

ORIGINAL PAPER

Peter Ventevogel · Gieljan De Vries · Willem F. Scholte · Nasratullah Rasa Shinwari · Hafizullah Faiz
Ruhullah Nassery · Wim van den Brink · Miranda Olf

Properties of the Hopkins Symptom Checklist-25 (HSCL-25) and the Self-Reporting Questionnaire (SRQ-20) as screening instruments used in primary care in Afghanistan

Accepted: 9 January 2007 / Published online: 13 February 2007

Abstract *Background* Recent epidemiological studies in Afghanistan using mental health questionnaires yielded high prevalence rates for anxiety and depression. *Objectives* To explore the validity in the Afghan cultural context of two mental health questionnaires, the Hopkins Symptom Checklist-25 (HSCL-25) and the Self-Reporting Questionnaire-20 (SRQ-20). *Methods* The two mental health questionnaires were compared against a 'gold standard' semi-structured psychiatric interview, the Psychiatric Assessment Schedule (PAS). All instruments were administered to a sample of 116 Pashto-speaking patients (53 men, 63 women) attending primary health care facilities in Eastern Afghanistan. *Results* Both HSCL-25 and SRQ-20 had modest properties to correctly identify mental disorders, with an AUC (area under the curve) of 0.73 and 0.72 respectively. The optimal cut-off points for this population are different from those often used in trans-cultural research. For women the optimal cut-off points are higher than usual (2.25 for the HSCL-25 and 17 for the SRQ-20). For men the cut-off point for the HSCL-25 is lower than usual (1.50) and for the SRQ-20 it was 10).

Conclusions This study underlines the necessity of validating instruments along with cultural context and gender. Earlier studies in Afghanistan may have overestimated the prevalence of mental disorders among women and underestimated the prevalence in men.

Key words screening – questionnaire – common mental disorders – primary care – gender differences – Afghanistan

Introduction

Several recent studies in Afghanistan have demonstrated very high levels of depressive and anxiety symptoms among the general population, especially in women [1, 22, 36]. These studies, as do many others in the aftermath of humanitarian emergencies in low income countries, used brief questionnaires administered by laypersons, to obtain a quick impression of the mental health status of the population. The authors were involved in one of these studies, a cross-sectional survey in which the Hopkins Symptom Checklist-25 (HSCL-25) was used [36]. The decision to use this instrument was pragmatic—people who are not trained mental health professionals can be easily instructed, and little time is lost to administration. These instruments were, however, not originally designed to distinguish between mental disorders and normal reactions to severe environmental stress, and have not been validated for use in Afghanistan. It remains unclear to what extent they can be used to estimate the prevalence of mental disorders in this context [4]. Therefore we felt the need to conduct this additional research. In this study, which should ideally have been done before the mentioned cross-sectional survey, we assessed the psychometric properties of

P. Ventevogel · N.R. Shinwari · H. Faiz
HealthNet TPO, Mental Health Programme
Jalalabad, Afghanistan

P. Ventevogel · G. De Vries · W.F. Scholte · W. van den Brink
M. Olf
Academic Medical Centre, Dept. of Psychiatry
Centre for Psychological Trauma
University of Amsterdam
Amsterdam, The Netherlands

P. Ventevogel (✉)
HealthNet TPO Burundi
1110 Bujumbura, Burundi
Fax: +257-242124
E-Mail: peterventevogel@yahoo.co.uk

R. Nassery
Directorate Mental Health
Ministry of Public Health
Kabul, Afghanistan

the Hopkins Symptom Checklist-25 (HSCL-25) and of another frequently used brief, lay-administered mental health questionnaires, the Self-Reporting Questionnaire-20 (SRQ-20), to detect persons with psychiatric disorders in a primary care setting in Eastern Afghanistan. As a gold standard we used a structured psychiatric interview conducted by a trained clinician from the same culture as the respondents, which is a procedure widely used in instrument validation in western and non-western populations [2, 5, 10, 11, 15, 19, 21, 25, 31, 33].

Methods

■ The study site and population

Nangarhar province in Eastern Afghanistan is ethnically dominated by the Pashtun, the largest and most conservative of Afghanistan's ethnic groups. The province is relatively well off economically due to its strategic location, near the Khyber Pass that connects Afghanistan with Pakistan. During the time of the Russian occupation of Afghanistan (1979–1987) the province was the scene of heavy resistance by *mujahedeen* forces. After the fall of the Soviet backed communist government, a prolonged fight for power broke out between several *mujahedeen* factions. During this period hundred of thousands of residents fled the ongoing destruction in the capital Kabul and were given temporary shelter in huge camps in Nangarhar. The period of Taliban government (1996–2001) was relatively peaceful but was characterized by considerable human rights violations and severe restrictions on the rights of women. During the US led attack on the Taliban in the fall of 2001, some parts of the province where training camps of Al Qaida were suspected (such as in the mountain region of Tora Bora) suffered from heavy bombardments. After the installation of the new government, attacks by insurgents on government, aid-organizations, and women's organizations continued.

