

Cross-cultural adaptation and validation of the Norwegian version of the Core Outcome Measures Index for low back pain

Kjersti Storheim · Jens Ivar Brox · Ida Løchting ·
Erik L. Werner · Margreth Grotle

Received: 21 July 2011 / Revised: 24 April 2012 / Accepted: 26 May 2012 / Published online: 14 June 2012
© Springer-Verlag 2012

Abstract

Purpose The Core Outcome Measures Index (COMI) is a short multidimensional scale covering all domains recommended to be included as outcome measures for patients with low back pain (LBP). The purpose of the present study was to translate and cross-culturally adapt the COMI into Norwegian and to test clinimetric properties of the Norwegian COMI version in patients with non-specific LBP recruited from various clinical settings.

Methods Ninety patients with non-specific LBP from primary care and hospital settings participated in the validation part and 61 also in the reproducibility part of the study (1 week apart). Acceptability, data quality, reproducibility and construct validity were investigated.

Results The questionnaire was well accepted and with little missing data and end effects. Reliability in terms of intraclass correlations (ICC) was satisfactory for the COMI index [0.89 (95 % CI 0.82–0.94)] and most single-core items. Agreement was acceptable for the COMI index [standard error of measurement ($SEM_{\text{agreement}}$) 0.80, minimal detectable change ($MDC_{\text{individual}}$) 2.21], but exceeded the minimal standard of acceptability in some of the individual core items. Construct validity was acceptable for the COMI index.

Conclusion The Norwegian version of the COMI index shows acceptable clinimetric properties in our patient population, but some of the sub-items had shortcomings. Our study, however, support the usefulness of the COMI index as an applicable stand-alone global scale when a light respondent burden is advisable.

Keywords Multidimensional scale · COMI · Clinimetric properties · Low back pain

Introduction

Patient reported outcome measures are today highly recommended when evaluating treatment success in clinical trials, routine quality management and registry systems [5, 22]. Several measures exist and the number and types of generic and disease-specific measures are growing [10, 22, 28]. Therefore, it has become a large challenge for clinicians and researchers to choose among a myriad of questionnaires. Moreover, comprehensive questionnaires covering numerous domains may become long, time-consuming and tiresome to complete, and result in reduced response rate and huge administration burden [23].

International panels of experts have proposed a set of five domains to be included in a standardized set of outcome measures for patients with spinal disorders: back-specific function, pain, generic health status, work disability and patients' satisfaction [5, 10]. Numerous full-scale questionnaires exist within each domain. However, few are short and easily administered. In order to alleviate patients and administrative burden and to increase the standardization of outcome measures within the field, a very parsimonious multidimensional six-item core set covering all recommended domains was proposed by the expert panel [10].

K. Storheim (✉) · I. Løchting · M. Grotle
Communication and Research Unit for Musculoskeletal Disorders (FORMI), Oslo University Hospital and University of Oslo, Building 37B, Nydalen, Box 4956, 0424 Oslo, Norway
e-mail: Kjersti.Storheim@oslo-universitetssykehus.no

K. Storheim · J. I. Brox
Department of Orthopaedics, Oslo University Hospital and University of Oslo, Oslo, Norway

E. L. Werner
Research Unit for General Practice, Uni Health, Bergen, Norway

This core set consists of six individual questions chosen from widely used instruments that has been studied and validated in former studies (such as the SF-36, the US Agency for Health Care Policy and Research, the World Health Organization quality of life—BREF, and the US National Health Interview Survey) and cover the key issues considered by the expert panel to be of greatest significance in low back pain (LBP) research. Deyo et al.'s [10] six-item multidimensional core set has been further developed by adding one more question that covers general quality of life and by establishing a composite index score (the Core Outcome Measures Index: COMI) [19]. The COMI is short and a very simple and easy-to-use scale and is recommended to promote standardization of outcome measurements in clinical trials, multicentre studies, clinical practice, quality improvement efforts and surgical registry systems, especially when it is more important to examine the perceptions of the majority in regard to a few key issues than to examine the outcome of just a selected few in great detail [19]. It has been used as outcome measure in various study designs [15, 16, 25] and is implemented as patient-based outcome in the International Spine registry [31]. The COMI has been translated into a growing number of languages [31], but published studies of psychometric properties of the COMI have mainly been conducted in surgically treated patients or in patients recruited from hospital settings [7, 12, 18–21]. Hence, the present study aimed to translate and cross-culturally adapt the COMI into Norwegian and to test clinimetric properties of the Norwegian COMI version in patients with non-specific low back pain recruited from various clinical settings.

Methods

Design

The study was carried out in two stages: the first stage included translation and cross-cultural adaptation of the COMI; the second stage was to test the clinimetric properties of the Norwegian COMI version in a cross-sectional validation study and a test–retest study within 1 week interval.

Translation and cross-cultural adaptation

The translation and cross-cultural adaptation was conducted according to recommendations from international guidelines (“Appendix”) [4]. The English and Norwegian versions are published in the “Appendix”.

