

## Diagnostic agreement between the Personality Diagnostic Questionnaire-4+ (PDQ-4+) and its Clinical Significance Scale

Natalia Calvo<sup>1,2</sup>, Fernando Gutiérrez<sup>3</sup> and Miguel Casas<sup>1,2</sup>

<sup>1</sup> Psychiatry Department. Hospital Vall d'Hebrón. CIBERSAM, <sup>2</sup> Universidad Autónoma de Barcelona and <sup>3</sup> Hospital Clinic Barcelona. IDIBAPS

### Abstract

**Background:** The Personality Diagnostic Questionnaire-4+ (PDQ-4+) is composed of a self-report and an interview, the Clinical Significance Scale, but no studies have reported joint findings. This study is the first to examine the diagnostic agreement between the Spanish version of the PDQ-4+ self-report and its corresponding interview. **Method:** The sample comprised 235 psychiatric outpatients who were assessed with both instruments. **Results:** The interview reduced to one half the number of diagnoses provided by self-report (83.4% to 38.3%; mean number of diagnoses 3.29 to .62). Diagnostic agreement was between fair and moderate (mean kappa .45 for PDQ-4+ total score). **Conclusions:** Findings suggest the utility of jointly administering the PDQ-4+ and its Clinical Significance Scale to screen for the presence or absence of personality disorders (PDs). Modifications in the diagnostic cut-offs for individual PDs and the PDQ-4+ total score may improve the efficacy of the instrument.

**Keywords:** PDD-4+; Clinical Significance Scale; diagnostic agreement; personality disorders.

### Resumen

**Concordancia diagnóstica entre el Personality Diagnostic Questionnaire-4+ (PDQ-4+) y su Escala de Significación Clínica.** **Antecedentes:** el Personality Diagnostic Questionnaire-4+ (PDQ-4+) está compuesto por un autoinforme y una entrevista, la Escala de Significación Clínica, pero ningún estudio ha sido publicado con resultados conjuntos. Este trabajo es el primero en examinar la concordancia diagnóstica entre la versión española del cuestionario PDQ-4+ y su correspondiente entrevista. **Método:** la muestra estaba formada por 235 pacientes psiquiátricos ambulatorios que fueron evaluados con ambos. **Resultados:** la entrevista reducía hasta la mitad el número de diagnósticos obtenido por el cuestionario (83,4% a 38,3%; número medio de diagnósticos de 3,29 a .62). La concordancia diagnóstica era de escasa a moderada (kappa media de .45 para la puntuación total del PDQ-4+). **Conclusiones:** nuestros datos sugieren la administración conjunta del PDQ-4+ y su Escala de Significación Clínica para el cribaje de presencia o ausencia de trastornos de personalidad (TPs). Modificaciones en los puntos de corte para TPs específicos y para la puntuación total del PDQ-4+ podrían mejorar la eficacia del instrumento.

**Palabras clave:** PDQ-4+; Escala de Significación Clínica; concordancia diagnóstica; trastornos de personalidad.

There is continuing debate regarding the most advisable method for assessing Personality Disorders (PDs): questionnaires or interviews. Currently, considering the absence of external criteria for validating diagnoses, interviews based on the *Diagnostic and Statistical Manual of Mental Disorders* (DSM, American Psychiatric Association, 1980-2000) are consensually considered the gold standard approach for diagnosis. They are said to discard false positive diagnoses better than self-reports, both by discriminating more accurately enduring traits from psychopathological states and better determining trait dysfunctionality (McDermutt & Zimmerman, 2005; Segal & Coolidge, 2003; Widiger & Samuel, 2005). However, their high cost severely limits their use within clinical settings (Aboraya, 2009). On the other hand, self-reports

are most widely used in the assessment of PDs, outperforming interviews in terms of their psychometric properties (Widiger & Samuel, 2005), but they are said to grossly over-diagnose (Hyler, Skodol, Kellman, Oldham, & Rosnick, 1990; Wang et al., 2012).

