Eur Spine J (2012) 21:101–114
DOI 10.1007/s00586-011-1921-4

ORIGINAL ARTICLE

Validity and responsiveness of the Core Outcome Measures Index (COMI) for the neck

C. D. Fankhauser · U. Mutter · E. Aghayev · A. F. Mannion

Received: 21 March 2011 / Revised: 17 May 2011 / Accepted: 9 July 2011 / Published online: 20 August 2011

Abstract
Purpose Patient-orientated outcome questionnaires are essential to evaluate treatment success. To compare different treatments, hospitals, and surgeons, standardised questionnaires are required. The present study examined the validity and responsiveness of the Core Outcome Measurement Index for neck pain (COMI-neck), a short, multidimensional outcome instrument.
Methods Questionnaires were completed by patients with degenerative problems of the cervical spine undergoing cervical disc arthroplasty before ($N = 89$) and 3 months after ($N = 75$) surgery. The questionnaires comprised the EuroQol-Five Dimension (EQ-5D), the North American Spine Society Cervical Spine Outcome Assessment Instrument (NASS-cervical) and the COMI-neck.
Results The COMI and NASS-cervical scores displayed no notable floor or ceiling effects at any time point whereas for the EQ-5D, the highest or lowest values were reached in around 32.5% of patients at follow-up. With one exception (symptom-specific well-being), the individual COMI items and the COMI summary score correlated to the expected extent ($R = 0.4$–0.8) with the scores of the chosen reference questionnaires. The area under the curve (AUC) generated by ROC analysis was significantly higher for the COMI (0.96) than for any other instrument/subscale when self reported treatment outcome was used as the external criterion, dichotomised as "good" (operation helped a lot/helped) versus "poor" (operation helped only a little/didn't help/made things worse). The COMI had a high effect size (standardised response mean; SRM) (2.34) for the good global outcome group and a low SRM for the poor outcome group (0.34). The EQ-5D and the NASS-cervical lacked this ability to differentiate between the two groups, showing less distinct SRMs for good and poor outcome groups.
Conclusions This study provides evidence that the COMI-neck is a valid and responsive questionnaire in the population of patients examined. Further investigations should examine its applicability in other patient groups with less severe neck pain or undergoing other treatment modalities.

Keywords COMI · Outcome · Spine surgery · Total disc arthroplasty

C. D. Fankhauser · U. Mutter · A. F. Mannion (✉)
Spine Center, Schulthess Klinik, Lengghalde 2, 8008 Zürich, Switzerland
e-mail: anne.mannion@kws.ch

E. Aghayev
Institute for Evaluative Research in Medicine, University of Bern, Stauffacherstrasse 78, 3014 Bern, Switzerland

Introduction

Neck pain is a very common symptom with a lifetime prevalence of around 50% [1]. Because so many patients are affected, approximately 1% of total health care expenditure is utilised in its treatment [2]. Many treatments for musculoskeletal disorders are carried out with the aim of improving the patient's quality of life and function. Patients, health insurances and governmental bodies increasingly expect appropriate documentation of the efficacy of medical treatment [3, 4]. As a result of outcomes research, indications can be optimised [5–7], therapy and predictors critically questioned, and success or deterioration measured [8, 9]. Patient safety has been improved by

the analysis of side effects and contraindications, and different surgeons and hospitals have been compared [10, 11]. Outcomes research tries to guide doctors and patients through the variety of treatment options, also in consideration of the costs of treatment.

Various options for monitoring treatment effects in neck pain have been examined. Physiological approaches involving, for example, the assessment of strength or range of motion, have by and large failed to serve as valid outcome parameters, because they do not relate well to the factors of importance to the patient, such as symptoms and function in daily life [12]. Social science approaches have more recently found their way into outcomes research and many patient self-rating questionnaires have been developed. These subjective patient-orientated questionnaires appear to be the most valid outcome measurements we have today, but they are only useful for systematic documentation if they are feasible for use in routine daily practice [13]. Questionnaires should be long enough to include all essential questions but demand as little time as possible for the patient to complete. Short questionnaires also reduce the workload of data management, making them easier to integrate into the existing infrastructure of an institution [14]. If the instruments are standardised and available in different languages, they allow national and worldwide comparisons of baseline status and treatment outcomes for a given disorder [14].

In 1998, Deyo et al. [14] recommended a set of six core questions as a parsimonious and valid instrument for assessing outcome in disorders of the lumbar spine. The questions evaluated the dimensions pain (axial and radiating pain), function, symptom-specific well-being, and disability (social and work). This core set showed excellent psychometric characteristics in patients with back pain undergoing either surgical or conservative management [15, 16] and multilingual versions were adopted for use in the Spine Tango system, the international spine surgery registry of Eurospine, the Spine Society of Europe (SSE) [17].

The set of questions was also adapted for the cervical spine, by enquiring about neck rather than back problems, and it too showed good validity and reliability [18]. However, the latter study included only patients with moderate symptoms undergoing conservative management, did not include a quality of life question, and did not examine the responsiveness of the questionnaire, which is one of the key elements of outcome instruments [19]. The aim of the present study was to further analyse the psychometric characteristics of the COMI-neck questionnaire in a group of patients undergoing surgery of the cervical spine, with a focus on its responsiveness compared with that of other well-established condition-specific and quality of life questionnaires.

Methods

Patients

The study represents a retrospective analysis of prospectively collected data. All patients who had undergone cervical spine disc replacement surgery at our hospital between May 2005 and March 2010 were eligible for inclusion in the study as long as they could understand written German, had fulfilled the indications for disc replacement surgery (aged between 18 and 65, no segmental kyphosis, degeneration in not more than two segments, unsuccessful conservative therapy for at least 3 months, suffering from cervical brachialgia, discogenic neck or shoulder pain or early stage cervical myelopathy) and had reached 3 months follow-up after their operation. Exclusion criteria were traumatic or neoplastic indications for surgery. After their consultation with the surgeon in which the decision to operate was made, patients filled out a booklet of baseline questionnaires containing the EuroQol-5D and the NASS-cervical (see below), in compliance with the SWISSspine registry for disc arthroplasty in Bern [20], part of the government-mandated prospective evaluation of disc arthroplasty outcomes in Switzerland. Three months after surgery, at the time of the clinical follow-up with the surgeon, the booklet of questionnaires was completed again. As part of our hospital's own in-house Spine outcomes registry a questionnaire containing the COMI (see below) was sent to the patient at home, preoperatively together with the information about their forthcoming hospital stay, and they were asked to complete it and hand it in during admission. Three months after surgery, a follow-up COMI was sent to them from the Research Department to complete and return by post.

Since the study was intended to compare the psychometric characteristics of the questionnaires themselves rather than report the outcome of the surgical procedure per se, the short-term follow-up of 3-months was considered unproblematic and in keeping with previous methodological studies [21].