Testing the clinimetric properties

This stage was performed on patients with low back pain recruited from different clinical settings in Oslo, Norway,

including three physiotherapy clinics (primary care), one outpatient rehabilitation clinic, one pain clinic (University Hospital) and one orthopaedic department (University Hospital). Eligible participants were patients with non-specific low back pain of more than 6 weeks duration aged >18 years and able to speak, read and write in Norwegian. Exclusion criteria were sciatica or “red flags”. Written informed consents were obtained from all patients. The Regional Ethical Committees for Medicine and Health and the Norwegian Data Inspectorate approved the study, which followed the Helsinki Declaration.

Patients completed the COMI together with socio-demographic variables and various reference scales. To assess test–retest reproducibility, the COMI was re-administered to patients consenting to participate also in the reproducibility-part of the study the next time they came to the treating physiotherapist or doctor, preferably 1 week after they first filled out the questionnaire. Additionally, a global question (six-point Likert scale) recording change in low back pain condition in the time interval was completed by the patients.

The study sample size was determined according to the recommendations from Terwee et al. [28] which suggest that at least 50 patients are necessary to test construct validity, test–retest reliability and ceiling/floor effects, whereas approximately 100 patients are needed to perform an internal consistency analysis.

Measures

The COMI consists of seven questions covering five domains. Domains, wording and response format for each single question are presented in Table 2. The COMI index score (range 0–10) is calculated by averaging transformed core-item scores from each domain [pain symptoms (two questions (back/leg pain)), back function (one question), symptom-specific well-being (one question), general well-being (one question) and disability [two questions (social/work disability)]. The transformation process is as follows: visual analogue scales for pain symptoms; highest value out of the two scores (back/leg pain) is noted, Likert scales (back function, symptom-specific well-being, general well-being, disability); category score marked by the patient are re-scored on a 0–10 scale (patients score –1 multiplied by 2.5). Likert scales for disability are first averaged before transformation. Higher numbers on the COMI index score indicate more symptoms [19, 21].

Reference scales (for construct validity): to evaluate construct validity the Roland Morris Disability Questionnaire (RMDQ) [24], EuroQol-5 Dimensions Index (EQ-5D index) [1], Brief Illness Perception Questionnaire (BIPQ) [6] and Hopkins Symptom Check List-25 (HSCL-25) [9] were used as reference scales.

Analysis

All data analyses were performed using SPSS version 15.0 (SPSS Inc., Chicago, IL). The clinimetric properties of the Norwegian COMI version were tested as follows.

Data quality

Here, missing data and end effects were described. Floor or ceiling effects were considered to be present if more than 15 % of the patients reported the lowest or the highest possible score, respectively [28].

Reproducibility

Test–retest analysis was applied on each of the five core items/domains, and on the composite COMI index score. Main analysis was performed on all patients participating on both test and retest. A supplementary analysis was performed in patients reporting a stable (unchanged) low back pain status from test to retest.

Agreement was expressed by the standard error of measurement (SEM) and minimal detectable change (MDC). $SEM_{\text{agreement}}$ was calculated by taking the square root of the error variance of an ANOVA analysis (obtained by taking $\sqrt{\text{within people residual mean square}}$) [28]. The $MDC_{\text{individual}}$ was calculated using the formula $MDC = 1.96 \times \sqrt{2} \times SEM$. The MDC represents the smallest score change which, with $P < 0.05$, can be interpreted as real change and not measurement error.

Reliability was assessed by the intraclass correlation coefficient ($ICC_{\text{agreement}}$) using a two-way random effects model [28]. Acceptable level was set to >0.70 [28]. In addition, weighted kappa with quadratic weighting was used for core items 3–5 since data are categorical and the categories are ordinal (Likert scales 1–5). Kappa values were categorized according to Altman: poor (0 to <0.2), fair (0.21–0.40), moderate (0.41–0.60) good (0.61–0.80) and very good (0.81–1.00) [2].

Construct validity

Construct validity was assessed by testing the correlation between each single core item (domains) and full scales or single questions of matching reference scales, and the correlation between COMI index score and matching reference scales according to a priori formulated hypothesis (Table 4).

Pain symptoms (highest of back/leg pain) were validated against question number 5 from the BIPQ (How much do you experience symptoms from your back problems?) [6], item number 4 (pain/discomfort) from EQ-5D [1] and the full scale of RMDQ [24].

Back function was validated against the full scale of RMDQ [24] and item number 3 (usual activities) from EQ-5D [1].

Symptom-specific well-being was validated against the full scales of the EQ-5D [1] and HSCL 25, and question number 1 from the BIPQ (how much does your back problem affect your life?) [6].

General well-being was validated against the full scale of EQ-5D [1] and to question number 1 from the BIPQ (see above) [6].

Disability (mean of social/work disability) was validated against RMDQ full scale [24] and item number 3 (usual activities) from EQ-5D [1].

Core Outcome Measures Index score was validated against full scales of EQ-5D [1], RMDQ [24] and HSCL-25 [9].

Spearman's rho was used in all correlation analysis since the COMI and reference scales represent a variety of scale types. Additionally, some items were not normally distributed. Correlation coefficients under 0.3, between 0.3 and 0.6 and over 0.6 were considered low, moderate and high, respectively [3].

