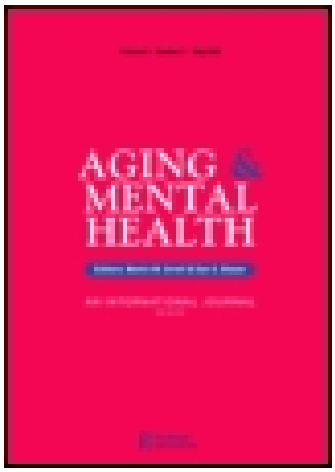

Aging & Mental Health
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/camh20

Diagnostic accuracy of the Gerontological Personality Disorder Scale (GPS) in Dutch general practice

Krystle A.P. Penders[a], Gina Rossi[b], Job F.M. Metsemakers[a], Inge G.P. Duimel-Peeters[ac] & Sebastiaan J.P. van Alphen[bd]

[a] Department of Family Medicine, School for Public Health and Primary Care, Maastricht University (UM), Maastricht, The Netherlands

[b] Faculty of Psychology & Educational Sciences, Department of Clinical & Lifespan Psychology, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[c] Department of Integrated Care, Maastricht University Medical Centre (MUMC), Maastricht University (UM), Maastricht, The Netherlands

[d] Department of Old Age Psychiatry, Mondriaan Hospital, Heerlen-Maastricht, The Netherlands

Published online: 16 Feb 2015.

To cite this article: Krystle A.P. Penders, Gina Rossi, Job F.M. Metsemakers, Inge G.P. Duimel-Peeters & Sebastiaan J.P. van Alphen (2015): Diagnostic accuracy of the Gerontological Personality Disorder Scale (GPS) in Dutch general practice, Aging & Mental Health, DOI: 10.1080/13607863.2015.1008989

To link to this article: http://dx.doi.org/10.1080/13607863.2015.1008989

Aging & Mental Health, 2015
http://dx.doi.org/10.1080/13607863.2015.1008989

Routledge
Taylor & Francis Group

Diagnostic accuracy of the Gerontological Personality Disorder Scale (GPS) in Dutch general practice

Krystle A.P. Penders[a]*, Gina Rossi[b], Job F.M. Metsemakers[a], Inge G.P. Duimel-Peeters[a,c] and Sebastiaan J.P. van Alphen[b,d]

[a]*Department of Family Medicine, School for Public Health and Primary Care, Maastricht University (UM), Maastricht, The Netherlands;* [b]*Faculty of Psychology & Educational Sciences, Department of Clinical & Lifespan Psychology, Vrije Universiteit Brussel (VUB), Brussels, Belgium;* [c]*Department of Integrated Care, Maastricht University Medical Centre (MUMC), Maastricht University (UM), Maastricht, The Netherlands;* [d]*Department of Old Age Psychiatry, Mondriaan Hospital, Heerlen-Maastricht, The Netherlands*

(*Received 18 September 2014; accepted 1 January 2015*)

Objective: Personality disorders (PDs) often remain unrecognized in older adults by doctors in general practice. Therefore, this study evaluated the diagnostic accuracy of a screening instrument, the Gerontological Personality Disorder Scale (GPS), in a Dutch general-practice population of older adults.
Method: The psychometric properties of the GPS patient (GPS-pv) and informant (GPS-iv) versions were assessed in a sample of 302 (144 male) patients (average age: 69.9 years) and 302 (124 male) informants (average age: 64.7 years), respectively, using an informant-based personality questionnaire (the *Hetero-Anamnestische Persoonlijkheidsvragenlijst*) as a reference criterion.
Results: The internal consistency (average item correlation) of the subscale and total scores of the GPS-pv and GPS-iv were .12 (HAB), .16 (BIO), and .10 (total); and .16 (HAB), .15 (BIO), and .12 (total), respectively. The test–retest reliability was strong for both the GPS-pv (r_s = .56 [HAB], r_s = .67 [BIO], r_s = .66 [total]) and the GPS-iv (r_s = .52 [HAB], r_s = .65 [BIO], r_s = .68 [total]) versions. The sensitivity and specificity of the GPS-pv were .83 and .27, respectively, with a cutoff score of ≥1. Raising the cutoff score to ≥2, the sensitivity dropped to .59, whereas the specificity rose to .57. For the GPS-iv, a cutoff score of ≥3 maximized the sensitivity (.78) and specificity (.65).
Conclusion: The diagnostic accuracy of the GPS-iv was preferable to that of the GPS-pv. This is the first psychometric study to use the GPS as an age-specific screening instrument for PDs.

Keywords: personality disorders; elderly; diagnostic efficiency; reliability; general practice

Introduction

The publication of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychological Association [APA], 2013) has changed little with regard to the classification of personality disorders (PDs). In contrast to many researchers' opinions (e.g., Widiger & Trull, 2007), a clear revision of the section on PDs in the DSM-5 (APA, 2013) did not occur; instead, Section II completely retained the DSM-IV (APA, 2000) categories and criteria. A growing number of researchers have been strongly stating the importance of investigating PDs in older adults (Agronin & Maletta, 2000; Oltmanns & Balsis, 2011). They have argued that a deeper understanding and recognition of PDs in older adults is urgently needed because PDs are a widespread problem in this population. Although both scientific and clinical attention to PDs are slowly shifting toward older adults (i.e., individuals >60 years), the DSM-5 (APA, 2013) continues to focus solely on younger adults. Despite the ongoing discussion concerning the feasibility of including criteria assessing PD in older adults in the DSM (Balsis, Woods, Gleason, & Oltmanns, 2007; Segal, Coolidge, & Rosowsky, 2006), the DSM is frequently used in older adults worldwide because there is no better alternative (van Alphen et al., 2015).

The recognition of PDs that is a multifaceted construct is of great clinical relevance because the presence of a PD may interfere with the diagnosis of other clinical disorders. From a treatment perspective, PDs are a complicating factor in two ways. First, patients with PDs are more complex because they are more resistant to the effects of treatment (cluster A), less compliant (cluster B), and tend to make extreme demands related to their care (cluster C) (van Alphen, Derksen, Sadavoy, & Rosowsky, 2012, p. 806), which leads to an increase in healthcare consultations (Jackson & Burgess, 2004). Second, PDs adversely affect treatment outcomes, resulting in less improvement and an increased risk for relapse and readmission (Stevenson, Brodaty, Boyce, & Byth, 2011). One explanation might be when a PD is present, the symptoms of comorbid mental disorders are more severe and chronic (Stevenson et al., 2011). Moreover, individuals with PDs tend to have (1) lower self-perceived quality of life and general health; (2) decreased physical, social, and cognitive functioning; (3) lower self-esteem; (4) less life satisfaction; (5) decreased well-being; (6) a lower number of, and

*Corresponding author. krystle.penders@maastrichtuniversity.nl

satisfaction with, social support; and (7) an increased predisposition to other mental disorders compared to individuals without PDs (e.g., Condello, Padoani, Uguzzoni, Caon, & De Leo, 2003; Frankenburg & Zanarini, 2006; Jackson & Burgess, 2004; Stevenson et al., 2011). Moreover, PDs are highly prevalent, with rates between 3% and 13% in a population of community-dwelling older adults (Ames & Molinari, 1994; Silberman, Roth, Segal, & Burns, 1997; Weissman, 1993).