Nonetheless, the question might be not whether questionnaires or interviews are the right method, but whether accurate PD assessments need to combine them. Although both methods agree poorly with each other (Duijsens, Bruinsma, Jansen, Eurelings-Bontekoe, & Diekstra, 1996; Guy, Poythress, Douglas, Skeem, & Edens, 2008; Perry, 1992; Zimmerman, 1994), a literature review suggests that self-reports and interviews provide complementary information (Blackburn, Donnelly, Logan, & Renwick, 2004; Hopwood et al., 2008; Miller, Bagby, & Pilkonis, 2005). In this line, it has been suggested that the most efficient strategy in clinical settings is the initial administration of a questionnaire, followed by an interview-based assessment when self-reported information is positive (Widiger & Samuel, 2005). However, combining both requires a better understanding of their mutual relationship, currently unknown. We need to know to what degree diagnostic outputs differ; whether these differences can be improved by

simply changing the diagnostic thresholds; whether, contrarily, questionnaires and interviews measure distinct constructs; and whether these outcomes are the same for all individual disorders.

The Personality Diagnostic Questionnaire-4+ (PDQ-4+; Hyler, 1994) has been strongly recommended instead of a screening instrument (Abdin et al., 2011; Widiger & Samuel, 2005). The PDQ-4+ is a self-report, assessing ten specific PDs plus two PDs proposed in the *DSM-IV* (American Psychiatric Association, 1994), Appendix B. These 99 items, true-false format, literally reflect a single DSM-IV diagnostic criterion. Besides, a brief structured interview (10-15 minutes), the Clinical Significance Scale, follows the self-report and either confirms or does not confirm the diagnosis for each individual PD scoring over threshold. Unlike others, this interview directly reflects the principal DSM-IV general criteria for PD assessing whether: (a) the trait is enduring (criterion D for DSM-IV); (b) it is present in the absence of a psychopathological state, the effects of a substance or any medical condition (criteria E and F); and (c) it leads to distress or impairment (criterion C). This format better reflects the diagnostic procedure proposed by the DSM-IV and emphasized by the future DSM-5, but the increasing understanding that the intensity of a trait (i.e., personality) and its eventual harmfulness (i.e., disorder) can and must be distinguished (Hopwood, Thomas, Markon, Wright, & Krueger, 2012; Livesley, 2001; Parker & Barrett, 2000; Wakefield, 2008). So, using a single instrument that combines questionnaire and interview can avoid a pervasive problem of disagreeing diagnoses. Despite the apparent advantages informed by various authors (Abdin et al., 2011; de Reus, van den Berg, & Emmelkamp, 2011), there is no published research on diagnostic agreement with the Clinical Significance Scale of the PDQ-4+.

The main objective of this study was to examine the diagnostic agreement between the Spanish version of the PDQ-4+ self-report and its corresponding interview, the Clinical Significance Scale. Specifically, we wish to determine: (a) in which disorders the interview reduces prevalence; (b) the diagnostic indices Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value; and (c) the agreement diagnosis (kappa). A second goal was to examine the accuracy with which the questionnaire can predict interview-based diagnoses and to analyze the different diagnostic thresholds in order to improve its efficacy.

## Method

### Participants

The sample was comprised of 235 outpatients, consecutively attended at the Psychology Department of a General Teaching Hospital in Barcelona, Spain. All of them completed the PDQ-4+ and the Clinical Significance Scale.

Their mean age was 32.3 years ( $SD = 10.9$ , range 17-82 years). Regarding gender, 58.5% of them were men. Moreover, 76.2 ( $n = 179$ ) patients received at least one DSM-IV Axis I diagnosis. The most frequent diagnoses were Anxiety Disorders (65.9%,  $n = 155$ ) and Mood Disorders (56.2%,  $n = 132$ ).

### Instruments

The Personality Diagnostic Questionnaire-4+ (PDQ-4+; Hyler, 1994) has been partly described above. The total PDQ-4+ self-report score provides an index of overall personality disturbance:

for scores under 20, PD is discarded; scores of 20-30 require further assessment; and scores above 30 indicate a probable PD diagnosis.

The interview-based Clinical Significance Scale determines whether each self-reported disorder that reaches the diagnostic threshold fulfills the general criteria for PD: duration, state-independence, functional impairment, and distress. Like its previous versions (PDQ and PDQ-R), the PDQ-4+ has proven to have suitable psychometric properties both in its original version (Hyler, 1994) and in its adaptation to other languages and cultures, and in clinical and nonclinical samples (Abdin et al., 2011; Ha, Kim, Abbey, & Kim, 2007; Fossati et al., 1998; Fossati, Porro, Maffei, & Borroni, 2012; Hopwood et al., 2013; Kim, Choi, & Cho, 2000; Wang et al., 2012; Wilberg, Dammen, & Friis, 2000; Yang et al., 2000). The psychometric properties of the Spanish version of the self-report have been published previously, with acceptable findings in clinical (Calvo, 2007; Calvo et al., 2002, 2012) and nonclinical samples (Fonseca-Pedrero, Lemos-Giráldez, Paino, & Muñiz, 2011; Fonseca-Pedrero, Santarén-Rosell, Paino, & Lemos-Giráldez, 2013).

