

Responsiveness of the Oswestry Disability Index and the Roland Morris Disability Questionnaire in Italian subjects with sub-acute and chronic low back pain

Marco Monticone · Paola Baiardi · Carla Vanti ·
Silvano Ferrari · Paolo Pillastrini · Raffaele Mugnai ·
Calogero Foti

Received: 28 January 2011 / Revised: 6 July 2011 / Accepted: 24 July 2011 / Published online: 8 August 2011
© Springer-Verlag 2011

Abstract

Introduction An ability to assess longitudinal changes in health status is crucial for the outcome measures used in treatment efficacy trials. The aim of this study was to verify the responsiveness of the Italian versions of the Oswestry Disability Index (ODI) and the Roland Morris Disability Questionnaire (RMDQ) in subjects with subacute or chronic low back pain (LBP).

Material and methods At the beginning and end of an 8 week rehabilitation programme, 179 patients completed a booklet containing the ODI, the RMDQ, a 0–10 numerical

rating scale (NRS), and the 36-item Short-Form Health Survey (SF-36). A global perception of change scale was also completed at the end of the programme, and collapsed to produce a dichotomous outcome (i.e. improved vs. not improved). Responsiveness was assessed by means of distribution methods [minimum detectable change (MDC); effect size (ES); standardised response mean (SRM)] and anchor-based methods (ROC curves).

Results The MDC for the ODI and RMDQ was, respectively, 13.67 and 4.87; the ES was 0.53 and 0.68; and the SRM was 0.80 and 0.81. ROC analysis revealed an area under the curve of 0.71 for the ODI and 0.64 for the RMDQ, thus indicating discriminating capacity; the best cut-off point for the dichotomous outcome was 9.5 for the ODI (sensitivity 76% and specificity 63%) and 2.5 for the RMDQ (sensitivity 62% and specificity 55%). These estimates were comparable between the subacute and chronic subjects. Both the ODI and the RMDQ moderately correlated with the SF-36 and NRS (Spearman's and Pearson's correlation coefficients of >0.30).

Conclusion The Italian ODI and RMDQ proved to be sensitive in detecting clinical changes after conservative treatment for subacute and chronic LBP. Our findings are consistent with those published in the literature, thus allowing cross-cultural comparisons and stimulating cross-national studies.

IRB approval. Our Institutional Review Board approved the study, which was conducted in conformity with ethical and humane principles of research.

M. Monticone (✉)

Physical Medicine and Rehabilitation Unit, Scientific Institute of Lissone (Milan), Salvatore Maugeri Foundation, IRCCS, Via Monsignor Bernasconi 16, 20035 Lissone, MI, Italy
e-mail: marco.monticone@fsm.it

P. Baiardi

Consorzio Valutazioni Biologiche e Farmacologiche,
Pavia University, Salvatore Maugeri Foundation,
IRCCS, Pavia, Italy

C. Vanti

School of Physiotherapy, Bologna University, Bologna, Italy

S. Ferrari

Manual Therapy Sciences, Padova University, Padua, Italy

P. Pillastrini · R. Mugnai

Department of Internal Medicine, Geriatrics and Nephrology,
University of Bologna, Bologna, Italy

C. Foti

Chair of Physical and Rehabilitation Medicine,
Tor Vergata University of Rome, Rome, Italy

Keywords Low back pain · Responsiveness ·
Oswestry Disability Index · Roland Morris Disability
Questionnaire · Outcome measures

Introduction

Surveys of patient self-reported health and function have become useful means of assessing low back pain (LBP) outcomes that have replaced physiological measurements,

which have proved to have little relevance for patients with back symptoms [1].

The Oswestry Disability Index (ODI) and the Roland Morris Disability Questionnaire (RMDQ) are the primary condition-specific health status measures (i.e. outcome instruments that focus on the specific symptoms or functional impact of a particular condition) for the assessment of LBP-related disability [2, 3]. The ODI is a self-administered, 10-item questionnaire: the first section rates the intensity of pain and the others describe its disabling effect on typical daily activities. The score for each item ranges from 0 to 5, and the sum of the ten scores is expressed as a percentage of the maximum score and thus ranges from 0 (no disability) to 100 (maximum disability) [4]. The RMDQ is a self-administered questionnaire derived from the Sickness Impact Profile that consists of 24 items reflecting a variety of daily living activities; each item is scored 1 if declared applicable to the respondent and 0 if not, and so the total score can vary from 0 (no disability) to 24 (severe disability) [5]. Both questionnaires were originally developed in English, but they have been culturally adapted in various languages and have satisfactory psychometric properties (internal consistency, reproducibility and validity) in a wide variety of situations [2].

