

## Reliability and validity of the cross-culturally adapted Italian version of the Core Outcome Measures Index

A. F. Mannion · M. Boneschi · M. Teli ·  
A. Luca · F. Zaina · S. Negrini · P. J. Schulz

Received: 9 November 2010 / Revised: 7 February 2011 / Accepted: 20 February 2011 / Published online: 17 March 2011  
© Springer-Verlag 2011

**Abstract** Patient-orientated outcome questionnaires are essential for the assessment of treatment success in spine care. Standardisation of the instruments used is necessary for comparison across studies and in registries. The Core Outcome Measures Index (COMI) is a short, multidimensional outcome instrument validated for patients with spinal disorders and is the recommended outcome instrument in the Spine Society of Europe Spine Tango Registry; currently, no validated Italian version exists. A cross-cultural adaptation of the COMI into Italian was carried out using established guidelines. 96 outpatients with chronic back problems (>3 months) were recruited from five practices in Switzerland and Italy. They completed the newly translated COMI, the Roland Morris disability (RM), adjectival pain rating, WHO Quality of Life (WHOQoL), EuroQoL-5D, and EuroQoL-VAS scales. Reproducibility was assessed in a subgroup of 63 patients who returned a second questionnaire

within 1 month and indicated no change in back status on a 5-point Likert-scale transition question. The COMI scores displayed no floor or ceiling effects. On re-test, the responses for each individual domain of the COMI were within one category in 100% patients for “function”, 92% for “symptom-specific well-being”, 100% for “general quality of life”, 90% for “social disability”, and 98% for “work disability”. The intraclass correlation coefficients (ICC<sub>2,1</sub>) for the COMI back and leg pain items were 0.78 and 0.82, respectively, and for the COMI summary index, 0.92 (95% CI 0.86–0.95); this compared well with 0.84 for RM, 0.87 for WHOQoL, 0.79 for EQ-5D, and 0.77 for EQ-VAS. The standard error of measurement (SEM) for COMI was 0.54 points, giving a “minimum detectable change” for the COMI of 1.5 points. The scores for most of the individual COMI domains and the COMI summary index correlated to the expected extent (0.4–0.8) with the corresponding full-length reference questionnaires ( $r = 0.45–0.72$ ). The reproducibility of the Italian version of the COMI was comparable to that published for the German and Spanish versions. The COMI scores correlated in the expected manner with existing but considerably longer questionnaires suggesting adequate convergent validity for the COMI. The Italian COMI represents a practical, reliable, and valid tool for use with Italian-speaking patients and will be of value for international studies and surgical registries.

---

A. F. Mannion (✉)  
Spine Center Division, Department of Research  
and Development, Schulthess Klinik, Lengghalde 2,  
8008 Zurich, Switzerland  
e-mail: anne.mannion@kws.ch

M. Boneschi · P. J. Schulz  
Institute of Communication and Health,  
Università della Svizzera Italiana, Lugano, Switzerland

M. Teli  
Department of Spinal Surgery, Galeazzi Orthopaedic Institute,  
Via Galeazzi 4, 20161 Milan, Italy

A. Luca  
Spine Center Division, Schulthess Klinik, Lengghalde 2,  
8008 Zurich, Switzerland

F. Zaina · S. Negrini  
Italian Scientific Spine Institute (ISICO), Milan, Italy

**Keywords** Back pain · Outcome questionnaire ·  
Cross-cultural adaptation · Reliability · Validity

### Introduction

In the last two decades, outcome assessment in musculo-skeletal medicine has undergone something of a paradigm

shift, moving away from imaging and objective indices of function and towards patient self-rated evaluation [4]. In order to promote larger, multinational studies and encourage the use of international registries, it is essential that valid instruments are available in a range of different languages. This also facilitates the standardisation and pooling of data when performing meta-analyses of the results of research carried out in different countries [6].

The Core Outcome Measures Index (COMI) comprises a short set of questions used to assess the impact of spinal disorders on multiple patient-orientated outcome domains. It is based on a set of individual items selected from established questionnaires and recommended for standardised use by an international group of experts in the field [9]. With slight modifications, the set of questions was adapted to produce an outcome instrument in the German language [16, 17] and in Spanish [10] for use in patients with back problems, and in the English language for patients with neck pain [32]. These studies revealed that the COMI was a reliable, valid, and responsive instrument, showing comparable psychometric properties in the different language versions [10, 16, 17, 32]. This, coupled with its brevity, makes it appealing for use in large-scale international investigations where maximum participation is desired. The instrument is gaining increasing popularity within the scientific community, being developed in other languages [34] and adapted for different medical conditions [29], and its use is foreseen in Registries of surgical and conservative spinal treatment throughout Europe and the rest of the world [15, 20, 26, 34].

The aims of the present study were to carry out a cross-cultural adaptation of the COMI for use with Italian-speaking patients and to investigate its psychometric properties in a group of patients presenting with chronic low back pain at rheumatology and orthopaedic practices within the Italian-speaking region of Switzerland and in Italy.

## Materials and methods

### The Core Outcome Measures Index

The COMI is a self-administered multidimensional instrument that consists of seven items to assess the extent of the patient's back pain and leg pain, difficulties with functioning in everyday life, symptom-specific well-being, general quality of life, and social and work disability (Appendix 1). The questionnaire is completed in reference to the patient's status "in the last week" for all but the two

disability items (which instead refer to the last 4 weeks). Leg pain and back pain are assessed on 0–10 graphic rating scales and all other items on 5-point adjectival scales. In each case, a higher score indicates worse status. Scores for each domain and a summary index score are calculated. For the latter, the "worst pain" score is firstly taken, as the higher of the two pain scale scores (back and leg). For the other items, each incremental "step" is given 2.5 points so that they range from 0 (best status) to 10 (worst status), analogous to the pain scale. The scores for social disability and work disability are averaged to form one disability score. A summary index score from 0 (best health status) to 10 (worst health status) can then be computed by averaging the values for the five subscales (worst pain, function, symptom-specific well-being, general quality of life, and disability) [16, 17].

