Reliability and Responsiveness of Patient-Reported Outcome Measures of Neck Disability to Physical Therapy: Comparison of the Copenhagen, Northwick Park, and Neck Bournemouth Questionnaires and the Neck Disability Index

Natalija Stefanovitch-Lawbuary, BSc, MB ChB,[a] Rouin Amirfeyz, MSc, MD, FRCS(T&O),[b] Rick Lovell, BSc, MCSP,[c] and Gordon Bannister, MD, MChOrth, FRCS(Ed Orth)[a]

Abstract

Objective: The purpose of this study was to identify the external and internal reliability and responsiveness of the validated patient-reported outcome measures (PROMs) of neck pain to a standardized regimen of physiotherapy administered acutely after mild whiplash injury using the clinically significant improvement components of the Patient Global Impression of Change (PGIC) as the outcome measure.

Methods: Eighty-six patients with neck pain alone were referred for physiotherapy within 2 weeks of whiplash injury. They completed the Copenhagen, Northwick Park (NP), and Neck Bournemouth (NBQ) questionnaires and the Neck Disability Index (NDI) before starting and after treatment when they also completed the PGIC. Treatment comprised deep soft tissue massage, myofascial releases, muscle energy techniques, joint articulation and manipulation techniques, and a home exercise program. The duration of treatment was between 3 and 6 weeks. A PGIC of 6 or 7 was considered to be clinically significant improvement.

Results: The external reliability of the PROMs was >0.7 and internal >0.87. All components of the PROMs contributed to the final score except headache in the Copenhagen and upper-limb dysesthesia in the NP. The most reliable questionnaire was the NBQ, which was significantly more responsive than the Copenhagen ($P = .008$). The NBQ was slightly more responsive than the NDI and NP. The NBQ and NDI were successfully completed more frequently than the NP and Copenhagen.

Conclusion: The NP, NDI, and NBQ are all reliable and responsive measures of change after physiotherapy for neck pain after acute whiplash injury. (J Manipulative Physiol Ther 2019;42:104-107)

Key Indexing Terms: *Neck Pain; Patient Reported Outcome Measures*

Introduction

Neck pain has a substantial impact on the health and quality of life of affected individuals and represents a significant financial burden on society.[1] Its prevalence in the UK population is 16.5% in women and 10.7% in men.[2,3] Manipulative physical therapy improves symptoms in the short term,[4] but there is difficulty in comparing the results of different interventions because outcome measures in the literature are not standardized. Many excellent studies have used variants of visual analog scores, but these are generally lengthy research instruments too detailed for general clinical use and less sensitive than the Neck Disability Index (NDI)[5] and the Neck Bournemouth questionnaire (NBQ),[6] which are disease-specific patient-

[a] Department of Orthopaedic Surgery, University of Bristol, Bristol, United Kingdom.
[b] University Hospital Bristol, Bristol, United Kingdom.
[c] Physio World, Solihull, Birmingham, Birmingham, United Kingdom.
Corresponding author: Gordon Bannister, MD, MChOrth, FRCS(Ed Orth), 19, Cranbrook Road, Redland, Bristol BS6 7BL, UK. Tel.: +000441179248440.
(e-mail: *gordon-bannister@tiscali.co.uk*).
Paper submitted September 28, 2017; in revised form May 1, 2018; accepted March 7, 2019.
0161-4754

https://doi.org/10.1016/j.jmpt.2019.03.007

reported outcome measures (PROMs) of neck pain. If interventions are to be compared reliably and standardized, validated PROMs should be used. There are 4 well-established, disease-specific PROMs of neck pain. These are the Copenhagen Neck Functional Disability Scale,[7] Northwick Park Neck Pain Questionnaire (NP),[8] NDI,[5] and NBQ.[6] There are 15 domains in the Copenhagen each with 3 possible responses, 9 in the NP each with a Likert scale of 5 responses, 10 in the NDI each with 6 response categories, and 7 in the NBQ each scored on an 11-point rating scale.

There have been comparisons of some of these questionnaires before,[9,10] but the relative reliability and responsiveness of these 4 questionnaires has not been established in one cohort of patients, and this would be useful to know to optimize the outcome measures used in reporting neck pain and disability and its management. Manipulative physical therapy improves symptoms in the short term,[4] and we considered that symptoms of neck pain were most likely to change with this intervention giving the best opportunity to test responsiveness to treatment. The aim of this study was to establish the relative reliability, responsiveness of and patient compliance in completing these 4 PROMs, standardizing the patients' clinical presentation and physical therapy as far as possible.

Patients and methods

Patients

Patients were recruited from one of a syndicate of 23 private physiotherapy clinics (Physio World) throughout England and Wales. They were selected from those presenting with acute whiplash-associated disorder (WAD) 1 or 2.[11] They were invited to take part in the study by their treating physiotherapist, and their consent was accepted verbally in accordance with the advice and requirements of the institutional research ethics committee. Patients attended within 2 weeks of the road traffic collision associated with onset of their symptoms. Patients with low back or widespread pain were excluded as the PROMs under investigation were specific to the neck. One hundred nineteen patients entered the study, of which 33 completed the first but not the second set of data. These 33 were excluded, leaving 86 with a complete data set. Of these, there were 31 men and 55 women with a mean age of 42 years.