While there is a growing interest in PDs in older adults, there are several barriers to accurate screening and assessment. One significant barrier is the paucity of age-specific and well-validated instruments (Edelstein et al., 2007; Oltmanns & Balsis, 2011). This may lead to reliance on instruments that are not suited for older adults, resulting in denial of the aging process and environmental related changes known to occur in older adults (APA, 2014). This may result in over-diagnosis or under-diagnosis of PDs (Balsis et al., 2007). Another more general obstacle is the predominant reliance on self-report data, which is known to have limitations (e.g., Klonsky, Oltmanns, & Turkheimer, 2002). A more accurate view on PDs might be acquired also by obtaining informant information.

In response to the lack of instruments assessing PD in older adults, van Alphen and colleagues developed the Gerontological Personality Disorder Scale (GPS), an age-specific screening instrument for PDs in older adults consisting of both a patient (GPS-pv) and an informant version (GPS-iv). Psychometric properties of the GPS-pv in elderly psychiatric outpatients yielded fair to good internal consistency, test–retest reliability, and criterion validity, when using clinical diagnoses of experienced medical doctors and psychologists as a reference criterion (van Alphen, Engelen, Kuin, Hoijtink, & Derksen, 2006). In a multicenter study by Tummers, Hoijtink, Penders, Derksen, and van Alphen (2011), the GPS items were added to an experimental pool of items formulated by a Dutch expert panel on PDs in older adults by identifying the items that were the most predictive of PDs (based on multidisciplinary diagnosis) in an elderly psychiatric outpatient population. Of this item pool, the GPS item, 'I have sometimes said to my family or friends that I don't want to live any longer,' combined with the experimental item, 'I like to be in control,' was the most accurate (65%) in predicting the presence or absence of PDs.

The aim of this study was to evaluate the internal consistency, test–retest reliability, inter-rater reliability, and diagnostic accuracy of both the GPS versions in older adults in Dutch general practice. Given the disability caused by PDs, the high prevalence of PDs, and PDs' complicating effect on treatment, it is of great clinical value for the general practitioner (GP) to have a reliable and validated brief instrument based on the DSM-5 definition of PD to use with older adults (van Alphen, 2006). Such an instrument will enable GPs to objectify the presence of PDs and tailor treatment to accommodate the specific needs of individual patients. This may enhance the odds of a positive response to treatment and circumvent treatment dropout.

Methods

Participants

From 2009 to 2012, 302 older, adult patients receiving care from one of the five participating general practices (Heerlen: $n = 47$, Schinveld: $n = 48$, Hoensbroek: $n = 84$, Landgraaf: $n = 39$, and Ubachsberg: $n = 84$) in the south of the Netherlands were included. These general practices (i.e., first-line primary care) were representative of practices throughout the Netherlands.

To be eligible to participate, individuals had to be 60 years or older and have an informant who was also willing to participate. In addition, to guarantee adequate and reliable responding, older adults with severe psychiatric disorders, cognitive dysfunctions, and major attentional problems as a result of sedation and/or alcohol use, intellectual disabilities, or a life expectancy of less than three months were excluded. The presence of the exclusion criteria was determined by a two-part process. The GPs made an initial evaluation of eligible patients; then, the following assessments were conducted during a home visit: the national list of illnesses from the *Centraal Bureau voor de Statistiek* (Central Office of Statistics in the Netherlands, CBS), the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), the Geriatric Depression Scale (GDS; Yesavage et al., 1983), the Alcohol Use Disorder Identification Test (AUDIT; Babor, Higgings-Biddle, Saunders, & Monteiro, 2001), and the Brief Symptom Inventory (BSI; De Beurs, 2008). Older adults with a GDS score of $\geq$20, BSI subscale scores for anxiety, phobia, and psychoticism categorized as 'very high,' or MMSE score of $\leq$24 were excluded. After exclusion (for an overview, see Figure 1), the data from 302 patient-informant pairs were retained for analysis. Additionally, 104 patient-informant pairs were willing to complete a follow-up assessment.

Approximately half of the 302 patients were male ($n = 144$). The average age was 69.9 years (range = 60−91). Of the 302 informants, 124 were male. Their average age was 64.7 years (range = 32−89). Table 1 shows demographic information for both the patients and informants.

Recruitment

Ethics approval was granted from the Medical Ethical Review Commission of the Academic Hospital Maastricht (AzM), the Netherlands (approval no. MEC 09-4-060).

Prior to patient recruitment, the GPs and practice support staff (PSS) from the participating general practices were given in-depth information about the content and procedure of the study. Subsequently, the GPs and PSS were asked to check their electronic patient databases and, using the formulated inclusion and exclusion criteria, list the eligible patients.

These patients received a written invitation from their GP with a reply coupon, stamped envelope, and information regarding the study. Patients willing to participate

