Severity Indices of Personality Problems–Short Form in Old-Age Psychiatry: Reliability and Validity

Barbera E. van Reijswoud, Inge Debast, Arjan C. Videler, Gina Rossi, Jill Lobbestael, Daniel L. Segal & Sebastiaan P. J. van Alphen

To cite this article: Barbera E. van Reijswoud, Inge Debast, Arjan C. Videler, Gina Rossi, Jill Lobbestael, Daniel L. Segal & Sebastiaan P. J. van Alphen (2020): Severity Indices of Personality Problems–Short Form in Old-Age Psychiatry: Reliability and Validity, Journal of Personality Assessment, DOI: 10.1080/00223891.2020.1743710

To link to this article: https://doi.org/10.1080/00223891.2020.1743710

Published online: 08 Apr 2020.

Submit your article to this journal

Article views: 171

View related articles

View Crossmark data
Severity Indices of Personality Problems—Short Form in Old-Age Psychiatry: Reliability and Validity

Barbera E. van Reijswoud1, Inge Debast2, Arjan C. Videler3, Gina Rossi2, Jill Lobbestael4, Daniel L. Segal5, and Sebastiaan P. J. van Alphen

1Mondriaan Institute for Mental Healthcare, Heerlen, Netherlands; 2Department of Clinical and Lifespan Psychology, Free University Brussels, Brussels, Belgium; 3Breburg Institute for Mental Health Care, Breda, The Netherlands; 4Department of Clinical Psychology Sciences, Maastricht University, Maastricht, The Netherlands; 5Department of Psychology, University of Colorado

ABSTRACT
The Severity Indices of Personality Problems (SIPP; Verheul et al., 2008) is a popular self-report questionnaire that measures severity of maladaptive personality functioning. Two studies demonstrated the utility of the short form (SIPP–SF) among older adults but validation in clinical settings is lacking. Therefore, we examined the psychometric properties of the SIPP–SF in a large sample of older adult Dutch outpatients (N = 124; age range = 60–85 years, M = 69.8, SD = 5.3). The SIPP–SF domains showed good to excellent internal reliability (Cronbach’s α = .75–.91) and effectively discriminated between participants with and without a personality disorder, as assessed with the Structured Clinical Interview for DSM–IV Axis II Personality Disorders (SCID–II). Convergent validity of the SIPP–SF was examined with instruments for measuring personality pathology among older adults (Informant Personality Questionnaire [HAP]; Gerontological Personality Disorders Scale [GPS]). The GPS generally correlated with the SIPP–SF domains in expected directions, with small to large effect sizes. For the HAP, only 1 scale correlated with all SIPP–SF domains. No associations were found between the SIPP–SF and psychiatric symptomatology as measured by the Brief Symptom Inventory (BSI). The SIPP–SF appears to be a promising instrument for assessing maladaptive personality functioning among older adult outpatients.

The assessment of personality disorders (PDs) in later life, age 65 years and older, is known to be particularly problematic because many of the diagnostic criteria for PDs in the Diagnostic and Statistical Manual of Mental Disorders (5th ed. [DSM–5]; American Psychiatric Association, 2013) are not adequately attuned to older adults and the unique life context of old age (e.g., Rossi et al., 2014; Segal et al., 2006). As a result, almost a third of the PD symptoms defined in DSM–5 are expressed differently in later life (Balsis et al., 2007). This problem negatively affects the reliability, validity, and utility of the PD construct in older adults, and frequently leads to misdiagnosing PDs in later life (Debast et al., 2017).

The prevalence of PDs among older adults in the general population is estimated around 8% (Schuster et al., 2013). Among older psychiatric outpatients, prevalence rates between 5% and 33% have been reported and the prevalence of (comorbid) PDs in older psychiatric inpatients has been estimated between 7% and 80% (Van Alphen et al., 2012). These rates represent significant problems, because PDs in old age are associated with a lower quality of life and impaired relationships (Segal et al., 2006), greater psychiatric comorbidity (Schuster et al., 2013), and more medical treatment (Friedman et al., 2013). It is critically important to detect PDs in older people because there is accumulating evidence for the efficacy of psychotherapeutic treatment of PDs in adults (Cristea et al., 2017; Dixon-Gordon et al., 2011; Stoffers et al., 2012). Recently, two studies have supported the efficacy of schema therapy for reducing PD symptoms in older adults. Such findings have increased optimism among those working with patients with PDs in later life (Videler et al., 2014; Videler et al., 2018).