The Core Outcome Measures Index for the neck (COMI-neck)

The COMI-neck is a short, self-administered outcome instrument consisting of just seven questions to evaluate the five dimensions pain, neck-related function, symptom-specific well-being, general quality of life and disability (social and work). Apart from the two disability items, which refer to the last 4 weeks, all items relate to how the patient felt in the last week. The two pain items use a 0–10 graphic rating scale; all other items use a 5-point adjectival scale. The higher the score, the worse the patient's status.

Table 1 COMI items, response options, and scoring

Dimension	COMI item	Response options and scoring
Pain	1. How severe was your neck pain in the last week?	Response options: 10-point graphic rating scale, "no pain" to "worst pain that I can imagine"
	2. How severe was your arm/shoulder pain in the last week?	The higher of the two pain scores ("COMI high pain") (0–10) is used to represent the dimension "pain" in calculating the COMI index score
Function	3. During the past week, how much did your neck problem interfere with your normal work (including both work outside the home and housework)?	Response options: 5-point adjectival scale, "not at all" to "extremely" 1 = 0 points 2 = 2.5 points 3 = 5.0 points 4 = 7.5 points 5 = 10.0 points
Symptom-specific well-being	4. If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?	Response options: 5-point adjectival scale, "very satisfied" to "very dissatisfied" 1 = 0 points 2 = 2.5 points 3 = 5.0 points 4 = 7.5 points 5 = 10.0 points
Quality of life	5. Please reflect on the last week. How would you rate your quality of life?	Response options: 5-point adjectival scale, "very good" to "very poor" 1 = 0 points 2 = 2.5 points 3 = 5.0 points 4 = 7.5 points 5 = 10.0 points
Disability (social and work)	6. During the past 4 weeks, how many days did you cut down on the things you usually do (work, housework, school, recreational activities) because of your neck problem? 7. During the past 4 weeks, how many days did your neck problem keep you from going to work (job, school, housework)?	Response options: 5-point adjectival scale, "none" to "more than 21 days" 1 = 0 points 2 = 2.5 points 3 = 5.0 points 4 = 7.5 points 5 = 10.0 points The average score for these two items is used to represent the dimension "disability" in calculating the COMI index score.
COMI summary score		Average of all dimensions, scored 0–10

The scoring for each dimension [15] is summarised in Table 1. For the summary score the average of the scores for all five dimensions (each transformed 0–10) is taken [15]. According to the categories described by Beaton and Schemitsch [22] the COMI is a condition-specific questionnaire but includes also generic and so-called additional elements.

At the 3-month follow-up, the same questions were presented with an additional question to evaluate the global outcome of treatment [23]. The question enquired: "Overall, how much did the treatment that you received (the operation) help your neck problem? …" and was answered with a 5-point Likert scale (operation helped a lot, helped, helped only little, did not help, made things worse). This global outcome question was dichotomised into "good" (operation helped, helped a lot) and "poor" (operation helped only little, did not help, made things worse) for further analyses. Although "helped only little"

Table 2 Overview of the items in the NASS-cervical questionnaire and the items making up each of the subscales

Question	Item content
C1	Frequency: neck pain
C2	Frequency: arm pain
C3	Frequency: numbness/tingling in arm/hand
C4	Frequency: weakness in arm/hand
C5	Bothersome: neck pain
C6	Bothersome: arm pain frequency
C7	Bothersome: numbness/tingling in arm/hand
C8	Bothersome: weakness in arm/hand
C9	Activity limitation: getting dressed
C10	Activity limitation: lifting
C11	Activity limitation: walking/running
C12	Activity limitation: sitting
C13	Activity limitation: standing
C14	Activity limitation: sleeping
C15	Activity limitation: social activities
C16	Activity limitation: travelling
C17	Activity limitation: sexual activity
C18	Frequency: stiffness in legs when walking
C19	Frequency: shaking in legs when walking
NASS-cervical 1 Pain&disability [28, 29]	Mean(C1+C5+C9+C10+C11+C12+C13+C14 +C15+C16+C17)
NASS-cervical 2 Neurology score 1 [28]	Mean(C2+C3+C4+C6+C7+C8+C18+C19)
NASS-cervical 3 Pain score	Mean(C1+C2+C5+C6)
NASS-cervical 4 Neurology score 2 [29]	Mean(C3+C4+C7+C8+C18+C19)
NASS-cervical 5 Disability score	Mean(C9+C10+C11+C12+C13+C14+C15+C16+C17)

is still a positive outcome, the cut-off point for "good" was placed higher than this, since clinically this is not considered a satisfactory outcome for elective surgery [15, 23].

Assessment of the psychometric properties of the COMI-neck

Questionnaire battery

To evaluate the COMI's construct validity the following reference instruments were used: the EuroQol-Five Dimension (EQ-5D), the EuroQol-visual analogue scale (EQ-VAS), the North American Spine Society Cervical Spine Outcome Assessment Instrument (NASS-cervical) and two 0–10 visual analogue scales for neck pain and for arm pain.

The EQ-5D measures health-related quality of life [24]. It is a standardised, widely used, generic questionnaire and consists of the five items, i.e. mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each item is rated on a 3-point adjectival scale. The EQ-VAS is used to quantify the 'overall health state', with the patient indicating his/her current health status on a 0–100 VAS. In the present study, a horizontal scale was used in preference to the vertical scale used in the original version, for ease of layout. The EQ-5D summary index scores [ranging from −0.594 (worse than death) to 1 (best possible health)] were calculated using the unweighted method of Prieto and Sacristán [25].

The NASS-cervical is a pain and disability questionnaire developed by the North American Spine Society Outcome Assessment Task Force and is a region-specific measure [26, 27]. It is based on questions from the Oswestry Disability Index concerning dressing, lifting, walking, sitting, standing, sleeping, participating in social life, travelling, and sexual activity plus eight additional questions about the frequency and bothersomeness of pain (neck, arm), sensory disturbances and motor disturbances. Several summary scores exist, which differ in relation to the questions chosen to form the average value (see Table 2); however, the "pain&disability" dimension appears to be the most frequently used subscale and the remaining subscales have not been widely researched.

Additional medical history and surgical variables describing the study group were extracted from the Spine Tango Spine Surgery Registry [17].

Statistical analysis

The following "missing" rules were applied in the case of missing data: for COMI and EQ-5D, no missing data were allowed because they consist of only one item per domain. For the NASS-cervical a minimum of 80% items had to have been completed for the two main domains (pain&disability; neurology). We used this completion rate as the acceptable minimum for questionnaires in general (Elfering, personal communication) because there appeared to be no consensus in the literature from previous authors working with the NASS questionnaire [28, 29]. The score was then derived using the average of the values for the items that had been completed, to replace the missing values.

