Cross-cultural adaptation and assessment of the reliability and validity of the Core Outcome Measures Index (COMI) for the Brazilian-Portuguese language

L. H. F. Damasceno · P. A. G. Rocha ·
E. S. Barbosa · C. A. M. Barros · F. T. Canto ·
H. L. A. Defino · A. F. Mannion

Received: 14 September 2011 / Accepted: 27 November 2011 / Published online: 15 December 2011
© Springer-Verlag 2011

Abstract

Purpose The use of patient-orientated questionnaires is of utmost importance in assessing the outcome of spine surgery. Standardisation, using a common set of outcome measures, is essential to aid comparisons across studies/in registries. The Core Outcome Measures Index (COMI) is a short, multidimensional outcome instrument validated for patients with spinal disorders. This study aimed to produce a Brazilian-Portuguese version of the COMI.

Methods A cross-cultural adaptation of the COMI into Brazilian-Portuguese was carried out using established guidelines. 104 outpatients with chronic LBP (>3 months) were recruited from a Public Health Spine Medical Care Centre. They completed a questionnaire booklet containing the newly translated COMI, and other validated symptom-specific questionnaires: Oswestry Disability Index (ODI) and Roland Morris disability scale (RM), and a pain visual analogue scale. All patients completed a second questionnaire within 7–10 days to assess reproducibility.

Results The COMI summary score displayed minimal floor and ceiling effects. On re-test, the responses for each individual domain of the COMI were within 1 category in 98% patients for the domain ‘function’, 96% for ‘symptom-specific well-being’, 97% for ‘general quality of life’, 99% for ‘social disability’ and 100% for ‘work disability’. The intraclass correlation coefficients (ICC_{2,1}) for COMI pain and COMI summary scores were 0.91–0.96, which compared favourably with the corresponding values for the RM (ICC, 0.99) and ODI (ICC, 0.98). The standard error of measurement for the COMI was 0.6, giving a “minimum detectable change” (MDC95%) of approximately 1.7 points i.e., the minimum change to be considered “real change” beyond measurement error. The COMI scores correlated as hypothesised (Rho, 0.4–0.8) with the other symptom-specific questionnaires.

Conclusions The reproducibility of the Brazilian-Portuguese version of the COMI was comparable to that of other language versions. The COMI scores correlated in the expected manner with existing but longer symptom-specific questionnaires suggesting good convergent validity for the COMI. The Brazilian-Portuguese COMI represents a valuable tool for Brazilian study-centres in future multicentre clinical studies and surgical registries.

L. H. F. Damasceno · C. A. M. Barros
Santa Casa e Asilo dos Pobres, Batatais, Brazil

P. A. G. Rocha · E. S. Barbosa · C. A. M. Barros
Centro Universitário Claretiano, Batatais, Brazil

F. T. Canto · H. L. A. Defino
Ribeirão Preto School of Medicine,
University of São Paulo,
São Paulo, Brazil

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

Keywords Self-rated outcome · Spine · Back pain ·
Cross-cultural adaptation · Brazilian-Portuguese

Introduction

It is now generally accepted that the evaluation of outcome in musculoskeletal medicine should focus on patient self-rated assessment [4]. Valid instruments, available in a range of different languages, are necessary to promote multinational studies and encourage the use of international

registries. To reduce the burden to the patient, and be practical for use in busy clinical settings, such instruments should also be short, easily scored and freely available in the public domain. The use of common instruments facilitates the standardisation and pooling of data when performing meta-analyses of the results of research carried out in different countries [5].

The Core Outcome Measures Index (COMI) is a patient-orientated instrument that has been adopted as the questionnaire of choice in the Spine Society of Europe's "Spine Tango" Registries of surgical and conservative spinal treatment, now operational throughout Europe and the rest of the world [15, 21, 24, 32]. The COMI comprises a short set of questions used to assess the impact of spinal disorders on multiple patient-orientated outcome domains including pain (axial and radiating pain), function, symptom-specific well-being, and disability (social and work). It is based on a set of individual items selected from established questionnaires and recommended for use by an international group of experts in the field of low back pain [7]. With slight modifications, the set of questions was adapted to produce a back outcome instrument in the German [17, 18] and Spanish [9] languages, and has recently been successfully adapted for the Italian [16] and French [10] languages. An analogous version for patients with neck pain/neck problems has been evaluated in the English [30] and German [8] languages. All the aforementioned studies showed the COMI to be a reliable, valid and highly responsive instrument with comparable psychometric properties for its different language versions [9, 17, 18, 30]. This, coupled with its brevity, makes it appealing for use in large-scale international investigations aimed at maximum participation. The instrument is gaining increasing popularity within the scientific community, not only via its use in Registries but also recently being adapted for different medical conditions [26].