■ Study design

In June 2004 we assessed five rural basic health centers in Nangarhar Province run by HealthNet TPO, a non-governmental organization specialized in health care and psychosocial assistance in post conflict areas. In each health facility a sample of persons older than 15 years was drawn from the registration book in which each patient's data has to be entered before being seen by the health care staff. Sampling ratios at the health centers differed due to variations in the number of people attending (average sampling ratio was 1:3). The purpose of the study was explained to each potential participant by the local study coordinator (NSR). The literacy rate of the Afghan population is very low: 28.1% for ages 15 and older in 2004 [38]. Therefore informed consent was obtained from each respondent by reading aloud an explanatory text and then asking for participation. Using verbal instead of written consent is often a necessity when conducting research in low-income countries with high rates of illiteracy and a high level of distrust toward signing documents.

The SRQ-20 and HSCL-25 were administered by trained lay interviewers of the same sex as the participant. Subsequently a mental health professional held a clinical interview with each participant, on the site. The interviewer used a semi-structured clinical psychiatric interview, which contained no information about the scores on the HSCL-25 and SRQ-20. The mental health professionals were all male and therefore accompanied by a female *chaperone* for those female participants who did not want to talk to a male in a one-to-one situation. This survey formed part of a larger research project for which formal review and approval has been

given by the medical ethical committee of the University of Amsterdam, Amsterdam, the Netherlands. Local program directors, their boards, and local authorities approved the research procedures, which were consistent with the Declaration of Helsinki [44].

■ Instruments

Hopkins Symptom Checklist-25 (HSCL-25)

The Hopkins Symptom Checklist-25, derived from the 90-item Symptom Checklist (SCL-90) [8], is a screening tool designed to detect symptoms of anxiety and depression. It is composed of a 10-item subscale for anxiety and a 15-item subscale for depression, with each item scored on a Likert scale from 1 (*not at all*) to 4 (*extremely*). The period of reference is the past month. Originally developed as a self-report symptom inventory it is often used as an interviewer administered scale in settings with non-literate populations. A cut-off point of 1.75 was found to be optimal in the UK [42]. In a clinic sample of Indochinese refugees in the USA this finding was replicated and this cut-off point became widely accepted in refugee settings and in cross-cultural research [25, 26]. Few validation studies of the HSCL-25 with non-western populations have been done. For Vietnamese refugees in the USA sensitivity and specificity for detection of DSM-III major depression were estimated as 88% and 73%, based on a cut-off score of 1.75 for 'caseness' [15]. In a population of human immuno-deficiency virus (HIV) positive pregnant women in Tanzania a significantly lower value (1.06 with sensitivity 89% and specificity 80%) was found to be the optimum cut-off point [19]. In Nepal, Thapa and Hauff [37] calculated a sensitivity of 87% and specificity of 60% against DSM-IV mild depression diagnosed using the Composite International Diagnostic Interview (CIDI; area under the curve of 0.79).

Self-Reporting Questionnaire-20 (SRQ-20)

The Self-Reporting Questionnaire-20 items (SRQ-20) was developed by the World Health Organization (WHO) as a screening tool for common mental disorders in primary health care settings, especially in developing countries [13]. When patients are literate it can be self-administered; but in developing countries it is usually administered by lay interviewers. The instrument consists of 20 yes/no questions about common mental health symptoms such as anxiety, depressive symptoms, and psychosomatic complaints. The SRQ-20 has been used in numerous settings [43]. Cut-off points vary considerably depending on setting (community, primary care, hospital) and culture. A cut-off point of 8 is widely used [14]. Among primary care attenders in India the most appropriate cut-off score was found to be 12 [43]. In a community sample in the Punjab province of Pakistan, a validation study led to a cut-off score of 9 [35]. As far as we know, no validation study for Afghanistan has been performed. The SRQ was administered among Afghan refugees in Pakistan using a cut-off point of 13; however, this value was not empirically validated among the study population [32].

Psychiatric Assessment Schedule (PAS)

A semi-structured psychiatric interview was conducted using the Psychiatric Assessment Schedule. This instrument uses selected questions from the Present State Examination (PSE [41]) as screening items, subsequently followed by the appropriate ICD-10 research diagnostic criteria. The PAS assesses systematically depressive disorders, anxiety disorders (obsessive-compulsive disorder, panic disorder, agoraphobia, social phobia, specific phobia, generalized anxiety disorder) and somatization disorder. The instrument has never been used in Afghanistan but is well known in the region through its use in various areas in neighboring Pakistan [16, 27–29, 34]. The original English version was expanded with a section for post-traumatic stress disorder (PTSD). This was done within the format of the original instrument, during a workshop of the Afghan

research team and the team from Rawalpindi Psychiatric Institute that had earlier developed Urdu and Chitrali PAS versions for Pakistan.

■ Instrument translation

All instruments were translated into Pashto as spoken in Nangarhar province. The HSCL-25 was translated in 2002, in preparation of the epidemiological survey [36], following the principles described by van Ommeren et al. [39] which include translation to Pashto and back translation to English by separate groups of bilingual clinicians. Discrepancies between the various translations were subjected to a panel discussion with the involved clinicians and translators. Each item was then examined to ensure face validity and a transfer of conceptual meaning. Final changes to the Pashto versions were made after field-testing that included focus group discussions. The SRQ-20 was translated according to the same thorough guidelines in 2003, and was subsequently used in training activities for primary health care staff.

