Eur Spine J (2016) 25:2078–2086
DOI 10.1007/s00586-016-4543-z

ORIGINAL ARTICLE

Confirmatory factor analysis of the neck disability index, comparing patients with whiplash associated disorders to a control group with non-specific neck pain

Charles Philip Gabel[1] · Antonio Cuesta-Vargas[2] · Sebastian Barr[3] · Stephanie Winkeljohn Black[3] · Jason W. Osborne[6] · Markus Melloh[4,5]

Received: 11 July 2014 / Revised: 21 March 2016 / Accepted: 21 March 2016 / Published online: 4 April 2016

Abstract Purpose The neck disability index (NDI) as a 10-item patient reported outcome (PRO) measure is the most commonly used whiplash associated disorders (WAD) assessment tool. However, statistical rigor and factor structure are not definitive. To date, confirmatory factor analysis (CFA) has not examined whether the factor structure generalizes across different groups (e.g., WAD versus non-WAD). This study aimed to determine the psychometric properties of the NDI in these population groups.
Methods This study used CFA to analyze NDI baseline-data for WAD ($n = 804$; 69 % females) and non-WAD ($n = 963$; 67 % females), each for the full sample and separate genders. Invariance analyses examined the NDI structure across the four groups.
Results Across both populations and gender subgroups the one-factor solutions consistently showed better model fit over two-factor solutions. The NDI was best characterized as one-dimensional and invariant across multiple sub-groups.
Conclusion The NDI remains a valid PRO in WAD populations that provides acceptable measurement of neck status that is appropriate for basic functional assessment across genders. However, it is recommended that both clinicians and researchers initiate the transition toward more rigorous and less ambiguous PRO measurement tools for WAD patients and research. This future graduated movement toward other PROs should consider both regional PROs and computerized decision support systems, initially measured concurrently with the NDI to establish ways to convert existing scored data prior to their singular use.

Keywords Neck Pain · Whiplash · Clinimetrics · Neck disability index · Factor Analysis

✉ Charles Philip Gabel
cp.gabel@bigpond.com

Antonio Cuesta-Vargas
acuesta@uma.es

Sebastian Barr
sebastian.barr@louisville.edu

Stephanie Winkeljohn Black
stephanie.winkeljohn@louisville.edu

Jason W. Osborne
jwo@clemson.edu

Markus Melloh
markus.melloh@zhaw.ch

1 Private Researcher, Coolum Physiotherapy, Coolum Beach, QLD, Australia

2 Department of Psychiatry and Physiotherapy, Faculty of Medicine, Malaga University, Malaga, Spain

3 Department of Counselling and Human Development, University of Louisville, Louisville, KY, USA

4 Centre for Health Sciences, School of Health Professions, Zurich University of Applied Sciences, Winterthur, Switzerland

5 Centre for Medical Research, University of Western Australia, Nedlands, Australia

6 Department of Mathematical Sciences and the Graduate School, Clemson University, Clemson, SC, USA

Introduction

Neck problems that arise from whiplash are due to an acceleration–deceleration force transferring its energy to the cervical spine. The symptoms that develop are subsequently referred to as 'whiplash associated disorder' (WAD) [1]. These symptoms have considerable consequences for individuals, insurer groups and public health [2]. The associated functional status and change due to WAD can be quantified through the use of patient-reported outcome (PRO) measures [3] where the ten-item neck disability index (NDI) [4] is the most commonly used PRO. Furthermore, the risk of delayed recovery can also be determined and this can be achieved through the use of bio-psycho-social screening instruments [5] and the presence of impaired self-reported pre-injury health [6]. This risk of delayed recovery can lead to a chronic symptom profile that resembles a disorder that is functional somatic rather than being organic pain [7], The quantification of the impact of conditions such as WAD on the individual's functional status is an area in urgent need of further research [8]. The original WAD classification system by the Quebec Task Force was based upon the presenting signs and symptoms. However, the usefulness of this system has been questioned as the majority of sufferers with ongoing symptoms fell into one category. This led to recommendations for a modification of the system with proposals for further sub-classifications. This has resulted in a modified system with changes and re-classification that evolve the original system to one that has greater conformity with the bio-psycho-social model and accounts for measurable disturbances in motor, sensory and psychological dysfunction [9].

