



ORIGINAL ARTICLE

Outcome measurement in functional somatic syndromes: SF-36 summary scores and some scales were not valid

Andreas Schröder^{a,*}, Eva Oernboel^a, Rasmus W. Licht^b, Michael Sharpe^c, Per Fink^a

^aThe Research Clinic for Functional Disorders and Psychosomatics, Aarhus University Hospital, Noerrebrogade 44, 8000 Aarhus C, Denmark

^bMood Disorders Research Unit, Aarhus University Hospital, Skovagervej 2, 8240 Risskov, Denmark

^cPsychological Medicine Research, School of Molecular and Clinical Medicine, University of Edinburgh, Royal Edinburgh Hospital, EH10 5HF Edinburgh, UK

Accepted 22 March 2011

Abstract

Objective: This study aimed to test the validity of the 36-item Short-Form Health Survey (SF-36) scales and summaries in patients with severe functional somatic syndromes (FSS), such as fibromyalgia and irritable bowel syndrome.

Study Design and Setting: One hundred twenty patients with severe FSS enrolled in a randomized controlled trial filled in the SF-36 questionnaire. We tested for data quality, central scaling assumptions, and agreement with the conceptual model.

Results: Most SF-36 scales were found to be valid; however, three scales (role physical, role emotional, and general health) did not satisfy predefined criteria for construct validity, internal consistency, or targeting to the sample. The correlations between SF-36 scales differed considerably from those reported in the general population. As a consequence, the SF-36 summaries, physical component summary (PCS) and mental component summary (MCS), did not accurately reflect their underlying scales and were negatively correlated ($r = -0.46$, 95% CI [-0.60 to -0.31]).

Conclusion: Although the SF-36 is a valuable instrument to assess perceived health in patients with severe FSS, there are problems with some of the scales and with the scoring procedure of the summaries. The SF-36 PCS may, therefore, not accurately measure the physical health status of these patients. Alternative summary measures are needed. © 2011 Elsevier Inc. All rights reserved.

Keywords: Quality of life; Outcome measurement; SF-36; Physical component summary; Validity; Functional somatic syndromes

1. Introduction

Functional somatic syndromes (FSS), for example, fibromyalgia, chronic fatigue syndrome, and so on, are prevalent worldwide and in all medical settings [1,2]. They constitute a burden for sufferers [3], they are costly for society [4,5], and their management and treatment is a major challenge to health care providers [1]. FSS show substantial overlap [6–14] and share the core feature of patients' suffering from persistent painful and disabling physical symptoms in the absence of identifiable and sufficiently explanatory structural or other specified pathology [1,2,9,15–18]. Objective measures of disease severity are lacking, and subjective measures of symptom severity and of perceived physical health are therefore of particular interest in outcome measurement in FSS.

At present, the large number of divergent outcome measures used in clinical studies investigating treatment strategies in FSS hamper systematic reviews and meta-analyses [1,19–22]. It has been argued that this diversity in outcome measurement forms an essential obstacle for

evidence-based improvement of care in FSS [1], and there is growing consensus about the need for global patient-rated outcome measures for functioning and health-related quality of life in FSS research [17,23].

The most widely used [24] measure to assess perceived physical health is the Medical Outcomes Study 36-item Short-Form Health Survey (SF-36) [25,26] physical component summary (PCS). The SF-36 is a generic instrument used to assess health-related quality of life on eight dimensions of physical and mental health [27]. The PCS measures patients' overall physical health by summarizing all SF-36 scales, giving strongest weight to the scales physical functioning (PF), bodily pain (BP), role physical (RP), and general health (GH) [28] (Fig. 1). According to the conceptual model underlying the SF-36, a low PCS score indicates limitations in self-care and daily activities, suffering from severe pain, and poor general health [28].

The SF-36 PCS offers several advantages for outcome measurement in FSS. First, the SF-36 is extensively validated in the general population [29–31], primary care [30], and several medical conditions [27]. Second, the SF-36 is increasingly used in studies investigating FSS.

* Corresponding author. Tel.: +45-89494310; fax: +45-89494340.

E-mail address: andreas.schroeder@aarhus.rm.dk (A. Schröder).

What is new?**Key findings**

- In patients with severe functional somatic syndromes (FSS), three 36-item Short-Form Health Survey (SF-36) scales (role physical, role emotional, and general health) with considerable contribution to the SF-36 physical component summary (PCS) failed basic scaling assumptions.
- The component summaries PCS and MCS, originally created as independent measures of physical and mental health, were negatively correlated in these patients.
- Principal component analysis revealed a stronger association of the SF-36 vitality scale with the physical component than with the mental component.

What this adds to what was known?

- Our findings question the validity of the SF-36 summaries PCS and MCS in patients with severe FSS. Particularly, the SF-36 PCS may not accurately measure the physical health status of these patients.

What is the implication, what should change now?

- SF-36 summaries and some SF-36 scales should be interpreted with caution in patients with severe FSS. Alternative summary measures and alternative scales to measure role limitations and general health perceptions should be developed and tested.

A MEDLINE search for the years 2004 to 2008 revealed 67 clinical trials in fibromyalgia, irritable bowel syndrome, or chronic fatigue syndrome with SF-36 scales or summaries as outcome measures. This facilitates comparison of treatment effects across different treatments and in different FSS. Third, compared with single SF-36 scales, the PCS has some statistical advantages [28]: Scores are normally distributed without any floor or ceiling effects, the reliability in the general population is high, and because several SF-36 scales are affected in FSS [32–36], the PCS has a high power to detect treatment effects.

On the other hand, the use of the PCS as outcome measure in FSS may be afflicted with shortcomings and pitfalls. No study has so far examined the validity of the SF-36 scales and summaries in patients with FSS, but a number of studies in other clinical conditions indicate that it may be problematic to apply a generic instrument like the SF-36 to a clinical sample because items or scales may not be well targeted [37–39]. This regards especially the scales RP and role emotional (RE) that are often hampered

by very high ceiling effects [37,40–42]. Moreover, a study in patients with chronic fatigue syndrome found indications that these scales may be particularly sensitive to patients' symptom attributions [35] and, therefore, should be interpreted with caution in trials investigating ill-defined medical conditions. Both problems may affect the performance of the PCS, because the RP and RE scales contribute substantially, but with opposite loadings, to the summary score (Fig. 1). Finally, the scoring procedure for the summary measures PCS and MCS was derived using a national sample of US adults, and studies in different patient populations have shown that this scoring procedure can distort results and produce inconsistency between single SF-36 scales and the summaries [43–49].

2. Aim

On this background, we aimed to evaluate the validity of the SF-36 scale and summary scores in patients with severe FSS using data from a randomized clinical trial.

3. Methods*3.1. Participants*

Among 278 patients referred to our Clinic with diagnoses of FSS or multiple medically unexplained physical symptoms, 147 were screened positive to participate in the trial and underwent a thorough clinical assessment. Of them, 127 patients reached criteria for chronic and severe bodily distress syndrome [10,50], a novel empirically based diagnostic category that has recently been shown to encompass several FSS and diagnostic analogues [50]. In its severe type, bodily distress syndrome requires functional somatic symptoms from at least three of four bodily systems and moderate to severe impairment in daily living [50]. A total of 120 patients accepted participation in the trial and were randomly assigned to receive either specialized treatment or enhanced usual care. The trial is registered with ClinicalTrials.gov, number NCT00132197, where a more detailed description can be found.