Methods

At their initial consultation, patients completed the Copenhagen, NP, NDI, and NBQ before starting treatment. The order of the questionnaires was randomized to avoid fatigue bias in their completion.

The physiotherapy was provided by a number of practitioners from the same network of clinics. They adopted standardized principles of manual therapy tailored to each patient's individual symptoms. During treatment sessions, they administered deep soft tissue massage, myofascial releases, muscle energy techniques and joint articulation and manipulation techniques. Outside of treatment sessions, patients were instructed to follow a home exercise program taught during the treatment sessions. The aim of this was to maintain mobility of the soft tissues and joints after treatment and improve range of motion where restricted. Advice on self-management of activity levels, posture, and the use of ice and heat was given.

At the end of treatment, patients completed the Copenhagen, NP, NDI, and NBQ and the Patient Global Impression of Change (PGIC)[12] as described by Hurst and Bolton[12] (2004). A PGIC of 6 or 7 was considered to be a clinically significant improvement.[12]

Ethics and Consent

Research ethics approval for the study was obtained from the University of Bristol Medical Research Ethics Committee (reference number 101148). Patients were given an information sheet to read before participating in the study. This gave the contact details of the researchers and emphasized that patients were free to withdraw at any time. Confidentiality was emphasized because a number of these patients would be engaged in personal injury litigation. Formal written consent was not taken; rather, with widespread involvement of a large number of clinics, verbal consent was assumed if patients took part.

Statistical Methods

There was no power calculation because it has not been the convention to do so in previous studies. The internal consistency of reliability of the PROMs was assessed by Cronbach α to measure the correlations between different items on the same test. This instrument assessed if individual questions produce equal measures toward the general construct of each questionnaire.

The responsiveness was assessed by the association of changes in PROM scores measured against the anchor (PGIC). The Student t test and χ-square test with a Bonferroni correction were used where indicated. The statistical package used was SPSS version 13 (SPSS Inc, Chicago, Illinois).

Results

The external reliability of the PROMs was greater than 0.7 and internal reliability 0.87 or more. All items of the PROMs contributed to the final score except headache in the Copenhagen and upper-limb dysesthesia in the NP. The most reliable of the PROMs was the NBQ. The NBQ was significantly more responsive than the Copenhagen (P = .008) and slightly better than the NDI and NP. The NBQ and NDI were successfully completed more frequently than the NP and the Copenhagen (Table 1). On the PGIC,[12] all

Table 1. *Failure to Complete Questionnaires*

Questionnaire	Failure to Complete (%)
Neck Bournemouth Questionnaire	2.12
Neck Disability Index	2.72
Northwick Park Questionnaire	4.09
Copenhagen Questionnaire	6.06

but 5 participants reported noticeable improvement or better (Table 2).

Table 2. *Outcome of PGIC*

PGIC Score	Raw Numbers	% of Total
7	38	44
A great deal better		
Considerable improvement		
Has made all the difference		
6	25	29
Better		
Definite improvement		
Real and worthwhile difference		
5	19	22
Somewhat better		
Slight but noticeable change		
4 or less	4	5
Slightly better		
Not really any difference		

PGIC, Patient Global Impression of Change.

Discussion

This is the first study to compare multiple neck PROMs using the PGIC. Hurst and Bolton[12] used the PGIC to assess the NBQ. Gay et al[7] used a visual analog scale when comparing the NDI and NBQ. The PGIC[12] gave the participants more guidance than the visual analog scale. The 73% of participants who made a clinically significant improvement is slightly higher than the 63% reported by Hurst and Bolton.[12] This may reflect the case selection bias toward milder injury or the nature of the physiotherapy intervention. Case selection is unlikely. Whiplash-associated disorder grades 1 and 2 constitute most of the presentations with acute whiplash injury[13] and are associated with a shorter time to return to work than WAD grade 3[14] and fewer symptoms,[13] although the latter observation has not been reproduced by other authors.[15,16]

Our observations were similar to those reported by Gay et al,[9] who also found that the NBQ was slightly superior to the NDI. Although we found that all items of the 4 questionnaires contributed except for headaches in the Copenhagen and upper-limb dysesthesia in the NDI, Pickering et al[10] with a similarly sized sample, identified neck pain and dysfunction in the NDI, NP, and Copenhagen. The lack of contribution of upper-limb dysesthesia in our study relates to the participants' selection on the basis of their presentation as WAD 1 and 2.

Strengths and Limitations

The strength of this study is the relatively homogeneous clinical presentation, the standardization of the physiotherapy, and the wide geographical spread of the participants. The limitations of this study are the responsiveness, the rate of retrieval, and the inherent difficulty of data collection from a number of locations.

