



Research Article

DOI: 10.36959/362/489

Measurement of Depression and Anxiety in Chronic Low Back Pain and Fibromyalgia: A Validity Comparison of Four Patient-Reported Outcome Measures

Felix Angst, MD, MPH¹, Prof. Thomas Benz, PhD^{1,2}, Susanne Lehmann, RS¹, Prof. Erich Seifritz, MD³, Prof. Peter S. Sándor, MD¹ and Katrin Angst, MD³



¹Research Department, Rehaklinik Bad Zurzach, ZURZACH Care Group, Bad Zurzach, Switzerland

²ZHAW Zurich University of Applied Sciences, School of Health Sciences, Institute of Physiotherapy, Winterthur, Switzerland

³Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital Burghölzli, University of Zürich, Zürich, Switzerland

Abstract

Background: Depression and anxiety are highly prevalent in chronic pain and predictive of course. The aim was to evaluate and compare the content validity and cross-sectional and longitudinal construct validity (responsiveness) of the depression and anxiety scales of four patient-reported outcome measures in chronic low back pain and fibromyalgia, before and after a standardized pain program.

Methods: Longitudinal cohort study (n = 118 to 257) using qualitative content analysis, cross-sectional intra-individual correlations, responsiveness (by effect sizes, ESs) and score differences to population-based norms (by standardized mean differences, SMDs).

Results: The Short Form 36 (SF-36) Mental health (5 items) and the Multidimensional Pain Inventory (MPI) Negative mood (3 items), including the core symptoms, are more responsive than the Hospital Anxiety and Depression Scale (HADS) (7 items each) and the Symptom CheckList-90R (SCL-90R) (10 & 13 items), which cover additional content. For example, ES = 0.535 (MPI), 0.466 (SF-36), 0.383 (SCL-90R Depression), 0.169 (SCL-90R Anxiety) in low back pain. Pair-wise scale correlations at baseline ranged from 0.589 to 0.790 comparable to the literature. The SCL-90R Depression had the lowest SMD = -2.362. The number of items of the scales correlated with the ESs by -0.792.

Conclusions: All four instruments showed high content and construct validity. The shorter scales (SF-36, MPI) are more specific but less sensitive, whereas the longer scales (HADS, SCL-90R) are more sensitive but less specific. This reciprocity allows scope for optimization. We recommend using the generic SF-36 and the condition-specific SCL-90R to combine best specificity and sensitivity.

Keywords

Depression, Anxiety, Pain, Fibromyalgia, Low back pain, SF-36, SCL-90

Abbreviations

ACR: American College of Rheumatology; BMI: Body Mass Index; CI: Confidence Interval; EK AG: Ethics (K)committee of the canton Aargau, Switzerland; ES: (baseline-standardized) effect size; FSQ: Fibromyalgia Survey Questionnaire; HADS: Hospital Anxiety and Depression Scale; HRQL: Health-Related Quality of Life; MPI: (West Haven-Yale) Multidimensional Pain Inventory; N: Number (of subjects); PROM: Patient-Reported Outcome Measure; ROC: Receiver Operating Characteristic; RCT: Randomized Controlled Trial; SCL-90R: Symptom CheckList-90Revised; SCL-90S: Symptom CheckList-90Standard; SF-36: Short Form 36; SMD: Standardized Mean Difference; ZISP: Zurzach Interdisciplinary Schmerz (Pain) Program

Introduction

Depression and anxiety are complex syndromes characterized by various signs and symptoms [1,2] (Table 1).

As the two main affective health dimensions, they co-occur with and influence other health disorders, especially chronic pain conditions, and are potential predictors of a disease's course and the effects of therapy [3,4]. According to ICD-

Table 1: Content and construct coverage: four scales examined by DSM-5 and ICD-11 defined syndromes.

Depression Symptoms (DSM-5, ICD-11)	SF-36	MPI	HADS	SCL-90R
Depressed mood	+	+	+	+
Diminished interest			+	+
Sorrow	+		+	+
Hopelessness			(+)	+
Listlessness			+	+
Joylessness	+		+	+
De-/Increased appetite/weight loss/gain				
Slowing down of thought				
Diminished ability to think or concentrate				
Indecisiveness				
Feeling worthless	(+)			
Feeling of excessive or inappropriate guilt				
Reduced self-confidence				+
Feeling of loneliness				+
Reduction of physical movement			+	+
Fatigue/loss of energy				+
Intolerance of performance				+
Insomnia/sleep disorder				
Agitation	+	(+)	+	+
Petulance		+		+
Loss of libido				+
Morning low				
Suicide ideation/plan/attempt				+
Anxiety Symptoms	SF-36	MPI	HADS	SCL-90R
Anxiety/worry		+	+	+
Panic attacks			+	+
Nervousness	+		(+)	+
Trembling				+
Edginess or restlessness			+	+
Irritability			+	+
Numbness/dizziness				
Loss of concentration/mind goes blank				
Depersonalization/derealization				
Loss of control				
Sweating				
Heart palpitation				+
Angina pectoris				
Muscle aches or soreness				
Insomnia/sleep disorder				
Fatigue				
Dyspnoea				
Abdominal pain			+	

11, “chronic primary musculoskeletal/widespread pain is chronic pain (...) that is characterized by significant emotional distress (anxiety, anger/frustration or depressed mood) or functional disability ...”, a definition that applies to both low back pain and fibromyalgia [2]. Anxiety is especially highly associated with low back pain [4,5], whereas depression and catastrophizing are highly prevalent in, and important predictors of, fibromyalgia [3,6-8].

