Cross-cultural adaptation and psychometric validation of the Dutch version of the Core Outcome Measures Index for the back (COMI-back) in patients undergoing surgery for degenerative disease of the lumbar spine

Pravesh S. Gadjradja,b,*, Mehrman Chalakia,c, Maurits W. van Tulderd, Biswadjiets. Harhangia,e

a Department of Neurosurgery, Park MC, Rotterdam, the Netherlands
b Department of Neurological Surgery, Weill Cornell Brain and Spine Center, New York-Presbyterian Hospital, New York, NY, USA
c Department of Neurosurgery, Leiden University Medical Center, Leiden, the Netherlands
d Department of Human Movement Sciences, Faculty of Behavioral and Movement Sciences, Amsterdam Movement Sciences Research Institute, Vrije Universiteit Amsterdam, the Netherlands
e Department of Neurosurgery, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

ARTICLE INFO
Keywords:
Core outcomes measurement index
Low back pain
Patient reported outcome measure
Dutch
Validation

ABSTRACT
Introduction: Patient-reported outcome measures (PROMs) are the preferred outcomes measured in patients with lumbar spinal degenerative diseases. As PROMs can be lengthy and therefore pose a burden to patients and researchers, short and standardized PROMs are needed, such as the Core Outcome Measures Index (COMI).

Research question: Is the Dutch version of the COMI-back a reliable and responsive PROM to measure outcomes in lumbar degenerative spinal surgery?

Methods: After translating and cross-cultural adapting the COMI-back into Dutch, patients who were on the waiting-list for lumbar decompression surgery in a secondary referral center, were enrolled in the validation study. Patients completed a baseline booklet consisting of the COMI-back, likert scales measuring back and leg pain, the Roland-Morris Disability questionnaire, the EuroQol-5 dimensions and the WHO-Quality-of-Life-BREF questionnaire to test construct validity. Within 2 weeks and before undergoing surgery, patients completed the COMI-back again, to measure test-retest stability. Three months after surgery, a global treatment outcome (GTO) question and the COMI-back were completed to test responsiveness.

Results: The COMI-back was successfully translated and adapted into Dutch. One-hundred-thirty-five patients completed the baseline booklet, 93 the test-retest questionnaire and 102 the responsiveness questionnaire. The COMI-summary score and four of five COMI-domains, showed good to very good correlation to the reference questionnaires (ρ > 0.41). The COMI-back showed a good test-retest stability with an intraclass correlation coefficient of 0.81 for the summary score. Furthermore, the standard error of agreement was 0.65 and the minimal detectable change was 1.8 points. The ROC-curve showed an area under the curve of 0.89 (95% CI 0.82 to 0.95).

Conclusion: The Dutch version of the COMI-back has satisfactory psychometric properties and is a reliable and responsive patient-reported outcome measure in patients undergoing surgery for lumbar degenerative disease.

1. Introduction
In degenerative spinal surgery, patient-reported outcome measures (PROMs) are the outcome measure of choice when outcomes of treatments are measured or compared between patients or patient groups (Deyo et al., 1998). Throughout the years multiple PROMs have been validated and are used in studies involving patients undergoing lumbar spine surgery (Ostelo et al., 2004; Trompeenaars et al., 2005). These PROMs can cover one or more outcome domains such as pain or quality of life (Chiarotto et al., 2015). Because large studies are usually interested in multiple domains, patients are subsequently subjected to multiple PROMs which can be burdensome (Chiarotto et al., 2015). The Core

Abbreviations: PROM, Patient-reported outcome measure; COMI, Core Outcomes Measures Index; RMDQ, Roland-Morris Disability Questionnaire; EQ5D, EuroQol-5 dimensions; WHOQoL-BREF, World Health Organization Quality of Life-BREF questionnaire; ICC, Intraclass correlation coefficient; SEM, Standard errors of measurement agreement; MDC, Minimum detectable change; SRM, Standardized response mean; ROC, Receiver-operating characteristic.