Despite these changes there has been little or no change in the PROs used to measure WAD in both the clinical and research setting [10] or the effectiveness of the process of measurement and clinical practice guidelines implementation on patient outcomes and costs of care [11]. Several PROs have been developed and trialed including: the whiplash disability questionnaire (WDQ) [12, 13], the whiplash specific activity and participation list (WAL) [14], the functional rating index (FRI) [15] and the spine functional index [16]. However, these tools either lack the practicality of the NDI, as they are more than one-page in length and require more than 2–3 min to complete, are lacking in rigorous review of their psychometric properties [5, 17, 18], or have yet to make an impact since publication, so their adoption by clinicians and researchers is limited at this stage. Another area of consideration, progress and research that reflects the digital age and twenty-first century has been the use computerized adaptive testing (CAT) and computerized decision support systems (CDSS) [19, 20]. However, despite the different available PROs and options, the 10-item NDI has remained the tool of choice for researchers investigating neck-related conditions.

It must be noted, however, that several publications have questioned the NDI's suitability as a stand-alone measure [10, 21, 22] due to conflicting constructs [4, 23, 24] and a two-factor structure [25, 26]. A PRO must exhibit a one-dimensional factor structure if a single-summated score is to be used [3]. These studies have employed exploratory factor analysis (EFA), a statistical method that provides an initial or 'exploratory' indication as to whether the questions within a PRO are representative of one or multiple factors or constructs. This EFA process requires a smaller sample and indicates the direction for a more thorough confirmatory factor analysis (CFA)—a statistical process that uses assumptions, modeling and estimations to determine whether the structure of the questionnaire is multi-faceted or if it is a single unitary structure [27]. More recent research on the NDI using Rasch analysis [28–30]—a statistical technique which uses different process to CFA to indicate whether there is equal informativeness between items to create a single "true" score, has questioned the NDI's ten items and their ability to represent even-scaled intervals and to produce a single "true" score. In these outpatient three studies all author groups suggested the NDI should be shortened by removal of specific questions. The initial Rasch analysis study by van der Velde et al. [25] found an eight-item PRO gave an optimal even-scaled score in a pooled two-cohort North American non-specific neck pain population, whereas Johansen et al. [30] found a seven-item PRO in a single population of neck pain patients treated at a Norwegian University hospital clinic, and most recently Walton et al. [28] found a five-item PRO was optimal in a pooled four-cohort community based North American population of mechanical specific and non-specific neck pain.

By contrast, no study to date has considered the NDI factor structure in a whiplash specific population using CFA. There are only two studies [23, 26] that have analyzed the NDI factor structure using this preferred standard of CFA. Both studies concluded that a one-factor structure was acceptable. However, the Hains [23] study had methodological inconsistencies that appeared inconsistent with CFA, whereas the Gabel [26] study detailed its methodology and used a more rigorous approach. This latter study also found that a two-factor model did approach significance, particularly for males, where question items separated into constructs of mental and physical function, which reflected the concerns of previous researchers [21, 24]. Recommendations were made that further CFA research is needed that analyzes the NDI

Table 1 Demographic data of patients from the seven contributing centers

Study source	Sample total *n*	WAD status	Sub-group *n*	Gender (% female)	Age Mean ± SD	% WAD	Compensation status in sample
1. University Sunshine Coast, Queensland, Australia	104[a]	WAD	42	69.8	39.1 ± 13.0	40.4	Permitted
		Non-WAD	62	49.2	42.9 ± 12.1		
2. Malaga University, Malaga, Spain	339[a]	WAD	234	63.0	35.6 ± 13.5	69.0	Not permitted
		Non-WAD	105	60.4	54.8 ± 8.4		
3. University of Queensland, Queensland, Australia	685[a]	WAD	378	71.9	38.2 ± 11.5	55.2	Permitted
		Non-WAD	307	77.5	41.3 ± 11.0		
4. Mobile Spine and Rehabilitation, Alabama, USA	57[a]	WAD	14	53.3	44.8 ± 11.2	24.6	Not Permitted
		Non-WAD	43	73.3	51.0 ± 13.5		
5. Melbourne Whiplash Centre, Victoria, Australia	355	WAD	59	78.7	38.2 ± 11.1	16.6	Permitted
		Non-WAD	296	64.4	41.0 ± 11.6		
6. Complejo Hospitalraio, Ciudad de Jaen, Spain	136	WAD	65	66.7	38.8 ± 13.9	47.8	Not Permitted
		Non-WAD	71	56.0	40.4 ± 14.2		
7. University of Newcastle, New South Wales, Australia	79[a]	WAD	0	Not applicable	Not applicable	0	Permitted
		Non-WAD	79	68.7	46.1 ± 17.7		
Totals	1755	WAD	792	68.7	37.2 ± 12.9	45.1	N/A
		Non-WAD	963	67.3	43.0 ± 14.8		

All studies were prospective and conducted within a physiotherapy outpatients setting with participants referred consecutively by a medical practitioner from a convenience sample

[a] Original groups that had publications that included data pertaining to this study [5, 15, 23, 29–33])

factor structure in large samples and different population groups, including those with WAD.

