.73	0.79
Langbehn 1999[11]	52	46.1	IOWA	11	I	SIDP-IV	0.59	0.79	0.95	0.64
Stern 2000[12]	90	15.5	IIP-PD	28	Q	SIDP-R	0.63	0.88	0.93	0.87
Nurnberg 2000[13]	1342	42.9	DSM Selec.	15	Q	PDE/SIDP-R/SCID-II	0.73	0.87	0.90	0.84
Trull 2001[14]	103	34.9	IOWA	11	I	SIDP-R	0.62	0.83	0.69	0.91
Moran 2003[15]	60	55.0	SAP-AS	8	I	SCID-II	0.80	0.90	0.94	0.85
Morse 2007[16]	70	84.0	IIP-PD	28	Q	SIDP-IV	0.17	0.60	0.58	0.73
Morse 2007[16]	81	44.0	IIP-PD	28	Q	SIDP-IV	0.22	0.61	0.67	0.56
Morse 2007[16]	70	84.0	IOWA	19	Q	SIDP-IV	0.27	0.76	0.80	0.55
Morse 2007[16]	81	44.0	IOWA	19	Q	SIDP-IV	0.21	0.59	0.78	0.44
Median (weighted)							0.56 (0.73)	0.78 (0.87)	0.79 (0.90)	0.76 (0.84)
PDQ										
Dubro 1988[17]	56	41.0	PDQ	163	Q	SIDP	0.51	0.75	0.87	0.68
Pfohl 1989[18]	45	35.5	PDQ	163	Q	SIDP	0.72	0.87	0.94	0.83
Zimmerman 1990[19]	697	13.4	PDQ	163	Q	SIDP	0.33	0.86	0.35	0.94
Hyler 1992[20]	59	30.5	PDQ-R	152	Q	SCID-II /PDE	0.32	0.61	1.00	0.43
Van Velzen 1999[21]	137	41.6	PDQ-R	133	Q	SCID-II	0.26	0.59	0.80	0.40
Davison 2001[22]	62	79.0	PDQ-4+	99	Q	SCID-II	0.50	0.79	0.77	0.84
Median (weighted)							0.42 (0.33)	0.77 (0.86)	0.84 (0.35)	0.76 (0.94)
TCI										
Starcevic 1995[23]	48	31.2	TPQ	100	Q	SCID-II	0.86	0.94	1.00	0.91
Gutiérrez 2002[24]	74	60.8	TCI	240	Q	SCID-II	0.56	0.78	0.77	0.79
Morse 2007[16]	70	84.0	TCI-SD	44	Q	SIDP-IV	0.17	0.60	0.58	0.73
Morse 2007[16]	81	44.0	TCI-SD	44	Q	SIDP-IV	0.14	0.59	0.33	0.80
Gutiérrez 2007[25]	205	29.7	TCI	240	Q	SCID-II	0.36	0.77	0.37	0.93
Median (weighted)							0.36 (0.36)	0.77 (0.77)	0.58 (0.37)	0.80 (0.91)
SAP										
Walters 2004[26]	57	64.9	SAP	–	I	SCID-II	0.28	0.68	0.78	0.50
Mann 1999[27]	90	28.8	SAP	–	I	IPDE (ICD)	0.40	0.68	0.47	0.97
Median (weighted)							0.34 (0.40)	0.68 (0.68)	0.63 (0.47)	0.74 (0.97)
DSM based large Inst.										
Ekselius 1994[28]	69	53.6	SCID-S	124	Q	SCID-II	0.61	0.81	0.87	0.75
Jacobsberg 1995[29]	260	23.8	SCID-S	124	Q	SCID-II	0.29	0.76	0.98	0.47
Duijsens 1996[30]	108	20.4	VKP	174	Q	IPDE	0.17	0.53	0.82	0.44
Lenzenweger 1997[31]	258	8.1	IPDE-S	250	Q	IPDE	0.20	0.64	1.00	0.61
Ottosson 1998[32]	138	65.9	DIP-Q	140	Q	DIP-I	0.61	0.81	0.83	0.76
Median (weighted)							0.29 (0.29)	0.76 (0.76)	0.87 (0.98)	0.61 (0.61)
MCMI										
Dubro 1988[17]	56	41.0	MCMI	175	Q	SIDP	0.40	0.68	0.96	0.48
Soldz 1993[33]	97	77.3	MCMI-II	175	Q	PDE	0.26	0.60	0.81	0.43
Marlowe 1997[34]	110	77.0	MCMI-II	175	Q	SCID-II	0.19	0.78	0.96	0.20
Fdz.-Montalvo 2006[35]	50	22.0	MCMI-II	175	Q	IPDE	0.17	0.58	0.72	0.53
Fdz.-Montalvo 2006[35]	55	7.2	MCMI-II	175	Q	IPDE	0.52	0.89	1.00	0.88
Median (weighted)							0.26 (0.26)	0.68 (0.68)	0.96 (0.96)	0.48 (0.43)
Others										
Starcevic 2000[36]	112	52.6	SCL-90R	90	Q	SCID-II	0.70	0.85	0.72	0.98
Dubro 1988[17]	56	41.0	MMPI	566	Q	SIDP	0.43	0.71	0.78	0.67
Total										
Median (weighted)							0.40 (0.50)	0.76 (0.81)	0.80 (0.82)	0.73 (0.84)

