Measuring the Outcome of Health Resort Programs

MARTIN WEIGL, THOMAS EWERT, JUERGEN KLEINSCHMIDT, and GEROLD STUCKI

ABSTRACT. ***Objective***. To evaluate the metric properties and practicability of valid, internationally available outcome instruments in the special setting of health resort programs.

Methods. A cohort study in a convenience sample of patients with low back pain, upper back pain, conditions of the lower extremities, and conditions of the upper extremities was conducted. Their functioning and health were assessed before and after a health resort program by the disease-specific North American Spine Society (NASS) instruments Lumbar NASS and Cervical NASS; WOMAC Osteoarthritis Index; Disabilities of Arm, Shoulder and Hand Questionnaire; and the general instrument, Medical Outcome Study Short Form-36 (SF-36).

Results. Completeness on the scale level ranged between 1% and 10%. Criterion validity of condition-specific instruments was confirmed by stronger associations of the pain and function scales to the Physical Health component of the SF-36 ($r = -0.59$ to -0.79, $p < 0.001$ for all scales) than to the Mental Health component ($r = -0.11$, NS, to $r = -0.42$, $p < 0.001$). Reliability (Cronbach's alpha coefficient) was higher than 0.8 for all scales of condition-specific instruments and for 6 of 8 SF-36 scales. Floor and ceiling effects ranged between 0% and 7%. The condition-specific instruments demonstrated a good responsiveness with an effect size ranging between 0.28 and 0.55 and with a standardized response mean between 0.32 and 0.94. The responsiveness of most SF-36 scales was similar, but the Physical Function scale showed a lower responsiveness than the condition-specific scales.

Conclusion. The evaluated instruments can be recommended for use in clinical trials that assess the outcome of health resort programs. (J Rheumatol 2006;33:764–70)

Key Indexing Terms:
BALNEOLOGY EXERCISE THERAPY MUSCULOSKELETAL DISEASES
HEALTH RESORTS OUTCOME ASSESSMENT (HEALTH CARE) QUALITY OF LIFE

Spa therapy has a long tradition in many European countries and in Israel as a credible medical treatment[1,2]. Treatment programs at spas combine treatment with local resources such as thermal water or mineral water with physical medicine interventions and traditional medicine to optimize functioning and health. The goals of health resort programs are to overcome impairments, activity limitations, and participation restrictions (a rehabilitative strategy) and to prevent further loss of functioning (a preventive strategy)[3]. Although the condition may not be cured or prevented, spa therapy can improve function and minimize disability.

From the Department of Physical Medicine and Rehabilitation, Ludwig-Maximilians-University, Munich, Germany; and the Institute for Health and Rehabilitation Sciences (former Institute of Medical Balneology and Climatology), Ludwig-Maximilians-University, Munich, Germany.

Supported by the Bayerischer Heilbäderverband (BHV; Bavarian Spa Association); the municipalities and spa administrations of Bad Füssing, Bad Wörishofen, and Bad Kissingen; and the Sebastian Kneipp Institut GmbH.

M. Weigl, MD, MPH; T. Ewert, MA, Department of Physical Medicine and Rehabilitation, Ludwig-Maximilians-University Munich; J. Kleinschmidt, PhD, Institute for Health and Rehabilitation Sciences, Ludwig-Maximilians-University Munich; G. Stucki, MD, MS, Department of Physical Medicine and Rehabilitation, and Institute for Health and Rehabilitation Sciences, Ludwig-Maximilians-University Munich.

Address reprint requests to Prof. G. Stucki, Department of Physical Medicine and Rehabilitation, University Hospital Munich, Marchioninistr. 15, D-81377 Munich, Germany.
E-mail: gerold.stucki@med.uni-muenchen.de

Accepted for publication November 23, 2005.

Some controlled clinical trials have shown the effectiveness of spa therapy in reducing pain and improving physical function and quality of life in patients with low back pain[4], osteoarthritis (OA)[5], rheumatoid arthritis[6,7], and ankylosing spondylitis[8]. Combined spa exercises showed a favorable cost-effectiveness compared with standard treatment alone in patients with ankylosing spondylitis[9]. A systematic review on the efficacy of balneotherapy for OA of the knee concluded that this therapy has short-term benefits for pain relief and function[10]. However, a Cochrane review on balneotherapy and its efficacy in treating rheumatoid arthritis concluded that a firm conclusion on the effectiveness of balneotherapy cannot be drawn, because the number of high-quality studies was small and several studies had methodological flaws[11]. One of the authors' recommendations was that new research should use outcome measures that are relevant to patients and are adequate and responsive to the study treatment.

