

# Cross-cultural adaptation and validation of the Neck Bournemouth Questionnaire in the Italian population

Tommaso Geri · Alessio Signori · Silvia Gianola ·  
Giacomo Rossetini · Gisel Grenat ·  
Giovanni Checchia · Marco Testa

Accepted: 15 September 2014 / Published online: 21 September 2014  
© Springer International Publishing Switzerland 2014

## Abstract

**Purpose** The aim of this study was to cross-culturally adapt the Neck Bournemouth Questionnaire in Italian (NBQ-I) and validate its psychometric properties in a sample of subjects with chronic neck pain.

**Methods** The NBQ-I was developed according to international standards. The psychometric testing included the content validity, assessed by considering the relevance and comprehensiveness of the items, the structural validity by factor analysis, the construct validity and the responsiveness by mean of hypotheses testing process comparing, respectively, the test scores and the changes scores of the NBQ-I with those of the Neck Pain and Disability Scale, the Numerical Rating Scale for Pain, the EuroQoL 5 Dimension Index and its numerical rating scale. Finally, we

tested the reliability by internal consistency (Cronbach's  $\alpha$ ) and the interpretability by calculating the Minimal Clinical Important Difference (MCID).

**Results** The NBQ-I had acceptable psychometric characteristics. A total of 96 subjects with chronic neck pain completed the questionnaire's administration. The high relevance and comprehensiveness of the items pointed out acceptable face validity. The construct validity analysis was based on the structural validity, which revealed a two factors structure explaining 69.2 % of variance, and on the hypotheses testing process, which showed a moderate validity. The internal consistency was acceptable ( $\alpha = 0.89$ ). The responsiveness, assessed with the hypotheses testing process, was moderate. The MCID was 5.5 points.

**Conclusions** The validation process revealed acceptable psychometric properties of the NBQ, whose Italian version can be used for research and clinical purposes.

---

T. Geri (✉) · G. Rossetini · M. Testa  
Department of Neuroscience, Rehabilitation, Ophthalmology,  
Genetics, Maternal and Child Health, University of Genova,  
Campus of Savona, Savona, Italy  
e-mail: tommaso.geri@gmail.com

A. Signori  
Section of Biostatistics, Department of Health Sciences,  
University of Genova, Genoa, Italy

S. Gianola  
Clinical Epidemiology Unit, IRCCS Galeazzi Orthopaedic  
Institute, Milan, Italy

S. Gianola  
Department of Health Science, Center of Biostatistics for  
Clinical Epidemiology, University of Milano-Bicocca, Monza,  
Italy

G. Grenat · G. Checchia  
Physical Medicine and Rehabilitation Department, Santa Corona  
Hospital, Pietra Ligure, Italy

**Keywords** Neck pain · Questionnaires · Outcome assessment · Neck Bournemouth Questionnaire · Validation studies

## Introduction

Neck pain (NP) is a common musculoskeletal disorder causing a limitation in activities ranging from 2 to 11 % in the population suffering from this condition [1]. A meta-analysis on the prognosis of NP reports that pain and disability are still present 1 year after onset [2]. The subsequent health-care utilization and work absenteeism lead to high economic and societal costs [3–5] whose reduction may be achieved by further understanding of how this disease affects the quality of life.

The complex etiopathogenesis of NP [1] mandates a biopsychosocial approach like the one purposed by the

international classification of functioning, disability and health (ICF) [6] which is best investigated through the use of multi-dimensional questionnaires. The cross-cultural adaptation and validation of questionnaires in different languages, besides allowing their use among several countries, may help the development of common strategies to limit the burden of NP across populations.

Among the questionnaires used for measuring the outcomes in people with NP, Ferreira et al. [7] found the Neck Disability Index [8], the Neck Pain and Disability Scale (NPDS) [9], and the Neck Bournemouth Questionnaire (NBQ) [10] to fit suitably the ICF categories. The NBQ is a short-form questionnaire constituted by seven items representing important aspects of the biopsychosocial model. These are encoded in the ICF framework as emotional functioning, sensation of pain, housework, managing daily routine, remunerative and non-remunerative employment, community life, and recreation. Each item is scored on a 0–10 numerical scale, where zero represents absence of limitation, for a total of 70 points [10].

As the NBQ holds validated version in a population of chronic NP patients in English [10], German [11], and French [12] languages and in a population of patients with whiplash associated disorders in Dutch [13], the aim of the present study was to cross-culturally adapt and validate this questionnaire in a sample of Italian chronic NP patients.

## Materials

This study was approved by the Ethical Committee of the Azienda Sanitaria Locale 2 Savonese (no. 650-04/07/2013), and permission was obtained from the designer of the original version. The questionnaire was administered at the beginning and at the end of a 4-week physiotherapy treatment program constituted by re-training of breathing (5'), self-mobilization (10'), stretching (10') and strengthening (10') of the neck and shoulder girdle muscles, and self-massage in the pain area (5'). The validation of the NBQ-I was structured using the taxonomy and terminology suggested by the COSMIN [14].

### Setting and subjects

The participants were recruited in a convenience way among patients attending the outpatient physiotherapy service of the Santa Corona Hospital and affiliated centers from September 2012 to September 2013. The inclusion criteria were: age >18 years, the ability to read and speak Italian fluently, and chronic non-specific NP (>3 months) assessed by a medical doctor specialized in physical and rehabilitation medicine with 20 years of experience. The exclusion criteria were: specific NP, psychiatric and mental

deficits, central or peripheral neurological signs, systemic illness, clinical instability (cardiac, respiratory, vascular), and vertebral surgery. Ten patients per item were chosen to estimate the sample size [15].

