

Normative data for the Dutch version of the Penn State Worry Questionnaire

Colin van der Heiden*, Peter Muris**, Arjan E. R. Bos**, Henk van der Molen** and Martijn Oostra**

Worry is a common symptom in various psychiatric problems and the key symptom of generalised anxiety disorder (GAD). The Penn State Worry Questionnaire (PSWQ) is the most widely used self-report scale for measuring worry. The present study provides normative data for the Dutch version of the PSWQ for a large community sample and a clinically referred sample of patients with GAD. Norms are not only provided for the original 16-item version, but also for an abbreviated 11-item version, which only consists of the positively worded items and has been shown to be a promising alternative to the full-length version. The percentile scores obtained for the community sample and the clinical GAD sample did not show much overlap, and this appeared true for the full-length as well as the abbreviated version of the PSWQ. These normative data seem suitable for differentiating between normal and abnormal manifestations of worrying and for evaluating the efficacy of treatments for GAD. (*Netherlands Journal of Psychology*, 65, 69-75.)

Keywords: worry; generalised anxiety disorder; self-report scale; normative data

Worry is a common symptom in various psychiatric problems, including anxiety, mood, and eating disorders (Harvey, Watkins, Mansell, & Shafran, 2004). In addition, excessive and uncontrollable worrying is viewed as the cardinal feature of generalised anxiety disorder (GAD; American Psychiatric Association, 2000). To as-

* Outpatient Treatment Centre PsyQ, Rotterdam and Institute of Psychology, Erasmus University Rotterdam

** Institute of Psychology, Erasmus University Rotterdam

Correspondence to Colin van der Heiden, PsyQ, Van Vollenhovestraat 3 (2nd floor), 3016 BE Rotterdam, the Netherlands, e-mail: c.vanderheiden@rijnmond.psyq.nl

Submitted 11 February 2009; revision accepted 29 March 2009.

sess the frequency, intensity, and uncontrollability of worry and to evaluate the efficacy of treatments for GAD, psychometrically sound assessment tools are strongly needed, as well as representative normative data for such measures. In the case of worry and GAD, the Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990) has emerged as the most widely used self-report measure, in both research and clinical practice.

The PSWQ was originally developed as a unifactorial measure, with 11 positively worded items (e.g., 'I worry about projects until they are all done') and five negatively worded items (e.g.,

'I never worry about anything'). Factor analytic studies have indicated that a two-factor model (with the positively worded items loading on the first factor and the negatively worded items loading on the second factor) provided a better fit for the data than the hypothesised one-factor model (Brown, Antony, & Barlow, 1992; Stöber, 1995; Beck, Stanley, & Zebb, 1995; Van Rijsoort, Emmelkamp, & Vervaeke, 1999). However, more recent research has shown that the negatively worded items form a method factor that should be considered a statistical artifact and is not really meaningful (Brown, 2003; Hazlett-Stevens, Ullman, & Craske, 2004; Van der Heiden, Muris, Bos, & Van der Molen, submitted). As such, it has been argued that the negatively worded items may undermine the psychometric qualities of the PSWQ (Woods, 2006), and so it has been suggested to remove these items from the scale and to employ an abbreviated scale that only includes the positively worded items (Hazlett-Stevens et al., 2004; Fresco, Heimberg, Mennin, & Turk, 2002). There is indeed some evidence indicating that such a shortened version of the PSWQ is just as or even more reliable and valid than its full-length counterpart (Hazlett-Stevens et al., 2004; Fresco et al., 2002).

While many studies have investigated the reliability and validity of the PSWQ (see for a review: Startup & Erickson, 2006), research presenting normative data for this measure is sparse. So far, only one investigation has provided norms for the PSWQ in a community sample (Gillis, Haaga, & Ford, 1995). Further, normative data in clinical populations are limited to a handful of studies that reported cut-off scores that may be helpful for discriminating GAD patients from other patient groups (Behar, Alcaine, Zuellig, & Borkovec, 2003; Fresco, Mennin, Heimberg, & Turk, 2003). The present study was designed to provide normative data for the Dutch version of the PSWQ for a sample from the general population and a clinical sample of patients with GAD. Given its potential we also provided cut-off scores for the shortened 11-item version of the PSWQ, for which currently no normative data exist.