* Corresponding author. Department of Neurosurgery, Park MC, Hoofdweg 90, 3067 GH, Rotterdam, the Netherlands.
E-mail address: p.gadjradj@erasmusmc.nl (P.S. Gadjradj).

https://doi.org/10.1016/j.bas.2021.100004
Received 14 July 2021; Accepted 6 August 2021
Available online 18 August 2021
2772-5294/© 2021 The Authors. Published by Elsevier B.V. on behalf of EUROSPINE, the Spine Society of Europe, EANS, the European Association of Neurosurgical Sciences. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
Outcome Measures Index is a PROM measuring five domains using only seven questions (Mannion et al., 2009). These domains are (1) pain; (2) function; (3) symptom-specific well-being; (4) quality of life and (5) disability.

The COMI-back has been cross-culturally adapted and validated in several languages and among different patient groups and has become the PROM of choice for the The Spine Tango, the Eurospine Registry (Abeldaidem et al., 2020; Damasceno et al., 2012; Klemencics et al., 2016; Komesh et al., 2020; Mannion et al., 2016; Nagata et al., 2020; Qiao et al., 2013). A validation among Dutch patients or patients who can speak or read Dutch in the Netherlands, undergoing lumbar spine surgery, however, has not been conducted yet. Therefore, the purpose of the current study was to translate, cross-culturally adapt and validate the COMI-back in Dutch among patients undergoing surgery for lumbar degenerative spinal disease.

2. Methods

2.1. COMI-back

The COMI-back is a patient-reported outcome measure, that was intended to measure seven domains through seven questions. These domains are (1) low back pain; (2) leg pain; (3) function; (4) symptom-specific well-being; (5) quality of life; (6) social disability and (7) work disability. Low back pain and leg pain are measured on a numeric rating scale ranging from 0 to 10 with 10 indicating the “worst pain imaginable”. The other 5 domains are scored on a scale ranging from 0 to 5 with 5 indicating the “worst status”. These scores can be recalculated to scores ranging from 0 to 10 points. All seven domains can be used to calculate an overall COMI-summary score by selecting the highest pain score between low back pain and leg pain (worst pain score) and taking the average from work and disability scores (disability score). Then the COMI-summary score is calculated as the average of the worst pain score, the disability score, function, symptom-specific well-being and quality of life, and consequently ranges from 0 (best clinical status) to 10 (worst clinical status).

2.2. Translation process and cross-cultural adaptation

The English version of the COMI-back was translated to a Dutch version using established guidelines (Beaton et al., 2000). In brief, two native Dutch speakers (one involved in the field of spine surgery and one translator as layman) translated the English version to a Dutch version, independently. Afterwards, both translations were compared together with the original English version, which resulted in a first Dutch version. Subsequently, two native English speakers (one colleague and one layman) performed the back translation, independently. Both back translations were compared with the original English version.

All discrepancies were discussed during a research meeting, consisting of a clinician fluent in both languages and two bilingual researchers. All discrepancies were resolved in a group discussion and a pre-final Dutch version of the COMI-back was created. After institutional review board approval, this version was pilot tested among a subset of patients on a waiting list for lumbar decompression surgery. The pilot-testing did not result in any relevant interpretation issues and the pre-final version was finalized.

2.3. Patients

To further test the psychometric properties of the Dutch version of the COMI-back, the questionnaire was further validated among patients enrolled from two secondary referral centers for low complex spinal surgery. After their consultation with the neurosurgeon and after an indication for surgery was determined, patients were asked to participate in the current study. Patients were included if they were (1) at least 18 years of age; (2) a candidate for lumbar decompression surgery due to lumbar spinal stenosis and/or lumbar disk herniation. Exclusion criteria were (1) pregnancy; (2) infections or malignancy; (3) neurodegenerative disease; (4) lack of cooperation; (5) unable to understand Dutch and (6) serious psychopathology (to the discretion of the neurosurgeon).

2.4. Study procedures

After acquiring written informed consent, patients were requested to fill in a baseline booklet consisting of multiple PROMs. Within two weeks, and before undergoing surgery, patients were requested to return the retest questionnaire. Finally, three months after surgery, patients were asked to return the responsiveness questionnaire. All questionnaires were filled in by the patients without any assistance.