### Translation and cross-cultural adaptation

This process was performed according to well-accepted guidelines [16]. Two Italian mother-tongue independent translators having an excellent knowledge of English and different cultural profiles (medical and humanities) provided the forward translations of the NBQ and reported about any problem arisen during the translation. To gain the development of a culturally adapted version, one of the investigators synthesized the two forward translations in a common version on the basis of the reports and personal opinions of the two translators. Afterward, two independent bilingual translators with English as their mother tongue back-translated the common version, with the aim of reflecting the common English wording, and reported about any problem emerged. They were selected because living in Italy, unaware about the explored concepts and out of medical background. Then, a bilingual committee composed by two clinicians (physiotherapists), and the translators obtained the pre-final version after reviewing all the translations and the reports and considering all the items to reach semantic, idiomatic, experiential and conceptual equivalence. Thirty-two subjects with chronic NP filled up the pre-final version and underwent a semi-structured interview about problems in wording and answering. All the findings were re-evaluated by the committee, although no further adjustment was required. Finally, a coordinating committee reviewed the pre-final version and all the documents produced during each step. Hence, the final Italian version of the questionnaire (NBQ-I) was adopted.

### Comparator instruments

The outcome measures used to test the construct validity and responsiveness of the NBQ-I were the Italian versions of the Neck Pain and Disability Scale (NPDS-I) [17] and of the European Quality of Life 5 Dimensional scale (EQ5D) [18], the Numerical Rating Scale for pain intensity (NRS-PAIN) [19], and the Global Perceived Effect (GPE).

The NPDS-I has 20 items, each scoring between 0 (normal function) and 5 (the worst possible situation) and is divided in three sub-scales: neck dysfunction related to general activities, neck dysfunction related to activities of the cervical spine and neck pain, and cognitive-behavioral aspects. Its total score ranges from 0 (no disability) to 100 (maximum disability) [17]. The EQ5D is divided in five qualitative scales related to five dimensions of the quality of life (mobility, self-care, activities of daily living, pain,

and anxiety/depression). Each dimension is graded on a three-category ordinal scale where the lowest score represents absence of the problem while the highest a condition extremely compromised. According to the instrument's score, we obtained a numerical index (IND-EQ5D) where 1,000 correspond to the best perception of quality of life. In addition, a numerical rating scale (0–100 points where 0 is the worst possible condition) measures the health status (NRS-EQ5D) [18]. The NRS-PAIN ranks the pain intensity on a 11-point rating scale ranging from 0 (no pain at all) to 10 (the worst imaginable pain) [19]. The GPE was administered after the 4-week physiotherapy treatment to measure the effect of the intervention on patients' health status perception. This Likert scale had five response options (+2 = very much improved, +1 = much improved, 0 = No change, -1 = much worse, -2 = very much worse).

### Data analysis

The subjects' clinical and sociodemographic characteristics were described using mean values and standard deviations or counts and percentages. The data analysis was performed using the statistical software R [20]. Specific analyses for each measurement property are described below [14].

#### Content validity

During the first administration, subjects completed a semi-structured interview provided to test the content validity. The subjects' judgement was used to investigate the relevance of the items for the study population, therefore, obtaining an assessment of the face validity. The comprehensiveness was evaluated considering the number of missing or multiple responses [21].

#### Structural validity

An Exploratory Factor Analysis (EFA) to test the dimensionality of the scale was performed after having considered if the Bartlett's test of sphericity, assessing if the correlation matrix of data is an identity matrix with all items unrelated, was significant and the Kaiser–Meyer–Olkin (KMO) criterion, which tests the sampling adequacy to ensure that the scale items are relevant for factorial analysis, was  $\geq 0.80$ . After this assessment, the Cattell's Scree Test was used to determine the number of extracting factors (eigenvalues  $> 1$ ). We used the Maximum Likelihood (ML) method to extract factors and the Promax oblique method to obtain rotated solutions. Community of each item and its loading on the owning factor were reported and discussed after extraction to verify the amount of variance explained by the extracted factors for each original variable.

Subsequently, we performed a Confirmatory factor analysis (CFA) using the ML method to test the hypothesized model with the number of factors detected using EFA. The fit indices used to evaluate the model fit were the chi-square ( $\chi^2$ ) test, which indicates a good fit when the comparison between the fitted model with the saturated model that fits the covariances perfectly is not significant, and the Root Mean Square Error of Approximation (RMSEA), which should be close to zero and however  $\leq 0.05$  for a good fit. Further, the lower and upper bounds of RMSEA 90 % confidence interval should be, respectively, lower than 0.05 and higher than 0.10 [22]. Both the Comparative Fit Index (CFI) and the Tucker–Lewis Index (TLI) should be higher than 0.90 or 0.95 for, respectively, acceptable or good model fit [23].

For possible model mis-specifications, the expected parameter changes (EPC) and modification index (MI) were calculated if the fitting of the original model was unsatisfactory.

#### Hypotheses testing for construct validity and responsiveness

The construct validity was assessed by means of hypotheses testing using the correlations between the observed test scores of the questionnaires in both the administrations. Also, the validity of responsiveness was assessed by means of the same hypotheses testing process used for the construct validity, with the only difference that the correlations were based on the change scores of the questionnaires, calculated as the difference between the test scores obtained in the two administrations. The Pearson's  $r$  or Spearman's  $\rho$  were the correlation coefficients used according to data distribution checked with the Shapiro–Wilk test.

