AUTHOR	Widbom-Kolhanen, Sara; Pernaa, Katri; Saltychev, Mikhail
TITLE	Reliability and validity of the Neck Disability Index among patients undergoing cervical surgery
YEAR	2022
DOI	10.1097/MRR.0000000000000540
VERSION	Author's accepted manuscript
CITATION	Widbom-Kolhanen, Saraa; Pernaa, Katri I.b; Saltychev, Mikhailc. Reliability and validity of the Neck Disability Index among patients undergoing cervical surgery. International Journal of Rehabilitation Research 45(3):p 273-278, September 2022. | DOI: 10.1097/MRR.0000000000000540

Reliability and Validity of the Neck Disability Index (NDI) Among Patients Undergoing Cervical Surgery

Short title: NDI reliability and validity

Sara S. Widbom-Kolhanen MD[1], Katri I. Pernaa MD, PhD[2], Mikhail Saltychev MD, PhD[3]

[1] Department of Surgery, Satasairaala Hospital, Pori, Finland

[2] Department of Orthopedics, Turku University Hospital and University of Turku, Turku, Finland

[3] Department of Physical and Rehabilitation Medicine, Turku University Hospital and University of Turku, Turku, Finland

ADDRESS FOR CORRESPONDENCE

Sara Widbom-Kolhanen, Department of Surgery, Satasairaala Hospital, Sairaalantie 3, 28500 Pori, Finland; tel.: +358 2 62771; e-mail: sswiko@utu.fi

CONFLICT OF INTEREST AND SOURCE OF FUNDING

None declared

ACKNOWLEDGEMENTS

None declared

Word count: 2,636

ABSTRACT

Objectives To explore the internal consistency and factor structure of the Neck Disability Index (NDI) among patients undergoing surgery on cervical spine.

Methods This was an observational retrospective cohort study among 392 patients undergoing a cervical surgery of any kind in a university hospital between 2018 and 2021. The patients responded to repeated surveys preoperatively and 3, 12 and 24 months postoperatively. The reliability and validity of the Neck Disability Index were investigated using Cronbach's alpha and factor analysis.

Results The internal consistency of the Neck Disability Index was found to be good 0.86. The exploratory factor analysis demonstrated unidimensionality. The correlations between the main factor "disability" and all the individual items of the Neck Disability Index were at least moderate. The highest correlations were observed for "pain intensity", "reading", "driving" and "recreation".

Conclusions The Neck Disability Index is internally consistent and unidimensional scale when applied to a population of patients undergoing cervical surgery.

KEYWORDS: Neck Disability Index; Neck Pain; Cervical Surgery; Reliability; Validity

INTRODUCTION

The valid and reliable assessment of the level of functioning is a prerequisite of successful rehabilitation after surgery. The Neck Disability Index (NDI) is one of the most widely used and validated measures to assess self-rated disability due to neck pain and outcome of cervical spine surgery [1, 2]. It has been created as a modification of the Oswestry Low Back Pain Disability Index [3]. The NDI has been translated into several languages and recommended by many international clinical guidelines [1].

Making successful clinical decisions requires valid tools with sound psychometric properties [4]. In 2002, a systematic review comparing five standard neck pain scales has praised NDI for its validity and recommended its use in different populations with neck pain [5]. In 2008, the developer of the NDI has reviewed the 17-year history of NDI emphasizing its reliability, validity and internal consistency [1]. However, in some studies, factor analysis has shown inconsistent findings concerning the structural validity of NDI. Reviewing the results of 37 studies, MacDermid et al. have found the NDI reliable and valid for evaluating disability in both acute and chronic neck pain, questioning however the unidimensionality of NDI since some studies had suggested a two-factor structure. [3]

In 2008, Cleland et al. have observed low test-retest reliability and construct validity of the NDI [6]. In a letter to the editor [7], however, Vernon has pointed a few limitations of the study, including violating a test-retest assumption, variance in subjects with minimal change and a too short time interval [8]. Also, Schellingerhout et al. have questioned the reliability of the NDI and claimed that previous studies might suffer from poor methodology [9]. Additionally, Hung et al. have raised doubts concerning the multidimensionality and a high floor effect of the NDI [10]. In 2019, Young et al. have suggested that the NDI might perform inconsistently depending on different characteristics of studied populations, e.g., patients with or without radicular symptoms. They have suggested that the psychometric properties of the NDI should be verified in different populations [4]. Recently, Jones and Sterling have addressed the existing disagreement regarding the reliability of the NDI [11]. A need for further research on the properties of the NDI across different groups and settings has been suggested by several studies [3, 8, 10, 11].