■ Training of study staff

A team of two male and two female interviewers who had no background in mental health care administered the HSCL-25 and SRQ-20. Two of the interviewers had also participated in the earlier epidemiological study [36]. They received a 3-day training in Jalalabad in administering the instrument (elements of the training: ‘explanation and discussion of each item’, ‘learning how to interview with role playing and real patients’, and a ‘1-day field test’). The psychiatric interviews were carried out by five mental health professionals (subsequently referred to as ‘MHPs’) who were all native Pashto speakers. Three were Afghan medical doctors (NRS, HF, and RN) with extensive experience as mental health supervisors in the primary mental health care project of HealthNet TPO in Nangarhar; two were psychiatrists from Peshawar, Pakistan. A 4-day training in the use of the PAS was organized in the Institute of Psychiatry in Rawalpindi, Pakistan by two senior Pakistani psychiatrists and a Dutch psychiatrist (PV). The training consisted of detailed discussions of the different items of the instrument, (using the conceptual definitions of symptoms in the glossary of the PSE as a reference point), role-playing, and group interviewing of typical patients. Subsequently the inter-rater reliability among the mental health professionals was assessed through 52 independent interviews using 12 different Pashto-speaking patients from Lady Reading Hospital Postgraduate Medical Institute in Peshawar, Pakistan. The level of agreement among raters was high: 96% for the main diagnosis, and 78% for second diagnoses. Cohen’s kappa’s for diagnoses were calculated for different combinations of raters, and ranged from 0.76 to 1.00.

■ Statistical analysis

Factor analyses were performed to check the cross-cultural measurement equivalence of the HSCL-25 and the SRQ-20. Measurement equivalence is based on the concept of construct or theoretical validity, which is defined as the correlation of an observed variable with some theoretical construct (latent variable) of interest [18]. Measurement equivalence refers to the equivalence of theoretical validities across populations. Evidence for measurement equivalence is a psychometric prerequisite for the comparison of prevalence rates or mean scores of (sub)scales [7].

Factor analyses with principal axing factoring extraction were performed to uncover the covariances between items (e.g., latent constructs). To facilitate interpretation of the factor structures varimax rotations were performed on the initial factor solutions. The revealed factor structures were visually inspected.

We used independent *t*-tests to compare the continuous instrument scores between men and women. For the comparison of the number of psychiatric diagnoses between men and women, the

Pearson Chi-square statistic was used. The Pearson’s correlation coefficient was calculated to examine the relationship between the HSCL-25 and SRQ-20 scales.

Receiver operating characteristic (ROC) curves were used to explore the optimal cut-off scores for the HSCL-25 and the SRQ-20 in the current sample with the psychiatric diagnoses as established with the PAS as the gold standard. ROC curves plot the sensitivity (on the *y*-axis) against 1—specificity (on the *x*-axis) of each possible cut-off point. Each ROC curve is characterized by an area under the curve (AUC) indicating the overall accuracy of the questionnaire (over the whole range of possible cut-off points to distinguish correctly between a case and a non-case). The AUC can range from 0.0 to 1.0. An AUC of 1.0 indicates a perfect prediction, and 0.5 indicates that the ability of a questionnaire to correctly identify a case is equal to chance prediction. The AUC is used to compare the validity of the two screeners over the total range of scores. Calculations are based on the empirically derived values and not on interpolations from the binormal distribution. In addition, sensitivity, specificity, and AUC were calculated for the HSCL-25 scale and its subscales for depressive disorders and anxiety disorder, and the SRQ-20 scale.

Differences were considered statistically significant when the two-sided *P*-value < 0.05. To test whether the criterion value of the instrument at hand exceeded chance level (AUC > 0.5), a one-side *P*-value was used.

Results

■ Socio demographic data

The sample was composed of 116 patients (53 men; 63 women) visiting a health center. The male respondents ranged in age from 18 to 80 year-old with a mean of 33 years (SD = 14.8). Women respondents ranged in age between 17 and 57 years with a mean of 29 years (SD = 9.3). The sample differed in several aspects from the cross-sectional population sample of the epidemiological study [36]. The current sample consisted of health care users in the eastern part of the province where HealthNet TPO supports the governmental health services. Our sample is ethnically more homogeneous (100% Pashtun compared to 92% in the epidemiological survey), more rural, less well educated, and more often unemployed. Unsurprisingly, male and female respondents differed on some background variables such as educational level and occupation. On all other variables no statistically significant differences were found (see Table 1).

■ Outcomes on HSCL-25, SRQ-20, and PAS

Factor analysis of the HSCL-25 items (principal axing factoring extraction and varimax rotation) revealed a two-factor model (factors ‘depression’ and ‘anxiety’) that explained 48% of the items variances. A factor analysis of the SRQ-20 items revealed a two-factor model (factors ‘common disorders’ and ‘social disability’) that explained 39% of the items variances. The factor structure of the items of both instruments agreed with the factor structures reported in the literature. Thus we found support for the measurement equivalence of the HSCL-25 and the SRQ-20.