Our analysis for both the construct validity and the responsiveness concerned primarily the absolute magnitude of scores of the entire questionnaires. Our a priori hypotheses were that the correlations of the NBQ-I with the NPDS-I and with the NRS-PAIN were positive with a magnitude higher than 0.60, while the correlations of the NBQ-I with the IND-EQ5D and with the NRS-EQ5D were negative and ranging between  $-0.30$  and  $-0.60$ . Further, we tested the relative magnitude of the entire questionnaires and their sub-scales. Consequently, we generated a list of hypotheses that stated which was the higher between two correlations and the respective magnitude. The rationale for comparing two correlations between questionnaires or their sub-scales was that more similar are the constructs supposed to be measured less would be their difference, and vice versa. A difference of 0.1 point was attributed when questionnaires or subscales measuring the same construct were present in both terms of the correlation. An example is when the subscale 1 of the NBQ was compared with subscales of the NPDS (e.g., hypotheses

10 and 11, Tables 3, 4). A difference of 0.2 points was deemed when in one term of the relative correlation there was the NRS-PAIN. A difference of 0.3 points was hypothesized when the questionnaire or subscales measuring different construct were compared. For example, the correlation of subscales measuring the same construct of the NBQ-I and the NPDS-I was considered 0.3 points higher than the correlation between the sub-scales of NBQ-I and the sub-scales of the IND-EQ5D (e.g., hypotheses 13–16, Table 3) because the latter measures constructs concerning basic aspects of quality of life which were supposed to be not directly affected by the presence or absence of NP (e.g., the three response options for item 1, which measure mobility, are: no difficulty in walking and some difficulty in walking and bedridden). The hypotheses testing of the construct validity and the responsiveness were both rated according to the criteria proposed by de Boer et al. [24], which state the validity of these psychometric properties is high if less than 25 % of the hypotheses are refuted, moderate if 25–50 % is refuted, and poor if more than 50 % of the hypotheses are refuted.

#### Internal consistency

This property, calculated after the factor analysis, was assessed using Cronbach's  $\alpha$  and values ranging between 0.70 and 0.90 were judged acceptable [25]. Correlation of each item with its own subscale total score (item-test), with the total score of remaining items (item-rest) and inter-items correlation were also reported.

#### Interpretability

The total scores on the pre- and post-treatment conditions and the change scores were summarized using descriptive statistics. The floor and ceiling effects were deemed present when more than 15 % of the patients received the lowest or highest possible scores [26]. Once the questionnaire was completed, the time needed was registered.

The Minimal Clinical Important Difference (MCID) was computed using the ROC curves [27] by comparison of NBQ change scores with the gold standard represented by the subject's GPE. The treatment was considered beneficial for responses on GPE of +2 and +1. We considered an area under the curve (AUC) of at least 0.70 to be adequate [28].

## Results

### Subjects

A total of 108 subjects participated in the study, of which 80 were women (74.1 %) and 28 were men (25.9 %). The mean age was 51.5 (SD = 13.6) and the mean duration of

**Table 1** Sociodemographic characteristics

Variables	Numbers	%
Sex		
Men	28	25.93
Women	80	74.07
Marital Status		
Not reported	3	2.78
Married	65	60.19
Unmarried	40	37.04
Education		
Elementary	7	6.48
Middle school	19	17.59
Upper school	53	49.07
University	29	26.85
Employment		
Not reported	2	1.85
Unemployed	2	1.85
Student	4	3.70
Employed	61	56.48
Self-employed	9	8.33
Pensioner	24	22.22
Housewife	6	5.56
Smoking		
Yes-current smoker	20	18.52
No	88	81.48
Symptoms irradiation		
Not reported	1	0.93
Yes	46	42.59
No	61	56.48
Drugs		
Anxiolytics/antidepressants	3	2.78
Analgesics	28	25.93
Muscle relaxant	4	3.70
NSAIDs	3	2.78
None	61	56.48
Poly-assumption	9	8.34
Comorbidities		
Heart disease	6	5.56
Respiratory disease	1	0.93
Endocrine disease	8	7.41
Enteric disease	5	4.63
Anxiety/depression	5	4.63
Other musculoskeletal system diseases	12	11.11
None	45	41.67
Poly-disease	26	24.11

NSAIDs Non Steroidal Anti-Inflammatory Drugs, % percentage

symptoms was 12.3 months (SD = 7.5). On average, the sample's BMI was 24.0 (SD = 3.7). The sociodemographic characteristics are reported in Table 1.

At the end of the questionnaire's administration, a drop out rate of 9.2 % was observed. These subjects abandoned the treatment for personal reasons. The response rates to the GPE were 23.1 % (+2), 45.4 % (+1), 16.7 % (0), 2.8 % (−1), and 12.0 % did not answer (of which 9.2 % dropped-out).

#### Translation and cross-cultural adaptations

The cross-cultural adaptation of the NBQ in the Italian population was reached without any problem. All the items were translated linearly and the experts agreed during all phases of the process, even though a consensus about semantic, idiomatic and conceptual equivalence was needed.