A narrative review by Godil et al. has investigated the properties of the NDI as a measure of postoperative outcome predictor. Comparing six common patient-reported outcome measures (PROMs) among patients undergoing primary anterior cervical discectomy and fusion (ACDF) for neck and arm pain followed up to 12 months, the NDI has been found to be the most valid and responsive measure of improvement after surgery for pain and disability. [2] Same research group has previously established the minimum clinically important difference (MCID) of the NDI after anterior cervical discectomy and fusion between 16% and 28%. [12] Studying predictors of outcome after ACDF, Anderson et al. have observed higher initial NDI as a positive predictor of good postoperative outcome [13]. Goyal et al. have recently studied 118 patients after surgery due to cervical spondylotic myelopathy and have found that only two domains of the NDI – limitations in work and recreation – were significant predictors of postoperative improvement [14]. Similarly, Steinhaus et al. have previously found that although all the NDI domains improve significantly after surgery, only work, recreation and pain intensity were independent predictors of surgery success [15].

Although the NDI has been used for three decades, the evidence on its structure validity and reliability has been inconsistence. It has been suggested that the properties of NDI should further be studied in different populations and settings. The evidence on the reliability and validity of NDI among patients undergoing cervical surgery is scarce. The objective of this study was to explore the internal consistency and factor structure of NDI among patients undergoing surgery on cervical spine.

METHODS

The data were obtained from the ongoing study among patients undergoing cervical surgery of any kind between June 21, 2018 and August 17, 2021 in a university hospital. All the patients, who have undergone more than one procedure during a follow-up were excluded. The patients responded to a survey within two months before the surgery. The survey contains questions on demographics and the severity of disability. Patients were included if their procedure codes were among following: NAG40, ABC60, ABC21, NAG41, ABC30, ABC10, NAG42 or ABC50, according to the Nordic Classification of Surgical Procedures (NCSP), version 1.15. The ethics board of university hospital district has approved the study protocol.

Age was defined in full years at the time of surgery. Body mass index (BMI) was defined as a body weight divided by a squared height and expressed in kg/m^2. The duration of pain was defined in years at the time of surgery. Pain intensity was assessed by using a visual analogue scale 0 to 100 points with 0 indicating no pain and 100 indicating most possible pain.

The NDI is a questionnaire containing 10 items covering disability caused by neck pain. Each item is assessed on a six-level ordinal scale with '0' describing 'no limitation' and '5' describing 'extreme limitation or an inability to function'. The total score is a percentage calculated by the sum of all answers divided by 50 (the maximum possible number of points) and multiplied by 100 as follows: 'Total score = (∑item scores/50) x 100'. The equation is adjusted when the responses to one or more items are missing. A score of 0% represents the highest possible level of functioning and independence while a score of 100% represents the lowest level of functioning with total dependence.

The normality of distribution of total NDI score was reviewed graphically using a histogram. While the distribution was close to normal curve, medians and interquartile ranges (IQR) were reported along with means and standard deviations (SDs) due to a relatively small sample size.

While there is no consensus on a sufficient size for a factor analysis, the common rule of thumb is that there should be at least 10 respondents per one item and n=300 should usually be sufficient for almost any model [17].

Statistical analysis

Internal consistency

The internal consistency of the NDI was assessed with Cronbach's alpha. Alpha ≥ 0.9 was considered excellent, ≥ 0.8 good, ≥ 0.7 acceptable, ≥ 0.6 questionable, ≥ 0.5 poor, and <0.5 was considered unacceptable. The sensitivity test was conducted excluding one item at a time.

Exploratory Factor Analysis

Both exploratory and confirmatory factor analyses were used. In an ideal situation, these two analyses should be performed on two different samples drawn from the same population. This was not possible in this study. Thus, the available sample was randomly split into two subsamples for conducting the two types of factor analysis balancing on gender on age.

This study employed exploratory factor analysis to approximate the construct structure of the NDI. The goal was to determine if the NDI measures only one latent trait (e.g., disability) or if there are other possible significant latent variables affecting the results as well. The results were analyzed both numerically and graphically. Exploratory factor analysis (principal factors) was applied with a minimum eigenvalue for retention set at >1.0 (Kaiser's rule) [18]. Retained and excluded factors were also explored visually on a scree plot and accompanied by a parallel analysis.