The baseline booklet consisted of (1) questions regarding patients’ demographics; (2) the Dutch version of the COMI-back; (3) two five-point likert-scales measuring back and leg pain; (4) the Roland-Morris Disability Questionnaire; (5) the EuroQol-5 dimension questionnaire; and (6) the World Health Organization Quality of Life-BREF questionnaire. The Roland-Morris Disability Questionnaire (RMDQ) is a 24-question PROM measuring functional disability due to low back pain from 0 (no disability) to 24 (maximal disability) (Ostelo et al., 2004). The EuroQol-5 dimensions questionnaire (EQSD) is a PROM consisting of 5 multiple choice answers and a 0 to 100 VAS measuring quality of life (MV et al., 2016). The World Health Organization Quality of Life-BREF questionnaire (WHOQOL-BREF) is a 26-question PROM covering four domains. The crude scores on the WHOQol-BREF can be transformed to a 0 to 100 (best quality of life) summary score (Tronpenaars et al., 2005).

The retest and responsiveness questionnaires contained the COMI-back and a global treatment outcome question (GTO) that is measuring self-reported recovery on a 7-point likert scale.

2.5. Analysis

All analyses were conducted using SPSS (version 24) with a p-value <0.05 indicating statistical significance. Missing data were handled as customary for each specific PROM. When the COMI-summary score is calculated, no missing answers were allowed for the analysis (Ostelo et al., 2004; Tronpenaars et al., 2005; MV et al., 2016). To assess the psychometric properties of the COMI-back, (1) floor/ceiling effects; (2) construct validity; (3) test-retest stability and (4) responsiveness were measured (Tronpenaars et al., 2005).

Floor/ceiling effects depict the percentage of patients which have the worst (floor) and best (ceiling) score on the PROM and do not reflect any meaningful deterioration or improvement in clinical status on the PROM. Descriptive statistics were used to measure floor/ceiling effects, and floor/ceiling effects were considered to be present if they were >15% and to be detrimental if they were >75%.

Construct validity involves the correlation between the new instrument, the COMI-back, and other validated PROMs. The Pearson’s correlation coefficient (r) was used to evaluate the construct validity. A coefficient was considered to depict an excellent correlation with a r ≥ 0.81, very good correlation with a 0.61 ≤ r < 0.81, a good correlation with a 0.41 ≤ r < 0.61, a poor correlation with a 0.21 ≤ r < 0.40 and no correlation with a r < 0.21 (Cohen, 1988; Steiner and Norman, 1995). It was hypothesized that all domains, except for the domain symptom-specific well-being, had a good correlation (r ≥ 0.41) for their reference questionnaires (see Table 3). Based on the literature, symptom-specific well-being was expected to have a fair correlation with the RMDQ and WHOQOL-BREF (physical function) (Nagata et al., 2020; Storheim et al., 2012; Genevay et al., 2012).

To assess reproducibility, test-retest stability was calculated. Reproducible tests should lead to similar test results with repeated testing when no change in clinical status occurs. To calculate the test-retest stability, only patients that indicated “no change in symptoms” on the GTO were included for this analysis. In these patients, the intraclass correlation coefficient (ICC) was calculated with corresponding 95% confidence intervals.
interval for all COMI-domains and the COMI-summary score. An ICC > 0.7 was considered to indicate acceptable reproducibility in a study sample of minimal 50 study participants (Terwee et al., 2007). Furthermore, standard errors of measurement agreement (SEM) and minimum detectable change (MDC) were calculated.

Responsiveness is the ability of an instrument to detect change in a clinical condition when it occurs. Responsiveness was measured in three ways (1) comparing scores on the COMI-back at baseline and at three months after surgery, using t-tests; (2) calculating the standardized response mean (SRM); (3) construct a receiver-operating characteristic (ROC) curve to evaluate sensitivity and specificity anchored on the GTO. An area under the curve of at least 0.70 was considered to be acceptable and the optimal cut-off score was determined using the Youden-method (Terwee et al., 2007; Youden, 1950).

3. Results

The COMI-back was successfully translated and is shown in the supplementary material. Items for discussion were mostly focused on cross-culture adaptations of items as “buttocks” and “pins and needles”. After performing the backward translations, no additional modifications were made to the COMI-back. Fig. 1 gives an overview of the study procedures. A total of 135 patients were included; 93 patients returned a re-test questionnaire and 102 patients returned the responsiveness questionnaire.