The ability to assess longitudinal changes in health status is crucial for the outcome measures used in treatment efficacy trials, such as the minimum detectable change (MDC; sensitivity, or the smallest change in score that probably reflects a true change rather than a measurement error) and the minimal clinically important difference (MCID, or the smallest difference in score that patients perceive as being beneficial) [6]. Determining these values (also called responsiveness) is not only important in clinical decision making (individual level), but also for power calculations, sample size estimates and cost evaluations in clinical research (group level) [7].

Responsiveness data are available for the English versions of the ODI and the RMDQ, and some of the translated questionnaires for LBP and other spinal conditions [2]. The Italian versions of the ODI and RMDQ have been psychometrically analysed and found to have similar properties to those of other versions [8, 9], but their responsiveness has not yet been determined and this limits their use for clinical and research purposes.

The primary aim of this study was to determine the responsiveness of the two questionnaires in a large population of Italian subjects with sub-acute or chronic common LBP using the distribution-based and anchor-based methods mainly suggested in the current literature [10, 11]. The secondary aim was to compare the results with existing data in order to evaluate the possibility of making cross-cultural comparisons and conducting cross-national research studies.

Methods

This research was part of an observational study that was approved by the Institutional Review Board of our research hospital; the patients gave their written consent to participate.

Subjects

Outpatients referred to our Rehabilitation Unit and three affiliated rehabilitation centres were enrolled between September 2009 and June 2010. The inclusion criteria were diagnosis of sub-acute or chronic common LBP, age of 18–70 years, and ability to read and speak fluent Italian; the exclusion criteria were acute common LBP, specific causes of LBP including disc herniation, canal stenosis, spinal deformity, fracture, spondylolisthesis, infections, central or peripheral neurological signs, systemic illness, and psychiatric or neuropsychological deficits. Patients with recent myocardial infarctions, cerebrovascular events, or chronic lung or renal diseases were also ruled out by case history and excluded.

The socio-demographic and clinical characteristics of the enrolled patients were investigated using a specific schedule.

Procedures

All the participants were provided written information concerning the questionnaires and procedures by three research assistants. Those satisfying the entry criteria underwent an 8 week rehabilitation programme that included exercises aimed at improving postural control, strengthening and stabilising back muscles, and stretching; cognitive-behavioural principles targeted on fear avoidance beliefs, catastrophising, coping strategies and illness behaviours were also used as part of a bio-psychosocial approach to LBP. This conservative programme was the same for all of the enrolled subjects.

The Italian versions of the RMDQ and the ODI [8, 9] were administered to all the patients as part of a comprehensive pre- and post-rehabilitation assessment that included evaluations of pain, the quality of life and the global perceived effect (GPE).

Pain intensity was assessed using a 0–10 numerical rating scale [12], and the quality of life by means of the Italian version of the Short-Form Health Survey questionnaire (SF-36), with the eight domain scores being calculated on the basis of the User's Manual for the Italian version [13, 14]. Global perception of change at the end of treatment was determined using a five-level Likert scale, which had two improvement levels (much better = 1, better = 2), one no change level (approximately the

same = 3) and two worsening levels (a little worse = 4, worse = 5) [15].

Statistical procedures

The levels of the global perception of the condition scale were collapsed to produce a dichotomous variable: improved (much better and better) and not improved (approximately the same, a little worse and worse).

Responsiveness was determined using distribution and anchor-based methods [10]: the former included the minimum detectable change (MDC), effect size (ES, also using Guyatt's approach), and the standardised response mean (SRM).

The MDC was calculated by multiplying the standard error of the measurements (SEM) by the z-score associated with a 95% level of confidence and the square root of 2, which reflects the additional uncertainty introduced using difference scores based on measurements made at two time points (pre- and post-rehabilitation assessment). The SEM indicates the precision of the outcome measure, and was estimated by taking the square root of the within-subject variance of the patients categorised as "unchanged" (GPE scores 3, 4 or 5). As only the unchanged patients were assessed, there was a more than 95% chance that no real change had occurred in the patients whose change scores were less than or equal to the MDC, and a less than 5% chance that no real change had occurred in the patients whose change scores were more than the MDC.