Confirmatory Factor Analysis

This study employed confirmatory factor analysis to verify the construct structure of the NDI seen in an exploratory factor analysis. Confirmatory factor analysis extends the abilities of an exploratory factor analysis including into a model a measurement error. The estimation procedure used the maximum likelihood method considering covariances supplied as input being unbiased. For simplicity, the estimates were reported in standardized form as correlation coefficients. A correlation <0.2 was considered poor, from 0.21 to 0.4 fair, from 0.41 to 0.6 moderate, from 0.61 to 0.8 substantial, and >0.8 perfect.

To assess how well the model matches the observed data, the root mean square error of approximation (RMSEA) was used. First, the model fit was tested assuming there were no covariances between unique factors. After that, the modification indices suggested by the software were used to add covariances between factors (double-headed arrows in Figure 1) one at a time, each time testing the lower 90% confidence limit (90% CL) of RMSEA closeness to 0.05 and upper 90% CL closeness to 0.10. The probability of RMSEA been <=0.05 was also reported. Every insertion was considered plausible if it made logical sense and did not violate the assumption that the common and the unique factors are uncorrelated. After achieving the acceptable RMSE value, no further covariances were imputed and the overall goodness of fit was assessed using a chi-square test for difference between the used model and a saturated model (a model with theoretically perfect fit). The results were accompanied by the Akaike's information criterion, the Bayesian information criterion, the comparative fit index, the Tucker-Lewis index, the standardized root mean squared residual and the coefficient of determination.

All analyses were conducted using Stata/IC Statistical Software: Release 16. College Station (StataCorp LP, TX, USA).

RESULTS

The data were available for 392 patients (Table 1). Their average age was 54.9 (11.3) years. Of the patients, 202 (52%) were women and 190 (48%) were men. The average NDI score was 44.3 (17.0) points. Of 392 procedures, 294 (70%) was anterior fusion of cervical spine without fixation (NAG40), 45 (11%) decompression of cervical spinal cord (ABC60) and 30 (8%) anterior decompression of cervical spine with insertion of interbody fixating implant (ABC21). The most frequent reasons for the surgery were “M50 Cervical disc disorders” (38%), “M47 Spondylosis” (34%) and “G99 Other disorders of nervous system (15 %).

Internal consistency

The Cronbach ‘s alpha was good 0.86 (lower 95% CL 0.84) (Table 2). All the items demonstrated good item-test an item-rest correlations. Also, excluding one item at a time did not improve alpha.

Exploratory factor analysis

The exploratory factor analysis demonstrated unidimensionality of the NDI. A single factor retained with eigenvalue 5.31 (Tables 3 and 4 and Figure 1). Item loadings on this retained factor were good for all 10 items.

Confirmatory factor analysis

The 1-factor model of confirmatory factor analysis showed acceptable fit (Table 5). The covariances of measurement errors was imputed for items #2, #3, #7 and #10. The correlations between the main factor “disability” and the individual items varied from moderate 0.51 to substantial 0.78. The highest correlations were observed for items #1 “pain intensity”, #4 “reading”, #8 “driving” and #10 “recreation”. Path diagram of confirmatory factor analysis is shown in Figure 2.

DISCUSSION

This observational cohort study among almost 400 patients undergoing different cervical surgical procedures investigated the reliability and validity of the NDI. The internal consistency of the NDI was found to be good. Excluding one item at a time did not improve alpha, verifying strong interrelatedness among the items of NDI. Exploratory factor analysis found the NDI to be a unidimensional scale with a one-factor structure. Using a one-factor model, the structure of NDI was analyzed using a confirmatory factor analysis. The correlations between the main factor "disability" and all the individual items of NDI were at least moderate. The highest correlations were observed for items #1 "pain intensity", #4 "reading", #8 "driving" and #10 "recreation".

The generalizability of the results might be affected by several issues. This study was conducted in a single university-based spine clinic. The cohort represented varying clinical conditions and procedures. As an example, indications of ACDF might include disc protrusion, single foraminal stenosis or spinal canal stenosis. However, the majority of the procedures were performed due to varying grade of degenerative cervical spine disease. The exploratory and confirmatory factor analyses were conducted on dependent samples.

