NormQuest

Reference Values for ROM Instruments and Questionnaires

Chapter 5

Reference values for anxiety questionnaires

Yvonne W.M. Schulte-van Maaren
Erik J. Giltay
Albert M. van Hemert
Frans G. Zitman
Margot W.M. de Waal
Ingrid V.E. Carlier

Journal of Affective Disorders (2013) 150 (3), 1008-1018

ABSTRACT

Background: The monitoring of patients with an anxiety disorder can benefit from Routine Outcome Monitoring (ROM). As anxiety disorders differ in phenomenology, several anxiety questionnaires are included in ROM: Brief Scale for Anxiety (BSA), PADUA Inventory Revised (PI-R), Panic Appraisal Inventory (PAI), Penn State Worry Questionnaire (PSWQ), Worry Domains Questionnaire (WDQ), Social Interaction, Anxiety Scale (SIAS), Social Phobia Scale (SPS), and the Impact of Event Scale-Revised (IES-R). We aimed to generate reference values for both 'healthy' and 'clinically anxious' populations for these anxiety questionnaires.

Methods: We included 1295 subjects from the general population (ROM reference-group) and 5066 psychiatric outpatients diagnosed with a specific anxiety disorder (ROM patient-group). The MINI was used as diagnostic device in both the ROM reference group and the ROM patient group. To define limits for one-sided reference intervals (95^{th} percentile; P_{95}) the outermost 5% of observations were used. Receiver Operating Characteristics (ROC) analyses were used to yield alternative cut-off values for the anxiety questionnaires.

Results: For the ROM reference-group the mean age was 40.3 years (SD=12.6), and for the ROM patient-group it was 36.5 years (SD=11.9). Females constituted 62.8% of the reference-group and 64.4% of the patient-group. P_{95} ROM reference group cut-off values for reference versus clinically anxious populations were 11 for the BSA, 43 for the PI-R, 37 for the PAI Anticipated Panic, 47 for the PAI Perceived Consequences, 65 for the PAI Perceived Self-efficacy, 66 for the PSWQ, 74 for the WDQ, 32 for the SIAS, 19 for the SPS, and 36 for IES-R. ROC analyses yielded slightly lower reference values. The discriminative power of all eight anxiety questionnaires was very high.

Limitations: Substantial non-response and limited generalizability.

Conclusions: For 8 anxiety questionnaires, the BSA, PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R, a comprehensive set of reference values was provided. Reference values were generally higher in women than in men, implying the use of gender-specific cut-off values. Each instrument can be offered to every patient with MAS disorders to make responsible decisions about continuing, changing or terminating therapy.

INTRODUCTION

Anxiety disorders are characterized by pervasive, persistent, anxious affective states. The DSM-IV recognizes various specific types of anxiety disorders: panic disorder (PD); phobic disorders (i.e., agoraphobia (AD), social phobia (SoPD), and specific phobia (SpPD)); obsessive-compulsive disorder (OCD); acute stress disorder (ASD); posttraumatic stress disorder (PTSD); and generalized anxiety disorder (GAD). Anxiety disorders frequently occur as comorbid disorders. The current global prevalence of anxiety disorders is 7.3% (4.8–10.9%), ranging from 5.3% (3.5–8.1%) in African cultures to 10.4% (7.0–15.5%) in Euro/Anglo cultures [1]. Lifetime prevalence rates in the Netherlands are 19.6% for any anxiety disorder, 3.8% for PD, 0.9% for AD, 9.3% for SoPD, 0.9% for OCD, 7.4% for PTSD, and 4.5% for GAD [2-4].

Routine outcome monitoring (ROM) is the assessment of treatment outcome at regular intervals in order to monitor patients' progress during treatment. Alongside generic questionnaires completed by all patients, patients who meet the criteria for a particular disorder can be administered disorder-specific questionnaires [5,6] The correct interpretation of ROM results for making clinical decisions about continuing, altering, or terminating treatment requires reliable ratings from reference populations [7]. These ratings can be used to determine whether a patient's level of symptoms falls within the normal range of values following treatment (e.g., whether a treated patient is now no different from normal controls with respect to the severity of anxiety symptoms).

Important issues regarding reference values appear in the literature. First, when data tend toward a non-Gaussian distribution, non-parametric percentile scores provide more appropriate reference values compared to parametric means and standard deviations (SDs) [8,9], and to weighted cut-off values calculated by the Jacobson & Truax method [10]. In that case, the 95th percentile (P_{95}) of the reference-group and the 5th percentile (P_5) of the patient-group commonly serve as reference values [9]. Second, when both reference data and patient data are available, Receiver Operating Characteristics (ROC) analyses can be used to provide cut-offs. The optimal trade-off between sensitivity and specificity, the point of (near) equality, leads to the optimal number of false results (i.e., false positives plus false negatives) [11], depending on the prevalence of the disorder in the general population. It is of note that this applies to disorders that are not very rare. Third, reference values are often established in healthy populations [9]. Absolute health does not exist but is a relative statement. Health should nevertheless be clearly defined, a priori, via inclusion and exclusion criteria [12-14]. Kendall et al., [15] stated that excluding with MDD participants from the reference group if they exhibit elevated levels of the target psychopathology, might lead to creating a nonrepresentative, "supernormal" sample. When comparing the patient group with a supernormal reference group an overly stringent criterion with unreasonable narrow reference intervals would be the result [16]. The inclusion of all possible participants in the

reference group, including those who may currently be experiencing elevated levels of psychopathology is therefore preferable. The goal is to generate a sample that is representative of the general community population [15]. This is in line with a statistical definition of normality, as opposed to a medical definition, both proposed by Wakefield [17]. The statistical perspective of normality is based on the distribution of scores from the population, including all individuals who are not currently treated in secondary care, with extreme scores considered as deviant. The medical perspective excludes individuals with psychopathology from the reference group. A similar definition of disease was given by Cohen [18]: "quantitative deviations from the normal". Fourth, to reduce the amount of uncertainty and random error, (sub)sample sizes of at least 120 are needed [8].

Symptoms of anxiety are suitable for self-rating because anxious persons in general tend to have rather realistic perception and insight (relative to other psychopathological conditions) [19]. We focused on 8 anxiety questionnaires that are often implemented in ROM (Table 5.1). These questionnaires are the self-rated PADUA Inventory Revised (PI-R), Panic Appraisal Inventory (PAI), Penn State Worry Questionnaire (PSWQ), Worry Domains Questionnaire (WDQ), Social Interaction, Anxiety Scale (SIAS), Social Phobia Scale (SPS), and the Impact of Event Scale-Revised (IES-R). Finally, the Brief Scale for Anxiety (BSA) is an observer-rated scale

For healthy control groups, reference values (in the form of means and SDs) have been published for the following questionnaires: PI-R [20,21], PSWQ [22-30], WDQ [26-28,30], both SIAS and SPS [22,31,32], and IES-R [33]. To our knowledge, no reference values have been reported for the BSA and the PAI. For patient groups, means and SDs were published for the BSA [34,35], the PI-R [20,21,36], the PAI [37-39], the PSWQ [22,23,25,40], the WDQ [40], both the SIAS and SPS [22,31], and the IES-R [33,41-44]. However, because of the strong positively skewed distribution of total scores in healthy populations, such as our ROM reference-group, the assumption of a normal distribution is unlikely to be satisfied [8,9]. Reference values should preferably be based on a distribution-free percentile or ROC methodology.

In previous studies, cut-off values (i.e., clinical thresholds) were assessed for the PI-R [21], the PSWQ [23], and the IES-R [33] [45]. Gender differences were reported previously for the PSWQ and WDQ [25,26], the SIAS and the SPS [31,32], and the IES-R [43] healthy control groups. All of these studies reported higher mean values for women than for men. Characteristics of previous studies on reference values are summarized in Table 5.1. The aim of this study was to establish reference values for the BSA, PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R. These reference values included percentile scores, ROC-based cut-off values, and the more commonly reported means and SDs. We compared a sample of 1295 subjects from the general population with a sample of 5066 outpatients suffering from anxiety disorders. A special contribution of the current study is that a healthy (but not necessarily symptom-free) reference-group was included, alongside a well-defined psychiatric patient-group and that both sample sizes were large.

METHODS

Participants

Our analyses of reference values were based on two study samples: a ROM reference-sample from the general population (i.e., the ROM reference-group) and a ROM sample of psychiatric outpatients diagnosed with at least one anxiety disorder (i.e., ROM patient-group).