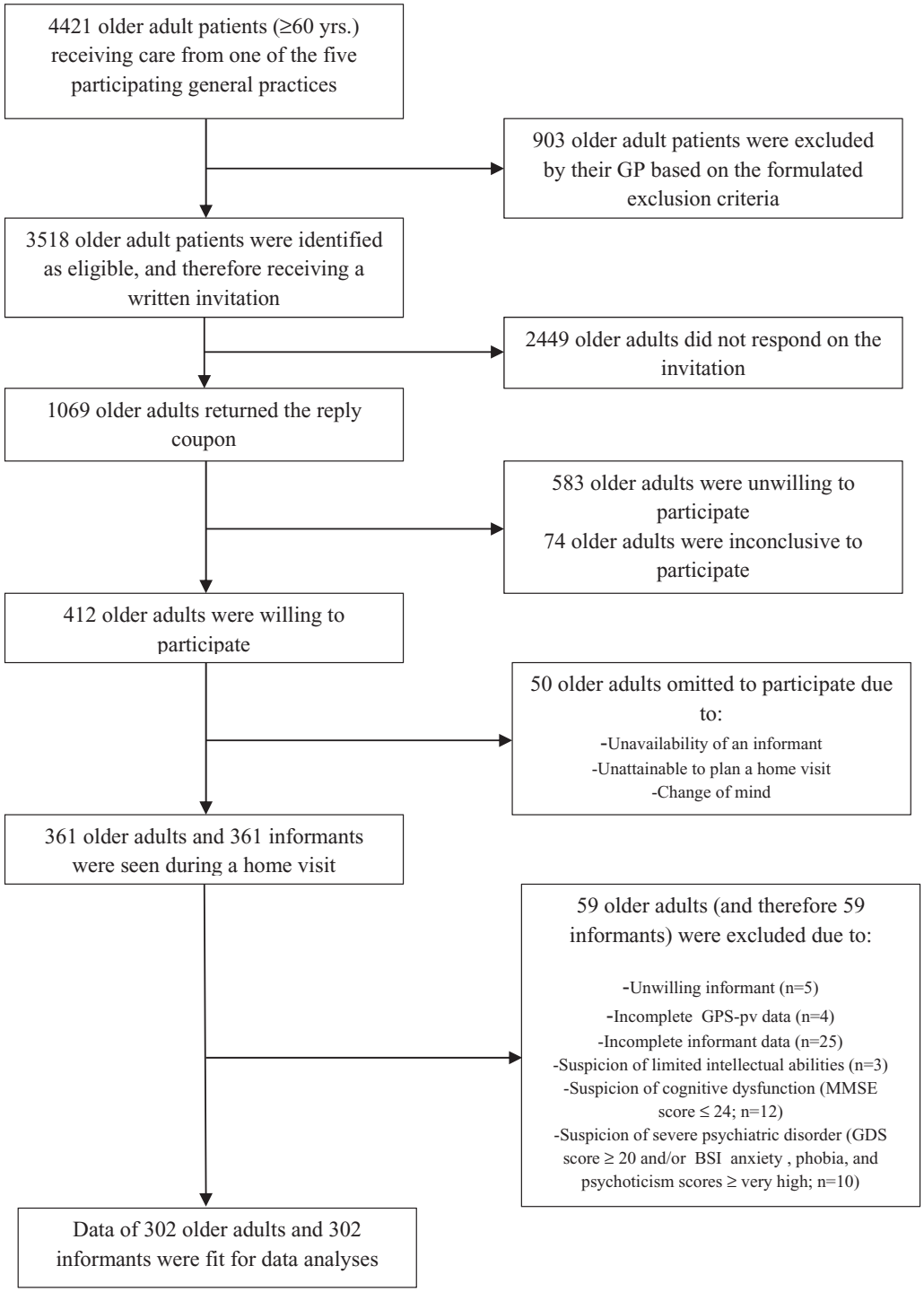

Figure 1. Flow chart response and exclusion of older adult patients and informants.

received a phone call from the researcher, who provided additional information and inquired about the availability of an informant. After receiving informed consent orally from both the patient and informant, a home visit was scheduled.

Measures

The following instruments were used. Instruments 1–6 were administrated to the patients and instruments 6 and 7 were completed by the informants.

CBS list

The CBS list is a Dutch questionnaire addressing the 24 most prevalent somatic conditions/complaints in the Netherlands according to the Dutch Central Office of Statistics (CBS, n.d.). The 25 items are scored with either an 'yes' or a 'no' response.

Mini-Mental State Examination (MMSE)

The MMSE is a brief questionnaire addressing cognitive impairment (Folstein et al., 1975). Scores range from 0 to

Table 1. Demographic features of patients and informants.

	Patients		
	No PS*	PS*	Informants
Age			
Years (sd)	69.6 (7.4)	71.5 (7.3)	64.7 (11.3)
Range	60.0-91.0	60.0-88.0	32.0-89.0
Gender (%)			
Men	116 (45.3)	28 (60.9)	124 (41.1)
Women	140 (54.7)	18 (39.1)	178 (58.9)
Marital status (%)			
Single	9 (3.5)	4 (8.7)	–
Married/civil partnership	200 (78.1)	32 (69.6)	–
Cohabit	7 (2.7)	1 (2.2)	–
Two-household family (LAT)	5 (2.0)	2 (4.3)	–
Divorced	3 (1.2)	2 (4.3)	–
Widowed	30 (11.7)	5 (10.9)	–
Missing data	2 (0.8)	0 (0.0)	–
Housing (%)			
Dwelling house	252 (98.4)	46 (100.0)	–
Sheltered accommodation	1 (0.4)	0 (0.0)	–
Elderly home	1 (0.4)	0 (0.0)	–
Other	2 (0.8)	0 (0.0)	–
Educational level (%)			
Elementary school	26 (10.1)	5 (10.9)	20 (6.6)
School of domestic science/trade school	110 (43.0)	20 (43.5)	102 (33.8)
Senior secondary vocational ED	50 (19.5)	9 (19.6)	83 (27.5)
Senior general secondary ED/pre-university ED	10 (3.9)	2 (4.3)	8 (2.6)
Higher professional ED	47 (18.4)	6 (13.0)	63 (20.9)
University	13 (5.1)	4 (8.7)	26 (8.6)
Nature of relationship (%)			
Partner		238 (78.8)	
Sibling		8 (2.7)	
Child		38 (12.6)	
Grandchild		1 (0.3)	
Brother/sister-in-law		1 (0.3)	
Friend		5 (1.7)	
Son/daughter-in-law		2 (0.6)	
Other		9 (3.0)	
Duration of relationship in years (%)			
1–5			8 (2.6)
6–10			9 (3.0)
11–15			6 (2.0)
16–20			7 (2.3)
21–25			10 (3.3)
26–30			7 (2.3)
31–35			11 (3.6)
36–40			35 (11.6)
41–45			43 (14.2)
46–50			58 (19.2)
51–55			42 (13.9)
56–60			34 (11.3)
61–65			15 (5.0)
66–70			4 (1.3)
71–75			2 (0.7)
Missing data			11 (3.6)

Notes: – = data were not obtained.
*Classification based on score on HAP score (4 = high/very high subscale scores; 2 = high/very high subscale scores; and 6 = above average subscale scores).

30, and a score ≥25 indicates normal cognitive functioning.

Geriatric Depression Scale (GDS)

The GDS is an age-specific self-report measure concerning depressive symptomatology (Yesavage et al., 1983) consisting of 30 yes/no questions. Scores range from 0 to 30, and higher scores indicate the presence of more depressive symptoms. The GDS has adequate psychometric properties (Yesavage et al., 1983). In the current sample, the internal consistency reliability was $\alpha = .80$.

Alcohol Use Disorders Identification Test (AUDIT)

The AUDIT is a self-report measure concerning hazardous alcohol use, harmful alcohol use, and dependence symptoms (Babor et al., 2001), and consists of 10 items. Scores range from 0 to 40, and higher scores indicate a higher risk of alcohol-related problems. The AUDIT is a widely recommended, reliable, and valid instrument (Babor et al., 2001; Gómez et al., 2006). In the current sample, the internal consistency reliability, as measured with average item correlation (AIC), was .18 with $\alpha = .56$.

The Brief Symptom Inventory (BSI)

The BSI (De Beurs, 2008) is a brief version of the Symptom Checklist-90, a multidimensional self-report measure concerning psychopathology. It consists of 53 items that assess somatic complaints, cognitive problems, interpersonal sensitivity, depressed mood, anxiety, hostility, phobia, paranoid thoughts, and psychoticism. Each item is scored on a five-point Likert scale ranging from 'entirely not' to 'very much.' The BSI is a reliable and valid instrument (De Beurs, 2008). In the current sample, the internal consistency reliabilities (AIC) ranged from .06 with $\alpha = .28$ (psychoticism) to .22 with $\alpha = .54$ (hostility), and with $\alpha = .86$ for the total scale.

Gerontological Personality Disorders Scale (GPS)

The GPS is a brief and age-specific screening instrument designed to measure PDs in older adults in mental health care (van Alphen et al., 2006). The GPS measures the presence or absence of a PD based on general definitions in the DSM-IV and DSM-5, respectively (APA, 2000; APA, 2013). Its construction was based on literature review and case studies, including the Delphi study of 53 Dutch experts in PDs in older adults (van Alphen, 2006). The only difference between the GPS-pv and GPS-iv is the perspective from which the questions are asked (first versus third person, respectively). There are 16 yes/no statements addressing both habitual behaviors (HAB) and biographical information (BIO; i.e., life characteristics). Table 2 shows the GPS content.