3.1. Demographics

Table 1 gives an overview of the baseline characteristics of the included patients. Patients had a mean age of 59.3 ± 15.4 years and 74.0% was married or in a relationship. Overall, 44.6% of the included patients had a paid job. The majority of the patients (45.4%) only had received lower education, while 24.6% had received middle level and 30.0% higher education. At baseline, 72.0% of the patients used pain medication.

3.2. Acceptability and floor/ceiling effect

On average, patients completed the COMI-back in four and a half minutes. Table 2 gives an overview of the percentage missing data of the COMI-back at baseline and the floor/ceiling effects at baseline and at 3-months follow-up. Missing data ranged from 0 to 4.4% on the different COMI-domains. The domains low back pain, leg pain, social and functional disability, and the summary score had missing items. At baseline, the domains function (31.9%) and the social- (54.5%) and work (41.4%) disability showed high floor effects. Only the domain symptom-specific well-being (29.5%) showed high ceiling effects at baseline. At three months, the domains symptom-specific well-being (29.5%) and the domains social- (27.6%) and work (29.7%) disability had high floor effects. Furthermore, high ceiling effects were present at the domain of low back pain (16.2%), leg pain (21.0%) and social- (24.8%) and work (38.6%) disability at three months follow-up.

Fig. 1. Flowchart of study procedures.
3.3. Construct validity

Table 3 gives an overview of the correlation tests of the COMI-back with the other questionnaires tested at baseline. The COMI domains low back pain and leg pain showed a very good correlation with the likert scales for low back pain (0.77) and leg pain (0.64). The COMI domain function showed a good correlation with the RMDQ (0.41) and WHOOQL-BREF Physical function (–0.47). The domain symptom-specific well-being showed a poor correlation with the RMDQ (0.27) and WHOOQL-BREF Physical function (–0.33), and no correlation with the WHOOQL total score (–0.18). The COMI domain quality of life showed good correlation with the EQSD (–0.49) and EQSD-VAS (–0.45), but poor correlation with the WHOOQL-BREF (–0.40). The COMI domains social- and work disability showed good correlations with the RMDQ (0.47 and 0.53, respectively) and with the WHOOQL-BREF (both –0.47). Finally, the overall COMI summary score showed a good correlation with the EQSD (–0.51) and the WHOOQL-BREF (–0.47), and a very good correlation with the RMDQ (0.63) and WHOOQL-BREF Physical function (–0.65).

3.4. Test-retest reliability

Of the 93 patients that returned the test-retest questionnaire, 60 experienced no change in symptoms based on the GTO-question. Table 4 gives an overview of the results of the test-retest reliability analyses. Except for the domains function (0.53), symptom-specific well-being (0.51) and quality of life (0.51), all domains (0.75–0.88) and the summary score (0.81) showed good test-retest stability. The COMI summary score had a SEM of 0.65 and a MDC of 1.8.

Table 2
Floor and ceiling effects at baseline and at 3-months of follow-up.

|                      | Min-max | Missing at baseline (%) | Floor effects % (N) | Ceiling effects % (N) | Floor effects % (N) | Ceiling effects % (N) |
|----------------------|---------|-------------------------|---------------------|----------------------|---------------------|----------------------|
| COMI low back pain   | 0–10    | 2 (1.5%)                | 3 (2.3%)            | 3 (2.3%)             | 0                   | 17 (16.2%)           |
| COMI leg pain        | 0–10    | 2 (1.5%)                | 8 (6.0%)            | 0                    | 1 (1.0%)            | 22 (21.0%)           |
| COMI function        | 0–10    | 0                       | 43 (31.9%)          | 4 (3.0%)             | 12 (11.4%)          | 11 (10.5%)           |
| COMI symptom-specific well-being | 0–10 | 0                       | 110 (81.5%)         | 0                    | 31 (29.5%)          | 12 (11.4%)           |
| COMI quality of life | 0–10    | 0                       | 17 (12.6%)          | 1 (0.7%)             | 8 (7.6%)            | 5 (4.8%)             |
| COMI social disability | 0–10 | 1 (0.7%)                | 73 (54.3%)          | 3 (2.2%)             | 29 (27.6%)          | 26 (24.8%)           |
| COMI work disability | 0–10    | 2 (1.5%)                | 55 (41.4%)          | 19 (14.3%)           | 30 (29.7%)          | 39 (38.6%)           |
| COMI summary score   | 0–10    | 6 (4.4%)                | 1 (0.8%)            | 0                    | 0                   | 1 (1.0%)             |