A total of 1295 participants aged 18 to 65 years (mean age=40.3 years; SD=12.6; 62.8% females) were included in the ROM reference-group, as part of the 'Leiden Routine Outcome Monitoring Study' [6,46]. A representative general population sample was randomly selected from the registration systems of eight general practitioners (GPs) in the region of Leiden, the Netherlands. In the Netherlands, 99.9% of the general population is registered with a GP [47]. The aim was to recruit an apparently psychiatrically healthy reference-group (but not necessarily symptom-free). Therefore, persons who were receiving treatment for psychiatric disorders and/or alcohol or drugs dependency during the six months prior to assessment were excluded. Additional exclusion criteria were hearing impairment or limited cognitive or language abilities (i.e., aphasia, severe dyslexia or dementia; illiteracy or insufficient mastery of the Dutch language). To ensure that the group was demographically comparable to the ROM patient-group, the ROM reference-group was matched for gender, age and urbanization-level (62.3% urban). Participants in the ROM reference-group were assessed in a similar way to the ROM patient-group, except that those in the ROM reference-group completed every disorder-specific questionnaire. As noted previously, the response rate of the ROM reference-group recruitment was 37.1% [6,48], perhaps due to the extensive number of questionnaires which needed to be completed by participants. The BSA was completed by the majority of the ROM reference-group (n=1291), the self-report questionnaires were completed by 50% of the ROM reference-group (due to time-constraints).

The ROM patient-group consisted of a sample of 5066 psychiatric outpatients, aged between 18 and 65 years (mean age=39.3, SD=12.3; 61.0% females), who were diagnosed with and treated for anxiety disorders at the Leiden University Medical Center (LUMC) Department of Psychiatry or the Rivierduinen specialized mental healthcare centres. Baseline assessment was part of the usual ROM procedure. On average, 80% of the patients with a tentative diagnosis of mood-, anxiety- and/or somatoform (MAS) disorder were assessed with ROM in the study period [46]. The BSA was completed by the majority of the ROM patient-group (n=4368), the self-report questionnaires were completed by those who were diagnosed with the relevant anxiety disorder.

To diagnose psychopathology in a standardized manner according to the DSM-IV, a diagnostic interview with the Mini-International Neuropsychiatric Interview plus (MINI-Plus 5.0.0.) [49,50] was done in all participants.

Procedures and questionnaires

Procedures for the web-based ROM program of the LUMC Department of Psychiatry are

described in detail elsewhere [46,51]. For the current study, we used baseline ROM assessments that comprised a standardized diagnostic interview (Dutch version of the Mini-International Neuropsychiatric Interview Plus, version 5.00-R: MINI-Plus) [49,50], the gathering of sociodemographic and socioeconomic data, observer-rated scales, and self-report questionnaires. The assessments were performed by specially trained and constantly supervised research nurses in outpatient clinics of the LUMC and Rivierduinen. Table 5.1 presents the description of each questionnaire, including domains, subscales, ratings, and score-ranges, as well as the respective ROM sample sizes. Sample sizes were determined by participants that completed the particular questionnaire (and not by presence of a particular anxiety disorder). The MINI-Plus was used to establish the presence of Axis I diagnoses according to the DSM-IV.

The Medical Ethical Committee of the LUMC approved the general study protocol associated with ROM, in which ROM was administered as part of the routine treatment process for patients. It involved a comprehensive protocol (titled "Psychiatric Academic Registration Leiden database") which safeguarded the anonymity of patients and persons in the reference-group and ensured proper handling of the ROM data. At intake, patients were informed that the data would be used for research purposes, but only in anonymized form. If patients object to such use, their data were removed. The Medical Ethical Committee of the LUMC approved the regulations and agreed with this policy. In addition, persons in the ROM reference-group signed informed consent for the purpose of this study.

Table 5.1: Anxiety questionnaires used in Routine Outcome Monitoring

Questionnaire {Abbreviation}	Domain	number of items	Rating	Range for score	Our sample sizes Reference / Patient-group	Range for sample sizes in previous studies Reference/Patient-group	References
Brief Scale for Anxiety (vCPRS subscale) {BSA (vCPRS)}	General anxiety	10	0=symptom is absent; 6=symptom is totally dominant	0-60	1291 / 4368	- / 50-101	[47] [30] [29]
PADUA Inventory revised {PI-R}	Obsessive Compulsive Disorder		0=not at all; 4=very much		651 / 657	76-430 / 30-222	[48-50] [15] [16]
Impulses		7		0-28			
Washing		10		0-40			
Checking		7		0-28			
Rumination		11		0-44			
Precision		6		0-24			
Total		41		0-164			
Panic Appraisal Inventory {PAI}	Panic Disorder				630 / 1392	- / 35-47	[32,34] [33]
Anticipated panic		15	0=no chance of panic occurrence; 100=definite panic occurrence	0-100 (average score)			
Perceived consequences of panic:			0=not at all troubling; 10=extremely troubling				
Physical		5		0-50			
Social		5		0-50			
Loss of control		5		0-50			
Total		15		0-150			

Table 5.1: continued

Questionnaire {Abbreviation}	Domain	number of items	Rating	Range for score	Our sample sizes Reference / Patient-group	Range for sample sizes in previous studies Reference/Patient-group	References
Perceived self-efficacy in coping with panic		15	0=not confident at all; 100=completely confident	0-100 (average score)			
Penn State Worry Questionnaire {PSWQ}	Generalized Anxiety Disorder: Excessive and uncontrollable (pathological) worry	16	1=not at all typical of me; 5=very typical of me		651 / 893	32–1138 / 60–436	[20,23,35] [18,51] [24]
Worry Domains Questionnaire {WDQ}	Generalized Anxiety Disorder: Non-pathological worry		0=not at all; 4=extremely		649 / 887	136–432 / -	[23,25,52] [53]
Relationships		4		0-16			
Lack of confidence		5		0-20			
Aimless Future		8	0=no chance of panic occurrence; 100=definite panic occurrence	0-32			
Work incompetence		3	0=not at all troubling; 10=extremely troubling	0-12			
Financial		4		0-16			
Physical Health		6		0-24			
Total		30		0-120			

Table 5.1: continued

Questionnaire {Abbreviation}	Domain	number of items	Rating	Range for score	Our sample sizes Reference / Patient-group	Range for sample sizes in previous studies Reference/Patient-group	References
Social Interaction and Anxiety Scale {SIAS}	Social Phobia	20	0= not at all characteristic or true of me; 4- extremely characteristic or true of me	0-80	651 / 1231	21–482 / 13-165	[17,27]
Social Phobia Scale {SPS}	Social Phobia	20	0= not at all characteristic or true of me; 4- extremely characteristic or true of me	0-80	651 / 1237	21–482 / 13-165	[17,27]
Impact of Event Scale – Revised {IES-R}	Traumatic Events		0=not at all; 4=extremely		1272 / 390	154 / 120–4167	[39,54] [55];[28] [37]
Intrusions		8		0-32			
Avoidance		8		0-32			
Hyperarousal		6		0-24			
Total		22		0-88			
Mini International Neuropsychiatric Interview Plus 5.0.0. {MINI Plus 5.0.0}	General Pathology				1295/5066		[45,46]

Statistical analyses

Analyses were performed separately for the ROM reference-group and the patient-group, while ROC and internal consistency analyses were conducted using data from both groups combined. In both groups, participants who had more than one missing value per subscale were excluded. This allowed us to conduct a robust evaluation of the use of the anxiety questionnaires. Sociodemographic and psychopathological variables were analyzed using descriptive statistics (percentages in the case of categorical variables, means and SDs for the continuous variables). Cut-off values indicating an optimal discrimination threshold between 'healthy' and 'diseased' were obtained by ROC analyses. We chose to allow sensitivity and specificity to be equal, taking into account the trade-off between the two [11]. The discriminatory power of the questionnaire (sub) scales was assessed with the associated areas under the ROC curve (AUCs). AUC's over 0.75 were considered clinically useful, with 0.85 showing moderate discriminatory power and 0.95 very high power [52]. The 5^{th}, 25^{th}, 50^{th} (i.e. median), 75^{th}, and 95^{th} percentile scores were calculated. The central 95% of the distribution in reference-groups is commonly used in cases of non-Gaussian distributions [12,53]. The remaining 5% was categorized as 'abnormal' [54]. We chose to categorize the top 5% of the reference-group (95^{th} percentile scores, P_{95}) as 'abnormal' because the lowest 2.5% (functioning 'abnormally' good) cannot be identified in general population samples; the studied anxiety questionnaires merely assess the level of dysfunctionality and not the level of 'health' or normal functionality. Likewise, we regarded the bottom 5% of the patient-group (5^{th} percentile scores, P_5) as indistinguishable from people in the normal range. Furthermore, means and SDs were calculated. Reference values were calculated for all participants combined, as well as for men and women separately. To test our decision not to exclude those individuals in the ROM reference-group with a current psychiatric diagnosis, we performed a sensitivity analysis. The internal consistency of the questionnaires was evaluated using Cronbach's alpha for the total scores and the subscores (with >0.70 indicating adequate internal consistency) [55]. For all analyses, SPSS version 20.0 was used (SPSS Inc, Chicago, Illinois).

RESULTS

Sociodemographic and psychopathological characteristics

The sociodemographic and psychopathological characteristics of the ROM reference-group and patient-group are shown in Table 5.2.