Because the age neutrality of many DSM PD criteria can be considered doubtful (Balsis et al. [2007] estimated that 29% of criteria display measurement bias), two age-specific personality measurement instruments have been developed for assessing PDs in geriatric psychiatry (Rossi et al., 2014). The first one is the Hetero-Anamnestische Persoonlijkheidsvragenlijst (Informant Personality Questionnaire; HAP; Barendse et al., 2013), which is a Dutch informant questionnaire. The HAP items are based on detecting premorbid maladaptive and dysfunctional personality traits in retrospect. The psychometric properties of the HAP are good (see Barendse et al., 2013). The second age-specific instrument is the Dutch Gerontological Personality Disorders Scale (GPS; van Alphen et al. 2006). The GPS is a screening instrument used to detect PDs in older adults, and thus more generally captures the presence of PD pathology. The GPS consists of

CONTACT Barbera E. van Reijswoud barbera.evanreijswoud@derooysewissel.nl FPP “De Horst,” Heerdenweg 25, 6224 LA Maastricht, The Netherlands.

© 2020 Taylor & Francis Group, LLC
both a patient version and an informant version. Sensitivity and specificity of the GPS in samples of older adults both in psychiatric and general practice populations have been reported to be reasonably good (Penders et al., 2015; Van Alphen et al., 2006). Although these two instruments can be used to screen for personality pathology in later life, neither were developed or validated for the use of detecting changes in components of personality functioning, for instance due to treatment, and hence cannot be used for assessing treatment efficacy in terms of personality functioning. The alternative model for PDs in Section III of the DSM–5 differentiates the severity of impaired personality functioning (Criterion A) from the presence of maladaptive personality traits (Criterion B). Criterion A is defined as impairment in the self and in the capacity for interpersonal functioning, which is dimensionally based. These core components of personality dysfunction have been found to discriminate between patients with and without a (traditionally diagnosed) PD (Berghuis et al., 2013). Criterion A describes a dimensional view of personality functioning. Indeed, this view fits with the finding that dimensional models are more useful than categorical ones for assessing dysfunctional traits and behavior patterns in older people with PDs (Van Alphen et al., 2013), because assessment of dimensional traits has been found to be less age biased (Oltmanns & Balsis, 2011). It should be noted, though, that not all researchers support the dimensional model (see, e.g., Shedler et al., 2010) but instead favor the traditional categorical model.

The Severity Indices of Personality Problems (SIPP–118; Viersprong, 2006) and its short form (SIPP–SF; Verheul et al., 2008) is a promising instrument for assessing DSM–5 Criterion A (Bastiaansen et al., 2013). The SIPP was developed to differentiate between normal and clinical populations and to measure structural personality changes in treatment studies. The instrument provides a set of five reliable and valid indexes of core components of (mal-)adaptive personality functioning that seems to be sensitive to change following treatment of patient populations (Johansen et al., 2016). The SIPP–118 has also shown to be a promising instrument for measuring personality pathology in adolescents (Feenstra et al., 2011). One main caveat, however, is that the SIPP–118 has not been formally validated in older adults. Moreover, self-report questionnaires that include a large number of items and detailed semistructured interviews are relatively time-consuming and intensive for older adults (Van Alphen et al., 2006). For these reasons, shorter versions of self-report questionnaires are preferable in older psychiatry and clinical geropsychology. The SIPP–SF has about half the number of items of the SIPP–118 (i.e., 60 items instead of the original 118), and it has been found to show good psychometric properties in a community sample with an overall mean age of 25 years (Ro & Clark, 2009).

Notably, two studies in community-dwelling older adults demonstrated the utility of the SIPP–SF. The first study demonstrated the construct validity of the SIPP–SF in both older and younger adults by demonstrating a factorial structure of five higher order domains (Rossi et al., 2017). In older adults, personality functioning, as measured by the SIPP–SF, was more strongly associated with pathological traits of the alternative model for PDs (namely psychotism, disinhibition, antagonism, and dissocial behavior) than in younger adults. The second study showed that the SIPP–SF was an age-neutral instrument for measuring three out of four domains (self-control, identity integration, and social concordance) of personality functioning that closely correspond to Criterion A of DSM–5. The SIPP–SF domains of self-control and identity integration capture the self-dimension, whereas the SIPP–SF domains social concordance and relational functioning capture the interpersonal dimension as represented by the levels of personality functioning of the DSM–5 Section III model (Debast et al., 2018). In both these studies, the SIPP–SF was compared with instruments that measure PDs as described in DSM–5 Section III, yet no previous study used a categorical instrument for assessing DSM Section II PDs, like the Structured Clinical Interview for DSM–IV Axis II personality disorders (SCID–II; First et al., 1997).