Results

A total of 90 patients were included in the study. Sixty-one participated in the test–retest study of which 59 had complete COMI scores at both measurements. The time interval between test and re-test was median 7 days (range 1–31 days). Study sample characteristics are shown in Table 1.

Data quality

All data were normally distributed, except from “symptom specific well-being”. There was relatively little missing data (Table 2). Most frequently the question asking for leg pain was missing. Floor and ceiling effects were not present in the composite index score, but there were end effects in item two (leg pain/sciatica), item four (symptom-specific well-being), item six (social disability) and seven (work disability) (Table 2). However, when taking the highest out of pain symptoms and averaging the two disability items (according to the procedure for calculating COMI index score), end effects were not more present for pain symptoms (lowest and highest 2.2 and 1.1 %, respectively) and for disability (lowest and highest 10.0 and 8.9 %, respectively).

Reproducibility

There were no systematic differences between the test and retest scores. Table 3 shows the results for agreement and test–retest reliability for the COMI. The COMI index score and most of the core items showed good test–retest reliability. Reliability according to the ICCs was above the recommended minimum standard for all core items, with the

Table 1 Patient demographic characteristics and clinical status

	Validity study (<i>n</i> = 90)	Test–retest study (<i>n</i> = 61)
Sex (female, no) (%)	52 (57.8)	32 (52.5)
Age in years (mean (SD))	47.6 ± 11.7	49.2 ± 11.2
Recruited from (no, %)		
Primary health care physiotherapy	30 (33.3)	14 (23.0)
Outpatient rehab clinic	24 (26.7)	22 (36.1)
Orthopaedic dept university hospital	30 (33.3)	21 (34.4)
Pain clinic university hospital	6 (6.7)	4 (6.6)
Employment status (no, %)		
Working	40 (44.4)	24 (39.3)
Sick listed	18 (20.0)	15 (24.6)
Pension (disability, retirement)	25 (27.8)	18 (29.5)
Former back surgery (yes, no, %)	33 (36.7)	23 (37.7)
Duration current episode (all) (weeks) (mean (SD))	503 ± 683	572 ± 689
<3 months (no, %)	19 (21.1)	11 (18.0)
>3 months (no, %)	71 (78.9)	50 (82.0)
Back pain (NRS) (0–10) (mean (SD))	4.8 ± 2.1	4.9 ± 2.1
Disability (RMDQ) (0–24) (mean (SD))	7.6 ± 5.2	7.2 ± 5.1
HSCL-25 (1–4) (mean (SD))	1.61 ± 0.49	1.64 ± 0.49
EQ-5D (mean (SD)) ^a	0.54 ± 0.31	0.55 ± 0.29

NRS numeric rating scale, RMDQ Roland–Morris disability questionnaire, HSCL-25 Hopkins symptom check list

^a Higher score indicate better health

exception of the item symptom-specific well-being (0.68). Weighted Kappa was moderate to good and ranged from 0.51 to 0.68. Again, the item symptom-specific well-being showed the weakest result. The ICC for the COMI index score was 0.89 (95 % CI 0.82–0.94) and the SEM and MDC was 0.80 and 2.21, respectively. MDC % for the COMI index score and all core items, however, exceeded 20 %.

Sensitivity analysis of the 34 patients who scored “no change” at the global rating scale at re-test revealed similar results in all reproducibility analysis. Another sensitivity analysis excluding patients with short (<5 days) and long time interval (>14 days) between test and re-test neither changed the results.

Construct validity

The hypotheses were mostly confirmed by the correlations between core items/COMI index score and reference scales, except from correlation between the COMI index score and RMDQ/HSCL where the correlation coefficients were somewhat higher than expected (Table 4).

Discussion

Overall, this study shows that the Norwegian version of the COMI holds acceptable standard in all clinimetric tests of

the COMI index in patients with non-specific LBP recruited from different clinical settings, suggesting that this short multidimensional scale covering all key issues of greatest interest in LBP research may serve as a useful stand-alone questionnaire when a lighter respondent burden than the alternative battery of full-scale questionnaires to measure relevant domains is required. Some of the subscales, however, did not reach optimal standard in terms of agreement and validity.

Only two former methodological studies of the COMI have included conservatively treated patients [7, 19]. However, the forerunner for the COMI, Deyo et al.’s six-item multidimensional core set, has previously been tested for clinimetric properties in LBP patients with good results [11]. Further, a pilot study describing the use of COMI as a documentation instrument for conservative treated neck- or LBP patients in the international spine registry exists [14], in addition to a validation study of a neck-version of Deyo et al.’s core set [30]. Kessler et al. [14] found the COMI to be useful and feasible, and White et al. [30] concluded that the core neck pain questionnaire was reliable and valid. Three recent studies testing the clinimetric properties of COMI in Italian [18], French [12] and Brazilian–Portuguese [7] speaking LBP populations all found the COMI to hold good psychometric properties in patients mainly recruited from hospital settings. Finally, an adaptation of