In summary, the NDI is the assessment instrument of choice for clinicians investigating neck problems after an acceleration–deceleration injury of the neck (WAD). Currently, the NDI is presenting results as a single score. An instrument that uses a single score must have only one underlying dimension (e.g., physical symptoms or mental symptoms); otherwise, it has to be considered as invalid. In the current literature there is conflicting evidence on whether the NDI has one or two dimensions—so called one- or two-dimensional factor structure.

The hypothesis of this study was that: analysis of the NDI factor structure using CFA in an adequately large sample of WAD and non-WAD patients, would reveal a single factor structure, and that this structure would be consistent across test two competing models of single versus dual-factor structure across both male and female-only subgroups.

Methods

Participants

To ensure the sample population was both diverse and sufficiently large, seven separate research groups were approached through their primary researcher. Each group had institutional ethical approval and some data was extracted from published studies [5, 16, 24, 31–35]. This enabled a secondary analysis of the de-identified-data as pooled samples which were then combined in a consistent format to provide a preliminary homogenous sample of symptomatic neck patients (total $n = 1817$, WAD $n = 804$, Non-WAD $n = 1013$) with 62 cases removed due to missing responses or missing fields to leave the final sample ($n = 1755$, WAD $n = 792$, age 37.2 ± 12.9 years, 68.7 % females; and Non-WAD $n = 963$, age 43.0 ± 14.8 years, 67.3 % females (Table 1).

Assessment tools

The NDI is a ten-item ordinal-scaled questionnaire with six responses per question scored with increasing severity from zero (no disability) to five (most severe disability) [4, 36]. The scores are added to provide a total maximum of 50 points then multiplied by two for a percentage scale. Severity levels are indicated by cut-off point scores with ≤8 NDI-points (16 %) for no disability/recovered or mild disability, >8–28 NDI-points for moderate disability and >28 NDI-points (56 %) for severe disability [4, 34, 37].

One-factor versus two-factor solutions

Analyses were conducted to determine if a one-factor or two-factor model fit the data better within different groups of participants (WAD males, WAD females, non-WAD males, non-WAD females). The alternative two-factor model, previously proposed and validated in a pooled non-specific neck pain sample, placed the ten NDI items into two factors: six items within a physical-symptoms factor (personal care, lifting, work, driving, sleep and recreation) and four items within a mental-symptoms factor (pain, headache, concentration and reading) [26].

For each of these eight models, modification indices (Wald indices) were analyzed to determine whether error terms could be allowed to covary in order to improve model fit. Generally, a parameter with a Wald exceeding 4.0 is considered legitimate to examine for estimation, but only if the two items both relate to the same latent variable. Thus, in the two-factor model errors loading on different factors were not correlated regardless of the Wald statistic. Error terms that were allowed to correlate were identical in models for both groups (i.e., if error term between items #1 and 2 were correlated in one group, they were correlated in all models for all groups).

The pooled data were analyzed through CFA using SPSS' AMOS software using maximum likelihood extraction (MLE) with a sampling of several common fit indices to evaluate the model's appropriateness. These indices were the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the likelihood ratio test (LRT) statistic (reported as CMIN but interpreted as Chi-square). For the RMSEA, values equal to or less than 0.08 reflect a reasonable fit, whereas values equal to or less than 0.05 indicate an excellent fit. For the CFI, values vary along a continuum of 0–1 in which values equal to or greater than 0.90 are considered satisfactory, and 0.95 or higher reflect an excellent fit [38].

Multi-group analyses

These were conducted in order to determine whether the one-factor model fits the data equally well for: (a) whiplash and non-whiplash participants and (b) male and female participants. Additionally analyses compared the fit of this model between gender groups within whiplash participants only and non-whiplash participants only. For each multi-group analysis, the CFA model first was run with no between-group constraints (i.e., the unconstrained model). This initial model was then compared with a series of models that contained increasingly strict invariance constraints: the second model constrained factor loadings to be equal across groups, the third constrained measurement intercepts, the fourth constrained measurement residuals, and the fifth constrained factor variance. Error terms were not allowed to correlate in these invariance analyses for the sake of simplicity.