Method

Participants and procedure

In order to obtain a large and representative sample of the general population, we approached participants in three ways. A sample of 455 participants was drawn from the Dutch population by Flycatcher, a full-service online research company. Effort was made to match this sample to the demographic profile of the adult Dutch population. Participants completed the PSWQ as part of a larger survey in return for a small financial reward. In addition, 340 participants filled in the PSWQ, after the question-

naire had been distributed by e-mail within the personal networks of the first author and several colleagues, friends, and family, and within the networks of some of the participants themselves (i.e., snowball sampling method). Examples of networks where the PSWQ was administered were sport clubs (e.g., soccer club), church communities, and companies (e.g., cleaning company). Finally, 47 participants were approached at two shopping malls in Rotterdam, the Netherlands, with the request to complete the PSWQ. Participants in the last two groups did not receive a reward for their participation. The overall sample consisted of 842 participants (379 men and 463 women; mean age = 43.6 years, $SD = 15.8$; range 16-84 years), none of whom were actually being treated for a psychological disorder and thus were defined as 'non-clinical' (Kendall & Sheldrick, 2000; Sabshin, 1989).

Participants in the clinical sample of GAD patients completed the PSWQ as part of the standard assessment at PsyQ, a mental health care organisation specialised in cognitive-behaviour therapy. The diagnosis of GAD was established using the Structured Clinical Interview for DSM-IV (SCID I: First, Spitzer, Gibbon, & Williams, 1994), which was administered by a trained psychologist. In a second diagnostic interview, the diagnosis of GAD was confirmed by another experienced clinician. In total, 102 patients took part in this study (26 men and 76 women; mean age = 34.2 years, $SD = 10.5$, range 19-66 years).

Questionnaire

As already described in the introduction, the PSWQ was designed to assess the intensity, excessiveness and uncontrollability of worry (Meyer et al., 1990). Respondents are instructed to indicate for each of the 16 items how applicable they are to them, using a five-point scale ranging from 'not at all typical of me' to 'very typical of me'. A total score is calculated by summing all items, after recoding the scores on the five negatively worded items. As such, scores range from 16 to 80, with higher scores representing a stronger tendency to worry. Psychometric properties of the original English version are good (Meyer et al., 1990), and this is also true for the Dutch version of the scale (Kerkhof, Hermans, Figee, Laeremans, Pieters, & Aardema, 2000; Van Rijsoort, Vervaeke, & Emmelkamp, 1997; Van Rijsoort et al., 1999). In the present study, Cronbach's alpha values for the full-length PSWQ were 0.92 in the community sample and 0.83 in the clinical sample of GAD patients. For the shortened version these values were 0.93 and 0.87, respectively.

Results

The Statistical Package for the Social Sciences (SPSS) was used to compute mean scores, stan-

dard deviations, and percentile scores (i.e., cut-off scores for 10 equal groups) for the full-length PSWQ and the abbreviated 11-item PSWQ in both the community and the clinical GAD sample.

For the full-length version of the PSWQ, the range of total scores in the non-clinical sample was 17-80 ($M = 42.4, SD = 11.8$), for the abbreviated PSWQ the range was 11-55 ($M = 25.9, SD = 9.4$). Significant sex and age differences were found for both the full-length and the abbreviated version of the PSWQ. Participants younger than 45 years (PSWQ: $M = 43.1, SD = 11.8$; abbreviated PSWQ: $M = 26.6, SD = 9.3$) displayed significantly higher PSWQ scores than participants aged 45 and above (PSWQ: $M = 41.4, SD = 11.6$; abbreviated PSWQ: $M = 25.2, SD = 9.5$) [$t(840)$'s being 2.08, $p < 0.05$ and 2.27, $p < 0.05$ respectively]. Further, women (PSWQ: $M = 44.1, SD = 12.0$; abbreviated PSWQ: $M = 27.1, SD = 9.7$) reported significantly higher levels of worry than men (PSWQ: $M = 40.2, SD = 11.1$; abbreviated

PSWQ: $M = 24.5, SD = 8.8$) [$t(840)$'s being 4.89, $p < 0.01$ and 4.09, $p < 0.01$ respectively].

In the clinical sample of patients with GAD the range of total scores was 37-80 ($M = 67.1, SD = 8.8$) for the full-length PSWQ, and 17-55 ($M = 44.9, SD = 7.8$) for the abbreviated version. Here, no significant sex and age differences were found.

Percentile scores for the full-length PSWQ are presented for both the community and the clinical sample in Table 1. Age- and sex-specific norms for the community sample are displayed in Table 2. For the abbreviated PSWQ, these normative data are provided in Tables 3 and 4. It is noteworthy that for both versions of the PSWQ, there was hardly any overlap in the percentile scores for the community and clinical samples (see Tables 1 and 3). A score of 58 or 59 on the full-length version, and a score between 35 and 39 on the abbreviated scale appeared to differentiate between the two samples, and thus seems to indicate the boundary between normal and pathological worrying.