Participants in the ROM reference-group and the ROM patient-group were comparable with respect to mean age and similar with respect to gender distribution. For the ROM reference-group the mean age was 40.3 years (SD=12.6), for the ROM patient-group it was 36.5 years (SD=11.9). Females constituted 62.8% of the reference-group and 64.4% of the patient-group. Those in the ROM reference-group were more often married relative to those in the ROM patient-group and they were less often living alone. Those in the ROM

reference-group were more often married relative to those in the ROM patient-group and they were less often living alone. Those in the ROM reference-group also showed higher levels of education relative to those in the ROM patient-group. Furthermore, work-related disability and unemployment were less prevalent in the ROM reference-group. Fewer participants in the ROM reference-group were of ethnic origin (defined as oneself not being born in the Netherlands or both parents not being born in the Netherlands). Of the ROM reference-group 9.3% had at least one anxiety disorder and 5.2% met criteria for a psychiatric disorder in addition to an anxiety disorder as diagnosed with the MINI-Plus. There was a high rate of psychopathological co-morbidity (i.e., psychopathology in addition to psychopathological anxiety) among participants in the ROM patient-group (55.6%).

REFERENCE VALUES

Percentile scores

Table 5.3 presents the reference values of the eight anxiety questionnaires for the ROM reference-group and the ROM patient-group. For the ROM reference-group, the distribution of each total score and sub score was positively skewed. Mental health was also demonstrated for the ROM reference-group by the substantial percentage of participants (5-25%) having the lowest possible scores (e.g., 5% for the BSA, PAI, SPS, and 25% for the IES-R).

Analyses of gender indicated that both healthy and women with anxiety disorders showed more symptoms of anxiety relative to the men, both in the ROM reference- and ROM patient-groups (see Supplementary Tables 1 through 6).

Table 5.2.: Sociodemographic and psychiatric characteristics of the ROM reference (n=1295) patient (n=5066) groups.

	ROM reference group (n= 1295)		ROM patient group (n=4627)	
Gender: - n (%)				
Male	482	(37.2)	1806	(35.7)
Female	813	(62.8)	3260	(64.4)
Age in years: - mean (± SD)	40.3	(12.6)	36.5	(11.9)
Male	41.2	(12.6)	37.8	(11.9)
Female	39.7	(12.6)	35.8	(11.8)
Marital status[1]: - n (%)				
Married/cohabitating	890	(68.7)	2206	(43.5)
Divorced/separated/widow	78	(6.0)	539	(10.6)
Single	327	(25.3)	1744	(34.4)
Housing situation[1]: - n (%)				
Living alone	201	(15.5)	982	(19.4)
Living with partner	902	(69.7)	2259	(44.6)
Living with family	192	(14.8)	1248	(24.6)
Educational status1,3: - n (%)				
Lower	295	(22.8)	1867	(36.9)
Higher	1000	(77.2)	2619	(51.7)
Employment status[1]: - n (%)				
Employed part-time	509	(39.3)	1033	(20.4)
Employed full-time	554	(42.8)	986	(19.5)
Unemployed/retired	197	(15.2)	1298	(25.6)
Work-related disability	35	(2.7)	1172	(23.1)
Ethnic background[1]: - n (%)				
Dutch	1150	(88.8)	3505	(69.2)
Other ethnicity	145	(11.2)	982	(19.4)
MINI diagnoses: - n (%)				
Currently None	1174	(90.7)	0[2]	
Anxiety disorder (single)	54	(4.2)	2246	(44.3)
Anxiety disorder (comorbidity)	18	(1.4)	2820	(55.6)
Other psychiatric disorder	49	(3.8)	0[2]	

SD denotes standard deviation
[1] Data not available for 128 (2.4%) to 640 (11.8%) of patients
[2] Selection criterion
[3] Lower education: primary or vocational school: Higher education: college or university

Table 5.3: Percentile scores and mean values for Routine Outcome Monitoring anxiety disorder questionnaires in the ROM reference (n=1295) and patient (n=5066) groups.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
			(n=1291)							(n=4368)		
Brief Scale for Anxiety (BSA)	0	1	3	6	11	3.91 ± 3.92	6	12	16	21	28	16.36 ± 6.78
PADUA Inventory Revised (PI-R)			(n=651)							(n=657)		
Impulses	0	0	0	1	4	0.84 ± 1.70	0	1	4	8	15	5.09 ± 5.18
Washing	0	0	1	3	11	2.27 ± 3.83	0	1	6	17	32	9.86 ± 10.72
Checking	0	1	3	6	12	4.07 ± 4.04	1	8	14	20	26	13.95 ± 7.60
Rumination	1	3	7	11	18	7.71 ± 5.69	10	19	24	29	38	23.87 ± 8.23
Precision	0	0	1	2	6	1.57 ± 2.20	0	3	6	11	18	7.38 ± 5.72
Total	2	7	13	22	43	16.46±13.30	20	40	58	78	106	60.15 ±26.21
Panic Appraisal Inventory (PAI)			(n=630)							(n=1392)		
Anticipated panic	0	1	7	17	37	10.82±12.16	14	32	47	62	82	47.42 ±20.32
Perceived consequences of Panic:												
-Physical	0	0	0	2	17	2.87 ± 6.62	0	7	18	31	44	19.52 ± 13.99
-Social	0	0	0	4	14	3.01 ± 5.72	0	6	14	26	40	16.68 ± 12.79
-Loss of Control	0	0	0	3	17	3.13 ± 5.86	2	9	17	27	40	18.43 ±11.89
-Total	0	0	2	11	47	9.01 ± 15.03	10	31	52	75	108	54.63 ±29.84

Table 5.3: continued.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Perceived self-efficacy in coping with panic	0	7	21	36	65	24.19±21.30	29	49	62	76	90	61.48 ±18.41
			(n=651)						(n=893)			
Penn State Worry Questionnaire (PSWQ)	22	30	36	47	66	39.52±13.19	**48**	**62**	**69**	**74**	**79**	**66.95 ± 9.92**
Worry Domains Questionnaire (WDQ)			(n=649)						(n=887)			
Relationships	4	4	4	6	9	5.27 ± 1.99	4	7	10	14	18	10.28 ± 4.47
Lack of Confidence	5	5	7	9	14	7.65 ± 3.30	7	12	16	20	24	15.95 ± 5.15
Aimless Future	8	8	10	12	19	11.05 ± 4.00	9	15	21	27	35	21.07 ± 7.79
Work Incompetence	3	3	4	6	9	4.77 ± 1.99	3	6	8	11	14	8.47 ± 3.37
Financial	4	4	5	8	13	6.51 ± 3.04	4	7	10	15	19	10.86 ± 4.92
Health	**6**	**6**	**7**	10	15	8.46 ± 3.31	7	11	14	19	26	15.20 ± 5.94
Total	33	34	39	49	74	43.72±13.62	44	65	81	97	120	81.82 ±23.66
			(n=651)						(n=1231)			
Social Interaction and Anxiety Scale (SIAS)	1	6	10	17	32	12.50 ± 9.34	18	33	44	54	68	43.70 ±14.92
			(n=651)						(n=1237)			
Social Phobia Scale (SPS)	0	2	4	8	19	6.04 ± 6.57	11	22	33	47	64	35.06 ±16.76

Table 5.3: continued.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Impact of Event Scale – Revised (IES-R)[1]			**(n=1272)**						**(n=390)**			
Intrusions	0	0	1	5	15	3.51 ± 5.03	5	15	20	24	31	19.52 ± 7.36
Avoidance	0	0	0	4	14	2.72 ± 4.80	5	13	18	22	29	17.30 ± 6.96
Hyperarousal	0	0	0	2	8	1.77 ± 3.14	4	11	15	18	22	14.39 ± 5.39
Total	0	0	2	11	36	7.99 ± 11.97	19	43	53	62	78	51.20 ±17.12

SD denotes standard deviation. [1] IES-R scores are sum scores: to yield average scores, divide by number of items

In a sensitivity analysis, we excluded the 9.7% of participants in the ROM reference-group who had a MINI-diagnosis. Among the remaining 1161 participants we found that the median of the changes of the mean scores of the eight anxiety questionnaires was –8% (interquartile range: –5% to –13%). The median of the changes of the P_{95} scores was –9% (interquartile range: –7% to –12%).

To facilitate comparability with the international literature, we also provided means and SDs in Table 5.3. However, we consider these reference values as less valid given that the distributions of all (sub) scores were positively skewed in the ROM reference-group (Figure 1).

Receiver operating characteristic (ROC) curves

Cut-off values, defined by equal sensitivity and specificity, were calculated with ROC analyses (see Table 5.4). The discriminative power of the eight anxiety questionnaires is depicted in Figure 5.1.

ROC analyses, used to discriminate between health and disease, yielded the following cut-off values: 8.5 for the BSA total score, 30,5 for the PI-R total score, 23.5 for the PAI Anticipated Panic subscale score, 21.5 for the total of the PAI Perceived Consequences, and 43.5 for the PAI Perceived Self-efficacy subscale. The cut-off values were as follows: 55.5 for the PSWQ, 55.5 for the WDQ total scale, 24.5 for the SIAS, 14 for the SPS, and 27.5 for the IES-R total scale. AUC values indicated very high discriminatory power for the BSA, the SIAS, the SPS, and the IES-R. Two subscales, PI-R Washing and WDQ Financial, showed clinically useful discriminatory power. All other (sub) scales proved to have moderate discriminatory power. Sensitivity and specificity exceeded 85% for most (sub) scales; for PI-R subscales and WDQ subscales sensitivity and specificity were somewhat lower.