### Translation and cross-cultural adaptation

The translation and cross-cultural adaptation of the original English version of the COMI into Italian was carried out in accordance with previously published guidelines [2, 11]. These guidelines describe the process currently recommended by the American Academy of Orthopaedic Surgeons (AAOS) Outcomes Committee.

### Translation and synthesis

Two native Italian speakers (T-1, T-2) carried out independent translations of the COMI from English to Italian. T-1 was familiar with the concepts being examined and the clinical content of the questionnaires. T-2 was a layperson who was not familiar with the specific concept being investigated (the "naïve" translator). The different profiles of the two translators assured good agreement and accuracy with the original English version in terms of both the clinical content and the appropriateness of the terminology. The two translations were compared with one another and with the original English version. After discussing any discrepancies that had arisen, a consensus was finally reached, and the two versions were synthesised to form one common Italian version, T-12.

### Back-translation

Two native English speakers with Italian as their second language (BT-1, BT-2) carried out a back-translation of the Italian version (T-12) into English. Neither of the back-translators was familiar with the subject matter of the questionnaire; both were blind to the English original, and each carried out their translation independently. A third

person (native English with a knowledge of Italian) compared the two back-translations with each other and with the original-questionnaire and highlighted any conceptual errors or gross inconsistencies in the content of the translated versions, in preparation for the expert committee meeting.

#### *Expert committee*

An expert committee was formed consisting of both translators, one of the back-translators, one Italian-speaking outcomes research assistant, one bilingual clinician (rheumatologist), and one native English clinical research scientist. The group examined the translations, the back-translations, and the notes made in carrying out/comparing the translations, and consolidated these to produce a “pre-final” version of the Italian COMI. The task of this expert committee was to assure semantic and idiomatic equivalence (i.e. to check for ambiguous words or inappropriately translated colloquialisms) and experiential and conceptual equivalence (i.e. to address any peculiarities specific to the cultures examined) between the Italian and English versions of the questionnaire. For all parts of the questionnaire (instructions, items, and response options) consensus was eventually found between the members of the committee. All stages of the translation process, and any discrepancies, problems, or difficulties encountered, were documented in written form.

#### *Test of the pre-final version*

The questionnaire was given to ten Italian-speaking people (back patients and friends/colleagues) as a test of the pre-final version. They were probed regarding their general comments on the questionnaire (layout, wording, ambiguities, ease of understanding, etc.). The findings from this phase of the adaptation process (face validity of the questionnaire) were evaluated before the final Italian version of the COMI was produced and subject to further psychometric testing.

#### Assessment of the psychometric properties of the Italian version of the COMI

##### *Questionnaire battery*

Patients were asked to complete a questionnaire booklet, which contained the Italian version of the COMI and additional questionnaires intended to assess the COMI's construct validity (convergent and divergent; see later). The full-length scales used for comparison were, as far as

available in Italian, the same as those used in the original COMI validation study [16] and comprised: (1) pain intensity in the last week, rated on a 5-point verbal rating (adjectival) scale (no pain, a little, moderate, severe, extreme pain) pain [12]; (2) the Italian version [23] of the Roland Morris (RM) disability questionnaire [27], which enquires as to whether back pain hinders the performance of 24 activities of daily living (today), with possible responses of “yes” and “no” (scored 0–24 points); (3) the Italian version [8] of the World Health Organisation Quality of Life Questionnaire (brief version) WHOQOL-BREF [33]. The WHOQOL-BREF consists of 26 items measuring four domains considered to contribute to overall quality of life: psychological, physical, social, and environmental well-being. Each domain is scored 4 (best status) to 20 (worst status); (4) the Euroqol-Five Dimensions (EQ-5D) and the Euroqol-“visual analogue scale” (EQ-VAS) for overall health state [this was used instead of the “Psychological General Well-Being Index” (PGWB) which was used in the original COMI validation study [16] since the PGWB was not available in Italian]. The EQ is a standardised instrument for use as a measure of health outcome; it is applicable to a wide range of health conditions and treatments [5, 25] and has been validated in Italian [28]. It comprises five single items—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression—each rated with a three-point adjectival scale, and a 0–100 scale commonly referred to as a “visual analogue scale” (but numbered and presented as a vertical scale) for ‘overall health state’. Summary index scores (ranging from –0.59 to 1) were computed using the unweighted method described by Prieto and Sacristán [24].

Additional questions concerned sociodemographic and pain-related variables: age, gender, educational level, work status, work heaviness, sick leave, duration of current episode, and length of current sick leave.