Responsiveness of an instrument refers to the magnitude of change in scores associated with a given change in the health status. For group comparisons, the greater the responsiveness of an outcome measure, the fewer subjects required to detect a significant treatment effect[12]. A variety of statistical methods such as effect size (ES) or standardized response mean (SRM) have been used to assess responsiveness, and no single one is superior[13-15]. Responsiveness should be considered a highly contextualized attribute of an instrument, rather than a static property[16].

We evaluated the metric properties and the practicability of valid, internationally available outcome instruments in the special setting of a health resort program. The specific aims were to determine the criterion validity, floor and ceiling effects, reliability, 2 indices of responsiveness (ES, SRM), and the number of missing values of the following instruments: the condition-specific North American Spine Society (NASS) Lumbar and Cervical NASS; the Western Ontario McMaster University Osteoarthritis Index (WOMAC); the Disabilities of Arm, Shoulder, and Hand Questionnaire (DASH); and the general health related quality of life Medical Outcome Study Short Form-36 Survey (SF-36), in the setting of a 2–4 week outpatient health resort program.

MATERIALS AND METHODS

Study design, patients. Our study on metric properties of outcome measures was designed as an integral part of an ongoing prospective cohort study that evaluates the outcome quality of health resorts programs. A convenience sample of patients at the health resorts Bad Woerishofen and Bad Füssing in Bavaria, Germany, entered the study between July 2002 and September 2004. Patients included had one of the following health conditions: pain in the upper back, pain in the lower back, conditions of the upper extremity, conditions of the lower extremity; were between 18 and 80 years old; stayed at least 14 days at the health resort; had sufficient German language skills to complete the assessment tools; and signed an informed consent.

Data collection. Patients were informed about the study by collaborating spa doctors. In the first week after arrival a set of questionnaires was distributed and collected by the spa administration: a sociodemographic questionnaire, the Self-administered Comorbidity Questionnaire (SCQ)[17]; a condition-specific questionnaire, and the SF-36 general health questionnaire.

At the end of treatment, i.e., after a 2 to 4 week health resort program, the patients again completed the condition-specific questionnaire, the SF-36[18-21], and an adapted client satisfaction questionnaire[22]. The satisfaction questionnaire was not analyzed for this study. All instruments were scannable. Completeness of questionnaires was not checked before scanning.

Study intervention. Patients received usual health resort programs that were individually tailored by spa doctors. Treatment options included bathing and swimming in thermal water, aquatic exercises, mud application, medical massage therapy, group exercise therapy, walking, manual therapy, cryotherapy, electrotherapy, learning relaxation techniques, and diet.

Measures

Self-administered Comorbidity Questionnaire[17]. The SCQ asks: "Do you have any of the following problems?": heart disease, high blood pressure, lung disease, diabetes, ulcer or stomach disease, kidney disease, liver disease, anemia or other blood disease, cancer, depression, arthritis, or back pain. For each problem the questionnaire asks, "Do you receive treatment for it?" and "Does it limit your activities?", as proxies for disease severity and burden of disease. An individual can receive a maximum of 3 points for each medical condition. Higher scores indicate a higher level of comorbidity. For our study, we excluded cohort-defining diseases from the comorbidities. We counted only the number of comorbidities, but did not calculate a total score.

Short Form-36 Health Survey.The SF-36 is a generic instrument that contains 36 items to assess health related quality of life[18-21]. It yields an 8-scale profile of scores (Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health), as well as summary physical and mental scores. All scores are from 0 (worst health status) to 100 (best health status).

The SF-36 has proven useful in comparing general and specific populations, estimating the relative burden of different diseases, differentiating the health benefits resulting from a wide range of different treatments, and screening individual patients[23].