3.2. Instrument and measures

Patients filled in the Danish version [51] of the SF-36 questionnaire a few days before the clinical assessment, which was carried out by two experienced psychiatrists with at least 4 years of training in internal medicine and surgery and two senior residents in psychiatry.

3.2.1. The measurement model of the SF-36

The SF-36 consists of 35 items, which are used to assess eight health domains (Fig. 1). One additional item measures change of health status over the past year and is not scored. Likert's method of summated ratings [52] is used to calculate scale scores, that is, item responses are summed

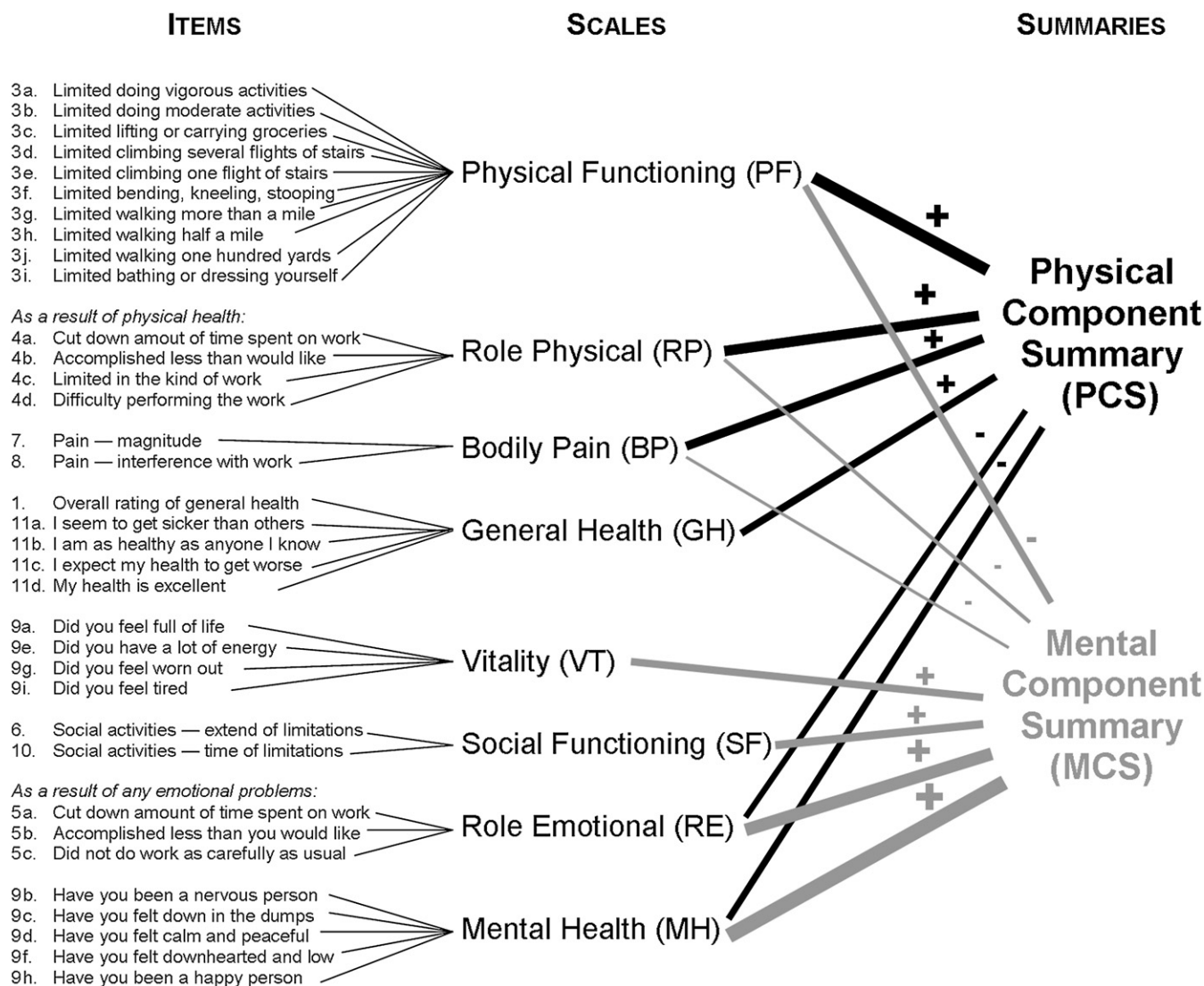


Fig. 1. Measurement model of the SF-36 (adapted from Ware et al. [28]). Lines indicate contribution of items to scales and of scales to summary measures. The thickness of the lines connecting scales and summaries indicates how much a single scale contributes to the summary measure, that is, the magnitude of the factor score coefficients by which z scores of scales are multiplied (range -0.23 to 0.49). Signs indicate positive or negative factor score coefficients. Negligible factor score coefficients (-0.02 to 0.03) are not shown. *Abbreviation:* SF-36, 36-item Short-Form Health Survey.

without weighting or standardization. Sum scores of the eight scales are then transformed linearly to values from 0 (worst health) to 100 (best health). Although the identical range of all scales may assume identical metrics, this is not the case (i.e., the same score does not have the same meaning across scales). To overcome this, the developers propose norm-based scoring, which means that the summed scores are presented as z scores (mean 0, standard deviation [SD] 1) or T scores (mean 50, SD 10), and hence individual scores are reported in SD units [28].

The summary measures physical and mental component summary (PCS and MCS) are derived from z scores by using means and SDs of the corresponding scales of the US population. z scores are then multiplied with their respective factor score coefficients and summed (Fig. 1). Finally, these aggregated scores are transformed into T scores [28,53].

We calculated scale and summary scores according to the Danish manual [51], which uses means and SDs from the general US population and identical weights as the original version. PCS and MCS are therefore comparable to those in US studies, but it is noteworthy that the Danish population shows slightly better MCS (54.0) than the US population [54].

3.2.2. The section on physical and dissociative symptoms of the SCAN

The Schedules for Clinical Assessment in Neuropsychiatry (SCAN) is a semistructured instrument endorsed by the World Health Organization to assess physical symptoms and emotional complaints [55]. To explore patients' physical symptoms, we used a revised version of Section 2, covering a total of 88 different physical complaints. All

available data (clinical interview, routine physical examination, blood counts, and complete case records from general practitioners, ambulatory care, and hospital wards) were used to designate physical symptoms as either “functional” (i.e., not attributable to a conventionally defined medical disease) or attributable to a medical disease or side effects of medication. Items where a clear decision could not be made were excluded from the analysis. Diagnoses of FSS were made according to previously reported diagnostic algorithms [50].

3.3. Analysis

3.3.1. Analysis—step 1: Are SF-36 scales valid outcome measures in FSS?

We tested the SF-36 scales for several scaling assumptions that must be satisfied according to Likert’s method of summated ratings [52] and the developers of the SF-36 [27]. Furthermore, we evaluated data quality and the extent to which the relationship between items and scales was consistent with the conceptual model. All analyses were carried out in Stata version 9.0 [56].

To be regarded as valid outcome measures in patients with FSS, SF-36 scales must satisfy the following criteria [57–59]:

1 High data quality

Missing percentages for items and scales were calculated to estimate data quality. It was regarded excellent (i.e., items are acceptable to patients and easy to answer), if missing values were below 5% for items and below 3% for scales.

2 Parallelism of items

Response distribution, means, and SDs were estimated for each item. We regarded items of a scale as being parallel if they showed similar estimates for means and SDs.

