Classifying Whiplash Recovery Status Using the Neck Disability Index: Optimized Cutoff Points Derived From Receiver Operating Characteristic

Arthur C. Croft, PhD, DC, MSc, MPH^{a,*}, Julie A. Workman, DC^b,
Michael P. Szatalowicz, DC^c, Philip E. Roberts, DC^d, Leonard R. Suiter, DC^e

^a Director, Spine Research Institute of San Diego, San Diego, CA

^b Private Practice, Richland, WA

^c Private Practice, Brookfield, WI

^d Private Practice, Fort Smith, AR

^e Private Practice, Ballwin, MO

Received 28 October 2015; received in revised form 10 March 2016; accepted 25 March 2016

Key Indexing Terms:

Whiplash injuries;
Surveys and
questionnaire;
Disability evaluation

Abstract

Objective: Researchers often use Neck Disability Index (NDI) scores to classify recovery status in whiplash patients. The purpose of this study was to investigate the optimal cutoff point score for the NDI as a mechanism for differentiating recovery from nonrecovery after whiplash.

Methods: Subjects (N = 123) who had previously sustained whiplash injuries were recruited from 12 clinics. Subjects rated themselves as being recovered (36%) or nonrecovered (64%). This state variable was compared with their NDI score as test variable using the receiver operating characteristic statistic. The area under the receiver operating characteristic curve and optimized cutoff points were computed for the whole group and also dichotomized for sex and age.

Results: The mean NDI score for the recovered group was 7.8. It was 27.1 for the nonrecovered group. The cutoff point that optimized sensitivity and specificity for the whole group was an NDI score of 15. For women, it was 19; for older persons, it was 21.

Conclusion: The optimal NDI score cutoff point for differentiating the recovery state after whiplash is 15. Misclassification errors are likely when using lower values.

© 2016 National University of Health Sciences.

Introduction

In the 1960s and 1970s, patient self-reports of pain and dysfunction were generally regarded by researchers and clinicians as being too subjective to serve as

* Corresponding author at: 95 Tunapuna Ln, Coronado, CA 92118. Tel.: +1 619 423 5475; fax: +1 619 423 3084.

E-mail address: arthur.croft59@gmail.com (A. C. Croft).

reliable measures of disability. Accordingly, they sought more objective forms of disability assessment such as measures of the ranges of motion of joints, physical strength as assessed by manual muscle testing, and the results of orthodox orthopedic and neurological tests. However, the boundary between hard (ie, objective or measurable) and soft (ie, subjective self-report) data is often blurred.¹ Range of motion measurement is associated with a degree of uncertainty arising from the limited precision in recording the ranges, as well as the temporal variability in voluntary range of motion associated with the subject's day-to-day experience of pain. Manual muscle testing has been shown to be lacking in sensitivity. Trained examiners have been unable to detect changes in muscular strength until the loss in strength approaches 50%.²

Disability rating assessment methodologies of the past have been grounded on assumptions that do not always coincide with real-life conditions. For example, although a loss of range of motion is usually considered an indicator of impairment, in some cases, reduced spinal mobility may be associated with reductions in pain and disability scores.³ Conversely, joint instability is often associated with pain and impairment but can also coincide with greater than normal joint motion. Accordingly, Feinstein⁴ argued that the crucial attribute of hardness is simply the reliability and reproducibility of a finding.

Fairbank et al⁵ developed a questionnaire for the assessment of disability related to low back pain. Known today as the Revised Oswestry Disability Index, it is widely used both in the clinical setting and in research.⁶ It is a self-assessment questionnaire patients can quickly complete in a few minutes, and it can be easily scored by therapists or physicians. The format is categorical, but each category is ordinal. As such, the scale is not arithmetically isomorphic, and the scores cannot be considered linearly correlated with disablement.