Because the SIPP appears to be a useful instrument for measuring personality pathology among adolescents and adults, it is important to investigate the utility of this instrument among older adults because this would offer an opportunity to measure core components of personality across the life span. The SIPP–SF seems to be a promising instrument to measure the core components of personality functioning in older adults and to measure changes in personality functioning due to treatment. Because research on the reliability and validity of the SIPP–SF in older adults is limited, and completely lacking in clinical samples, the aim of this study was to investigate the psychometric properties of the SIPP–SF in a clinical sample of older adults, namely psychiatric outpatients in the Netherlands. We examine (a) the reliability of the five domains of the SIPP–SF; (b) the criterion validity of the SIPP–SF by comparing scores on the SIPP–SF between patients with and those without a SCID–II PD diagnosis and assessing the nonredundant contribution of the scales in discriminating individuals with and without PDs; (c) associations between severity of personality pathology as measured by the SIPP–SF and DSM Section II PDs as measured by the SCID–II (categorical and dimensional); (d) the convergent validity with the SIPP–SF and instruments developed specifically for assessing personality pathology in older adults, by relating scores on the SIPP–SF with scores on the HAP and GPS; and (e) associations between SIPP–SF scores and psychiatric symptomatology as measured by the Brief Symptom Inventory (BSI; Derogatis, 1975).

**Method**

**Participants**

Patients were recruited from two mental health institutes in the Netherlands. The first sample was collected at the old age Psychiatry Department of Mondriaan, including the Clinical Center of Excellence for Personality Disorders in Older Adults. The second sample was collected at PersonaCura, an expertise center for PDs in later life of
Mental Health Center Breburg. The Medical Ethics Review Committee Zuyderland-Zuyd (METC-Z) gave approval for the research. The exclusion criteria were severe cognitive problems or dementia (defined as a Mini-Mental State Examination score \( \leq 24/30 \)), severe psychotic or bipolar problems, significant intellectual problems (an IQ measured or estimated as below 80), and the presence of alcohol or drug addiction or use during testing. In Sample 1, eight patients did not meet the criteria (too young, low intelligence, or withdrawing informed consent later on), resulting in 99 older adult patients with a diverse range of psychiatric problems, significant intellectual problems (an IQ measured or estimated as below 80), and the presence of alcohol or drug addiction or use during testing. In Sample 1, eight patients did not meet the criteria (too young, low intelligence, or withdrawing informed consent later on), resulting in 99 older adult patients with a diverse range of psychiatric problems, all of whom had at least one DSM-5 classification. In Sample 2, there were nine patients who did not meet the criteria (too young), resulting in 25 older adult patients who were referred for personality problems to Breburg Mental Health Center, as can be seen in Table 1. At the old-age Psychiatry Department of Mondriaan, the assessment battery included the SCID-II, SIPP-SF, HAP, GPS, and BSI. At Breburg, the same instruments were used with the exception of the BSI, which was not included. Of Sample 1, four participants did not complete the BSI. In both samples, 12 participants did not complete the GPS. In 42 cases no informant was available for the informant questionnaire (HAP). In both samples, gender was more or less equally divided, the majority was of average education, and most participants were married or living together. See Table 1 for further demographic information.

**Table 1. Overview of the samples, with a total of \( n = 124 \) participants.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>% female</th>
<th>Mean age (SD)</th>
<th>% educational level</th>
<th>% marital status</th>
<th>Assessment battery</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99</td>
<td>51.5</td>
<td>70.6 (5.3)</td>
<td>Low</td>
<td>Married/living together</td>
<td>SCID–II, SIPP–SF, HAP, GPS</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>60.0</td>
<td>66.9 (4.6)</td>
<td>Average</td>
<td>Single/divorced/widow(er)</td>
<td>SCID–II, SIPP–SF, HAP, GPS</td>
</tr>
</tbody>
</table>

Note. SCID–II = Structured Clinical Interview for DSM–IV Axis II Personality Disorders; SIPP–SF = Severity Indices of Personality Problems–Short Form; HAP = Hetero-Anamnestische Persoonlijkheidsvragenlijst (Informant Personality Questionnaire); GPS = Gerontological Personality Disorders Scale; BSI = Brief Symptom Inventory.

Respondents indicate the extent to which they agree with statements over a time frame of the last 3 months. The response categories range from 1 to 4 and are described as **fully disagree**, **partly disagree**, **partly agree**, or **fully agree**. An example of an item from the self-functioning scale is "Sometimes I get so overwhelmed that I can’t control my reactions.” An example of an item from the interpersonal functioning scale is "I tend to think of myself as a loner.” The scores are clustered into five domains, with higher domain scores indicating more adaptive functioning and lower scores indicating more maladaptive personality functioning. In a prior study, Cronbach’s alpha values ranged from \( \alpha = .83 \) (social concordance) to \( \alpha = .89 \) (self-control and identity integration; Ro & Clark, 2009).