### Procedure

The study was approved by the hospital ethics committee. All patients agreed to participate voluntarily and provided written informed consent after receiving a complete explanation of the study. Exclusion criteria were psychosis, cognitive disorders and a current severe affective disorder.

The psychopathological assessment was carried out in 2 sessions by a clinical psychologist. The PDQ-4+ was completed individually. Patients who reached the diagnostic threshold, fulfilling the general criteria for PD, were assessed with the Clinical Significance Scale.

### Data analysis

SPSS 15.0 and AMOS 7.0 were used for all the analyses. Prevalences of self-reported PDQ-4+ scales were analyzed using the DSM-IV thresholds. The Clinical Significance Scale interview confirmed the diagnosis for screening-positive disorders, leading to dichotomous present/absent outcomes.

The capacity of the self-reported PDQ-4+ scores to predict interview-based PD diagnoses was analyzed through two different procedures. First, logistic regressions were performed in order to ascertain whether each interview-based categorical diagnosis was exclusively, or at least preferentially, predicted by its corresponding self-reported scale. Second, alternative cut-off points were selected, based on receiver operating characteristic (ROC) curves and kappa statistics, and diagnostic indices were then calculated: Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, Hit Ratio, and Cohen's kappa for chance-corrected diagnostic agreement. Sensitivity is the proportion of confirmed PD participants who had screened positive. Specificity is the proportion of confirmed non-PD participants who had screened negative. Positive predictive value is the proportion of positive screenings that the interview confirmed as PD. Negative predictive value is the proportion of negative screenings that the interview confirmed as non-PDs. Generally, kappa indexes under .40 indicate little-fair agreement; .41 to .60, moderate agreement; .61 to .80, substantial agreement; and .81 to 1.00 perfect agreement. As the

PDQ-4+ design only allows for people who screened positive to be interviewed for confirmation, false negatives must be assumed to be zero. To assess the effects of an eventual verification bias (Begg & Greenes, 1983), a variable number of false negatives (up to 10% of the total negative screenings) were simulated for each disorder, and diagnostic indices were recalculated in order to test their robustness. A greater percentage of false negatives was considered improbable, as self-reports systematically tend toward false-positives (Clark & Harrison, 2001; Widiger & Samuel, 2005).

Results

Table 1 (left) presents the prevalences of the PDQ-4+ scales using DSM-based cut-offs. Obsessive-Compulsive, Depressive, Avoidant and Borderline PDs had prevalences over 40%, whereas Antisocial was the least prevalent disorder (4.7%). Of patients, 83.4% had at least one PD diagnosis. We examined the degree to which the Clinical Significance Scale interview reduced the prevalence of PDs. After the interview, total prevalence dropped to 38.3%. Moreover, the interview-based prevalence for individual PDs ranged from .9% (Histrionic) to 16.6% (Avoidant). Also, the mean number of diagnoses per subject dropped from 3.29 (*SD* = 2.47) by self-report to .62 (*SD* = .92) by the interview. That is, the interview produced a reduction of 81.1% in the total number of diagnoses and a reduction of 54.1% in the number of subjects with any diagnosis. If the interview is considered the gold standard, this entails percentages of over-diagnosis ranging from 62.5 to 93.8% for individual disorders.

Next, the extent to which each self-reported scale can predict its corresponding interview-based diagnosis was analyzed, as good predictions would indicate simple differences in threshold levels between methods. Logistic regression results revealed that seven interview-based disorders were exclusively predicted by their respective self-reported score, with low to moderate Nagelkerke *R*<sup>2</sup> coefficients (*R*<sup>2</sup><sub>N</sub>). They were the Paranoid (.22), Schizoid (.54), Schizotypal (.24), Histrionic (.48), Narcissistic (.48), Borderline (.34) and Dependent (.57) disorders. On the other

hand, the Antisocial (.32), Avoidant (.58), Obsessive-Compulsive (.25) and Depressive (.24) disorders were mainly predicted by their own scales, although other non-corresponding scales entered into the equation. Only the Negativistic disorder was exclusively predicted by a non-corresponding scale (Paranoid, .57). The mean Nagelkerke *R*<sup>2</sup> for the most predictive scale was .40.