In 1991, Vernon and Mior⁷ modified the Revised Oswestry Disability Questionnaire so that it would be responsive to disability related to neck pain and named it the Neck Disability Index (NDI). The NDI was subsequently validated in a whiplash patient population and achieved a high degree of internal consistency, reliability, and responsiveness.⁷ In subsequent years, a number of other questionnaires have been developed for evaluating neck pain. These include the Whiplash Disability Questionnaire,^{8–10} the Functional Rating Index,¹¹ the Northwick Park Neck Pain Questionnaire,^{9,10} the Neck Pain and Disability Scale,⁹ the Copenhagen

Neck Functional Disability Index,^{9,12,13} and the EQ-5D.¹⁴ To date, the NDI has been the most extensively used questionnaire in clinical trials and prognosis studies.^{9,13,15–43}

The general format follows that of the Oswestry questionnaire, with a 10-item design. These items or subscales include Pain Intensity, Personal Care, Lifting, Reading, Headaches, Concentration, Work, Driving, Sleeping, and Recreation. Each has 6 possible choices, with the first representing the normal state and carrying a score of 0, and the last representing the greatest degree of suffering or disability and carrying a score of 5. The 10 subscales are summed and multiplied by 2, thereby providing a range of 0%-100% disability.

Although the theoretical perfect NDI score would be 0, it is likely that many adults will endorse 1 or more of the NDI subscales in the lower or moderate ranges. Several authors have used the NDI to classify subjects' recovery status in clinical trials or outcome studies. Whereas some authors, using the NDI within a whiplash population to establish recovery, have set the threshold for recovery at 2% or less for headaches and at 1% or less for all other symptoms,³⁹ in 3 other published studies, recovery from whiplash injury based on the NDI score was set at a score of less than 8%.^{44–47}

Notwithstanding published cutoff points for NDI scores, limited normative data have been available until recently. The mean background score obtained from a large adult Japanese study was 6.98.⁴⁸ Another group has recently published normative data for Indian adolescent boys (3.59) and girls (4.92).⁴⁹ The purpose of this study was to investigate the optimal cutoff point for NDI score for a group of American adults who had suffered whiplash injury, using their self-assessment of recovery as the state variable or criterion standard.

Methods

Clinic patients from 12 clinics throughout the United States were recruited to participate in this study. Inclusion criteria included a whiplash injury resulting from a motor vehicle collision at least 24 weeks prior to recruitment, at least 3 treatment sessions with a licensed health care practitioner for the injury, fluency in English, and the requirement that they were at least 18 years of age when the injury occurred. Exclusionary criteria included prior cervical spine surgery or serious

spinal disease/disorder, chronic neck pain not related to a whiplash injury, or pending litigation. Persons who suffered spinal cord injury, spinal fracture, or spinal dislocation in their motor vehicle collision were also excluded. In total, 123 subjects were enrolled (55 men, 68 women). Of these, 44 (36%) self-rated themselves as having recovered and 79 (64%) self-rated themselves as not recovered. The ages of volunteers were normally distributed, with a range from 18 to 68 years. Fifty percent of the volunteers were younger than 40 years. Informed consent was obtained, and institutional review board approval was obtained from Walden University.

All subjects completed an NDI and general information questionnaire which included a statement as to their self-assessment concerning their recovery status after their whiplash injury (recovered or not recovered). The state values (recovered or not recovered) were then compared with the test values (NDI scores) using the receiver operating characteristic (ROC) statistic (IBM SPSS Statistics, Version 22).

Results

In the present study, 4 volunteers self-rated themselves as nonrecovered even though their NDI scores were less than 8 (ie, false negatives). Twenty-one claimed to have recovered but had scores of 8 or higher (ie, false positives). Twenty-three claiming to have recovered had NDI scores less than 8 (ie, true positives), and 75 had scores of 8 or higher who were not recovered (ie, true negatives).

The mean NDI score in our recovered group was 7.8. The mean score of nonrecovered persons was 27.1. Men generally had lower scores than women, with a mean recovered group NDI score of 5.5 and a nonrecovered group score of 27.8. For women, the scores were 9.5 and 26.5, respectively. Women have been shown to have consistently higher Revised Oswestry Disability Index scores as well.⁶

The area under the ROC curve (AUC) for the 123 subjects as a group was .910, the standard error was .025, and the asymptotic significance level was $P < .0001$, with a 95% confidence interval of .856 to .964 (Fig 1). By tradition, the abscissa represents the value of $1 - \text{specificity}$, whereas the ordinate represents sensitivity.

The ROC curve is also plotted for men and women in Fig 2. For men, the AUC was .935, the standard error was .031, and the asymptotic significance was $P < .0001$, with a 95% confidence interval of .875 to .995.