##### *Patients*

Ninety-six patients with chronic LBP (>3 months) were recruited from five practices in the Italian-speaking part of Switzerland (rheumatology and manual medicine practices) and in Italy (an orthopaedic practice). Inclusion criteria were: non-specific low back pain or a low back problem due to disc herniation, spinal stenosis, or spinal deformity causing back pain or referred pain for more than 3 months, and ability to understand written Italian. Exclusion criteria were: low back pain due to fracture, cancer, infection, or inflammatory diseases. Patients were recruited from the rheumatology/manual medicine

practices upon attendance for consultation. Patients from the orthopaedic practice were recruited by a consultant spinal surgeon following selection, from his own database, of surgical and non-surgical cases meeting the inclusion criteria. After providing their informed consent, patients were asked to complete the questionnaire booklet and return it to the study administration office. Once the completed questionnaire was received back at the office, the patient was sent out a second booklet to be completed and again posted back as soon as possible. The second booklet also contained a transition question evaluating any perceived change in back status since the first booklet (5-point Likert scale: better, a little better, no change, a little worse, worse) [3]. Of the 96 patients recruited, 93 (97%) returned a second questionnaire, 86 of them within 1 month of the first (which in the present study was considered the maximum acceptable interval for test–retest analysis). Of these 86 patients, 63 reported no change in their back pain status. Hence, the data of 96 patients (see Table 1 for patient characteristics) were used for the analyses of floor/ceiling effects and construct validity, and the data of 63 patients [38 women, 25 men; mean (SD) age 55 (14) years] were used for the assessment of questionnaire reproducibility.

The study was approved by the corresponding Ethics committees of the Swiss and Italian institutions.

### Statistical analysis

Scores for each instrument were calculated as per their authors' instructions and applying the following rules for missing data: no missings were allowed for COMI or EQ-5D since these have just one item per domain; for the WHOQoL, a minimum of 80% answers were required for each domain/questionnaire [33] and for the Roland Morris, similarly 80% (Elfering, personal communication).

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. Floor/ceiling effects >70% are considered to be adverse [14] and <15–20%, ideal [1, 19]. 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 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 [31]. One

**Table 1** Patient characteristics

Total number	96
Sex (male/female)	37/59
Age, mean $\pm$ SD (range)	55.1 $\pm$ 15.2 (21–91)
LBP before this episode	
Yes	63 (66%)
No	32 (33%)
Missing	1 (1%)
Duration of current episode	
3–6 months	37 (39%)
>6 and <18 months	26 (27%)
>18 months	27 (28%)
Missing	6 (6%)
Normal work	
Retired	30 (31%)
No paid work	6 (6%)
On benefits	10 (11%)
Employed	46 (48%)
Unemployed	2 (2%)
Missing	2 (2%)
Length of current sick leave	
Not applicable	18 (19%)
Not on sick leave	38 (40%)
<7 weeks	9 (9%)
7 weeks–3 months	3 (3%)
>3 and <6 months	3 (3%)
>6 and <18 months	9 (9%)
>18 months	3 (3%)
Missing	13 (14%)
Educational level	
Obligatory	6 (6%)
Secondary education	28 (29%)
University education	45 (47%)
Higher degree	16 (17%)
Missing	1 (1%)
Type of work	
Sedentary	33 (34%)
Physical	34 (36%)
Mixture of sedentary and physical	27 (28%)
Missing	2 (2%)

type of construct validity, convergent validity, requires that different measures of the same or similar construct agree to an acceptable extent [1], and in the present study, this was evaluated using Spearman Rank correlation coefficients corrected for ties. It was hypothesised (based on the validation studies for the original COMI and as recommended by Streiner and Norman [30] for measures of the same/similar attributes) 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 (listed in Table 3) and between the COMI summary index score and RM, WHOQOL-physical and EQ-5D summary index scores. As a measure of divergent validity, correlations  $<0.4$  were expected for the COMI summary index score and the social, environmental, and psychological items of the WHOQOL.

Reproducibility indicates the extent to which the same results are obtained on repeated administration of the given instrument when no change is expected. For the COMI 5-point ordinal scales, reproducibility (stability) of measures was assessed by examining the proportion of participants recording test–retest differences for each item within a reference value of  $\pm 1$  point (where at least 90% was considered acceptable) [21, 29]. For scales/items yielding approximately normally distributed values (pain scales, COMI summary score, Roland Morris), one-way repeated measures ANOVA was used to assess the differences in means for the repeated trials and to determine the intraclass correlation coefficient (ICC; model ICC<sub>agreement 2,1</sub>) and their 95% confidence intervals. ICCs can range from 0 to 1; greater than 0.7 in groups of at least 50 patients are generally considered to indicate acceptable reliability [31]. Standard errors of measurement SEM<sub>agreement</sub> were used to indicate the absolute measurement error (“agreement” [31]) and to calculate the minimum detectable change (MDC 95%) for the instruments, i.e. the degree of change required in an individual’s score in order to establish it (with a given level of confidence) as being a real change, over and above measurement error. At the 95% confidence level, this is defined as  $1.96 \times 2 \times \text{SEM}$  which is equivalent to  $2.77 \times \text{SEM}$ . The ICCs and SEMs were determined for all scales in order to provide some perspective for interpreting the corresponding values for the COMI itself.

## Results

### Cross-cultural adaptation of the COMI

The Italian version of the COMI is presented in Appendix 2. Few difficulties arose during its adaptation: (a) Translation of “how many days...*cut down on* the things you usually do” (social disability). At first, the word “*rinunciare*” was chosen in the consensus Italian version, but the English back translation revealed this to be closer to “avoid” or “renounce doing” something, rather than just “cutting down/reducing”. After discussion, this was changed to “... ridurre le sue attività abituali”; (b) Translation of “how many days...*keep you from going to*

*work...*” (work disability). At first, this was translated as “...*non ha potuto svolgere la sua attività lavorativa...*” in the consensus Italian version, and the English back translation suggested “*impossible to do your work*” which did not focus sufficiently on the notion of failing *to go to* work, i.e., taking days off. After discussion, this was changed to “...*ha impedito di andare al lavoro*”.