While some patient-reported outcome measures (PROMs) are specifically designed for depression and anxiety, other, generic, instruments also include affective health scales. However, there has been little examination of the comparative strengths and limitations of most of these instruments [9] and there is no data from a direct comparison of the four PROMs used together in our study. These were: The Short-Form 36 (SF-36) Mental health scale; the West Haven-Yale

Multidimensional Pain Inventory (WHYMPI, abbreviated to MPI) Negative mood scale; both scales of the Hospital Anxiety and Depression Scale (HADS), and the Symptom Checklist 90Revised (SCL-90R) Anxiety and Depression scales.

The aim was to evaluate and compare the content, cross-sectional and longitudinal construct validity of the scales for depression and anxiety of four PROMs in chronic low back pain and fibromyalgia before and after our inpatient pain program. We used qualitative content analysis, quantitative intra-individual correlations, sensitivity to change (responsiveness) and comparison to normative, population-based severity levels.

Materials and Methods

Patients and data sampling

Patients were consecutively referred by their family physician, internist or psychiatrist to the Rehaklinik Bad Zurzach, Switzerland for a four-week inpatient pain program between 1998 and 2023. The inclusion criteria were as follows [10,11]: 1) Age 18-90 years. 2) Chronic (≥ 3 months) low back pain according to ICD-11, or fibromyalgia syndrome according to the diagnostic criteria of the American College of Rheumatology (ACR) [2,12,13]. The ACR 1990 criteria defined fibromyalgia by 11/18 positive “tender points” (definition used for the ZISP 1 sample, see below) and the ACR 2011/2016 criteria by use of the Fibromyalgia Survey Questionnaire (FSQ) (applied in the ZISP 2 sample, see below) [12,13]. The ACR 2016 criteria have been described in detail in an earlier paper [10]. The study was approved by the ethics committee of Aarau, Canton Aargau, Switzerland (EK AG 2008/026) and written informed consent was obtained from all study participants.

Intervention

The Zurzach Interdisciplinary Schmerz (Pain) Program (ZISP) is a well-established, thoroughly evaluated, comprehensive interdisciplinary pain program. Inpatient rehabilitation consists of education sessions, active physiotherapies, various operant and cognitive behavioral therapies, and medication adaptation, delivered in more than 100 hours of interventions over four weeks. The ZISP has been extensively described in an earlier publication [11,14].

Measures

In chronic pain, it is more useful to measure affective symptoms on the severity continuum of the relevant syndrome than to classify patients according to the diagnostic thresholds defined in the diagnostic manuals [3,10,11,14-16]. In the first phase of our study 1998-2009 (ZISP 1 sample), the HADS, the SF-36 version 1, and the MPI were applied [15]. The HADS was one of the few affective health instruments then available and was designed especially for non-psychiatric settings [17]. In the revised measurement set applied since 2010 (ZISP 2 sample), the HADS was replaced by the SCL-90R, and the SF-36 version 1 by the updated version 2. There were two reasons for this adjustment. First, the HADS depression scale was less responsive than the affective health scales of

the MPI and the SF-36 ([18]: see Table 2). Second, we wanted to catch more signs and symptoms of the syndromes than was possible with the HADS. The validated German versions of all four instruments were used. The items of the four scales examined with their content and construct are listed in Table 2.

The SF-36 is the most widely used PROM of generic health-related quality of life (HRQL), and its validity has been demonstrated in numerous settings [19-22]. The SF-36 mental health scale has five items relating to affective symptoms, with the answers graded according to 5 (version 1) and 6 (version 2) levels from “always” to “never”. Deceptively simple but able to cover a substantial part of the depression construct (Table 2), the items compose a scale with excellent criterion, content and construct validity [3]. For example, the SF-36 Mental health scale correlated to the Mental Health Inventory by 0.96 [21]. The 6-item response levels (version 1) correlated by 0.998 to 7 predefined depression criteria [3,22]. Normative levels, specific for sex and age (5-year categories), are available from the German population survey of 1998 ($n = 6945$) [20]. The regression coefficients were personally obtained from the survey’s authors at the time of publication, allowing calculation of the norm scores stratified by no/at least one comorbid condition.

The MPI assesses pain and its consequences in terms of symptoms/disability, activity, behavior, mood, and social relationships and was designed and validated for chronic pain conditions [23]. The Negative mood scale of the MPI is composed of 3 items that assess probands’ mood, anxiety and irritability on a 7-level scale (extremely low to extremely high) [23,24] (Table 2). We found no studies that compared the MPI Negative mood to other affective scales and no normative population data.

The HADS is a short self-rating scale of anxiety and depression with 7 items for each condition; it was designed especially for non-psychiatric settings [17,25-27]. The questionnaire has a long history of application in psychology and psychiatry and has been well-tested in large population and patient surveys. An overview of the correlations to various affect scales can be found on page 28 of the German manual [26]. We used the sex-specific and age-specific norms (10-year categories) from a representative German population survey ($n = 4410$) [27].