The interview about comprehension issues delivered to each participant after the administration of the pre-final version highlighted no significant flaws in the questionnaire. Therefore, the coordinating committee adopted the final version of the NBQ-I without any changes. The 32 subjects were enrolled in further administration of the questionnaires for the validation of its psychometric properties.

#### Content validity

The questionnaire had acceptable face validity because a considerable amount of subjects judged the questionnaire relevant for their health problem (78.7 %) or for other people with neck pain (87.9 %). Further, the response rates to the question "How much is your health problem represented in the NBQ questionnaire?" on a Likert scale with answers' options "A lot", "Enough", "Normal", "A little" and "Not at all" were 10.1, 46.3, 25.9, 13.9, and 0.0 %, respectively (3.7 % did not answer). The distribution of the total scores showed the absence of missing or multiple responses.

#### Structural validity

A KMO coefficient of 0.83 and statistical significance ( $p < 0.001$ ) of Bartlett's test for sphericity allowed to perform the factor analysis. Only two factors had an eigenvalue higher than 1 and the first explained 56.6 % of all variance, while the second 12.6 %. Table 2 shows the item-factor loadings using oblique Promax rotation and the single item communalities. Items 1, 2, 3, 6, and 7 had a high load on factor 1 and items 4 and 5 on factor 2. All items but 7 had communality higher than 60 %. The two factors showed a correlation of 0.52. The subsequent CFA revealed an optimum fit of the model with the two factors observed in explorative analysis. In fact, a not significant  $\chi^2$  test ( $\chi^2 = 16.37$ ;  $p = 0.23$ ), a RMSEA of 0.049 [90 % CI 0.00–0.11; test for RMSEA  $\leq 0.05$  ( $p$  value) = 0.46], a

**Table 2** Loadings obtained from the factor analysis

Items	Factor 1	Factor 2	Communalities
1	<b>0.82</b>	−0.01	0.65
2	<b>0.97</b>	−0.05	0.89
3	<b>0.82</b>	0.10	0.76
4	0.12	<b>0.85</b>	0.84
5	−0.03	<b>0.92</b>	0.82
6	<b>0.73</b>	0.10	0.61
7	<b>0.50</b>	0.06	0.28

The bold point out which item is belonging to each factor

CFI of 0.993, and a TLI of 0.989 were detected, all indicating a good fit of the hypothesized model.

The factors were named pain and functioning (factor 1; items 2, 1, 3, 6, and 7) and anxiety and depression (factor 2; items 5 and 4).

#### Hypotheses testing for construct validity

The results of the hypotheses testing process are reported in the fourth and sixth column of Table 3 for the condition before and after treatment, respectively. Before treatment, the expected absolute magnitudes of correlations were all respected but the one with IND-EQ5D. Regarding the 41 hypotheses concerning the relative magnitude of the entire questionnaires and their sub-scales, 15 were rejected. In total, the hypotheses testing process rejected 35.5 % of the generated hypotheses.

After treatment, the expected absolute magnitudes were all confirmed. The hypotheses testing process about the expected directions of the relative magnitude rejected 14 (31.1 %) out of 45 hypotheses.

Considering the percentage of hypotheses refuted in the pre- and post-treatment conditions, the construct validity was moderate.

#### Hypotheses testing for responsiveness

The results of the hypotheses testing for this property are reported in Table 4. The hypothesized absolute magnitudes were all rejected but the correlation with the change score of IND-EQ5D. Among the 41 hypotheses on the relative magnitude, 17 were refuted. In total, 44.4 % of the generated hypotheses were rejected and the responsiveness was rated moderate.

#### Internal consistency

Cronbach's  $\alpha$  was 0.89 (95 % CI 0.84–0.92) for the entire questionnaires. When analysing the two sub-scales, the coefficients were 0.88 (95 % CI 0.83–0.92) and 0.90 (95 % CI 0.86–0.94) for factor 1 and 2, respectively.