Table 1 Socio-demographic characteristics of the study sample ($N = 116$)

		No. (%) of respondents					
		Male ($N = 53$)		Female ($N = 63$)		Total ($N = 116$)	
		N	%	N	%	N	%
Age	17–24	18	(34)	12	(19)	30	(26)
	25–34	14	(27)	15	(24)	29	(25)
	35–44	10	(19)	12	(19)	22	(19)
	45–80	11	(21)	24	(38)	35	(30)
Marital status	Single	6	(11)	6	(10)	12	(10)
	Married	47	(89)	51	(81)	98	(85)
	Widowed once	0	(0)	6	(10)	6	(5)
Education	None	29	(55)	55	(89)	84	(73)
	Primary (0–6 years)	11	(21)	0	(0)	11	(10)
	Middle (6–9 years)	7	(13)	0	(0)	7	(6)
	High (>9 years)	3	(6)	0	(0)	3	(3)
Job	Religious school	3	(6)	7	(11)	10	(9)
	Yes	40	(76)	1	(2)	41	(36)
	No	10	(19)	0	(0)	10	(9)
	Student	3	(6)	0	(0)	3	(3)
Previous mental illness	Housewife	0	(0)	61	(98)	61	(53)
	Yes	11	(21)	22	(34.9)	33	(28)
	No	42	(79)	41	(65.1)	83	(72)

Table 2 presents results of the SRQ-20 and HSCL-25 including relevant subscales. We used independent t -tests to compare the continuous instrument scores between men and women. For the comparison of the number of psychiatric diagnoses between men and women, the Pearson Chi-square statistic was used. Mean scores on the HSCL-25 were 2.07 ($SD = 0.62$) for the total scale with significant gender differences (men: mean = 1.59; women: mean = 2.47; $P < 0.001$). For the SRQ-20 the mean for endorsed items was 16 (out of 20) for women and 9 for men. The correlation between the two screening instruments was 0.82,

indicating a high overlap between them. Table 3 gives an overview of the diagnosis with the PAS. Many men ($24/53 = 43\%$) and many women ($38/63 = 60\%$) had a common mental disorder, mainly depressive and anxiety disorders.

■ HSCL-25 and SRQ-20 as screener for psychopathology

The primary analysis focused on the ability of the HSCL-25 and SRQ-20 to detect any psychiatric dis-

Table 2 Outcomes on screening instruments HSCL-25, SRQ-20 for both males and females

	Male		Female		Total		Differences between men and women	
	Mean	SD	Mean	SD	Mean	SD	T	P -value
HSCL total	1.59	(0.40)	2.47	(0.46)	2.07	(0.62)	10.95	<0.001
HSCL general anxiety	1.65	(0.45)	2.66	(0.54)	2.19	(0.71)	10.76	<0.001
HSCL depression	1.55	(0.40)	2.35	(0.48)	1.99	(0.60)	9.54	<0.001
SRQ_20	9.25	(4.71)	16.38	(3.42)	13.12	(5.39)	9.43	<0.001

Table 3 Clinical psychiatric diagnoses with PAS for both males and females

	Male		Female		Total		Differences between men and women	
	N	%	N	%	N	%	χ^2	P -value
Depressive disorder	15	28.3	33	52.4	48	41.4	6.880	0.009
Anxiety disorder (incl. PTSD)	15	28.3	13	20.6	28	24.1	0.924	0.336
Post-traumatic stress disorder	1	1.9	3	4.8	4	3.4	0.715	0.398
Somatization disorder	2	3.8	0	0.0	2	1.7	2.419	0.120
Any disorder	23	43.4	38	60.3	61	52.6	3.306	0.069

Table 4 Properties of HSCL-25 and SRQ-20 to detect psychopathology as measured by clinical interview with PAS

Sex	Instrument/scale	AUC	LL	UL	Z-value	P-value
Male	HSCL-25 Total Scale	0.78	0.62	0.88	4.35	<0.01
	HSCL-25 Depression Scale	0.79	0.60	0.90	4.02	<0.01
	HSCL-25 Anxiety Scale	0.81	0.65	0.90	4.92	<0.01
	SRQ-20	0.74	0.56	0.85	3.31	<0.01
Female	HSCL-25 Total Scale	0.67	0.51	0.79	2.40	0.01
	HSCL-25 Depression Scale	0.65	0.49	0.77	2.09	0.02
	HSCL-25 Anxiety Scale	0.65	0.48	0.78	1.97	0.02
	SRQ-20	0.73	0.58	0.83	3.62	<0.01
Total	HSCL-25 Total Scale	0.73	0.62	0.81	4.78	<0.01
	HSCL-25 Depression Scale	0.74	0.64	0.82	0.52	<0.01
	HSCL-25 Anxiety Scale	0.61	0.49	0.71	2.03	0.02
	SRQ-20	0.72	0.62	0.80	4.79	<0.01

AUC: area under curve; LL: 95% confidence interval lower limit; UL: 95% confidence interval upper limit; Z-value: the Z-value compares the AUC to 0.5, since the AUC of a 'useless' criterion equals 0.5; P-value: one-sided probability value expressing whether the criterion is better than chance level

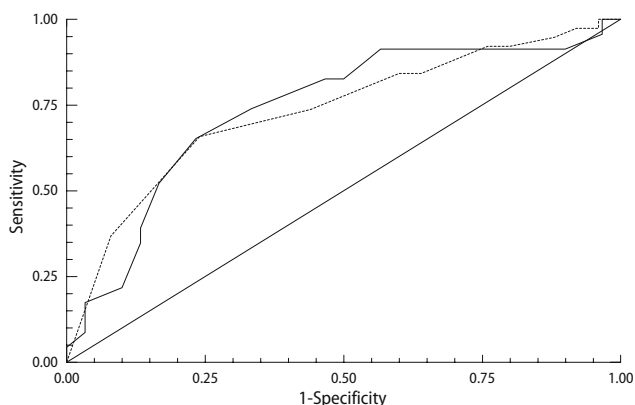


Fig. 1 ROC curve of the classification of any disorder for the SRQ-20. The *solid line* displays the curve found for male subjects, the *dashed line* displays the curve found for female subjects