Alternative cut-offs were then established for the self-reported scales based on kappa statistics and ROC curves, and diagnostic indices were calculated (Table 1, right). For 9 of 12 self-reported scales, the best cut-off was located one or two criteria above the original DSM-IV threshold. Specificity (*M* = .90, range: .73-.98) and negative predictive value (*M* = .98, range: .94-1.00) were good to excellent. On the contrary, sensitivity (*M* = .73, range .50 to 1.00) and positive predictive value (*M* = .27, range .06 to .61) were often unsuitable. Furthermore, whereas hit rates (*M* = .89, range .74 to .98) were good, kappa indices (*M* = .32, range: .10-.58, all *ps*<.001) indicated that many successful classifications, albeit significant, were not far from those expected by chance. Additional analyses simulating a variable number (up to 10%) of false negatives in order to control for verification bias produced severe losses of sensitivity, but had practically no effect on the remaining indices.

Regarding the PDQ-4+ total score, the optimal cut-off for the presence of PD was found to be 35, five points above the proposed threshold. Despite this, diagnostic values were still in the moderate range: sensitivity = .81, specificity = .67, positive predictive value = .60, negative predictive value = .85, hit rate = .72 and Kappa = .45 (*p*<.001).

Discussion and conclusions

Overall, using the PDQ-4+ self-report, our results found a prevalence of PDs that is within the expected range in clinical samples and consistent with the literature (83.4%, mean number of diagnoses 3.29) (de Reus et al., 2013; Fossati et al., 1998; Fossati et al., 2012; Wang et al., 2012; Wilberg et al., 2000). In agreement with previous studies, Obsessive-Compulsive, Borderline,

Table 1  
Prevalence of self-reported and interview-based PDs in the PDQ-4+, and prevalence, diagnostic indices, and agreement for alternative cut-off points (n = 235)

	Self-reported diagnoses using DSM cut-offs			Interview-based diagnoses		Alternative cut-offs									
						Self-reported diagnoses using alternative cut-offs			Diagnostic indices					Agreement	
	Cut off	n	(%)	n	(%)	Cut off	n	(%)	S	SP	PPV	NPV	Hit rate	k	p
Paranoid	4	80	(34.0)	6	(2.6)	6	21	(8.9)	.50	.92	.14	.99	.91	.19	<.001
Schizoid	4	25	(10.6)	4	(1.7)	5	11	(4.7)	.75	.97	.27	.99	.96	.39	<.001
Schizotypal	5	48	(20.4)	6	(2.6)	5	48	(20.4)	1.00	.82	.13	1.00	.82	.19	<.001
Antisocial	4	11	(4.7)	3	(1.3)	4	17	(7.2)	1.00	.94	.18	1.00	.94	.28	<.001
Borderline	5	95	(40.4)	15	(6.4)	7	30	(12.8)	.60	.90	.30	.97	.89	.34	<.001
Histrionic	5	32	(13.6)	2	(.9)	7	5	(2.1)	.50	.98	.20	1.00	.98	.28	<.001
Narcissistic	5	25	(10.6)	3	(1.3)	6	7	(3.0)	.67	.98	.29	1.00	.97	.39	<.001
Avoidant	4	105	(44.7)	39	(16.6)	6	44	(18.7)	.69	.91	.61	.94	.88	.58	<.001
Dependent	5	48	(20.4)	18	(7.7)	6	21	(8.9)	.67	.96	.57	.97	.94	.58	<.001
Obsessive	4	137	(58.3)	30	(12.8)	5	79	(33.6)	.80	.73	.30	.96	.74	.31	<.001
Depressive	5	120	(51.1)	17	(7.2)	7	37	(15.7)	.53	.87	.24	.96	.85	.26	<.001
Negativistic	4	47	(20.0)	3	(1.3)	4	47	(20.0)	1.00	.81	.06	1.00	.81	.10	<.001

Note: S= Sensitivity; SP= Specificity; PPV= Positive Predictive Value; NPV= Negative Predictive Value; k= kappa