Fig 1. The ROC curve for the 123 subjects in the study. With a sensitivity of .818 and a specificity of .810, the score of 15 would be the optimum cutoff point. AUC, area under the ROC curve; NDI, Neck Disability Index; ROC, receiver operating characteristic.

For women, the AUC was .884, the standard error was .046, and the asymptotic significance was $P < .0001$, with a 95% confidence interval of .795 to .974.

The 123 subjects were then dichotomized on the basis of age into a younger than 40 years group and a 40 years and older group (Fig 3). For the younger than 40

Fig 2. The ROC curve for male subjects and female subjects analyzed separately. The dashed line is the ROC curve for the entire group collectively. AUC, area under the ROC curve; ROC, receiver operating characteristic.

group, the AUC was .898, the standard error was .048, and the asymptotic significance was $P < .0001$, with a 95% confidence interval of .804 to .993. For the 40 and older group, the AUC was .944, the standard error was .028, and the asymptotic significance was $P < .0001$, with a 95% confidence interval of .890 to .999.

On the basis of this study, a reasonable optimal cutoff point for the classification of persons on the basis of NDI scores would be 15, which would provide a sensitivity of .818 and a specificity of .810. For women, the best cutoff point was 19, and for persons older than 40, it was 21.

Discussion

The ROC curve was first developed by electrical engineers during World War II as a means of determining the accuracy of radar to differentiate items of concern for more benign ones. Subsequently, it has been adopted for evaluating the sensitivity and specificity of diagnostic tests.

In this data set, the ROC curve rises rapidly toward the upper left-hand corner of the graph, indicating a low false-positive and false-negative rate. In general, the closer the AUC is to 1, the better the test. A value of .5 is the theoretical null value and would be represented by a diagonal line from the lower left to upper right parts of the graph.

In a recent Japanese study of 1200 persons registered with an Internet research company, Kato et al.⁴⁸ reported a

mean NDI score 6.98, which is quite close to the 7.8 mean score we measured from our whiplash injury recovered group. They also suggested that an optimal cutoff point would be 15 for a general population and then dichotomized this number on the basis of sex, suggesting a score of 13 for men and 17 for women. And for persons in their 60s and 70s, they recommended a score of 21. Although this was a larger study of a general population, their conclusions are quite close to ours, which were based on a smaller data set of whiplash-injured persons.

It is noteworthy that the optimized scores are reported as odd numbers, whereas it is not possible to obtain an odd-numbered NDI score because the total of the subscales is always doubled. The odd numbers are merely a function of IBM's *SPSS* program, which was also used by Kato et al.⁴⁸ From our analysis of the full 123 subjects, an NDI score of 13 had a sensitivity of .795 and a specificity of .861, which is reasonably close to the optimized score of 15 (.818 and .810). It is clear, however, that using a cutoff point of under 8 will result in a large misclassification error. In our study, for example, at a cutoff score of 7, the sensitivity and specificity were .523 and .949, and at a cutoff score of 9, the numbers were .591 and .937, which would result in a large type II error.

Limitations

This was a case-control design and the subjects who volunteered were self-selected, although there would have been no obvious ulterior motivation to participate. Subjects were asked in the general information questionnaire to answer (yes or no) the question as to whether they had recovered from their whiplash injury. The term *recovery*, however, was not formally defined, and it is conceivable that some subjects could have interpreted it as meaning something other than the more orthodox meaning (ie, the resolution of the symptoms or problems caused by the injury). It is also possible that, in some instances, subjects might have discounted minor ongoing symptoms when characterizing themselves as having recovered vs not recovered.

There are, of course, many human conditions, diseases, and disorders other than whiplash injury that can contribute to the score of the NDI. Because we did not have access to or inquire about preinjury NDI scores, it is conceivable that some of the subjects characterizing themselves as having recovered, despite NDI scores of 8 or higher, would have endorsed the same or nearly the same responses prior to their injury. But this missing information also limits the interpretation of the NDI in many if not most of the whiplash studies that have used the NDI.

Fig 3. The ROC curve for subjects dichotomized on the basis of age. The dashed line is the ROC curve for the entire group. AUC, area under the ROC curve; ROC, receiver operating characteristic.