Internal consistency

The internal consistencies of the total scales and subscales of the questionnaires (for all subjects combined) are shown in Table 5.4. The total scales and subscales of all seven self-rating questionnaires showed excellent internal consistencies, with the exception of WDQ subscale Work Incompetence which possessed adequate internal consistency. The internal consistency of the BSA was also adequate.

Table 5.4: Internal consistency and cut-off values in the ROM reference (n=1295) and patient (n=5066) groups for Routine Outcome Monitoring anxiety disorder questionnaires.

	Number of items	Cronbach's Alpha	N	ROC analysis cut-off	Area under the Curve	Sensitivity / specificity (%)
Brief Scale for Anxiety (BSA)	10	0.78	5659	8.5	0.95	87.7 / 87.8
PADUA Inventory Revised (PI-R)			1308			
Impulses	7	0.84		1.5	0.80	68.7 / 81.3
Washing	10	0.95		1.5	0.73	70.2 / 62.5
Checking	7	0.92		6.5	0.86	79.7 / 77.1
Rumination	11	0.93		13.5	0.94	87.5 / 86.3
Precision	6	0.83		2.5	0.83	75.9 / 77.6
Total	41	0.96		30.5	0.94	86.1 / 86.2
Panic Appraisal Inventory (PAI)			2202			
Anticipated panic	15	0.93		23.5	0.94	87.2 / 86.6
Perceived consequences of panic						
-Physical	5	0.89		4.5	0.89	84.1 / 82.9
-Social	5	0.86		4.5	0.86	80.2 / 78.9
-Loss of Control	5	0.84		6.5	0.90	82.6 / 84.1
-Total	15	0.92		21.5	0.93	86.3 / 86.3
Perceived self-efficacy in coping with panic	15	0.96		43.5	0.90	83.1 / 83.4
Penn State Worry Questionnaire (PSWQ)	16	0.95	1544	55.5	0.93	87.4 / 86.5

Table 5.4: continued

	Number of items	Cronbach's Alpha	N	ROC analysis cut-off	Area under the Curve	Sensitivity / specificity (%)
Worry Domains Questionnaire (WDQ)			1536			
Relationships	4	0.87		6.5	0.85	75.5 / 82.4
Lack of Confidence	5	0.91		10.5	0.90	82.5 / 84.4
Aimless Future	8	0.89		13.5	0.88	80.0 / 80.0
Work Incompetence	3	0.79		5.5	0.82	76.6 / 71.3
Financial	4	0.90		7.5	0.76	69.2 / 72.1
Health	6	0.87		10.5	0.86	75.3 / 80.4
Total	30	0.96		55.5	0.92	85.6 / 85.4
Social Interaction and Anxiety Scale (SIAS)	20	0.96	1882	24.5	0.96	88.9 / 89.2
Social Phobia Scale (SPS)	20	0.96	1888	14.0	0.96	90.0 / 90.0
Impact of Event Scale – Revised (IES-R)			1662			
Intrusions	8	0.96		10.5	0.95	88.2 / 88.4
Avoidance	8	0.94		9.0	0.95	87.7 / 87.9
Hyperarousal	6	0.94		6.5	0.96	92.3 / 91.5
Total	22	0.98		27.5	0.96	91.3 / 91.4

The optimal cut-off derived by the ROC analysis is defined by equal sensitivity and specificity scores

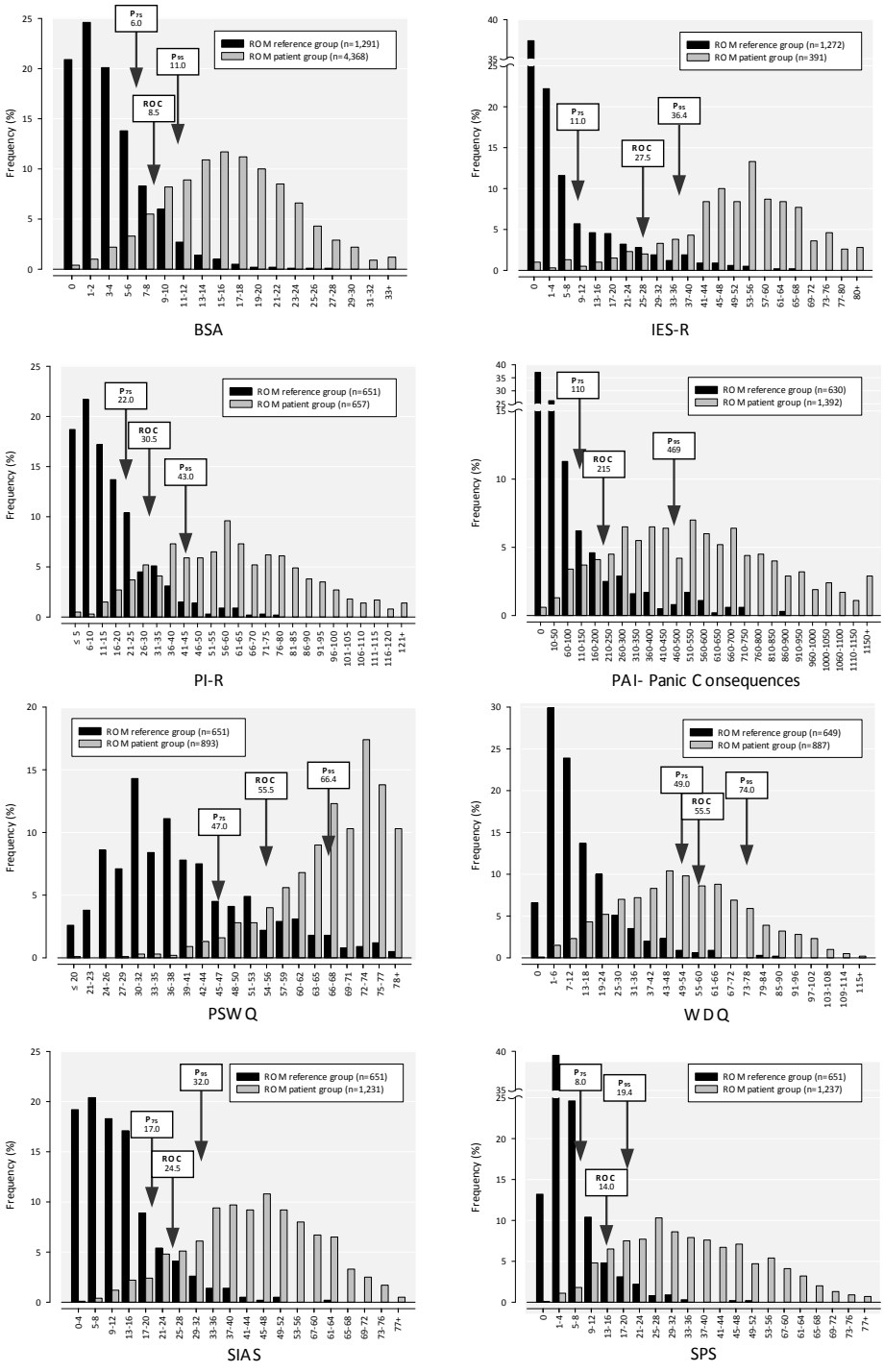

Figure 5.1: Distribution of the scores of the Brief Scale for Anxiety (BSA), PADUA Inventory Revised (PI-R), Panic Appraisal Inventory (PAI) subscale Panic Consequences, Penn State Worry Questionnaire (PSWQ), Worry Domains Questionnaire (WDQ), Social Interaction and Anxiety Scale (SIAS), Social Phobia Scale (SPS), and Impact of Event Scale-Revised (IES-R). Three types of cut-off values are depicted: the 75th percentile score (P_{75}), the 95th percentile score (P_{95}) and the Receiver Operating Characteristics (ROC) cut-off value defined by equal sensitivity and specificity.

DISCUSSION AND CONCLUSION

We reported reference values for a broad range of anxiety questionnaires in two large samples from 'healthy' and 'psychiatrically ill' populations. P_{95} values of the ROM reference-group, cut-off values based on ROC analysis, and P_5 values of the ROM patient-group yielded closely related values. P_{95} values of the ROM reference-group were the highest, ROC values were slightly lower, and P_5 values of the ROM patient-group were the lowest. A pervasive gender-specific pattern in reference values was observed, with higher reference values in women than in men in the ROM reference-group.