1 I=Interview, Q=Questionnaire

useful detector of non-cases. The self-report format of PDQ and TCI economizes professional time.

The characteristics of the remaining instruments reviewed make them inadequate for screening purposes. The SAP is a non-structured interview with no established cut off, which makes its standardization difficult. The agreement indexes of MCMI and of the large DSM-based instruments hardly differ from that expected by chance. Finally, MMPI and SCL-90 only have one published study each.

Although the results provide an initial guide for choosing the best screening instrument, some methodological clarifications are necessary. First, the predictive capacity of most of the screenings was studied in samples with high prevalence of PD (M = 42.9%). Thus, care should be taken when generalizing it, to the general population. Second, the purpose of some studies is not to maximize Kappa agreement but rather sensitivity, so Kappa may underestimate the true capacity of the instrument. Finally, it should be remembered that the diagnostic interviews used as standard also show low agreements with one another,[9] which decreases the diagnostic reliability of the screenings.

We can conclude that the brief interviews have better capacity to detect the presence of PD. However, this does not imply that they are the instruments of choice for other uses. Dimensional models of personality such as the TCI or the DAPP have obtained better empiric support than the DSM itself[37] and have been proposed as alternatives to it. The proposal of basing the diagnosis of PD not on the intensity of certain personality traits but rather on the degree of maladaptation they produce, that is, on biographical impairment measures, is also acquiring progressive support.

REFERENCES

1. Ozer DJ, Benet-Martinez V. Personality and the prediction of consequential outcomes. Ann Rev Psychol 2006;57:401-21.
2. Lenzenweger MF, Lane MC, Loranger AW, Kessler RC. DSM-IV personality disorders in the National Comorbidity Survey Replication. Biol Psychiatry 2007;62:553-64.
3. Moran P, Jenkins R, Tylee A, Blizard R, Mann A. The prevalence of personality disorder among UK primary care attenders. Acta Psychiatr Scand 2000;102:52-7.
4. Skodol AE, Gunderson JG, McGlashan TH, Dyck IR, Stout RL, Bender DS, et al. Functional impairment in patients with schizotypal, borderline, avoidant, or obsessive-compulsive personality disorder. Am J Psychiatry 2002;159:276-83.
5. Mulder RT. Personality pathology and treatment outcome in major depression: A review. Am J Psychiatry 2002;159:359-71.
6. Frayne SM, Halanych JH, Miller DR, Wang F, Lin H, Pogach L, et al. Disparities in diabetes care: impact of mental illness. Arch Intern Med 2005;165:2631-8.
7. Moran P, Rendu A, Jenkins R, Tylee A, Mann A. The impact of personality disorder in UK primary care: a 1-year follow-up of attenders. Psychol Med 2001;31:1447-54.
8. Schafer S, Nowlis DP. Personality disorders among difficult patients. Arch Fam Med 1998;7:126-9.
9. Zimmerman M. Diagnosing personality disorders. A review of issues and research methods. Arch Gen Psychiatry 1994;51:225-45.
10. APA. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. Washington DC: American Psychiatric Press, 2000.
11. Langbehn DR, Pfohl BM, Reynolds S, et al. The Iowa Personality Disorder Screen: Development and preliminary validation of a brief screening interview. J Pers Disord 1999;13:75-89.
12. Stern BL, Kim Y, Trull TJ, Scarpa A, Pilkonis P. Inventory of Interpersonal Problems - Personality Disorder scales: Operating characteristics and confirmatory factor analysis in nonclinical samples. J Pers Assess 2000;74:459-71.
13. Nurnberg HG, Martin GA, Somoza E, Coccaro EF, Skodol AE, Oldham JM, et al. Identifying personality disorders: Towards the development of a clinical screening instrument. Compr Psychiatry 2000;41:137-46.
14. Trull TJ, Amdur M. Diagnostic efficiency of the Iowa personality disorder screen items in a nonclinical sample. J Pers Disord 2001;15:351-7.
15. Moran P, Leese M, Lee T, Walters P, Thornicroft G, Mann A. Standardised Assessment of Personality-Abbreviated Scale (SAPAS): Preliminary validation of a brief screen for personality disorder. Br J Psychiatry 2003;183:228-32.
16. Morse JQ, Pilkonis PA. Screening for personality disorders. J Pers Disord 2007;21:179-98.
17. Dubro AF, Wetzler S, Kahn MW. A comparison of three self-report questionnaires for the diagnosis of DSM-III personality disorders. J Pers Disord 1988;2:256-266.
18. Pfohl B, Blum N, Zimmerman M, Stangl D. Structured Interview for DSM-III-R Personality SIDP-R. Iowa City, Iowa: Department of Psychiatry, University of Iowa, 1989.
19. Zimmerman M, Coryell WH. Diagnosing personality disorders in the community. A comparison of self-report and interview measures. Arch Gen Psychiatry 1990;47:527-31.
20. Hyler SE, Skodol AE, Oldham JM, Kellman HD, Doidge N. Validity of the Personality Diagnostic Questionnaire-Revised: A replication in an outpatient sample. Compr Psychiatry 1992;33:73-7.
21. Van Velzen CJM, Luteijn F, Scholing A, van Hout WJPJ, Emmelkamp PMG. The efficacy of the Personality Diagnostic Questionnaire-Revised as a diagnostic screening instrument in an anxiety disorder group. Clin Psychol Psychother 1999;6:395-403.
22. Davison S, Leese M, Taylor PJ. Examination of the screening properties of the Personality Diagnostic Questionnaire 4+ (PDQ-4+) in a prison population. J Pers Disord 2001;15:180-94.
23. Starcevic V, Uhlenhuth EH, Fallon S. The Tridimensional Personality Questionnaire as an instrument for screening personality disorders: Use in patients with generalized anxiety disorder. J Pers Disord 1995;9:247-53.
24. Gutiérrez F, Sangorrín J, Martín-Santos R, Torres X, Torrens M. Measuring the core features of personality disorders in substance abusers using the Temperament and Character Inventory (TCI). J Pers Disord 2002;16:344-59.
25. Gutiérrez F, Navinés R, Navarro P, García-Esteve L, Subirá S, Torrens M, Martín-Santos R. What do all personality disorders have in common? Ineffectiveness and uncooperativeness. Compr Psychiatry 2008;49:570-8.
26. Walters P, Moran P, Choudhury P, Lee T, Mann A. Screening for personality disorder: A comparison of personality disorder as-