Table 2 Missing data, end effects and floor- and ceiling effects for the CORE items

Core items (domains, response format and wording) (<i>n</i> = 90)	Missing data, <i>n</i> (%)	Mean (SD)	Lowest (%)	Highest (%)
Pain symptoms (0–10, visual analogue scale)				
How severe was your back pain in the last week? 0 = no pain and 10 = worst pain that I can imagine	3 (3.3)	4.79 (2.10)	2.2	0
How severe was your leg pain (sciatica)/buttock pain in the last week? 0 = no pain and 10 = worst pain that I can imagine	6 (6.7)	3.79 (2.86)	15.6	1.1
Back function (1–5, Likert scale)				
During the past week, how much did your back problems interfere with your normal work (including both work outside the home and housework)? (1 = not at all, 2 = a little bit, 3 = moderately, 4 = quite a bit, 5 = extremely)	0	3.31 (0.97)	3.3	8.9
Symptom-specific well-being (1–5, Likert scale)				
If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it? (1 = very satisfied, 2 = somewhat satisfied, 3 = neither satisfied nor dissatisfied, 4 = somewhat dissatisfied, 5 = very dissatisfied)	0	4.21 (1.01)	1.1	53.3
General well-being (1–5, Likert scale)				
Please reflect on the last week. How would you rate your quality of life? (1 = very good, 2 = good, 3 = moderate, 4 = bad, 5 = very bad)	0	2.94 (0.85)	4.4	2.2
Disability (1–5, Likert scale)				
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 back problem? (1 = none, 2 = between 1 and 7 days, 3 = between 8 and 14 days, 4 = between 15 and 21 days, 5 = more than 22 days)	1 (1.1)	2.85 (1.33)	14.4	16.7
During the past 4 weeks, how many days did your back problem keep you from going to work (job, school, house work)? (1 = none, 2 = between 1 and 7 days, 3 = between 8 and 14 days, 4 = between 15 and 21 days, 5 = more than 22 days)	1 (1.1)	2.57 (1.32)	24.4	12.2

Table 3 Test–retest reliability results for each core item/domain and the composite index score (COMI index score)

Domain/core items (<i>n</i>)	Range	Mean (SD) first test	Mean (SD) re-test	SEM	MDC	MDC (%)	ICC (95 % CI)	Kappa W (95 % CI)
Pain symptoms (61)	0–10 (highest) ^a	5.41 (2.15)	5.48 (2.18)	1.12	3.11	31.1	0.85 (0.74–0.91)	NA
Back function (61)	1–5	3.38 (0.90)	3.15 (1.0)	0.52	1.45	29.0	0.81 (0.67–0.89)	0.68 (0.53–0.08)
Symptom-specific well-being (61)	1–5	4.11 (1.07)	4.07 (1.14)	0.77	2.13	42.6	0.68 (0.47–0.81)	0.51 (0.26–0.77)
General well-being (61)	1–5	2.90 (0.75)	2.79 (0.91)	0.52	1.43	28.6	0.76 (0.60–0.86)	0.61 (0.45–0.77)
Disability (59)	1–5 (mean) ^b	2.76 (1.20)	2.62 (1.22)	0.57	1.56	31.2	0.88 (0.79–0.93)	NA
COMI index score (59)	0–10	5.66 (1.76)	5.39 (1.92)	0.80	2.21	22.1	0.89 (0.82–0.94)	NA

SEM_{agreement} $\sqrt{\text{within people residual mean square}}$, MDC_{individual} ($\sqrt{\text{within people residual mean square}} \times 2.77$, ICC_{agreement} two-way random effects model (absolute agreement)

MDC % MDC as percentage of maximum score, Kappa W weighted kappa with quadratic weighting, CI confidence interval, NA not assessed

^a Highest pain of back/leg pain

^b Mean of social/work disability

the COMI for patients operated for inguinal hernia (COMI hernia) has been found to have good psychometric properties [26]. This supports the use of brief core

scales as an alternative for a battery of full-scale questionnaires not only in low back pain but also in various fields in medicine.