Hetero-Anamnestische Persoonlijkheidsvragenlijst (HAP)

The HAP (Barendse & Thissen, 2006) is a Dutch informant-based personality questionnaire that measures the

Table 2. Items of the Gerontological Personality Disorder Scale (GPS).

		Yes	No
	Habitual behavior (HAB)		
1	I don't like growing older because I become less attractive	1	0
2	I often worry about my health	1	0
3	I'm often concerned about my memory	1	0
4	I hope that others solve my problems	1	0
5	I'm often afraid of losing those who care for me, such as members of the family or my partner	1	0
6	I'm often taken advantage of by others	1	0
7	I find it difficult to fend for myself	1	0
	Biographical information (BIO)		
1	In my life I've been to see the doctor for many vague physical complaints	1	0
2	I have sometimes said to my family or friends that I don't want to live any longer	1	0
3	In the past I've been admitted to a psychiatric institution or convalescent home because of nerves	1	0
4	At important times in my life I've had a lot of trouble with nerves, stress or moodiness	1	0
5	In the past I've already had treatment from a psychiatrist or psychologist	1	0
6	I have sometimes tried to end my life	1	0
7	At the most I've only had 1 acquaintance or friend in my life	1	0
8	In my life I've not been very interested in sexual contact	1	0
9	In the past I've often taken tranquilizers and/or sleeping pills	1	0

premorbid personality of the older adult from the informant's perspective. The 62 age-neutral items address 10 content scales: socially avoidant behavior, uncertain behavior, vulnerability in interpersonal relationships, somatizing behavior, disorderly behavior, rigid behavior, perfectionistic behavior, antagonistic behavior, self-satisfied behavior, and unpredictable and impulsive behavior. HAP scales and items are based on the biosocial-learning model (Millon, 1996; Millon & Everly, 1985). Each statement is written in past tense, third person singular and is scored on a three-point Likert scale with responses of 'no,' 'more or less,' and 'yes.' To control for possible confounding as a result of the informants' feelings of sympathy or antipathy toward the patient, items are formulated to assess and correct for positive and negative response tendencies. The HAP has good internal consistency, good-to-excellent inter-rater reliability, good test–retest reliability, and good construct and concurrent validity (Barendse, Thissen, Oei, Rossi, & van Alphen, 2013). In the current sample, the internal consistency reliabilities (AIC) ranged from .19 with $\alpha = .47$ (rigid behavior) to .38 with $\alpha = .69$ (disorderly behavior), and with Cronbach's $\alpha = .84$ for the total scale.

In this study, the HAP, which relies solely on informant information, served as the reference criterion. According to the HAP manual, four high/very high subscale scores or the combination of two high/very high subscale scores and six above average subscale scores indicate the presence of a PD (Barendse & Thissen, 2006, p. 65).

Procedure

Prior to data collection, the patient, informant, and researcher signed the consent form. The informant then received a set of questionnaires and was asked to complete them in a separate room to prevent response bias.

In an interview with the patient, demographic information was collected. Next, several questionnaires were administrated to ensure that the exclusion criteria were not met. Finally, the GPS-pv was completed by the patient.

In order to determine the test–retest reliability, a follow-up assessment consisting of the GPS-pv or the GPS-iv and the HAP (for the informants) was planned for four weeks later.

Statistical analyses

All statistics were performed using the SPSS 19.0 package.

Reliability

Internal consistency was examined using both Cronbach's α and AIC. The AIC is considered to be superior to Cronbach's α because it measures internal consistency independently of the number of scale items. We considered an AIC above .15 as acceptable (Clark & Watson, 1995). Spearman's correlations (r_s) were calculated to determine the test–retest and inter-rater reliability of the GPS-pv and GPS-iv. The following rule was used to interpret the strength of the relationships: $r_s = .10-.29$ indicates a small effect, $r_s = .30-.49$ indicates a medium effect, and $r_s = .50-1.0$ indicates a large effect (Cohen, 1988a).

Diagnostic accuracy

The diagnostic accuracy of the GPS-pv and GPS-iv was assessed by calculating classical diagnostic validity statistics (Hsu, 2002), such as sensitivity (sens; the proportion of people with the disorder who are detected by the test),

specificity (spec; the proportion of people without the disorder who are correctly identified by the test), positive predictive power (PPP; the probability that the disorder is present given the test is positive), negative predictive power (NPP; the probability that the disorder is absent given the test is negative), and the overall correct classification (OCC; the proportion correctly classified) for each possible cutoff score. As PPP and NPP are influenced by the prevalence of the disorder (Meehl & Rosen, 1955), the prevalence rate was also reported. Additionally, the incremental validity of positive (IPPP) and negative (INPP) test scores was calculated; values greater than zero indicate that positive or negative test-based findings are more informative for diagnostic decisions than the disorder's prevalence rate (e.g., Rossi & Sloore, 2008; Streiner, 2003). For an example calculation, see Rossi and Sloore (2008).

The area under the curve (AUC) was retrieved by performing receiver-operating-characteristic curve (ROC) analyses. The associated curve represents the ratio of sensitivity and specificity, according to the cutoff score used to differentiate individuals with and without disorders. The AUC can be interpreted as the probability that the test will yield a higher value for a randomly chosen individual with the disorder than for a randomly chosen individual without the disorder (Agresti, 2002). An AUC value greater than .70 is considered to be fair to excellent (Bewick, Cheek, & Ball, 2004).

Predictive validity

To examine the value of the GPS items in predicting the presence or absence of PDs, binary logistic regression analyses using Garson's method for testing the difference between any two nested models (Garson, 2013, p. 84) were conducted. The items' predictive power was evaluated by calculating endorsement rates and Cohen's effect sizes (d) for the difference in response proportion (Cohen, 1988a). Subsequently, validity statistics were calculated for the most predictive items of the GPS-pv and GPS-iv.

Results

Reliability

Internal scale reliabilities

In the general practice setting, internal consistency as estimated by Cronbach's α was .48 (HAB), .61(BIO), and .63 (total) for the GPS-pv, and .57 (HAB), .58 (BIO), and .68 (total) for the GPS-iv, which are all below the minimum required level of .70 (DeVellis, 2003). However, the internal consistency based on AIC values was .12 (HAB), .16 (BIO), and .10 (total) for the GPS-pv, and .16 (HAB), .15 (BIO), and .12 (total) for the GPS-iv. The correlation between the GPS-pv BIO subscale and both GPS-iv subscales was above the minimum level of .15 (Clark & Watson, 1995).

Test–retest reliability

The GPS-pv test–retest reliability, as assessed with Spearman's correlations ($n = 104$), was $r = .56$ (HAB), $r = .67$ (BIO), and $r = .66$ (total), and all were significant ($p = .00$), indicating a large effect size (Cohen, 1988b, pp. 79–81). The GPS-iv test–retest reliability was $r_s = .52$ (HAB), .65 (BIO), and .68 (total). All the correlations were significant ($p = .00$) and showed a large effect (Cohen, 1988b, pp. 79–81).