Avoidant and Depressive PDs were the most prevalent diagnoses (Ha, Kim, Abbey, & Kim, 2007). However, in the absence of studies using the Clinical Significance Scale allowing comparison, the interview-based prevalence of any PD (38.3%) was close to recent estimates in clinical samples (31.4% in Zimmerman, Rothschild, & Chelminski, 2005; 31.9% in Wang et al., 2012). Nevertheless, this similarity may be banal, as other studies have reported prevalences ranging from 11% to 72%, depending on the instrument, sample, or DSM version used (de Reus et al., 2013; Zimmerman et al., 2005). The differences found in diagnostic prevalence between the PDQ-4+ and the interview, though impressive, are similar to those usually reported (Abidin et al., 2011; de Reus et al., 2011; Wang et al., 2012). For instance, after its interview, the overall number of diagnoses was reduced by one fifth, and the number of diagnosed subjects to less than one half. Concerning individual PDs, only between 6.2% and 37.5% of the subjects initially screened as positive were finally considered to have a disorder after the interview.

Concordance between the PDQ-4+ and its interview ranged from poor to moderate, as reported using any other combination of questionnaire and interview (Gárriz & Gutiérrez, 2009). The PDQ-4+ self-report seems well suited for discarding individual PDs within clinical samples: 98% of the participants who screened negative were indeed free from personality psychopathology (Negative Predictive Value), whereas nearly 90% of the non-PD participants were successfully screened as healthy (Specificity). Furthermore, all diagnoses, except for Negativistic PD, are mainly, or even uniquely, predicted by their corresponding self-reported scales. In contrast, a mean  $R^2_N$  of .40 and a mean kappa of .32 for individual PDs suggest that much of the variation in the diagnosis is unrelated to the self-reported scores. The kappa index for the presence of any PD (.45) was also moderate, a poor outcome, considering that our attempt was to maximize kappa. Our results only partially outperform those reported in the literature. Brief screenings such as the Iowa Personality Disorder Screen (IPDS) or the Inventory of Interpersonal Problems (IIP-PD) show superior agreement (median kappa = .56); different versions of the Personality Diagnostic Questionnaire (PDQ) present median kappa of .42; the Temperament and Character Inventory (TCI) has a median kappa of .36; the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II) and the International Personality Disorder Examination (IPDE) screening tools have a median kappa of .29; and the different versions of the Millon Clinical Multiaxial Inventory (MCMI) presents a median kappa of .26 (Gárriz & Gutiérrez, 2009).

A weak association between self-report and interview is far more problematic than a simple difference in threshold. The evidence suggests that the inability of self-reports to predict final diagnoses would really constitute more of a conceptual issue than just measurement concerns. Discordance among methods is partly attributable to a flawed underlying model, the DSM. Indeed, disagreement between interviews and self-reports is similar to that between any two interviews, so the method is exonerated (Clark & Harrison, 2001). Agreement usually increases whenever disorders are measured dimensionally instead of categorically (Miller et al., 2005; Skodol, Oldham, Rosnick, Kellman, & Hyler, 1991). Second, agreement is not the same across different traits. In our study, Cluster C disorders showed the best kappa (.58), whereas Negativistic, Paranoid and Schizotypal PDs showed the worst one (.10 to .19). Similar heterogeneity has been found in other

studies, where Avoidant PD has shown higher agreement (Clark, Livesley, & Morey, 1997). Therefore, our research highlights that instruments cannot be better than the model they operationalize, and further improvement of our instruments necessarily entails reexamining our taxonomy (Clark et al., 1997).