The aims of the present study were to carry out a cross-cultural adaptation of the COMI for use with Brazilian-Portuguese speaking patients and to investigate its psychometric properties in a group of patients presenting with chronic low back pain at a Public Health Spine Medical Care Center.

Materials and methods

The Core Outcome Measures Index (COMI)

The COMI is a self-administered multidimensional instrument that consists of seven questions 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.

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 five-point response 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 calculated as the higher of the two pain scale (back and leg) scores, and ranges from 0–10 points. For the other items, the five-point scale is rescaled (0 = 0 points, 1 = 2.5 points, 2 = 5.0 points, 3 = 7.5 points, 4 = 10 points) so that the score for the item ranges from 0 to 10, analogous to the pain scale. The scores for social disability and work disability are averaged to form one disability score. The COMI summary score, ranging from 0 (best health status) to 10 (worst health status), is then computed by averaging the values for the 5 subscales (worst pain, function, symptom-specific well-being, general quality of life, and disability) [17, 18].

Translation and cross-cultural adaptation

The translation and cross-cultural adaptation of the original English version of the COMI into Brazilian-Portuguese was carried out in accordance with previously published guidelines [3, 11]. These guidelines describe the process currently recommended by the American Academy of Orthopaedic Surgeons (AAOS) Outcomes Committee.

Translation and synthesis

Two native Brazilian-Portuguese speakers (T-1, T-2) with different job profiles (a spine surgeon, familiar with the concepts being examined and the clinical content of the questionnaires, and an English teacher, the "naïve" translator) carried out independent translations of the COMI from English to Brazilian-Portuguese. 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 reached and the two versions were synthesised to form one common Brazilian-Portuguese version, T-12.

Back-translation

Two native English speakers with Brazilian-Portuguese as their second-language (BT-1, BT-2) carried out a back-translation of the Brazilian-Portuguese 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 Brazilian-Portuguese) 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, 1 bilingual clinician (spine surgeon) and 1 native English clinical research scientist. The discussions took place online/through email contact. 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 Brazilian-Portuguese 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 Brazilian-Portuguese 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 Brazilian-Portuguese speaking friends/colleagues and back patients to appraise 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 Brazilian-Portuguese version of the COMI was produced and subjected to further psychometric testing.

Assessment of the psychometric properties of the Brazilian-Portuguese version of the COMI

Questionnaire battery

Patients were asked to complete a short questionnaire booklet, which contained the Brazilian-Portuguese version of the COMI and additional validated back-specific

questionnaires intended to assess the COMI’s construct validity. The instruments used for comparison measured: (1) back pain intensity in the last week, rated on a 0–10 visual analogue scale pain [12]; (2) back-related disability, assessed using the Brazilian-Portuguese version [23] of the Roland Morris (RM) disability questionnaire [25], 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) back-related pain and disability, assessed using the Brazilian-Portuguese version of the Oswestry Disability Index (ODI) [29], which measures the extent of patient’s back pain and difficulty in carrying out (today) nine different activities of daily life, with an adjusted score from 0 to 100.

Additional questions concerned sociodemographic (age, gender and work heaviness) and pain-related variables.

Patients

Patients were recruited from non-surgical candidates presenting to the Public Health Outpatients Spine Medical Care unit. All were examined by an experienced spine surgeon. Inclusion criteria were: chronic, non-specific low back pain (diagnosed according to the European Guidelines for Chronic Low Back Pain [1]) for more than 3 months and ability to understand written Brazilian-Portuguese. For the latter, patients were questioned about their general reading skills, e.g. in relation to being able to read a newspaper, since the population comprises a large number of illiterate and only partially literate people. Exclusion criteria were: low back pain due to fracture, cancer, infection or inflammatory diseases (“specific” causes of LBP according to the Guidelines [2]), and involvement in a work-compensation claim. Patients were told that their participation would not influence their current or subsequent treatment. Those that agreed to participate gave their signed informed consent. There were 29 men and 89 women, with a mean (SD) age of 42.8 (SD 14, range 19–88) years. Their jobs were rated as very strenuous, e.g. metal worker, mason, cabinet maker/woodworker ($N = 10$; 8.5%), somewhat strenuous, e.g. housekeeper, housemaid, teacher, cook, dressmaker, driver, police officer ($N = 75$; 63.5%) or mostly sedentary, e.g. unemployed, retired, office worker, manager, student, sales assistant ($N = 33$; 28.0%). Seventy-six (64%) reported that back pain was their main problem, 13% leg/buttock pain, 21% neurological disturbances, and 2% other.