sessment by patient and informants. Int J Methods Psychiatr Res 2004;13:34-9.
27. Mann AH, Raven P, Pilgrim J, Khanna S, Velayudham A, Suresh KP, et al. 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;29:985-9.
28. Ekselius L, Lindstrom E, von Knorring L, Bodlund O, Kullgren G. SCID-II interviews and the SCID Screen questionnaire as diagnostic tools for personality disorders in DSM-III-R. Acta Psychiatr Scand 1994;90:120-3.
29. Jacobsberg L, Perry S, Frances A. Diagnostic agreement between the SCID-II screening questionnaire and the Personality Disorder Examination. J Pers Assess 1995;65:428-33.
30. Duijsens IJ, Bruinsma M, Jansen SJT, Eurelings-Bontekoe EHM, Diekstra RFW. Agreement between self-report and semi-structured interviewing in the assessment of personality disorders. Pers Indiv Diff 1996;21:261-70.
31. Lenzenweger MF, Loranger AW, Korfine L, Neff C. Detecting personality disorders in a nonclinical population. Application of a 2-stage procedure for case identification. Arch Gen Psychiatry 1997;54:345-51.
32. Ottosson H, Bodlund O, Ekselius L, von Knorring L, Kullgren G, Lindström E, et al. DSM-IV and ICD-10 personality disorders: A comparison of a self-report questionnaire (DIP-Q) with a structured interview. Eur Psychiatry 1998;13:246-53.
33. Soldz S, Budman S, Demby A, Merry J. Diagnostic agreement between the Personality Disorder Examination and the MCMI-II. J Pers Assess 1993;60:486-99.
34. Marlowe DB, Husband SD, Bonieskie LM, Kirby KC. Structured interview versus self-report test vantages for the assessment of personality pathology in cocaine dependence. J Pers Disord 1997;11:177-90.
35. Fernandez-Montalvo J, Landa N, Lopez-Goni JJ, Lorea I. Personality disorders in alcoholics: A comparative pilot study between the IPDE and the MCMI-II. Addict Behav 2006;31:1442-8.
36. Starcevic V, Bogojevic G, Marinkovic J. The SCL-90-R as a screening instrument for severe personality disturbance among outpatients with mood and anxiety disorders. J Pers Disord 2000;14:199-207.
37. Widiger TA, Simonsen E. Alternative dimensional models of personality disorder: Finding a common ground. J Pers Disord 2005;19:110-30.