3 Construct validity

Corrected item-own scale correlations (i.e., the correlation between each item and the total score computed from the remaining items in that scale) were calculated to test for convergent and discriminant validity of each item. We scored items as definite scaling success when item-own correlations exceeded item-other correlations by at least two standard errors (SEs) of a correlation coefficient ($SE = 2 * 1/\sqrt{N}$), and item-own correlations at the same time exceeded 0.40 [60]. Because our sample size was small, we accepted one item-other correlation to a closely related scale (according to the conceptual model [28]) within the range of two SEs. Otherwise, items were scored as probable scaling success (item-own correlations were not exceeding item-other correlations by two SEs in two or more cases) or probable scaling failure (item-other correlations were exceeding item-own correlations in at least one case).

4 Internal consistency

Cronbach’s alpha (i.e., the weighted average correlation among all items in a scale) was calculated to estimate internal consistency. We regarded this criterion satisfied if Cronbach’s alpha exceeded 0.70 [57,58].

5 Targeting to the sample

Floor and ceiling effects were evaluated from the distribution of scores. We regarded scales well targeted to patients with severe FSS if floor and ceiling effects were negligible (<15%) [58].

3.3.2. Analysis—step 2: Is the SF-36 PCS a valid measure of patients’ perceived physical health in FSS?

To be regarded as a valid outcome measure in patients with FSS, the SF-36 PCS should satisfy the following criteria [58]:

1 Correlation matrix should be concordant with the conceptual model

Pearson correlations of scales and summaries were calculated and compared with the correlations reported by Ware and Kosinski [28]. According to the model, the PCS correlates strongly with the PF, RP, and BP scales and weakly with the RE and MH scales. Furthermore, the PCS is expected to be uncorrelated with the MCS [28].

2 Extreme PCS scores should correctly summarize scale scores

Extremely low PCS values (i.e., values below the profile range, where all scale scores are equal to 0 [61]) indicate extremely low perceived physical health and should therefore be related to lower values of PF, RP, and BP and not higher values of RE and MH. We tested for this criterion by comparing all patients below the profile range of the SF-36 PCS (<20.1 [61]) with patients in the range from 20.1 through 30.0 by means of the *t*-test.

3 Results from principal component analysis (PCA) should be concordant with the conceptual model

A PCA of the SF-36 scales with varimax rotation was carried out as described by the developers of the SF-36 [28]. Eigenvalues and the scree plot were examined to estimate the optimum number of components to rotate. Factor loadings on the two orthogonally rotated components were compared with those reported in the manual [28].

4. Results

4.1. Sample characteristics

Table 1 summarizes the characteristics of the study population. Patients were predominantly female (79%), mean age of 36 years. A considerable proportion (22%)

Table 1. Sample characteristics

Participants	
Number	120
Female, <i>n</i> (%)	95 (79)
Age (yr)	35.8 (6.4)
Illness characteristics	
Impairment, <i>n</i> (%)	
Moderate	32 (27)
Severe	88 (73)
Duration of functional disorder (yr)	7.0 (3–13)
Current psychiatric comorbidity ^a , <i>n</i> (%)	40 (33)
Functional somatic symptoms (number) ^b	32.5 (8.9)
Sociodemographics, <i>n</i> (%)	
Education	
Basic school (7th–10th)	67 (56)
Further education	53 (44)
Work status	
Employed or student	47 (39)
Unemployed	47 (39)
Disability pension (including flexible jobs) ^c	26 (22)

Data are *n* (%), mean (standard deviation [SD]), or median (interquartile range [IQR]).

^a Current major depressive episode, dysthymia, or anxiety disorder (*Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* [DSM-IV diagnoses]).

^b Functional somatic (i.e., medically unexplained) symptoms during the past 2 years according to the Schedules for Clinical Assessment in Neuropsychiatry (SCAN)-interview and review of clinical records.

^c Special jobs for employees with permanently markedly reduced working ability that includes compensation to the employer.

was in receipt of disability pensions, and another 39% were unemployed and had been so in most cases for several years. As expected, most patients had multiple FSS diagnoses. The illnesses were long lasting (median, 7 years), and one-third suffered from current comorbid anxiety or depression. According to the SCAN diagnostic instrument, patients suffered from a mean of 33 functional somatic symptoms (range, 13–58).

4.2. Data quality

Missing items (including those set to be missing because of logical errors [27]) were low for all items (<5%) (Table 2); hence the large majority of patients were capable of understanding and answering the questions. However, scale scores for the RP and RE scales could only be estimated for 116 patients (96.7%); so these scales did not satisfy our criterion for excellent data quality. In each of the eight scales, the items had similar SDs and roughly similar means, with the exception of the PF scale where some items are ordered by severity. Almost all response options were endorsed, with the exception of four items: no patient expressed very good or excellent health (items 1 and 11d), a maximum of energy (item 9e) or being extremely nervous (item 9b). Given the clinical characteristics of the sample, these findings were not unexpected;

however, they may indicate a reduced sensitivity of the instrument in patients with severe FSS.

4.3. Are SF-36 scales valid outcome measures in FSS?

Fig. 2 gives the distribution, means, and Danish norms of the SF-36 scale scores. Severe FSS had a major impact on both physical and mental health-related quality of life with average decrements from Danish norms of one to two SDs. The RE and RP scales showed pronounced floor and ceiling effects and were not normally distributed. Skewness was considerable for the RP scale only.

Correlations of items with scales and scales with summaries are reported for the 116 patients with complete data on all scale and summary scores:

The vast majority of items showed definite scaling success (Table 3). However, scaling success for the items 4a, 4c, 11a, and 11c was uncertain; and two items (1 and 3a) showed probable scaling failure: Item 1 correlated with the PF, RP, and BP scales more than with the GH scale but showed no correlation with the RE and MH scales, indicating that patients did not recognize their emotional problems as a part of their overall health—although many of them experienced emotional problems. Item 3a was more strongly correlated with the BP scale than with the PF scale, indicating that patients were limited in vigorous activities mostly because of their pain.

At least half of the items in the RP and GH scales could not be rated as definite scaling successes. In addition, the internal consistency criterion for these two scales was not satisfied (Cronbach's alpha below 0.70, Table 4).

4.4. Is the SF-36 PCS a valid measure of patients' perceived physical health in FSS?

Table 4 gives the correlation matrix of scales and summaries. Pearson correlations among SF-36 scales ranged from –0.12 to 0.60 with 14 of 28 (50%) correlation coefficients not exceeding 0.30, which is in contrast with the conceptual model and results from the general population (Tables 3.9 and 3.10 in [28]). In addition, the purest mental health scales (RE and MH) were not at all correlated with the physical health scales (PF, RP, and BP) and, as a consequence of their negative weighting in the scoring procedure for the summaries, they showed substantial negative correlation with the PCS. Conversely, the MCS was only weakly correlated with the physical health scales. The GH scale showed substantial correlations solely with vitality (VT) and social functioning (SF), whereas correlations with the physical health scales were weak. Finally, PCS and MCS were negatively correlated ($r = -0.46$, 95% CI [–0.60 to –0.31]), indicating a negative relationship of physical and mental health in FSS. This counterintuitive negative correlation was even stronger when we restricted the analysis to patients with comorbid anxiety and depression ($r = -0.68$, 95% CI [–0.82 to –0.45]) but remained

Table 2. Descriptive statistics for SF-36 items ($N = 120$)