This study confirmed previous research of the NDI as a valid and reliable measure. Internal consistency was assessed using Cronbach's alpha. Estimated previously, the NDI alpha has generally been exceeding 0.85 [3]. A cross-sectional study by Hains et al. has found evidence on the one-factor structure of the NDI [19]. Also, in that study, all the NDI items have correlated positively with coefficients varying from 0.54 to 0.84 similarly to our study (0.51-0.78). Based on internal consistency, unidimensional properties and ability to distinguish patients Saltychev et al. have previously recommended the NDI for use when selecting patients for rehabilitation, setting rehabilitation goals and measuring outcome of intervention [20]. Conversely, using a Rasch analysis, van der Velde et al. have stated that the NDI might not be unidimensional, though the unidimensionality could be achieved by removing two items "headaches" and “lifting” creating a new revised version (NDI-8) [21]. Another Rasch approved 5-item version of the NDI have been developed by Walton and MacDermid (NDI-5) [22]. Also, a Rasch analysis performed by Hung et al. have found the NDI being multidimensional stating the superiority of Rasch analysis over the classical test theory used by the majority of previous research [10]. The differences between those results and the present findings might be explained by the different psychometric behavior across diverse populations. This speculation is supported by a suggestion made by MacDermid et al. – observing a one- or two-factor structure might be due to different weight, which pain and disability may achieve across different pathologies and samples [3].

The national Swedish Spine Register (SweSpine) has used the NDI as a measure of self-rated disability and outcome of cervical surgery and it is also coming for use in the Finnish National Spine Register (FinSpine) to evaluate the prognostics, outcome and effectiveness of cervical surgery. Based on the reports from the United States, Norway and Finland, the rate of surgery for degenerative cervical spine disease has increased dramatically [16], and therefore, it is important that the tool used to assess the changes in functioning level after surgery is reliable and consistent.

Future research might reveal in more detail the properties of NDI among cohorts with different cervical pathologies and diverse surgical procedures. The test-retest reliability of the NDI in the population of interest was left outside the scope of the present study as the repeated responses in similar conditions (e.g., two or more responses before the surgery) were not available. Also, the psychometric properties of NDI in the studied or similar population request an analysis using the item response theory or Rasch.

The results suggested that the NDI is internally consistent and unidimensional scale when applied to a population of patients undergoing cervical surgery.

TABLES AND FIGURES

Table 1. Basic characteristic of the sample

Variables	Mean	Standard deviation	Median	IQR	N
Age, years	54.9	11.3	55.0	15.0	392
Body mass index (BMI), kg/m^2	28.2	5.5	27.7	6.3	392
Pain severity, points	53.8	28.5	59.0	48.0	215
Neck Disability Index, points					
Total score	44.3	17.0	46.0	22.0	338
Item 1	2.3	1.1	2.0	1.0	392
Item 2	1.4	1.1	1.0	1.0	391
Item 3	2.3	1.4	3.0	2.0	390
Item 4	2.2	1.2	2.0	2.0	388
Item 5	1.8	1.4	2.0	2.0	392
Item 6	1.7	1.1	2.0	1.0	389
Item 7	2.8	1.5	3.0	2.0	371
Item 8	2.3	1.3	3.0	2.0	361
Item 9	2.3	1.3	2.0	2.0	390
Item 10	2.8	1.4	3.0	2.0	376
				%	N
Gender					
Men				48%	190
Women				52%	202
Duration of neck pain before surgery					
< 6 weeks				6%	21
6-12 weeks				10%	36
3-12 months				35%	128
>12 months				50%	184
Total				100%	369

Table 2. Internal consistency of NDI (Cronbach's alpha)

Item		n	Sign	Item-test correlation	Item-rest correlation	Average interitem covariance	Alpha
Item 1	Pain intensity	392	+	0.72	0.65	0.62	0.84
Item 2	Personal care	391	+	0.63	0.54	0.64	0.84
Item 3	Lifting	390	+	0.62	0.50	0.62	0.85
Item 4	Reading	388	+	0.74	0.66	0.60	0.83
Item 5	Headaches	392	+	0.48	0.34	0.67	0.86
Item 6	Concentration	389	+	0.69	0.60	0.63	0.84
Item 7	Work	371	+	0.68	0.56	0.59	0.84
Item 8	Driving	361	+	0.77	0.69	0.58	0.83
Item 9	Sleeping	390	+	0.59	0.48	0.64	0.85
Item 10	Recreation	376	+	0.77	0.69	0.58	0.83
Total score						0.62	0.86[1]