Accordingly, it has been suggested that a considerable part of disagreement would be attributable to self-reports and interviews measuring different constructs. For example, we observed that many self-reported traits in our study did not reflect personality but rather the individual's current condition: transient psychopathological states, such as anxiety or distress, which were subsequently excluded by the interview. Self-reported instruments are more accurate than observable behaviors in assessing subjective experiences, especially those entailing undesirable consequences for others (Blackburn et al., 2004; Hopwood et al., 2008; Miller, Campbell, Pilkonis, & Morse, 2008). This would suggest that interviews are more capable of discriminating the enduring traits than the defining PD traits. However, personality pathology has been revealed to be more variable and fluid than stated in previous literature: the course of pathological traits fluctuates significantly over periods as brief as six years (Zanarini, Frankenburg, Hennen, Reich, & Silk, 2005) and they are, in fact, less enduring than many Axis I disorders (Shea & Yen, 2003). As for taxonomists, they still need to clarify which duration turns a state into a trait. In this sense, whereas the PDQ-4+ interview stringently requires traits to be lifetime, the SCID-II requires five years and the DIPD, two. On the other hand, a notable amount of self-reported traits are discounted by the interview because they do not cause distress or impairment. For example, our Obsessive-compulsive patients often reported that over-control decreases, instead of increasing, their distress, providing them with considerable advantages. Indeed, these patients have shown minimal evidence of impairment elsewhere (Costa, Samuels, Bagby, Daffin, & Norton, 2005). Paranoid subjects see others as malevolent, so hyper-vigilance and mistrust seem to be the adequate response for this group. Similarly, many antisocial and narcissistic subjects are more detrimental to others than to themselves, and are overall benefited by their behavioral style (Mealey, 1995). Besides, we should abandon the idea that an extreme trait inevitably, or frequently, leads to a dysfunction. The relationship between trait intensity and maladaptation has almost been a premise in the field, despite convincing argumentation to the contrary (Livesley, 2001; Livesley & Jang, 2000; Parker & Barrett, 2000; Wakefield, 2008). The issue of which assessment method is better is probably meaningless until these conceptual issues are clarified. The task of determining whether the presence of certain traits or their detrimental consequences should decide the diagnosis, who (the patient, the clinician, or others) is the competent judge of such detriment, and how it should be operationalized is under way.

This study presents some limitations. The sample was recruited at an outpatient setting at a University Hospital and was relatively small. Additionally, the sample showed considerable comorbidity with Axis I disorders. Therefore, our results do not necessarily extend to those reported in other clinical and nonclinical samples. Moreover, diagnostic instruments for PD have usually shown poor convergence with each other, so our findings need further replication beyond the PDQ-4+ and its interview. Future research should investigate the relationship in different clinical and nonclinical samples and compare our results with other interviews and self-reports measures to generalize the findings.

Despite these limitations, this study found that the routine administration of the Clinical Significance Scale following the self-report seems warranted. The PDQ-4+ interview could be considered overall as an indicator of personality pathology severity following the general PD criteria. It unequivocally links the interview with the evaluation of durability, state-independence, and harmfulness for the future DSM-5. Therefore, the Clinical Significance Scale interview provided complementary information about PDs, reducing false-positive diagnoses typical of the self-report PDQ-4+. Our findings indicate that, with some thresholds for specific PDs (one or two criteria) and total score (5 points above), the PDQ-4+ could be modified or changed to improve the diagnosis of PDs. These changes are likely to increase the validity

and reliability of the instrument and probably the diagnostic agreement.

#### Acknowledgements

There are no conflicts of interest with other projects.

The work was partially supported by public funds from the Pla Director de Salut Mental i Addiccions (Generalitat de Catalunya Health Department) and grants from the Obra Social - Fundació "La Caixa".

This work was partially supported by a grant from the Spanish Ministerio de Educación y Ciencia (FIS 03/0464) awarded to F. Gutiérrez.