The SCL-90R is probably the oldest PROM in use worldwide; it covers a broad spectrum of mental HRQL. Although its origins date back to 1917, the first official version was not published until 1973, since when it has a long history of application and validation [28-31]. The 1973 version was translated into German by Angst J, et al. in 1976 and used to screen for the severity of mental distress in the selection of the sample for the Zurich study, a longitudinal epidemiologic study of the general population [28,29]. That 1976 Swiss Standard German version is the word for word translation of the SCL-90R published by the developer Derogatis LR, which is still in use [30]. A Standard High German language version of the SCL-90 (SCL-90S) was published in 2014 [31].

Table 2: Scales and item content of the four PROMs used.

Tool	Scale	Items
SF-36	Mental health	Have you been very nervous?
	5 items	Have you felt so down in the dumps that nothing could cheer you up?
	5 response levels	
		Have you felt calm and peaceful?
		Have you felt downhearted and depressed?
MPI	Negative mood	Rate your overall mood during the past week
	3 items	During the past week how irritable have you been?
	7 response levels	During the past week how tense or anxious have you been?
		Have you been happy?
HADS	Depression	I still enjoy the things I used to enjoy
	7 items	I can laugh and see the funny side of things
	4 response levels	I feel cheerful
		I feel as if I am slowed down
		I have lost interest in my appearance
		I look forward with enjoyment to things
		I can enjoy a good book or radio or TV program
SCL-90R	Depression	Loss of sexual interest or pleasure
	13 items	Feeling low in energy or slowed down
	5 response levels	Thoughts of ending your life
		Crying easily
		Feeling of being trapped or caught
		Blaming yourself for things
		Feeling lonely
		Feeling blue
		Worrying too much about things
		Feeling no interest in things
		Feeling hopeless about the future
		Feeling everything is an effort
		Feelings of worthlessness
HADS	Anxiety	I feel tense or 'wound up'
	7 items	I get a sort of frightened feeling as if something awful is about to happen
	4 response levels	
		Worrying thoughts go through my mind
		I can sit at ease and feel relaxed
		I get a sort of frightened feeling like 'butterflies' in the stomach
		I feel restless as I have to be on the move
	I get sudden feelings of panic	
SCL-90R	Anxiety	Nervousness or shakiness inside
	10 items	Trembling
	5 response levels	Suddenly scared for no reason
		Feeling fearful
		Heart pounding or racing
		Feeling tense or keyed up
		Spells of terror or panic
		Feeling so restless you couldn't sit still
		Feeling that familiar things are strange or unreal
	Feeling pushed to get things done	

The earlier German version of 1976 was chosen for this study because of the closeness of the translation of the content of the anxiety items to the original version and because the current SCL-90S version was not available at that time (start of ZISP 2 in 2010). The differences are small however, which allowed us to use the normative data from the representative German population survey (n = 2025), stratified by sex and 10-year age groups [31].

Analysis

The samples were descriptively characterized by socio-demographic and disease-relevant data, using a standardized form [3]. To determine the scales, participants had to have responded to 3/5 items of SF-36 Mental health, 2/3 items of MPI Negative mood, 6/7 items of both HADS scales, 9/13 items of SCL-90R Depression, and 7/10 items of SCL-90R Anxiety [19,24,26,31]. All scales are calculated as unweighted means of the component items. All scores were scaled from 0 = no depression/anxiety to 100 = maximum depression/anxiety, i.e., high scores mean a heavy burden/many symptoms; this is the scaling originally provided for the MPI, the HADS and the SCL-90R and applying to most PROMs that measure affective health [24,26,31]. The scores of the SF-36, where 0 = worst health, maximum symptoms/disability and 100 = best health, no symptoms, were therefore converted by deducting the original score from 100 [19].

The content and construct of the scales were qualitatively analyzed by allocating specific symptom content according to ICD-11/DSM5 to the items of the scales used (Table 1: + = symptom addressed) [1,2]. Quantification of the construct overlaps was provided by parametric Pearson correlations [11,32]. Detailed information regarding types of validity and their meaning can be found in [11,32].

Differences between patients' scores and the norms (between groups) were quantified by standardized mean differences (SMDs). The SMD equals the inter-individual score difference divided by the "pooled" standard deviation of the two samples; it is identical to "Cohen's delta" and is the state of the art for quantifying differences, e.g. in randomized controlled trials [33,34]. A negative SMD reflects a higher burden of depression/anxiety than expected from the population norm. Expressing the distance between the patients' and the norm scores, the SMD serves as a measure of discriminant validity [32].

As an intra-individual (within group) parameter to measure change over time, the (baseline-standardized) effect size (ES) is equal to the mean score difference between the baseline (entry into the pain program) and the follow-up scores (discharge) divided by the standard deviation of the baseline scores; it is identical to the "Glass's delta" [33-35]. A positive ES reflects improvement between entry and discharge. The ES quantifies the responsiveness, also known as sensitivity to change, which is in fact a measure of the specificity of the scale's construct [32].

Standardized parameters are dimensionless, allow comparison between scores with different scaling, are most often symmetrically or normally distributed, and result in more stable levels at the end of the closed scales (0-100)

[33,34]. For both parameters, 95% confidence intervals (CI) were reported. If zero = no effect/no difference is excluded from the 95% CI, this means that there is statistical significance on the one-sided type I error $p = 0.025$. The corresponding type I error p is the same as obtained by the Student's t-test.