The ES is a standardised measure of change over time that is calculated by dividing the difference between the pre- and post-test scores by the pre-test standard deviation (SD); in the case of Guyatt's approach, the difference is divided by the pre-test SD calculated only on stable subjects whose clinical status remained unchanged. The ES therefore represents individual change in terms of the number of pre-test SDs. It has been suggested that ES values of 0.20, 0.50, and 0.80, respectively, represent small, moderate and large changes.

The SRM (also referred to as the responsiveness-treatment coefficient or efficacy index) is the ratio between individual change and the SD of that change. It has been suggested that SRM values of 0.20, 0.50, and 0.80, respectively, represent small, moderate and large changes.

As an anchor-based method, we selected receiving operating curves (ROCs), which are very useful indicators of the relationship between a measure and an external indicator of change, such as the GPE. Responsiveness is described in terms of sensitivity (the probability that the measure correctly classifies patients who demonstrate change when an external criterion of clinical change is used) and specificity (the probability that the measure correctly classifies patients who do not demonstrate change

using the external criterion). Sensitivity and specificity of each value of change in the measure are calculated and used to plot a ROC. The values for sensitivity and false-positive rates (1-specificity) are plotted on the y and the x axis of the curve, and the area under the ROC represents the probability that a measure correctly classifies patients as improved or unchanged. This area theoretically ranges from 0.5 (no discriminating accuracy) to 1.0 (perfect accuracy). The point on the ROC curve closest to the upper left corner of the figure was taken as the MCID, which indicates the change score associated with the least misclassification. The ability of the measure to classify subjects as improved or not improved correctly was estimated and is described in terms of accuracy.

The distribution- and anchor-based methods were used considering the sample as a whole, and the two subgroups of subacute and chronic patients.

Responsiveness was also investigated by means of correlation analyses with external criteria (the SF-36 physical subscales, NRS and GPE). We tested the correlations between the outcome measures at both time points (pre- and post-rehabilitation assessment), including the GPE at follow-up. Moreover, the change scores in the ODI and RMDQ were correlated with GPE by estimating Spearman's rank order correlation coefficients, and the change scores in the SF-36 (physical activity, physical role and pain subscales) and NRS by estimating Pearson's correlation coefficients.

The Italian SPSS statistical software, version 18, was used for the statistical calculations.

Results

Subjects

Two hundred and fifteen patients were addressed, of whom 21 (10%) refused to participate. Of the 194 selected subjects, 15 dropped out before starting the rehabilitation sessions due to logistic problems (7), economic difficulties (3) or personal problems (5), and so the final study population consisted of 179 subjects (112 females, 62.6%, and 67 males, 37.4%) with a mean age of 47.7 ± 12.3 years and a median duration of pain of 6 months (interquartile range: 4 months). Table 1 shows the other sociodemographic characteristics of the study patients.

Procedures

The study procedures were well accepted by all of the patients, who did not raise any specific questions during the instruction phase or the administration of the questionnaires. None of the clinical procedures led to any problems

Table 1 Socio-demographic characteristics of the sample (*n* = 179)

Variables	Data
Age (years)	47.7 ± 12.3
Gender (M/F)	67 (37.4%)/112 (62.6%)
Married (yes/no)	121 (69.5%)/53 (30.5%)
Employed	
Yes	139 (77.7%)
No	22 (12.3%)
Retired	18 (10.1%)
Education	
Primary school	3 (1.7%)
Secondary school	30 (16.8%)
Higher education	94 (52.5%)
Degree	52 (29.1%)
Smokers (yes/no)	45 (25.1%)/134 (74.9%)
LBP duration (months)	6 (4)
Limb involvement (yes/no)	116 (64.8%)/62 (34.6%)
Drugs	
Antianxious/antidepressants	17 (9.5%)
Pain-killers	50 (27.9%)
Muscle-relaxants	16 (8.9%)
NSAIDs	45 (25.1%)
Non-spinal comorbidities (yes/no)	29 (16.2%)/150 (83.8%)