**Table 3** Hypotheses testing—construct validity pre-treatment and post-treatment

Hypothesized directions		Hypothesized magnitude	Observed magnitude PRE	Test PRE	Observed magnitude POST	Test POST
<i>Absolute correlations of entire questionnaires</i>						
1	NBQ with NPDS $\geq$	0.60	0.67	T	0.70	T
2	NBQ with NRS-PAIN $\geq$	0.60	0.63	T	0.73	T
3	NBQ with IND-EQ5D in a range between	(−0.30, −0.60)	−0.66	F	−0.52	T
4	NBQ with NRS-EQ5D in a range between	(−0.30, −0.60)	−0.43	T	−0.33	T
<i>Relative correlations between the entire questionnaires</i>						
5	NBQ with NPDS $\geq$ NBQ with NRS-PAIN	0.20	0.04	F	−0.04	F
6	NBQ with NPDS $\geq$ NBQ with IND-EQ5D	0.30	0.01	F	0.17	F
7	NBQ with NPDS $\geq$ NBQ with NRS-EQ5D	0.30	0.24	F	0.36	T
8	NBQ with NRS-PAIN $\geq$ NBQ with IND-EQ5D	0.20	−0.03	F	0.21	T
9	NBQ with NRS-PAIN $\geq$ NBQ with NRS-EQ5D	0.20	0.20	T	0.40	T
<i>Relative correlations between sub-scales (in brackets) of the questionnaires</i>						
10	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with NPDS (3)	0.10	0.02	F	0.09	F
11	NBQ (1) with NPDS (2) $\geq$ NBQ (1) with NPDS (3)	0.10	0.12	T	0.12	T
12	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with NRS-PAIN	0.20	−0.04	F	−0.08	F
13	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (1)	0.30	0.21	F	0.39	T
14	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (2)	0.30	0.30	T	0.36	T
15	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (4)	0.30	0.10	F	0.38	T
16	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (5)	0.30	0.42	T	0.57	T
17	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with NRS-EQ5D	0.30	0.27	F	0.43	T
18	NBQ (1) with NRS-PAIN $\geq$ NBQ (2) with NRS-PAIN	0.10	0.35	T	0.36	T
19	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (1)	0.20	0.25	T	0.47	T
20	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (2)	0.20	0.34	T	0.44	T
21	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (3)	0.20	0.09	F	0.30	T
22	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (5)	0.20	0.46	T	0.65	T
23	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with NRS-EQ5D	0.20	0.32	T	0.51	T
24	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (1)	0.10	0.15	T	0.17	T
25	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (2)	0.10	0.25	T	0.15	T
26	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (4)	0.10	0.05	F	0.16	T
27	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (5)	0.10	0.36	T	0.35	T
28	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with NRS-EQ5D	0.10	0.22	T	0.21	T
29	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with EQ5D (1)	0.10	0.10	T	0.01	F
30	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with EQ5D (2)	0.10	0.20	T	−0.01	F
31	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with EQ5D (5)	0.10	0.31	T	0.19	T
32	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with NRS-EQ5D	0.10	0.17	T	0.05	F
33	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NPDS (1)	0.10	0.14	T	0.03	F
34	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NPDS (3)	0.10	0.10	T	0.17	T
35	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NRS-PAIN	0.20	0.22	T	0.08	F
36	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (1)	0.30	0.24	F	0.25	F
37	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (2)	0.30	0.40	T	0.23	F
38	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (3)	0.30	0.22	F	0.21	F
39	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (4)	0.30	0.22	F	0.22	F
40	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NRS-EQ5D	0.30	0.16	F	0.12	F
41	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (1)	0.10	0.30	T	0.27	T
42	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (2)	0.10	0.45	T	0.25	T
43	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (3)	0.10	0.27	T	0.23	T
44	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (4)	0.10	0.27	T	0.24	T
45	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with NRS-EQ5D	0.10	0.21	T	0.14	T
Total amount of refuted hypotheses				16/45	14/45	
Percentage of refuted hypotheses				35.5 %	31.1 %	

*EQ5D* EuroQoL 5 Dimensions, *F* False, *IND-EQ5D* Index-EuroQoL 5 Dimensions, *NBQ* Neck Bournemouth Questionnaire, *NPDS* Neck Pain and Disability Scale, *NRS-PAIN* Numerical Rating Scale for pain intensity, *NRS-EQ5D* Numerical Rating Scale of the EuroQoL 5 Dimensions for health status, *T* True

The item-test correlation was higher than 0.80 for all items but 7 ( $r = 0.69$ ) and, similarly, the item-rest correlations were higher than 0.70 (for item 7:  $r = 0.53$ ). Most of the correlations between items were higher than 0.50 with lower correlation between both items 4 and 5 and all other items and between item 7 and all other items with values ranging from 0.40 to 0.50 except for items 4 and 5 ( $r = 0.29$  and  $0.32$ , respectively).

### Interpretability

The median and Interquartile Range (IQR) of the entire questionnaire (maximum score = 70) in the pre- and post-treatment conditions were, respectively, 29 (IQR = 14–37.5) and 15 (IQR = 8–24.2). The change score had mean of 9.4 and median of 8 with IQR ranging from 1 to 21.

Factor 1 (maximum score = 50) had, respectively, median 21 (IQR = 10–28.5) and 12 (IQR = 6–19) in the pre- and post-treatment conditions. Its change score had mean 7.1 and median 5 (IQR = 1–14.5).

The median and IQR for factor 2 (maximum score = 20) were 6 (IQR = 2–12) in the pre-treatment condition and 4 (IQR = 1–6) in the post-treatment condition. The change score had mean of 2.4 and median 2 (IQR = 0–5). Factor 2 further showed a floor effect after the treatment because 19.4 % reached the lowest score. The distribution of test and change scores is summarized in Table 5.

The time needed to fulfil the whole questionnaire had median 1' 67" (IQR = 1'15"–2'25").

The results of the ROC analysis indicated an MCID of 5.5 points. This cut-off score was the most appropriate to detect a patient-reported improvement with a sensitivity of 0.75, a specificity of 0.60, and an area under the ROC curve of 0.72 (95 % CI 0.61–0.83).

### Discussion

The cross-cultural adaptation of the NBQ in Italian showed the absence of any major difficulty during the forth and back translation processes. The review of the expert committee obtained further semantic, idiomatic and conceptual equivalence. The test of the pre-final version confirmed that a reasonable translation was reached because comprehension and equivalence issues were avoided. Also, the results of the face validity pointed out the representation of NP problems in the questionnaire.

The construct validity was firstly investigated through the analysis of structural validity. The factor analysis revealed the NBQ-I is based on a formative model composed by two different unidimensional reflective sub-scales dealing with pain and functioning on the one hand and

anxiety and depression on the other. Further, comparison with other versions was not feasible since this is the first study reporting on this property of the NBQ. Although this questionnaire is claimed to cover several aspects of people's quality of life because it has items asking about them, the results of factor analysis revealed a robust two factors structure. Our recommendation is therefore, to rely on the results of structural analysis of a questionnaire rather than on the content of its single items when considering the number of sub-scales present in a multi-dimensional questionnaire.