Between-group differences were tested for statistical significance by the chi-square test for categorical frequency data and the Student's t-test for continuous data. Sample sizes of $n \geq 100$ provide symmetrically distributed data, especially for score differences, with small 95% confidence intervals and allow the use of parametric statistics and significance tests [32,36]. All analyses were performed with the statistical software package IBM SPSS 29.01 for Windows® (SPSS Inc., Chicago, IL, USA).

Results

Patients

The typical participant with low back pain was female (76.9% in sample 1 and 63.8% in sample 2), aged 47.0-48.3 years, slightly overweight (median BMI = 25.7 and 27.4), living with a partner (about 3/4), educated to vocational training level (> 50%), and had 2 (median) comorbidities (Table 3). A typical patient with fibromyalgia was female (96.6% and 85.5%), aged 46 and 49 years, slightly overweight (median BMI = 25.3 and 28.2), living with a partner (about 3/4), educated to vocational training level, and had a median of 2 (ZISP 1) and 4 (ZISP 2) comorbid conditions.

In the more recent sample, the ZISP 2, there were statistically significantly more men, more comorbidities, more overweight participants, and more pain than in the ZISP 1 sample in both conditions. Living situation and education attainment levels were comparable.

Content and construct coverage of the four scales

The qualitative analysis of the content and construct of the items is summarized in Table 1. The SF-36 Mental health and the MPI Negative mood questionnaires focus mainly on depressive symptoms and to a lesser extent on anxiety. The HADS depression scale covers the most important depressive symptoms: depressed mood, diminished interest, sorrow, hopelessness, listlessness and reduction in physical movement. The anxiety scale probes anxiety/worry, panic attacks, nervousness, edginess/restlessness, and irritability. The SCL-90R covers the same content as the HADS but in addition includes important depressive symptoms, such as reduced self-confidence, feeling of loneliness, fatigue/loss of energy, intolerance of performance, loss of libido, and items on the vegetative symptoms of anxiety, such as trembling and heart palpitation.

Descriptive score data before and after the pain program

At admission to the pain program, the participants' scores on the SF-36 Mental health scale lay mid-range (0 = best-100 = worst), with negligible differences between the ZISP

Table 3: Socio-demographic and disease-relevant data.

Characteristic	Low back pain					Fibromyalgia				
	ZISP 1	(n = 173)	ZISP 2	(n = 257)	p	ZISP 1	(n = 118)	ZISP 2	(n = 131)	p
Female	133	76.9	164	63.8	< 0.001	114	96.6	112	85.5	0.003
Living with partner/o. person	133	76.9	187	72.8	0.169	89	75.4	104	79.4	0.454
Educational attainment level					0.219					0.132
Basic school (8-9 years)	53	30.6	61	23.7		33	28.0	28	21.4	
Vocational training	89	51.4	138	53.7		66	55.9	69	52.7	
College/high school/university	31	17.9	58	22.6		19	17.9	34	26.0	
Comorbidities					0.030					< 0.001
none	16	9.2	20	7.8		16	13.6	6	4.6	
1	42	24.3	58	22.6		34	28.8	9	6.9	
2	47	27.2	51	19.8		40	33.9	22	16.8	
3	36	20.8	45	17.5		18	15.3	23	17.6	
≥ 4	32	18.5	83	32.3		10	8.5	71	54.2	
Age (years): mean (sd)	47.0	(12.7)	48.3	(11.8)	0.273	46.0	(10.0)	49.0	(9.9)	0.014
minimum-maximum	21.4	-79.7	20.5	-77.1		19.8	-67.3	18.2	-65.8	
BMI (kg/m²): mean (sd)	25.7	(4.8)	27.4	(5.3)	0.001	25.3	(4.6)	28.2	(6.2)	< 0.001
minimum-maximum	15.6	-44.7	17.3	-46.6		17.4	-41.1	16.9	-52.1	
MPI Pain severity: mean (sd)	65.2	(19.9)	75.1	(14.9)	< 0.001	65.0	(18.7)	78.2	(14.4)	< 0.001

Legend: n: number of subjects; sd: standard deviation; o: other; BMI: Body Mass Index; kg: kilogram; m: meter; MPI: Multidimensional Pain Inventory (0 = no, 100 = maximal pain); ZISP = Zuzach Interdisciplinary (Schmerz) Pain Program (samples 1 and 2); p = type I error of the comparison of sample ZISP 1 versus ZISP 2

1 and 2 samples: $p = 0.300$ in low back pain and $p = 0.919$ in fibromyalgia, t-test; not shown in Table 4 (Table 4). The same was true on the MPI Negative mood scale ($p = 0.559$ and 0.439). All baseline score levels reflected a far higher burden than expected from the population norms, which reached statistical significance: SMDs between -0.904 to -1.691 pooled standard deviations in low back pain, -1.641 to -2.362 in fibromyalgia, zero excluded by all 95%-CIs.

Compared to the SCL-90R, the HADS found more severe depression in low back pain (mean 45.1 versus 38.6, $p = 0.001$) but comparable depressive symptoms in fibromyalgia (46.4 vs. 50.5, $p = 0.126$). Anxiety scores were much higher on the HADS than on the SCL-90R in both conditions (48.8 vs. 25.0 and 55.4 vs. 36.4, both $p < 0.001$). These findings remained consistent at discharge.