The Dutch version of the SCID–II (First et al., 1997; Weertman et al., 2000) was used to formally diagnose DSM–5 Section II PDs. The SCID–II is a semistructured interview that includes coverage of all 10 specific DSM–5 PDs as well as PD not otherwise specified (other specified PD in DSM-5). The interview contains 134 open-ended questions and begins with questions about behavior and relationships of the patient. Thereafter, items assess the diagnostic criteria for each of the 10 standard PDs, organized one by one. For example, the SCID–II consists of questions like these: “When you are out in the public and see people talking, do you often feel that people are talking about you?” Each PD criterion is rated as 1 (absent or false), 2 (subthreshold), or 3 (threshold or true). In this study, all clinicians conducting SCID–II interviews were extensively trained to ensure the quality of interviewing. The training was provided by the main researcher by giving oral education about the instrument to the clinicians individually in 60-min sessions. Subsequently, all clinicians observed two interviews done by an experienced interviewer, and then all clinicians performed two interviews under supervision before doing the interviews independently. The main researcher was available for consultation for the clinicians during the study. The SCID–II has shown good interrater reliability for the presence or absence of a PD in previous research (Lobbestael et al., 2011), especially among trained interviewers.

The Dutch HAP (Barendse et al., 2013) is an informant questionnaire with items that are based on detecting...
premorbid maladaptive personality traits. The HAP was developed and validated for use in old-age psychiatry and nursing homes. The HAP consists of 62 items that are retrospectively assessed. Scores are provided on 10 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. There are three response categories: yes, more or less, and no. In the instructions of the HAP, a distinction is made between current psychiatric symptoms or problems and the person’s premorbid personality. The psychometric qualities of the HAP are good. The internal consistency of the 10 scales is good (αs = .63–.85), Akaike’s information criterion [AICs] = .23–.53); the interrater and test–retest reliabilities are good to excellent (interclass correlation coefficients [ICCs] = .60–.98); and the construct validity, as evidenced through factor analyses, showed the same factor structure in both nursing home residents and older adult psychiatric patient populations (Barendse et al., 2013). In this study, Cronbach’s alpha values ranged from α = .44 (unacceptable) to .80 (good). Additionally, the average interitem correlations (AIC) were calculated to correct for the small numbers of items in the subscales. An AIC above .15 is considered to be acceptable (Clark & Watson, 1995). All AICs were above .15, specifically ranging from .15 (Rigid Behavior) to .47 (Somatizing Behavior). Descriptive statistics for the HAP are provided in Table 2.

The GPS (Van Alphen et al., 2006) is a screening instrument to detect PDs in older adults. The GPS consists of a patient version and an informant version. Both versions consist of two scales: habitual behavior (GPS-HAB) and biographical information (GPS-BIO). The GPS-HAB scale assesses habitual behaviors that reflect the expression of a number of PD features. In the GPS-BIO scale important and recurrent events or decisions in life are linked to the presence or absence of DSM–5 PDs. The internal consistency (Cronbach’s alpha) of the two scales ranges from poor (GPS-HAB α = .57) to acceptable (GPS-BIO α = .77; Van Alphen, 2006). The test–retest reliability of the GPS-HAB and the GPS-BIO subscales were moderate (Spearman’s r = .72) and excellent (Spearman’s r = .89), respectively. Sensitivity and specificity of the GPS patient version in an older psychiatric outpatient population was shown to be fair with sensitivity and specificity levels around 70% (Van Alphen et al., 2006). In this study, only the GPS patient version was used.

In this study, Cronbach’s alpha for the total score was α = .71 (acceptable), for the GPS-HAB scale α = .52 (poor), and for the GPS-BIO α = .71 (acceptable). In addition, the average interitem correlation (AIC) was calculated for the subscales to correct for the different numbers of items in the subscales. AICs were .14 (GPS-HAB) and .20 (GPS-BIO). The AIC of the GPS-BIO scale was above the minimum level of .15 (Clark & Watson, 1995). Descriptive statistics for the GPS are provided in Table 2.

The Dutch version of the BSI (Derogatis, 1975; translated by De Beurs, 2006) was used to measure symptomatic distress. The BSI consists of 53 self-report items covering nine symptom dimensions: somatization, obsession-compulsion, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. The BSI also contains three global indexes of distress: Positive Symptom Distress Index, Positive Symptom Total, and Global Severity Index (GSI). Only the GSI was used in this study. The GSI is a measure for overall psychological distress reflecting the average score of all item responses. Scores range from 1 to 5, with higher scores indicating a higher level of psychological and emotional distress. The GSI was used because it integrates all scales with different kinds of symptoms and it is useful for measuring symptomatic distress for patients with diverse psychopathology as included in our study. A sample item on this measure is “feeling no interest in things.” Respondents rate each item for the past 7 days on a 5-point Likert scale: 0 (not at all), 1 (a little bit), 2 (moderately), 3 (quite a bit), and 4 (extremely). Reliability of the Dutch version is good (Cronbach’s α = .71–.85) and the factorial structure is comparable to that of the original version (De Beurs, 2006). In this study, Cronbach’s alpha of the GSI scale was excellent (α = .97). Descriptive statistics for the GSI are provided in Table 2.