Table 4 Construct validity: a priori formulated hypothesis

COMI-domain	Hypothesis	Correlation value ^a	Hypothesis confirmed?
Pain symptoms	Pain symptoms were expected to be moderately to highly correlated to question number 5 (back symptoms) from the BIPQ and to the item for pain/discomfort from EQ-5D Questionnaire. Since RMDQ measure different aspects of pain-related disability, a moderate correlation was expected	BIPQ: 0.51 EQ-5D pain: 0.46 RMDQ: 0.51	BIPQ: yes EQ-5D: yes RMDQ: yes
Back function	We expected moderate to high correlation between the domain “back function” and RMDQ and the item for usual activities from EQ-5D	RMDQ: 0.54 EQ-5D function: 0.61	RMDQ: yes EQ-5D: yes
Symptom specific well-being	Symptom-specific well-being was expected to be moderately to highly correlated to question number 1 from BIPQ since they are both appear to measure the same construct, and moderately correlated to EQ-5D full scale and HSCL-25 since they measure slightly different aspects of health	BIPQ: 0.34 EQ-5D full: -0.43 HSCL-25: 0.37	BIPQ: yes EQ-5D: yes HSCL-25: yes
General well-being	We expected this question to be moderately to highly correlated to EQ-5D full scale since they both measure the same construct, and moderately correlated to question no 1 from BIPQ measuring slightly different aspects of health	EQ-5D full: -0.69 BIPQ: 0.60	EQ-5D: yes BIPQ: yes
Disability	The disability questions in the COMI scale asks for days out of action and the RMDQ and the item for usual activities from EQ-5D Questionnaire asks for patients’ capacity related to daily functional activities. Hence, moderate correlation between the COMI-question and those reference scales was expected	RMDQ: 0.47 EQ-5D function: 0.47	RMDQ: yes EQ-5D: yes
COMI index	The COMI index and EQ-5D are both scales consisting of items covering a mix of health aspects and a high correlation was expected. RMDQ and HSCL are more “clean” scales measuring pain-related activity and psychological phenomena, respectively and were expected to correlate moderately to the COMI index	EQ-5D full: -0.71 RMDQ: 0.64 HSCL-25: 0.68	EQ-5D: yes RMDQ: no HSCL-25: no

All correlations are significant at the 0.01 level (2-tailed)

Correlation coefficients under 0.3 = low, between 0.3 and 0.6 = moderate, over 0.6 = high

BIPQ brief illness perception questionnaire, EQ-5D EuroQol-5 dimensions index, RMDQ Roland-Morris disability questionnaire

^a Spearman’s rho

Checking the data quality of the present study revealed that all single core items were normally distributed, except from the question dealing with symptom-specific well-being, in which the patients strongly tended to score in the higher end. End effects were also seen for the individual pain- and disability questions, but there were no floor/ceiling effects when the items were combined in the composite index score. These findings are in line with previous studies [7, 12, 18, 19]. The question regarding leg pain had the highest proportion of missing data (6.7 %), and many of the patients scored in the lower end indicating no leg pain (15.6 %). This finding probably reflects our sample, in which we excluded patients with sciatic symptoms. Mannion et al. [19] included LBP patients with and without leg pain and did not report this phenomenon.

In the present study, the Norwegian version of the COMI showed acceptable reproducibility for the composite

COMI index and most of the individual core items for a median time between test and retest of 7 days. This time interval was chosen since it minimizes the possible memory effect and the chance of change in low back pain status among patients. Patients completed several instruments which might also reduce the possibility to remember their previous responses. Furthermore, we also consider the patients who took part in the retest study to be representative for this population as they were similar to the total cohort with regard to all baseline characteristics. Intraclass correlations and weighted kappa was above the recommended minimum standard for all single core items, with the exception of the item symptom-specific well-being where ICC was just below the recommended minimum standard. Reliability was also good for the composite COMI index score. Our results in terms of ICC are in line with former reliability studies of the COMI back [7, 12, 18, 19]

and somewhat better than for COMI hernia [26]. Compared to full scale questionnaires our results indicate that the reliability of the composite COMI index score is at least equal to back pain-specific questionnaires [24] and to generic questionnaires frequently used in the field of low back pain [17]. Agreement expressed by SEM and MDC were, however, not optimal for some of the individual core items, but acceptable for the sum scale and only marginally poorer than in former studies of the COMI [7, 12, 19, 26] and within the range of that reported for other LBP outcome instruments [8, 19]. Still, an MDC% of 22.1 % for the COMI index means that a change of more than 2.2 points at the COMI index need to be observed to be judged as “a real change”.

Test of construct validity of the COMI found moderately to high correlations for all tests performed, indicating that both single-core items and the composite index score showed acceptable construct validity within this field where gold standards are not available. It should, however, be mentioned that the correlation between the core items “pain symptoms” and “back function” and our reference scales were somewhat weaker in our study compared to in Mannion et al.’s study [19] but similar to recent studies of Italian and French-speaking back pain populations [12, 18] and slightly better than in Brazilian–Portuguese populations [7]. To our surprise, the correlation between “pain symptoms” in the COMI, and RMDQ and back symptoms in the BIPQ were identical. We hypothesized that back symptoms in the BIPQ might better correlate with the core-item “pain symptoms” in COMI than with RMDQ. Our expectations for correlation between the subscale pain symptoms from the BIPQ, a questionnaire assessing illness perception, might have been overestimated. Further, there is considerable overlap between domains like pain, back function and disability in the patient population included in the present study [29] and tools limited to the evaluation of one single domain in back patients often have a mixed content reflecting various constructs [13]. Correlations were within the lower bound for the domain “symptom specific well-being”, which is in line with all other methodological studies of the COMI [7, 12, 18, 19] and its forerunner [11]. Since the reliability also was weakest for this particular COMI item, it may be questioned whether this item should be kept in the COMI. However, Mannion argued that it may deliver unique information that may be of importance to the multidimensional nature of the overall index and recommended its continued inclusion in the core-set [19]. Still, concerns about the rationale behind summarizing these five domains into a single global score representing a new undefined domain have been raised [11]. Mannion et al. [19] introduce the possibility of examining change for each domain separately. This might be applicable at the patient level in daily clinical practice