Scale	Item	Response option frequency distribution (%) ^a							Item score ^b		
		Missing	1	2	3	4	5	6	Mean	SD	Skew
PF	3a	0	79.2	12.5	8.3	— ^c	—	—	1.29	0.61	1.93
	3b	0	36.7	40.0	23.3	—	—	—	1.87	0.77	0.23
	3c	0	21.7	48.3	30.0	—	—	—	2.08	0.72	-0.12
	3d	0	37.5	32.5	30.0	—	—	—	1.93	0.82	0.14
	3e	1.7	9.2	35.8	53.3	—	—	—	2.45	0.66	-0.79
	3f	0.8	13.3	48.3	37.5	—	—	—	2.24	0.68	-0.33
	3g	1.7	30.8	28.3	39.2	—	—	—	2.08	0.84	-0.16
	3h	0	5.0	23.3	69.2	—	—	—	2.66	0.57	-1.46
	3i	2.5	3.3	18.3	75.8	—	—	—	2.74	0.51	-1.86
	3j	0.8	4.2	25.8	69.2	—	—	—	2.66	0.60	-1.37
RP	4a	3.3	65.0	31.7	—	—	—	—	1.33	0.47	0.73
	4b	3.3	85.0	11.7	—	—	—	—	1.12	0.33	2.33
	4c	2.5	76.7	20.8	—	—	—	—	1.21	0.41	1.40
	4d	3.3	84.2	12.5	—	—	—	—	1.13	0.34	2.21
BP	7	0.8	2.5	2.5	15.8	25.8	33.3	19.2	2.70	1.23	0.53
	8	0	8.3	11.7	15.8	45.8	18.3	—	2.48	1.23	0.98
GH	1	0	0	0	10.8	53.3	35.8	—	1.79	0.73	0.72
	11a	0	10.0	20.8	24.2	18.3	26.7	—	3.31	1.33	-0.15
	11b	3.3	2.5	9.2	14.2	31.7	39.2	—	2.01	1.08	0.93
	11c	0.8	6.7	14.2	47.5	11.7	19.2	—	3.23	1.12	0.09
	11d	0.8	0	4.2	6.7	25.8	62.5	—	1.52	0.80	1.57
VT	9a	0.8	0.8	1.7	11.7	24.2	40.8	20.0	2.36	1.04	0.69
	9e	0.8	0.0	1.7	4.2	19.2	31.7	42.5	1.90	0.97	0.93
	9g	0.8	10.8	25.0	19.2	25.8	14.2	4.2	3.20	1.37	0.11
	9i	0.8	25.0	33.3	25.0	11.7	2.5	1.7	2.38	1.16	0.74
SF	6	0	16.7	20.0	29.2	24.2	10.0	—	3.09	1.23	0.04
	10	0.8	6.7	15.8	36.7	25.0	15.0	—	3.26	1.11	-0.15
RE	5a	4.2	33.3	62.5	—	—	—	—	1.65	0.48	-0.64
	5b	2.5	51.7	45.8	—	—	—	—	1.47	0.50	0.12
	5c	3.3	34.2	62.5	—	—	—	—	1.65	0.48	-0.61
MH	9b	0.8	0	10.8	9.2	16.7	32.5	30.0	4.62	1.30	-0.73
	9c	0.8	0.8	6.7	9.2	14.2	25.0	43.3	4.87	1.30	-1.00
	9d	0.8	0.8	10.0	16.7	29.2	30.0	12.5	2.84	1.20	0.34
	9f	1.7	2.5	12.5	18.3	25.8	29.2	10.0	3.98	1.28	-0.34
	9h	0.8	0.8	11.7	15.8	32.5	30.8	7.5	2.96	1.15	0.38

Abbreviations: SF-36, 36-item Short-Form Health Survey; PF, physical functioning; BP, bodily pain; RP, role physical; MH, mental health; GH, general health; VT, vitality; SF, social functioning; RE, role emotional.

^a From raw questionnaires.

^b Calculated after item scores reversed and recalibrated where appropriate (higher scores indicate better health).

^c Response option not available.

significant in the subgroup of patients who had no psychiatric comorbidity ($r = -0.28$, 95% CI [-0.47 to -0.06])

Nine patients had extremely low PCS scores, that is, values below the profile range of <20.1 [61]. Compared with scale scores of the 49 patients with a PCS score in the range 20–30, patients with extremely low PCS scores had significantly lower scores on the SF-36 scales, PF and RP, and significantly higher scores on the RE and MH scales (t -test with unequal variances, all $P < 0.05$). Differences for the remaining scales were not significant. Hence, in patients with FSS, extremely low PCS scores did not indicate extremely poor physical health, but rather a combination of poor PF, good MH, and illness perceptions attributing role limitations to a physical disease and not emotional problems.

PCA with varimax rotation extracted two components with eigenvalues exceeding unity (Fig. 3). The scree plot supported the existence of two higher order factors. However, the two components explained only 58.8% of the variance in all SF-36 scales, which is considerably less than the 80% recommended to generate summaries on the basis of SF-36 scales. In addition, the factor loadings differed markedly from those reported by Ware and Kosinski from the general US population (Figure 3.2 in [28]). Specifically, the VT scale loaded primarily on the physical component, whereas PF and BP showed weak negative loadings on the mental component. For the GH scale, only a small proportion (33%) of the reliable variance was explained by the two components. To summarize, PCA of the correlations of SF-36 scales revealed a structure of health quite different

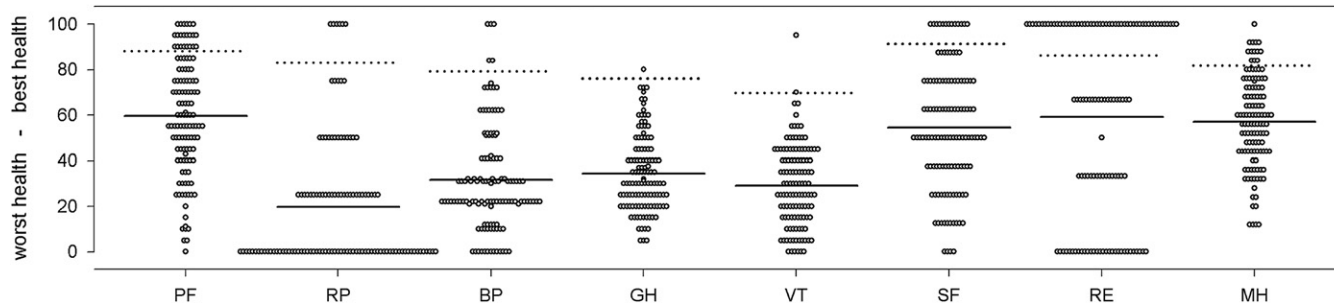


Fig. 2. Dot plot of the SF-36 scales. Solid lines indicate means, dotted lines indicate scale norms (means of the general Danish population [51]). *Abbreviation:* SF-36, 36-item Short-Form Health Survey.

from that underlying the scoring algorithm for the SF-36 PCS and MCS.

5. Discussion

In this cross-sectional study, we aimed to test the validity of the SF-36 scales and summaries in patients with severe FSS. Our main finding was that the structure of self-rated health in patients with severe FSS differed substantially from that of the general population. We found not only reduced performance of single scales but also serious problems with the summary measures PCS and MCS. According to the SF-36 manual [28], a weak or zero correlation of the PCS and MCS is an indicator of the validity of the orthogonal factor rotation used in the construction of the summary measures and hence a basic prerequisite for their use. However, we demonstrated an almost strong negative correlation between these summaries in patients with severe FSS. This questions the validity of the summary measures PCS and MCS in patients with severe FSS.