### Statistical analyses

All statistical analyses were performed using SPSS 22.0. First, the internal reliability of the SIPP–SF scores was analyzed using Cronbach’s alpha, AIC, and interscale Pearson correlations (effect size r). Second, criterion validity of the SIPP–SF between patients with and those without a PD, as assessed by the SCID–II, was tested with independent sample t tests. Bonferroni correction was used to correct for familywise error rates. The significance level for the analyses was set at $p = .01 (.05/5)$. Effect sizes were computed by Cohen’s d. In addition, to examine the value of the SIPP–SF

| Table 2. Descriptive statistics for the HAP, GPS and BSI. |
|-------------|--------------------|-----------------|-----------------|-------|-------|-------|
| Scale       | Minimum | Maximum | M     | SD    | Variance |
| HAP          |         |         |       |       |         |
| SOC          | .00     | 10.00   | 2.96  | 2.73  | 7.42   |
| UNC          | .00     | 10.00   | 3.64  | 2.89  | 8.33   |
| VUL          | .00     | 12.00   | 5.83  | 3.31  | 10.95  |
| SOM          | .00     | 8.00    | 2.43  | 2.15  | 6.18   |
| DIS          | .00     | 8.00    | 2.19  | 2.41  | 5.82   |
| RIG          | .00     | 8.00    | 3.77  | 2.08  | 4.32   |
| PER          | .00     | 8.00    | 4.24  | 2.40  | 5.77   |
| ANT          | .00     | 18.00   | 6.09  | 4.04  | 16.34  |
| SEL          | .00     | 10.00   | 2.27  | 2.07  | 4.29   |
| GPS          |         |         |       |       |         |
| BIO          | .00     | 8.00    | 3.83  | 2.15  | 4.77   |
| HAB          | .00     | 7.00    | 3.00  | 1.68  | 2.81   |
| Total score  | .00     | 14.00   | 6.83  | 3.15  | 9.91   |
| BSI          | .02     | 10.00   | 2.96  | 2.73  | 7.42   |

Note. HAP = Hetero-Anamnestiche Persoonlijkheidsvragenlijst (Informant Personality Questionnaire); GPS = Gerontological Personality Disorders Scale; BSI = Brief Symptom Inventory; SOC = Socially Avoidant Behavior; UNC = Uncertain Behavior; VUL = Vulnerability in Interpersonal Relationships; SOM = Somatizing Behavior; DIS = Disorderly Behavior; RIG = Rigid Behavior; PER = Perfectionistic Behavior; ANT = Antagonistic Behavior; SEL = Self-Satisfied Behavior; UNP = Unpredictable and Impulsive Behavior.
scales in predicting the presence or absence of PDs, a binary logistic regression analysis was conducted. Third, associations between severity of personality pathology as measured by the SIPP–SF and by the SCID–II (categorical, as in number of diagnosable PDs and dimensional, as in the number of endorsed PD criteria) were analyzed by Pearson correlations (effect size $r$). Fourth, convergent validity of the SIPP–SF was evaluated by calculating Pearson correlations (effect size $r$) between the SIPP–SF and both the HAP scales and GPS scores. Finally, Pearson correlations (effect size $r$) were calculated to evaluate associations between SIPP–SF domain scales and psychiatric symptomatology using the GSI scale of the BSI.

## Results

The SCID–II findings showed that 93 participants were diagnosed with one or more PD(s), whereas 31 participants were not diagnosed with a PD. This means that 75% of the participants were diagnosed with one or more PD(s). About 40% had one PD, and 34.6% had two or more PDs. In the percentage of cases of PDs, the other specified PD (OSPD) was the most common diagnosis (60.5%). The most common specific PD diagnosis was obsessive-compulsive PD (16.9%), followed by borderline PD (14.5%), avoidant PD (13.7%), antisocial PD (7.3%), dependent PD (5.6%), narcissistic PD (4.0%), paranoid PD (3.2%), and schizoid PD and schizotypal PD (both 0.8%). No diagnosed cases of histrionic PD were found. Subsamples with specific PDs were too small to allow statistics for specific PDs. Therefore, the broad distinction between having a PD or not was used in further analyses. Besides PDs, several comorbid psychiatric problems were present in the sample. Specifically, the number of non-PD DSM-5 diagnoses varied from none to five. The most common diagnosis was depressive disorder.

### Research Question 1: Internal reliability of SIPP–SF scores

The five domains of the SIPP–SF scores showed Cronbach’s alpha values ranging from $\alpha = .75$ to .91 with a mean estimated alpha value of $\alpha = .82$ (see Table 3). The AIC ranged from .20 (relational capacities) to .46 (identity integration), with a mean AIC of .31. These values indicate acceptable to excellent reliability of all the domain scales. Intercorrelations between the domains ranged from $r = .39$ (between identity integration and social concordance) to .75 (between self-control and social concordance), with a median correlation of $r = .51$ (see Table 4). These positive correlations of medium and large effect sizes confirm the general homogeneity of the SIPP–SF.