Patients were asked to complete the questionnaire booklet, unassisted by the care provider. It was checked for completion and a second appointment was scheduled for 7 days after the first. Those patients who did not attend the scheduled second visit were contacted immediately to be reminded, such that a maximum of 10 days elapsed

between the two visits. On the follow-up visit the patient was asked to complete a second booklet, again unassisted, and hand it back before leaving the outpatient spine care unit. There was no change in the patient's treatment between the first and second visits; it was explained to the participants that treatment would only begin after the second visit, upon which the results of various exams such as X-rays, MRIs, blood tests, their questionnaires, etc., would be checked. All 118 patients who were recruited completed a second questionnaire within maximum 10 days of the first. As the volunteers filled in the booklet whilst in the spine care unit, the questionnaires could be checked for missing answers and fully completed at the time of their visit; just a few items remained missing, where patients considered the question not applicable.

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, since it has just one item per domain; for the ODI and RM it was considered that a minimum of 80% answers should be required (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 [2, 20]. Floor and ceiling effects were determined for all the 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 [28]. One type of construct validity, convergent validity, requires that different measures of the same or a similar construct agree to an acceptable extent [2] and in the present study this was evaluated using Spearman Rank correlation coefficients (ρ ; ρ) corrected for ties. It was hypothesised (based on the validation studies for the original COMI and as recommended by Streiner and Norman [27] for measures of the same/similar attributes) that correlation coefficients would range from 0.4 to 0.8 for the relationships between the COMI summary index score and RM and ODI scores.

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 five-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 ± 1 point (where at least 90% was considered acceptable) [22, 26]. For scales/items with approximately normally distributed interval data (pain scales, COMI summary score, Roland Morris, ODI), 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_{agreement 2,1}) and their 95% confidence intervals. Intraclass correlation coefficients (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 [28]. Standard errors of measurement SEM_{agreement} were used to indicate the absolute measurement error ("agreement" [28]) and to calculate the minimum detectable change (MDC95%) 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 $1.96 \times \sqrt{2} \times \text{SEM}$, 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 Brazilian-Portuguese version of the COMI is presented in "Appendix". Only one difficulty arose during its adaptation: translation of "back". At first, the word "*coluna*" was chosen in the consensus Brazilian-Portuguese version, but the English back-translation revealed this to be too medical–technical, coming back as "spine". After discussion, this was changed to "*costas*".

Upon conclusion of the main validation study, one slight change to the wording for one item was made, for improved clarity. For the work disability question, the original translation of "how many days...keep you from going to work (job, school, housework)?" was "...*te impiediu de fazer algo (trabalho, tarefas domésticas, escola, lazer)*". It was later felt that this might not focus sufficiently on the notion of *failing to go to work*, i.e., taking days off. After discussion, this was modified to "...*te impiediu de ir ao trabalho (emprego, escola, tarefas domésticas)*."

Missing data

There were no missing answers for the COMI. Only a couple of items in some of the other questionnaires had

Table 1 Floor and ceiling effects for the instruments in the baseline evaluation of 118 patients

Instrument	Floor effects (worst status) (%)	Ceiling effects (best status) (%)
COMI LBP	5.9	0.8
COMI LP	7.6	15.3
COMI worst pain of leg and back	11.0	2.5
COMI function	2.5	19.5
COMI symptom-specific well-being	66.0	2.5
COMI quality of life	6.8	5.9
COMI social disability	11.9	45.8
COMI work disability	12.7	54.2
Pain VAS (back or leg)	3.4	0
COMI summary score	1.0	0
Roland Morris score	1.7	0.8
Oswestry Disability Index score	0.8	0.8

Bold indicate scores from scales with more than one item

occasional missing answers, where patients felt the question was not applicable (e.g. for the question on sex in the ODI), but this was never to an extent that prevented calculation of the sum score using the missing-data rule (see “[Materials and methods](#)”).