Responsiveness, longitudinal construct validity

The score changes between entry and discharge showed moderate effects in low back pain, with ESs between 0.354 and 0.535, except for SCL-90R anxiety (Table 4). The effects were stronger in fibromyalgia, especially in the ZISP 1 sample (e.g. MPI Negative mood: 0.732 vs. 0.542 in ZISP 2, $p < 0.001$, p values not shown in Table 4) and in anxiety (HADS vs. SCL-90R: 0.458 vs. 0.165, $p < 0.001$). In depression, the HADS and the SCL-90R measured comparable improvements: Low back pain ES = 0.354 vs. 0.383 and 0.477 vs. 0.411 in fibromyalgia.

Discriminant construct validity

The greatest differences between the patients' and the norm scores were measured by the SCL-90R Depression scale. Its SMD of -1.691 in low back pain was significantly lower than the -1.327 measured on the SF-36 Mental health ($p < 0.001$, p values not shown in Table 4) and the SCL-90R Anxiety scale (SDM = -1.115 , $p < 0.001$).

In fibromyalgia, the SMD of -2.362 on SCL-90R Depression was significantly lower than the -1.692 on the SF-36 Mental health ($p < 0.001$) and -1.641 ($p < 0.001$) on the SCL-90R Anxiety scales (Table 4, ZISP 2). All other pair-wise comparisons, especially those with the HADS scales, showed statistically comparable levels.

Convergent construct validity measured by cross-sectional correlations, compared with levels in the literature

In the construct of depression, the correlations of the four examined scales showed high levels and varied on comparable levels for both conditions (Pearson correlations: 0.568-0.768, Table 5). The SF-36 correlated slightly more strongly with the HADS and the SCL-90R (0.734-0.740) than did the MPI (0.596-0.715). The same was true for the construct of anxiety SF-36: 0.653-0.735 and MPI: 0.589-0.700.

The correlation levels of SF-36 Mental health were comparable to those reported in the literature, except for

Table 4: Descriptive score data: baseline (entry), follow-up (discharge) and differences baseline to follow-up.

		Norm		Entry		Discharge		Difference		ES (entry-discharge)			SMD (entry-norm)		
		m	sd	m	sd	m	sd	m	sd	m	95%	CI	m	95%	CI
LBP	ZISP 1 (n = 173)														
SF-36	Mental health	31.7	3.7	53.3	20.4	45.0	19.7	8.2	16.4	0.402	0.282	0.523	-0.904	-1.127	-0.682
MPI	Negative mood			56.4	23.6	44.3	22.5	12.1	21.3	0.515	0.379	0.650			
HADS	Depression	21.7	4.0	45.1	21.5	37.5	23.6	7.6	15.6	0.354	0.246	0.463	-1.510	-1.750	-1.270
HADS	Anxiety	23.4	2.0	48.8	21.0	43.2	20.1	5.7	14.9	0.272	0.165	0.378	-1.701	-1.949	-1.454
LBP	ZISP 2 (n = 257)														
SF-36	Mental health	31.5	3.9	51.2	20.6	41.6	20.7	9.6	18.2	0.466	0.357	0.574	-1.327	-1.519	-1.135
MPI	Negative mood			57.7	22.1	45.8	23.7	11.9	24.0	0.535	0.402	0.668			
SCL-90R	Depression	12.0	1.9	38.5	22.1	30.1	21.5	8.5	15.1	0.383	0.299	0.467	-1.691	-1.893	-1.489
SCL-90R	Anxiety	7.6	1.5	25.0	22.0	21.3	21.4	3.7	13.3	0.169	0.095	0.243	-1.115	-1.302	-0.929
FM	ZISP 1 (n = 118)														
SF-36	Mental health	33.1	2.8	58.3	17.9	46.0	17.1	12.4	16.4	0.692	0.525	0.858	-1.967	-2.281	-1.654
MPI	Negative mood			61.3	20.4	46.3	22.3	14.9	21.2	0.732	0.542	0.922			
HADS	Depression	21.3	2.9	46.4	19.5	37.1	21.6	9.3	15.2	0.477	0.335	0.620	-1.794	-2.098	-1.490
HADS	Anxiety	24.1	1.4	55.4	20.5	45.8	21.7	9.4	17.1	0.458	0.306	0.610	-2.145	-2.468	-1.823
FM	ZISP 2 (n = 131)														
SF-36	Mental health	33.3	3.2	58.1	20.4	47.9	21.0	10.2	15.7	0.500	0.367	0.633	-1.692	-1.976	-1.408
MPI	Negative mood			63.3	21.0	51.9	24.1	11.4	21.3	0.542	0.367	0.718			
SCL-90R	Depression	12.2	1.9	50.5	22.8	41.1	25.3	9.4	16.1	0.411	0.289	0.534	-2.362	-2.690	-2.034
SCL-90R	Anxiety	7.7	1.6	36.4	24.6	32.4	25.6	4.1	13.6	0.165	0.069	0.260	-1.641	-1.932	-1.351

Legend: LBP: Low Back Pain; FM: Fibromyalgia; ZISP: Zurich Interdisciplinary (Schmerz) Pain Program (samples 1 and 2); n: number; ES: (baseline) Effect Size; SMD: Standardized Mean Difference; CI: Confidence Interval; M: Arithmetic Mean; Sd: Standard Deviation. Scaling: 0 = no; 100 = maximal symptoms.

the outstanding and almost perfect correlation to the Mental Health Inventory (MHI) of 0.96 [21]. Depression and anxiety on the HADS correlated almost equally by 0.680 and 0.663 on the HADS and by 0.790 and 0.779 on the SCL-90R. All correlations between the two samples (compare superscript 1 to 2 and 3 to 4 in Table 5) and the two conditions (compare superscript 1 to 3 and 2 to 4 in Table 5) showed comparable levels and no significant differences ($p > 0.105$).