Continuous variables: mean values ± standard deviation; discrete variables: frequency (percentages); LBP duration: median value (interquartile range)

Table 2 GPE distributions

Measure	GPE category	Mean ± standard deviation		
		T0	T1	T1–T0
ODI	Improved	25.4 ± 15.1	12.7 ± 12.1	–12.67 ± 11.21
	Stable	27.7 ± 17.6	22.0 ± 17.4	–5.73 ± 9.54
	Total	26.8 ± 16.6	17.9 ± 16.0	–8.91 ± 11.2
RMDQ	Improved	5.69 ± 3.73	1.79 ± 2.80	–3.93 ± 4.00
	Stable	6.80 ± 4.63	4.64 ± 4.05	–2.17 ± 3.08
	Total	6.36 ± 4.30	3.42 ± 3.83	–2.94 ± 3.63

and all of the patients completed the rehabilitation programme. No specific issues were raised by the patients or the physiotherapists involved in the rehabilitation training.

Psychometric properties

The dichotomisation of the GPE showed that 77 subjects (43%) improved and 102 subjects (57%) did not (Table 2).

When considering the sample as a whole, the MDC was 13.67 for the ODI and 4.87 for the RMDQ. The ES for the ODI was moderate (0.53), and slightly decreased when

Table 3 Results arising from the distribution-based and anchor-based methods

Method	Value		
	Total	Subacute	Chronic
ODI			
Minimum detectable change (MDC)	13.67	15.35	12.72
Effect size (ES)	0.53	0.73	0.44
Effect size (Guyatt)	0.46	0.52	0.38
Standardised response mean (SRM)	0.80	0.93	0.76
Optimal cut-off point (AUC; sensitivity; specificity)	9.5 (0.71; 76; 63)	9.0 (0.70; 66; 68)	9.5 (0.70; 81; 59)
RMDQ			
Minimum detectable change (MDC)	4.87	4.74	4.88
Effect size (ES)	0.68	0.84	0.59
Effect size (Guyatt)	0.58	0.74	0.48
Standardised response mean (SRM)	0.81	0.95	0.74
Optimal cut-off point (AUC; sensitivity; specificity)	2.5 (0.64; 62; 55)	2.5 (0.65; 56; 57)	2.5 (0.60; 65; 53)

Guyatt’s approach was used (0.46); the ES for the RMDQ was higher (0.68), but also slightly decreased when Guyatt’s approach was used (0.58). Both the ODI and the RMDQ had larger SRMs (respectively, 0.80 and 0.81).

ROC analysis showed that the area under the curve was 0.71 (95% CI: 0.64–0.79) for the ODI and 0.64 (95% CI: 0.55–0.72) for the RMDQ. The curve of each measure was to the left above the diagonal, showing some discriminating ability; the ODI curve was closer to the upper left than that of the RMDQ. The best ODI threshold discriminating the improved and non-improved subjects was 9.5, which led to a sensitivity of 76% and a specificity of 63%; the best RMDQ threshold was 2.5 (sensitivity 62% and specificity 55%).

The accuracy of the ODI and RMDQ was, respectively, 71 and 64%.

The results in the subacute and chronic patient groups were comparable with those of the sample as a whole, with the exception of the ES and SRM estimates, which were slightly higher in the subacute patients.

All the results are summarised in Table 3 and the ROC plots are shown in Fig. 1.

The analyses made using external responsiveness criteria showed that both the ODI and the RMDQ moderately correlated with the SF-36 physical subscales and the NRS. GPE moderately correlated with the ODI, but less with the RMDQ. When considering the variables at baseline and post-rehabilitation, the correlations were confirmed with slightly higher correlation coefficients than those estimated