North American Spine Society Cervical and Lumbar Spine Outcome Instrument. The NASS is a condition-specific instrument with specific modules for low back pain and neck pain[24-26]. The original version consists of 11 questions about pain and disability, and 6 (Lumbar NASS) and 8 (Cervical NASS) questions about neurogenic symptoms. Scores on the subscales pain and disability and neurogenic symptoms both range from 1 (excellent functioning/health) to 6 (worst functioning/health). We used an adapted version without the question about impairment of sexual life.

Western Ontario and McMaster Universities Osteoarthritis Index. The WOMAC is a condition-specific (i.e., OA) multidimensional measure of pain (5 items), stiffness (2 items), and physical functional ability (17 items)[27-29]. All 24 WOMAC items are rated in a numerical rating scale ranging from 0 ("no symptoms/no limitation") to 10 ("maximal symptoms/maximal limitation"), which is the format used in the German validation study[30].

Disabilities of Arm, Shoulder, and Hand Questionnaire. The DASH covers symptoms of functioning and health of the entire upper extremity. It consists of a 30-item Disability/Symptom scale, scored from 0 to 100 (0 = excellent functioning/health, 100 = worst functioning/health)[31-33]. We used an adapted version without the question about impairment of sexual life. The optional sports/arts module and the work module were not analyzed for this study.

Analyses

Practicability. With respect to determining practicability, we focused on the completeness of the total set of instruments as well as on the completeness of the single instruments on the scale level.

Criterion validity. The criterion validity of condition-specific instruments was tested using Pearson correlation of pain and function scales of these instruments with the Physical Health component and the Mental Health component of the SF-36.

Reliability. Internal consistency is a type of reliability based on inter-item correlations. The most common value for the internal consistency is Cronbach's alpha.

Floor and ceiling effects. A floor effect is present when patients show the worst possible score of an instrument. In this situation the scale cannot detect a worsening. A ceiling effect is a situation where patients show the best possible score of an instrument and, accordingly, the instrument fails to detect an improvement.

Responsiveness. Two indices for responsiveness were calculated: ES and SRM. ES is the mean change in score divided by the standard deviation of baseline scores. The SRM is the mean change in score divided by standard deviation of individuals' changes in score. Since the 2 different statistical methods for evaluating responsiveness may result in different ranking of an instrument to other instruments or in different recommendations for the use of an instrument, we used both methods[15]. If both methods show good responsiveness, we can be relatively confident that this finding is true. Paired t tests were used to determine the significance of changes in scores before and after treatment ($p < 0.05$).

RESULTS

Patients. The baseline characteristics for 439 patients are shown in Table 1. In all health conditions the Physical Health component of the SF-36 was worth compared to the normal population of the same age. The lowest score was found for patients with conditions of the lower extremity (35.6) and the highest for patients with upper back pain (42.4). The mean score of study patients was 38.5 versus 44.8 in the normal population. In the Mental Health component, patients with upper back pain had the lowest scores and patients with lower extremity conditions had the highest scores. The mean Mental Health component score did not differ significantly from the normal population (study population: 51.7, normal: 53.0)[19].

Table 1. Patient characteristics.

	Total	Lower Back	Upper Back	Lower Extremity	Upper Extremity
No. at baseline	439	178	105	116	40
Age, mean (SD) yrs	66.0 (8.1)	66.0 (8.3)	64.4 (9.6)	67.6 (6.7)	65.1 (6.4)
Female, %	52.4	52.2	60	48.3	45
SF-36 Physical Health, mean (SD)	38.5 (10.0)	38.0 (9.9)	42.4 (9.1)	35.6 (10.2)	38.7 (9.4)
SF-36 Mental Health, mean (SD)	51.7 (10.5)	52.4 (9.5)	49.1 (12.2)	53.2 (9.7)	50.9 (10.8)
Sport > 1 h/week, %	65.0	60.0	69.3	70.6	60.5
High school, %	15.3	14.5	19.4	12.1	17.5
Comorbidities, no.					
0, %	30.1	33.7	27.5	14.7	41.9
1, %	36.9	39.9	37.5	36.2	32.4
≥ 2, %	33.0	26.4	35.0	49.1	25.7

The percentage of patients with 2 or more comorbidities was higher for patients with lower extremity conditions (49.1%) than for other health conditions.