Association of scale length and responsiveness

The number of items of the scales correlated with the score differences (entry to discharge) by -0.762 (95%-CI: -0.916,-0.415) and with the ESs by -0.792 (-0.927,-0.476) on the 16 data points (4 scales, 2 samples, 2 diagnoses; Spearman rank correlations; data from Table 4).

Discussion

We examined and compared the content and construct validity of the SF-36, the MPI, the HADS, and the SCL-90R scales in the measurement of the syndromes of depression and anxiety in two chronic pain disorders and two comparable samples (having the same severity levels on the SF-36 Mental health and the MPI Negative mood).

Based on the qualitative analysis, the SF-36 Mental health and the MPI Negative mood scales focus mainly on depression but they also cover some anxiety symptoms. The HADS scales gather further important signs and symptoms and the SCL-90R captures most content in both domains. Quantitatively, the SF-36, the HADS and SCL-90R Depression yielded the highest correlations (construct overlap, convergent validity), and the MPI and the SF-36 the highest levels of responsiveness (specificity). In addition, the SCL-90R Depression scale showed the highest discriminant validity, with the largest gap between the patients' and the population norm scores.

A statistically significant inverse association (correlation -0.792) was found between the number of items, included symptoms of a scale and the scales' responsiveness (ESs) in both conditions and both samples. This may be the effect of a degree of dilution by less depression-defining symptoms covered by the SCL-90R and the HADS (e.g. reduction of physical movement). The most important and disease-defining symptoms, such as depressed mood, diminished interest/sorrow and anxiety itself are assessed by all four instruments. Those "core" symptoms dominate the short scales of the SF-36 and the MPI. The additional symptoms of the HADS and SCL-90R may be less prevalent and therefore scored as less severe by many participants. Calculated as the unweighted mean of all items, the longer scale thus results in a lower score than the shorter scale with its focus on the disease-defining core symptoms only.

From this perspective, the SF-36 Mental health and the MPI, while being more specific (responsive) in the measurement of depression and anxiety, are also less sensitive. Conversely, the SCL-90R, which measures the broadest symptom spectrum, is comparatively the most sensitive but least specific instrument, followed by the HADS. This is consistent with the well-known phenomenon of the inverse and reciprocal relationship of

sensitivity and specificity, as is clearly demonstrated in the receiver operating characteristic (ROC) analysis [37].

Overall, our correlation levels were comparable to those reported in the literature and showed no statistical differences [26,38-44]. While there is no "gold standard" to measure anxiety and depression, the numerous studies that have evaluated the validity of the four PROMs compared in the present study have proven their high criterion, content and construct validity; our results are in line with them [32].

The correlations tended to be slightly higher within depression than within anxiety, which may indicate that the anxiety syndrome is more difficult to capture through symptoms other than anxiety itself and feelings of panic, through vegetative adrenergic symptoms for example, which are unspecific. In anxiety, the "dilution effect" described above was most obvious when comparing the SCL-90R to the HADS. The SF-36 showed slightly higher correlations to the MPI, HADS and SCL-90R (0.653 to 0.735) than the MPI to the HADS and SCL-90R (0.568 to 0.700). However, the highest correlation in our data was between SCL-90R Depression and Anxiety (0.790), which demonstrates that the two syndromes are closely linked and co-occur frequently in the person affected [45].

The most important strengths of our study are the comparison of different instruments covering the same construct, the large sample sizes, and the comparison of the patients' disease severity to population-based norms. The score data of the SF-36 and the MPI in the two samples, ZISP 1 and 2, show comparable severity levels of depression and anxiety, allowing indirect comparison of the HADS to the SCL-90R, which was used in only one of the samples. Intra-individual comparison of responsiveness is not affected by bias caused by inter-individual differences. Furthermore, the notable disease severity of our inpatients -all SMDs presented large distances to the norms- prevented floor and ceiling effects at the end of the closed scale (0-100), which might affect the measurement of change during the course. Finally, comparative correlation data were reviewed in an extensive body of literature.

A weakness of our study may be the limited generalizability of our inpatient results to less severely affected persons, such as outpatients, and to very severely affected inpatients, who were excluded from the assessment because of physical, mental or language limitations. The socio-demographic and disease-relevant characteristics of the two samples, ZISP 1 and 2, differed with regard to sex, body weight, the number of comorbidities, and pain severity. Moreover, historically the definition of fibromyalgia changed between the two settings from number of tender points to application of the FSQ. Both reasons limited the direct comparison of the HADS with the SCL-90R.

Conclusions

All four PROMs examined have been convincingly validated in the numerous settings documented in the literature; they performed well in measuring the complex syndromes of depression and anxiety in chronic pain inpatients. The short

scales of the SF-36 and the MPI focus on the core signs of the two syndromes, whereas the longer scales of the HADS and especially the SCL-90R assess additional items that are less syndrome-defining. This leads to the finding that the shorter scales are more specific but less sensitive, while the longer scales are more sensitive but less specific. This confirmation of the well-known reciprocal relationship of the two measurement characteristics allows optimization in the choice of instrument best adapted to the scope of planned research. We recommend using the SF-36 as a generic and the SCL-90R as a condition-specific tool, in order to combine maximum sensitivity and specificity.