Although the majority of the SF-36 scales showed good data quality and satisfied basic scaling assumptions in the study sample, this was not the case for the scales RP, GH, and RE. The RP and RE scales showed marked floor and ceiling effects, which also has been reported in other studies [37,40–42]. These effects can be explained, at least partly, by the dichotomous format (yes vs. no) of the items (4a–d, 5a–c, Fig. 1) that measure these concepts, and this format has consequently been changed to a five-point Likert scale in the newer version (version 2) of the SF-36 [62]. However, what was more specific for our sample was the combination of marked floor effects in the RP scale and marked ceiling effects in the RE scale. This may be explained by the fact that the items of the RE and RP scales force patients specifically to attribute their limitations in daily living to either physical or mental health problems [35] because patients are asked two questions at the same time [58]: (1) Have you experienced difficulties performing your work and daily activities? and (2) Are these difficulties a result of your physical health or of any emotional problems? Our results indicate that many patients with severe FSS regard their illness as a pure physical problem, whereas others have difficulties to attribute their limitations

to either physical or mental health problems, which may explain the reduced data quality and low internal consistency of the RP scale. A study in patients with schizophrenia found similar problems with the RE and RP scales and suggested one common scale to measure disability [63], which would be less affected by patients' individual understanding of their illness. This, together with a five-point Likert scale, may also solve the problems of the RE and RP scales in patients with FSS.

The GH scale's low internal consistency in patients with severe FSS was caused by two items that are measuring patients' future health expectations (items 11a and 11c, Fig. 1) rather than patients' current experience of their health. These two items were neither substantially correlated with their own scale nor with any other SF-36 scale. They were also the reason why the GH scale was correlated only with VT and SF and not—as it would be in line with the conceptual model—with PF, RP, and BP. To our knowledge, this has not been reported in other clinical populations. One solution would be to leave the items 11a and 11c out in the calculation of the GH score. However, because of the frequent rating of very poor overall health (item 1), which is typical for patients with severe FSS [32–36], this modified GH scale would probably be hampered by marked floor effects. In its current form, the GH scale should be interpreted with caution in patients with severe FSS.

Our study is a further case example of the limitations of calculating the summary measures PCS and MCS without careful examination whether their use is justified in a specific sample. Studies in a variety of clinical conditions have demonstrated that the scoring algorithms that are derived from PCA in the general US population [28] are often not valid in clinical samples [37–40,63]; and Hann and Reeves [48] have reported disease-specific correlations of the SF-36 scales with underlying dimensions of physical or mental health. Others, however, have found the relationship of the SF-36 dimensions of physical and mental health to be stable across different health states and in the normal population of several countries [54]. To establish whether the negative correlation of PCS and MCS in patients with severe FSS might be an artifact of the sampling of patients by means of the novel unifying diagnosis bodily distress

Table 3. Item-to-scale correlations ($N = 116$)

Scale	Item	SF-36 scales							
		PF	RP	BP	GH	VT	SF	RE	MH
PF	3a	0.55	<i>0.42</i>	0.62	0.11	0.36	0.22	-0.20	-0.15
	3b	0.73	0.45	0.47	0.22	0.33	0.28	-0.03	-0.08
	3c	0.70	0.44	0.46	0.24	0.34	0.31	-0.03	-0.09
	3d	0.74	0.37	0.45	0.20	0.36	0.20	-0.07	-0.08
	3e	0.78	0.36	0.27	0.07	0.29	0.23	-0.11	-0.05
	3f	0.50	0.25	<i>0.33</i>	0.10	0.22	0.25	0.02	0.06
	3g	0.76	0.35	0.32	0.09	0.25	0.22	-0.09	-0.03
	3h	0.70	0.24	0.13	0.08	0.14	0.21	-0.14	-0.08
	3i	0.67	0.22	0.13	0.10	0.21	0.23	-0.08	-0.04
	3j	0.58	0.18	0.27	0.12	0.19	0.28	-0.23	-0.08
RP	4a	<i>0.35</i>	0.45	<i>0.28</i>	0.17	<i>0.35</i>	<i>0.31</i>	0.19	0.12
	4b	0.32	0.57	0.24	0.19	0.36	0.36	0.24	0.13
	4c	<i>0.36</i>	0.41	<i>0.31</i>	0.18	<i>0.30</i>	0.18	0.00	-0.03
	4d	0.29	0.54	0.29	0.17	0.35	0.28	0.11	-0.02
BP	7	0.41	0.26	0.74	0.15	0.23	0.07	-0.09	0.02
	8	0.48	0.44	0.74	0.29	0.41	0.23	-0.03	0.05
GH	1	0.44	0.45	0.40	0.34	0.45	<i>0.28</i>	0.05	0.07
	11a	-0.04	<i>0.11</i>	0.00	0.27	<i>0.10</i>	<i>0.24</i>	<i>0.21</i>	<i>0.17</i>
	11b	0.21	0.19	0.19	0.63	0.25	0.25	0.11	0.11
	11c	-0.02	0.01	0.05	0.32	<i>0.19</i>	0.11	<i>0.25</i>	<i>0.20</i>
	11d	0.24	0.17	0.32	0.61	0.31	0.23	0.11	0.10
VT	9a	0.30	0.45	0.20	0.25	0.66	0.46	0.25	0.43
	9e	0.30	0.37	0.24	0.23	0.60	<i>0.47</i>	0.15	0.36
	9g	0.24	0.33	0.28	0.28	0.59	0.24	0.29	0.29
	9i	0.31	0.31	0.35	0.35	0.54	<i>0.38</i>	0.15	0.15
SF	6	0.24	0.33	0.11	0.29	0.41	0.76	0.28	0.35
	10	0.37	0.38	0.21	0.33	0.49	0.76	0.24	0.37
RE	5a	0.00	0.22	0.00	0.20	0.31	0.30	0.68	<i>0.58</i>
	5b	-0.19	0.07	-0.12	0.16	0.18	0.14	0.68	0.46
	5c	-0.11	0.18	-0.06	0.24	0.22	0.28	0.65	<i>0.50</i>
MH	9b	-0.26	-0.17	-0.21	0.10	0.02	0.07	<i>0.45</i>	0.58
	9c	-0.06	0.08	0.01	0.17	0.33	0.44	<i>0.55</i>	0.71
	9d	0.06	0.13	0.17	0.20	0.40	0.31	<i>0.43</i>	0.60
	9f	-0.04	0.10	0.05	0.20	0.37	0.35	0.52	0.77
	9h	-0.01	0.16	0.15	0.19	0.43	0.36	0.41	0.69

Abbreviations: SF-36, 36-item Short-Form Health Survey; PF, physical functioning; BP, bodily pain; RP, role physical; MH, mental health; GH, general health; VT, vitality; SF, social functioning; RE, role emotional.

Figures in bold indicate item correlations with own scale corrected for overlap. Italics indicate items (and the relevant correlations), where scaling success is uncertain, that is, item-own correlations are not exceeding item-other correlations + 0.1856 (2 SE) in at least two cases. Bold italics indicate items (and the relevant correlations) with higher correlation with other than with own scale (probable scaling failure). No items show definite scaling failure.

syndrome [50] in our study, we contacted two other research groups working with FSS. They reported small negative correlations of PCS and MCS for patients with multisomatoform pain disorder ($r = -0.19$, 95% CI [-0.33 to -0.06], H. Sattel, personal communication, 2009) and irritable bowel syndrome [34] ($r = -0.20$, 95% CI [-0.32 to -0.08], B. Tomeson, personal communication, 2008). In subgroups of patients with irritable bowel syndrome with and without comorbid mental disorders, respectively [64], the negative correlations were even higher, ranging from -0.25 to -0.28 (B. Tomeson, personal communication). Hence, the finding of a negative correlation of PCS and MCS in this study is supported by studies in a variety of FSS.