Inter-rater reliability

The agreement between the GPS-pv and GPS-iv, as measured with Spearman's correlations, was $r = .32$ (HAB), $r = .54$ (BIO), and $r = .46$ (total), and all were significant ($p = .00$); the effect size ranged from medium to large (Cohen, 1988a).

Diagnostic accuracy

Table 3 shows the classical diagnostic validity statistics, chance-adjusted statistics, and odds measures for both the GPS versions for the different cutoff scores. These statistics show that for the GPS-pv, there was no optimal cutoff score to discriminate between the presence and absence of PDs. The choice of cutoff score is at the expense of either sensitivity or specificity. The results indicate that the GPS-pv is better at excluding than identifying PDs. For detecting the presence versus absence of a PD, an ROC curve was calculated using sensitivity and one-specificity values. The AUC for the GPS-pv was only .58 (95% CI [.48, .66]) and not significant ($p = .10$; 95% CI [.48, .66]).

For the GPS-iv, a cutoff score of ≥ 3 maximizes the sensitivity and specificity. The results indicated a sensitivity of .78 and a specificity of .65 (see Table 3). The AUC for the GPS-iv showed a fair diagnostic accuracy (.77; 95% CI [.69, .84]; $p < .00$).

Predictive validity

Logistic regression was performed to assess the impact of the GPS items on the likelihood that the patient had a PD. The model for the GPS-pv contained all 16 items as predictor variables. A total of 302 cases were analyzed, and the full model significantly predicted the PD status (omnibus chi-square = 32.88; $df = 16$; $p < .008$). The model accounted for between 10.3% and 18.0% of the variance in the PD status, and successfully predicted 98.4% of the patients without PDs. However, only 8.7% of the predictions of patients with PDs were accurate. Overall, 84.8% of the predictions were correct. Table 4 gives the coefficients, Wald statistic, associated degrees of freedom, and probability values for each of the predictor variables. These show that only the items HAB4 and BIO5 reliably predicted the PD status. The value of the coefficient for HAB4 reveals that patients who gave a positive response to this item were 8.5 times more likely to have a PD than those who gave a negative response. Patients who agreed with item BIO5 were approximately three times more likely to have a PD than those who disagreed.

A logistic regression analysis was performed with PD status as the dependent variable and the 16 GPS-iv items as predictor variables. The full model containing all

Table 3. Diagnostic validity statistics for the GPS-pv and the GPS-iv and for the most predictive items of the GPS-pv[1] and the GPS-iv[2].

	sens	spec	PPP	NPP	OCC	IPPP	INPP
GPS-pv							
Cutoff ≥00	1.00	.00	.15		.15	.00	
Cutoff ≥01	.83	.27	.17	.90	.36	.02	.05
Cutoff ≥02	.59	.57	.20	.88	.57	.04	.04
Cutoff ≥03	.30	.71	.16	.85	.65	.01	.00
Cutoff ≥04	.22	.84	.20	.86	.75	.05	.01
Cutoff ≥05	.17	.91	.26	.86	.80	.11	.01
Cutoff ≥06	.07	.95	.19	.85	.82	.04	.00
Cutoff ≥07	.04	.97	.20	.85	.83	.05	.00
Cutoff ≥08	.04	.98	.33	.85	.84	.18	.00
Cutoff ≥09	.02	1.00	.50	.85	.85	.35	.00
Cutoff ≥10	.00	1.00		.85	.85		.00
GPS-iv							
Cutoff ≥00	1.00	.00	.15		15	.00	
Cutoff ≥01	.94	.25	.18	.96	.35	.03	.11
Cutoff ≥02	.89	.47	.23	.96	.54	.08	.11
Cutoff ≥03	.78	.65	.28	.94	.67	.13	.10
Cutoff ≥04	.59	.79	.33	.91	.76	.18	.60
Cutoff ≥05	.44	.88	.39	.90	.81	.24	.05
Cutoff ≥06	.39	.91	.45	.89	.83	.30	.05
Cutoff ≥07	.26	.96	.52	.88	.85	.37	.03
Cutoff ≥08	.11	.98	.56	.86	.85	.40	.01
Cutoff ≥09	.09	1.00	1.00	.86	.86	.85	.01
Cutoff ≥10	.07	1.00	1.00	.86	.86	.85	.01
Cutoff ≥11	.00	1.00		.85	.85		.00
GPS-pv[1]							
Cutoff ≥00	1.00	.00	.15		.15	.00	
Cutoff ≥01	.44	.84	.32	.89	.78	.17	.04
Cutoff ≥02	.02	.99	.33	.85	.84	.18	.00
GPS-iv[2]							
Cutoff ≥00	1.00	.00	.15		.15	.00	
Cutoff ≥01	.91	.52	.25	.97	.58	.10	.12
Cutoff ≥02	.65	.86	.45	.93	.83	.30	.08
Cutoff ≥03	.28	.98	.68	.88	.87	.53	.04
Cutoff ≥04	.07	1.00	1.00	.86	.86	.85	.01

Notes: Base rate of PD is 15.23% (46/302); [1]most predictive items of the GPS-pv are HAB4 and BIO5; [2]most predictive items of the GPS-iv are HAB2, HAB4, HAB6, and BIO1.

Table 4. Logistic regression analyses predicting the likelihood of having a pd based on the GPS items.

	B	SE	Wald	*p*	Odds ratio	95% CI
GPS-pv						
HAB 1	−1.06	0.89	1.44	.23	0.35	[0.06, 1.96]
HAB 2	0.07	0.49	0.02	.89	1.07	[0.41, 2.80]
HAB 3	−0.23	0.48	0.22	.64	0.80	[0.31, 2.06]
HAB 4	2.14	0.52	16.89	.00	8.51	[3.06, 23.62]
HAB 5	−0.33	0.38	0.76	.38	0.72	[0.34, 1.51]
HAB 6	−0.31	1.45	0.05	.83	0.73	[0.04, 12.51]
HAB 7	−0.54	0.61	0.78	.38	0.59	[0.18, 1.92]
BIO 1	0.75	0.49	2.38	.12	2.12	[0.82, 5.49]
BIO 2	−1.34	1.29	1.07	.30	0.26	[0.02, 3.32]
BIO 3	−0.10	1.09	0.84	.36	0.37	[0.04, 3.10]
BIO 4	0.89	0.49	3.32	.07	2.42	[0.94, 6.28]
BIO 5	1.13	0.50	5.07	.02	3.09	[1.16, 8.24]
BIO 6	0.53	1.37	0.15	.70	1.70	[0.12, 24.77]
BIO 7	1.17	0.79	2.21	.14	3.22	[0.69, 15.01]
BIO 8	−0.89	0.68	1.74	.19	0.41	[0.11, 1.54]
BIO 9	−1.44	1.01	2.03	.16	0.24	[0.03, 1.72]
Constant	−2.03	0.27	56.07	.00	0.13	
GPS-iv						
HAB 1	−0.26	0.61	0.18	.67	0.77	[0.24, 2.56]
HAB 2	1.52	0.45	11.56	.00	4.55	[1.90, 10.91]
HAB 3	0.02	0.48	0.00	.96	1.02	[0.40, 2.61]
HAB 4	1.35	0.50	7.32	.01	3.85	[1.15, 10.24]
HAB 5	0.17	0.41	0.16	.69	1.18	[0.53, 2.63]
HAB 6	1.56	0.50	9.70	.00	4.75	[1.78, 12.66]
HAB 7	0.37	0.51	0.53	.47	1.45	[0.54, 3.93]
BIO 1	1.64	0.46	12.48	.00	5.14	[2.07, 12.76]
BIO 2	−1.08	1.04	1.09	.28	0.34	[0.05, 2.58]
BIO 3	−1.56	0.96	2.65	.10	0.21	[0.03, 1.38]
BIO 4	0.19	0.45	0.18	.67	1.21	[0.50, 2.94]
BIO 5	0.82	0.65	1.61	.21	2.23	[0.64, 8.13]
BIO 6	1.87	1.48	1.59	.21	6.46	[0.36, 117.57]
BIO 7	−0.07	0.63	0.01	.92	0.94	[0.27, 3.23]
BIO 8	−0.56	0.57	0.97	.33	0.57	[0.19, 1.75]
BIO 9	−0.45	0.72	0.39	.53	0.64	[0.16, 2.61]
Constant	−3.64	0.43	72.42	.00	0.03	