Declarations

Ethics approval and consent to participate and for publication

The study was approved by the ethics committee of Aarau, Canton Aargau, Switzerland (EK AG 2008/026) and written informed consent was obtained from all study participants.

Consent for publication

Not applicable. Data contain no personal data.

Availability of data and material

All data and materials are free available. Please contact the corresponding author for data requests.

Competing interests

There are no conflicts of interest or competing interests to declare.

Funding

There are no sources of funding to declare.

Authors' contributions

FA planned the study, analyzed and interpreted the data and wrote the study report. As an experienced study nurse, SL helped in study planning, collected the data, and contributed to analyze and interpret the data. PS provided the resources for carrying out the study, helped in planning the study and in interpreting the data. TB, experienced physiotherapist, contributed to plan the study, in interpreting the data and to write the study report. KA and ES are experienced psychiatrists, provided important clinical information and helped in interpreting the data. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by the Zurzach Rehabilitation Foundation SPA, Bad Zurzach, Switzerland. We thank all patients for their voluntary participation in the study. We further thank Elizabeth Kyrke for the linguistic editing of the text.

References

1. American Psychiatric Association (2013) *DSM-5: Diagnostic and statistical manual of mental disorders*. (5th edn), Washington, D.C: American Psychiatric Publishing.
2. World Health Organization (2022) *International Classification of Diseases, Eleventh Revision (ICD-11)*.
3. Angst F, Benz T, Lehmann S, et al. (2020) Extended overview of the longitudinal pain-depression association: A comparison of six cohorts treated for specific chronic pain conditions. *J Affect Disord* 273: 508-516.
4. Leeuw M, Goossens MEJB, Linton SJ, et al. (2007) The fear-avoidance model of musculoskeletal pain: Current state of scientific evidence. *J Behav Med* 30: 77-94.
5. Angst F, Angst J, Ajdacic-Gross V, et al. (2017) Epidemiology of back pain in young and middle-aged adults: A longitudinal population cohort survey from age 27-50 years. *Psychosomatics* 58: 604-613.
6. Angst F, Lehmann S, Sandor PS, et al. (2022) Catastrophizing as a prognostic factor for pain and physical function in the multidisciplinary rehabilitation of fibromyalgia and low back pain. *Eur J Pain* 26: 1569-1580.
7. Gota CE, Kaouk S, Wilke WS (2017) The impact of depressive and bipolar symptoms on socioeconomic status, core symptoms, function and severity of fibromyalgia. *Int J Rheum Dis* 20: 326-339.
8. Salaffi F, Di Carlo M, Carotti M, et al. (2019) The impact of different rheumatic diseases on health-related quality of life: A comparison with a selected sample of healthy individuals using SF-36 questionnaire, EQ-5D and SF-6D utility values. *Acta Biomed* 89: 541-557.
9. Lam RW, Michalaak EE, Swinson RP (2006) *Assessment scales in depression and anxiety – CORPORATE*. (Servier Edn), CRC Press.
10. Angst F, Benz T, Lehmann S, et al. (2021) Common and contrasting characteristics of the chronic soft-tissue pain conditions fibromyalgia and lipedema. *J Pain Res* 14: 2931-2941.
11. Benz T, Lehmann S, Elfering A, et al. (2021) Comprehensiveness and validity of a multidimensional assessment in patients with chronic low back pain: A prospective cohort study. *BMC Musculoskelet Disord* 22: 291.
12. Wolfe F, Clauw DJ, Fitzcharles M-A, et al. (2016) 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Semin Arthritis Rheum* 46: 319-329.
13. Wolfe F, Smythe HA, Yunus MB, et al. (1990) The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. *Arthritis Rheum* 33: 160-172.
14. Angst F, Verra M, Lehmann S, et al. (2009) Clinical effectiveness of an interdisciplinary pain management programme compared with standard inpatient rehabilitation in chronic pain: A naturalistic, prospective controlled cohort study. *J Rehabil Med* 41: 569-575.
15. Angst F, Brioschi R, Main CJ, et al. (2006) Interdisciplinary rehabilitation in fibromyalgia and chronic back pain: A prospective outcome study. *J Pain* 7: 807-815.
16. Wong WS, Chen PP, Yap J, et al. (2011) Assessing depression in patients with chronic pain: A comparison of three rating scales. *J Affect Disord* 133: 179-187.
17. Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. *Acta Psychiatr Scand* 67: 361-370.
18. Angst F, Verra ML, Lehmann S, et al. (2008) Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. *BMC Med Res Methodol* 8: 26.