for the establishment of a patient profile, treatment evaluation and quality assurance purposes. The interpretation should, however, be done with caution since the results of the present study indicate limitations in clinimetric properties for some individual items. Lower measurement error in the COMI index score suggest the use of the index score on group level evaluation in clinical settings, multicentre studies and registries when it is desirable to assess the impact of spinal disorders on multiple patient-orientated outcome domains quickly and with high respondent rate. Former studies have found the COMI to hold rather high level of internal consistency [11, 19], although the scale is multidimensional and does not purport to measure a defined underlying construct. In patients with non-specific LBP, domains like pain, back function and work loss is closely linked and influenced by multiple factors [29]. This may explain the relatively high level of reported homogeneity for this multidimensional scale and support the use of the COMI index as a global measure for back patients.

Our limit for an acceptable level of validity was set to 0.3 [3]. This is somewhat more liberal than other studies testing the validity of COMI who all define values above 0.4 to be acceptable [7, 12, 18]. Using 0.4 as an acceptable limit for construct validity implies that correlations between the subscale “symptom specific well-being” and BIPQ and HSCL-25 fall outside what should be interpreted as sufficient validity. However, weak correlations do not only rely on how close the scales are with regard to underlying construct, but also on the reliability of the scales being compared [27]. The subscale “symptom specific well-being” was one out of two sub-scales with weakest reliability in our study. Substantially higher correlations between other sub-scales and BIPQ and between HSCL-25 and the COMI index may indicate that low correlation at least partly may be explained by limited reliability of this particular sub-scale. On the other hand, it might also indicate that the subscale “symptom specific well-being” measure a different phenomena than its reference scales.

One should consider that constantly introducing new tools may result in little shared understanding of what certain results mean, what their clinical relevance may be, or how the patient populations and results of different studies may compare [10]. However, the COMI is a result of an initiative taken by international panels of experts requesting better standardization of outcome measures within the field, and as an outcome in quality management or studies of treatment effectiveness when it is more important to examine the perceptions of the majority with regard to a few key issues than to examine the outcome of just a selected few in great detail [10, 19]. The introduction of a short core-set may facilitate many types of comparison and pooling of data [10, 19]. Further, the administration

burden is less when the scale is short and easy-to-use and patient compliance is likely to decline when batteries of long questionnaires are administered [23].

Strengths of the present study are that all measurement properties of health status questionnaires proposed by Terwee et al. [28] possible within our design are examined and that our sample size was satisfactory for all parts of the study. Further, this is one of the first studies of psychometric properties of the COMI conducted in a language other than German or English and we had very little missing data. Our study includes patients recruited from various clinical settings, primary care included, in contrast to former studies mainly recruiting patients from specialized health care. One limitation is the lack of responsiveness to change data for the COMI, which was impossible to obtain within our design. This will be assessed in a future study. Further, our reproducibility analysis both includes patients reporting a change in LBP condition from test to re-test and patients filling out the re-test questionnaire outside the pre-decided time interval. Sensitivity analysis of these two divergences seems, however, did not influence the results.

The Norwegian version of the of the COMI index shows acceptable clinimetric properties in terms of acceptability, data quality, test–retest reliability and construct validity when used in a sample of patients with non-specific low back pain recruited from various clinical settings, although some of the individual core items did not provide optimal results in terms of agreement and validity. Our study, however, supports the clinical appropriateness of the COMI index and it can be recommended for application in non-specific low back pain patients in primary and secondary health care. The high prevalence and socioeconomic burden of low back pain may further support the use

of a short core-set in order to reduce administrative and respondent burden in settings where batteries of full-scale questionnaires are redundant.

Acknowledgments The authors acknowledge clinicians at Hans and Olaf physiotherapy institute, Hjelp24NIMI, Stadion Fysikalske, Aker University Hospital (pain clinic), Friskvernsenteret and the National Hospital (Orthopaedic Department) for recruiting patients for the study. We also thank all patients participating in the study.

Conflict of interest None.

Appendix

The translation process: the translation and cross-cultural adaptation was conducted according to recommendations from international guidelines [4]. The original COMI version was first “forward” translated into Norwegian by two independent translators whose mother tongue is Norwegian and with different profiles (one clinician and one philologist). A consensus of the Norwegian translation was made before it was translated back into English by two native-English speaking translators, who were blinded to the original English version. In a formal meeting the translators, one health professional and the researchers in our research group reviewed all translations and discussed possible discrepancies until consensus on a final version of the COMI was achieved. The final translated Norwegian version was reviewed by the first patients included in the clinimetric study. Since they had no problems with reading, interpreting and filling in the questionnaire, no changes were made. The English and Norwegian versions follow below.

Back problems can lead to back pain and/or pain in the legs/buttocks, as well as to sensory disturbances such as tingling, 'pins and needles', or numbness in any of these regions.