Invariance between the groups on a particular parameter is supported when there is no statistical or meaningful difference between a model without a parameter constrained to be equal across groups and the model with that parameter constrained. In this case, the more parsimonious (more constrained) model is retained and compared to the subsequent model with an additional constraint to assess equivalence across groups on this newly constrained parameter.

Assessing competing models

There are multiple indicators of whether models are equivalent. The most common is a Chi-square-based LRT comparing the overall goodness of fit Chi-square values between the two models. However, because Chi-square is known to be highly (perhaps inappropriately) sensitive to unimportant differences in large samples [39], other indicators were also examined to determine invariance [38]. Other indicators in this analysis included ΔCFI, which was calculated by subtracting the comparative fit index (CFI) of models being compared (a $\Delta \text{CFI} > 0.01$ can indicate failure of invariance across models; [38 p242, 40] note that Meade et al. [39] recommend considering $\Delta \text{CFI} > 0.002$ as significant). Finally, the akaike information criterion (AIC) was used to evaluate the parsimony of the two models being compared. A lower AIC demonstrates better model fit to the data [41].

If there is a significant and substantial difference between two models, the model with fewer constraints is retained, indicating that the parameter in question is not invariant across groups. Results from the multi-group analyses, including goodness of fit statistics for all evaluated models and the standardized regression weights of the final, best fitting models, are reported.

In summary, this study investigated the NDI's factor structure in a whiplash and a non-whiplash population, also considering female and male subgroups. This was in order to determine whether the NDI had a one-dimensional factor structure that was robust to injury status and gender differences, or demonstrated a two-dimensional factor structure when used in certain populations.

Results

Evaluating one-factor and two-factor solutions

When the model fit of proposed one- and two-factor models were compared for each subgroup (WAD males, WAD females, non-WAD males, non-WAD females) this confirmed that the one-factor solution was in fact the *best* fit for the particular demographics. Although both models

Table 2 Summary of model comparisons between one-factor and two-factor solutions for the NDI

Model	*N*	RMSEA[a]	CFI	χ^2	*df*	Significant difference between models?[b]
Comparison within WAD males subgroup						
One-factor	249	0.025	0.996	33.488	29	$p = 0.00219956$
Two-factor	249	0.044	0.988	45.727	31	
Comparison within WAD females subgroup						
One-factor	543	0.022	0.997	33.115	26	$p = 0.00000135$
Two-factor	543	0.047	0.986	70.712	32	
Comparison within Non-WAD males subgroup						
One-factor	309	0.000	1.00	24.533	26	$p = 0.000024$
Two-factor	309	0.049	0.968	55.665	32	
Comparison within Non-WAD females subgroup						
One-factor	654	0.016	0.997	25.703	22	$p = 0.00000002$
Two-factor	654	0.049	0.968	77.553	30	

Note For all models modification indices (MI) were examined and the model was optimized to fit that sample to the extent allowed. In the one-factor model all errors were allowed to covary if the MI was greater than 4.00. In the two-factor model, errors were allowed to covary with errors from other items within the same factor, but not across factors

[a] Root mean square error of approximation. Closer to 0.000 is better

[b] Calculated as a Chi-square (χ^2_2–χ^2_1) with *df* (df_2–df_1); $p < 0.05$ indicates significant difference

had acceptably good fit within all four subgroups, the one-factor model fit the data significantly better than the two-factor solution (Table 2).

Comparing WAD and non-WAD participants

Multi-group analysis comparing WAD participants ($n = 792$) and non-WAD participants ($n = 963$) indicated a slightly lowered fit once factor loadings were constrained to be equal across groups ($\Delta\chi^2_{\text{Unconstrained vs. Metric}}$ $(9) = 32.777$, $p < 0.001$; ΔCFI $= 0.004$; Table 3). While there were some differences in standardized factor loadings between the two groups (Table 4), these differences were not accompanied by a significant change in CFI. Given that the NDI is designed as a basic measure it is unlikely this small invariance effect will be clinically meaningful. This is supported by the CFA results confirming that a one-factor model with all ten items was the best fit for all groups (Table 2).

Comparing male and female participants

Multi-group analysis comparing males ($n = 558$) and females ($n = 1197$) indicated equivalence across item factor loadings ($\Delta\chi^2_{\text{Unconstrained vs. Metric}}$ $(9) = 13.138$, $p < .156$; ΔCFI $= 0.00$; see Table 3), suggesting that the standardized regression weights (Table 4) are statistically similar between male and female participants overall.