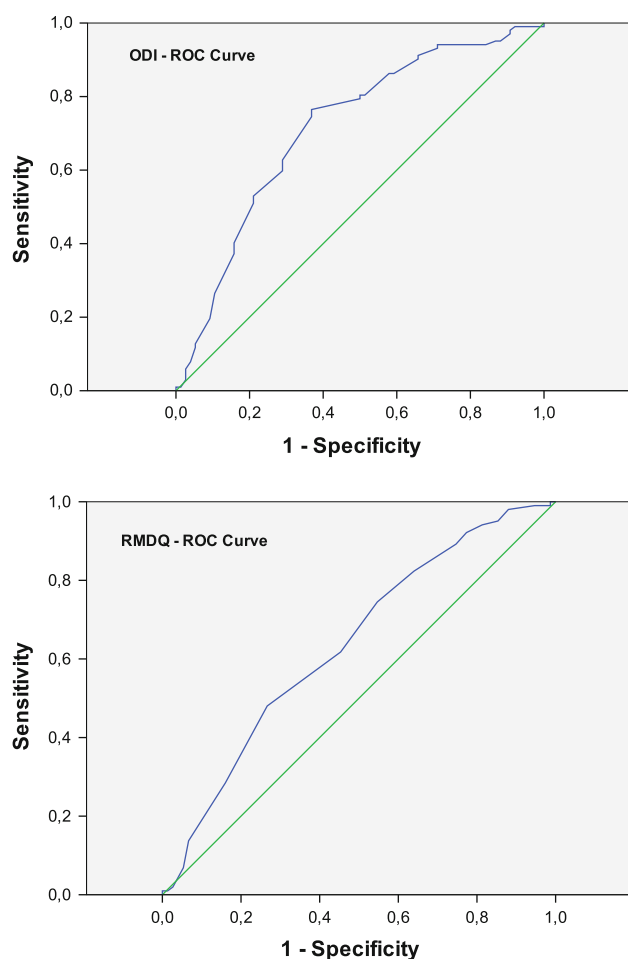


Fig. 1 ROC characteristics of the ODI and RMDQ

on the differences between the two time points, as expected. Higher correlation levels were also observed between ODI and RMDQ. Table 4 shows the full details.

Discussion

This paper describes the responsiveness of the ODI and RMDQ in a population of Italian subjects with sub-acute and chronic common LBP.

The literature is full of papers concerning the responsiveness of a measure, but there is also considerable confusion about the meaning and interpretation of the word. Responsiveness addresses the idea of clinical importance and is defined as the ability of a measure to capture clinically relevant changes over time. It is crucially useful in clinical trials and practice. It also includes the sensitivity of a measure: i.e. its ability to detect any change statistically. A number of approaches have been used to estimate responsiveness, but there is still no consensus as to which is the best [10]. Furthermore, the use of different methods often leads to large variations in the estimates within the

Table 4 Correlation analyses with external responsiveness criteria

	Correlation coefficient	Type of correlation
ODI		
Baseline ODI vs.		
VAS	0.466*	Pearson
SF-36 phys act	-0.613*	Pearson
SF-36 phys role	-0.486*	Pearson
SF-36 pain	-0.541*	Pearson
RMDQ	0.737*	Pearson
Post-rehabilitation ODI vs.		
GPE	0.451*	Spearman
VAS	0.607*	Pearson
SF-36 phys act	-0.672*	Pearson
SF-36 phys role	-0.449*	Pearson
SF-36 pain	-0.516*	Pearson
RMDQ	0.697*	Pearson
RMDQ		
Baseline RMDQ vs.		
VAS	0.502*	Pearson
SF-36 phys act	-0.669*	Pearson
SF-36 phys role	-0.499*	Pearson
SF-36 pain	-0.576*	Pearson
ODI	0.737*	Pearson
Post-rehabilitation RMDQ vs.		
GPE	0.506*	Spearman
VAS	0.671*	Pearson
SF-36 phys act	-0.692*	Pearson
SF-36 phys role	-0.515*	Pearson
SF-36 pain	-0.640*	Pearson
ODI	0.697*	Pearson
$\Delta T1-T0$		
ODI (T1-T0) vs.		
GPE	0.431*	Spearman
VAS (T1-T0)	0.539*	Pearson
SF-36 phys act	-0.404*	Pearson
SF-36 phys role	-0.332*	Pearson
SF-36 pain	-0.476*	Pearson
RMDQ (T1-T0) vs.		
GPE	0.287*	Spearman
VAS (T1-T0)	0.474*	Pearson
SF-36 phys act	-0.401*	Pearson
SF-36 phys role	-0.365*	Pearson
SF-36 pain	-0.517*	Pearson

* $p < 0.001$

same study, and there are also large variations when the same method is used to assess different studies [10]. The lack of agreement concerning a preferred index makes it difficult to compare responsiveness between studies

problematic [16], which can be considered indirect proof of the need to improve the methodology [17].