The results of this study may provide some guidance for researchers and clinicians, but it should be expected that the results derived from other groups of patients will vary, and we encourage others to repeat this study.

The ROC AUC was greater than .900 in all but 1 subgroup analysis, indicating that the NDI had good overall sensitivity and specificity. However, although some authors have used a score of < 8 as a proxy for recovery status after whiplash injury, our study suggests that the optimized cutoff point should be 15 for men and for a general demographic, 19 for women, and as high as 21 for older populations of men and women. These optimized points are based on the combination of sensitivity and specificity. Our recommended cutoff points compare with those of a larger (N = 1200) Japanese study of persons registered with a research company and not selected for having experienced a whiplash injury: 15 for a general population, 13 for men, 17 for women, and 21 for older adults.

One limitation to the interpretation of the NDI in whiplash or other injury studies, or in a clinical setting, is the uncertainty concerning what preinjury NDI scores might have been. One potential solution might be to dichotomize the NDI into a before/after format in which subjects could provide responses to subscales based on their precrash and postcrash symptoms and impairments.

Conclusion

Although the optimal or perfect NDI score is 0, population studies have indicated that scores of generally healthy asymptomatic persons range from 4 to 5 in children to 7 in adults. Our goal was to investigate the optimal cutoff point for NDI score for a group of American adults who had suffered whiplash injury using their self-assessment of recovery as the state variable or criterion standard. The results of our investigation indicate that the optimal NDI score cutoff point for differentiating the recovery state after whiplash is 15. Misclassification errors are likely when using lower values.

Funding Sources and Potential Conflicts of Interest

This study was funded by the Spine Research Institute of San Diego. No conflicts of interest were reported for this study.

References

1. Deyo R. Measuring the functional status of patients with low back pain. *Arch Phys Med Rehabil* 1988;69:1044–53.
2. Ambroz A, Zucher R, Ambroz C. Strength testing in pain assessment. *Pract Pain Manag* 2006;42–7.
3. Lankhorst GJ, Van de Stadt RJ, Van der Korst JK. Natural history of idiopathic low back pain: three-year follow-up study of spinal motion, pain and functional capacity. *Rehabil Med* 1985;17:1–4.
4. Feinstein AR. Clinical biostatistics XLI. Hard science, soft data, and challenges of choosing clinical variables in research. *Clin Pharmacol Ther* 1977;22:485–98.
5. Fairbank JCT, Davies JB, Couper J, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66(8):271–3.
6. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine* 2000;25(22):2940–53.
7. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991;14(7):409–15 [PubMed PMID: 1834753].
8. Willis C, Niere KR, Hoving JL, Green S, O'Leary EF, Buchbinder R. Reproducibility and responsiveness of the Whiplash Disability Questionnaire. *Pain* 2004;110(3):681–8 [PubMed PMID: 15288409].
9. Pinfold M, Niere KR, O'Leary EF, Hoving JL, Green S, Buchbinder R. Validity and internal consistency of a whiplash-specific disability measure. *Spine* 2004;29(3):263–8 [PubMed PMID: 14752347].
10. Pho C, Godges J. Management of whiplash-associated disorder addressing thoracic and cervical spine impairments: a case report. *J Orthop Sports Phys Ther* 2004;34(9):511–9 [discussion 20–3. PubMed PMID: 15493519].
11. Cameron ID, Rebbeck T, Sindhusake D, Rubin G, Feyer A-M, Walsh J. Legislative change is associated with improved health status in people with whiplash. *Spine* 2008;33(4):250–4.
12. Vernon H. Assessment of self-rated disability, impairment, and sincerity of effort in whiplash-associated disorder. *J Musculoskelet Pain* 2000;8:155–67.
13. Stewart M, Maher CG, Refshauge KM, Bogduk N, Nicholas M. Responsiveness of pain and disability measures for chronic whiplash. *Spine* 2007;32(5):580–5.
14. Pink J, Petrou S, Williamson E, Williams M, Lamb SE. Properties of patient-reported outcome measures in individuals following acute whiplash injury. *Health Qual Life Outcomes* 2014;12:38, <http://dx.doi.org/10.1186/1477-7525-12-38> [Epub 2014/03/15, PubMed PMID: 24625124; PubMed Central PMCID: PMC3984688].
15. Busse JW, Dufton JA, Kilian BC, Bhandari M. The impact of non-injury-related factors on disability secondary to whiplash associated disorder type II: a retrospective file review. *J Manipulative Physiol Ther* 2004;27(2):79–83 [PubMed PMID: 14970807].
16. Miettinen T, Leino E, Airaksinen O, Lindgren KA. The possibility to use simple validated questionnaires to predict long-term health problems after whiplash injury. *Spine* 2004;29(3):E47–51 [PubMed PMID: 14752363].
17. Falla D, Bilenkij G, Jull G. Patients with chronic neck pain demonstrate altered patterns of muscle activation during