order. Both HSCL-25 and SRQ-20 performed moderately; the area under the curve measured 0.73 for the HSCL-25 and 0.72 for the SRQ-20 (see Table 4, Figs. 1, 2). When analyzed separately for men and women the HSCL-25 showed a tendency to perform better in men (AUC 0.78) than in women (AUC 0.67). Additional separate analyses were also made for the HSCL-25 depression subscale to detect any depressive disorder, and HSCL-25 anxiety subscale to detect any anxiety disorder. For men, the SRQ-20 scale performed moderately (AUC 0.74), while the HSCL-25 depression and anxiety scales did reasonably well (respectively AUC 0.79, AUC 0.81). For women the AUCs for the HSCL-25 subscales are low (both AUC 0.65). For all analyses, except for that of the HSCL-25 anxiety scale ($P = 0.05$), the scales still differ significantly from chance level (AUC 0.5). Table 5 gives the sensitivity, specificity, positive and negative predictive values, kappa's, and percent agreement of the scales with different cut-off points. In evaluating the HSCL-25 and SRQ-20 as potential screeners for psychopathology the most appropriate cut-off points is a trade-off between a high sensitivity with an acceptable specificity. Depending on the purpose of the instru-

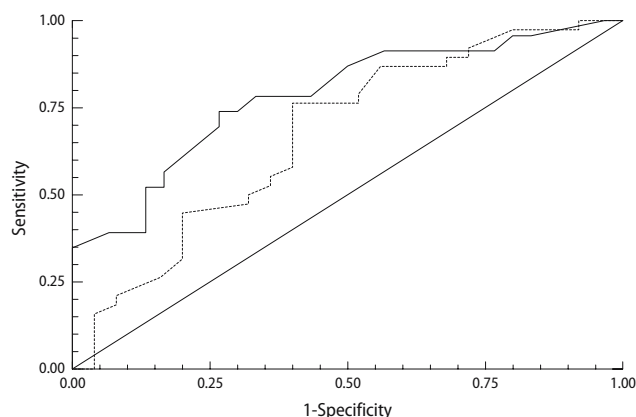


Fig. 2 ROC curve of the classification of any disorder for the HSCL-25. The *solid line* displays the curve found for male subjects, the *dashed line* displays the curve found for female subjects

ment (here used as a screening instrument) the optimal cut-off point is the one with a high rate of sensitivity and good specificity. For the HSCL-25 this optimal cut-off was 2.00 (sensitivity 0.69, specificity 0.67). However, when men and women were analyzed separately the cut-off point for men had to be lowered to 1.50; that for women had to be elevated to 2.25. The optimum cut-off point for the SRQ-20 was 10 for men and 17 for women.

Discussion

The aim of the study was to assess whether frequently used, easy to administer questionnaires were able to detect clinical cases of anxiety and depression. It proved difficult to identify an optimal cut-off point for the HSCL-25 or SRQ-20 with both a reasonable positive predictive value and a satisfactory degree of sensitivity. Our study corroborates the conclusions of a study with the HSCL-25 with Afghan refugees in Japan where the usual cut-off point of the HSCL-25 would have led to overestimation of the rate of

Table 5 Properties of HSCL and SRQ-20 with different cut-off points

Sex	Scale	Cut-off	Sens.	Spec.	PPV	NPV	κ	% Agree	Case
Male	HSCL-25	1.50 ^a	0.78	0.67	0.64	0.80	0.44	71.7	18/23
	Total Scale	1.75	0.39	0.87	0.69	0.65	0.27	66.1	9/23
	HSCL-25	1.25	0.93	0.24	0.33	0.90	0.13	45.3	14/15
	Depression Scale	1.50 ^a	0.80	0.70	0.52	0.90	0.44	73.5	14/15
		1.75	0.53	0.89	0.67	0.83	0.46	79.3	7/15
	HSCL-25	1.50	0.93	0.50	0.42	0.95	0.32	62.2	15/15
	Anxiety Scale	1.75 ^a	0.67	0.76	0.53	0.85	0.40	73.6	15/15
		2.00	0.47	0.92	0.70	0.81	0.43	79.2	10/15
	SRQ-20	9	0.83	0.53	0.58	0.80	0.34	66.0	19/23
		10 ^a	0.74	0.67	0.63	0.77	0.40	69.8	17/23
		11	0.65	0.77	0.68	0.74	0.42	71.7	15/23
	Female	HSCL-25	1.75	0.97	0.16	0.64	0.80	0.11	63.5
Total Scale		2.00	0.92	0.28	0.66	0.70	0.22	66.7	35/38
		2.25 ^a	0.79	0.48	0.70	0.60	0.28	66.6	30/38
HSCL-25		1.75	0.97	0.13	0.55	0.80	0.11	57.1	32/33
Depression Scale		2.00 ^a	0.91	0.37	0.61	0.79	0.28	65.1	30/33
		2.25	0.64	0.57	0.62	0.59	0.20	60.3	30/33
HSCL-25		1.75	1.00	0.10	0.21	1.00	0.01	26.9	12/13
Anxiety Scale		2.50	0.92	0.30	0.24	0.94	0.08	41.3	12/13
		2.75 ^a	0.58	0.60	0.26	0.86	0.10	58.7	12/13
SRQ-20		16	0.84	0.40	0.67	0.63	0.08	61.9	36/38
		17 ^a	0.74	0.56	0.72	0.58	0.30	66.6	28/38
		18	0.66	0.76	0.81	0.59	0.30	66.6	28/38
Total	HSCL-25	1.75	0.75	0.55	0.65	0.67	0.30	65.6	46/61
	Total Scale	2.00 ^a	0.69	0.67	0.70	0.66	0.36	68.1	42/61
		2.25	0.54	0.76	0.72	0.60	0.30	64.6	33/61
	HSCL-25	1.50	0.93	0.31	0.49	0.86	0.30	62.0	44/48
	Depression Scale	1.75 ^a	0.84	0.48	0.53	0.80	0.37	67.3	40/48
		2.00	0.70	0.65	0.59	0.75	0.41	70.7	35/48
	HSCL-25	1.75	0.86	0.30	0.28	0.88	0.11	48.3	22/28
	Anxiety Scale	2.00 ^a	0.75	0.43	0.29	0.85	0.13	54.3	19/28
		2.25	0.60	0.57	0.30	0.82	0.16	58.6	19/28
	SRQ-20	10	0.89	0.40	0.62	0.76	0.29	65.6	54/61
		11 ^a	0.84	0.47	0.64	0.72	0.31	66.4	51/61
		12	0.79	0.51	0.64	0.68	0.31	66.4	51/61