Upon conclusion of the main validation study, a slight clarification to the wording of the “function” item was made because there had been some question as to whether the original Italian translation for “housework” had for some people implied only the kind of work that a professional can do at home (e.g., consultancy, computer programming, etc.) as opposed to work around the house [cleaning, DIY (“do it yourself”), cooking, washing, etc.], which was the intended meaning. This was hence clarified by replacing the initial wording in brackets at the end of the item, “*considerando sia il lavoro fuori casa che quello in casa*”, with “*come il lavoro fuori casa e/o le faccende domestiche*”.

### Missing data

Data were generally very complete for the 96 questionnaires: there were missing answers for 1–14% of the demographic/pain history items (see Table 1), 1 (1%) patient for each of the EQ-5D items (and the summary index score and VAS general health status) and in up to 3 (3%) patients for the individual COMI items and COMI summary score. For the Roland Morris, three patients (3%) had too many missing answers to allow valid calculation of a score and for the WHOQOL, missing items led to missing domain scores ranging from 1 (1%) for WHOQOL-physical up to 11 (11%) for WHOQOL-social.

### Floor and ceiling effects

The floor effects (worst status) and ceiling effects (best status) for each of the questionnaire items/scales are shown in Table 2.

Minimal floor effects were found for the COMI items pain, function, and quality of life (0–2%), but higher values were found for symptom specific well-being, and social and work disability (15–24%). A low ceiling effect (0–6.3%) was found for most of the individual COMI items; however, ceiling effects were 19% for leg pain, 34% for social disability, and 56% for work disability. The EQ-5D items showed generally low floor effects (0–4%) except for pain (10%), but ceiling effects were high (30–67%) for all domains other than pain (8%).

**Table 2** Floor and ceiling effects for all the instruments in the 96 patients completing the first evaluation

Instrument	Floor effects (worst status) (%)	Ceiling effects (best status) (%)
COMI LBP	0	1.1
COMI LP	1.1	19.1
COMI worst pain (leg or back)	1.1	0
COMI function	2.1	6.3
COMI symptom-specific well-being	24.0	0
COMI quality of life	1.0	2.1
COMI social disability	20.0	33.7
COMI work disability	15.1	55.9
<i>COMI summary score</i>	<i>0</i>	<i>0</i>
<i>Roland Morris score</i>	<i>2.2</i>	<i>3.2</i>
EQ-5D mobility	0	46.9
EQ-5D self-care	0	66.7
EQ-5D usual activities	4.2	30.2
EQ-5D pain	10.4	8.3
EQ-5D anxiety/depression	1.1	56.8
<i>EQ-5D summary index score</i>	<i>0</i>	<i>6.3</i>
EQ-5D VAS general health	0	0
<i>WHOQoL physical</i>	<i>0</i>	<i>0</i>
<i>WHOQoL psychological</i>	<i>0</i>	<i>0</i>
<i>WHOQoL social</i>	<i>0</i>	<i>1.2</i>
<i>WHOQoL environmental</i>	<i>0</i>	<i>0</i>
<i>WHOQoL whole score</i>	<i>0</i>	<i>0</i>

Italicised rows indicate scores from scales with more than one item

Considering the multiple-item questionnaires, there were minimal floor effects (0–2%) for the COMI summary score, the Roland Morris disability score, and the domains/whole score of the WHOQoL; ceiling effects for these scales were similarly low (0–3%) for all except the EQ-5D summary index score (6.3%).

### Construct validity

The correlation coefficients for the relationship between the scores for each item of the COMI and its corresponding full-length questionnaire are shown in Table 3.

All but one of the hypotheses concerning the convergent validity of the COMI items (coefficients 0.4–0.8 with the corresponding full instruments) could be confirmed. A good correlation was found between the COMI worst pain score and the adjectival pain scale scores ( $\rho = 0.67$ ). Correlations of 0.54–0.66 were found between the COMI function item scores and the full-length function/disability scales (RM and WHOQOL physical). The scores for COMI

symptom-specific well-being showed a correlation of  $-0.45$  with the WHOQOL physical scale scores, but their correlation with the WHOQOL-BREF whole scores was just  $-0.35$ . COMI general quality of life showed correlations of 0.52–0.63 with the global quality of life scale. There was a correlation of 0.60 between COMI disability and the RM and WHOQOL physical. The correlation between the summary index score of the COMI and the each of the full instrument whole scores was 0.63–0.72. Indicating reasonable divergent validity, correlations  $\leq 0.4$  were found for the COMI summary index score and the social, environmental, and psychological items of the WHOQOL.

### Reproducibility

The mean duration between the first and the second questionnaire was 10.4 (SD 6) days.

Differences in response to each domain on the COMI were  $\pm 1$  category in 100% patients for the domain ‘function’, 92% for ‘symptom-specific well-being’, 100% for ‘general quality of life’, 90% for ‘social disability’, and 98% for ‘work disability’, hence all satisfying the stability criterion of  $\geq 90\%$  suggested by Nevill et al. [21].

Table 4 shows the mean (SD) scores on the two test occasions, the ICC and SEMs for each of the scales.

There was no systematic bias (i.e. significant difference in mean scores from test to re-test) in the scores for the COMI summary index although the COMI back pain and worst pain items showed slightly but significantly lower values at the second assessment, as did the Roland Morris score (Table 4). The ICCs for COMI pain and COMI summary index scores were 0.78–0.92; this compared favourably with the corresponding values for the full-length scales (0.67–0.88) (Table 4). The SEM and MDC 95% values for each of the scales are also shown in Table 4. The SEM for the COMI summary index score was 0.54 and the MDC 95%, 1.5 points. Expressed as a percentage of the maximum score range for the given scale, the SEMs were similar for all scales, being approximately 5–12%.