We are aware of three mechanisms that may explain our main finding. First, the problems with single items and scales that we have demonstrated in this study may contribute to the lack of validity of the PCS in patients with severe FSS. This regards especially the scales RE, RP, and GH, which are highly influenced by patients' causal attributions and expectations about their health, because patients' own explanations of their symptoms play a key role in their interpretation of the underlying items [35]. This may not only influence the reported level of disability for single scales but also have pronounced effects on the performance of the summaries. As one easily can calculate [61], two patients with an almost identical SF-36 profile, but greatly divergent causal attributions of their illness (patient 1: RP

Table 4. Correlation matrix for SF-36 subscales and component summaries

SF-36	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
PF	(0.91) ^a	—	—	—	—	—	—	—	—	—
RP	0.46	(0.69)	—	—	—	—	—	—	—	—
BP	0.48	0.38	(0.85)	—	—	—	—	—	—	—
GH	0.19	0.24	0.24	(0.66)	—	—	—	—	—	—
VT	0.36	0.46	0.35	0.36	(0.78)	—	—	—	—	—
SF	0.32	0.37	0.17	0.33	0.48	(0.86)	—	—	—	—
RE	<i>-0.12</i>	0.18	<i>-0.07</i>	0.23	0.28	0.28	(0.82)	—	—	—
MH	-0.08	0.07	0.04	0.21	0.38	0.38	0.60	(0.86)	—	—
PCS	0.82	0.61	0.69	0.33	0.34	0.21	<i>-0.35</i>	<i>-0.33</i>	—	—
MCS	-0.23	0.08	<i>-0.12</i>	0.25	0.43	0.52	0.83	0.87	<i>-0.46</i>	—

Abbreviations: SF-36, 36-item Short-Form Health Survey; PF, physical functioning; BP, bodily pain; RP, role physical; MH, mental health; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; PCS, Physical component summary; MCS, Mental component summary.

Note: Weak or zero correlations (between -0.3 and 0.3) are in bold. Substantial negative correlations (exceeding -0.3) are in italics. All correlations $>|0.3|$ are highly significant ($P < 0.001$).

^a Cronbach's alpha in parentheses. Alpha coefficients below 0.70 are in italics.

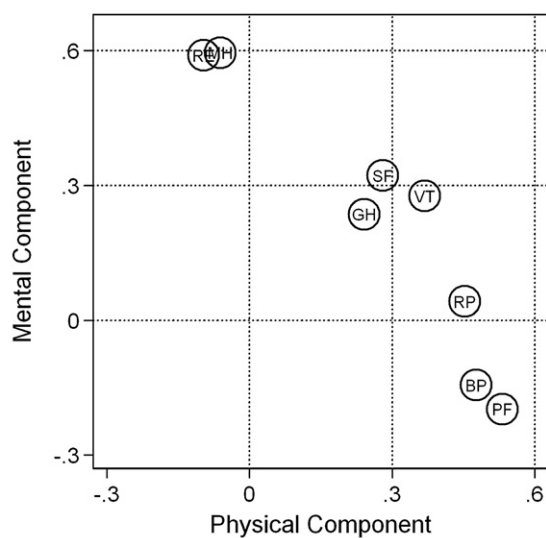
score 0, RE score 100; patient 2: RP score 100, RE score 0), would show a difference of 16 points in their PCS score.

Second, the negative correlation of PCS and MCS may be explained simply by very low scale values in our sample, which exacerbate the weaknesses of factor score coefficients that are based on orthogonal factor rotation. Our study revealed that low PF in combination with good MH resulted in extremely low PCS scores. This finding may not be specific for patients with FSS. It may, instead, be a general result of the scoring procedure. As Taft et al. [61] have convincingly demonstrated, the SF-36 summaries are negatively correlated (and hence not valid) at their extremes as a result of the orthogonal scoring procedure [43]. Consequently, we would expect similar problems with extremely low summary scores in other clinical conditions. However, when we excluded extremely low PCS scores

from the analysis, the negative correlation of PCS and MCS remained substantial ($r = -0.40$, 95% [CI -0.55 to -0.23]).

Third, some researchers are convinced that the construction of summaries based on orthogonal factor solution may distort results especially in situations where patients experience complex health problems that result in both physical disability and emotional distress [47–49,65,66], as is the case in severe FSS [1]. In a large population-based study, Hann and Reeves [48] compared summary measures based on orthogonal and oblique factor rotation. They found that score differences were particularly evident among patients with mental illness and among patients with comorbid mental and physical diseases, often affecting 40% or more of the sample. This is in accordance with our own finding that the negative correlation of PCS and MCS was especially manifested in the subgroup of patients with comorbid psychiatric disorders and further supported by two studies in primary care: a trial assessing treatment effects in depressed patients reported moderate negative correlation ($r = -0.30$, 95% CI $[-0.40$ to $-0.19]$) of PCS and MCS [67]. Our own reanalysis of data of a large trial in primary care, the functional illness in primary care study [68,69], revealed small to moderate negative correlation of PCS and MCS in subgroups of patients with a *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV) diagnosis of anxiety or depression ($r = -0.28$, 95% CI $[-0.43$ to $-0.11]$), in patients with any DSM-IV somatoform disorder ($r = -0.23$, 95% CI $[-0.36$ to $-0.09]$), and, finally, in patients reaching criteria for severe bodily distress syndrome ($r = -0.30$, 95% CI $[-0.61$ to $-0.02]$). A recent study demonstrated low validity of the SF-36 PCS and MCS in patients with schizophrenia [63]. Overall, there is substantial evidence that orthogonally constructed summary measures do not accurately reflect patients' physical and mental health [45,70] in patients with mental disorders and in patients with mental and physical comorbidity [43,44,46–48,63,66].

To conclude, our main finding of a distorted structure of self-rated health in patients with severe FSS, as compared



Rotation: orthogonal varimax

Fig. 3. PCA with orthogonal rotation of the correlations of the SF-36 scales (the same procedure as used by Ware et al. [28]). *Abbreviations:* PCA, principal component analysis; SF-36, 36-item Short-Form Health Survey.

with the general population, may be explained by a combination of psychometric problems and low-content validity of single scales (RP, RE, and GH), well-known problems with the orthogonal scoring procedure for the SF-36 summaries, and complex interacting physical and mental health problems in this group of patients.

Several strategies have been proposed to aggregate summary measures in situations where the original scoring algorithm for the PCS and MCS inaccurately summarizes scale scores: (1) oblique rotation of two [66,71] or three factors [31] derived from all eight scales; (2) oblique rotation of two factors derived from the five most “pure” scales of physical and mental health (PF, BP, RP, RE, and MH) [48]; (3) simple sum [63] or weighted sum [26] of (a) PF, RP, BP, and GH; and (b) SF, VT, RE, and MH to calculate physical health and MH components, respectively; and (4) new construction of scales based on item response theory [70,72]. Our study supports the existence of two distinct higher order factors of self-rated health in FSS, which can be interpreted as dimensions of physical and mental health. This finding is well established through numerous empirical studies [37,38,63,72,73]. However, none of the reported proposals can immediately be recommended for use in patients with FSS because all would include the scales RP, RE, and GH. We found these scales to be hampered not only by psychometric problems but also by low content validity.