We used both distribution-based and anchor-based methods: the former are based on statistical measures, and the latter on an external criterion or “anchor” defining the important change. Both approaches led to substantially similar results: considering the sample as a whole, the responsiveness of the ODI ranged from 9 to 14 points, and that of the RMDQ from 2.5 to 5 points; considering the subacute and chronic subjects separately, the responsiveness of the ODI ranged from 9 to 15 points in the former, and from 9 to 13 points in the latter, whereas that of the RMDQ ranged from 2.5 to 5 points in both cases without any significant clinical differences.

These results are also comparable with those of other published studies. Authors in other countries have reported ODI values of 4–23 in the case of sub-acute/chronic LBP [18], 4–15 in the case of acute/chronic LBP [2, 19, 20], and 12.8 in the case of post-surgical treatment [21]. The RMDQ values range from 2.5 to 6 in acute/chronic LBP [2, 19], whereas a precise estimate of 3.5 has been found in the case of sub-acute/chronic LBP [22] and post-surgical treatment [23].

However, as in the case of previous studies, some of our methods were dependent on the patient-reported outcome, i.e. subjective perceived global effect used as an external criterion. The inclusion of patient-reported outcomes when evaluating responsiveness is important as it supplements efficacy evaluation based only on clinician judgement or laboratory tests. On the other hand, it is clear that accuracy of GPE is crucial as it may affect the reliability of the findings. We assessed GPE using a five-point Likert scale, and then considered all of the patients with a score other than 1 or 2 as “unchanged”. Nevertheless, clinically important changes would probably have been better discriminated using a seven-point scale, as indicated in other papers [18, 20, 23]. GPE was the external criterion used in the ROC analysis, but it may also have affected the estimates of the MDC and ES according to Guyatt’s method in which the intrinsic variation of the phenomenon was based on the patients categorised as “unchanged”.

The results of the ROC method, as indicated by the area under the curve, were moderate for both the ODI (0.71) and the RMDQ (0.64); the estimates for the subacute and chronic groups were comparable. These figures are in line with those reported in the literature: ODI values ranging from 0.72 to 0.80 in the case of acute/sub-acute and chronic LBP [18, 20, 24–26], and RMDQ values from 0.69 to 0.93 in the same populations [18, 20, 22, 25–28].

The optimal cut-off points estimated on the basis of ROC analysis were 9.0–9.5 for ODI and 2.5 for RMDQ, both of which are in line with those published by other authors [18, 21, 22]. However, the estimates may have

been affected by the dichotomous “unchanged” and “improved” classification required by the method and the subsequent division of the sample into sub-groups because the greater the imbalance between the sub-groups, the less reliable the estimates. Especially when the data are not normally distributed, the ROCs of sub-groups are not smooth and the optimal cut-off points tend to vary.

The ES statistics provided the same information as the ROC estimates in a manner that was easier to calculate. ES can be evaluated as a signal-to-noise ratio as the mean change in the measure is divided by the standard deviation of the change. The same considerations concerning the ease of calculation and interpretation can be applied to the SRM. We estimated moderate ESs and large SRMs for both the ODI and the RMDQ, when the sample was considered as a whole. When dividing the subjects into sub-acute and chronic groups, the former showed slightly higher estimates than the latter, probably because of the more stable condition of chronic patients. The ODI ESs reported in the literature vary from small (0.37 in chronic LBP) [20] to large (0.87 for post-surgical treatment [29]; 0.88–1 in sub-acute/chronic LBP [27]; 1.05 in chronic LBP [30]), and the reported SRM is large (0.84 in chronic LBP [30]). In the case of the RMDQ, the published ESs range from small (0.44 in chronic LBP [20]) to moderate (0.70–0.74 in sub-acute/chronic LBP [27]). Moreover, when analysing ES, it is also important to consider follow-up periods as potential moderators of responsiveness estimates: unlike in our study in which the follow-up was short, ES values tend to increase when the re-test period is longer (3–12 months) [24, 27, 29].