Floor and ceiling effects

Floor and ceiling effects were given by the proportion of individuals obtaining scores equivalent to the worst status and the best status, respectively, for each item and scale investigated. This indicates the proportion for whom, respectively, no meaningful deterioration or improvement in their condition could be detected since they are already at the extreme of the range. Due to the different scoring polarity of the questionnaires, for the COMI and NASS the highest scores represented floor effects (worst status) and the lowest scores, ceiling effects (best status); the converse was true for the EQ-5D and EQ-VAS scores. Floor/ceiling effects >70% are considered to be adverse and <15–20%, ideal [30, 31]. Floor and ceiling effects were determined for all scales, in order to provide some perspective for interpreting the corresponding values for the COMI.

Construct validity

Construct validity addresses the extent to which a questionnaire's scores relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured [32]. The relationships between the COMI items/summary score and other questionnaire items or scores describing similar dimensions were examined using Pearson's correlation coefficients. The following pairs of COMI items and corresponding items/questionnaires were examined:

- the COMI "high pain" score and the higher of the two (arm or neck pain) 0–10 VAS pain scores and the NASS-cervical pain score;
- the COMI-neck (arm) pain and the respective neck (arm) pain VAS;
- the COMI item "neck function" and the NASS-cervical pain&disability and NASS-cervical disability score;
- the COMI item "symptom-specific well being" and the EuroQol-5D and EQ-VAS;
- the COMI item "general quality of life" and the EuroQol-5D and EQ-VAS;
- the COMI "disability" average score and the NASS-cervical pain&disability and NASS-cervical disability score.

The correlations between the COMI summary score and all summary scores of the EQ-5D and the NASS-cervical were also examined [33].

Based on the validation studies for the original COMI and as recommended by Streiner and Norman [33] for measures of the same/similar attributes it was hypothesised that correlation coefficients would range from 0.4 to 0.8 for the relationships between the individual COMI items and their corresponding full-length questionnaires and between the COMI summary index score and NASS-cervical and EQ-5D summary index scores.

Responsiveness

Responsiveness refers to the ability of an instrument to show small but clinically important changes [21]. Beaton et al. [34] emphasise the importance of using different measures of responsiveness. In the present study we used three different approaches to compare questionnaire responsiveness.

Firstly, the effect size (standardised response mean; SRM) for the different questionnaires was calculated by taking the group mean of all the individual changes scores and dividing this by the standard deviation of these change scores [35]. An effect size (or SRM) of 0.2 is regarded as small, 0.5 as moderate and 0.8 as large [36, 37]. This SRM allows a group-level interpretation of the study population undergoing treatment [34].

Secondly, unpaired *t* tests were used to detect significant differences between change scores (pre-treatment to the 3-month follow-up) for the good and the poor outcome groups (dichotomised as described above). In addition, the SRMs were determined and compared for "good" and "poor" outcome groups separately.

Thirdly, Receiver Operating Characteristics (ROC) curves were plotted. The responsiveness of a questionnaire can be analysed in an analogous manner to the evaluation of a diagnostic test [21]. The score-change for the questionnaire represents the diagnostic test and this is examined in relation to the "global outcome of surgery", which is taken to represent the "gold standard" or external criterion [21]. The resulting ROC curve displays the sensitivity and specificity for detecting a "good outcome" of several possible change-score cut-off points. The area under the ROC curve (AUC) describes how close the ROC plot compares to a perfect test

discriminating with 100% sensitivity and 100% specificity (AUC = 1.0) [38]. An AUC of 0.93, for example, means that a randomly selected patient from the "good" group has a greater change score than that of a randomly selected patient from the "poor" group 93% of the time. The sum of specificity and sensitivity was maximised by calculating the Youden index (Youden index = Sensitivity + Specificity − 1) [39]. According to Beaton et al.'s classification [34] determination of the cut-off score in this manner allows individual-level interpretation for the observed questionnaires, which facilitates the monitoring of change in individual patients. The analyses were conducted using PASW Statistics 18.0 (SPSS Inc., Chicago, IL, USA) and MedCalc (MedCalc Software, Mariakerke, Belgium) and statistical significance was accepted at the $P < 0.05$ level.

Results

Over the period of study, 134 patients were eligible for inclusion. However, only 89 patients of these had completed and returned all 3 questionnaires at baseline (Table 3). 14/89 (15.7%) patients were lost to follow-up leading to a follow-up group of 75 patients. The data from all 89 patients at baseline were used for the analysis of floor and ceiling effects and construct validity. The data from the 75 patients with follow-up questionnaires were used for the calculations of responsiveness and follow-up floor and ceiling effects. At follow-up, it was not possible to calculate a NASS-cervical "pain&disability" subscale score for one patient or a "neurology" subscale score for another, due to there being fewer than 80% items completed in the subscale (see "Methods"). The demographic, medical history and surgical variables describing the whole study group ($N = 89$) are shown in Table 4.

Floor and ceiling effects

Table 5 shows the percentage floor effects (worst status) and ceiling effects (best status) for each of the instruments. The COMI summary score, NASS-cervical pain&disability score, neurology score 1 and pain score, and the EQ-VAS each showed low (<15%) floor and ceiling effects at both baseline and follow-up.

At follow-up there were high but not adverse ceiling effects for the NASS-cervical neurology score 2 and the NASS-cervical disability score (33 and 26%, respectively) and for EQ-5D (33%). Similarly, all the individual COMI items displayed high ceiling effects at follow-up (19–48%). At baseline the COMI items "function", "disability" and "symptom specific well-being" showed high to adverse floor effects (29, 35 and 83%, respectively). Some of the individual items of the EQ-5D (mobility, self-care, and anxiety/depression) had very high ceiling effects at baseline (51–65%) and adverse (89–92%) ceiling effects at follow-up. EQ-5D pain had 37% floor effects at baseline and a similar percentage of ceiling effects at follow-up.

Construct validity

The relationships between each of the COMI item scores and the corresponding questionnaire scores are shown in Table 6. The COMI summary score showed moderate to high correlations with the EQ-5D, EQ-VAS, NASS-cervical pain&disability, NASS-cervical pain and NASS-cervical disability scores (−0.60 to 0.73) but low correlations with the two NASS-cervical neurology scores (0.24–0.37).

Correlation coefficients of 0.47–0.63 were found for the relationship between the COMI-neck pain item score and the NASS-cervical pain scale, NASS pain questions 1 and 5 and the two VASs for neck pain. Similar correlations (0.54–0.72) were found for the various measures of arm pain. The COMI item "function" correlated well (0.60) with the NASS-cervical pain&disability and NASS-cervical disability scores. Generally low correlations (−0.27 to −0.31) were found between the COMI item "symptom specific well-being" and the EQ-5D and EQ-VAS scores. COMI "general quality of life" scores showed correlations of −0.50 to −0.59 with the EQ-5D and the EQ-VAS scores. COMI "disability" scores showed correlations of 0.56–0.57 with the NASS-cervical pain&disability and the NASS-cervical disability scores.