Limitations include that power calculations were not done before designing the study because they had not been used previously.[9,10,12] As in previous studies,[12] we used the standardized response mean (change of scores after and before treatment), which takes variance into account. There was limited variance because 73% of the participants experienced considerable improvement, so there was limited difference in responsiveness among the different PROMs.

We lost complete sets of data in 28% of participants because the clinical practitioners treating the patients did not have the time to check the responses. Some clinics returned many more samples than others, suggesting further loss of potential participants. Further work with a similar case mix comparing this physiotherapy regimen with a different one or an untreated control group would clarify the merits of the treatment administered to the patients in this study.

Conclusion

The Northwick Park, NDI, and NBQ are all reliable and responsive to changes after physical therapy for neck pain after acute whiplash injury. Historically, the NDI has been widely used, and this facilitates comparison with previous studies. The NBQ was the most reliable and responsive, was shorter, and was completed slightly more reliably than the other PROMs.

Funding sources and conflicts of interest

Funding was received from Department of Orthopaedics, University of Bristol, United Kingdom, Gordon Bannister discretionary fund. No conflicts of interest were reported for this study.

Contributorship information

Concept development (provided idea for the research): N.S.-L., G.B.
Design (planned the methods to generate the results): N.S.-L., G.B.
Supervision (provided oversight, responsible for organization and implementation, writing of the manuscript): G.B.
Data collection/processing (responsible for experiments, patient management, organization, or reporting data): N.S.-L., R.A., R.L., G.B.
Analysis/interpretation (responsible for statistical analysis, evaluation, and presentation of the results): R.A.
Literature search (performed the literature search): N.S.-L., G.B.
Writing (responsible for writing a substantive part of the manuscript): N.S.-L., G.B.
Critical review (revised manuscript for intellectual content, this does not relate to spelling and grammar checking): N.S.-L., R.A., R.L., G.B.

Practical Applications

- The findings indicate that 3 PROMs of neck pain are accurate and reproducible outcome measures of physical interventions.
- The work supports the NBQ and confirms the value of the NDI.
- Patient-reported outcome measures should be used before and after intervention to assess their outcomes and efficacy.

References

1. Borghouts JA, Koes BW, Vondeling H, Bouter LM. Cost-of-illness of neck pain in The Netherlands in 1996. *Pain.* 1999;80(3):629-636.
2. Urwin M, Symmons D, Allison T, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis.* 1998;57(11):649-655.
3. Webb R, Brammah T, Lunt M, Unwin M, Allison T, Symmons D. Prevalence and predictors of intense, chronic, and disabling neck and back pain in the UK general population. *Spine.* 2003;28(11):1195-1202.
4. Korthals-de Bos IB, Hoving JL, van Tulder MW, et al. Cost effectiveness of physiotherapy, manual therapy, and general practitioner care for neck pain: economic evaluation alongside a randomised controlled trial. *Brit Med J.* 2003;326(7395):911.
5. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther.* 1991; 14(7):409-415.
6. Bolton JE, Humphreys BK. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II. Psychometric properties in neck pain patients. *J Manipulative Physiol Ther.* 2002;25(3):141-148.
7. Jordan A, Manniche C, Mosdal C, Hindsberger C. The Copenhagen Neck Functional Disability Scale: a study of reliability and validity. *J Manipulative Physiol Ther.* 1998; 21(8):520-527.
8. Leak AM, Cooper J, Dyer S, Williams KA, Turner-Stokes L, Frank AO. The Northwick Park Neck Pain Questionnaire, devised to measure neck pain and disability. *Br J Rheumatol.* 1994;33(5):469-474.
9. Gay RE, Madson TJ, Cieslak KR. Comparison of the Neck Disability Index and the Neck Bournemouth Questionnaire in a sample of patients with chronic uncomplicated neck pain. *J Manipulative Physiol Ther.* 2007;30(4):259-262.
10. Pickering PM, Osmotherly PG, Attia JR, McElduff P. An examination of outcome measures for pain and dysfunction in the cervical spine: a factor analysis. *Spine.* 2011;36 (7):581-588.
11. Spitzer WO, Skovron ML, Salmi LR, et al. Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders. *Spine.* 1995;20(8 suppl):1S-73S.
12. Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. *J Manipulative Physiol Ther.* 2004;27(1):26-35.
13. Hartling L, Brison RJ, Arden C, Pickett W. Prognostic value of the Quebec classification of whiplash-associated disorders. *Spine.* 2001;26(1):36-41.
14. Suissa S, Harder S, Veilleux M. The relation between initial symptoms and signs and the prognosis of whiplash. *Eur Spine J.* 2001;10(1):44-49.
15. Kivioja J, Jensen I, Lindgren U. Neither the WAD-classification nor the Quebec Task Force follow-up regimen seems to be important for the outcome after a whiplash injury. A prospective study of 186 consecutive patients. *Eur Spine J.* 2008;17(7):930-935.
16. Atherton K, Wiles NJ, Lecky FE, Hawes SJ, Macfarlane GJ, Jones GT. Predictors of persistent neck pain after whiplash injury. *Emerg Med J.* 2006;23(3):195-201.