We also investigated responsiveness in terms of the correlations between baseline and post-rehabilitation outcome measures and between pre–post treatment changes in the ODI and RMDQ and the related changes in GPE, the SF-36 physical sub-scales, and a pain evaluation. This kind of responsiveness (which is also known as external responsiveness) reflects the extent to which changes in a measure over a specific time relate to corresponding changes in one or more reference measures [10]. In this context, the measure is not of primary interest in and of itself because what is important is the relationship between the change in the measure and the change in the external standard, and the change in the standard is generally accepted as a change in the condition of the patient. It is worth noting that external responsiveness only depends on the external standard, and not on the studied treatment or patient-reported outcome. As a result, it can be applied in a wider range of settings than the other forms of responsiveness. Our estimated correlation coefficients for the ODI and RMDQ in relation to SF-36 and the NRS were generally moderate, thus supporting the capacity of both to reflect changes in perceived effect, the quality of life, and

pain. These results are in line with previously published findings [31, 32]. In detail, the ODI–RMDQ correlations reported in the literature vary from 0.60 to 0.81 [33, 34]; ODI–NRS and RMDQ–NRS correlations, respectively, vary from 0.36 to 0.78 [34, 35] and from 0.32 to 0.73 [36, 37]; and the ODI and RMDQ correlations with SF-36 reported in the literature are moderate to high when assessing physical domains [33, 38, 39].

In conclusion, although better standard methods should be identified in order to address the issue of relevant changes, our study revealed ranges of ODI and RMDQ responsiveness in an Italian population with sub-acute or chronic LBP. These findings should be considered confirmatory as they are largely in line with other published figures, and we recommend taking them into account when evaluating patient improvement or planning clinical trials because of their ability to detect efficacious treatments.

Acknowledgments The authors would like to thank Silvia Borghi, Annalisa Generali, Caroline O’ Reilly, Sergio Parazza and Antonio Romeo for their assistance, and Kevin Smart for his help in preparing the English version of this paper.

Conflict of interest None.

References

- Revicki D, Hay RD, Cella D, Sloan J (2006) Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *JCE* 61:102–109
- Roland M, Fairbank J (2000) The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. *Spine (Phila Pa 1976)* 25(24):3115–3124
- Cleland J, Gillani R, Bienen EJ, Sadosky A (2011) Assessing dimensionality and responsiveness of outcomes measures for patients with low back pain. *Pain Pract* 11(1):57–69
- Fairbank JCT, Pynsent PB (2000) The Oswestry Disability Index. *Spine (Phila Pa 1976)* 25:2940–2953
- 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 (Phila Pa 1976)* 8(2):141–144
- Crosby RD, Kolotkin RL, Williams GR (2003) Defining clinically meaningful change in health-related quality of life. *JCE* 56:395–407
- de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LX (2006) Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 4:54
- Padua R, Padua L, Ceccarelli E, Romanini E, Zanolli 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(2):126–129
- Monticone M, Baiardi P, Ferrari S, Foti C, Mugnai R, Pillastrini P, Vanti C, Zanolli G (2009) Development of the Italian Version of the Oswestry Disability Index, ODI-I. A cross-cultural adaptation, reliability and validity. *Spine (Phila Pa 1976)* 34(19):2090–2095
- Husted JA, Cook RJ, Farewell VT, Gladman DD (2000) Methods for assessing responsiveness: a critical review and recommendations. *JCE* 53:459–468
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knola DL, Dekker J, Bouter LM, de Vet HCW (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *JCE* 60:34–42
- Huskinson EC (1974) Measurement of pain. *Lancet* 2(7889): 1127–1131
- Apolone G, Mosconi P (1998) The Italian SF-36 Survey: translation, validation and norming. *JCE* 51(11):1025–1036
- Apolone G, Mosconi P, Ware J (2000) Questionario sullo stato di salute SF-36. Manuale d’uso e guida all’interpretazione dei risultati. [SF-36 quality of life questionnaire. User’s Manual and guide to the interpretation of results]. Milan, Guerini e Associati Ed. (In Italian)
- Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ (2010) Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *JCE* 63:760–766
- Schmitt JS, Richard P, Di Fabio RP (2004) Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *JCE* 57:1008–1018
- Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Kelvin P, Jordan KP, Croft P, de Vet HCW (2010) Mind the MIC: large variation among populations and methods. *JCE* 63:524–534
- Beurskens AJHM, de Vet HCW, Koke AJA (1996) Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 65:71–76
- Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korf M, Bouter LM, de Vet HCW (2008) Interpreting change scores for pain and functional status in low back pain towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 33(1):90–94
- Coelho RA, Siqueira FB, Ferreira PH, Ferreira ML (2008) Responsiveness of the Brazilian–Portuguese version of the Oswestry Disability Index in subjects with low back pain. *Eur Spine J* 17:1101–1106
- Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY (2008) The minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and Pain Scales. *Spine J* 8(6):968–974
- Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Cano A, Muriel A, Zamora J, Gil del Real MT, Gestoso M, Mufraggi N (2007) Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine (Phila Pa 1976)* 32(25):2915–2920
- Ostelo RW, de Vet HC, Knol DL, van den Brandt PA (2004) 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *JCE* 57:268–276
- Walsh TL, Hanscom B, Lurie JD, Weinstein JN (2003) Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine (Phila Pa 1976)* 28(6):607–615
- Grotle M, Brox JI, Vøllestad NK (2004) Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine (Phila Pa 1976)* 29(21):E492–E501
- Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82(1):8–24
- Frost H, Lamb SE, Stewart-Brown S (2008) Responsiveness of a patient specific outcome measure compared with the Oswestry