Alternative aggregate measures to replace the SF-36 PCS in patients with severe FSS should be based on valid SF-36 scales that measure core aspects of perceived physical health in this patient group. On the basis of our study, and according to others [17], the scales PF and BP fulfill these requirements and are therefore important candidates for such an aggregate score. However, the VT scale may also represent important aspects of perceived physical health in FSS. Fatigue and lack of energy are frequent complaints in patients with severe FSS, also in the absence of psychiatric comorbidity. We found the VT scale to be correlated primarily to patients’ physical health, a finding that is supported by a recent study that demonstrated condition-specific loadings of the VT scale on factors of physical and mental health [48]. Further studies are needed to sort out whether the VT scale should be included in a summary measure of perceived physical health in FSS. In the construction of such an alternative summary score, it is crucial to avoid methods that produce negative factor score coefficients.

Several limitations should be mentioned. First, it is uncertain whether the results of our study can be generalized to patient groups with mild and less chronic FSS because all patients included in this study had chronic conditions. However, the study sample can be regarded as representative for patients with severe and chronic FSS because patients were referred from primary and secondary care, several medical specialties, and urban and rural parts of Denmark. In addition, only few patients refused to participate.

Second, our sample is rather small to question the validity of a well-established outcome measure, and our findings therefore require replication.

6. Conclusion

Based on the present study, only five SF-36 scales (PF, BP, VT, SF, and MH) are valid outcome measures in patients with severe FSS. The remaining scales (RP, RE, and GH) are hampered by psychometric problems and low content validity. The component summaries PCS and MCS may not be valid in patients with severe FSS, and alternative summaries are needed to measure perceived physical health in this patient group. Our results need to be confirmed in larger samples and in patients with less severe and less chronic illness.

References

- [1] Henningsen P, Zipfel S, Herzog W. Management of functional somatic syndromes. *Lancet* 2007;369:946–55.
- [2] Fink P, Rosendal M. Recent developments in the understanding and management of functional somatic symptoms in primary care. *Curr Opin Psychiatry* 2008;21:182–8.
- [3] Whitehead LC. Quest, chaos and restitution: living with chronic fatigue syndrome/myalgic encephalomyelitis. *Soc Sci Med* 2005;62:2236–45.
- [4] Barsky AJ, Orav EJ, Bates DW. Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. *Arch Gen Psychiatry* 2005;62:903–10.
- [5] Hansen MS, Fink P, Sondergaard L, Frydenberg M. Mental illness and health care use: a study among new neurological patients. *Gen Hosp Psychiatry* 2005;27:119–24.
- [6] Nimnuan C, Rabe-Hesketh S, Wessely S, Hotopf M. How many functional somatic syndromes? *J Psychosom Res* 2001;51:549–57.
- [7] Simon G, Gater R, Kisely S, Piccinelli M. Somatic symptoms of distress: an international primary care study. *Psychosom Med* 1996;58:481–8.
- [8] Whitehead WE, Palsson O, Jones KR. Systematic review of the comorbidity of irritable bowel syndrome with other disorders: what are the causes and implications? *Gastroenterology* 2002;122:1140–56.
- [9] Aaron LA, Buchwald D. A review of the evidence for overlap among unexplained clinical conditions. *Ann Intern Med* 2001;134(9 Pt 2):868–81.
- [10] Fink P, Toft T, Hansen MS, Ornbol E, Olesen F. Symptoms and syndromes of bodily distress: an exploratory study of 978 internal medical, neurological, and primary care patients. *Psychosom Med* 2007;69:30–9.
- [11] Deary IJ. A taxonomy of medically unexplained symptoms. *J Psychosom Res* 1999;47:51–9.
- [12] Sullivan PF, Smith W, Buchwald D. Latent class analysis of symptoms associated with chronic fatigue syndrome and fibromyalgia. *Psychol Med* 2002;32:881–8.
- [13] Schur EA, Afari N, Furberg H, Olarte M, Goldberg J, Sullivan PF, et al. Feeling bad in more ways than one: comorbidity patterns of medically unexplained and psychiatric conditions. *J Gen Intern Med* 2007;22:818–21.
- [14] Yunus MB. Central sensitivity syndromes: a new paradigm and group nosology for fibromyalgia and overlapping conditions, and the related issue of disease versus illness. *Semin Arthritis Rheum* 2008;37:339–52.
- [15] De Gucht V, Fischler B. Somatization: a critical review of conceptual and methodological issues. *Psychosomatics* 2002;43:1–9.