Note: d$f = 1$.

predictors was statistically significant ($n = 302$; omnibus chi-square = 71.80; d$f = 16$; $p < .000$), indicating that the model was able to distinguish between informants who did and did not report the presence of a PD in their patients. The model as a whole explained between 21.2% and 36.9% of the variance in the PD status, with an overall correct classification of 88.7%. Of the patients without a PD, 98.4% was accurately predicted. However, only slightly more than one-third (34.8%) of the predictions of the presence of a PD were correct. As shown in Table 4, 4 of the 16 items made a unique, statistically significant contribution to the model (HAB2, HAB4, HAB6, and BIO1). The coefficients of these four items all show that if the informant responded positively to them, patients were four to five times more likely to have a PD.

For both the GPS-pv and GPS-iv, additional analyses were performed. These consisted of logistic regression analyses using the likelihood ratio test for testing the difference between two nested models (Garson, 2013, p. 84) and the calculation of endorsement rates. For both questionnaires, these analyses identified the same items as the biggest contributors to PD predictions (Supplementary tables, available online), except for BIO5 on the GPS-pv, which failed to show a large distinction.

Logistic regression analyses identified GPS-pv items HAB4 and BIO5 and GPS-iv items HAB2, HAB4, HAB6, and BIO1 as those with the most predictive value. Validity statistics were also calculated for these items. Table 3 shows the classical diagnostic validity statistics for the

shortened GPS-pv and GPS-iv assessments, given different cutoff scores.

The validity statistics for the two GPS-pv items show that although these combined items insufficiently identified patients with PDs, they were able to detect patients without PDs. In contrast to the original GPS-pv, the short version had a significant ($p = .00$) AUC (.64; 95% CI [.51, .73]). The validity statistics for the four GPS-iv items indicate that a cutoff score of ≥ 1 provides optimal screening for PDs given a sensitivity of .91 and a specificity of .52. The AUC was .82 (95% CI [.75, .89]) and was significant at $p = .00$.

Discussion

Empirical research on PDs in older adults is a young field, particularly in the area of general practice. This is the first study to assess the value of a screening instrument for PDs in older adults (the GPS) in Dutch general practice. The test–retest reliability of the GPS subscales and the total scale was strong and yielded significance for both versions. Based on the diagnostic accuracy statistics, the GPS-iv is preferable to the GPS-pv; sensitivity and specificity were 78% and 65%, respectively, for the GPS-iv and 83% and 27%, respectively, for the GPS-pv. This might be due to the nature of the applied reference criterion, which is purely informant-based and relies on just one diagnostic source. It is plausible that the GPS-pv would have performed better if the reference criterion also had included self-report. Thus, there is an urgent need for the development of a self-report reference criterion. Additionally, six GPS items (HAB4 and BIO5 from the GPS-pv and HAB2, HAB4, HAB6, and BIO1 from the GPS-iv) yielded the most predicted validity. However, the diagnostic accuracy of these items needs cross validation in community-dwelling older adult populations.

The present results differ from those of van Alphen and colleagues (2006). They found that the diagnostic accuracy of the GPS-pv was better than that of the GPS-iv. However, that study differed from the present study with respect to target population (psychiatric elderly outpatients versus community-dwelling older adults) and informants (exclusively children versus diverse, related informants). Another important difference relates to the reference criterion. In the present study, the GPS was validated with an informant-based instrument, whereas in van Alphen et al. (2006), the presence or absence of PD was based on a clinical diagnosis, which is independent of both the patient's and the informant's perspectives.

The results of the present study speak to the value of involving multiple sources of information because it included informant information in addition to self-report in the assessment of PDs in older adults. Relying solely on self-report has serious limitations (e.g., APA, 2014; Klonsky et al., 2002). By including informant information, some of the limitations can be overcome. First, the use of informant data solves issues related to the older adults' motivation and ability to report PDs. Individuals, including older adults, may be consciously or unconsciously driven to present themselves more favorably as a result of the socially undesirable characteristics associated with PDs (Cooper, Balsis, & Oltmanns, 2014). In addition, they may be unaware of their maladaptive behaviors due to the distorted self-perceptions that characterize many PDs (Segal et al., 2006) or the ego-syntonic nature of PDs (Klonsky et al., 2002). Second, older adults may lack the insight needed to accurately report their personality and their interpersonal relationships as a result of cognitive decline (including dementia) and unfamiliarity with assessment situations (APA, 2014). Third, informant data minimize the cognitive and somatic load associated with frailty in older adults. In addition, in general practice, usage of informant information is becoming more acceptable as a way to screen for dementia in older adults (Brodaty et al., 2002). The GPS-iv could have a similar role in the detection of PDs.

There are several limitations to this study. First, there was a low response rate. Although over 3500 older adults were invited to participate, only 30.4% returned the reply, and, of those, only 38.5% were willing to participate. Consequently, the results may not be representative of the target population. Nonetheless, recruitment of older adults in research studies is often difficult (Gregson et al., 1997). The low response rate in the present study is comparable to that in studies that used similar methods (e.g., Edelman, Yang, Guymon, & Olson, 2013).

Second, we used the HAP as the reference criterion. This instrument is solely based on informant information, and, therefore, is not ideal for validating a self-report screening instrument, such as the GPS-pv. However, when the current study began, no self-report-based reference criterion that specifically addresses older adults by taking the aging process and environmental related changes into account was available (Rossi, van den Broeck, Dierckx, Segal, & van Alphen, 2014). The HAP is one of the few validated and reliable instruments which addresses older adults' personality and uses age-neutral items to consider age-related issues (Barendse et al., 2013). Moreover, the HAP is a promising instrument for the measurement of PDs as defined by the DSM-5: an expert study showed that the HAP items largely adhered to the DSM-5 criteria (Barendse, Rossi, & van Alphen, 2014).