- Disability Index v2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. *Spine (Phila Pa 1976)* 33(22):2450–2457
28. Stratford PW, Binkley JM, Riddle DL, Guyatt GH (1998) Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Phys Ther* 78(11):1186–1196
 29. Mannion AF, Junge A, Grob D, Dvorak J, Fairbank JC (2006) Development of a German version of the Oswestry Disability Index. Part 2: sensitivity to change after spinal surgery. *Eur Spine J* 15:66–73
 30. Changulani M, Shaju A (2009) Evaluation of responsiveness of Oswestry low back pain disability index. *Arch Orthop Trauma Surg* 129(5):691–694
 31. Hashimoto H, Komagata M, Nakai O, Morishita M, Tokuhashi Y, Sano S, Nohara Y, Okajima Y (2006) Discriminative validity and responsiveness of the Oswestry Disability Index among Japanese outpatients with lumbar conditions. *Eur Spine J* 15:1645–1650
 32. Wittink H, Turk DC, Carr DB, Sukiennik A, Rogers W (2004) Comparison of the redundancy, reliability, and responsiveness to change among SF-36, Oswestry Disability Index, and multidimensional pain inventory. *Clin J Pain* 20(3):133–142
 33. Grotle M, Brox JI, Vollestad NK (2003) Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. *J Rehabil Med* 35:241–247
 34. Yakut E, Duger T, Oksuz C, Yorukan S, Ureten K, Turan D, Firat T, Kiraz S, Kırdı N, Kayıhan H, Yakut Y, Guler C (2004) Validation of the Turkish Version of the Oswestry Disability Index for patients with low back pain. *Spine (Phila Pa 1976)* 29(5):581–585
 35. Mannion AF, Junge A, Fairbank JC, Dvorak J, Grob D (2005) Development of a German version of the Oswestry Disability Index. Part 1: cross-cultural adaptation, reliability, and validity. *Eur Spine J* 15:55–65
 36. Máaroufi H, Benbouazza K, Faik A, Bahiri R, Lazrak N, Abouqal R, Amine B, Hajjaj-Hassouni N (2007) Translation, adaptation, and validation of the Moroccan version of the Roland Morris Disability Questionnaire. *Spine (Phila Pa 1976)* 32(13):1461–1465
 37. Monteiro J, Faísca L, Nunes O, Hipólito J (2010) Roland Morris disability questionnaire, adaptation and validation for the Portuguese speaking patients with back pain. *Acta Med Port* 23(5):761–766
 38. Mousavi SJ, Parnianpour M, Mehdian H, Montazeri HA, Mobini B (2006) The Oswestry Disability Index, the Roland-Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: Translation and Validation Studies of the Iranian Versions. *Spine (Phila Pa 1976)* 31(14):E454–E459
 39. Fujiwara A, Kobayashi N, Saiki K (2003) Association of the Japanese Orthopaedic Association Score With the Oswestry Disability Index, Roland-Morris Disability Questionnaire, and Short-Form 36. *Spine (Phila Pa 1976)* 28:1601–1607