The correlations for all the change scores showed slightly lower coefficients ($r = -0.22$ to -0.60 than for

Table 3 Overview of the number of questionnaires handed out and returned, preoperatively and at follow-up

The numbers marked in bold indicate the preoperative and follow-up groups used for analysis

	Number of questionnaires			
	Handed out	Returned preoperatively	Returned follow-up	Returned preoperatively and follow-up
COMI	134	130	115	114
NASS cervical	134	90	86	78
EuroQol	134	92	85	80
All questionnaires	134	**89**	79	**75**

Table 4 Study sample characteristics of baseline and follow-up population

	Baseline	Follow-up
Total number	89	75
Sex (male/female)	41 (46%)/48 (54%)	43 (57%)/32 (43%)
Age mean ± SD (range)	46.0 ± 8.4 (22.3-61.9)	49.0 ± 8.4 (24-64)
Diagnosis/number of patients		
Degenerative disease	88 (99%)	74 (99%)
Fracture/trauma	1 (1%)	1 (1%)
Most severely affected segment/vertebral body (frequencies)	C2/3 (1), C3/4 (3), C4/5 (7), C5/6 (51), C6 (1), C6/7 (26)	C2/3 (1), C3/4 (3), C4/5 (5), C5/6 (43), C6/7 (23)
Goal of surgery*		
Pain relief	82	70
Functional improvement	36	26
Neurological improvement	30	27
Complications		
Surgical complications	0	0
General complications	0	0
Number of disc prostheses		
1 disc prosthesis	85 (96%)	71 (95%)
2 disc prosthesis	4 (4%)	4 (5%)
Morbidity state		
ASA1 (no disturbance)	54 (60.5%)	43 (57.3%)
ASA2 (mild/moderate)	30 (34.5%)	27 (36.0%)
ASA3 (severe)	5 (5%)	5 (6.7%)

* More than one goal possible

the corresponding correlations of the absolute scores at baseline (Table 6).

Responsiveness

The global outcome ratings were distributed as follows: 58 (77.3%) helped a lot, 10 (13.3%) helped, 7 (9.3%) helped only little, 0 (0%) did not help, 0 (0%) made things worse. Hence the "good outcome" group consisted of 68 patients (90%) and the "poor outcome" group of 7 (10%).

There was a significant ($P < 0.001$) difference in the mean COMI change-scores for the good and poor outcome groups. Four out of the five NASS-cervical scores and the EQ-VAS (but not the EQ-5D) also showed significant differences between the scores for the good and poor outcome groups (Table 7, Fig. 1). The effect sizes (SRMs) giving information about the responsiveness or sensitivity to change for each of the instruments are compared in Fig. 1. The COMI showed the greatest difference between the SRMs for the good and poor outcome groups (2.34 and 0.34, respectively), i.e. it showed the best ability to discriminate between outcome groups, having a very high SRM in the good outcome group and a low SRM in the group with a poor outcome (Table 7). All the NASS subscales and EQ-5D scales showed smaller SRM differences between the good and poor outcomes indicating a worse discriminative ability.

Figure 2 shows the ROC curves for each of the questionnaires. The COMI summary score is the closest to the top left corner, i.e., shows the best discriminative function. This is also shown by the data for the AUC which was 0.96 for the COMI and significantly ($P < 0.05$) higher than the AUCs for all the other questionnaires (Table 8). The EQ-VAS showed a slightly greater AUC than the EQ-5D summary score but the difference was not significant. An improvement of 2.7 or more points in the COMI summary score predicted a good outcome with a sensitivity of 83.3% and specificity of 100% (Youden index 0.83). Summarising, with all three of the methods applied to examine responsiveness, the COMI showed the best ability to discriminate between good and poor outcomes.

Discussion

Patients, health insurances and governmental bodies increasingly expect outcome research to be carried out to evaluate the effectiveness of treatment and the performance of individual health professionals and hospitals. Short questionnaires like the COMI are ideal for the longitudinal assessment of treatment outcomes [40] and have various advantages over longer instruments, such as easier administration and higher completion rates [13]. Nonetheless, it is also important to use questionnaires with

Table 5 Floor effects (worst status) and ceiling effects (best status) for each of the questionnaire items/scales

Instrument	Preoperatively (89 patients) %		Follow-up (75 patients) %	
	Floor effects (worst health)	Ceiling effects (best health)	Floor effects (worst health)	Ceiling effects (best health)
COMI summary score	1.1	0.0	0.0	13.3
COMI neck pain	4.5	7.9	0	26.7
COMI arm pain	3.4	5.6	1.3	37.3
COMI high pain	4.5	2.2	1.3	18.7
COMI function	29.2	4.5	0.0	34.7
COMI symptom-specific well-being	83.3	0.0	8.0	37.7
COMI quality of life	16.9	1.1	0.0	29.3
COMI disability average	34.8	10.1	8.0	48.0
NASS-cervical Pain&disability score	0.0	0.0	0.0	10.8
NASS-cervical Neurology score 1	0.0	0.0	0.0	6.7
NASS-cervical Pain Score	10.1	1.1	1.3	12.0
NASS-cervical Neurology score 2	1.1	0.0	0.0	33.3
NASS-cervical Disability score	3.4	0.0	0.0	25.7
EQ-5D summary score	0.0	0.0	0.0	32.5
EQ-5D mobility	1.1	50.6	0.0	90.5
EQ-5D self-care	0.0	65.2	0.0	91.9
EQ-5D usual activities	15.7	16.9	1.4	67.6
EQ-5D pain	37.1	1.1	2.7	36.0
EQ-5D anxiety/depression	4.5	58.4	0.0	89.3
EQ-gen health VAS	2.3	1.1	0.0	13.3

adequate psychometric properties. For instruments designed to be used in longitudinal assessments i.e. as outcome instruments, responsiveness and validity are two of the most essential criteria [19, 41]. The single COMI items and the COMI summary score showed good correlations with the corresponding fuller questionnaires, indicating adequate construct validity, and the COMI also demonstrated good responsiveness.

Floor and ceiling effects

There were observable floor and/or ceiling effects (depending on the time-point of assessment) for the single items of all questionnaires examined in the present study. Some, notably the EQ-5D items mobility, anxiety/depression, and self-care at follow-up, and the COMI item symptom-specific well-being at baseline, even exceeded the critical level of 70% [40]. This might suggest that the responsiveness of these items would be limited because further change in an even more extreme direction might not be measurable. As a consequence of the critical ceiling effects for the EQ-5D individual items, the summary score of the EQ-5D also showed relatively high ceiling effects at follow-up. Despite there being some floor and ceiling effects for some single COMI items, the COMI-neck summary score showed no critical floor and ceiling effects.