The mean PI-R score for our ROM reference-group (M=16.5; SD=13.3) was lower than the mean PI-R scores reported previously, ranging from 21.6 [21] to 37.7 [20]. The mean PSWQ score for the ROM reference-group (M=39.5; SD=13.2) was comparable to the mean PSWQ scores reported by other researchers, ranging from 34.9 to 49.5 [23-28,56], suggesting that our reference-group showed normal levels of pathological worry. The mean WDQ score for the ROM reference-group (M=43.7; SD=13.6) was slightly higher compared to the mean WDQ scores reported in the literature, where it ranged from 24.8 to 38.1 [26-28,30]. This could be explained by only a few participants in our reference-group that showed a high level of non-pathological worry, within the positively skewed distribution. For the ROM reference-group the mean SIAS score (M=12.5; SD=9.3) was slightly lower than the mean SIAS scores reported in other studies, ranging from 14.3 to 19.9 [22,31,32]. The mean SPS score for the ROM reference-group (M=6.0; SD=6.6) was slightly lower than the mean SPS scores reported in literature, ranging from 6.3 to 14.4 [22,31,32]. For the ROM reference-group the mean IES-R score (M=8.0; SD=12.0) was much lower than the mean IES-R score reported by Creamer et al. (M=40.0; SD=23.1) [33]. In sum, the mean scores for our ROM reference-group tended to be lower than the mean scores reported by other researchers, suggesting that our reference-group was relatively healthy. It should however be taken into account that the highly skewed distributions precluded a valid comparison of mean values. For the ROM reference-group the mean scores for the PI-R, the PSWQ, and the IES-R were well below the clinical thresholds as used by other researchers [21,23,33]. This indicated no or only mild anxiety, similar to the previous results. The conducted sensitivity analyses showed slightly lower cut-off values for the reference-group with individuals with a current psychiatric diagnosis excluded. However, these individuals were chosen to be included, in order to prevent producing too strict cut-off values. This would lead to fewer patients considered recovered when P_{95} cut-off scores are used. The high internal consistencies of the PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R are in accordance with previous studies [25,26,32,36,39,44].

There were some notable differences among the previously published and the present reference values. Health perceptions and health problem expressions vary between cultures [57]. Furthermore, there are differences in study design (e.g., mode of questionnaire administration) [58-60], socio-economic status [58,61], physical functionality [61], health

status varying with area of residence [59], or clinical severity [58,61]. Furthermore, different language versions of the same questionnaire have to measure the same underlying construct where all aspects of this construct (e.g., domain, operational mode, semantics, and psychometric properties) should be similar [60,62]. Two versions of the same questionnaire can be equally sensitive to a given change in functional status yet assign different scores to a given level of distress [61]. Therefore, our reference values should be used with caution in different settings. Further research should evaluate cross-country variability of reference values.

It is noteworthy that a consistent pattern was observed in the 75th and 95th percentile scores of the ROM reference-group, the ROC cut-off values, and the 5th percentile scores of the patient-group. That is, they overlapped considerably, with P_{95} of the ROM reference-group being slightly highest, followed by the ROC cut-off values. The 5th percentile scores of the ROM patient-group had similar values compared to the 75th percentile scores of the ROM reference-group. These values were lower than the 95th percentile scores and ROC cut-off values. This pattern is very similar to the pattern we observed for ROM generic questionnaires [48]. In contrast, for the ROM mood questionnaires the 5th percentile of the ROM patient-group had similar values compared to the 95th percentile of the ROM reference-group [63]. This suggests that there is relatively more subsyndromal anxiety as compared to subsyndromal depression in the ROM reference-group. Mild anxiety may be considered a normal human experience. The ROC cut-off values were rather consistent with the cut-off values derived by other researchers for the PSWQ (55.5 versus 52.3 [23] and for the IES-R (27.5 versus 33 [33].

Furthermore, on average, men from the ROM reference-group scored lower on all eight anxiety scales than did the women from the ROM reference-group. Respectively, for men and women, cut-off (P_{95}) values were 10 and 12 for the BSA, 38 and 44 for the PI-R, 27 and 39 for the PAI Anticipated Panic, 27 and 52 for the PAI Perceived Consequences, 71 and 62 for the PAI Perceived Self-efficacy, 61 and 70 for the PSWQ, 61 and 77 for the WDQ, 27 and 34 for the SIAS, 14 and 22 for the SPS, and 29 and 38 for IES-R. It may be too early to recommend gender-specific reference values because more research is needed in reference populations. Nevertheless, it was striking that reference values from a non-anxious population showed a clinically important gender effect. Most previous studies did not stratify for gender, but those which did [25,26,31,32,43] reported higher means for women than for men, similar to our results.

The results of our study have several clinical implications. The excellent performance of the questionnaires suggests that our reference values are appropriate for various objectives: 1) decisions about treatment termination and referral back to primary care (using the P_{95} of the ROM reference-group); 2) identification of people who may benefit from referral by primary care to specialized mental health care (using the P_5 of the ROM Patient-group), and even 3) diagnostics (using the ROC cut-off values). Regarding diagnostics, these cut-off values might aid in screening for various anxiety disorders, although clinical judgment

and validated diagnostic tools remain the gold standard (e.g., MINI [49,50], Composite International Diagnostic Interview [CIDI; [64]], the Structured Clinical Interview for DSM-III-R [SCID, [65]]). Moreover, cut-off values may be used to classify anxiety. When making decisions about treatment termination or referral to primary care, specificity has to be high [66]. The 75th percentile scores of the ROM reference-group result in few false positives for 'health'. For referral from primary care to specialized mental health care, cut-off values with a high sensitivity are more appropriate, and for that purpose we recommend ROC-based cut-offs or 5th percentile scores from the ROM patient-group because they result in few false positives for 'disease'.

The present study has several strengths. The assessment procedures for both groups were standardized and of high quality (ascertained by training and supervision). Furthermore, the ROM reference-group was large, it was clearly defined, and it resembled the patient-group in all relevant respects (age, gender, level of urbanization) other than those under investigation (i.e., level of psychopathology). The rather precise estimates arising out of the current study are probably attributable to the large sample size. Additionally, the reference-group probably represents the general population quite well. GP registers were used to recruit the reference-group and in the Netherlands the GP registration rate is very high. The ROM patient-group was large as well. Finally, stratification of the ROM reference-group into more homogeneous gender-subgroups may have reduced variation among subgroups, leading to gender-specific reference values, which can be used in clinical practice.

A limitation of the present study includes the relatively high non-response rate in the ROM reference-group, which may have introduced potential selection bias. Additionally, the generalizability of this study is limited by the nature of our ROM reference-group in that it included Dutch-speaking people aged between 18 and 65 years. Reference values may not automatically be applicable to other ethnic groups, to children, and to the elderly. Finally, it is important to recognize that population-based reference values should not be applied rigidly. The choice of cut-off values remains arbitrary and dependent on one's goal (e.g., for confirmation of a diagnosis, specificity should be high and the 95th percentile would be more appropriate than the 75th percentile of the ROM reference-group).

In conclusion, this large-scale population-based study provides reference values and reliability coefficients for the BSA, PI-R, PAI, PSWQ, WDQ, SIAS, SPS, and IES-R. These values increase the utility of these questionnaires, inasmuch as they can be employed as ROM questionnaires to facilitate the assessment of severity of anxiety disorder symptoms. To make responsible decisions about continuing, changing, or terminating therapy, any of these questionnaires can be offered to every patient with MAS disorders. Additionally, these reference values are suitable for indicating which patients have recovered enough to be referred back from specialized mental health care to primary care.

Reference List

1. Baxter AJ, Scott KM, Vos T, Whiteford HA. (2012) Global prevalence of anxiety disorders: a systematic review and meta-regression. *Psychol Med,* 1-14.

2. Bijl RV, Ravelli A, Van Zessen G. (1998) Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol*, 33 (12), 587-595.

3. De Graaf R, Ten Have M, Van Gool C, Van Dorsselaer S. (2012) Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc Psychiatry Psychiatr Epidemiol,* 47 (2), 203-213.

4. De Vries GJ, Olff M. (2009) The lifetime prevalence of traumatic events and posttraumatic stress disorder in the Netherlands. *J Trauma Stress,* 22 (4), 259-267.

5. De Beurs E, Den Hollander-Gijsman ME, Van Rood YR, Van der Wee NJ, Giltay EJ, Van Noorden MS, Van der Lem R, Van Fenema EM, Zitman FG. (2011) Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clin Psychol Psychother*, 18, 1-12.

6. Schulte-van Maaren YWM, Carlier IVE, Giltay EJ, Van Noorden MS, De Waal MW, Van der Wee NJ, Zitman FG. (2012) Reference values for mental health assessment instruments: objectives and methods of the Leiden Routine Outcome Monitoring Study. *J Eval Clin Pract*.

7. Kazdin AE. (2008) Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *Am Psychol*, 63 (3), 146-159.

8. Reed AH, Henry RJ, Mason WB. (1971) Influence of statistical method used on the resulting estimate of normal range. *Clin Chem*, 17 (4), 275-284.

9. Solberg HE. (2008) Establishment and use of reference values. Burtis CA, Ashwood ER, Bruns DE, editors. Fundamentals of clinical chemistry. 6[14], 229-238. St. Louis, Missouri, Saunders Elsevier.