3.5. Responsiveness

Using the dichotomized likert-scale for recovery, 57.3% of the patients were fully recovered at three months follow-up. Patients that recovered had a mean COMI-summary score of 3.2 ± 2.0 compared to 6.6 ± 1.8 (mean difference –3.3, p < 0.001). The SRM was 1.1 overall. The SRM for a good outcome was 2.0 while the SRM for a poor outcome was 0.65. Fig. 2 depicts the ROC-curve of the COMI-summary score anchored for recovery. The area under the curve was 0.89 with 95% CI (0.82–0.95). The optimal cut-off was 2.0 points and yielded a sensitivity of 69% and specificity of 89%.

4. Discussion

This study presents the cross-culturally adapted and validated Dutch version of the COMI-back. The COMI-back was validated among 135 patients who underwent surgery for degenerative lumbar spinal disease. The COMI-back had only a small percentage of missing answers and the COMI-summary score had negligible floor and ceiling effects. One domain had detrimental floor effects at baseline, while the COMI-summary score had no floor and ceiling effects. Furthermore, the COMI-summary score and all COMI domains, except for the domain symptom-specific well-being, showed good to very good correlation to the reference questionnaires. Test-retest stability was good for the summary score and responsiveness analysis showed a good discriminatory ability of the COMI-back between patients with a good and patients with a poor outcome.

Recently, a Dutch version of the COMI-back was cross-culturally adapted and validated among Dutch-speaking Belgian patients.
suffering from low back pain with or without leg pain (Van Lerbeirgh et al., 2018). It is unknown of these patients were candidates for surgery or conservative treatment. Both Dutch versions show many similarities, but also some cultural differences. For instance, words like “gelieve” (please)”, “job (job)”, “omwille (because)”, are more commonly or only used in the Dutch that is spoken in Flanders, Belgium. The ‘Flemish Dutch’ validation show that the COMI-back is a valid and reliable patient-reported outcome measure in patients with at low back pain while the current ‘Holland Dutch’ validation confirms this among patients on a waiting list for lumbar decompression surgery. Furthermore, the current study also demonstrated good responsiveness of the COMI-back.

In the current study, the domains function and disability showed high floor effects, while symptom-specific well-being showed detrimental floor effect at baseline. Although floor and ceiling effects are unfavorable for PROMs in general, these effects are not unknown to the COMI (Damasceno et al., 2012; Klemencsics et al., 2016; Storheim et al., 2012; Miekisiak et al., 2013). The unexpected high floor effect in the domain symptom-specific well-being, may also be partially explained by our study population, as invalidating pain or disability is part of the indication for surgery. Nevertheless, the COMI-summary score showed negligible floor and ceiling effects and therefore only a small impact is expected of the floor and ceiling effects of the individual domains.

Except for one domain, all COMI-domains and the summary score showed good to very good correlation with the index questionnaires. Both the (very) good and poor correlations shown in the current study are comparable to the correlations found in validation studies in other languages of the COMI-back (Abdeldaiem et al., 2020; Damasceno et al., 2012; Klemencsics et al., 2016; Storheim et al., 2012; Van Lerbeirgh et al., 2018; Miekisiak et al., 2013; Kim et al., 2018). Furthermore, also the test-retest reliability and MDC-scores calculated in the current study seem to be in line with the literature (Abdeldaiem et al., 2020; Damasceno et al., 2012; Klemencsics et al., 2016; Komesh et al., 2020).

The responsiveness of the COMI-back has only been assessed in a few other studies (Klemencsics et al., 2016; Mannion et al., 2005). In the validation study of the Hungarian version of the COMI-back, the responsiveness was tested in 159 patients, six months after they underwent surgery for the lumbar spine. In a German validation study, similarly, responsiveness’ was tested in a surgical population (Mannion et al., 2005). The ROC-curves of the Hungarian and German versions had an area under the curve of 0.84 and 0.82, respectively. In the current study an area under the curve of 0.89 was found, indicating comparable and good discriminatory abilities of the COMI-back.