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 1.

Acceptable floor effects (2–13%) were found for the COMI individual items pain, function, quality of life and social and work disability, but higher values were found for symptom-specific well-being (66%). A low ceiling effect (0–6.3%) was found for the individual COMI items back pain, worst pain, symptom-specific well-being, and general quality of life; however, ceiling effects were 20% for function, 46% for social disability and 54% for work disability.

Considering the multiple-item questionnaires, there were minimal floor effects (0–2%) for the COMI summary score, ODI, and RM; ceiling effects for these scales were similarly low (0–1%).

Construct validity

The correlation coefficients for the relationship between the COMI item/summary scores and the corresponding full-length questionnaires are shown in Table 2.

Table 2 Correlations between some of the COMI single items and the full-length pain/disability reference questionnaires

Core-index-items	Reference scales	ρ
Pain symptoms	Pain VAS for back or leg pain	0.72
Back function	Roland and Morris	0.37
	Oswestry Disability Index	0.46
Disability	Roland and Morris	0.51
	Oswestry Disability Index	0.52
COMI summary score ^a	Roland and Morris	0.57
	Oswestry Disability Index	0.64

ρ values in italics indicate those where the pre-defined hypothesis for the extent of the correlation could not be confirmed

^a 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); see text for further details

Only one of the hypotheses concerning the convergent validity of the COMI items (coefficients 0.4–0.8 with the corresponding full instruments) could not be confirmed, namely the relationship between the “back function” item of the COMI and the RM score ($r = 0.37$). A good correlation was found between the COMI worst pain score and the pain VAS for “back/leg pain” ($r = 0.72$). The correlation between the summary index score of the COMI and each of the full symptom-specific instruments’ scores was $r = 0.57$ – 0.64 .

Reproducibility

Differences in response to each domain on the COMI were ± 1 category in 98% patients for the domain ‘function’, 96% for ‘symptom-specific well-being’, 97% for ‘general quality of life’, 99% for ‘social disability’ and 100% for ‘work disability’, hence all satisfying the stability criterion of $\geq 90\%$ suggested by Nevill et al. [22].

Table 3 shows the mean (SD) scores on the two test occasions, and the ICC, SEM and MDC95% 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 back pain, leg pain, worst pain or COMI summary index scores, or for the RM scores; only the ODI scores showed a significant ($p = 0.01$), but only very slight difference between the two test occasions. The ICCs for COMI pain and COMI summary index scores were 0.91–0.96, which compared favourably with the corresponding values for the RM (ICC, 0.99) and ODI (ICC; 0.98) scales. The SEM for the COMI summary index score was 0.60 and the MDC95%, 1.66 points. Expressed as a percentage of the maximum score range for the given scale,

Table 3 Test–retest reliability results for the COMI summary index score, COMI individual pain scores and the full reference scale scores

Instrument	No. of items	Range	M1	M2	<i>P</i>	ICC	95% CI _{ICC}	SEM	SEM%	MDC95%
COMI summary index score	5	0–10	5.4 (1.9)	5.3 (2.0)	0.44	0.91	0.87–0.93	0.60	6.0	1.66
COMI back pain	1	0–10	6.4 (2.4)	6.3 (2.5)	0.37	0.93	0.90–0.95	0.65	6.5	1.79
COMI leg pain	1	0–10	5.1 (3.2)	5.1 (3.1)	0.92	0.96	0.94–0.97	0.67	6.7	1.86
COMI worst pain	1	0–10	6.9 (2.2)	6.8 (2.3)	0.14	0.93	0.90–0.95	0.62	6.2	1.71
Roland Morris disability	24	0–24	11.7 (5.7)	11.7 (5.9)	0.92	0.99	0.98–0.99	0.63	6.3	1.76
Oswestry Disability Index	10	0–100	29.6 (14.0)	30.4 (14.4)	0.01	0.98	0.96–0.98	2.24	2.2	6.19

M1, *M2* mean value at first and second assessment, *P* significance of difference between mean values on the two occasions, *ICC* intraclass correlation coefficient ($ICC_{2,1}$), *CI_{ICC}* 95% confidence intervals for the ICC, *SEM* standard error of measurement, *SEM%* SEM as percentage of maximum score, *MDC95%*, minimum detectable change score

the SEMs were similar for all instruments, being approximately 2–7%.