[1] Lower 95% CI 0.84

Table 3. Exploratory factor analysis – item loadings

Items	Factor 1	Uniqueness
Item 1	0.80	0.36
Item 2	0.78	0.39
Item 3	0.68	0.54
Item 4	0.82	0.32
Item 5	0.50	0.75
Item 6	0.80	0.36
Item 7	0.82	0.33
Item 8	0.86	0.26
Item 9	0.61	0.62
Item 10	0.83	0.32

Table 4. Parallel analysis

Factors	Eigenvalues	Eigenvalues averaged over 10 replications	Difference
1	5.75	0.67	5.08
2	0.55	0.51	0.03
3	0.22	0.36	-0.15
4	0.18	0.22	-0.04
5	0.04	0.11	-0.07
6	-0.05	0.01	-0.05
7	-0.07	-0.08	0.01
8	-0.11	-0.17	0.06
9	-0.14	-0.26	0.12
10	-0.15	-0.34	0.19

Table 5. Confirmatory factor analysis – goodness of fit

Fit statistic	Value
Likelihood ratio	
chi2_ms(30)	51.8
p > chi^2	0.008
Population error	
Root mean squared error of approximation (RMSEA)	0.065
RMSEA 90% CI, lower bound	0.03
RMSEA 90% CI, upper bound	0.09
Probability RMSEA <= 0.05	0.191
Information criteria	
Akaike's information criterion (AIC)	5,159
Bayesian information criterion (BIC)	5,269
Baseline comparison	
Comparative fit index (CFI)	0.97
Tucker-Lewis index (TLI)	0.95
Size of residuals	
Standardized root mean squared residual (SRMR)	0.05
Coefficient of determination (CD)	0.88

Figure 1. Scree plot of exploratory factor analysis.