Some argue that Likert scales with only five categories or the EQ-5D-style scale with only three categories may not be able to detect small but important changes [33]. As an alternative, 7-point or 10-point rating scales (similar to the pain VAS) have been recommended. However, in other studies, the 5-point Likert scale has been shown to display almost identical responsiveness to the 0–10 VAS, with the added advantage of being easier to administer and interpret [42]. The three-category response scale of the EQ-5D showed extremely high floor and ceiling effects and this likely diminished its responsiveness. This problem is known to the developers of the instrument and its further evaluation has led to the establishment of a 5-point scale for the EQ-5D similar to that used in the COMI [43–45]. Floor and ceiling effects are highly population-dependent [40]. The present study involved patients undergoing cervical spine surgery, who typically suffer from severe functional restrictions, neurological deficits and high pain preoperatively (see Table 4) and who generally have only minimal symptoms after treatment. This would be commensurate with greater floor effects preoperatively and greater ceiling effects postoperatively. It is likely that a group of patients with less severe symptoms undergoing conservative therapy for neck pain would not show as many floor effects at baseline or ceiling effects after treatment.

Table 6 Correlation coefficients describing the relationships between the COMI single items/summary score and reference items or full-length questionnaires at baseline in the 89 patients

COMI item	Reference instrument	Pearson r
Pain Symptoms		
How severe was your neck pain in the last week?	NASS-cervical Question C1	0.567**
	NASS-cervical Question C5	0.562**
	NASS-cervical neck pain VAS	0.629**
	NASS-cervical pain score	0.470**
How severe was your arm pain in the last week?	NASS-cervical Question C2	0.592**
	NASS-cervical Question C6	0.544**
	NASS-cervical arm pain VAS	0.715**
	NASS-cervical Pain score	0.593**
COMI high pain	NASS-cervical VAS high pain	0.643**
	NASS-cervical Pain score	0.548**
Function	NASS-cervical Pain&disability score	0.594**
During the past week, how much did you neck problem interfere with you normal work (including both work outside the home and housework)?	NASS-cervical Disability score	0.596**
Symptom-specific well-being	EuroQol-5D summary index	−0.268*
If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?	EuroQol-VAS	−0.310**
Quality of life	EuroQol-5D summary index	−0.589**
Please reflect the last week. How would you rate your quality of life?	EuroQol-VAS	−0.501**
Social and work disability average	NASS-cervical pain&disability score	0.564**
During the past 4 weeks, how many days did you cut down on the things you usually do because of your neck problem (social disability)/did your neck problem keep you from going to work (work disability)	NASS-cervical disability score	0.566**
COMI summary score	EuroQol-5D summary index	−0.596**
	EuroQol-VAS	−0.545**
	NASS-cervical pain&disability score	0.729**
	NASS-cervical neurology score 1	0.371**
	NASS-cervical pain score	0.570**
	NASS-cervical neurology score 2	0.235*
	NASS-cervical disability score	0.703**
ΔCOMI change score	ΔEuroQol-5D summary index	−0.425**
	ΔEuroQol-VAS	−0.387**
	ΔNASS-cervical pain&disability score	−0.601**
	ΔNASS-cervical neurology 1 score	−0.388**
	ΔNASS-cervical pain score	−0.596**
	ΔNASS-cervical neurology 2 score	−0.217
	ΔNASS-cervical disability score	−0.573**

* $P < 0.05$ level

** $P < 0.01$ level

Δ, difference between 3-month follow-up score and score before the operation

For further information on the NASS items, see Table 2

Many patients in the present study reported a good outcome and displayed high EQ-5D and low COMI and NASS-cervical scores at 3 months postoperatively. Generally speaking, a very high predominance of good outcomes might suggest that participation had been selective and be indicative of response bias. However, in the

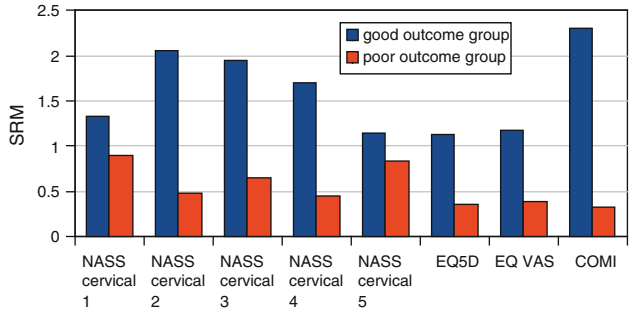

Fig. 1 Standardised response means (SRMs) for the good and poor outcome groups for each instrument, highlighting the ability of the instrument to discriminate between the groups. The higher the SRM in the good outcome group, the lower the SRM in the poor outcome group (should be close to zero) and the greater the difference between the SRMs for the two groups, the more discriminative is the instrument

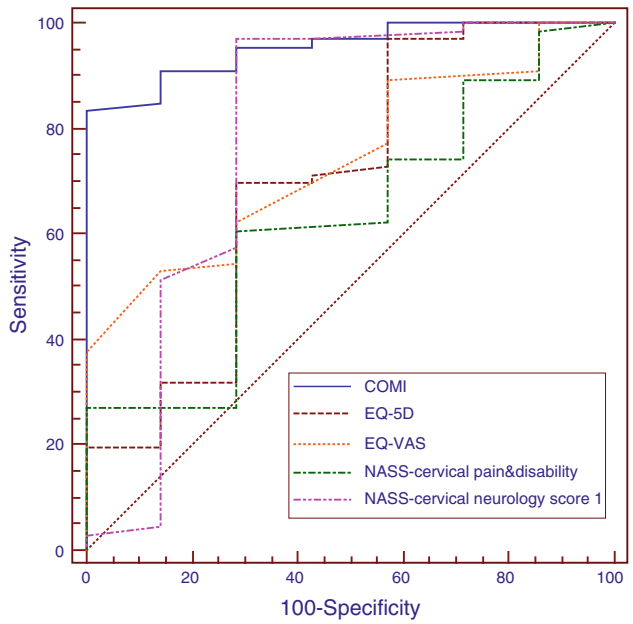

Fig. 2 Receiver operating characteristics curves for the different instruments. As an external criterion the global outcome question was chosen. See Table 8 for further details

present study we believe that it was simply the effective surgery that led to this distribution of outcomes. We deduce this from the fact that, in the larger group of patients that completed the COMI but not the other questionnaires (see Table 3), the % good outcomes [87.8% (101/115) patients; detailed data not shown] was similar to the value in the smaller group (90.6%, in 75 patients) that completed all three questionnaires preoperatively and at follow-up, and who were used in the comparisons of instrument responsiveness. The size of the follow-up group was lower due to the poorer rate of completion of the NASS and EQ-5D questionnaires. Whether this was the result of the different (and less "local") administrative system used to collect the data or the greater length of the questionnaire battery cannot be ascertained. High completion rates are essential to feel confident in measuring the benefit of the treatment, unbiased by selective participation.

Construct validity

As in the original validation study [15], the individual COMI items showed a good correlation with their reference scales with the exception of "symptom specific well-being". A possible explanation for this finding, namely that this item delivers unique important information for the summary score and should therefore continue to be included in the instrument, has been discussed before [15]. There was a much stronger correlation between the COMI and the NASS-cervical pain&disability score (and their respective change scores) than between the COMI and the NASS neurology score. This behaviour of the neurology score was also described by Stoll et al. [27] who found no correlation between this subscale and all SF-36 subscales. It is likely explained by the lack of any specific neurology assessment in the COMI and in the SF-36.