- 1 Which of the following problems troubles you **the most**? Please tick **ONE BOX** only.
- ₁ back pain
 - ₁ leg/buttock pain
 - ₂ sensory disturbances in the back/leg/buttocks e.g. tingling, 'pins and needles', numbness
 - ₃ none of the above

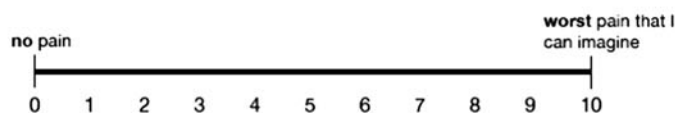
- 2 For the following 2 questions (2a and 2b) we would like you to indicate the severity of your pain, by marking a cross on the line from 0 to 10 (where "0"=no pain, "10"=the worst pain you can imagine). There are separate questions for back pain and for leg pain (sciatica)/buttock pain.



- 2a How severe was your **back pain** in the last week?



- 2b How severe was your **leg pain (sciatica)/buttock pain** in the last week?



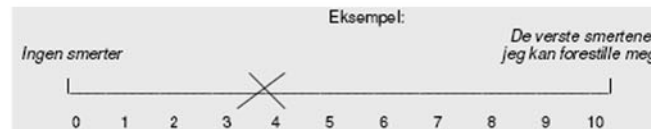
- 3 During the **past week**, how much did your back problem **interfere with your normal work** (including both work outside the home and housework)?
- ₁ not at all
 - ₁ a little bit
 - ₂ moderately
 - ₃ quite a bit
 - ₄ extremely
- 4 If you had to spend **the rest of your life with the symptoms you have right now**, how would you feel about it?
- ₁ very satisfied
 - ₁ somewhat satisfied
 - ₂ neither satisfied nor dissatisfied
 - ₃ somewhat dissatisfied
 - ₄ very dissatisfied
- 5 Please reflect **on the last week**. How would you rate your quality of life?
- ₁ very good
 - ₁ good
 - ₂ moderate
 - ₃ bad
 - ₄ very bad
- 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 back problem?
- ₁ none
 - ₁ between 1 and 7 days
 - ₂ between 8 and 14 days
 - ₃ between 15 and 21 days
 - ₄ more than 22 days
- 7 **During the past 4 weeks**, how many days did your back problem **keep you from going to work** (job, school, housework)?
- ₁ none
 - ₁ between 1 and 7 days
 - ₂ between 8 and 14 days
 - ₃ between 15 and 21 days
 - ₄ more than 22 days

Ryggproblemer kan forårsake smerter i ryggen og/eller i beina/baken, samt forårsake sanseforstyrrelser som kribling, nummenhet og prikking og stikking i disse områdene.

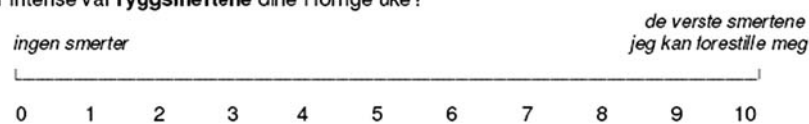
1 Hvilket av følgende problemer plager deg **mest**?
Velg bare **ett** av alternativene.

- ₁ Ryggsmerter
₂ Smerter i beina/baken
₃ Sanseforstyrrelser i ryggen/beina/baken som f.eks. kribling, nummenhet og prikking og stikking
₄ Ingen av de ovennevnte

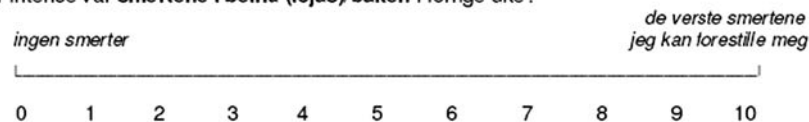
2 For de 2 følgende spørsmålene (2a og 2b) ønsker vi at du angir hvor intense smertene dine er ved å sette et kryss på linjen fra 0 til 10 (hvor "0" = ingen smerter, "10" = de verste smertene du kan tenke deg). Det er separate spørsmål for ryggsmerter og for smerter i beina (isjas)/baken.



2a Hvor intense var **ryggsmerterne** dine i forrige uke?



2b Hvor intense var **smertene i beina (isjas)/baken** i forrige uke?



3 I løpet av **siste uke**, i hvilken grad **forstyrret** ryggproblemene deg i ditt **vanlige arbeid** (omfatter både arbeid utenfor hjemmet og husarbeid) (sett kryss).

- ₁ Ikke i det hele tatt
₂ Litt
₃ I moderat grad
₄ I stor grad
₅ I svært stor grad

4 Hvis du måtte leve resten av **livet med de symptomene du har akkurat nå**, hva ville du synes om det?

- ₁ Svært fornøyd
₂ Litt fornøyd
₃ Verken fornøyd eller misfornøyd
₄ Litt misfornøyd
₅ Svært misfornøyd

5 Tenk over hvordan **siste uke** har vært.
Hvordan vil du karakterisere din livskvalitet?

- ₁ Svært god
₂ God
₃ Moderat
₄ Dårlig
₅ Svært dårlig

6 I løpet av **de siste 4 ukene**, hvor mange dager **gjorde du mindre enn du pleier** (jobb, husarbeid, skole, fritidsaktiviteter) på grunn av ryggproblemene?