10. Jacobson NS, Truax P. (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol,* 59 (1), 12-19.

11. Fan J, Upadhye S, Worster A. (2006) Understanding receiver operating characteristic (ROC) curves. *CJEM*, 8 (1), 19-20.

12. Geffre A, Friedrichs K, Harr K, Concordet D, Trumel C, Braun JP. (2009) Reference values: a review. *Vet Clin Pathol*, 38 (3), 288-298.

13. Katayev A, Balciza C, Seccombe DW. (2010) Establishing reference intervals for clinical laboratory test results: is there a better way? *Am J Clin Pathol*, 133 (2), 180-186.

14. Sasse EA, Doumas BT, Miller WG, D'Orazio P, Eckfeldt JH, Evans SA, et al. (2000) How to define and determine reference intervals in the clinical laboratory; approved guideline-Second edition. NCCLS document C28-A2 . 20[13], 1-38. Wayne, PA, NCCLS.

15. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. (1999) Normative comparisons for the evaluation of clinical significance. *J Consult Clin Psychol,* 67 (3), 285-299.

16. Horn PS, Feng L, Li Y, Pesce AJ. (2001) Effect of outliers and nonhealthy individuals on reference interval estimation. *Clin Chem,* 47 (12), 2137-2145.

17. Wakefield JC. (1992) The concept of mental disorder. On the boundary between biological facts and social values. *Am Psychol*, 47 (3), 373-388.

18. Cohen H. (1981) The evolution of the concept of disease. Caplan AL, Engelhardt HT, McCartney JJ, editors. Concepts of health and disease: Interdisciplinary perspectives. 209-220. Reading, MA, Addison-Wesley.

19. Prusoff BA, Klerman GL, Paykel ES. (1972) Concordance between clinical assessments and patients' self-report in depression. *Arch Gen Psychiatry,* 26 (6), 546-552.

20. Besiroglu L, Yucel AM, Boysan M, Eryonucu B, Gulec M, Selvi Y. (2005) [The assessment of obsessive-compulsive symptoms: the reliability and validity of the Padua inventory in a Turkish population]. *Turk Psikiyatri Derg*, 16 (3), 179-189.

21. Van Oppen P, Hoekstra RJ. (1995) The structure of obsessive-compulsive symptoms. *Behav Res Ther,* 33 (1), 15-23.

22. Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH. (1997) Validation of the social interaction anxiety scale and the social phobia scale across the anxiety disorders. *Psychological Assessment*, 9 (1), 21-27.

23. Dupuy JB, Beaudoin S, Rheaume J, Ladouceur R, Dugas MJ. (2001) Worry: daily self-report in clinical and non-clinical populations. *Behaviour Research and Therapy*, 39 (10), 1249-1255.

24. Fresco DM, Heimberg RG, Mennin DS, Turk CL. (2002) Confirmatory factor analysis of the Penn State Worry Questionnaire. *Behav Res Ther*, 40 (3), 313-323.

25. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. (1990) Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther*, 28 (6), 487-495.

26. Robichaud M, Dugas MJ, Conway M. (2003) Gender differences in worry and associated cognitive-behavioral variables. *J Anxiety Disord,* 17 (5), 501-516.

27. Stober J. (1998) Reliability and validity of two widely-used worry questionnaires: Self-report and self-peer convergence. *Personality and Individual Differences*, 24 (6), 887-890.

28. Van Rijsoort S, Vervaeke G. (1999) The Penn State Worry Questionnaire and the Worry Domains Questionnaire: structure, reliability and validity. *Clinical Psychology & Psychotherapy,* 6 (4), 297-307.

29. Zlomke KR. (2009) Psychometric properties of internet administered versions of Penn State Worry Questionnaire (PSWQ) and Depression, Anxiety, and Stress Scale (DASS). *Computers in Human Behavior*, 25 (4), 841-843.

30. Verkuil B, Brosschot JF, Thayer JF. (2007) Capturing worry in daily life: are trait questionnaires sufficient? *Behav Res Ther*, 45 (8), 1835-1844.

31. Heimberg RG, Mueller GP, Holt CS, Hope DA, Liebowitz MR. (1992) Assessment of Anxiety in Social-Interaction and Being Observed by Others - the Social-Interaction Anxiety Scale and the Social Phobia Scale. *Behavior Therapy*, 23 (1), 53-73.

32. Mattick RP, Clarke JC. (1998) Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther,* 36 (4), 455-470.

33. Creamer M, Bell R, Failla S. (2003) Psychometric properties of the Impact of Event Scale - Revised. *Behav Res Ther,* 41 (12), 1489-1496.

34. Mattila-Evenden M, Svanborg P, Gustavsson P, Asberg M. (1996) Determinants of self-rating and expert rating concordance in psychiatric out-patients, using the affective subscales of the CPRS. *Acta Psychiatr Scand*, 94 (6), 386-396.

35. Tyrer P, Owen RT, Cicchetti DV. (1984) The brief scale for anxiety: a subdivision of the comprehensive psychopathological rating scale. *J Neurol Neurosurg Psychiatry,* 47 (9), 970-975.

36. Anholt GE, Van Oppen P, Cath DC, Smit JH, Van Dyck R, Van Balkom AJLM. (2009) Measuring obsessive-compulsive symptoms: Padua Inventory-Revised vs. Yale-Brown Obsessive Compulsive Scale. *J Anxiety Disord*, 23 (6), 830-835.

37. De Beurs E, Smit JH, Comijs HC. (2005) Kort Instrumenteel: De Paniek Opinie Lijst (POL). De Betrouwbaarheid en validiteit van een cognitieve maat voor paniekstoornis. *Tijdschrift voor Gedragstherapie*, 2, 141-154.

38. Feske U, De Beurs E. (1997) The Panic Appraisal Inventory: psychometric properties. *Behav Res Ther,* 35 (9), 875-882.

39. Telch MJ, Brouillard M, Telch CF, Agras WS, Taylor CB. (1989) Role of cognitive appraisal in panic-related avoidance. *Behav Res Ther*, 27 (4), 373-383.

40. Kerkhof A, Hermans D, Figee 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. *Tijdschrift voor Gedragstherapie,* 2, 135-154.

41. Beck JG, Grant DM, Read JP, Clapp JD, Coffey SF, Miller LM, Palyo SA. (2008) The impact of event scale-revised: psychometric properties in a sample of motor vehicle accident survivors. *J Anxiety Disord*, 22 (2), 187-198.

42. Morina N, Bohme HF, Ajdukovic D, et al. (2010) The structure of post-traumatic stress symptoms in survivors of war: confirmatory factor analyses of the Impact of Event Scale--revised. *J Anxiety Disord,* 24 (6), 606-611.

43. Mystakidou K, Tsilika E, Parpa E, Galanos A, Vlahos L. (2007) Psychometric properties of the Impact of Event Scale in Greek cancer patients. *J Pain Symptom Manage,* 33 (4), 454-461.

44. Weiss DS, Marmar CR. (1997) The Impact of Event Scale-Revised. Wilson JP, Keane TM, editors. Assessing psychological trauma and PTSD: A handbook for practitioners. 399-411. New York, Guildford Press.

45. Sondergaard HP, Ekblad S, Theorell T. (2003) Screening for post-traumatic stress disorder among refugees in Stockholm. *Nord J Psychiatry,* 57 (3), 185-189.

46. Van Noorden MS, Giltay EJ, Den Hollander-Gijsman ME, Van der Wee NJ, Van Veen T, Zitman FG. (2010) Gender differences in clinical characteristics in a naturalistic sample of depressive outpatients: the Leiden Routine Outcome Monitoring Study. *J Affect Disord*, 125 (1-3), 116-123.

47. Poortvliet MC, Lamkadden M, Deville W. Niet op naam ingeschreven (NONI) bij de huisarts. Inventarisatie en gevolgen voor de ziekenfondsverzekerden. Utrecht: NIVEL; 2005.

48. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Hemert AM, De Waal MW, Van Noorden MS, Giltay EJ. (2012) Reference values for generic instruments used in Routine Outcome Monitoring: the Leiden Routine Outcome Monitoring Study (in press). *BMC Psychiatry.*

49. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*, 59 Suppl 20, 22-33.

50. Van Vliet IM, De Beurs E. (2007) The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders. *Tijdschr Psychiatr,* 49 (6), 393-397.

51. Carlier IVE, Meuldijk D, Van Vliet IM, Van Fenema EM, Van der Wee NJ, Zitman FG. (2012) Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract,* 18 (1), 104-110.

52. Barnabei L, Marazia S, De CR. (2007) Receiver operating characteristic (ROC) curves and the definition of threshold levels to diagnose coronary artery disease on electrocardiographic stress testing. Part I: The use of ROC curves in diagnostic medicine and electrocardiographic markers of ischaemia. *J Cardiovasc Med (Hagerstown),* 8 (11), 873-881.