Practicability and completeness. Regarding the sequence of the questionnaires, patients had to first fill out the SF-36 and then the condition-specific questionnaires, but some patients filled out only the SF-36. Among patients with low back pain, 5.1% did not start the Lumbar NASS, among those with upper back pain 4.8% did not start the Cervical NASS, and among patients with conditions of the lower extremity, 2.6% did not start the WOMAC. All patients with upper back pain filled out at least one question of the DASH.

The percentages of missing values on the scale level are presented in Table 2. They ranged between 1.0% for the Pain and Disability scale of the Cervical NASS and 10.0% for the Symptoms and Functioning scale of the DASH.

From the scales of the SF-36, only the Pain scale could be analyzed for all patients at baseline. The Role-Emotional scale was most often not calculable due to missing values (9.6%).

Table 2. Psychometric results.

	n	Missing Values (%)	Ceiling (%)	Floor (%)	Internal Consistency
Lumbar NASS (lower back pain)	169*				
Pain/disability		1.2	0.0	0.0	0.89
Neurogenic symptoms		5.3	8.1	0.0	0.88
Cervical NASS (upper back pain)	100*				
Pain/disability		1.0	0.0	0.0	0.82
Neurogenic symptoms		3.0	4.1	0.0	0.88
WOMAC (lower extremity)	113*				
Pain		4.4	1.9	0.9	0.81
Function		1.8	1.8	0.0	0.96
Stiffness		3.5	4.6	4.6	0.84
Global		8.0	0.0	0.0	0.97
DASH (upper extremity)	40*				
Symptoms and functioning		10.0	0.0	0.0	0.95
SF-36 (all conditions)	439				
Physical Functioning		2.3	4.2	0.5	0.90
Role-Physical		8.7	35.9	31.7	0.86
Pain		0.0	5.5	2.1	0.87
General Health		3.6	0.2	0.0	0.62
Vitality		4.8	0.5	0.2	0.80
Social Functioning		0.7	42.4	0.2	0.78
Role-Emotional		9.6	74.6	14.6	0.91
Mental Health		4.8	1.4	0.0	0.83

* Some patients answered only the SF-36, but not the condition-specific instruments. The analysis for missing values of the condition-specific instruments on the scale level included only patients who answered at least one question.

Internal consistency. The internal consistencies were above 0.70 for all scales of the SF-36, with the exception of the General Health scale, and above 0.80 for all scales of the condition-specific instruments (Table 2).

Floor and ceiling effects. Table 2 shows that in the SF-36 ceiling effects were common in the Role-Emotional scale, the Social Functioning scale and the Role-Physical scale, and floor effects were common in the Role-Physical scale and the Role-Emotional scale. The Role-Physical scale and the Role-Emotional scale showed a bimodal distribution, with peaks at both extremes of the scales. ES were therefore not reported.

Regarding the condition-specific questionnaires, the neurological scales of the Cervical and the Lumbar NASS and the Stiffness scale of the WOMAC showed slight ceiling effects. The WOMAC Stiffness scale also showed a slight floor effect.

Criterion validity. As expected, associations of the pain and function scales of the condition-specific instruments to the Physical Health component of the SF-36 were stronger (r = –0.61 to –0.75, $p < 0.001$ for all scales) than to the Mental Health component (r = –0.09, NS, to –0.48, $p < 0.001$; Table 3).

Responsiveness statistics. The responsiveness statistics for the condition-specific questionnaires and for the SF-36 are presented in Table 4 and Tables 5A and 5B.

For the measurement of physical function the Lumbar NASS Pain and Disability scale, the Cervical NASS Pain and Disability scale, the DASH, and WOMAC Function scales were more responsive than the SF-36 Physical Function scale. For the measurement of pain the Lumbar NASS Pain and Disability scale and the Cervical NASS Pain and Disability scale had lower ES but higher SRM than the SF-36 Pain scale. The WOMAC Pain scale had lower ES and lower SRM than the SF-36 Pain scale. The DASH had a higher ES and a higher SRM compared to the SF-36 Pain scale.

The ES and SRM of the SF-36 Mental Health component were not much smaller than those for the Physical Health component. The highest effects in the Mental Health component were for Mental Health and Vitality, with ES ≥ 0.40 for

Table 3. Correlations (Pearson) of condition-specific instruments with components of the SF-36.