### Research Question 2: The criterion validity by comparing scores on the SIPP–SF between patients with and those without a SCID–II PD diagnosis

The SIPP–SF scores on all domains showed statistically significant differences, at the .01 level, between patients with and without a PD, showing that patients with a PD scored lower on all scales compared to those without a PD (note that lower scores reflect greater maladaptive personality functioning). Effect sizes ($d$) ranged from .59 (social concordance) to .86 (relational capacities), indicating moderate to large differences, as can be seen in Table 3. Next, logistic regression was performed to assess the nonredundant impact of the SIPP–SF scales on the likelihood that the patient had a PD. The model for the SIPP–SF contained all five scales as predictor variables. A total of 124 cases were analyzed, and the full model significantly predicted the PD status: omnibus $\chi^2(5) = 23.346, p < .000$; Hosmer & Lemeshow $\chi^2(8) = 16.805, p < .05$. The model accounted for between 10.3% and 18.0% of the variance in the PD status, and successfully predicted 93.5% of the patients with PDs. In contrast, almost half (41.9%) of the predictions of patients without PDs were accurate. Overall, 80.6% of the predictions were correct, in comparison to 75.0% in the model only including the constant. Table 5 shows the coefficients, Wald statistic, and probability values for each of the predictor variables. These values show that none of the SIPP–SF scales showed nonredundant contributions in predicting the PD status.

### Research Question 3: Severity of personality pathology

As can be seen in Table 6, the number of DSM-5 PD criteria was negatively associated with SIPP–SF domains (medium effect sizes), showing a relationship between endorsed PD criteria and greater impairment of personality functioning. This was also seen for the number of present PDs.

---

### Table 3: Mean domain scores and standard deviations of patients without personality disorder (PD) and patients with PD as measured by the SCID–II, independent $t$-tests and effect size ($d$).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Patients without PD ($n = 31$)</th>
<th>Patients with PD ($n = 93$)</th>
<th>$t$ (134)</th>
<th>Effect size ($d$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>SD</td>
<td>$M$</td>
<td>SD</td>
</tr>
<tr>
<td>Self-control</td>
<td>.86</td>
<td>3.37</td>
<td>2.84</td>
<td>0.64</td>
</tr>
<tr>
<td>Identity integration</td>
<td>.91</td>
<td>3.12</td>
<td>2.52</td>
<td>0.75</td>
</tr>
<tr>
<td>Responsibility</td>
<td>.79</td>
<td>3.50</td>
<td>3.14</td>
<td>0.52</td>
</tr>
<tr>
<td>Relational capacities</td>
<td>.75</td>
<td>3.06</td>
<td>2.61</td>
<td>0.53</td>
</tr>
<tr>
<td>Social concordance</td>
<td>.79</td>
<td>3.35</td>
<td>3.04</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Note. Equal variances were assumed for all domains. SCID–II = Structured Clinical Interview for DSM–IV Axis II Personality Disorders. *$p < .01$ (using Bonferroni correction).

**Correlations are significant at the .01 level.**

### Table 4. Pearson correlations of the Severity Indices of Personality Problems—Short Form scales.

<table>
<thead>
<tr>
<th></th>
<th>Self-control</th>
<th>Identity integration</th>
<th>Responsibility</th>
<th>Relational capacities</th>
<th>Social concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-control</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Identity integration</td>
<td>.60**</td>
<td>—</td>
<td>.51**</td>
<td>—</td>
<td>.41**</td>
</tr>
<tr>
<td>Responsibility</td>
<td>.50**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Relational capacities</td>
<td>.55**</td>
<td>.61**</td>
<td>—</td>
<td>.41**</td>
<td>—</td>
</tr>
<tr>
<td>Social concordance</td>
<td>.75**</td>
<td>.39**</td>
<td>—</td>
<td>.53**</td>
<td>—</td>
</tr>
</tbody>
</table>

### Table 5. Logistic regression analyses predicting the likelihood of having a personality disorder based on the Severity Indices of Personality Problems—Short Form (SIPP–SF) scales.

<table>
<thead>
<tr>
<th>SIPP–SF scales</th>
<th>$B$</th>
<th>$SE$</th>
<th>Wald</th>
<th>$p$</th>
<th>Exp ($B$)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-control</td>
<td>−8.12</td>
<td>.641</td>
<td>1.605</td>
<td>.025</td>
<td>.444</td>
<td>[.126, 1.560]</td>
</tr>
<tr>
<td>Identity integration</td>
<td>−.246</td>
<td>.441</td>
<td>.312</td>
<td>.576</td>
<td>.782</td>
<td>[.330, 1.854]</td>
</tr>
<tr>
<td>Responsibility</td>
<td>−.663</td>
<td>.591</td>
<td>1.259</td>
<td>.262</td>
<td>.516</td>
<td>[.162, 1.640]</td>
</tr>
<tr>
<td>Relational capacities</td>
<td>−.832</td>
<td>.573</td>
<td>1.019</td>
<td>.346</td>
<td>.351</td>
<td>[.142, 1.338]</td>
</tr>
<tr>
<td>Social concordance</td>
<td>.250</td>
<td>.731</td>
<td>.117</td>
<td>.732</td>
<td>1.285</td>
<td>[.307, 5.381]</td>
</tr>
<tr>
<td>Constant</td>
<td>8.093</td>
<td>2.134</td>
<td>14.383</td>
<td>.000</td>
<td>3270.547</td>
<td></td>
</tr>
</tbody>
</table>