Responsiveness

For questionnaires that are to be used on a longitudinal basis, i.e. as outcome instruments, it is essential to know how well they are able to detect small but important changes [40, 46]. This information is used to inform clinical decisions and assist with the calculation of sample sizes in further studies. The *t* test results and the very low SRM for the poor outcome group and high SRM for the good outcome group indicated the excellent discriminative ability of the COMI. Examining the SRMs in the good and poor outcome groups separately was considered to be a fundamental necessity to see whether the questionnaires had the ability to differentiate between different global outcomes [34]. Evaluation of the SRM for the whole group alike fails to reveal whether an instrument also shows change where none is actually perceived by the patient. A responsive questionnaire should not show improvement or deterioration when none has occurred. This would not be an ideal characteristic for an outcome instrument. The NASS-cervical and the EuroQol showed less favourable SRM values than the COMI, and did not differentiate as well between good and poor outcome groups, suggesting they represented less responsive tools. A previous study [27] showed similar SRMs to those found in our study for the NASS pain&disability and the NASS neurology score 1 after conservative treatment over 3 weeks. A possible explanation for the greater responsiveness of the COMI might be the parsimonious choice of the COMI items, whereby only those that are most relevant to the condition

Table 7 Mean scores, standard deviation, SRM and *P* value (difference between outcome groups for the change score (baseline to follow-up)) of the different questionnaires split by the global outcome question or regarded as one group

Questionnaire	Mean Δ group outcome	SD Δ group outcome	SRM Δgroup outcome	Mean Δgood outcome	SD Δgood outcome	SRM Δgood outcome	Mean Δpoor outcome	SD Δpoor outcome	SRM Δpoor outcome	*P* value mean Δ good vs poor outcome
NASS-cervical Pain&disability	1.44	1.11	1.31	1.49	1.11	1.35	0.94	1.03	0.90	0.215
NASS-cervical neurology 1	1.68	0.97	1.73	1.78	0.86	2.07	0.71	1.46	0.49	0.005
NASS-cervical pain	2.35	1.36	1.72	2.48	1.27	1.95	1.11	1.71	0.65	0.011
NASS-cervical neurology 2	1.46	0.98	1.49	1.54	0.90	1.72	0.64	1.42	0.45	0.021
NASS-cervical disability score	1.29	1.14	1.14	1.34	1.16	1.17	0.79	0.95	0.84	0.241
COMI	4.2	2.61	1.86	4.78	2.08	2.34	0.55	1.64	0.34	0.000
EQ-5D	−0.35	0.34	−1.06	−0.38	0.33	−1.16	−0.13	0.39	−0.35	0.076
EQ-VAS	−32.10	29.65	−1.08	−34.29	29.20	−1.17	−10.71	27.15	−0.39	0.044

Table 8 Comparison of receiver operating curves for the different instruments

Questionnaire	Area	Standard error	95% confidence interval	Cut-off	Youden Index	Sensitivity	Specificity
COMI	0.95	0.03	0.88 to 0.99	2.70	0.83	83.3	100.0
NASS-cervical pain&disability	0.63	0.11	0.51 to 0.74	1.00	0.32	60.6	71.4
NASS-cervical neurology 1	0.79	0.14	0.68 to 0.87	0.37	0.68	97.0	71.4
NASS-cervical Pain	0.75	0.08	0.64 to 0.85	1.75	0.54	68.2	85.7
NASS-cervical neurology 2	0.76	0.08	0.65 to 0.86	>0.33	0.64	92.4	71.4
NASS-cervical disability	0.62	0.10	0.50 to 0.73	>1.89	0.30	30.3	100.0
EQ-VAS	0.74	0.08	0.63 to 0.84	<−40	0.39	53.0	85.7
EQ-5D	0.70	0.12	0.58 to 0.80	−0.17	0.41	69.7	71.4

are included, and the use in the summary score of the higher of the two pain scores (arm or neck pain), rather than either just neck pain or just radiating pain or an average of the two. Some patients suffer only from neck pain or only from arm pain. Pain is known to be one of the most responsive items in spinal surgery [47], and if the effect of intense pain in the most painful region is "diluted" by the averaging with pain scores for non-painful regions, then this will undoubtedly reduce the sensitivity of the pain item. Hyland [40] refers to this notion as shifting and non-shifting questions. In our study sample we observed that the items in the NASS-cervical lifting, walking, sitting, standing, stiffness, trembling and sexual activity, and the EuroQol items mobility, self-care and anxiety/depression had SRM values below 0.8, which indicates these were non-shifting elements (specific results not shown) and therefore likely diluted the average change-score.

The low responsiveness of the EuroQol compared with the COMI or NASS-cervical was not particularly unexpected, given that it is generic rather than condition-specific measure. The former are almost always less responsive, since the questions they contain are less specific to the condition in question and often contain non-shifting items (see above). As mentioned earlier, the EQ5D has only a 3-point scale with the two extremes effectively being "no problems" and "cannot do"; despite excellent treatment it is rare that patients change from the very worst to the best status or vice versa. The EuroQol is successfully used in cost-effectiveness analyses of treatment for spinal disorders [48] or to examine iatrogenic effects of treatment [49] but is not recommended for use as a standalone outcome instrument for specific conditions/disorders [50]. In ROC analysis, the EQ-VAS showed a greater AUC (0.74) than did the EQ-5D summary score (0.70) but with overlapping confidence intervals such that the two did not differ significantly. Interestingly, previous studies in patients with coronary heart disease, angina, stroke, diabetes, myocardial infarction, high blood pressure, joint pain, asthma have also shown that the single item EQ-VAS is more responsive than the EQ-5D summary score [51, 52].

The proximity of the COMI curve to the top left corner of the ROC curve and the high AUC value for the COMI

reflected its excellent ability to discriminate between the good and poor outcome groups. Its performance in this respect was better than either the NASS-cervical or EQ-5D. Previous studies calculating ROCs for the EQ-5D in 2 health surveys [53, 54] and in a study on the treatment of femoral neck fractures [55] observed similar AUCs (0.70–0.77) to that in the present study (0.70) However, unfortunately they did not calculate the Youden index or the corresponding values for sensitivity and specificity in detecting a good outcome or positive health status that would otherwise have allowed direct comparison with our data. To the best of our knowledge, no previous studies have carried out ROC analyses of the NASS-cervical in neck pain patients. The results of such analyses permit the monitoring and management of individual patients and are a fundamental element in Beaton et al.'s responsiveness classification [34]. In future studies, more attention should be paid to this useful analysis and outcome instruments should be evaluated for their sensitivity and specificity using the ROC method [56].