Discussion

The present study aimed to produce a cross-culturally adapted version of the COMI that would be valid and reliable for Brazilian-Portuguese speaking patients with back problems, and overall this aim appeared to be achieved. The process of cross-cultural adaptation, including translation and back-translation of the instrument, was carried out in accordance with established guidelines [3, 11] in an attempt to produce an adaptation of the questionnaire that would show a high degree of agreement with the original version. Overall, there were few problems translating the instrument and no missing data, and the psychometric characteristics of the COMI were comparable to those reported for the Spanish [9], German [17], Italian [16] and French [10] versions. Just a few items needed modification either to simplify the language and make it more patient-friendly (in relation to “spine” vs. “back”) or to clarify the concepts being investigated (absence from work).

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% [2, 20]), although they did not reach a level that would be considered adverse (>70%) for health-related quality of life questionnaires [14]. Interestingly, a similar phenomenon was seen for these same items in the German [17] Italian [16], Spanish [9] and French [10] versions of the COMI. However, as was the case for the other language versions, when the domain scores were combined to form the COMI summary score index, there were negligible floor and ceiling effects (<1%). Health-related quality of

life assessments often result in skewed distributions, and when the number of response categories is low, the proportion of responses at the extreme of the range naturally increases (with a dichotomous item by definition having *only* ceiling and floor effects). It has been suggested that high floor and ceiling effects are a threat to the responsiveness of an instrument, since they preclude the detection of improvement or worsening when it has, indeed, occurred. Interestingly, however, in three previous validation studies [8, 9, 17] the COMI summary scale was shown to be at least as responsive or even more responsive than other condition-specific instruments, having 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, in practice, the higher floor and ceiling effects are not so problematic, certainly not in terms of being a threat to responsiveness.

Construct validity

The individual core items of the Brazilian-Portuguese COMI were examined in relation to two multi-item condition-specific questionnaires, the Roland and Morris Disability questionnaire and the Oswestry Disability Index. Both of these have been established as being valid and reliable in the Brazilian-Portuguese language. When considering validity, it has been suggested that, in view of the fact that any measurement will have some associated error, 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 is to be assumed that either the reliability of one or the other measure is unacceptably low or that they are measuring different phenomena [27]. In the present study, the “function” item showed a correlation that was lower than hypothesised ($\rho = 0.37$) when examined in relation to the RM scores. However, its correlation with the ODI was better ($\rho = 0.46$), and confirmed the hypothesis on which construct validity was accepted ($\rho > 0.4$) for this domain. For the COMI summary index, the expected level of correlation with the longer condition-specific instruments was achieved (with coefficients of approx 0.6), confirming

the pre-defined hypothesis and concurring with the findings for the German (0.7–0.8) [17], Spanish (0.7–0.8) [9], Italian (0.6–0.7) [16] and French (0.6–0.7) [10] versions of the COMI back, and the English (0.5–0.6) [30] and German (0.6–0.7) [8] versions of the COMI neck.

Reproducibility

The test–retest reliability of the COMI was considered excellent, with ICCs for the individual pain scales being 0.93–0.96, and with an ICC for the COMI summary index score of 0.91. These ICCs were comparable to those previously reported for the COMI [9, 17] and compared well with those for the longer instruments evaluated here (0.98–0.99). The “minimum detectable change” (MDC95%) for the COMI summary index score was 1.7 points, which is similar to the values reported in previous language versions (German, 1.7 points [17]; Italian, 1.5 points [16]; French 2.0 points [10]). This MDC95% 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). The value of 17%, when expressed as a percent of the full-scale range (maximum value, 10 points), compares favourably with the values reported for other LBP outcome instruments [6]. Three studies have shown that the minimal *clinically important* difference (MCID) for the COMI is 2–3 points, depending on the external criterion used [8, 17, 19]. Assuming a similar MCID exists for the Brazilian-Portuguese version, then a clinically relevant change of 2–3 points (the “signal”) would far exceed the minimum detectable change of 1.7 points (the “noise”), rendering it an appropriate LBP outcome instrument [13].

The test–retest reliability or “stability” of the individual five-point response 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. [22], in which within-individual differences in responses on re-retest are calculated. These authors recommend that, when assessing the stability of five-point scales in self-report questionnaires, 90% participants should record test–retest differences within a reference value of ± 1 . In the present study, this was achieved by 96–100% patients for the individual COMI items, indicating good reliability.