Research Question 4: Convergent validity with personality pathology measures for older adults

All domain scores of the SIPP–SF correlated negatively with the GPS total score and GPS-BIO scale score with small (social concordance), medium (self-control, responsibility, relational capacities), and large (identity integration) effect sizes, as can be seen in Table 6. For the GPS-HAB scale score, the significant associations with self-control, identity integration, and responsibility were of medium effect size, and with relational capacities of small effect size. The association with social concordance was nonsignificant for the GPS-HAB scale score.

Regarding the HAP, the majority of HAP scales did not significantly correlate with the SIPP–SF domains, as can be seen in Table 6. Only one scale of the HAP, Unpredictable and Impulsive Behavior, correlated negatively with all SIPP–SF domains, yet only two correlations showed a medium effect (self-control and social concordance). The HAP scale Antagonistic Behavior correlated negatively with three SIPP–SF domains, showing a medium effect (self-control, relational capacities, and social concordance).

Research Question 5: Associations with psychiatric symptomatology

The domain scores of the SIPP–SF were not significantly correlated with the GSI scale of the BSI, as can be seen in Table 6.

Discussion

The overarching aim of this study was to investigate the psychometric properties of the SIPP–SF for assessing components of personality functioning among older adult psychiatric outpatients in the Netherlands. We found an acceptable to excellent internal consistency of all five SIPP–SF domain scores in this sample. For all SIPP–SF domains, a statistically significant difference was found between patients with and those without a PD, as classified with the SCID–II. More specifically, patients with a PD scored lower on all SIPP–SF scales compared to those without a PD, which was expected as lower SIPP–SF domain scores reflect greater maladaptive personality functioning. Furthermore, the SIPP–SF showed good criterion validity in predicting a PD. This implies that the SIPP–SF can adequately differentiate between patients with and without a PD and that SIPP–SF scales are meaningfully related to personality pathology among older adult outpatients. Also, the SIPP–SF scales were associated with the severity of personality pathology, given the negative correlations of the scales with the number of PDs and with the number of DSM–5 PD criteria. Moreover, all SIPP–SF domains correlated with the GPS-BIO subscale and total score of the GPS patient version, a screening instrument for PDs in older adults. However, the GPS-HAB scale failed to show significant correlations with the social concordance subscale.

The SIPP–SF did not correlate with most scales of the HAP. An explanation might be that the HAP is filled in by an informant and not by participants themselves. Research has shown that the informant sees a person and his or her pathology different than the participant sees himself or herself. This can be due to a lack of insight that a person has in the effect of one's behavior on others. Another possible explanation might be the unwillingness to disclose on a questionnaire or in an interview (Cruitt & Oltmanns, 2018). The scales from the HAP (ANT and UNP) that showed a negative correlation with medium effect with some, but not all, domains of the SIPP–SF, are scales that belong to the “impulsive and frustration tolerance” profile (Barendse & Thissen, 2006). For this profile, it is described that most informants experience this behavior as egocentric and unpleasant and high scores might indicate an antisocial, borderline, or passive-aggressive PD (Barendse & Thissen, 2006). This corresponds with the finding that self-informant concordance on PD traits is highest for Cluster B pathology, excluding narcissism (Klonsky et al., 2002) and this might explain the correlations on these specific scales (ANT and UNP) with the SIPP–SF. In addition, the HAP uses a lot of behavior descriptions, whereas the SIPP–SF includes questions about feelings and cognitions about oneself and other persons. Thus, the use of different kinds of questions in the instruments might measure different aspects of personality.

Contrary to our expectations, no correlations were found between the BSI GSI and the SIPP–SF domain scores. This
means that personality pathology as measured by the SIPP–SF was not associated with overall psychological distress as measured with the BSI GSI. This is in contrast with the work of Feenstra and colleagues (2011), who found meaningful correlations between the SIPP–118 and the GSI scale of the Symptom Checklist–90 (from which the BSI is derived) in adolescents. The variance of the GSI score showed no evidence for restricted range that would impede the correlation coefficient. One reason for this discrepancy can be that the study of Feenstra et al. (2011) used a more heterogeneous sample comprising high school students and inpatients in addition to outpatients. In a more heterogeneous sample, psychiatric symptoms are expected to show more variation due to lower and higher symptomatology, respectively. Alternatively, the absence of correlations between the SIPP–SF and GSI might indicate that the SIPP–SF is not greatly influenced by having other psychiatric symptomatology, at least in this sample with high rates of PDs. It is possible that, in the absence of a PD, the SIPP–SF is more sensitive to psychological distress. Further research with inpatient and nonpatient samples is needed to clarify this issue.