Limitations of the study

Our study has some limitations. The follow-up period was only 3 months. However, since the study was intended to compare the psychometric characteristics of the questionnaires themselves rather than report the outcome of the surgical procedure per se, the rather short 3-month period of follow-up was considered unproblematic, and in keeping with previous methodological studies [21]. Furthermore, such a follow-up period allowed the immediate effects of surgery to be assessed, in which most of the changes occur, and allowed the maximum number of datasets to be included; our decision was supported by the finding that previous studies of a similar nature have reported no significant change in outcome up to 2 years later [23, 57].

We did not evaluate the test–retest reliability of the COMI-neck in the present study, because White et al. [18] reported good reliability for the English version of the COMI-neck, and because the test–retest reliability of the German COMI-back, which is identical to the COMI-neck but for the fact that it enquires about back/leg symptoms rather than neck/arm symptoms, has also been confirmed [15]. However, the reliability of the COMI-neck might require further verification in relation to the specific patient group in which it is to be used in future studies.

The number of patients in the "poor outcome" group was rather low, which may limit the external validity of the responsiveness analysis. Nonetheless, in previous studies of the COMI-back [23] and the EQ-5D [53–55] similar results were obtained in terms of the responsiveness and minimal clinically important differences recorded.

From the initial 134 patients operated, only 75 could be included in the follow-up group, and this was predominantly the result of missing NASS-cervical/EQ5D questionnaires. We do not believe that this introduced a notable bias, though, since the outcome results for the COMI-neck in the larger group that completed only this questionnaire were similar to those reported for the group of 75 patients who completed all three (see earlier). Moreover, the baseline COMI scores for the larger group with a COMI but not the other questionnaires ($n = 130$) were similar to those for the group with all three questionnaires at baseline ($N = 89$) (detailed data not shown). In our spine centre we have observed how difficult the collection of questionnaires in daily practice can be without the employment of a dedicated study nurse/research assistant (the SWISSspine questionnaires were administered by the surgeon's secretary in conjunction with the SWISSspine registry for disc arthroplasty in Bern [20], whereas the COMI system is managed as part of an internal quality management system, run with dedicated staff from the Research Department).

A further ongoing problem in all responsiveness studies is the lack of an external gold standard for measuring treatment success. There is no consensus regarding the selection of an external criterion except that it should represent major clinical improvement or deterioration of health. Patient-orientated appraisals are widely accepted in the literature [21, 23, 35, 58–60], but other measurements, for example clinician and patient orientated assessment [59, 61] or return to full activity [21] may also be useful to examine in future studies. The global outcome criterion was included in the 3-month follow-up questionnaire that also contained the COMI. We examined whether this may have led to bias in that the global outcome was completed at the same time and under the same conditions as the COMI itself, and hence had a higher chance of being more closely related to it. However, there was a high correlation [$r = 0.7$ (data not shown)] between the highest pain score determined from almost identical single pain items in the SWISSspine and in the COMI instruments at follow-up, which would tend to suggest that this was unlikely the case.

None of the items in the COMI are weighted in the final score in relation to their perceived relative importance. The issue of weighting dimensions is an oft-discussed theme in the literature [33]. When the COMI was first developed the scores for the items were simply averaged for convenience and the excellent performance of the instrument resulted in the scoring being kept that way. An advantage of this method is of course its simplicity, in that it allows the quick and easy computation of the COMI summary score. Further studies might, however, examine whether other methods of computation would further improve its psychometric properties.

Conclusion

Our results demonstrate that the COMI-neck is a valid and responsive instrument for use in assessing the outcome of patients undergoing cervical spine surgery. Despite some large floor and ceiling effects for the individual items its responsiveness did not appear to be negatively affected: indeed, of the instruments examined, the COMI proved to be the best for discriminating between good and poor global outcomes. The COMI has the potential to serve as an outcome instrument not only for evaluating group outcomes in clinical trials, multicentre studies, routine quality management and surgical registry systems, but also for individual patient monitoring. In this way, the COMI can be used to enhance outcomes research, distinguish between useful and futile treatments, evaluate the performance of surgeons and hospitals and optimise the treatment of individual patients. Further analyses of the COMI-neck should be carried out in groups of non-surgically treated patients but, in view of the comparable performance of the COMI-back in both surgical and non-surgical groups [15], we are optimistic that the COMI-neck will perform just as well in non-surgical patients too.

Acknowledgments We are grateful to Gordana Balaban, Dave O'Riordan, Julian Amacker, Kirsten Clift, and Sara Preziosa, for their excellent work collecting the COMI and Spine Tango Surgical Registry data.

Conflict of interest None.

References

1. Fejer R, Kyvik K, Hartvigsen J (2006) The prevalence of neck pain in the world population: a systematic critical review of the literature. Eur Spine J 15:834–848
2. Borghouts J, Koes B, Vondeling H, Bouter L (1999) Cost-of-illness of neck pain in The Netherlands in 1996. Pain 80:629–636
3. Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine (Phila Pa 1976) 25:3100–3103
4. Strömqvist B (2002) Evidence-based lumbar spine surgery. The role of national registration. Acta Orthop Scand Suppl 73:34–39
5. Block A, Ohnmeiss D, Guyer R, Rashbaum R, Hochschuler S (2001) The use of presurgical psychological screening to predict the outcome of spine surgery. Spine J 1:274–282
6. Carragee E (2001) Psychological screening in the surgical treatment of lumbar disc herniation. Clin J Pain 17:215–219
7. Junge A, Dvorak J, Ahrens S (1995) Predictors of bad and good outcomes of lumbar disc surgery. A prospective clinical study with recommendations for screening to avoid bad outcomes. Spine (Phila Pa 1976) 20:460–468
8. Grob D, Porchet F, Kleinstück F, Lattig F, Jeszenszky D, Luca A, Mutter U, Mannion A (2010) A comparison of outcomes of cervical disc arthroplasty and fusion in everyday clinical practice: surgical and methodological aspects. Eur Spine J 19:297–306
9. Mannion A, Junge A, Elfering A, Dvorak J, Porchet F, Grob D (2009) Great expectations: really the novel predictor of outcome after spinal surgery? Spine (Phila Pa 1976) 34:1590–1599
10. Ahn H, Bhandari M, Schemitsch E (2009) An evidence-based approach to the adoption of new technology. J Bone Joint Surg Am 91(Suppl 3):95–98
11. Anderson P, Sasso R, Rouleau J, Carlson C, Goffin J (2004) The Bryan Cervical Disc: wear properties and early clinical results. Spine J 4:303S–309S
12. Deyo R (1988) Measuring the functional status of patients with low back pain. Arch Phys Med Rehabil 69:1044–1053
13. Bowling A (2005) Just one question: if one question works, why ask several? J Epidemiol Community Health 59:342–345
14. Deyo R, Battie M, Beurskens A, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976) 23:2003–2013
15. Mannion A, Elfering A, Staerkle R, Junge A, Grob D, Semmer N, Jacobshagen N, Dvorak J, Boos N (2005) Outcome assessment in low back pain: how low can you go? Eur Spine J 14:1014–1026
16. Ferrer M, Pellisé F, Escudero O, Alvarez L, Pont A, Alonso J, Deyo R (2006) Validation of a minimum outcome core set in the evaluation of patients with back pain. Spine (Phila Pa 1976) 31:1372–1379 discussion 1380
17. Melloh M, Staub L, Aghayev E, Zweig T, Barz T, Theis J, Chavanne A, Grob D, Aebi M, Roeder C (2008) The international spine registry SPINE TANGO: status quo and first results. Eur Spine J 17:1201–1209
18. White P, Lewith G, Prescott P (2004) The core outcomes for neck pain: validation of a new outcome measure. Spine (Phila Pa 1976) 29:1923–1930
19. Kirshner B, Guyatt G (1985) A methodological framework for assessing health indices. J Chronic Dis 38:27–36
20. Schluessmann E, Diel P, Aghayev E, Zweig T, Moulin P, Röder C, Group SR (2009) SWISSspine: a nationwide registry for health technology assessment of lumbar disc prostheses. Eur Spine J 18:851–861
21. Deyo R, Centor R (1986) Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 39:897–906
22. Beaton DE, Schemitsch E (2003) Measures of health-related quality of life and physical function. Clin Orthop Relat Res 90–105.
23. Mannion A, Porchet F, Kleinstück F, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 18(Suppl 3):374–379
24. Brooks R (2003) The measurement and valuation of health status using EQ-5D: a European perspective evidence from EuroQol BIOMED research programme. Kluwer, Dordrecht
25. Prieto L, Sacristán JA (2004) What is the value of social values? The uselessness of assessing health-related quality of life through preference measures. BMC Med Res Methodol 4:10
26. AAOS/NASS/SRS/CSRS/ORA/ASIA/COSS outcomes data collection questionnaires (1998) Cervical spine—baseline questionnaire, March 1998, Version 2.0
27. Stoll T, Huber E, Bachmann S, Baumeler H, Mariacher S, Rutz M, Schneider W, Spring H, Aeschlimann A, Stucki G, Steiner W (2004) Validity and sensitivity to change of the NASS questionnaire for patients with cervical spine disorders. Spine (Phila Pa 1976) 29:2851–2855
28. Sangha O, Wildner M, Peters A (2000) Evaluation of the North American Spine Society Instrument for assessment of health