	SF-36 Physical Health		SF-36 Mental Health	
	r	p	r	p
Lumbar NASS (lower back pain), n = 148*				
Pain and disability	–0.75	< 0.001	–0.16	0.050
Cervical NASS (upper back pain), n = 82*				
pain and disability	–0.56	< 0.001	–0.40	< 0.001
WOMAC (lower extremity), n = 89*				
Pain	–0.61	< 0.001	–0.11	0.323
Physical function	–0.67	< 0.001	–0.09	0.399
DASH (upper extremity), n = 40*				
Symptoms and functioning	–0.71	< 0.001	-0.48	0.008

* The number of patients is smaller than in Table 1, because only patients for whom all scales could be calculated were included.

Table 4. Responsiveness of condition-specific questionnaires.

	n	Before Treatment, mean (SD)	After Treatment, mean (SD)	Change, mean (SD)	ES	SRM
Lumbar NASS (lower back pain)	152*					
Pain and disability		2.78 (0.81)	2.46 (0.90)	0.32 (0.49)	0.39	0.65
Neurogenic symptoms		2.58 (1.08)	2.23 (1.06)	0.34 (0.70)	0.32	0.49
Cervical NASS (upper back pain)	92*					
Pain and disability		2.46 (0.76)	2.04 (0.67)	0.42 (0.44)	0.55	0.94
Neurogenic symptoms		2.42 (0.91)	1.99 (0.78)	0.43 (0.62)	0.48	0.70
WOMAC (lower extremity)	93*					
Pain		4.14 (2.27)	3.31 (2.12)	0.83 (1.89)	0.37	0.44
Function		4.16 (2.43)	3.48 (2.28)	0.68 (1.65)	0.28	0.41
Stiffness		4.82 (2.86)	3.97 (2.66)	0.85 (2.65)	0.30	0.32
Global		4.21 (2.31)	3.49 (2.17)	0.72 (1.61)	0.31	0.45
DASH (upper extremity)	31*					
Symptoms and functioning		36.37 (18.02)	29.17 (19.00)	7.20 (14.49)	0.40	0.50

* The number of patients is smaller than in Table 1 because only patients with calculable scores at both assessments were included. All scales had significant effects ($p < 0.05$).

Table 5A. Responsiveness of the SF-36.

	n	T0, mean (SD)	T1, mean (SD)	Change, mean (SD)	ES	SRM
Lower Back	135*					
Physical Function		65.88 (22.93)	69.47 (23.22)	3.59 (14.41)	0.16	0.25
Role-Physical		55.43 (42.10)	65.93 (42.72)	10.49 (40.48)	NA†	0.26
Pain		47.53 (21.61)	57.38 (21.73)	9.85 (17.61)	0.46	0.56
General Health		55.98 (15.76)	59.81 (16.22)	3.83 (13.13)	0.24	0.29
Vitality		55.01 (16.89)	62.11 (17.01)	7.10 (14.15)	0.42	0.50
Social functioning		82.59 (19.20)	88.06 (17.40)	5.46 (18.43)	0.28	0.30
Role-Emotional		86.17 (32.42)	88.39 (29.46)	2.22 (32.12)	NA†	0.07
Mental Health		71.97 (16.88)	79.17 (14.96)	7.20 (14.51)	0.43	0.50
Upper Back	80*					
Physical Function		75.39 (16.04)	80.60 (17.38)	5.21 (11.59)	0.32	0.45
Role-Physical		59.17 (40.54)	81.35 (32.13)	22.19 (39.69)	NA†	0.56
Pain		55.39 (22.14)	69.54 (21.69)	14.15 (17.13)	0.64	0.83
General Health		58.79 (18.92)	63.68 (17.82)	4.89 (14.43)	0.26	0.34
Vitality		55.92 (17.52)	64.77 (16.79)	8.86 (13.03)	0.51	0.68
Social functioning		83.44 (20.93)	89.69 (14.45)	6.25 (16.28)	0.30	0.38
Role-Emotional		72.08 (39.49)	89.17 (27.95)	17.08 (39.33)	NA†	0.43
Mental Health		68.48 (21.41)	77.55 (16.13)	9.08 (13.85)	0.42	0.66
Lower Extremity	85*					
Physical Function		62.16 (22.63)	64.93 (23.21)	2.78 (15.23)	0.12	0.18
Role-Physical		50.49 (42.94)	69.02 (40.39)	18.53 (39.31)	NA†	0.47
Pain		44.56 (20.36)	56.91 (19.21)	12.34 (18.69)	0.61	0.66
General Health		53.36 (14.96)	58.03 (14.17)	4.68 (11.83)	0.31	0.40
Vitality		54.55 (17.38)	61.51 (16.56)	6.96 (14.20)	0.40	0.49
Social functioning		83.09 (19.64)	90.59 (12.78)	7.50 (15.69)	0.38	0.48
Role-Emotional		83.53 (35.50)	91.37 (23.09)	7.85 (36.24)	NA†	0.22
Mental Health		71.79 (15.48)	81.45 (12.26)	9.66 (11.79)	0.62	0.82
Upper Extremity	27*					
Physical Function		68.39 (23.68)	73.37 (22.05)	4.98 (14.49)	0.21	0.34
Role-Physical		53.70 (45.84)	67.59 (42.07)	13.89 (34.20)	NA†	0.41
Pain		41.22 (17.18)	52.15 (19.69)	10.93 (17.29)	0.64	0.63
General Health		58.26 (17.83)	63.51 (16.24)	5.26 (11.31)	0.29	0.46
Vitality		57.84 (20.63)	61.60 (20.48)	3.76 (13.68)	0.18	0.28
Social functioning		84.72 (20.61)	90.28 (14.01)	5.56 (18.13)	0.27	0.31
Role-Emotional		70.37 (42.70)	88.89 (29.24)	18.52 (33.76)	NA†	0.55
Mental Health		71.85 (18.14)	79.48 (16.53)	7.63 (12.72)	0.42	0.60