Sens: sensitivity is the probability of a positive test result given the condition is present; Spec: specificity is the probability of a negative test result given the condition is absent; PPV: positive predictive value is the probability that the condition is present given the test is positive; NPV: negative predictive value is the probability that the condition is absent given the test is negative; % Agree: proportion agreement in percentages

^aOptimal threshold for use as a screening instrument

depression among the respondents and the authors conclude that the usual cut-off points would not even have been acceptable for screening purposes [17]. The implication of our findings is that inferences made from checklist research with the HSCL-25 or SRQ-20 in Afghanistan need to be interpreted with caution. This again highlights the need to calibrate and evaluate screening instruments in cross-cultural epidemiological research.

Some might argue that the use of an ICD-10 derived diagnosis in itself lacks cultural validity. International and presumably 'universal' diagnostic categories such as the ICD-10 and DSM-IV are to a large extent embedded in the culturally bound Euro-American psychiatric conceptualization of mental disorders and thus might not capture culturally unique patterns of distress [9]. Our goal was to assess the accuracy of screening questionnaires to detect mental disorders as defined by the ICD-10. We have not assessed the cross-cultural validity of the ICD-10 classification itself. Such an undertaking would have required a different type of research based on in-

depth ethnographic fieldwork, which was not feasible in the context of the current research. However, we feel that a diagnosis made by well-trained clinicians, being ethnic Pashtuns themselves and familiar with the cultural ways of expressing distress among Pashtun tribesmen, guarantees a minimal level of cultural competency in the clinical assessment procedure.

An interesting finding is the differences between men and women with regard to the predictive value of scores on the questionnaires. The HSCL-25 and its subscales showed a tendency to perform worse for women than for men, while this could not be observed for the SRQ-20. A possible explanation is that the HSCL-25 compared to the SRQ-20 has fewer questions about somatic equivalents of mental distress (such as 'Is your appetite poor?' or 'Do you have uncomfortable feelings in your stomach') and questions about social disfunctioning (such as 'is your daily work suffering'). Socially or somatically oriented questions are, among the Pashtun, less sensitive to gender-specific interpretations than 'psychologically oriented' questions about 'feeling sad' or 'crying much'. Indeed,

a questionnaire for depression and anxiety that was constructed within neighboring Pakistan included more somatic items and items about social functioning [30]. Similarly, in a sample of Mongolian women in the reproductive age, the SRQ-20 performed better than the Edinburgh Postnatal Depression Scale (EPDS), which does not contain somatic items [31].

Optimizing the cut-off points for men and women separately had opposite effects: for men the cut-off point had to be lowered while for women it had to be elevated. This may be related to the strong differences within traditional Pashtun culture between male and female modes of expression, differences so strong indeed that some speak of a 'schism between men's and women's emotional worlds' [12]. We concur with Miller et al. [23] who suggest that the use of self-reporting questionnaires as read aloud by interviewers might lead to underreporting of mental distress in Afghan men who have the tendency to downplay the frequency of certain expressions of distress (e.g., crying) in order to save face in the eyes of the surveyors. This suggests that earlier epidemiological studies about common mental disorders among Pashtun in Afghanistan [1, 22, 36] might have overestimated the prevalence in women and underestimated the prevalence in men.

Several factors might have contributed to the rather limited discriminative ability of the HSCL-25 and SRQ-20 in this setting. First, the self-reported symptoms might not represent psychopathology but rather general psychological distress, as was suggested by Bolton and Betancourt [4]. Second, respondents might have aggravated the severity of symptoms on the questionnaires, hoping to get attention for their suffering and possibly hoping to get better treatment. Respondents may have answered affirmative if they perceived an advantage in being seen as ill. This has been suggested in studies with the SRQ-20, in Guinea Bissau [6] and Ethiopia [20]. In the latter study an analysis of the affirmative responses on SRQ-20 items showed a large percentage of the affirmative answers to be invalid due to linguistic problems, a lack of conceptual clarity or deliberately affirmative answers in order to gain something from it. Third, it may be that the psychiatric diagnoses derived by the mental health professionals were biased because of the lack of female professional clinician interviewers. It is possible that women who endorsed certain symptoms to a same sex lay interviewer showed constraint in doing so in the presence of a male clinician. Without a doubt the use of mental health professionals of the same sex as the participant would have improved the quality of the data. However, Pashto-speaking female mental health professionals are extremely rare; therefore this was not feasible.