status in patients with chronic backache. Z Orthop Ihre Grenzgeb 138:447–451
29. Daltroy L, Cats-Baril W, Katz J, Fossel A, Liang M (1996) The North American spine society lumbar spine outcome assessment Instrument: reliability and validity tests. Spine (Phila Pa 1976) 21:741–749
30. Andresen EM (2000) Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil 81:S15–S20
31. McHorney CA, Tarlov AR (1995) Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 4:293–307
32. Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, Bouter L, de Vet H (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 60:34–42
33. Streiner DL, Norman GR (1995) Health measurement scales: a practical guide to their development and use. Oxford University Press, Oxford
34. Beaton DE, Bombardier C, Katz JN, Wright JG (2001) A taxonomy for responsiveness. J Clin Epidemiol 54:1204–1217
35. Beurskens A, de Vet H, Köke A (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 65:71–76
36. Cohen J (1988) Statistical power analysis for the behavioral sciences. Lawrence Earlbaum Associates, Hillsdale
37. Kazis L, Anderson J, Meenan R (1989) Effect sizes for interpreting changes in health status. Med Care 27:S178–S189
38. Zweig M, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39:561–577
39. Youden W (1950) Index for rating diagnostic tests. Cancer 3:32–35
40. Hyland M (2003) A brief guide to the selection of quality of life instrument. Health Qual Life Outcomes 1:24
41. Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998) Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 2:i–iv, 1–74
42. Bolognese J, Schnitzer T, Ehrich E (2003) Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. Osteoarthr Cartil 11:499–507
43. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ (2008) Comparing the standard EQ-5D three-level system with a five-level version. Value Health 11:275–284
44. Janssen MF, Birnie E, Bonsel GJ (2008) Quantification of the level descriptors for the standard EQ-5D three-level system and a five-level version according to two methods. Qual Life Res 17:463–473
45. Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S (2007) Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care 45:259–263
46. Deyo R, Diehr P, Patrick D (1991) Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 12:142S–158S
47. Walsh TL, Hanscom B, Lurie JD, Weinstein JN (2003) Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. Spine (Phila Pa 1976) 28:607–615
48. Fritzell P, Berg S, Borgström F, Tullberg T, Tropp H (2010) Cost effectiveness of disc prosthesis versus lumbar fusion in patients with chronic low back pain: randomized controlled trial with 2-year follow-up. Eur Spine J
49. Hyland M (1992) Selection of items and avoidance of bias in quality of life scales. Pharmacoeconomics 1:182–190
50. Guyatt GH, Feeny DH, Patrick DL (1993) Measuring health-related quality of life. Ann Intern Med 118:622–629
51. Johnson JA, Coons SJ, Ergo A, Szava-Kovats G (1998) Valuation of EuroQOL (EQ-5D) health states in an adult US sample. Pharmacoeconomics 13:421–433
52. Bharmal M, Thomas J (2006) Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. Value Health 9:262–271
53. Cunillera O, Tresserras R, Rajmil L, Vilagut G, Brugulat P, Herdman M, Mompart A, Medina A, Pardo Y, Alonso J, Brazier J, Ferrer M (2010) Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. Qual Life Res 19:853–864
54. Petrou S, Hockley C (2005) An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. Health Econ 14:1169–1189
55. Tidermark J, Bergström G (2007) Responsiveness of the EuroQol (EQ-5D) and the Nottingham Health Profile (NHP) in elderly patients with femoral neck fractures. Qual Life Res 16:321–330
56. Bombardier C, Hayden J, Beaton DE (2001) Minimal clinically important difference. Low back pain: outcome measures. J Rheumatol 28:431–438
57. Siepe CJ, Mayer HM (2009) Is the final outcome predictable following total lumbar disc replacement? Proceedings of the International Society for the Study of the Lumbar Spine, Miami, Florida, USA, 4–8 May 2009
58. Campbell H, Rivero-Arias O, Johnston K, Gray A, Fairbank J, Frost H, Trial UMSS (2006) Responsiveness of objective, disease-specific, and generic outcome measures in patients with chronic low back pain: an assessment for improving, stable, and deteriorating patients. Spine (Phila Pa 1976) 31:815–822
59. Stratford PW, Binkley JM, Riddle DL, Guyatt GH (1998) Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. Phys Ther 78:1186–1196
60. MacKenzie CR, Charlson ME, DiGioia D, Kelley K (1986) Can the Sickness Impact Profile measure change? An example of scale assessment. J Chronic Dis 39:429–438
61. Stratford PW, Binkley J, Solomon P, Gill C, Finch E (1994) Assessing change over time in patients with low back pain. Phys Ther 74:528–533