- [16] Prins JB, van der Meer JW, Bleijenberg G. Chronic fatigue syndrome. *Lancet* 2006;367:346–55.
- [17] Mease P, Arnold LM, Bennett R, Boonen A, Buskila D, Carville S, et al. Fibromyalgia syndrome. *J Rheumatol* 2007;34:1415–25.
- [18] Wessely S, Nimnuan C, Sharpe M. Functional somatic syndromes: one or many? *Lancet* 1999;354:936–9.
- [19] Whiting P, Bagnall AM, Sowden AJ, Cornell JE, Mulrow CD, Ramirez G. Interventions for the treatment and management of chronic fatigue syndrome: a systematic review. *JAMA* 2001;286:1360–8.
- [20] Zijdenbos IL, de Wit NJ, van der Heijden GJ, Rubin G, Quartero AO. Psychological treatments for the management of irritable bowel syndrome. *Cochrane Database Syst Rev* 2009;(1):CD006442.
- [21] Hauser W, Bernardy K, Uceyler N, Sommer C. Treatment of fibromyalgia syndrome with antidepressants: a meta-analysis. *JAMA* 2009;301:198–209.
- [22] Quartero AO, Meineche-Schmidt V, Muris J, Rubin G, de Wit N. Bulking agents, antispasmodic and antidepressant medication for the treatment of irritable bowel syndrome. *Cochrane Database Syst Rev* 2005;(2):CD003460.
- [23] Evans BW, Clark WK, Moore DJ, Whorwell PJ. Tegaserod for the treatment of irritable bowel syndrome and chronic constipation. *Cochrane Database Syst Rev* 2007;(4):CD003960.
- [24] Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000;17:13–35.
- [25] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- [26] Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ* 1993;2(3):217–27.
- [27] Ware JE Jr, Kosinski M, Gandek B. SF-36 health survey: manual & interpretation guide. Lincoln, RI: QualityMetric Incorporated; 1993, 2000.
- [28] Ware JE Jr, Kosinski M. SF-36 physical and mental health summary scales: a manual for users of version 1. 2nd edition. Lincoln, RI: QualityMetric Inc.; 2001.
- [29] McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247–63.
- [30] Brazier JE, Harper R, Jones NM, O’Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992;305:160–4.
- [31] Keller SD, Ware JE Jr, Bentler PM, Aaronson NK, Alonso J, Apolone G, et al. Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. *International Quality of Life Assessment. J Clin Epidemiol* 1998;51:1179–88.
- [32] Petrak F, Hardt J, Kappis B, Nickel R, Tiber EU. Determinants of health-related quality of life in patients with persistent somatoform pain disorder. *Eur J Pain* 2003;7:463–71.
- [33] Birtane M, Uzunca K, Tastekin N, Tuna H. The evaluation of quality of life in fibromyalgia syndrome: a comparison with rheumatoid arthritis by using SF-36 Health Survey. *Clin Rheumatol* 2007;26:679–84.
- [34] Creed F, Ratcliffe J, Fernandez L, Tomenson B, Palmer S, Rigby C, et al. Health-related quality of life and health care costs in severe, refractory irritable bowel syndrome. *Ann Intern Med* 2001;134(9 Pt 2):860–8.
- [35] Taillefer SS, Kirmayer LJ, Robbins JM, Lasry JC. Psychological correlates of functional status in chronic fatigue syndrome. *J Psychosom Res* 2002;53:1097–106.
- [36] van der Windt DA, Dunn KM, Spies-Dorgelo MN, Mallen CD, Blankenstein AH, Stalman WA. Impact of physical symptoms on perceived health in the community. *J Psychosom Res* 2008;64:265–74.
- [37] Baron R, Elashaal A, Germon T, Hobart JC. Measuring outcomes in cervical spine surgery: think twice before using the SF-36. *Spine* 2006;31:2575–84. Ref Type: Journal (Full).
- [38] Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke. Uses and abuses of the SF-36. *Stroke* 2002;33:1348–56. Ref Type: Journal (Full).
- [39] Hobart J, Freeman J, Lamping D, Fitzpatrick R, Thompson A. The SF-36 in multiple sclerosis: why basic assumptions must be tested. *J Neurol Neurosurg Psychiatry* 2001;71:363–70.
- [40] Jenkinson C, Hobart J, Chandola T, Fitzpatrick R, Peto V, Swash M. Use of the short form health survey (SF-36) in patients with amyotrophic lateral sclerosis: tests of data quality, score reliability, response rate and scaling assumptions. *J Neurol* 2002;249(2):178–83.
- [41] Reulen RC, Zeegers MP, Jenkinson C, Lancashire ER, Winter DL, Jenney ME, et al. The use of the SF-36 questionnaire in adult survivors of childhood cancer: evaluation of data quality, score reliability, and scaling assumptions. *Health Qual Life Outcomes* 2006;4:77.
- [42] Puhan MA, Gaspoz JM, Bridevaux PO, Schindler C, Ackermann-Liebrich U, Rochat T, et al. Comparing a disease-specific and a generic health-related quality of life instrument in subjects with asthma from the general population. *Health Qual Life Outcomes* 2008;6:15.
- [43] Taft C, Karlsson J, Sullivan M. Interpreting SF-36 summary health measures: a response—reply. *Qual Life Res* 2001;10:415–20.
- [44] Nortvedt MW, Riise T, Myhr KM, Nyland HI. Performance of the SF-36, SF-12, and RAND-36 summary scales in a multiple sclerosis population. *Med Care* 2000;38:1022–8.
- [45] Wilson D, Parsons J, Tucker G. The SF-36 summary scales: problems and solutions. *Soz Präventivmed* 2000;45:239–46.
- [46] Blanchard CM, Cote I, Feeny D. Comparing short form and RAND physical and mental health summary scores: results from total hip arthroplasty and high-risk primary-care patients. *Int J Technol Assess Health Care* 2004;20:230–5.
- [47] Schmitz N, Kruse J. The SF-36 summary scores and their relation to mental disorders: physical functioning may affect performance of the summary scores. *J Clin Epidemiol* 2007;60:163–70.
- [48] Hann M, Reeves D. The SF-36 scales are not accurately summarised by independent physical and mental component scores. *Qual Life Res* 2008;17:413–23.
- [49] Simon GE, Revicki DA, Grothaus L, VonKorff M. SF-36 summary scores: are physical and mental health truly distinct? *Med Care* 1998;36:567–72.
- [50] Fink P, Schröder A. One single diagnosis, bodily distress syndrome, succeeded to capture ten diagnostic categories of functional somatic syndromes and somatoform disorders. *J Psychosom Res* 2010;68:415–26.
- [51] Bjørner JB, Damsgaard MT, Watt T, Bech P, Rasmussen NK, Kristensen TS, et al. Danish SF-36 manual - a health status questionnaire. Dansk manual til SF-36-et spørgeskema om helbredsstatus. København, Denmark: LIF Lægemedelindustriforeningen; 2003.
- [52] Likert R. A technique for the measurement of attitudes. *Arch Psychol* 1932;22:55.
- [53] Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care* 1995;33(4 Suppl):AS264–79.
- [54] Ware JE Jr, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, et al. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. *International Quality of Life Assessment. J Clin Epidemiol* 1998;51:1167–70.
- [55] WHO. SCAN. Schedules for clinical assessment in neuropsychiatry, version 2.1. Geneva, Switzerland: World Health Organization, Division of Mental Health; 1998.
- [56] STATA. Software package, version 9.0. StataCorp LP, College Station, TX, USA. 2005.

- [57] Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193–205.
- [58] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- [59] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- [60] Ware JE, Harris WJ, Gandek B. MAP-R for Windows: multitrait/multi-item analysis program-revised user's guide. Boston, MA: Health Assessment Lab; 1997.
- [61] Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? *Qual Life Res* 2001;10:395–404.
- [62] Ware J, Kosinski M, Dewey JE. How to score: version 2 of the SF-36 Health Survey (Standard and Acute Forms). Lincoln, RI: QualityMetric Incorporated; 2005.
- [63] Leese M, Schene A, Koeter M, Meijer K, Bindman J, Mazzi M, et al. SF-36 scales, and simple sums of scales, were reliable quality-of-life summaries for patients with schizophrenia. *J Clin Epidemiol* 2008;61:588–96.
- [64] Creed F, Guthrie E, Ratcliffe J, Fernandes L, Rigby C, Tomenson B, et al. Does psychological treatment help only those patients with severe irritable bowel syndrome who also have a concurrent psychiatric disorder? *Aust N Z J Psychiatry* 2005;39:807–15.
- [65] Cunningham WE, Nakazono TT, Tsai KL, Hays RD. Do differences in methods for constructing SF-36 physical and mental health summary measures change their associations with chronic medical conditions and utilization? *Qual Life Res* 2003;12:1029–35.
- [66] Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. *Health Qual Life Outcomes* 2007;5:54.
- [67] Coulehan JL, Schulberg HC, Block MR, Madonia MJ, Rodriguez E. Treating depressed primary care patients improves their physical, mental, and social functioning. *Arch Intern Med* 1997;157:1113–20.
- [68] Toft T, Fink P, Oernboel E, Christensen K, Frostholm L, Olesen F. Mental disorders in primary care: prevalence and co-morbidity among disorders. Results from the functional illness in primary care (FIP) study. *Psychol Med* 2005;35:1175–84.
- [69] Toft T, Rosendal M, Ornbol E, Olesen F, Frostholm L, Fink P. Training general practitioners in the treatment of functional somatic symptoms: effects on patient health in a cluster-randomised controlled trial (the Functional Illness in Primary Care Study). *Psychother Psychosom* 2010;79:227–37.
- [70] Guthlin C, Walach H. MOS-SF 36-Structural equation modeling to test the construct validity of the second-order factor structure. *Eur J Psychol Assess* 2007;23(1):15–23.
- [71] Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. *Ann Med* 2001;33:350–7.
- [72] Chang CH, Wright BD, Cella D, Hays RD. The SF-36 physical and mental health factors were confirmed in cancer and HIV/AIDS patients. *J Clin Epidemiol* 2007;60:68–72.
- [73] Cella D, Chang CH, Wright BD, Von Roenn JH, Skeel RT. Defining higher order dimensions of self-reported health: further evidence for a two-dimensional structure. *Eval Health Prof* 2005;28:122–41.