#### References

- Abdin, E., Koh, K., Subramaniam, M., Guo, M.-E., Leo, T., Teo, C., ... Chong, S.A. (2011). Validity of the Personality Diagnostic Questionnaire-4 (PDQ-4+) among mentally ill prison inmates in Singapore. *Journal of Personality Disorders, 25*(6), 834-841.
- Aboraya, A. (2009). Use of structured interviews by psychiatrists in real clinical settings: Results of an open-question survey. *Psychiatry (Edgemont), 6*, 24-28.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4<sup>th</sup> ed.). Washington, DC: Author.
- Begg, C.B., & Greenes, R.A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics, 39*, 207-215.
- Blackburn, R., Donnelly, J.P., Logan, C., & Renwick, S.J.D. (2004). Convergent and discriminative validity of interview and questionnaire measures of Personality Disorder in mentally disordered offenders: A multitrait-multimethod analysis using confirmatory factor analysis. *Journal of Personality Disorders, 18*(2), 129-150.
- Calvo, N. (2007). Adaptació al castellà del Personality Diagnostic Questionnaire-4+: Propietats psicomètriques en població clínica i en estudiants [Spanish Adaptation of Personality Diagnostic Questionnaire-4+: Psychometric properties in clinical sample and students]. Doctoral Thesis. Department of Psychiatry and Legal Medicine. Faculty of Medicine. Autonomous University of Barcelona.
- Calvo, N., Caseras, X., Gutiérrez, F., & Torrubia, R. (2002). Adaptación Española del Personality Diagnostic Questionnaire-4+ (PDQ-4+) [Spanish adaptation of the Personality Diagnostic Questionnaire-4+]. *Actas Españolas de Psiquiatría, 30*, 7-13.
- Calvo, N., Gutiérrez, F., Andiñón, O., Caseras, X., Torrubia, R., & Casas, M. (2012). Psychometric properties of the Spanish version of the Personality Diagnostic Questionnaire-4+ (PDQ-4+) self-report in psychiatric outpatients. *Psicothema, 24*, 156-160.
- Clark, L.A., & Harrison, J.A. (2001). Assessment instruments. In W.J. Livesley (Ed.), *Handbook of personality disorders. Theory, research, and treatment* (pp. 277-306). New York: Guilford Press.
- Clark, L.A., Livesley, W.J., & Morey, L. (1997). Personality disorder assessment: The challenge of construct validity. *Journal of Personality Disorders, 11*, 205-231.
- Costa, P.T., Samuels, J., Bagby, R.M., Daffin, L., & Norton, H. (2005). Obsessive-compulsive personality disorder: A review. In M. Maj, H.S. Akiskal, J.E. Mezzich, & A. Okasha (Eds.), *Personality disorders* (pp. 405-439). Chichester, UK: Wiley.
- de Reus, R.J., van den Berg, J.F., & Emmelkamp, P.M.G. (2013). Personality Diagnostic Questionnaire-4+ is not useful as a screener in clinical practice. *Clinical Psychology and Psychotherapy, 20*(1), 49-54.
- 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.
- Fonseca-Pedrero, E., Lemos-Giraldez, S., Paino, M., & Muñiz, J. (2011). Schizotypy, emotional-behavioural problems and personality disorder traits in a non-clinical adolescent population. *Psychiatry Research, 190*, 316-321.
- Fonseca-Pedrero, E., Santarén-Rosell, M., Paino, M., & Lemos Giraldez, S., (2013). Cluster A maladaptive personality patterns in a non-clinical adolescent population. *Psicothema, 25*(2), 171-178.
- Fossati, A., Maffei, C., Bagnato, M., Donati, D., Donini, M., Fiorilli, M., ... Ansoldi, M. (1998). Brief communication: Criterion validity of the Personality Diagnostic Questionnaire-4+ (PDQ-4+) in a mixed psychiatric sample. *Journal of Personality Disorders, 12*, 172-178.
- Fossati, A., Porro, F. V., Maffei, C., & Borroni, S. (2012). Are the DSM-IV personality disorders related to mindfulness? An Italian study on clinical participants. *Journal of Clinical Psychology, 68*(6), 672-683.
- Gárriz, M., & Gutiérrez, F. (2009). Personality disorder screenings: A meta-analysis. *Actas Españolas de Psiquiatría, 37*(3), 148-153.
- Guy, L.S., Poythress, N.G., Douglas, K.S., Skeem, J.L., & Edens, J.F. (2008). Correspondence between self-report and interview-based assessment of antisocial PD. *Psychological Assessment, 20*, 47-54.
- Ha, J.H., Kim, E.E., Abbey, S.E., & Kim, T.S. (2007). Relationship between personality disorder symptoms and temperament in the young male general population of South Korea. *Psychiatry and Clinical Neurosciences, 61*, 59-66.
- Hopwood, C.J., Donnellan, M.B., Ackerman, R.A., Thomas, K.M., Morey, L.C., & Skodol, A.E. (2013). The validity of the Personality Diagnostic Questionnaire-4 Narcissistic Personality Disorder Scale for assessing pathological grandiosity. *Journal of Personality Assessment, 95*(3), 274-283.
- Hopwood, C.J., Morey, L.C., Edelen, M.O., Shea, M.T., Grilo, C.M., Sanislow, C.A., ... Skodol, A.E. (2008). A comparison of interview and self-report methods for the assessment of Borderline Personality Disorder criteria. *Psychological Assessment, 20*, 81-85.
- Hopwood, C.J., Thomas, K.M., Markon, K.E., Wright, A.G.C., & Krueger, R.F. (2012). DSM-5 Personality traits and DSM-IV personality disorders. *Journal of Abnormal Psychology, 121*, 424-432.
- Hyler, S.E. (1994). *PDQ-4+ Personality Diagnostic Questionnaire-4+*. New York: New York State Psychiatric Institute.
- Hyler, S.E., Skodol, A.E., Kellman, H.D., Oldham, J.M., & Rosnick, L. (1990). Validity of the Personality Diagnostic Questionnaire-Revised: Comparison with two structured interviews. *American Journal of Psychiatry, 147*, 1043-1048.
- Kim, D.I., Choi, M.R., & Cho, E.C. (2000). The preliminary study of reliability and validity on the Korean version of the Personality Disorder Questionnaire-4+ (PDQ-4+). *Journal of Korean Neuropsychiatric Association, 39*, 525-538.
- Livesley, W.J. (2001). Commentary on reconceptualizing personality disorder categories using trait dimensions. *Journal of Personality, 69*(2), 277-286.
- Livesley, W.J., & Jang, K.L. (2000). Toward an empirically based classification of personality disorder. *Journal of Personality Disorders, 14*, 137-151.