Some limitations of the study methodology have to be acknowledged. The Dutch language may have some cultural differences between the Dutch-speaking countries the Netherlands and Belgium. Therefore, this version may not be as applicable in other Dutch speaking regions, such as Surinam and the Netherlands Antilles. In the current study, we tried to prevent the use of region-specific language. Another limitation may be the use of the Pearson ρ to test the correlation as it may be debatable that the use of the Spearman ρ might have been more appropriate. Post-hoc analyses using the Spearman ρ coefficient, showed only small differences in correlation compared to the Pearson ρ coefficient and did not change the conclusion. Strengths of this study include the adherence to guidelines for the adaptation of PROMs, the appropriate sample sizes to perform the psychometric testing, the inclusion of a GTO-question while measuring the test-retest reliability, the inclusion of testing the responsiveness and the inclusion of patients from multiple clinics in the Netherlands. Specific strengths of applying the COMI-back in practice includes the short time needed to complete the questionnaire and the low burden to patients and researchers.

The Dutch version of the COMI-back has satisfactory psychometric properties and is a reliable and responsive patient-reported outcome measure in patients undergoing surgery for lumbar degenerative disease. As the use of the COMI-back is easy to understand and takes only about four and a half minutes to complete, we recommend using it as a patient-reported outcome measure in studies and registries focused on degenerative lumbar spinal surgery.

Declaration of competing interest

All authors (PG, MC, MvT and BSH) have no conflict of interest to disclose. ICMJE-forms of all authors are available upon request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bas.2021.100004.

References

Abdeldaiem, A., Saveeress, E.S.B., Shehab-Eldien, A.A., Mannion, A.F., Rehan Youssef, A., 2020. Cross-cultural adaptation and validation of the Arabic version of the Core Outcome Measures Index for the back (COMI-back) in patients with non-specific low back pain. Eur. Spine J 29, 2413–2430. https://doi.org/10.1007/s00586-020-06530-4.

Beaton, D.E., Bombardier, C., Guillemin, F., Ferraz, M.B., 2000. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 25, 3186–3191. https://doi.org/10.1097/00007632-200012150-00014.

Chiariotto, A., Deyo, R.A., Terwee, C.B., Boers, M., Buchbinder, R., Corbin, T.P., Costa, L.O., Foster, N.E., Grotle, M., Koes, B.W., Kovacs, F.M., Lin, C.W., Maher, C.G., Pearson, A.M., Peul, W.C., Schoene, M.L., Turk, D.C., van Tulder, M.W., Ostelo, R.W., 2015. Core outcome domains for clinical trials in non-specific low back pain. Eur. Spine J 24, 1127–1142. https://doi.org/10.1007/s00586-015-3892-3.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates.

Damasceno, L.H., Rocha, P.A., Barbosa, E.S., Barron, C.A., Canto, F.T., Defino, J.L., Mannion, A.F., 2012. Cross-cultural adaptation and assessment of the reliability and validity of the Core Outcome Measures Index (COMI) for the Brazilian-Portuguese language. Eur. Spine J 21, 1273–1282. https://doi.org/10.1007/s00586-011-21060.

Deyo, R.A., Battie, M., Beurskens, A.J., 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. https://doi.org/10.1097/00007632-199809150-00018.

Genevay, S., Codorasci, C., Marty, M., Rozenberg, S., De Goumoens, P., Faundez, A., Balague, F., Porchet, F., Mannion, A.F., 2012. Reliability and validity of the cross-culturally adapted French version of the Core Outcome Measures Index (COMI) in patients with low back pain. Eur. Spine J 21, 130–137. https://doi.org/10.1007/s00586-011-1992-2.

Kim, H.J., Yeom, J.S., Nam, Y., Lee, N.K., Heo, Y.W., Lee, S.Y., Park, J., Chang, B.S., Lee, C.K., Chun, H.J., Mannion, A.F., 2018. Validation and cross-cultural adaptation of the Korean version of the Core Outcome Measures Index in patients with