### Discussion

The present study aimed to produce an Italian version of the COMI that would be valid and reliable for Italian-speaking patients with back problems. The process of translating and back-translating the COMI was carried out in accordance with established guidelines [2, 11] in an attempt to produce an adaptation of the questionnaire that

**Table 3** Correlations between COMI domain single items and full-length reference questionnaires in the 96 patients completing the first evaluation

Core index items	Reference scales	$\rho$
Convergent validity		
Pain symptoms	Pain verbal rating scale	0.67
Back function	Roland and Morris	0.55
	WHOQOL-BREF physical health	-0.66
Symptom-specific well-being	WHOQOL-BREF physical health	-0.45
	WHOQOL-BREF whole score	<b>-0.35</b>
Quality of life	EQ-5D summary index	-0.63
	WHOQOL-BREF whole score	-0.52
Disability	Roland and Morris	0.60
	WHOQOL-BREF physical health	-0.60
COMI summary score <sup>a</sup>	Roland and Morris	0.63
	WHOQOL-BREF physical health	-0.72
	EQ-5D summary index	-0.67
Divergent validity		
COMI summary score <sup>a</sup>	WHOQOL-BREF social	-0.26
	WHOQOL-BREF environmental	-0.35
	WHOQOL-BREF psychological	<b>-0.40</b>

$\rho$  values in bold italics indicate those where the pre-defined hypothesis for the extent of the correlation could not be confirmed

<sup>a</sup> The summary score comprised the scores for five items: pain (worst, back or leg), back function, symptom-specific well-being, quality of life, and disability (average of social and work disability)

**Table 4** Test–retest reliability results for each of the domain index-items and the full reference scales for 63 patients who returned their questionnaire within 1 month and reported no change in their back problem since the first questionnaire

Instrument	No of items	Range	M1	M2	P	ICC	95% CI <sub>ICC</sub>	SEM	SEM%	MDC 95%
COMI summary index score	5	0 to 10	4.6 (1.9)	4.5 (1.9)	0.053	0.92	0.86–0.95	0.54	5.4	1.51
COMI back pain	1	0 to 10	5.0 (2.2)	4.5 (2.3)	0.005	0.78	0.64–0.87	1.07	10.7	2.95
COMI leg pain	1	0 to 10	3.7 (2.9)	3.8 (2.7)	0.70	0.82	0.71–0.89	1.20	12.0	3.32
COMI worst pain	1	0 to 10	5.5 (2.1)	5.0 (2.1)	0.002	0.82	0.69–0.89	0.93	9.3	2.58
Roland Morris Disability	24	0 to 24	10.5 (6.3)	9.1 (6.1)	0.002	0.84	0.72–0.91	2.49	10.4	6.90
EQ VAS general health	1	0 to 100	63.5 (18.2)	61.4 (16.8)	0.15	0.77	0.65–0.86	8.35	8.4	23.1
EQ-5D summary index	5	-0.59 to 1.0	0.56 (0.27)	0.56 (0.26)	0.99	0.79	0.67–0.87	0.12	7.7	0.33
WHOQOL-BREF physical health	7	4 to 20	12.7 (2.6)	12.8 (2.3)	0.88	0.88	0.80–0.92	0.86	5.3	2.37
WHOQOL-BREF psychological	6	4 to 20	13.7 (2.6)	13.6 (2.4)	0.36	0.88	0.81–0.93	0.87	5.5	2.42
WHOQOL-BREF social relationships	3	4 to 20	13.5 (2.8)	13.5 (2.5)	0.94	0.67	0.50–0.79	1.53	9.6	4.25
WHOQOL-BREF environment	8	4 to 20	13.6 (2.5)	13.3 (2.2)	0.05	0.84	0.75–0.90	0.96	6.0	2.65
WHOQOL-BREF whole	26	4 to 20	13.3 (2.1)	13.2 (1.8)	0.20	0.87	0.79–0.92	0.72	4.5	2.00

M1, M2 mean value at first and second assessment; P significance of difference between mean values on the two occasions; ICC intraclass correlation coefficient (ICC2,1); CI<sub>ICC</sub> 95% confidence intervals for the ICC; SEM standard error of measurement; SEM% SEM as percentage of maximum score; MDC 95% minimum detectable change score

would show a high degree of agreement with the original version. Overall, there were few problems translating the instrument, missing data were relatively infrequent (<3% for any given item), and the psychometric characteristics of the COMI were comparable to those reported for the Spanish [10] and German [16] versions. Just one item needed modification to clarify the notion of “housework/domestic duties” as opposed to “working at/from home”. Interestingly, the final version employed a similar

expression to that used in the COMI in Spanish, a language very close to Italian in both its vocabulary and sentence structure.