53. Solberg HE. (1989) Reference values. *Adv Clin Chem,* 27, 1-79.

54. Campbell WW, Robinson LR. (1993) Deriving reference values in electrodiagnostic medicine. *Muscle Nerve*, 16 (4), 424-428.

55. George D, Mallery P. (2003) SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston, Allyn & Bacon.

56. Brown TA, Antony MM, Barlow DH. (1992) Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behav Res Ther*, 30 (1), 33-37.

57. Kleinman A, Eisenberg L, Good B. (1978) Culture, illness, and care: clinical lessons from anthropologic and cross-cultural research. *Ann Intern Med,* 88 (2), 251-258.

58. Alonso J, Ferrer M, Gandek B, et al. (2004) Health-related quality of life associated with chronic conditions in eight countries: results from the International Quality of Life Assessment (IQOLA) Project. *Qual Life Res,* 13 (2), 283-298.

59. Bowling A, Bond M, Jenkinson C, Lamping DL. (1999) Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. *J Public Health Med,* 21 (3), 255-270.

60. Bullinger M, Anderson R, Cella D, Aaronson N. (1993) Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res*, 2 (6), 451-459.

61. Anderson RT, Aaronson NK, Wilkin D. (1993) Critical review of the international assessments of health-related quality of life. *Qual Life Res*, 2 (6), 369-395.

62. Herdman M, Fox-Rushby J, Badia X. (1998) A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*, 7 (4), 323-335.

63. Schulte-van Maaren YWM, Carlier IVE, Zitman FG, Van Hemert AM, De Waal MW, Van der Does AJW, Van Noorden MS, Giltay EJ. (2012) Reference values for major depression questionnaires: the Leiden Routine Outcome Monitoring Study. *Journal of Affective Disorders*.

64. World Health Organization. (1990) The Composite International Diagnostic Interview (CIDI). Authorized Core Version 1.0. Geneva, WHO.

65. Spitzer RL, Williams JBW, Gibbon M, First MB. (1988) Structured Clinical Interview for DSM-III-R - Patient Version (SCID-P, 4/I/88). New York, Biometrics Research Department, New York State Psychiatric Institute.

66. Bewick V, Cheek L, Ball J. (2004) Statistics review 13: receiver operating characteristic curves. *Crit Care,* 8 (6), 508-512.

Supplementary Tables

Supplementary Table 5.1: Sociodemographic and psychiatric characteristics of the ROM reference (n=1295) patient (n=5066) groups.

Supplementary Table 5.2: Percentile scores and mean values in the ROM reference (n=1291) and patient (n=4368) groups for the subscales and total score of the Brief Scale for Anxiety (BSA).

Supplementary Table 5.3: Percentile scores and mean values in the ROM reference (n=651) and patient (n=657) groups for the subscales and total score of the PADUA Inventory Revised (PI-R).

Supplementary Table 5.4: Percentile scores and mean values in the ROM reference (n=630) and patient (n=1392) groups for the subscales and total score of the Panic Appraisal Inventory (PAI).

Supplementary Table 5.5: Percentile scores and mean values in the ROM reference (n=651/649) and patient (n=893/887) groups for the subscales and total score of the Penn State Worry Questionnaire (PSWQ) and the Worry Domains Questionnaire (WDQ).

Supplementary Table 5.6: Percentile scores and mean values in the ROM reference (n=651) and patient (n=1231/1237) groups for the subscales and total score of the Social Interaction and Anxiety Scale (SIAS) and the Social Phobia Scale (SPS) and.

Supplementary Table 5.7: Percentile scores and mean values in the ROM reference (n=1272) and patient (n=390) groups for the subscales and total score of the Impact of Events Scale (IES-R).

Supplementary Table 5.1: Sociodemographic and psychiatric characteristics of the ROM reference (n=1295) patient and (n=5066) groups.

	ROM reference group (n=1295)				ROM patient group (n=5066)			
	females		males		females		males	
Gender - n (%)	813	(62.8)	482	(37.2)	3260	(64.6)	1806	(35.7)
Age in years: mean (± SD)	39.7	(12.6)	41.2	(12.6)	35.8	(11.8)	37.8	(11.9)
Marital status[1] - n (%)								
Married/cohabitating	552	(67.9)	338	(70.1)	1455	(44.6)	751	(41.6)
Divorced/separated/widow	59	(7.3)	19	(3.9)	401	(12.3)	138	(7.6)
Single	202	(24.8)	125	(25.9)	1047	(32.1)	697	(38.6)
Housing situation[1] - n (%)								
Living alone	132	(16.2)	69	(14.3)	547	(16.8)	435	(24.1)
Living with partner	560	(68.9)	342	(71.0)	1492	(45.8)	767	(42.5)
Living with family	121	(14.9)	71	(14.7)	864	(26.5)	384	(21.3)
Educational status1,3- n (%)								
Lower	189	(23.2)	106	(22.0)	1226	(37.6)	641	(35.5)
Higher	624	(76.8)	376	(78.0)	1676	(51.4)	943	(52.2)
Employment status[1]:- n (%)								
Employed part-time	428	(52.6)	81	(16.8)	854	(26.2)	179	(9.9)
Employed full-time	222	(27.3)	332	(68.9)	402	(12.3)	584	(32.3)
Unemployed/retired	140	(17.2)	57	(11.8)	909	(27.9)	389	(21.5)
Work-related disability	23	(2.8)	12	(2.5)	738	(22.6)	434	(24.0)
Ethnic background[1]: - n(%)								
Dutch	710	(87.3)	440	(91.3)	2259	(69.3)	1246	(69.0)
Other ethnicity	103	(12.7)	42	(8.7)	642	(19.7)	340	(18.8)
MINI diagnoses: - n (%)								
Currently None	723	(88.9)	451	(93.6)	0***		0***	
Anxiety disorder (single)	42	(5.2)	12	(2.5)	1446	(44.4)	800	(44.3)
Anxiety disorder (comorbidity)	15	(1.8)	3	(0.6)	1814	(55.6)	1006	(55.7)
Other psychiatric disorder (without anxiety)	32	(3.9)	16	(3.3)	0***		0***	

*Data not available for 128 (2.4%) to 640 (11.8%) of patients
** Lower education: primary or vocational school; Higher education: college or university
***Selection criterion

Supplementary Table 5.2: Percentile scores and mean values in the ROM reference (n=1291) and patient (n=4368) groups for the subscales and total score of the Brief Scale for Anxiety (BSA).

	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Brief Scale for Anxiety (BSA)												
All participants	0	1	3	6	11	3.91 ± 3.92	6	12	16	21	28	16.36 ± 6.78
- Men	0	0	2	4	10	2.99 ± 3.23	6	11	16	21	29	16.01 ± 6.97
- Women	0	1	4	7	12	4.45 ± 4.19	6	12	16	21	28	16.56 ± 6.67

SD denotes standard deviation.

Supplementary Table 5.3: Percentile scores and mean values in the ROM reference (n=651) and patient (n=657) groups for the subscales and total score of the PADUA Inventory Revised (PI-R).

	ROM reference group (n=651)						ROM patient group (n=657)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
PADUA Inventory Revised (PI-R)												
Impulses												
All participants	0	0	0	1	4	0.84 ± 1.70	0	1	4	8	15	5.09 ± 5.18
- Men	0	0	0	1	4	0.83 ± 1.45	0	2	5	8	17	5.87 ± 5.40
- Women	0	0	0	1	5	0.84 ± 1.83	0	1	3	7	15	4.63 ± 5.00
Washing												
All participants	0	0	1	3	11	2.27 ± 3.83	0	1	6	17	32	9.86 ± 10.72
- Men	0	0	0	2	9	1.72 ± 2.99	0	0	4	13	31	8.41 ± 9.88
- Women	0	0	1	3	13	2.60 ± 4.22	0	1	7	17	34	10.73 ± 11.12
Checking												
All participants	0	1	3	6	12	4.07 ± 4.04	1	8	14	20	26	13.95 ± 7.60
- Men	0	1	3	6	12	4.18 ± 4.06	2	9	15	21	26	14.85 ± 7.29
- Women	0	1	3	6	12	4.01 ± 4.04	1	7	13	19	26	13.41 ± 7.74
Rumination												
All participants	1	3	7	11	18	7.71 ± 5.69	10	19	24	29	38	23.87 ± 8.23
- Men	0	3	6	9	16	6.38 ± 4.80	10	18	24	30	38	23.84 ± 8.24
- Women	1	4	7	12	20	8.50 ± 6.02	10	19	24	29	38	23.89 ± 8.24
Precision												
All participants	0	0	1	2	6	1.57 ± 2.20	0	3	6	11	18	7.38 ± 5.72
- Men	0	0	1	2	6	1.56 ± 2.24	0	2	6	11	17	6.83 ± 5.48
- Women	0	0	1	2	6	1.58 ± 2.17	0	3	7	12	19	7.71 ± 5.85
Total												
All participants	2	7	13	22	43	16.46 ±13.30	20	40	58	78	106	60.15 ± 26.21
- Men	1	6	12	20	38	14.66 ±12.36	21	41	58	77	103	59.81 ± 25.65
- Women	2	8	14	24	44	17.54 ±13.73	20	39	58	79	107	60.36 ± 26.57

Supplementary Table 5.4. Percentile scores and mean values in the ROM reference (n=630) and patient (n=1392) groups for the subscales and total score of the Panic Appraisal Inventory (PAI).