- ₁ Ingen
₂ Mellom 1 og 7 dager
₃ Mellom 8 og 14 dager
₄ Mellom 15 og 22 dager
₅ Mer enn 22 dager

7 I løpet av **de siste 4 ukene**, hvor mange dager **hindret ryggproblemene deg fra** å arbeide (jobb, skole, husarbeid)

- ₁ Ingen
₂ Mellom 1 og 7 dager
₃ Mellom 8 og 14 dager
₄ Mellom 15 og 22 dager
₅ Mer enn 22 dager

References

- The EuroQol Group (1990) EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 16:199–208
- Altman DG (1991) *Practical statistics for medical research*. Chapman and Hall, London
- Andresen EM (2000) Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 81(12 Suppl 2):S15–S20
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186–3191
- Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 25:3100–3103
- Broadbent E, Petrie KJ, Main J, Weinman J (2006) The brief illness perception questionnaire. *J Psychosom Res* 60:631–637
- Damasceno LH, Rocha PA, Barbosa ES, Barros CA, Canto FT, Defino HL, Mannion AF (2011) Cross-cultural adaptation and assessment of the reliability and validity of the Core Outcome Measures Index (COMI) for the Brazilian–Portuguese language. *Eur Spine J* (Epub ahead of print)
- Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82:8–24
- Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L (1974) The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci* 19:1–15
- Deyo RA, Battie M, Beurskens AJ, 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* 23:2003–2013
- Ferrer M, Pellissier 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* 31:1372–1379
- Genevay S, Cedraschi C, Marty M, Rozenberg S, De GP, Faundez A, Balague F, Porchet F, Mannion AF (2011) Reliability and validity of the cross-culturally adapted French version of the Core Outcome Measures Index (COMI) in patients with low back pain. *Eur Spine J* 21(1):130–137
- Grotle M, Brox JI, Vollestad NK (2005) Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine Phila Pa* 30:130–140
- Kessler JT, Melloh M, Zweig T, Aghayev E, Roder C (2010) Development of a documentation instrument for the conservative treatment of spinal disorders in the International Spine Registry, Spine Tango. *Eur Spine J* 20:369–379
- Kleinstueck FS, Fekete T, Jeszenszky D, Mannion AF, Grob D, Lattig F, Mutter U, Porchet F (2011) The outcome of decompression surgery for lumbar herniated disc is influenced by the level of concomitant preoperative low back pain. *Eur Spine J* 20:1166–1173
- Lattig F, Grob D, Kleinstueck FS, Porchet F, Jeszenszky D, Bartanusz V, O’Riordan D, Mannion AF (2009) Ratings of global outcome at the first post-operative assessment after spinal surgery: how often do the surgeon and patient agree? *Eur Spine J* 18(Suppl 3):386–394
- Lurie J (2000) A review of generic health status measures in patients with low back pain. *Spine* 25:3125–3129
- Mannion AF, Boneschi M, Teli M, Luca A, Zaina F, Negrini S, Schulz PJ (2011) Reliability and validity of the cross-culturally adapted Italian version of the Core Outcome Measures Index. *Eur Spine J* (Epub ahead of print)
- Mannion AF, Elfering A, Staerckle R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, Boos N (2005) Outcome assessment in low back pain: how low can you go? *Eur Spine J* 14:1014–1026
- Mannion AF, Porchet F, Kleinstueck FS, 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
- Mannion AF, Porchet F, Kleinstueck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient’s perspective. Part 1: the Core Outcome Measures Index in clinical practice. *Eur Spine J* 18(Suppl 3):367–373
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 19:539–549
- Nakash RA, Hutton JL, Jorstad-Stein EC, Gates S, Lamb SE (2006) Maximising response to postal questionnaires—a systematic review of randomised trials in health research. *BMC Med Res Methodol* 6:5
- Roland M, Fairbank J (2000) The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. *Spine* 25:3115–3124
- Sobottke R, Rollinghoff M, Zarghooni K, Zarghooni K, Schluter-Brust K, Delank KS, Seifert H, Zweig T, Eysel P (2010) Spondylodiscitis in the elderly patient: clinical mid-term results and quality of life. *Arch Orthop Trauma Surg* 130:1083–1091
- Staerckle RF, Villiger P (2011) Simple questionnaire for assessing core outcomes in inguinal hernia repair. *Br J Surg* 98:148–155
- Streiner DL, Norman GR (1995) *Health measurement scales: a practical guide to their development and use*. Oxford University Press, Oxford
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60:34–42
- Waddell G (1998) *The back pain revolution*. Churchill Livingstone, London
- White P, Lewith G, Prescott P (2004) The core outcomes for neck pain: validation of a new outcome measure. *Spine* 29(Phila Pa 17):1923–1930
- Zweig T, Mannion AF, Grob D, Melloh M, Munting E, Tuschel A, Aebi M, Roder C (2009) How to Tango: a manual for implementing Spine Tango. *Eur Spine J* 18(Suppl 3):312–320