* The number of patients is smaller than in Table 1 because only patients with calculable scores at both assessments were included. All scales had significant effects ($p < 0.05$), except those in italics. † ES were not calculated for the Role-Physical and Role-Emotional of the SF-36 because these scales showed strong floor effects.

each health condition, except for Vitality, with an ES = 0.27 for patients with conditions of the upper extremity.

All condition-specific scales showed significant improvements (Table 4). All SF-36 scales showed significant improvement when all patients were combined (Table 5B). In the subgroups of patients the number of scales that showed significant benefits ranged between 5 (upper extremity) and 8 (upper back pain; Table 5A).

DISCUSSION

This study demonstrates that the German versions of the widely used condition-specific instruments Lumbar NASS, Cervical NASS, WOMAC, and DASH and the general health related quality of life instrument SF-36 are valid, reliable, and responsive measures in the specific setting of health resort programs. The significant effects suggest a pre-post treatment benefit for the patients in physical function, pain, and health related quality of life.

The context of health resorts programs differs from settings for which the metric properties of the instruments have been tested before. Obviously, the setting of a health resort program is considerably different from drug therapy or surgical therapy, but it is also different from inpatient rehabilitation. One difference is that patients who receive inpatient rehabilitation in Germany get reimbursed by national health insurance plans, except for a co-payment of 10 Euro per day for accommodation, whereas patients in outpatient health resort programs have to pay the majority of the total costs out of pocket. They get reimbursed only for doctor visits and interven-

Table 5B. Responsiveness of the SF-36 (all health conditions combined, n = 327*).

	T0, mean (SD)	T1, mean (SD)	Change, mean (SD)	ES	SRM
Physical Function	67.45 (21.89)	71.34 (22.49)	3.89 (13.97)	0.18	0.28
Role-Physical	54.92 (42.18)	70.64 (40.01)	15.72 (39.63)	NA†	0.40
Pain	48.16 (21.48)	59.80 (21.61)	11.64 (17.76)	0.54	0.66
General Health	56.17 (16.61)	60.60 (16.21)	4.43 (12.96)	0.27	0.34
Vitality	55.35 (17.44)	62.56 (17.12)	7.22 (13.86)	0.41	0.52
Social functioning	83.10 (19.78)	89.30 (15.30)	6.19 (17.15)	0.31	0.36
Role-Emotional	80.73 (36.36)	89.40 (27.44)	8.67 (35.65)	NA†	0.24
Mental Health	71.06 (17.84)	79.39 (14.75)	8.34 (13.53)	0.47	0.62

* The number of patients is smaller than in Table 1 because only patients with calculable scores at both assessments were included. All scales had significant effects ($p < 0.05$). † ES were not calculated for the Role-Physical and Role-Emotional of the SF-36 because these scales showed strong floor effects.

tions, but they have to cover costs for accommodation (except a small additional allowance of 13 Euro per day), and the accommodation is the biggest expense factor.