Floor and ceiling effects

For three of the individual COMI domains (symptom-specific well-being, social disability, and work disability), the percentages of patients indicating either the worst or

best possible status was greater than ideal (15–20% [1, 19]), but did not reach a level that would be considered adverse (>70%) for health-related quality of life questionnaires [14]. Further, when the domain scores were combined to form the COMI summary score index, there were no floor and ceiling effects at all. The assessment of health-related quality of life often results in skewed distributions, and when the number of response categories is low, the number of responses at the extreme of the range naturally increases (with a dichotomous item by definition having *only* ceiling and floor effects). The EQ-5D, which has just three response categories, also showed marked ceiling effects (30–67%) in the present study for four out of its five sub-domains. High floor and ceiling effects can threaten the responsiveness of an instrument since they can prevent improvement or worsening from being detected when it has indeed occurred. It might be assumed that the potential for ceiling and floor effects could be decreased and the responsiveness thereby increased by increasing the number of response options for a given item. However, an overview on this theme has reported that, first, humans are unable to discriminate much beyond seven levels, and, second, that responsiveness was quite similar between scales with 7-point response categories and those with as few as 4 points [22]. Hence, expanding the number of response categories would not necessarily make the COMI any more responsive. Interestingly, in both the previous validation studies [10, 16], the COMI was shown to be at least as responsive as other condition-specific instruments (with effect sizes >1.0) and even the individual items had moderate to large effect sizes of 0.52–0.84. Hence, it would appear that the higher floor and ceiling effects are not so problematic in practice.

#### Construct validity

As with the previous validation studies in other languages, each of the individual core items of the Italian COMI was examined in relation to a multi-item questionnaire established as being valid and reliable in the Italian language and addressing the same or a similar domain. In the context of validity, it has been suggested that any measurement will have some associated error, and as a result, correlations among measures of the same attribute should fall in the midrange of 0.4–0.8; if coefficients are any lower than 0.4, it must be assumed that either the reliability of one or the other measure is unacceptably low or that they are measuring different phenomena [30]. In keeping with the findings for both the Spanish and German versions of the COMI, only the symptom-specific well-being item failed to show a suitably high correlation with the full-length

questionnaires. Since the reliability of the item itself was good (with 92% responses  $\pm 1$  category on the two test occasions), we concur with previous authors that this item is likely delivering unique information, dissimilar to that of any other aspect of quality of life [16]. For all other individual COMI items and for the COMI summary index, the expected level of correlation with the longer instruments was achieved (with coefficients of 0.52–0.72), confirming our pre-defined hypotheses and concurring with the findings for the German ( $r = 0.68$ – $0.79$ ) [16] and Spanish ( $r = 0.67$ – $0.84$ ) [10] versions of the COMI-back and the English version of the COMI-neck ( $r = 0.48$ – $0.63$ ) [32].

#### Reproducibility

The test–retest reliability of the COMI was considered good, with intraclass correlation coefficients (ICCs) for the individual pain scales being 0.78 and 0.82, and with an ICC for the COMI summary index score of 0.92. These ICCs were similar to those previously reported for the COMI [10, 16], and they compared well with those for the longer instruments evaluated (0.77–0.84). The “minimum detectable change” (MDC 95%) for the COMI summary index score was 1.5 points, which is similar to the 1.7 points previously reported for the German version [16]. This value represents the minimum difference in an individual’s score required to state with 95% confidence that “real change” is responsible for the difference, as opposed to just measurement error (“noise” in the system). Expressed as a percent of the full-scale range (maximum value, 10 points), at 15% this is at the more favourable end of the range of values reported for other LBP outcome instruments [7]. The minimal clinically important difference (MCID) for the COMI is 2–3 points, depending on the external criterion used [16, 18]. If a similar MCID exists for the Italian version too, as is suggested by initial evaluations of the data collected in connection with the Spine Surgery registry, Spine Tango (unpublished data), then a clinically relevant change of 2–3 points (the “signal”) would far exceed the minimum detectable change of 1.5 points (the “noise”), confirming its suitability as a LBP outcome instrument [13].

The test–retest reliability or “stability” of the individual adjectival scale COMI items (function, symptom-specific well-being, quality of life and disability) was assessed using the simple but sensitive method recommended by Nevill et al. [21] for such 5-point psychometric scales, in which within-individual differences in responses are calculated. These authors recommend that, when assessing the stability of self-report questionnaires with 5-point scales, most participants (90%) should record test  $\pm$  retest differences within a reference value of  $\pm 1$ . In the present



study, this was achieved by 90–100% patients for the individual COMI items. In summary, adequate reliability was shown for both the individual items and the COMI index summary score.

#### Limitations of the study

Some limitations of the study are worthy of mention. The instruments were completed by patients living in different Italian-speaking geographical regions: South of Switzerland, North of Italy and Central/Southern Italy. There are no notable differences in the healthcare systems of Italy and Switzerland that should have biased the data, but people from these different areas use different Italian dialects in their daily language, which could potentially influence their interpretation or understanding of the questions. However, the main linguistic difference between these regions concerns their spoken language, and there are few grammatical or semantic differences in the use of the written language. In putting together this Italian version of the COMI, we used translators/back-translators from these different Italian-speaking regions and paid special attention to choosing words that were in common everyday use in all regions. Thus, we believe that the current version has wide applicability and should be easily understandable for all Italian speakers. Whilst we cannot rule out subtle differences in interpretation related to social or educational differences between Italian-speakers in Switzerland and Italy, there is no reason to believe that these would be any greater than the differences observed within a given region and across different regions in each individual country.