Limitations of the study

Some limitations of the study are worthy of mention. The adaptation of the COMI was carried out and tested in patients living in Brazil, speaking Brazilian-Portuguese. Thus we cannot automatically guarantee its success in other

Portuguese-speaking countries or in patients that speak Portuguese living in other countries. This version may serve as a useful basis for further adaptation into European Portuguese, but would require further evaluation from other research groups working in these lands, as recommended by Guillemín et al. [11]. One of the inclusion criteria in the present study was the ability to read Brazilian-Portuguese, since we wanted to evaluate the COMI as a self-evaluation instrument, completed independently by the patient. However, many of the Brazilian population are illiterate or only partially-literate. Further study of the psychometric properties of the COMI when administered by interview should be carried out to ensure that such patients can also be included for evaluation in future studies/registries.

Due to limited administrative resources for carrying out the present study, and to reduce the burden on the patients, we did not include a generic quality of life (QoL) instrument in the questionnaire booklet. Although this would have assisted to cross-validate the two items concerned with general quality of life or symptom-specific well-being and to provide divergent validity for some of the other items, some of these generic QoL instruments (such as the WHO-QoL-Bref or the SF36) can be rather onerous, and the benefits to be gained in the present study were not considered to outweigh the disadvantages of lengthening the questionnaire booklet. The COMI general QoL item is taken directly out of the WHOQoL questionnaire, a Brazilian-Portuguese version of which has already been validated [31] (<http://www.ufrgs.br/psiq>) and hence this item can be considered appropriate for inclusion. The COMI symptom-specific well-being item has, in previous language versions (German, Spanish, French or Italian), failed to correlate well with any of the commonly used quality of life questionnaires (either SF36, WHOQoL or EurQoL) and hence little was expected to be gained here by examining this correlation again. Indeed, this highly responsive item appears to be more closely related to symptom severity than to QoL per se [9, 17, 19].

In assessing test–retest reliability or the stability of questionnaire scores, there is no recommended “appropriate” or “best” time interval to use between repeated assessments and it is always a trade-off between minimising recall effects (which can occur with too short a time interval) and the likelihood of true change (too long a time interval); generally, 1–2 weeks is considered appropriate [28], as used in the present study. We did not use a transition question to monitor any perceived global change in condition on the part of the patient, which would otherwise have assisted in ensuring that the patients did indeed consider themselves to be in a stable, chronic state. However, no treatment was administered between questionnaire completions and the ICCs were all extremely high, suggesting that the failure to include such an assessment was

of little consequence. Perhaps in studies of this type it is suffice to take a group with confirmed chronic symptoms rather than expressly excluding patients based on their answers to the transition question (which itself can also be subject to measurement error). However, this would require further investigation.

In the present study, we made no assessment of the sensitivity to change or responsiveness of the Portuguese-Brazilian COMI after treatment. However, having undergone successful cross-cultural adaptation, the questionnaire is now being used in quality management and outcome projects in connection with the European Spine Surgery registry, Spine Tango, and further data to examine its responsiveness should hence accumulate rapidly.

In conclusion, we have established that the Brazilian-Portuguese version of the COMI displays good psychometric

characteristics, comparable to those of other language versions of the instrument. We recommend the adaptation of the COMI in other languages to promote its further use in international multicentre studies, routine quality management and surgical registry systems. The systematic documentation of spinal care in this manner is of benefit not only to drive and monitor improvements in quality but also in forming the basis of educational material that will allow realistic expectations of treatment to be established. This should ultimately lead to improved care for the individual patient with LBP.

Acknowledgments The authors would like to thank Andy Cumming, Mary Jo Haggerty and Silvana Claudia Saadi Salomão Ciccone for help with the translations.

Conflict of interest None.

Appendix

Problemas de coluna podem levar a dor nas costas e/ou nas pernas e nádegas, assim como distúrbios sensoriais tais como formigamento, “pontadas” ou dormência em qualquer uma dessas regiões.

1 - Qual dos seguintes problemas o incomoda mais? Por favor, marque somente uma resposta.

- ☐ Dor nas costas
- ☐ Dor na perna/nádega
- ☐ Distúrbios sensoriais nas costas, pernas ou nádegas (por exemplo, formigamento, “pontadas”, dormência)
- ☐ Nenhuma das acima

2 - Para as duas questões a seguir (2a e 2b) nós gostaríamos que você indicasse a gravidade da sua dor, marcando com um “X” na linha de 0 a 10 (onde “0” = sem dor e “10” = a pior dor que você pode imaginar). Há perguntas separadas para dor nas costas e dor na nádega/perna (ciática).