19. Bullinger M, Kirchberger I (1998) Fragebogen zum Gesundheitszustand: SF-36; Handanweisung [The SF-36 questionnaire to assess the health status. A manual]. Göttingen, Germany: Hogrefe, Verlag für Psychologie.
20. Kurth BM, Ellert U (2002) The SF-36 questionnaire and its usefulness in population studies: Results of the German Health Interview and Examination Survey 1998. *Soz Präventivmed* 47: 266-277.
21. Ware JE, Snow KK, Kosinski M, et al. (2004) SF-36 health survey: Manual and interpretation guide. (3rd edn), Boston: Health Institute, New England Medical Center, 693-718.
22. Ware JE, Kosinski M, Dewey JE (2000) How to Score Version 2 of the SF-36 Health Survey. (2nd edn), Lincoln, R.I.: Quality Metric Incorporated.
23. Kerns RD, Turk DC, Rudy TE (1985) The West Haven-Yale multidimensional pain inventory (WHYMPI). *Pain* 23: 345-356.
24. Flor H, Rudy TE, Birbaumer N, et al. (1990) The applicability of the West Haven-Yale multidimensional pain inventory in German-speaking countries. Data on the reliability and validity of the MPI-D. *Schmerz* 4: 82-87.
25. Bjelland I, Dahl AA, Haug TT, et al. (2002) The validity of the hospital anxiety and depression scale. An updated literature review. *J Psychosom Res* 52: 69-77.
26. Herrmann-Lingen C, Buss U, Snaith RP (2011) HADS-D: Hospital Anxiety and Depression Scale - Deutsche Version: ein Fragebogen zur Erfassung von Angst und Depressivität in der somatischen Medizin: Testdokumentation und Handanweisung [A Questionnaire to Assess Anxiety and Depressivity in Somatic Medicine-German version]. 3. Aufl. Bern, Switzerland: H. Huber.
27. Hinz A, Brähler E (2011) Normative values for the hospital anxiety and depression scale (HADS) in the general German population. *J Psychosom Res* 71: 74-78.
28. Angst J, Dobler-Mikola A (1984) The Zurich study. II. The continuum from normal to pathological depressive mood swings. *Eur Arch Psychiatry Neurol Sci* 234: 21-29.
29. Binder J, Dobler-Mikola A, Angst J (1981) An epidemiological study of minor psychiatric disturbances. *Soc Psychiatry* 16: 31-41.
30. Derogatis LR (1992) SCL-90-R: Administration, scoring & procedures manual-II for the R(revised) version and other instruments of the psychopathology rating scale series. Towson, MD: Clinical Psychometric Research.
31. Franke GH (2014) Symptom-Checklist-90-Standard: SCL-90-S. Göttingen, Germany: Hogrefe.
32. Angst F, Benz T (2020) Disabilities of the arm, shoulder and hand questionnaire (DASH). In: Maggino F, herausgeber, Encyclopedia of quality of life and well-being research. Cham: Springer International Publishing, 1-22.
33. Angst F, Aeschlimann A, Angst J (2017) The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 82: 128-136.
34. Borenstein M (2009) Effect sizes for continuous data. In: Cooper H, Hedges LV, Valentine JC, The handbook of research synthesis and meta-analysis. New York, NY, US: Russell Sage Foundation, Seite 222-236.
35. Kazis LE, Anderson JJ, Meenan RF (1989) Effect sizes for interpreting changes in health status. *Med Care* 27: S178-S189.
36. Lachin JM (1981) Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials* 2: 93-113.
37. Hedderich J, Sachs L (2016) Der diagnostische Test [The diagnostic test]. In: Angewandte Statistik: Methodensammlung mit R [Applied statistics]. Berlin, Heidelberg (Germany): Springer Spektrum, 180-190.
38. Brandt WA, Loew T, von Heymann F, et al. (2015) The ICD-10 Symptom Rating (ISR): Validation of the depression scale in a clinical sample. *Psychother Psychosom Med Psychol* 65: 234-237.
39. Duckro PN, Margolis RB, Tait RC (1985) Psychological assessment in chronic pain. *J Clin Psychol* 41: 499-504.
40. Findler M, Cantor J, Haddad L, et al. (2001) The reliability and validity of the SF-36 health survey questionnaire for use with individuals with traumatic brain injury. *Brain Inj* 15: 715-723.
41. Hardt J, Gerbershagen HU, Franke P (2000) The symptom checklist, SCL-90-R: Its use and characteristics in chronic pain patients. *Eur J Pain* 4: 137-148.
42. Lee K, Kim D, Cho Y (2018) Exploratory factor analysis of the beck anxiety inventory and the beck depression inventory-ii in a psychiatric outpatient population. *J Korean Med Sci* 33: e128.
43. Ruta DA, Hurst NP, Kind P, et al. (1998) Measuring health status in British patients with rheumatoid arthritis: Reliability, validity and responsiveness of the short form 36-item health survey (SF-36). *Br J Rheumatol* 37: 425-436.
44. Steer RA, Ball R, Ranieri WF, et al. (1997) Further evidence for the construct validity of the Beck depression Inventory-II with psychiatric outpatients. *Psychol Rep* 80: 443-446.
45. Kalin NH (2020) The critical relationship between anxiety and depression. *Am J Psychiatry* 177: 365-367.

***Corresponding author:** Felix Angst, Rehaklinik Bad Zurzach, Quellenstrasse 34, 5330 Bad Zurzach, Switzerland, Tel: +41-56-269-51-51, E-mail: fangst@vtxmail.ch

Accepted: August 06, 2024

Published online: August 08, 2024

Citation: Angst F, Benz T, Lehmann S, et al. (2024) Measurement of Depression and Anxiety in Chronic Low Back Pain and Fibromyalgia: A Validity Comparison of Four Patient-Reported Outcome Measures. *J Depress Anxiety Disord* 6(1):162-172

DOI: 10.36959/362/489