For logistic reasons, the method of patient selection and administration of the questionnaires differed slightly in the different practices (with patients being selected predominantly from an existing database in the orthopaedic centre and mainly upon consultation for care in the other centres). Further, most patients that were recruited from the rheumatology/manual medicine practices had mechanical non-specific LBP, whereas most of those from the orthopaedic practice were affected by specific causes of LBP for which they were undergoing or had undergone either surgical or non-surgical treatment. However, the admission criteria were identical in each case, and although the aetiology of their pain may have differed, all patients had a chronic back problem and exhibited the symptoms and functional difficulties being assessed by the questionnaires. For some patients, there was quite a long time between the two completions of the questionnaire, and the systematic changes in group mean scores for pain and Roland Morris disability suggested some improvement between the two assessments. This may have been the result of the well-

known statistical phenomenon of regression to the mean and/or may have reflected an inadvertent effect of simply seeing the doctor despite no reported change in global back status. There is no recommended best time interval to use between repeated assessments, and it is always a trade-off between minimising on the one hand recall effects and the other hand the likelihood of true change; generally, 1–2 weeks is considered appropriate [31]. We elected to use 1 month as our cut-off in order to allow for any delays in the sending and returning of questionnaires and to minimise the number of participants that would otherwise have been excluded by employing a shorter time interval. Using the transition question as well, we were able to eliminate the likelihood of including any patients with a wide variation in their back status, even if up to 1 month had passed since the first questionnaire was completed. Interestingly, further analysis using a 2-week cut-off did not eliminate the systematic change in mean scores and yielded similar reliability coefficients and SEMs. No formal assessment of the sensitivity to change or responsiveness of the Italian COMI was carried out within the confines of the present study. However, upon successful cross-cultural adaptation, the Italian COMI has been used in quality management and outcome projects in connection with the European Spine Surgery registry, Spine Tango, in two of the authors' institutions, and it will soon be implemented as the standard instrument for everyday use in another (also with non-surgical patients); hence, further data to examine its responsiveness should rapidly accumulate.

In conclusion, we have established that the Italian version of the COMI displays psychometric characteristics that are to all intents and purposes as good as those of corresponding full-length questionnaires and are comparable to those of other language versions of the instrument. We recommend the adaptation of the COMI in other languages and its continued, widespread use in multicentre studies, routine quality management and surgical registry systems. Improved documentation of spinal care in this manner should ultimately lead to an improved standard of care for the individual patient with LBP.

**Acknowledgments** The authors would like to thank Eurospine, the Spine Society of Europe, and the Schulthess Klinik, Zürich, for funding this work. We thank Elena Zaina for help with the translations, Gordana Balaban, Nik Maffioletti, Mario Bizzini and Franco Impellizzeri for their assistance in proof-reading the final Italian version, Vera Demalde for her help in collecting the data, and the secretaries of our clinical departments for assisting with questionnaire administration. We also thank the doctors who referred patients into the study: Maria Grazia Canepa, Nicola Keller, Guido Mariotti and Andrea Badaracco.

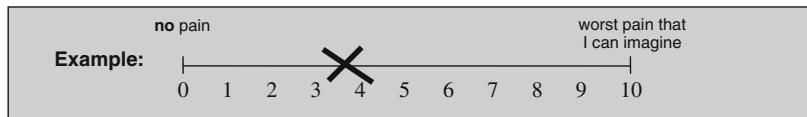
**Conflict of interest** None.

### Appendix 1: English version of the COMI

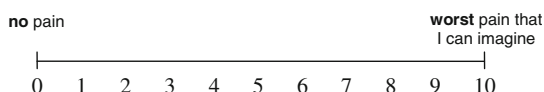
Back problems can lead to back pain and/or pain in the legs/buttocks, as well as to sensory disturbances such as tingling, 'pins and needles', or numbness in any of these regions.

For the following two questions (1a and 1b) 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).

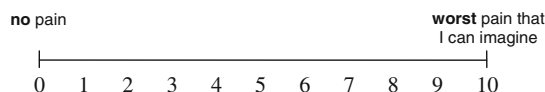
There are separate questions for **back pain** and for **leg pain (sciatica)/buttock pain**.



1a. How severe was your **back pain** in the last week?



1b. How severe was your **leg pain (sciatica)/buttock pain** in the last week?



2. During the **past week**, how much did your back problem **interfere with your normal work** (including both work outside the home and housework)?

- not at all
- a little bit
- moderately
- quite a bit
- extremely

3. If you had to spend **the rest of your life with the symptoms you have right now**, how would you feel about it?

- very satisfied
- somewhat satisfied
- neither satisfied nor dissatisfied
- somewhat dissatisfied
- very dissatisfied

4. Please reflect **on the last week**. How would you rate your quality of life?

- very good
- good
- moderate
- poor
- very poor

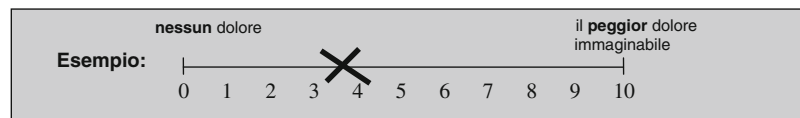
5. **During the past 4 weeks**, how many days did you **cut down on the things you usually do** (work, housework, school, recreational activities) because of your back problem?
- none
  - between 1 and 7 days
  - between 8 and 14 days
  - between 15 and 21 days
  - more than 21 days
6. **During the past 4 weeks**, how many days did your back problem **keep you from going to work** (job, school, housework)?
- none
  - between 1 and 7 days
  - between 8 and 14 days
  - between 15 and 21 days
  - more than 21 days

### Appendix 2: Italian version of the COMI

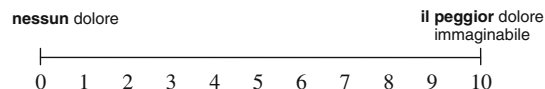
I problemi alla schiena possono portare a dolori alla zona lombare e/o alle gambe ed ai glutei, oltre che a disturbi come formicolio, puntura di aghi e spilli o riduzione della sensibilità in una di queste regioni.

Nelle prossime 2 domande (1a e 1b) indichi l'intensità del suo dolore, segnando una croce sulla linea da 0 a 10 (dove 0 = **nessun** dolore, 10 = il dolore **più forte** che lei possa immaginare).