Third, collateral information was not used to confirm a clinical diagnosis. The PD construct is complex because PDs are multifaceted. Moreover, from a practical point of view, the presence of a clinical diagnosis of PDs in older adults is rare in general practice. A longitudinal, expert, and all data (LEAD) standard (Spitzer, 1983) for the assessment of PDs is preferable (van Alphen et al., 2012). The LEAD standard uses longitudinal data to ultimately reach a consensus diagnosis among clinicians. The longitudinal data are gathered from observational, biographic, informant, test, and file data sources and from staff experiences with the patient. Although LEAD standards are implemented in mental health care services, they are difficult to use in general practice. Thus, our use of the HAP as a reference criterion for a general-practice population was based on practical considerations. However, it would be interesting to include other data, such as the frequency

of GP attendance or history of mental-health consultations, as possible features of PDs in future studies (Jackson & Burgess, 2004).

Fourth, there was heterogeneity of informants with respect to nature and duration of their relationships to the patient. Although the majority of the patients in the present study selected their spouses as informants, some chose other people, such as their (grand-) children or children-in-law. The difference in generations and the nature and duration of the relationship between the patient and informant could influence the agreement levels; in other words, observations by the various informants may be drawn from non-equivalent perspectives or sources. An analysis of these differences would be interesting.

Fifth, the internal consistency of the GPS-pv, and to a lesser degree GPS-iv, was limited. Although the homogeneity of the GPS was also measured with AICs, which is preferable to Cronbach's α when using a small-scaled instrument (Clark & Watson, 1995), the GPS-pv total score did not reach the .15 criterion. This is likely due to the abstract and multifaceted nature of the PD construct. Nevertheless, the other scales did reach sufficient AIC.

Sixth, there was relatively low specificity of both the GPS versions when a cutoff score with good sensitivity was suggested. However, when using opportunistic screening or a case-finding instrument (such as the GPS), a relatively higher sensitivity is preferred over specificity. Moreover, currently, there are no alternative brief instruments for screening PDs in older adults. Thus, it is not realistic to expect both high sensitivity and specificity when screening of PDs in older adults. Similar research in young adults, using more comprehensive instruments, was unable to do so because of the complexity of the PD construct (Tyrer et al., 2007).

Seventh, the GPS is intended as a screening-process tool for GPs. It can only be used to screen for the presence or absence of a PD based on the a-theoretical model of DSM-5. Specific DSM-5 PDs cannot be screened with the GPS.

Thus, additional research is needed to further validate the utility of the GPS in general practice, especially for the GPS-pv. There is a high need to replicate both the present study and van Alphen and colleagues (2006), once a criterion that allows both self and informant scoring is available for older adults. It would also be worthwhile to further investigate the predictive value of the GPS items (HAB4 and BIO5 for the GPS-pv; HAB2, HAB4, HAB6 and BIO1 for the GPS-iv). Research on GPS implementation in general practice is also warranted. Moreover, in future research, it is important to be even more attentive to recruitment issues for the target population. Hand delivering, rather than mailing, participant invitations may increase the response rate (Edelman et al., 2013).

This first attempt to provide GPs an age-specific screening instrument for PDs is innovative. This research represents an important step forward for GPs who intend to screen for PDs in their older adult patients but who lack the resources and time to systematically conduct extended diagnostic procedures (e.g., the LEAD standard). Screening for PDs in older adults that is based on the general DSM-5 criteria is an important first step for general practice.

Acknowledgements

The authors thank Peter Bartholet for assisting in planning the home visits for data collection. We also thank Kyra Plitscher, Amanda Haagmans and Petra de Vries for data collection assistance.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Care for the Elderly Programme from the Netherlands Organisation for Health Research and Development (ZonMw) [grant number 311070201].

Supplemental data

Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/13607863.2015.1008989

References

Agresti, A. (2002). *An introduction to categorical data analysis* (2nd ed.). New York, NY: Wiley.

Agronin, M.E., & Maletta, G. (2000). Personality disorders in late life: Understanding and overcoming the gap in research. *American Journal of Geriatric Psychiatry, 8*(1), 4–18.

American Psychological Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.; text rev.). Washington, DC: Author.

American Psychological Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric.

American Psychological Association. (2014). Guidelines for psychological practice with older adults. *American Psychologist, 69*(1), 34–65.

Ames, A., & Molinari, V. (1994). Prevalence of personality disorders in community living elderly. *Journal of Geriatric Psychiatry and Neurlogy, 7*(3), 189–194.

Babor, T.M., Higgings-Biddle, J.C., Saunders, J.B., & Monteiro, M.G. (2001). *AUDIT the alcohol use disorders identification test. Guidelines for use in primary care* (WHO/MSD/MSB/01.6a) (pp. 1–40). Retrieved from http://whqlibdoc.who.int/hq/2001/WHO_MSD_MSB_01.6a.pdf

Balsis, S.M.A., Woods, C.M., Gleason, M.E.J., & Oltmanns, T. F. (2007). Overdiagnosis and underdiagnosis of personality disorders in older adults. *American Journal of Geriatric Psychiatry, 15*, 742–753.

Barendse, H., Rossi, G., & van Alphen, S.P.J. (2014). Personality disorders in older adults: Expert opinion as a first step toward evaluating the criterion validity of an informant questionnaire (HAP). *International Psychogeriatrics, 26*(1), 173–174.

Barendse, H., & Thissen, T. (2006). *H.A.P. Hetero-Anamnestische Persoonlijkheidsvragenlijst handleiding*. Schijndel: Barendse & Thissen.

Barendse, H., Thissen, A.J.C., Oei, T.I., Rossi, G., & van Alphen, S.P.J. (2013). Psychometric properties of an informant personality questionnaire (the HAP) in a sample of older adults in the Netherlands and Belgium. *Aging and Mental Health, 17*(5), 623–629. doi: 10.1080/13607863.2012.756458

Bewick, V., Cheek, L., & Ball, J. (2004). Statistic review 13: Receiver operating characteristic curves. *Critical Care, 8*(6), 508–512. doi: 10.1186/cc3000

Brodaty, H., Pond, D., Kemp, N., Luscombe, G., Harding, L., Berman, K., & Huppert, F. (2002). The GPCOG: A new screening test for dementia designed for general practice. *Journal of American Geriatrics Society, 50*(3), 530−534.

Centraal Bureau voor de Statistiek (CBS). *CBS list*. Retrieved from http://www.cbs.nl/

Clark, L.A., & Watson, D. (1995). Constructing validity. Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309−319.

Cohen, J. (1988a). Differences between proportions. In J. Cohen (Ed.), *Statistical power analysis for the behavioral science* (2nd ed., pp. 179−213). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988b). The significance of a product moment rs. In J. Cohen (Ed.), *Statistical power analysis for the behavioral science* (2nd ed., pp. 75−108). Hillsdale, NJ: Lawrence Erlbaum Associates.

Condello, C., Padoani, W., Uguzzoni, U., Caon, F., & De Leo, D. (2003). Personality disorders and self-perceived quality of life in an elderly outpatient population. *Psychopathology, 36*, 78−83. doi: 10.1159/000070362.

Cooper, L.D., Balsis, S.M. A., & Oltmanns, T.F. (2014). A longitudinal analysis of personality disorder dimensions and personality traits in a community sample of older adults: Perspectives from selves and informants. *Journal of Personality Disorders, 28*(1), 151−165.