The absolute magnitude of correlation of the NBQ with the NPDS (pre-treatment = 0.67, post-treatment = 0.70) were in line with the coefficients provided by the German version (pre-treatment = 0.69, post-treatment = 0.80) [11]. However, we further assessed this property with the hypothesis testing process, which allowed us to gain further insights into the construct validity of the questionnaire. Consequently, we speculated on the relative magnitudes and directions that the correlations of the NBQ and its two sub-scales could have with the other instruments and their sub-scales. It was surprising the lack of difference between the correlation of NBQ/NPDS with the correlation of NBQ/NRS-PAIN or of the correlation between NBQ-factor 1/NPDS-factor 1, i.e., intended to be very similar, with the correlation NBQ-factor 1/NRS-PAIN. These results were confirmed with the post-treatment test scores as well. Therefore, we guessed that the NBQ-I and its factor 1, regarding pain and functioning, are well influenced by pain intensity.

The Cronbach's  $\alpha$  of the entire questionnaire ( $\alpha = 0.89$ ) was slightly less than the predefined threshold, even though it was similar to the values obtained in English ( $\alpha = 0.87$ ) and German ( $\alpha = 0.79$ ). This indicates a high interrelatedness of the items with a slight tendency to redundancy. The internal consistency calculated for the two sub-scales, derived from the factor analysis, revealed a similar pattern. While for factor 2 the high redundancy may be attributable to the overlapping of feelings like anxiety and depression, the results of the CFA may indicate item 7 as unnecessary in factor 1. Considering that the Cronbach's  $\alpha$  is strongly correlated with the length of the scale [25], the high coefficients obtained for both the full questionnaire and its sub-scales further support the high degree of homogeneity of the NBQ.

The hypotheses-testing method was chosen to evaluate the ability of the NBQ-I to detect change scores over time. Although the scores obtained in the other languages suggested a good responsiveness [10–12], our results indicated it was moderate. Despite the threshold used for the comparison between correlations, regarding both absolute and relative magnitude, might have been too high, the rationale for choosing the cut-off value of difference was

**Table 4** Hypotheses testing—responsiveness

Hypothesized directions		Hypothesized magnitude	Observed magnitude	Test
<i>Absolute correlations of entire questionnaires</i>				
1	NBQ with NPDS $\geq$	0.60	0.51	F
2	NBQ with NRS-PAIN $\geq$	0.60	0.58	F
3	NBQ with IND-EQ5D in a range between	(−0.30, −0.60)	−0.33	T
4	NBQ with NRS-EQ5D in a range between	(−0.30, −0.60)	−0.19	F
5	NBQ with GPE $\geq$	0.60	0.45	F
<i>Relative correlations between the entire questionnaires</i>				
6	NBQ with NPDS $\geq$ NBQ with NRS-PAIN	0.20	−0.06	F
7	NBQ with NPDS $\geq$ NBQ with IND-EQ5D	0.30	0.18	F
8	NBQ with NPDS $\geq$ NBQ with NRS-EQ5D	0.30	0.32	T
9	NBQ with NRS-PAIN $\geq$ NBQ with IND-EQ5D	0.20	0.25	T
10	NBQ with NRS-PAIN $\geq$ NBQ with NRS-EQ5D	0.20	0.39	T
<i>Relative correlations between sub-scales (in brackets) of the questionnaires</i>				
11	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with NPDS (3)	0.10	0.07	F
12	NBQ (1) with NPDS (2) $\geq$ NBQ (1) with NPDS (3)	0.10	0.17	T
13	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with NRS-PAIN	0.20	−0.13	F
14	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (1)	0.30	0.29	F
15	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (2)	0.30	0.38	T
16	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (4)	0.30	0.19	F
17	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with IND-EQ5D (5)	0.30	0.35	T
18	NBQ (1) with NPDS (1) $\geq$ NBQ (1) with NRS-EQ5D	0.30	0.34	T
19	NBQ (1) with NRS-PAIN $\geq$ NBQ (2) with NRS-PAIN	0.10	0.43	T
20	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (1)	0.20	0.52	T
21	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (2)	0.20	0.27	T
22	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (3)	0.20	0.49	T
23	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with EQ5D (5)	0.20	0.48	T
24	NBQ (1) with NRS-PAIN $\geq$ NBQ (1) with NRS-EQ5D	0.20	0.25	T
25	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (1)	0.10	0.15	T
26	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (2)	0.10	0.25	T
27	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (4)	0.10	0.05	F
28	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with EQ5D (5)	0.10	0.22	T
29	NBQ (1) with EQ5D (3) $\geq$ NBQ (1) with NRS-EQ5D	0.10	0.21	T
30	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with EQ5D (1)	0.10	0.10	T
31	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with EQ5D (2)	0.10	0.19	T
32	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with EQ5D (5)	0.10	0.16	T
33	NBQ (1) with EQ5D (4) $\geq$ NBQ (1) with NRS-EQ5D	0.10	0.15	T
34	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NPDS (1)	0.10	0.08	F
35	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NPDS (3)	0.10	0.19	T
36	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NRS-PAIN	0.20	−0.05	F
37	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (1)	0.30	0.27	F
38	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (2)	0.30	0.21	F
39	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (3)	0.30	0.06	F
40	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with EQ5D (4)	0.30	0.02	F
41	NBQ (2) with NPDS (2) $\geq$ NBQ (2) with NRS-EQ5D	0.30	0.00	F
42	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (1)	0.10	0.22	T
43	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (2)	0.10	0.17	T
44	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (3)	0.10	0.01	F
45	NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with EQ5D (4)	0.10	−0.02	F