- performance of a functional upper limb task. *Spine* 2004;29(13):1436–40 [PubMed PMID: 15223935].
18. Bunketorp L, Stener-Victorin E, Carlsson J. Neck pain and disability following motor vehicle accidents—a cohort study. *Eur Spine J* 2005;14(1):84–9 [PubMed PMID: 15241671].
 19. Kaale BR, Krakenes J, Albrektsen G, Wester K. Whiplash-associated disorders impairment rating: neck disability index score according to severity of MRI findings of ligaments and membranes in the upper cervical spine. *J Neurotrauma* 2005;22(4):466–75 [PubMed PMID: 15853463].
 20. Muller R, Giles LG. Long-term follow-up of a randomized clinical trial assessing the efficacy of medication, acupuncture, and spinal manipulation for chronic mechanical spinal pain syndromes. *J Manipulative Physiol Ther* 2005;28(1):3–11.
 21. Fernandez-de-las-Penas C, Fernandez-Carnero J, Palomeque del Cerro L, Miangolarra-Page JC. Manipulative treatment vs. conventional physiotherapy treatment in whiplash injury: a randomized controlled trial. *J Whiplash Relat Disord* 2004;3:73–90.
 22. Keating JL, Kent P, Davidson M, Duke R, McKinnon L, De Nardis R. Predicting short term response and non-response to neck strengthening exercise for chronic neck pain. *J Whiplash Relat Disord* 2005;4(1):43–55.
 23. Dvir Z, Gal-Eshel N, Shamir B, Pevzner E, Peretz C, Knoller N. Simulated pain and cervical motion in patients with chronic disorders of the cervical spine. *Pain Res Manag* 2004;9(3):131–6 [PubMed PMID: 15340583].
 24. Hurwitz EL, Morgenstern H, Vassilaki M, Chaing L-M. Frequency and clinical predictors of adverse reactions to chiropractic care in the UCLA neck pain study. *Spine* 2005;30(13):1477–84.
 25. Kumbhare DA, Balsor B, Parkinson WL, et al. Measurement of cervical flexor endurance following whiplash. *Disabil Rehabil* 2005;27(14):801–7.
 26. Treleaven J, Jull G, Lowchoy N. Standing balance in persistent whiplash: a comparison between subjects with and without dizziness. *J Rehabil Med* 2005;37(4):224–9 [PubMed PMID: 16024478].
 27. Treleaven J, Jull G, LowChoy N. Smooth pursuit neck torsion test in whiplash-associated disorders: relationship to self-reports of neck pain and disability, dizziness and anxiety. *J Rehabil Med* 2005;37(4):219–23 [PubMed PMID: 16024477].
 28. Bunketorp L, Carlsson J, Kowalski J, Stener-Victorin E. Evaluating the reliability of multi-item scales: a non-parametric approach to the ordered categorical structure of data collected with the Swedish version of the Tampa Scale for Kinesiophobia and the Self-Efficacy Scale. *J Rehabil Med* 2005;37(5):330–4 [PubMed PMID: 16208869].
 29. Sterling M, Jull G, Vicenzino B, Kenardy J, Darnell R. Physical and psychological factors predict outcome following whiplash injury. *Pain* 2005;114(1–2):141–8 [PubMed PMID: 15733639].
 30. Prushansky T, Gepstein R, Gordon C, Dvir Z. Cervical muscles weakness in chronic whiplash patients. *Clin Biomech (Bristol, Avon)* 2005;20(8):794–8 [PubMed PMID: 15996800].
 31. Armstrong BS, McNair PJ, Williams M. Head and neck position sense in whiplash patients and healthy individuals and the effect of the cranio-cervical flexion action. *Clin Biomech (Bristol, Avon)* 2005;20(7):675–84 [PubMed PMID: 15963617].
 32. Hooper RA, Frizzell JB, Faris P. Case series on chronic whiplash related neck pain treated with intraarticular zygapophysial joint regeneration injection therapy. *Pain Physician* 2007;10:313–8.