De Beurs, E. (2008). *Brief symptom inventory handleiding* [Brief symptom inventory manual]. Leiden: PITS B.V.

DeVellis, R.F. (2003). *Scale development: Theory and application* (2nd ed.). Thousand Oaks, CA: Sage.

Edelman, L.S., Yang, R., Guymon, M., & Olson, L.M. (2013). Survey methods and response rates among rural community dwelling older adults. *Nursing Research, 62*(4), 286−292.

Edelstein, B.A., Woodhead, E.L., Segal, D.L., Heisel, M.J., Bower, E.H., Lowery, A.J., & Stoner, S.A. (2007). Older adult psychological assessment. *Clinical Gerontologist, 31*(3), 1−34. doi: 10.1080/07317110802072108.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189−198.

Frankenburg, F.R., & Zanarini, M.C. (2006). Personality disorders and medical comorbidity. *Current Opinion in Psychiatry, 19*(4), 428−431. doi: 10.1097/01.yco.0000228766.33356.44.

Garson, D.G. (2013). *Logistic regression: Binary & multinomial*. Retrieved from www.statisticalassociates.com.

Gómez, A., Conde, A., Santana, J.M., Jorrín, A., Serrano, I.M., & Medina, R. (2006). The diagnostic usefulnessof the AUDIT and the AUDIT-C for detecting hazardous drinkers in the elderly. *Aging and Mental Health, 10*(5), 558−561.

Gregson, B.A., Smith, M., Lecouturier, J., Rousseau, N., Rodgers, H., & Bond, J. (1997). Issues of recruitment and maintaining high response rates in a longitudinal study of older hospital patients in England-Pathways to care study. *Journal of Epidemiology and Community Health, 51*(5), 541−548.

Hsu, L.M. (2002). Diagnostic validity statistics and the MCMI-III. *Psychological Assessment, 14*(4), 410−422. doi: 10.1037//1040-3590.14.4.410

Jackson, H.J., & Burgess, P.M. (2004). Personality disorders in the community: Results from the Australian National Survey of Mental Health and Well-Being Part III. *Social Psychiatry and Psychiatric Epidemiology, 39*, 765−776. doi: 10.1007/s00127-004-0821-x

Klonsky, E.D., Oltmanns, T.F., & Turkheimer, E. (2002). Informant-reports of personality disorder: relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9*, 300−311.

Meehl, P., & Rosen, A. (1955). Antedecent probability and the efficiency of psychometric signs, patterns, or cutting points. *Psychological Bulletin, 52*(3), 194−216.

Millon, T. (1996). Personality prototypes and their diagnostic criteria. In T. Millon, & G.L. Kierman (Eds.), *Contemporary directions in psychopathology: Towards the DSM-IV* (pp. 671−712). New York, NY: Guilford Press.

Millon, T., & Everly, G.S. (1985). *Personality and its disorders: A biosocial learning approach*. New York, NY: Wiley-Interscience.

Oltmanns, T.F., & Balsis, S.M.A. (2011). Personality disorders in later life: Questions about measurement, course, and impact of disorders. *Annual Review of Clinical Psychology, 7*, 321−349. doi: 10.1146/annurev-clinpsy-090310-120435.

Rossi, G., & Sloore, H. (2008). Diagnostic efficiency of the Dutch language version of the Millon Clinical Multiaxial Inventory-III. In I. Halvorsen & S. Olsen (Eds.), *New research on personality disorders* (pp. 193−211). New York, NY: Nova Science Publishers Inc.

Rossi, G., van den Broeck, J., Dierckx, E., Segal, D., & van Alphen, S.P.J. (2014). Personality assessment among older adults: The value of personality questionnaires unraveled. *Aging & Mental Health, 18*(8), 936−940. doi: 10.1080/13607863.2014.924089.

Segal, D.L., Coolidge, F.L., & Rosowsky, E. (2006). *Personality disorders and older adults: Diagnosis, assessment and treatment*. Hoboken, NJ: John Wiley & Sons.

Silberman, CS., Roth, L., Segal, D.L., & Burns, W.J. (1997). Relationship between the Millon Clinical Multiaxial Inventory-ll and Coolidge Axis II Inventory in chronically ill older adults: A pilot study. *Journal of Clinical Psychology, 53*(6), 559−566.

Stevenson, J., Brodaty, H., Boyce, P., & Byth, K. (2011). Personality disorder comorbidity and outcome: Comparison of three age groups. *Australian and New Zealand Journal of Psychiatry, 45*(9), 771−779. doi: 10.3109/00048674.2011.595685.

Streiner, D.L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment, 81*(3), 209−219.

Tummers, J.H.A., Hoijtink, J.H.A., Penders, K.A.P., Derksen, J.L.L., & van Alphen, S.P.J. (2011). Screening items for personality disorders in older adults: A multi-center study of psychiatric inpatients and outpatients in the Netherlands. *Clinical Gerontologist, 34*(1), 34−44. doi: 10.1080/07317115.2011.524601.

Tyrer, P., Coombs, N., Ibrahimi, F., Mathilakath, A., Bajaj, P., Ranger, M., . . . Din, R. (2007). Critical developments in the assessment of personality disorder. *The British Journal of Psychiatry, 190*(49), S51−S59.

van Alphen, S.P.J. (2006). *Diagnostiek van persoonlijkheidsstoornissen bij ouderen. Een bijdrage aan de ontwikkeling van een screeningsinstrument* [Diagnosis of personality disorders in older adults. A contribution to the construction of a screening instrument] (Doctoral thesis). Radboud Universiteit Nijmegen, Nijmegen. Retrieved from http://repository.ubn.ru.nl/bitstream/handle/2066/27409/27409.pdf.

van Alphen, S.P.J., Bolwerk, N., Videler, A.C., Tummers, J.H.A., van Royen, R.J.J., Barendse, H., . . . Rosowsky, E. (2012). Age-related aspects and clinical implications of diagnosis and treatment of personality disorders in older adults. *Clinical Gerontologist, 35*(1), 27−41. doi: 10.1080/07317115.2011.628368.

van Alphen, S.P.J., Derksen, J.L.L., Sadavoy, J., & Rosowsky, E. (2012). Features and challenges of personality disorders in late life. *Aging and Mental Health, 16*(7), 805−810. doi: 10.1080/13607863.2012.667781.

van Alphen, S.P.J., Dijk, V., Videler, A.C., Rossi, G., Dierckx, E., Bouckaert, F., & Oude Voshaar, R. (2015). Personality disorders in older adults: Emerging research issues. *Current Psychiatric Reports, 17*, 538−545. doi: 10.1007/s11920-014-0538-9.

van Alphen, S.P.J., Engelen, G.J.J.A., Kuin, Y., Hoijtink, H.J.A., & Derksen, J.L.L. (2006). A preliminary study of the diagnostic accuracy of the Gerontological Personality Disorders Scale (GPS). *International Journal of Geriatric Psychiatry, 21*(9), 862–868. doi: 10.1002/gps.1572.

Weissman, M.M. (1993). The epidemiology of personality disorders: A 1990 update. *Journal of Personality Disorders, 7* (Suppl 1), 44–62.

Widiger, T., & Trull, T. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62*, 71–83.

Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., & Leirer, V.O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17*(1), 37–49.