**Table 4** continued

Hypothesized directions	Hypothesized magnitude	Observed magnitude	Test
46 NBQ (2) with EQ5D (5) $\geq$ NBQ (2) with NRS-EQ5D	0.10	−0.05	F
Total amount of refuted hypotheses			21/46
Percentage of refuted hypotheses			45.6 %

*EQ5D* EuroQoL 5 Dimensions, *F* False, *GPE* Global Perceived Effect, *IND-EQ5D* Index-EuroQoL 5 Dimensions, *NBQ* Neck Bournemouth Questionnaire, *NPDS* Neck Pain and Disability Scale, *NRS-PAIN* Numerical Rating Scale for pain intensity, *NRS-EQ5D* Numerical Rating Scale of the EuroQoL 5 Dimensions for health status, *T* True

**Table 5** Description of scores

	Mean	SD	Median	1st Qu	3rd Qu	Minimum Score	Maximum score	Floor Effect (%)	Ceiling Effect (%)	<i>n</i>
Test score PRE	27.73	15.91	29.00	14.00	37.50	3	69	1.87	0.93	107
Test score PRE F1	20.53	12.04	21.00	10.00	28.50	0	50	0.93	1.87	107
Test score PRE F2	7.31	5.88	6.00	2.00	12.00	0	20	12.96	1.85	108
Test score POST	17.79	12.66	15.00	8.00	24.25	0	59	2.08	1.04	96
Test score POST F1	13.21	9.55	12.00	6.00	19.00	0	49	5.21	1.04	96
Test score POST F2	4.54	4.31	4.00	1.00	6.00	0	17	19.39	2.04	98
Change score	9.39	13.33	8.00	1.00	21.00	−29	39	1.05	1.05	95
Change score F1	7.07	10.14	5.00	1.00	14.50	−18	29	1.05	1.05	95
Change score F2	2.41	5.27	2.00	0.00	5.00	−12	19	1.02	1.02	98

*1st Qu* first quartile, *3rd Qu* third quartile, *F1* Factor 1, *F2* Factor 2, *n* number of subjects, *SD* Standard deviation

conservative and replied the hypotheses made for construct validity, as in the study of de Boer et al. [24].

The change score ( $x = 9.4$  points;  $SD = 13.3$ ) reported in this study is similar to that reported by Gay, Madson, and Cieslak [29] ( $x = 14$  points;  $SD = 11.9$ ) which performed an analogous treatment (e.g., three times/week of manual therapy techniques over 1 month plus home exercise program) in a comparable sample of subjects (e.g., chronic NP); therefore, it seems that the NBQ resembles the same measurement characteristics. The MCID of 5.5 points found in this study approximates the threshold suggested by Bolton and Humphreys [10]. However, other authors [30, 31] found a change in raw score of 13 points to detect an improvement in the patient's health status with specificity and sensitivity of 100 %.

Among the neck-specific questionnaires, the Neck Disability Index [8] has been recommended in two recent systematic reviews on the measurement properties of original [32] and translated [33] versions of neck-specific health-related quality of life instruments because it constituted the only instruments having all the measurement properties validated and with positive findings. However, the results of the present study add further contribution on the hypotheses testing and the responsiveness, strengthen

the internal consistency and provide novel evidence on the measurement properties content validity and structural validity of the NBQ, which now has all the measurement properties validated with performance comparable to those of the NDI. Therefore, it is plausible that the NBQ may be reconsidered in further updated reviews on the measurement properties of neck-specific questionnaires as an additional valid instrument to measure the quality of life in people suffering from neck pain.

Even though the validation process was made according to the principles stated by the COSMIN, few issues could limit the validity of the Italian version of the NBQ. Among all the psychometric properties evaluated the reliability is missing because, as the subjects were attending a physiotherapy service, it was considered unethical to interrupt or delay it in order to keep their condition stable enough to test the reliability of the questionnaire. Further, the threshold used to compare the relative magnitudes of correlation among different questionnaires for both the construct validity and the responsiveness may have been too high and this could have biased their real validity. Nevertheless, as the hypotheses testing process is a relatively new validation process, we tried to establish a simple rule that could serve as a reference for future validations of different

questionnaires. Among the outcome measures chosen in this validation process, we lacked the Neck Disability Index [8], one of the routinely used questionnaires across the literature concerning observational and intervention studies on neck pain, because it was not validated yet when our study began. Finally, the responsiveness of the NBQ may be considered not fully achieved because the outcome measures used to test it had the description of their responsiveness based on the significance of their change scores rather than on their validity.

## Conclusions

The NBQ has a two-factor structure whose construct validity and responsiveness are moderate. The internal consistency indicated the items of the NBQ have a high degree of interrelatedness. The results in change score obtained in the Italian population are similar to that in different populations and a clinical improvement is detected by the NBQ when the change score is greater than 5.5 points.

**Acknowledgements** We want to gratefully thank Rosaria Carioti for her organizational competences and the librarians of the University of Florence and Genua for their support in retrieving some of the cited articles.