Considering both the absence of patient-doctor communication (except for limited information given to patients before the study) and the inability to check the questionnaires before data were entered in the database, the completeness of the questionnaires was acceptable. Moreover, the method of data collection by the spa administration without immediate checking of each questionnaire, and the scanning of collected questionnaires, allowed for collection of a large number of patient data with limited human resources.

The expected strong correlations between pain and function scales of the condition-specific instruments and the SF-36 Physical Health component compared to weak correlations to the Mental Health component showed the criterion validity of the condition-specific instruments. This is in line with previous studies that analyzed correlations between Cervical NASS[25,26], Lumbar NASS[24,25], and the DASH[32] and mental health and physical health subscales of the SF-36.

The high internal consistency of all scales of the condition-specific instruments and the low floor and ceiling effects are consistent with previous studies[24-27,30-32]. The high internal consistency for the SF-36 scales, except for a relatively low consistency for the general health scale, is also in line with previous results in different groups of patients[21,23].

The responsiveness of the condition-specific scales was moderate. Compared to previous studies that evaluated inpatient rehabilitation, the ES were slightly smaller for the WOMAC scales[33] and about the same for the Cervical NASS[26]. The results of the SF-36 showed that not only the Physical Health scales, but also the Mental Health scales were responsive and showed significant effects. The improvement of health resort patients in the Mental Health scales may be explained by the holistic concept of health resort programs. The results confirm the concept of using a condition-specific instrument and an additional generic instrument in the outcome assessment of health resort programs.

The generalizibility of this study is limited by the fact that we studied a convenience sample of patients. The responsiveness of the evaluated instruments could be somewhat different in the general population of health resort patients. However, the mean age of 66.0 years and the slight majority of women (52.4%) is similar to the mean age at Bad Woerishofen in 1992-93 (64.4 yrs) and the typical distribution of women to men[34].

The evaluated instruments can be recommended for clinical trials that assess the outcome of health resort programs. Their use in clinical trials with longterm followup could answer the question of whether health resort programs are suitable to achieve and maintain improvement in functioning in patients with chronic rheumatologic health conditions.

ACKNOWLEDGMENT

We thank our partners from the collaborating municipalities, spa administrations, and organizations for the local implementation of the study: F. Gnan (President, Bavarian Spa Association), Alois Brundobler (Mayor, Bad Füssing), Klaus Holetschek (Mayor, Bad Wörishofen), K.H. Laudenbach (Mayor, Bad Kissingen), R. Weinberger (Director, spa administration, Bad Füssing), A. von Hohenegg (Director, spa administration, Bad Wörishofen), A. Kratz (Director, spa administration, Bad Kissingen), D. Jarosch (Business Manager, Sebastian-Kneipp-Instituts-GmbH); and all participating spa doctors.

REFERENCES

1. Van Tubergen A, van der Linden S. A brief history of spa therapy. Ann Rheum Dis 2002;61:273-5.
2. Bender T, Karagulle Z, Balint GP, Gutenbrunner C, Balint PV, Sukenik S. Hydrotherapy, balneotherapy, and spa treatment in pain management. Rheumatol Int 2005;25:220-4.
3. Stucki G, Kroeling P. Principles of rehabilitation. In: Hochberg MC, Silman AS, Smolen JS, Weinblatt ME, Weisman MH, editors. Rheumatology. 3rd ed. Edinburgh: Mosby; 2003:517-30.
4. Konrad K, Tatrai T, Hunka A, Vereckei E, Korondi I. Controlled trial of balneotherapy in treatment of low back pain. Ann Rheum Dis 1992;51:820-2.
5. Kovacs I, Bender T. The therapeutic effects of Cserkeszolo thermal water in osteoarthritis of the knee: a double blind, controlled, follow-up study. Rheumatol Int 2002;21:218-21.
6. Hall J, Skevington SM, Maddison PJ, Chapman K. A randomized and controlled trial of hydrotherapy in rheumatoid arthritis. Arthritis Care Res 1996;9:206-15.
7. Franke A, Reiner L, Pratzel HG, Franke T, Resch KL. Long term