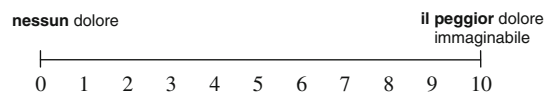
Le domande sono divise in: **mal di schiena** e **dolore alla gamba/gluteo**.



1a. Quale è stata l'intensità del **dolore alla schiena** nell'ultima settimana?



1b. Quale è stata l'intensità del **dolore alla gamba/gluteo** nell'ultima settimana?



2. **Durante la scorsa settimana** quanto i suoi problemi alla schiena hanno interferito con la sua **normale attività** (come il lavoro fuori casa e/o le faccende domestiche)?
- per nulla
  - poco
  - abbastanza
  - molto
  - moltissimo

3. Se dovesse trascorrere il **resto della vita con i disturbi che ha in questo preciso momento**, come si sentirebbe?
- molto soddisfatto
  - soddisfatto
  - né soddisfatto, né insoddisfatto
  - abbastanza insoddisfatto
  - molto insoddisfatto
4. Ripensando alla **scorsa settimana**, come giudicherebbe la sua qualità di vita?
- ottima
  - buona
  - accettabile
  - cattiva
  - pessima
5. **Nelle ultime quattro settimane**, per quanti giorni ha dovuto ridurre le sue **attività abituali** (lavoro, lavori di casa, scuola, attività del tempo libero) per colpa dei suoi problemi alla schiena?
- mai
  - da 1 a 7 giorni
  - da 8 a 14 giorni
  - da 15 a 21 giorni
  - per più di 21 giorni
6. **Durante le ultime quattro settimane**, per quanti giorni il suo problema alla schiena le **ha impedito** di andare al **lavoro** (lavoro, scuola, attività domestiche)?
- mai
  - da 1 a 7 giorni
  - da 8 a 14 giorni
  - da 15 a 21 giorni
  - per più di 21 giorni

## References

1. Andresen EM (2000) Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 81:S15–S20
2. 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
3. Beurskens AJHM, de Vet HCW, Köke AJA (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 65:71–76
4. Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 25:3100–3103
5. Brooks R (1996) EuroQol: the current state of play. *Health Policy* 37:53–72
6. Costa LO, Maher CG, Latimer J (2007) Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. *Spine* 32:1028–1037
7. Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82:8–24
8. De Girolamo G, Rucci P, Scocco P, Becchi A, Coppa F, D'Addario A, Darù E, De Leo D, Galassi L, Mangelli L, Marson C, Neri GLS (2000) Quality of life assessment: validation of the Italian version of the WHOQOL-Brief. *Epidemiol Psichiatr Soc* 9:45–55
9. Deyo RA, Battie M, Beurskens AJHM, 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
10. Ferrer M, Pellise 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* 31:1372–1379, discussion 1380
11. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 46:1417–1432

12. Haefeli M, Elfering A (2006) Pain assessment. *Eur Spine J* 15(Suppl 1):S17–S24
13. Hagg O, Fritzell P, Nordwall A, Group SLSS (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 12:12–20
14. Hyland ME (2003) A brief guide to the selection of quality of life instrument. *Health Qual Life Outcomes* 1:24
15. 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* (in press)
16. Mannion AF, Elfering A, Staerkle 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
17. Mannion AF, 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 1. The Core Outcome Measures Index (COMI) in clinical practice. *Eur Spine J* 18:367–373
18. Mannion AF, Porchet F, Kleinstück 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* 18:374–379
19. McHorney CA, Tarlov AR (1995) Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 4:293–307
20. Melloh M, Staub L, Aghayev E, Zweig T, Barz T, Theis JC, 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
21. Nevill AM, Lane AM, Kilgour LJ, Bowes N, Whyte GP (2001) Stability of psychometric questionnaires. *J Sports Sci* 19:273–278
22. Norman GR, Sloan JA, Wyrwich KW (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 41:582–592
23. Padua R, Padua L, Ceccarelli E, Romanini E, Zanolì G, Bondi R, Campi A (2002) Italian version of the Roland Disability Questionnaire, specific for low back pain: cross-cultural adaptation and validation. *Eur Spine J* 11:126–129
24. Prieto L, Sacristan 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
25. Rabin R, de Charro F (2001) EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 33:337–343
26. Roder C, Chavanne A, Mannion AF, Grob D, Aebi M, El-Kerdi A (2005) SSE Spine Tango—content, workflow, set-up. <http://www.eurospine.org-Spine> Tango. A European spine registry. *Eur Spine J* 14:920–924
27. Roland M, Morris R (1983) A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8:141–144
28. Savoia E, Fantini MP, Pandolfi PP, Dallolio L, Collina N (2006) Assessing the construct validity of the Italian version of the EQ-5D: preliminary results from a cross-sectional study in North Italy. *Health Qual Life Outcomes* 4:47
29. Staerkle RF, Villiger P (2011) Simple questionnaire for assessing core outcomes in inguinal hernia repair. *Br J Surg* 98(1):148–155
30. Streiner DL, Norman GR (1995) *Health Measurement Scales: a practical guide to their development and use*. Oxford University Press, Oxford
31. 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
32. White P, Lewith G, Prescott P (2004) The core outcomes for neck pain: validation of a new outcome measure. *Spine* 29:1923–1930
33. WHOQOL (1998) The World Health Organisation WHOQOL-BREF Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Soc Sci Med* 46:1569–1585
34. 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* 312(Suppl 3):312–320