Another striking feature of our sample of primary care attendees was the high percentage of people who had a mental disorder as assessed by a mental health professional. High prevalence rates for mental disorders among primary care patients have been found in

many cultural settings, but had not been established before in Afghanistan. The particularly high figures in our sample can probably, at least partly, be explained by the fact that the ICD-10 diagnostic criteria used do not require a significant social or occupational disability. Unfortunately, no data on associated social disability in this sample are available to put these figures in context.

Conclusion

The results of this study point to a rather limited usefulness of the HCSL-25 and SRQ-20 as screening tools for common mental disorders in the Afghan population. Earlier studies in Afghanistan using the HSCL-25 with the standard cut-off points might have overestimated the prevalence of mental disorder among women and underestimated the prevalence in men. The high prevalence of mental disorders in primary care patients in Nangarhar magnifies this problem since the properties of a screening instrument need to be extremely robust to be useful for diseases with high prevalence among the screening population. If the HCSL-25 and SRQ-20 would nevertheless be used as screening instruments different cut-off points for men and women have to be used. Rather than advocating the use of screeners to dichotomize the primary care attendees in 'probably mentally ill' and 'probably not mentally ill', we believe that overall attempts need to be made to increase the ability of primary health care staff in Afghanistan to identify depression and anxiety in their clinical encounters with patients. This involves training and supervision of Afghan doctors and nurses in primary mental health care skills, something that the Afghan government and institutional donors have high on their list of health care priorities [1, 24, 40].

■ **Acknowledgments** The authors would like to thank the interviewers, the management of HealthNet TPO in Jalalabad and the staff of the health care facilities where the study was based. We gratefully acknowledge the support of Prof. Saeed Farooq (Lady Reading Hospital, Postgraduate Medical Institute in Peshawar) and Dr. Khalid Saeed and Prof. Fareed Minhas (Institute of Psychiatry in Rawalpindi) for their help in the training and testing of the Psychiatric Assessment Schedule. Special thanks go to Drs. Shajaat Ali Khan and Mian Mokhtar ul-Haq, two Pakistani psychiatrists who were courageous enough to visit us in Afghanistan to conduct clinical interviews in the rural areas of Nangarhar Province. The Centers for Disease Control and Prevention (CDC; Atlanta, USA) provided funding to undertake this study.

References

1. Amowitz LL, Heisler M, Iacopino V (2003) A population-based assessment of women's mental health and attitudes toward women's human rights in Afghanistan. *J Womens Health* 12:577-587