2 a) Quão severa foi a sua dor nas costas na semana passada?

2 b) Quão severa foi a sua dor na nádega/perna (ciática) na semana passada?

3 - Durante a semana passada, quanto o seu problema nas costas interferiu no seu trabalho normal (incluindo trabalho fora de casa e as atividades domésticas)?

- ☐ Não interferiu
- ☐ Um pouco
- ☐ Moderadamente
- ☐ Muito
- ☐ Extremamente

4 - Se você tivesse que passar o resto da sua vida com os sintomas que você tem agora, como você se sentiria a respeito?

- ☐ Muito satisfeito
- ☐ Um pouco satisfeito
- ☐ Nem satisfeito, nem insatisfeito
- ☐ Um pouco insatisfeito
- ☐ Muito insatisfeito

5 - Por favor, pense sobre a semana passada. Como você avaliaria a sua qualidade de vida?

- ☐ Muito boa
- ☐ Boa
- ☐ Moderada
- ☐ Ruim
- ☐ Muito ruim

6 - Durante as últimas quatro semanas, em quantos dias você diminuiu as atividades que você geralmente faz (trabalho, tarefas domésticas, escola, lazer) por causa do seu problema nas costas?

- ☐ Nenhum
- ☐ Entre 1 e 7 dias
- ☐ Entre 8 e 14 dias
- ☐ Entre 15 e 21 dias
- ☐ Mais do que 21 dias

7 - Durante as últimas quatro semanas, por quantos dias o seu problema nas costas te impediu de ir ao trabalho (emprego, escola, tarefas domésticas)?

- ☐ Nenhum
- ☐ Entre 1 e 7 dias
- ☐ Entre 8 e 14 dias
- ☐ Entre 15 e 21 dias
- ☐ Mais que 21 dias

References

- Airaksinen O, Brox JJ, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G (2006) Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 15(Suppl 2):S192–S300
- Andresen EM (2000) Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 81: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
- 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
- Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82:8–24
- Deyo RA, Battie M, Beurskens AJHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korf M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine* 23:2003–2013
- Fankhauser C, Mutter U, Aghayev E, Mannion AF (2011) Validity and responsiveness of the Core Outcome Measures Index (COMI) for the neck. *Eur Spine J*. doi:10.1007/s00586-011-1921-4
- 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)
- Genevay S, Cedraschi C, Marty M, Rozenberg S, de Goumoens P, Faundez A, Balague F, Porchet F, Mannion A (2009) Reliability and validity of the cross-culturally adapted French version of the Core Outcome Measures Index (COMI). *Eur Spine J* 18:S478
- 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
- Haefeli M, Elfering A (2006) Pain assessment. *Eur Spine J* 15(Suppl 1):S17–S24
- 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
- Hyland ME (2003) A brief guide to the selection of quality of life instrument. *Health Qual Life Outcomes* 1:24
- Kessler JT, Melloh M, Zweig T, Aghayev E, Roder C (2011) 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
- 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*. doi:10.1007/s00586-011-1741-6
- 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
- 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
- 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
- McHorney CA, Tarlov AR (1995) Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 4:293–307
- 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
- Nevill AM, Lane AM, Kilgour LJ, Bowes N, Whyte GP (2001) Stability of psychometric questionnaires. *J Sports Sci* 19:273–278
- Nusbaum L, Natour J, Ferraz MB, Goldenberg J (2001) Translation, adaptation and validation of the Roland–Morris questionnaire—Brazil Roland–Morris. *Braz J Med Biol Res* 34:203–210
- 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
- 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
- Staerkle 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 Inc., 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
- Vigatto R, Alexandre NM, Correa Filho HR (2007) Development of a Brazilian Portuguese version of the Oswestry Disability Index: cross-cultural adaptation, reliability, and validity. *Spine (Phila Pa 1976)* 32:481–486
- White P, Lewith G, Prescott P (2004) The core outcomes for neck pain: validation of a new outcome measure. *Spine* 29:1923–1930
- WHOQOL (1998) The World Health Organisation WHOQOL-BREF Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Soc Sci Med* 46:1569–1585
- 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