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Panic Appraisal Inventory (PAI)												
Anticipated panic												
All participants	0	1	7	17	37	10.82 ±12.16	14	32	47	62	82	47.42 ± 20.32
- Men	0	1	5	14	27	8.72 ± 10.34	13	30	44	58	79	44.55 ± 19.94
- Women	0	2	7	19	39	12.05 ±12.98	15	33	49	64	84	49.13 ± 20.36
Perceived consequences of panic - social												
All participants	0	0	0	2	17	2.87 ± 6.62	0	7	18	31	44	19.52 ± 13.99
- Men	0	0	0	1	9	1.91 ± 5.23	0	7	17	30	43	18.81 ± 13.79
- Women	0	0	0	3	21	3.43 ± 7.25	0	8	18	31	45	19.95 ± 14.10
Perceived consequences of panic - social												
All participants	0	0	0	4	14	3.01 ± 5.72	0	6	14	26	40	16.68 ± 12.79
- Men	0	0	0	3	11	2.46 ± 4.26	0	6	13	25	37	15.63 ± 12.17
- Women	0	0	0	4	17	3.34 ± 6.41	0	6	15	27	41	17.31 ± 13.11
Perceived consequences of panic - Loss of control												
All participants	0	0	0	3	17	3.13 ± 5.86	2	9	17	27	40	18.43 ± 11.89
- Men	0	0	0	2	11	2.04 ± 4.01	1	8	16	25	38	17.08 ± 11.39
- Women	0	0	0	5	19	3.78 ± 6.64	2	10	18	28	41	19.22 ± 12.11

Supplementary Table 5.4: continued

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Perceived consequences of panic - Total												
All participants	0	0	2	11	47	9.01 ± 15.03	10	31	52	75	108	54.63 ± 29.84
- Men	0	0	2	8	27	6.40 ± 11.27	11	29	50	70	104	51.52 ± 28.00
- Women	0	0	2	14	52	10.55 ± 16.68	10	33	54	78	111	56.47 ± 30.75
Perceived self-efficacy in coping with panic												
All participants	0	7	21	36	65	24.19 ± 21.30	29	49	62	76	90	61.48 ± 18.41
- Men	0	5	19	34	71	23.68 ± 23.46	30	46	60	74	88	59.67 ± 18.09
- Women	0	7	23	37	62	24.49 ± 19.94	29	50	63	77	90	62.55 ± 18.52

SD denotes standard deviation.

Supplementary Table 5.5: Percentile scores and mean values in the ROM reference (n=651/649) and patient (n=893/887) groups for the subscales and total score of the Penn State Worry Questionnaire (PSWQ) and the Worry Domains Questionnaire (WDQ).

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Penn State Worry Questionnaire (PSWQ)			n=651							n=893		
All participants	22	30	36	47	66	39.52 ± 13.19	48	62	69	74	79	66.95 ± 9.92
- Men	20	27	32	41	61	35.07 ± 11.15	44	59	67	73	78	65.08 ± 10.25
- Women	24	32	40	51	70	42.17 ± 13.60	48	64	70	75	80	68.05 ± 9.56
Worry Domains Questionnaire (WDQ)			n=649							n=887		
Relationships												
All participants	4	4	4	6	9	5.27 ± 1.99	4	7	10	14	18	10.28 ± 4.47
- Men	4	4	4	5	8	4.85 ± 1.32	4	6	8.5	12	17	9.22 ± 4.11
- Women	4	4	5	6	10	5.53 ± 2.26	4	7	10	14	19	10.89 ± 4.56
Lack of Confidence												
All participants	5	5	7	9	14	7.65 ± 3.30	7	12	16	20	24	15.95 ± 5.15
- Men	5	5	6	8	12	6.65 ± 2.34	6	12	15	19	23	15.21 ± 4.93
- Women	5	5	7	10	16	8.25 ± 3.63	7	13	17	21	24	16.39 ± 5.23
Aimless Future												
All participants	8	8	10	12	19	11.05 ± 4.00	9	15	21	27	35	21.07 ± 7.79
- Men	8	8	9	12	16	10.58 ± 3.13	9	16	22	27	35	21.70 ± 7.55
- Women	8	8	10	13	20	11.34 ± 4.41	9	14	20	27	35	20.70 ± 7.91
Work Incompetence												
All participants	3	3	4	6	9	4.77 ± 1.99	3	6	8	11	14	8.47 ± 3.37
- Men	3	3	4	5	8	4.46 ± 1.56	3	5	8	11	14	8.27 ± 3.34
- Women	3	3	4	6	10	4.96 ± 2.18	3	6	9	11	14	8.58 ± 3.39

Supplementary Table 5.5: continued

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Financial												
All participants	4	4	5	8	13	6.51 ± 3.04	4	7	10	15	19	10.86 ± 4.92
- Men	4	4	5	7	11	6.21 ± 2.72	4	7	11	15	19	10.99 ± 4.73
- Women	4	4	6	8	14	6.68 ± 3.20	4	6	10	15	20	10.79 ± 5.02
Health												
All participants	6	6	7	10	15	8.46 ± 3.31	7	11	14	19	26	15.20 ± 5.94
- Men	6	6	7	9	14	8.13 ± 2.91	6	10	14	19	26	14.80 ± 5.65
- Women	6	6	7	10	17	8.66 ± 3.51	7	11	14	20	27	15.43 ± 6.10
Total												
All participants	33	34	39	49	74	43.72 ± 13.62	44	65	81	97	123	81.82 ± 23.66
- Men	30	34	38	45	61	40.88 ± 10.00	44	63	80	96	121	80.19 ± 22.93
- Women	30	35	40	52	77	45.43 ± 15.14	44	65	82	99	126	82.78 ± 24.04

SD denotes standard deviation.

Supplementary Table 5.6. Percentile scores and mean values in the ROM reference (n=651) and patient (n=1231/1237) groups for the subscales and total score of the Social Interaction and Anxiety Scale (SIAS) and the Social Phobia Scale (SPS) and.

	ROM reference group						ROM patient group					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
Social Interaction and Anxiety Scale (SIAS)			n=651						**n=1231**	n=1231		
All participants	1	6	10	17	32	12.50 ± 9.34	18	33	44	54	68	43.70 ± 14.92
- Men	1	5	9	15	27	11.22 ± 8.09	16	32	43	52	66	41.81 ± 14.67
- Women	1	6	11	18	34	13.27 ± 9.94	20	34	46	56	69	45.15 ± 14.96
Social Phobia Scale (SPS)			n=651						n=1237	n=1237		
Relationships												
All participants	0	2	4	8	19	6.04 ± 6.57	11	22	33	47	64	35.06 ± 16.76
- Men	0	1	3	7	14	4.54 ± 5.17	10	20	31	45	64	33.24 ± 16.59
- Women	0	2	5	9	22	6.93 ± 7.14	12	24	35	48	65	36.46 ± 16.76

SD denotes standard deviation.

Supplementary Table 5.7: Percentile scores and mean values in the ROM reference (n=1272) and patient (n=390) groups for the subscales and total score of **the Impact of Events Scale (IES-R).**

	ROM reference group (n=630)						ROM patient group (n=1392)					
	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD	P_5	P_{25}	P_{50} (median)	P_{75}	P_{95}	Mean ± SD
IES-R												
Intrusions												
All participants	0	0	1	5	15	3.51 ± 5.03	5	15	20	24	31	19.52 ± 7.36
- Men	0	0	1	4	13	2.81 ± 4.22	4	15	20	24	31	18.98 ± 7.30
- Women	0	0	1	6	16	3.92 ± 5.41	5	15	20	25	31	19.73 ± 7.39
Avoidance												
All participants	0	0	0	4	14	2.72 ± 4.80	5	13	18	22	29	17.30 ± 6.96
- Men	0	0	0	3	12	2.25 ± 4.10	5	13	17	21	26	16.83 ± 6.41
- Women	0	0	0	4	14	2.99 ± 5.14	4	13	18	22	29	17.48 ± 7.17
Hyperarousal												
All participants	0	0	0	2	8	1.77 ± 3.14	4	11	15	18	22	14.39 ± 5.39
- Men	0	0	0	2	7	1.52 ± 2.58	5	10	14	18	22	14.08 ± 5.09
- Women	0	0	0	2	10	1.91 ± 3.42	2	11	15	18	23	14.51 ± 5.52
Total												
All participants	0	0	2	11	36	7.99 ± 11.97	19	43	53	62	78	51.20 ±17.12
- Men	0	0	2	9	29	6.59 ± 10.03	20	40	52	61	74	49.89 ±16.24
- Women	0	0	3	13	38	8.82 ± 12.92	19	43	53	62	79	51.72 ±17.46

SD denotes standard deviation.