2. Aydin N, Inandi T, Yigit A (2004) Validation of the Turkish version of the Edinburgh Postnatal Depression Scale among women within their first postpartum year. *Soc Psychiatry Psychiatr Epidemiol* 39:483-486
3. Baingana F, Bannon I, Thomas R (2005) Mental health and conflicts: conceptual framework and approaches. Worldbank, Washington
4. Bolton P, Betancourt TS (2004) Mental health in postwar Afghanistan. *JAMA* 292:626-8
5. Bressan RA, Chaves AC, Shirakawa I, et al. (1998) Validity study of the Brazilian version of the Calgary Depression Scale for Schizophrenia. *Schizophr Res* 32:41-49
6. De Jong JTVM (1987) A descent into African Psychiatry. Royal Tropical Institute, Amsterdam
7. De Jong JTVM, Komproe IH, Spinazolla J, et al. (2005) DESNOS in 3 postconflict settings: assessing cross-cultural construct equivalence. *J Trauma Stress* 18:13-21
8. Derogatis LR, Lipmann RS, Rickels K (1974) The Hopkins symptoms checklist (HSCL): a self-report inventory. *Behav Sci* 19:1-15
9. Draguns JG, Tanaka-Matsumi J (2003) Assessment of psychopathology across and within cultures: issues and findings. *Behav Res Therapy* 41:755-776
10. Eberhard-Gran M, Eskild A, Tambs K, et al. (2001) Review of validation studies of the Edinburgh Postnatal Depression Scale. *Acta Psychiatr Scand* 104:243-249
11. Fawzi MCS, Murphy E, Pham T, et al. (1997) The validity of screening for posttraumatic stress disorder and major depression among Vietnamese former political prisoners. *Acta Psychiatr Scand* 95:87-93
12. Grima B (1993) The performance of emotion among Paxtun women. Oxford University Press, Karachi
13. Harding TW, Arango MV, Baltazar J, et al. (1980) Mental disorders in primary health care: a study of the frequency and diagnosis in four developing countries. *Psychol Med* 10:231-242
14. Harpham T, Reichenheim M, Oser R, et al. (2003) Measuring mental health in a cost-effective manner. *Health Policy Plan* 18:344-349
15. Hinton WLN, Chen YC, Tran CG, et al. (1994) Screening for major depression in Vietnamese refugees: a validation and comparison of two instruments in a health screening population. *J Gen Intern Med* 9:202-206
16. Husain N, Creed F, Tomenson B (2000) Depression and social stress in Pakistan. *Psychol Med* 30:395-402
17. Ichikawa M, Nakahara S, Wakai S (2006) Cross-cultural use of the predetermined scale cutoff points in refugee mental health research. *Soc Psychiatry Psychiatr Epidemiol* 41:248-250
18. Irvine SH, Carroll WK (1980) Testing and assessment across cultures: Issues in methodology and theory. In: Triandis HC, Berry JW (eds) *The handbook of cross-cultural psychology*, vol. 2. Allyn and Bacon, Boston, pp 127-180
19. Kaaya SF, Fawzi MCS, Mbwambo JK, et al. (2002) Validity of the Hopkins Symptom Checklist-25 amongst HIV-positive pregnant women in Tanzania. *Acta Psychiatr Scand* 106:9-19
20. Kortmann F, Ten Horn S (1988) Comprehension and motivation in responses to a psychiatric screening instrument: Validity of the SRQ in Ethiopia. *Br J Psychiatry* 153:95-101
21. Leung K, Lue B, Lee M, et al. (1998) Screening of depression in patients with chronic medical diseases in a primary care setting. *Family Pract* 15:67-75
22. Lopes Cardozo B, Bilukha OO, Crawford CA, et al. (2004) Mental health, social functioning, and disability in postwar Afghanistan. *JAMA* 292:575-584
23. Miller K, Omidian, P, Quraishy AS, et al. (in press) The Afghan Symptom Checklist: a culturally grounded approach to mental health assessment in a conflict zone. *Am J Orthopsychiatry*
24. Ministry of Public Health (2005) Basic package of health services, Kabul, Afghanistan
25. Mollica RF, Wyshak G, de Marneffe D, et al. (1987) Indochinese versions of the Hopkins Symptom Checklist-25: a screening instrument for the psychiatric care of refugees. *Am J Psychiatry* 144:497-500
26. Mollica RF, Lopes Cardozo B, Osofsky HJ, et al. (2004) Mental health in complex emergencies. *Lancet* 364:2058-2067
27. Mumford DB, Nazir M, Jilani FM, et al. (1996) Stress and psychiatric disorder in the Hindu Kush: a community survey of mountain villages in Chitral, Pakistan. *Br J Psychiatry* 168:299-307
28. Mumford DB, Saeed K, Ahmad I, et al. (1997) Stress and psychiatric disorder in rural Punjab: a community survey. *Br J Psychiatry* 170:473-478
29. Mumford DB, Minhas FA, Akhtar I, et al. (2000) Stress and psychiatric disorder in urban Rawalpindi: community survey. *Br J Psychiatry* 177:557-562
30. Mumford DB, Ayub M, Karim R, et al. (2005) Development and validation of a questionnaire for anxiety and depression in Pakistan. *J Affect Disord* 88:175-182
31. Pollock JI, Manaseki-Holland S, Patel V (2006) Detection of depression in women of child-bearing age in non-western cultures: a comparison of the Edinburgh Postnatal Depression Scale and the Self-Reporting Questionnaire-20 in Mongolia. *J Affect Disord* 92:267-271
32. Rahman A, Hafeez A (2003) Suicidal feelings run high among mothers in refugee camps: a cross-sectional survey. *Acta Psychiatr Scand* 108:392-393
33. Rait G, Burns A, Baldwin R, et al. (2001) Screening for depression in African-Caribbean elders. *Family Pract* 16:591-595
34. Saeed K, Gater R, Hussain A, et al. (2000) The prevalence, classification and treatment of mental disorders among attenders of native faith healers in rural Pakistan. *Soc Psychiatry Psychiatr Epidemiol* 35:480-485
35. Saeed K, Mubbashar SS, Dogar I, et al. (2001) Comparison of Self-Reporting Questionnaire and Bradford Somatic Inventory as screening instruments for psychiatric morbidity in community settings in Pakistan. *J Coll Physicians Surg Pak* 11:229-231
36. Scholte WF, Olf M, Ventevogel P, et al. (2004) Mental health problems following war and repression in Eastern Afghanistan. *JAMA* 292:585-593
37. Thapa SB, Hauff E (2005) Psychological distress among displaced persons during an armed conflict in Nepal. *Soc Psychiatry Psychiatr Epidemiol* 40:672-679
38. UNDP (2006) Human development index, Country Table Afghanistan. Available at: http://hdr.undp.org/hdr2006/statistics/countries/data_sheets/cty_ds_AFG.html. Accessed December 18, 2006
39. Van Ommeren M, Thapa S, Makaju R, et al. (1999) Preparing instruments for transcultural research: use of the translation monitoring form with Nepali-speaking Bhutanese refugees. *Transcult Psychiatry* 36:285-301
40. Ventevogel P, Nassery R, Azimi S, Faiz H (2006). Psychiatry in Afghanistan. *Int Psychiatry* 3(2):10-12
41. Wing JK, Cooper JE, Sartorius N (1974). The measurement and classification of psychiatric symptoms. Cambridge University Press, London
42. Winokur A, Winokur DF, Rickels K, et al. (1984) Symptoms of emotional stress in family planning service: stability over a four-week period. *Br J Psychiatry* 144:395-399
43. World Health Organization (1994) A user's guide to the Self Reporting Questionnaire (SRQ), Geneva
44. World Medical Association (2001) Declaration of Helsinki: ethical principles for medical research involving human subjects. *Bull World Health Organ* 79:373-374