efficacy of radon spa therapy in rheumatoid arthritis; a randomized sham-controlled study and follow-up. Rheumatology Oxford 2000;39:894-902.
8. Van Tubergen A, Landewe R, van der Heijde D, et al. Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial. Arthritis Rheum 2001;45:430-8.
9. Van Tubergen A, Boonen A, Landewe R, et al. Cost effectiveness of combined spa-exercise therapy in ankylosing spondylitis: a randomized controlled trial. Arthritis Rheum 2002;47:459-67.
10. Brosseau L, MacLeay L, Robinson V, et al. Efficacy of balneotherapy for osteoarthritis of the knee: A systematic review. Phys Ther Rev 2002;7:209-22.
11. Verhagen AP, Bierma-Zeinstra SM, Cardoso JR, de Bie RA, Boers M, de Vet HC. Balneotherapy for rheumatoid arthritis. Cochrane Database Syst Rev 2003;CD000518.
12. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol 2004;57:1008-18.
13. Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. Measuring clinically important changes with patient-oriented questionnaires. Med Care 2002;40 Suppl:II45-51.
14. Kazis ES, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27 Suppl:S178-89.
15. Wright JG, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol 1997;50:239-46.
16. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. J Clin Epidemiol 2001;54:1204-17.
17. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003;49:156-63.
18. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptional framework and item selection. Med Care 1992;30:473-83.
19. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Summary Scales: A user's manual. Boston: Health Assessment Lab; 1994:5.
20. Bullinger M. German translation and psychometric testing of the SF-36 Health Survey: preliminary results from the IQOLA Project. International Quality of Life Assessment. Soc Sci Med 1995;41:1359-66.
21. Bullinger M, Kirchberger I. SF-36 Fragebogen zum Gesundheitszustand. Handanweisung. Goettingen: Hogrefe; 1998.
22. Attkisson CC, Zwick R. The Client Satisfaction Questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5:233-7.
23. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey. Manual and interpretation guide. 2nd ed. Boston: National Health Institute, New England Medical Center; 1997.
24. Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH. The North American Spine Society Lumbar Spine Outcome Assessment Instrument: reliability and validity tests. Spine 1996;21:741-9.
25. Pose B, Sangha O, Peters A, Wildner M. Validation of the North American Spine Society Instrument for assessment of health status in patients with chronic backache [German]. Z Orthop Ihre Grenzgeb 1999;137:437-41.
26. Stoll T, Huber E, Bachmann S, et al. Validity and sensitivity to change of the NASS questionnaire for patients with cervical spine disorders. Spine 2004;29:2851-5.
27. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
28. WOMAC Osteoarthritis Index. A user's guide. London, Ontario: University of Western Ontario; 1995.
29. Bellamy N, Kean WF, Buchanan WW, Gerecz-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): post validation reapplication of the WOMAC osteoarthritis index. J Rheumatol 1992;19:153-9.
30. Stucki G, Meier D, Stucki S, et al. Evaluation of a German version of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index. Z Rheumatol 1996;55:40-9.
31. Hudak PL, Amadio PC, Bombardier C, and The Upper Extremity Collaborative Group (UECG). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. Am J Ind Med 1996;29:602-8.
32. Offenbaecher M, Ewert T, Sangha O, Stucki G. Validation of a German version of the disabilities of arm, shoulder, and hand questionnaire (DASH-G). J Rheumatol 2002;29:401-2.
33. Weigl M, Angst F, Stucki G, Lehmann S, Aeschlimann A. Inpatient rehabilitation for hip or knee osteoarthritis: 2 year follow-up study. Ann Rheum Dis 2004;63:360-8.
34. Pelka RB, Leuchtgens H, Albus T. Equivalence of outpatient and inpatient Kneipp cure and their respective efficiency — An observational study and 12 months' follow-up of 363 patients. Phys Med Rehab Kuror 1999;9:6-13.