Community sample		GAD patients	
Percentiles	Cut-off scores	Percentiles	Cut-off scores
10	28		
20	32		
30	35		
40	38		
50	42		
60	45		
70	48		
80	52		
90	59	10	58
		20	60
		30	62
		40	66
		50	69
		60	71
		70	72
		80	75
		90	77

Table 2		Cut-off scores for ten equal groups of the PSWQ scores in various age- and sex-based sub-groups of the community sample			
		<i>Women</i>		<i>Men</i>	
<i>Percentiles</i>		< 45 (n = 275)	≥ 45 (n = 188)	< 45 (n = 184)	≥ 45 (n = 195)
10		29	27	28	28
20		34	32	32	30
30		37	36	34	33
40		41	40	36	36
50		44	43	39	39
60		48	46	42	42
70		51	48	45	45
80		56	52	49	48
90		61	60	56	54

Table 3		Cut-off scores for ten equal groups of the abbreviated PSWQ in a community sample (n = 842) and a sample of clinically referred GAD patients (n = 102)			
<i>Community sample</i>		<i>Clinically referred GAD patients</i>			
<i>Percentiles</i>	<i>Cut-off scores</i>	<i>Percentiles</i>	<i>Cut-off scores</i>		
10	14				
20	17				
30	20				
40	23				
50	25				
60	28				
70	31				
80	34	10		35	
90	39	20		39	
		30		42	
		40		45	
		50		46	
		60		49	
		70		50	
		80		51	
		90		53	

Percentiles	Women		Men	
	< 45 (n = 275)	≥ 45 (n = 188)	< 45 (n = 184)	≥ 45 (n = 195)
10	16	13	15	14
20	19	17	17	16
30	22	20	20	18
40	24	24	21	20
50	27	26	24	23
60	30	29	26	25
70	33	31	29	29
80	36	34	32	32
90	41	40	38	37

Discussion

Statistical significance testing is traditionally the predominant way in research evaluating the efficacy and effectiveness of treatments for psychological disorders. While it is important to demonstrate that an intervention yields reductions in symptoms at a 'beyond chance' level, it is also essential to quantify the magnitude and meaningfulness of this improvement, which is generally referred to as 'clinical significance'. A statistically significant result could represent only modest benefits from therapy if the within-group variability is small or the sample size is large (Gillis et al., 1995). Therefore, methods have been developed for measuring the extent to which treatments produce clinically significant benefits. For example, Jacobson and Truax (1991) introduced the Reliable Change Index, which basically considers a patient to be 'recovered' when he/she displays improvement to a statistically reliable degree. Kendall and Grove (1988) proposed another possibility to compare data from treated individuals to normative data, typically collected from community samples. Such comparisons answer the question to what extent the treated individual, at post-treatment, is clinically equivalent to the normative group. Typically, these methods require normative data for outcome measures which possess good reliability and validity.

The present study provides such normative data for the PSWQ, the most widely employed self-report scale for measuring worry, the key feature of GAD. For this purpose, the PSWQ was administered in a large community sample as

well as a clinically referred sample of patients with GAD, for which currently no such normative data are available. Normative data were not only provided for the original 16-item version, but also for an abbreviated version from which the five negatively worded items have been removed. This shortened version of the PSWQ has been shown to be a promising alternative for the full-length version (Hazlett-Stevens et al., 2004; Fresco et al., 2002), and might overcome some of the problematic aspects of the use of reversed items. That is, negatively worded items can reduce the reliability and validity of a scale, and frequently form a separate method factor that has no substantive meaning (Woods, 2006), which was also shown to be the case with the PSWQ (Brown, 2003; Hazlett-Stevens et al., 2004; Van der Heiden et al., submitted). Thus, for the assessment of treatment effects it seems most appropriate to employ the normative data of the abbreviated version of the PSWQ. Nevertheless, it is recommended to administer the full-length version, as the negatively worded items may be helpful to disrupt a potential response bias when completing the scale (Marsh, 1996).

Positive features of this study are the adequate sample size of the community sample and the use of a well-diagnosed clinical sample of adult patients with GAD, the disorder for which worry is the key symptom. Furthermore, this is the first study that provides cut-off scores for the PSWQ for a clinically referred sample of patients with GAD, and the first to provide normative data for the abbreviated version of the PSWQ for both a community sample and a sample of GAD

patients. Besides these strengths, a number of limitations should also be mentioned. First, the community sample may not be fully representative for the demographic profile of the adult population. Second, with the exception of the Flycatcher subsample of the community sample, we do not know the response rate in the other subsamples, so in this respect the representativeness of these populations is questionable. Nevertheless, it should be noted that the current cut-off scores of the PSWQ compare really well with those obtained in previous research (Behar et al.,

2003; Fresco et al., 2003; Gillis et al., 2005), and so it seems reasonable to conclude that the normative data as provided in the current study are suitable for identifying high worrying individuals and for interpreting treatment outcome results. Further, to assist clinicians in assessing treatment progress, on which they can base their decision to continue, adjust or terminate treatment, norm groups for both the full-length and the abbreviated PSWQ in the community sample are presented in Table 5.