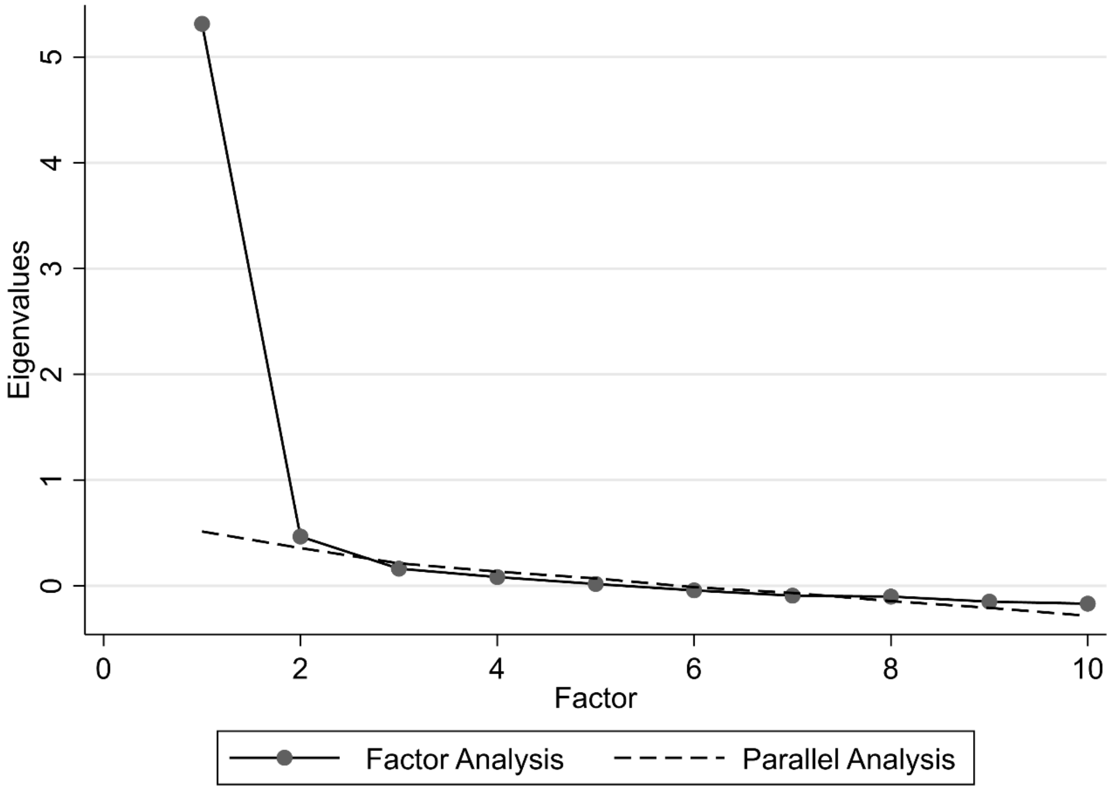
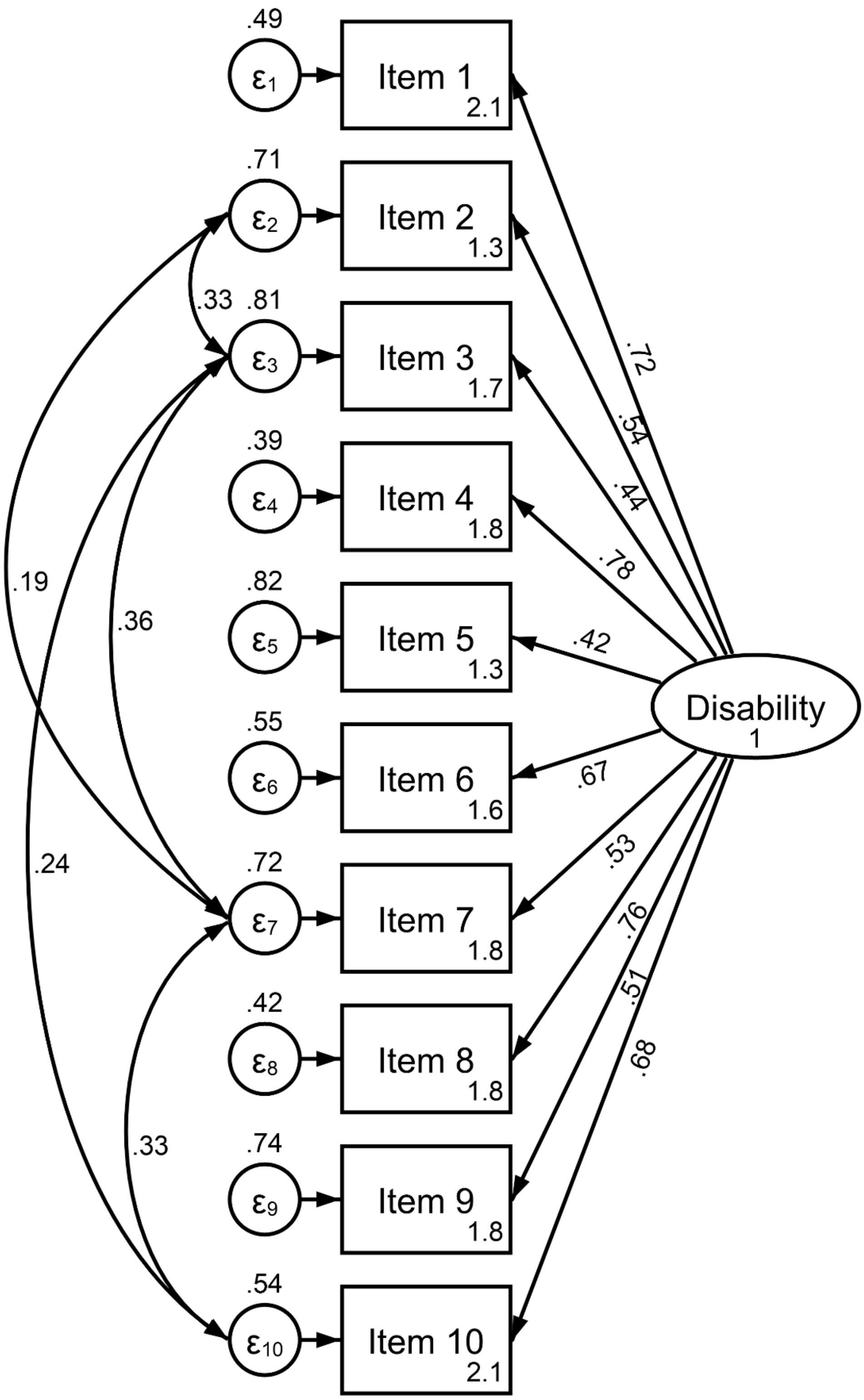


Figure 2. Path diagram of confirmatory factor analysis. Ɛ represents a measurement error; straight arrows between the latent trait (disability) and individual items represent correlations; curved arrows represent co-variance between individual items.