Our findings, however, indicate that the SIPP–SF is overall a good instrument to assess the severity of impaired personality functioning in older adult outpatients. Notably, it can differentiate between having a PD or not. The importance of detecting PDs in old age has become increasingly important because there is recent evidence of the efficacy of schema therapy for reducing PD symptoms in older adults (Videler et al., 2014; Videler et al., 2018). The SIPP–SF is a relatively short instrument (60 questions vs. 118 in the full version) and it is known that many older patients have difficulties with long instruments (Rossi et al., 2014). As such, the SIPP–SF might be called an older adult “friendly” instrument for PDs.

Several strengths of this study should be noted. First, this study included a relatively large sample for research with older adult outpatients. Studies in older adults are known for high dropout rates and recruitment difficulties (Provencher et al., 2014). Moreover, there was a high rate of PDs in our sample. Our total sample consisted of 124 older patients, of whom 75% were diagnosed with a PD. This high rate of PDs can be explained by the fact that in one sample (Breburg) only participants with a suspected PD were included, and both participating psychiatric institutions are expertise centers for PDs in older adults. Therefore, more people with PDs are likely referred to both institutions. This prevalence rate, however, is within the limits described by Van Alphen et al. (2012). Second, it was a strength that, in this study, almost all specific PDs were represented, except histrionic PD. Nevertheless, the number of specific PDs was too small for further statistical analyses and only the distinction between having a PD or not was used in our analyses, therefore causing no problem for absence or overrepresentation of specific PDs. Third, it was a strength that the SIPP–SF was investigated with instruments that are known to be applicable for older adults, like the BSI, or specifically designed for age-specific personality assessment in old age, like the GPS and the HAP (Rossi et al., 2014). The final strength was the fact that we compared two clinical groups of older psychiatric outpatients, one with a PD and one without a PD. Both clinical groups thus experienced general psychiatric distress. Due to differentiation and comparison of these groups, we decreased the odds that the SIPP–SF is measuring general psychiatric distress.
Despite many strengths, several important limitations were present. First, our sample did not include participants from the general population, where psychiatric symptoms are expected to be lower, or inpatients, where psychiatric symptoms are expected to be acute and higher. This might have given a different view on the associations with symptoms are expected to be acute and higher. This might refer to work can be perceived as “not suitable” to some older people. For instance, in the SIPP–SF one question is, “At work I get easily irritated about other people’s ways of doing things.” One can argue whether these items should be rewritten to general situations for use with older adults to capture all components of personality pathology in later life. Moreover, the domain “relational capacities” was found to lack age-neutrality in recent research due to a different degree of expression of the same underlying construct in the older age group (Debast et al., 2018). Third, the GPS-HAB scale showed poor internal consistency in this sample. This poor internal consistency might be explained by the fact that the GPS-HAB scale assesses habitual behaviors and consists of a short list of expressions that relate to behavior linked to various PDs and therefore the items are not necessarily correlated. Fourth, one benefit of the SIPP is that it is designed to capture personality change during treatment, but follow-up data were not assessed in this study. It is therefore a limitation of this study that test–retest reliability was not obtained, and this should be evaluated in future research. A final limitation was that we did not include interrater reliability for the SCID–II diagnoses. Although our raters were highly trained, formal interrater reliability should be established.

Conclusion

The SIPP–SF is a highly promising instrument to be used in geriatric psychiatry and clinical geropsychology for the measurement of core components of (mal)adaptive personality pathology. Two main advantages of the SIPP–SF for application in later life are the relatively short form of the instrument, which makes it more suitable for use in older adults, and the close correspondence with the concept of the severity of impaired personality functioning as operationalized in Criterion A of DSM–5 Section III. After all, treatment of PDs primarily aims at improving personality functioning. The best way of assessing improvement in personality functioning is by using an instrument that is designed to measure this construct, such as the SIPP–SF. Several research questions remain to be answered. For clinical use of the SIPP–SF, a cutoff score for the domains might be useful to indicate pathological personality functioning. Furthermore, as the SIPP–118 is known to be able to measure the treatment efficacy for PDs in both adults and adolescents (Feenstra et al., 2011; Verheul et al., 2008), further research on the capability of the SIPP–SF to assess treatment effects in older adults is desirable. Hopefully, our results will stimulate further research on older adults with PDs, and on the formal assessment of personality functioning among older adults with varying degrees of PD symptoms.

ORCID

Arjan C. Videler http://orcid.org/0000-0002-2175-3453

References