Table 5 Norm groups for the full-length and the abbreviated PSWQ

Norm group	Total score full-length PSWQ	Total score abbreviated PSWQ
Very low	<29	<17
Low	29-38	18-23
Average	39-48	24-31
High	49-59	32-39
Very high	>59	>39

Acknowledgement

We would like to thank Ellen Beelen, Clarina Butter, Daniëlle Flint, Joey Frankhuizen, Nathalie Gouweleeuw, Joop van der Heiden, Marcel

de Koning, Patrick Polak, Andrea Rau, Ella de Stigter, Yvonne van der Straaten and Daniëlle Veldhoven for their help in collecting the data for this study.

References

- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed. Text Revision). Washington, DC: American Psychiatric Association.
- Beck, J. G., Stanley, M. A., & Zebb, B. J. (1995). Psychometric properties of the Penn State Worry Questionnaire in older adults. *Journal of Clinical Geropsychology*, *1*, 33-42.
- Behar, E., Alcaine, O., Zullig, A. R., & Borkovec, T. D. (2003). Screening for generalized anxiety disorder using the Penn State Worry Questionnaire: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *34*, 25-43.
- Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy*, *41*, 1411-1426.
- Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour Research and Therapy*, *30*, 33-37.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1994). *Structured Clinical Interview for DSM-IV Axis-I disorders, research version, patient edition* (SCID-I/P). New York: Biometrics Research Department, New York State Psychiatric Interview.
- Fresco, D. M., Heimberg, R. G., Mennin, D. S., & Turk, C. L. (2002). Confirmatory analysis of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *40*, 313-323.
- Fresco, D. M., Mennin, D. S., Heimberg, R. G., & Turk, C. L. (2003). Using the Penn State Worry Questionnaire to identify individuals with generalized anxiety disorder: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *34*, 283-291.
- Gillis, M. M., Haaga, D. A. F., & Ford, G. T. (1995). Normative values for the Beck Anxiety Inventory, Fear Questionnaire, Penn State Worry Questionnaire and Social Phobia and Anxiety Inventory. *Psychological Assessment*, *7*, 450-455.
- Harvey, A., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford: Oxford University Press.
- Hazlett-Stevens, H., Ullman, J. B., & Craske, M. G. (2004). Factor structure of the Penn State Worry Questionnaire: Examination of a method factor. *Assessment*, *11*, 361-370.

- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.
- Kendall, P. C., & Grove, W. M. (1988). Normative comparisons in therapy outcome. *Behavioral Assessment, 10*, 147-158.
- Kendall, P. C., & Sheldrick, R. C. (2000). Normative data for normative comparisons. *Journal of Consulting and Clinical Psychology, 68*, 767-773.
- Kerkhof, A., Hermans, D., Figeo, A., Laeremans, I., Pieters, G., & Aardema, A. (2000). De Penn State Worry Questionnaire en de Worry Domains Questionnaire: Eerste resultaten bij Nederlandse en Vlaamse klinische en poliklinische populaties (The Penn State Worry Questionnaire and the Worry Domains Questionnaire: First results in Dutch and Flemish clinical populations). *Gedragstherapie, 33*, 135-146.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*, 810-819.
- Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*, 487-495.
- Sabshin, M. (1989). Normality and the boundaries of psychopathology. *Journal of Personality Disorders, 3*, 259-273.
- Startup, H. M., & Erickson, T. M. (2006). The Penn State Worry Questionnaire (PSWQ). In: G. C. L. Davey, & A. Wells (Eds.), *Worry and its psychological disorders: Theory, assessment and treatment* (pp. 265-283). Chichester, England: Wiley.
- Stöber, J. (1995). Worrying: A comparison of three questionnaires concerning everyday worries. *Zeitschrift für Differentielle und Diagnostische Psychologie, 16*, 50-63.
- Van der Heiden, C., Muris, P., Bos, A. E. R., & van der Molen, H. T. (submitted for publication). Factor structure of the Dutch version of the PSWQ in a community sample and in a sample of patients with generalized anxiety disorder.
- Van Rijsoort, S., Vervaeke, G., & Emmelkamp, P. (1997). De Penn State Worry Questionnaire en de Worry Domains Questionnaire: Eerste resultaten bij een normale Nederlandse populatie (The Penn State Worry Questionnaire and the Worry Domains Questionnaire: First results in a normal Dutch population). *Gedragstherapie, 30*, 121-128.
- Van Rijsoort, S., Emmelkamp, P., & Vervaeke, G. (1999). The Penn State Worry Questionnaire and the Worry Domains Questionnaire: Structure, reliability and validity. *Clinical Psychology and Psychotherapy, 6*, 297-307.
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28*, 189-194.