REFERENCES

1. Vernon H. The Neck Disability Index: state-of-the-art, 1991-2008. *J Manipulative Physiol Ther*. 2008;31:491-502.

2. Godil SS, Parker SL, Zuckerman SL, Mendenhall SK, McGirt MJ. Accurately measuring the quality and effectiveness of cervical spine surgery in registry efforts: determining the most valid and responsive instruments. *Spine J*. 2015;15:1203-9.

3. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. *J Orthop Sports Phys Ther*. 2009;39:400-17.

4. Young Ia Pt D, Dunning J Pt DPT, Butts R Pt P, Mourad F Pt DPT, Cleland Ja Pt P. Reliability, construct validity, and responsiveness of the neck disability index and numeric pain rating scale in patients with mechanical neck pain without upper extremity symptoms. *Physiother Theory Pract*. 2019;35:1328-35.

5. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine (Phila Pa 1976)*. 2002;27:515-22.

6. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. *Arch Phys Med Rehabil*. 2008;89:69-74.

7. Vernon H. The psychometric properties of the Neck Disability Index. *Arch Phys Med Rehabil*. 2008;89:1414-5; author reply 5-6.

8. Howell ER. The association between neck pain, the Neck Disability Index and cervical ranges of motion: a narrative review. *J Can Chiropr Assoc*. 2011;55:211-21.

9. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res*. 2012;21:659-70.

10. Hung M, Cheng C, Hon SD, Franklin JD, Lawrence BD, Neese A, et al. Challenging the norm: further psychometric investigation of the neck disability index. *Spine J*. 2015;15:2440-5.

11. Jones C, Sterling M. Clinimetrics: Neck Disability Index. *J Physiother*. 2021;67:144.

12. Parker SL, Godil SS, Shau DN, Mendenhall SK, McGirt MJ. Assessment of the minimum clinically important difference in pain, disability, and quality of life after anterior cervical discectomy and fusion: clinical article. *J Neurosurg Spine*. 2013;18:154-60.

13. Anderson PA, Subach BR, Riew KD. Predictors of outcome after anterior cervical discectomy and fusion: a multivariate analysis. *Spine (Phila Pa 1976)*. 2009;34:161-6.

14. Goyal DKC, Murphy HA, Hollern DA, Divi SN, Nicholson K, Stawicki C, et al. Is the Neck Disability Index an Appropriate Measure for Changes in Physical Function After Surgery for Cervical Spondylotic Myelopathy? *Int J Spine Surg*. 2020;14:53-8.

15. Steinhaus ME, Iyer S, Lovecchio F, Stein D, Ross T, Yang J, et al. Which NDI domains best predict change in physical function in patients undergoing cervical spine surgery? *Spine J*. 2019;19:1698-705.

16. Kotkansalo A, Leinonen V, Korajoki M, Korhonen K, Rinne J, Malmivaara A. Occurrence, Risk Factors, and Time Trends for Late Reoperations due to Degenerative Cervical Spine Disease: A Finnish National Register Study of 19 377 Patients Operated on Between 1999 and 2015. *Neurosurgery*. 2021;88:558-73.

17. Kyriazos T. Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*. 2018;09:2207-30.

18. Kaiser HF. The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*. 1960;20:141-51.

19. Hains F, Waalen J, Mior S. Psychometric properties of the neck disability index. *J Manipulative Physiol Ther*. 1998;21:75-80.

20. Saltychev M, Mattie R, McCormick Z, Laimi K. Psychometric properties of the neck disability index amongst patients with chronic neck pain using item response theory. *Disabil Rehabil*. 2018;40:2116-21.

21. van der Velde G, Beaton D, Hogg-Johnston S, Hurwitz E, Tennant A. Rasch analysis provides new insights into the measurement properties of the neck disability index. *Arthritis Rheum*. 2009;61:544-51.

22. Walton DM, MacDermid JC. A brief 5-item version of the Neck Disability Index shows good psychometric properties. *Health Qual Life Outcomes*. 2013;11:108.