Review Article

Challenging the norm: further psychometric investigation of the neck disability index

Man Hung, PhD[a,*], Christine Cheng, BS[b], Shirley D. Hon, BS[c], Jeremy D. Franklin, MA[d],
Brandon D. Lawrence, MD[b], Ashley Neese, BS[b], Chase B. Grover, BS[b], Darrel S. Brodke, MD[b]

[a]*University of Utah School of Medicine, Department of Orthopaedics, Department of Public Health, Division of Epidemiology, Huntsman Cancer Institute, 590 Wakara Way, Salt Lake City, UT 84108, USA*
[b]*University of Utah School of Medicine, Department of Orthopaedics, 590 Wakara Way, Salt Lake City, UT 84108, USA*
[c]*University of Utah College of Engineering, Department of Electrical and Computer Engineering, 50 S. Central Campus Dr., Rm. 3280 MEB, Salt Lake City, 84112, UT, USA*
[d]*University of Utah College of Education, Salt Lake City, UT, USA*

Abstract

BACKGROUND CONTEXT: The neck disability index (NDI) was the first patient-reported outcome (PRO) instrument specific to patients with neck pain, and it remains one of the most widely used PROs for the neck population. The NDI is an appealing measure as it is a short and well-known PRO measure. Currently, there are conflicting data on the performance and applicability of the NDI in patients undergoing either operative or nonoperative treatment for neck-related conditions.

PURPOSE: This study investigates the psychometric properties, performance, and applicability of the NDI in the spine patient population.

STUDY DESIGN: A total of 865 patients visiting a university-based spine clinic with neck complaints, with or without radiating upper extremity pain, numbness, or weakness were enrolled in the study. Visit types included new and follow-up visits to both operative and nonoperative treatments. Questionnaires were administered electronically on a tablet computer, and all patients answered all 10 questions of the NDI.

METHODS: Standard descriptive statistics were performed to describe the demographic characteristics of the patients. Rasch modeling was applied to examine the psychometric properties of the NDI.

RESULTS: The NDI demonstrated insufficient unidimensionality (ie, unexplained variance after accounting for the first dimension=9.4%). Person reliability was 0.85 and item reliability was 1.00 for the NDI. The overall item fit for the NDI was good with an outfit mean square of 1.03. The NDI had a floor effect of 35.5% and ceiling effect of 4.6%. The raw score to measure correlation of the NDI was 0.019.

CONCLUSIONS: Although the NDI had good person and item reliability, it did not demonstrate strong evidence of unidimensionality. The NDI exhibited a very large floor effect. Because of the poor raw score to measure correlation, the sum score should not be used in interpretation of findings. Despite great investment by physicians and other stakeholders in the NDI, this evaluation and previous research have demonstrated that the NDI needs further investigation and refinement. © 2015 Elsevier Inc. All rights reserved.

Keywords: NDI; Spine; Patient-reported outcomes; Measurement; Rasch; Orthopedics

Introduction

In 2006, nearly 22 million people in the United States sought treatment for spinal disorders [1]. Physicians have stressed the importance of more robust measures to assess the condition of patients with spine disorders. Traditionally, clinicians have relied on technology and clinical measures to assess patients. Recently, the perspective of the patient or how they feel has become a major interest of health-care systems, organizations, and physicians themselves. As a result, physicians are taking into account patient-reported outcome (PRO) measures when assessing treatment success [2,3]. As in other fields, instrument use varies from physician to physician. Furthermore, it is not clear which measures most accurately evaluate treatment effects for patients with spinal disorders. To understand the comprehensive condition of their patients, physicians need to incorporate valid and reliable measures. In conjunction with clinical measures, the perceptions and perspectives of the patient need to be clearly understood to identify appropriate treatments.

The neck disability index (NDI), also known as the Vernon Mior disability index, was first published in 1991 comprising 10 items (See Appendix 1) [4,5]. As one of the first PRO instruments specific to patients with neck pain, it remains the most widely used PROs for patients with neck disorders and has been previously tested for validity and reliability [6,7]. It has been shown to have construct validity and reasonable test-retest reliability [6]. It is also assumed to be a unidimensional scale [8,9], although past research has disputed this [10]. The NDI instrument is appealing to physicians and patients because it is relatively easy and quick to administer. The NDI is used throughout the world and has been translated into 19 languages [11]. Although it has been demonstrated to be valid, some research has raised questions about the performance and applicability of the NDI in patients undergoing either operative or nonoperative treatment for their neck-related condition. Furthermore, studies have investigated the psychometric properties of the NDI and have found problems with ceiling effects, floor effects, and dimensionality [10–14]. A systematic review of published studies on the NDI found evidence of contrasting measurement properties of the NDI [11]. While the majority of research examining the NDI has used classical test theory [5,8,11], this study seeks to use contemporary testing techniques, such as Rasch analysis, to further investigate the psychometric properties, performance, and applicability of the NDI in the cervical spine patient population.

Methods

Data collection

A total of 865 patients visiting a university-based spine clinic from June 2011 to May 2013 were asked to complete the NDI, a demographic questionnaire, and some outcome questions. Patient visits were because of primary neck complaints, with or without radiating upper extremity pain, numbness, or weakness. Both new and follow-up patients with operative and nonoperative treatments were included in the final analysis. Questionnaires were administered electronically on a tablet computer (iPad; Apple, Inc., Cupertino, CA, USA) before seeing the physician. The response rate was 100% as taking it is a standard of care measurement in this specific clinic. Patients younger than 18 years and non-English speakers were excluded.

This was a self-funded study and none of the participants received any compensation. Informed consent was not necessary because responding to the NDI was part of standard care in the clinic. However, institutional review board approval was obtained before data analysis.

Analytic approach

Descriptive statistics were conducted to examine the characteristics of the participants. Next, a Rasch item response theory model was used to assess the psychometric properties of the NDI, including the fit, dimensionality, reliability, coverage, and raw score to measure correlation.

The Rasch partial credit model (PCM) for polytomous data was performed in this study using the Winsteps software version 3.80.0 [15]. Winsteps implements the PCM with the joint maximum likelihood estimation method, also known as the unconditional maximum likelihood estimation, which allows estimation of both the item difficulties and person abilities simultaneously and does not assume any person distribution [16–18]. The PCM used in Winsteps models each item with its own structure using the following formulation: $\log (P_{nij}/P_{ni(j-1)}) = B_n - D_i - F_{ij}$, where P_{nj} is the probability that person n responding to item i is observed in response category j, B_n is the ability measure of person n, D_i is the item difficulty measure of item i, F_{ij} is the calibration measure of category j relative to response category j−1 for item i [15]. The Rasch model assumes that item difficulty is the main characteristic affecting person responses. Additionally, it produces person and item estimates along a logit scale that represents a unit interval scale.

The Rasch model represents an attractive approach for constructing instruments for measurement. Specifically, it enables both the items and persons to be measured on the same metric, allowing for meaningful comparison of scores. It also enables "transformation on the item and person data to convert the ordinal data to yield interval data," p. 29 [19]. Before using the Rasch model to evaluate measurement properties, fit of the data to the model must be established [20,21]. For details into how a Rasch model works and technical terms as defined by Rasch methodology, readers are encouraged to review references [10,20–24].

Fit

To evaluate the psychometric properties of the NDI, we investigated the fit of the data to a Rasch model. To indicate