- McDermutt, W., & Zimmerman, M. (2005). Assessment instruments and standardized evaluation. In J.M. Oldham, A.E. Skodol, & D.S. Bender (Eds.), *Textbook of personality disorders* (pp. 89-101). Washington DC: American Psychiatric Association.
- Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences*, 18(3), 523-599.
- Miller, J.D., Bagby, R.M., & Pilkonis, P.A. (2005). A comparison of the validity of the five-factor model (FFM) personality disorder prototypes using FFM Self-Report and interview measures. *Psychological Assessment*, 17, 497-500.
- Miller, J.D., Campbell, W.K., Pilkonis, P.A., & Morse, J.Q. (2008). Assessment procedures for Narcissistic Personality Disorder. *Assessment*, 15, 483-492.
- Parker, G., & Barret, E. (2000). Personality and personality disorder: Current issues and directions [Editorial]. *Psychological Medicine*, 30, 1-9.
- Perry, J.C. (1992). Problems and considerations in the valid assessment of personality disorders. *American Journal of Psychiatry*, 149, 1645-1653.
- Segal, D.L., & Coolidge, F.L. (2003). Structured interviewing and DSM classification. In M. Hersen & S.M. Turner (Eds.), *Adult psychopathology and diagnosis* (4<sup>th</sup> ed., pp. 72-103). New York: Wiley.
- Shea, M.T., & Yen, S. (2003). Stability as a distinction between Axis I and Axis II disorders. *Journal of Personality Disorders*, 17, 373-386.
- Skodol, A.E., Oldham, J.M., Rosnick, L., Kellman, H.D., & Hyler, S.E. (1991). Diagnosis of DSM-III-R personality disorders: A comparison of two structured interviews. *International Journal of Methods in Psychiatric Research*, 1, 3-26.
- Wakefield, J.C. (2008). The perils of dimensionalization: Challenges in distinguishing negative traits from Personality Disorders. *Psychiatric Clinics of North America*, 31, 379-393.
- Wang, L., Ross, C.A., Zhang, T., Dai, Y., Zhang, H., Tao, M., ... Xiao, Z. (2012). Frequency of borderline personality disorder among psychiatric outpatients in Shanghai. *Journal of Personality Disorders*, 26(3), 393-401.
- Widiger, T.A., & Samuel, D.B. (2005). Evidence-based assessment of personality disorder. *Journal of Personality Assessment*, 17, 278-287.
- Wilberg, T., Dammen, T., & Friis, S. (2000). Comparing Personality Diagnostic Questionnaire-4+ with Longitudinal, Expert, All Data (LEAD) standard diagnoses in a sample with a high prevalence of Axis I and Axis II disorders. *Comprehensive Psychiatry*, 41, 295-302.
- Yang, J., McCrae, R.R., Costa, P.T., Yao, S., Dai, X., Cai, T., & Gao, B. (2000). The cross-cultural generalizability of Axis II constructs: An evaluation of two personality disorder assessment instruments in the People's Republic of China. *Journal of Personality Disorders*, 14, 249-263.
- Zanarini, M.C., Frankenburg, F.R., Hennen, J., Reich, B., & Silk, K.R. (2005). The McLean Study of Adult Development (MSAD): Overview and implications of the first six years of prospective follow-up. *Journal of Personality Disorders*, 19, 505-523.
- Zimmerman, M. (1994). Diagnosing personality disorders. A review of issues and research methods. *Archives General of Psychiatry*, 51, 225-245.
- Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, 162, 1911-1918.