 33. Lamb SE, Gates S, Underwood MR, et al. MINT Study Team Managing Injuries of the Neck Trial (MINT): design of a randomised controlled trial of treatments for whiplash associated disorders. *BMC Musculoskelet Disord* 2007;8(7):1–7.
 34. Rebbeck TJ, Refshauge KM, Maher CG, Stewart M. Evaluation of the core outcome measure in whiplash. *Spine* 2007;32(6):696–702.
 35. Prushansky T, Handelzalts S, Pevzner E. Reproducibility of pressure pain threshold and visual analog scale findings in chronic whiplash patients. *Pain* 2007;23(4):339–45.
 36. Landers MR, Cheung W, Miller D, et al. Workers' compensation and litigation status influence the functional outcome of patients with neck pain. *Pain* 2007;23(8):676–82, <http://dx.doi.org/10.1097/AJP.0b013e31813d110e> [Epub 2007/09/22, 00002508-200710000-00006 [pii]. PubMed PMID: 17885346].
 37. Carroll A, Barnes M, Comiskey C. A prospective randomized controlled study of the role of botulinum toxin in whiplash-associated disorder. *Clin Rehabil* 2008;22:513–9.
 38. Periera MJ, Jull GA, Treleaven JM. Self-reported driving habits in subjects with persistent whiplash-associated disorder: relationship to sensorimotor and psychologic features. *Arch Phys Med Rehabil* 2008;89:1097–102.
 39. Geldman M, Moore A, Cheek L. The effect of pre-injury physical fitness on the initial severity and recovery from whiplash injury at six-month follow-up. *Clin Rehabil* 2008;22:364–76.
 40. Nieto R, Miro J, Huguet A. Disability in subacute whiplash patients. *Spine* 2008;33(18):E630–5.
 41. Elliott J, Sterling M, Noteboom JT, Darnell R, Galloway G, Jull G. Fatty infiltrate in the cervical extensor muscles is not a feature of chronic, insidious-onset neck pain. *Clin Radiol* 2008;63(6):681–7.
 42. Chien A, Eliav E, Sterling M. Whiplash (grade II) and cervical radiculopathy share a similar sensory presentation: an investigation using quantitative sensory testing. *Pain* 2008;24(7):595–603.
 43. Gabel CP, Burkett B, Neller A, Yelland M. Can long-term impairment in general practitioner whiplash patients be predicted using screening and patient-reported outcomes? *Rehabil Res* 2008;31(1):79–80.
 44. Sterling M, Jull G, Vicenzino B, Kenardy J. Sensory hypersensitivity occurs soon after whiplash injury and is associated with poor recovery. *Pain* 2003;104(3):509–17 [PubMed PMID: 12927623].
 45. Vetti N, Krakenes J, Eide GE, Rorvik J, Gilhus NE, Espeland A. Are MRI high-signal changes of alar and transverse ligaments in acute whiplash injury related to outcome? *BMC Musculoskelet Disord* 2010;11(1):260.
 46. Merrick D, Stalnackle B-M. Five years post whiplash injury: symptoms and psychological factors in recovered versus non-recovered. *BMC Res Notes* 2010;3:190.
 47. Sterling M, Elliott JM, Cabot PJ. The course of serum inflammatory biomarkers following whiplash injury and their relationship to sensory and muscle measures: a longitudinal cohort study. *PLoS One* 2013;8(10) [Epub 2013/10/23. PubMed PMID: 24147095; PubMed Central PMCID: PMC3798600].

-
48. Kato S, Takeshita K, Matsudaira K, Tonosu J, Hara N, Chikuda H. Normative score and cut-off value of the Neck Disability Index. *J Orthop Sci* 2012;17:687–93.
49. Koley S, Kaushal M, Sharma R. Correlations of neck disability index with anthropometric and neck biomechanical variables in Indian obese adolescents. *Ann Biol Res* 2014;5(2):51–4.