## References

- Hogg-Johnson, S., van der Velde, G., Carroll, L. J., Holm, L. W., Cassidy, J. D., Guzman, J., et al. (2008). The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)*, *33*(4 Suppl), S39–S51.
- Hush, J. M., Lin, C. C., Michaleff, Z. A., Verhagen, A., & Refshauge, K. M. (2011). Prognosis of acute idiopathic neck pain is poor: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation*, *92*(5), 824–829.
- Driessen, M. T., Lin, C. W., & van Tulder, M. W. (2012). Cost-effectiveness of conservative treatments for neck pain: A systematic review on economic evaluations. *European Spine Journal*, *21*(8), 1441–1450.
- Borghouts, J. A., Koes, B. W., Vondeling, H., & Bouter, L. M. (1999). Cost-of-illness of neck pain in The Netherlands in 1996. *Pain*, *80*(3), 629–636.
- Van Eerd, D., Cote, P., Kristman, V., Rezai, M., Hogg-Johnson, S., Vidmar, M., et al. (2011). The course of work absenteeism involving neck pain: a cohort study of Ontario lost-time claimants. *Spine (Phila Pa 1976)*, *36*(12), 977–982.
- World Health Organization. (2001). *ICF : International classification of functioning, disability and health/World Health Organization*. Geneva: World Health Organization.
- Ferreira, M. L., Borges, B. M., Rezende, I. L., Carvalho, L. P., Soares, L. P., Dabes, R. A., et al. (2010). Are neck pain scales and questionnaires compatible with the international classification of functioning, disability and health? A systematic review. *Disability and Rehabilitation*, *32*(19), 1539–1546.
- Vernon, H., & Mior, S. (1991). The Neck Disability Index: A study of reliability and validity. *Journal of Manipulative and Physiological Therapeutics*, *14*(7), 409–415.
- Wheeler, A. H., Goolkasian, P., Baird, A. C., & Darden, B. V. 2nd. (1999). Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. *Spine (Phila Pa 1976)*, *24*(13), 1290–1294.
- Bolton, J. E., & Humphreys, B. K. (2002). The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II. Psychometric properties in neck pain patients. *Journal of Manipulative and Physiological Therapeutics*, *25*(3), 141–148.
- Soklic, M., Peterson, C., & Humphreys, B. K. (2012). Translation and validation of the German version of the Bournemouth Questionnaire for Neck Pain. *Chiropractic & Manual Therapies*, *20*(1), 2.
- Martel, J., Dugas, C., Lafond, D., & Descarreaux, M. (2009). Validation of the French version of the Bournemouth Questionnaire. *Journal of the Canadian Chiropractic Association*, *53*(2), 102–120.
- Schmitt, M. A., de Wijer, A., van Genderen, F. R., van der Graaf, Y., Helders, P. J., & van Meeteren, N. L. (2009). The Neck Bournemouth Questionnaire cross-cultural adaptation into Dutch and evaluation of its psychometric properties in a population with subacute and chronic whiplash associated disorders. *Spine (Phila Pa 1976)*, *34*(23), 2551–2561.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745.
- Kline, P. (1993). *The handbook of psychological testing*. London: Routledge.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*, *25*(24), 3186–3191.
- Monticone, M., Baiardi, P., Nido, N., Righini, C., Tomba, A., & Giovanazzi, E. (2008). Development of the Italian version of the Neck Pain and Disability Scale, NPDS-I: cross-cultural adaptation, reliability, and validity. *Spine (Phila Pa 1976)*, *33*(13), E429–E434.
- Savoia, E., Fantini, M. P., Pandolfi, P. P., Dallolio, L., & Collina, N. (2006). Assessing the construct validity of the Italian version of the EQ-5D: preliminary results from a cross-sectional study in North Italy. *Health and Quality of Life Outcomes*, *4*, 47.
- Huskisson, E. C. (1974). Measurement of pain. *Lancet*, *2*(7889), 1127–1131.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., et al. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Medical Research Methodology*, *10*, 22.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). London: Sage Ltd.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- de Boer, M. R., Moll, A. C., de Vet, H. C., Terwee, C. B., Volker-Dieben, H. J., & van Rens, G. H. (2004). Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic and Physiological Optics*, *24*(4), 257–273.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

26. McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research*, 4(4), 293–307.
27. Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102–109.
28. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34–42.
29. Gay, R. E., Madson, T. J., & Cieslak, K. R. (2007). Comparison of the Neck Disability Index and the Neck Bournemouth Questionnaire in a sample of patients with chronic uncomplicated neck pain. *Journal of Manipulative and Physiological Therapeutics*, 30(4), 259–262.
30. Bolton, J. E. (2004). Sensitivity and specificity of outcome measures in patients with neck pain: detecting clinically significant improvement. *Spine (Phila Pa 1976)*, 29(21), 2410–2417.
31. Gemmell, H., & Miller, P. (2010). Relative effectiveness and adverse effects of cervical manipulation, mobilisation and the activator instrument in patients with sub-acute non-specific neck pain: results from a stopped randomised trial. *Chiropractic & Osteopathy*, 18, 20.
32. Schellingerhout, J. M., Verhagen, A. P., Heymans, M. W., Koes, B. W., de Vet, H. C., & Terwee, C. B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Quality of Life Research*, 21(4), 659–670.
33. Schellingerhout, J. M., Heymans, M. W., Verhagen, A. P., de Vet, H. C., Koes, B. W., & Terwee, C. B. (2011). Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Medical Research Methodology*, 11, 87.