Comparing male and female participants within the WAD group

Multi-group analysis comparing males within the WAD participant group ($n = 249$) and females within the WAD participant group ($n = 543$) produced a significant LRT but a minimal change in CFI ($\Delta\chi^2_{\text{Unconstrained vs. Metric}}$ $(9) = 21.025$, $p < 0.05$; ΔCFI $= 0.003$; Table 3). Similar to the analysis comparing WAD and non-WAD participants, while the factor loadings (Table 4) are significantly different between males with WAD injuries and females with WAD injuries, the clinical significance appears minimal. And again, this is supported by the good fit of the one-factor model with all ten items for the data for all subgroups. Thus, the metric model with regression weights constrained to be equal across groups was retained based on statistical analysis and author judgment; no further model comparisons were conducted.

Comparing male and female participants within the non-WAD group

Multi-group analysis comparing males ($n = 309$) and females ($n = 654$) within the non-WAD participant group indicated equivalence across item factor loadings ($\Delta\chi^2_{\text{Unconstrained vs. Metric}}$ $(9) = 4.925$, $p = 0.841$; Table 3), suggesting that the standardized regression weights (Table 4) are statistically similar between male and female participants with non-WAD injuries.

Table 3 Multi-group Comparisons

Model	χ^2	*df*	*P*	$\Delta\chi^2$	CFI[a]
Comparing WAD and non-WAD participants					
Unconstrained	398.117	70	<0.001		0.947
Metric	430.894	79	<0.001	32.777***, 9 *df*	0.943
Comparing male and female participants					
Unconstrained	393.013	70	<0.001		0.945
Metric	406.151	79	<0.001	13.138, 9 *df*	0.945
Comparing male and female participants within the WAD group					
Unconstrained	234.203	70	<0.001		0.958
Metric	255.288	79	<0.001	21.025*, 9 *df*	0.955
Comparing male and female participants within the non-WAD group					
Unconstrained	266.045	70	<0.001		0.912
Metric	270.970	79	<0.001	14.429, 9 *df*	0.914

Unconstrained model with no across-group constraints imposed, original model; *Metric* model with factor loadings constrained to be equal across groups

*** $p < 0.001$, * $p < 0.05$

[a] *CFI* Comparative Fit Index; closer to 1.00 indicates better fit

In summary, our findings confirmed for the NDI a one-dimensional factor structure in the whiplash and the non-whiplash population across both gender subgroups.

Discussion

Main findings

This study is the first to assess the factor structure of the NDI with CFA in sufficiently large samples of WAD and non-WAD patients to allow comparative analysis for males and females within each group. The findings indicated that, although both one- and two-factor structures provided well-fitting models across subgroups, the one-factor solution showed significantly better model fit than the two-factor solution in all groups and, consequently, is preferable. Importantly, invariance analyses indicated preferred one-factor structure met metric invariance requirements, supporting the use of the scale with all subgroups tested.

Consequently, we conclude that practitioners and researchers may use the NDI as a single construct that produces a single summated score—where all items are added equally (i.e., without weights) to provide the total score [42, 43]. This is in contrast to other PROs such as the WDQ [12] and WAL [14] that seem to be multidimensional in structure [13, 44], which can violate the assumptions made when creating a single summated score [16, 45, 46]. Furthermore, practicality in the form of length, completion and scoring time must also be considered if the obstacles of user burden, that impede the uptake of PROs, are to be minimized [17, 18]. In this study we were fortunate to have access to an unusually large sample compared to many other studies in this area. This more robust sample, combined with the use of modern confirmatory methods, should give readers more confidence in the generalizability and replicability of the findings.

This study supports the quantification, or score of the patient's functional status on the NDI, representing a single dimension and equally for males and females, as well as WAD and non-WAD populations. This finding would also be independent of the presence of any occult fractures as these have been shown to be present up to 36 % of the time but did not affect the development of persistent pain after whiplash [47]. The score can then be compared to previous measures made at baseline prior to the rehabilitation or earlier in the course of the disability for all types of patients, including those involved in a compensation process. This comparison in scores can determine whether a selected intervention or management process has been effective and assist in re-evaluation of the original primary diagnosis. This patient reported quantification of status would supplement that obtained from other sources such as body pain diagrams for symptom distribution and location using either paper or electronic methods [48].

This use of the NDI as a ten-item PRO can be made from the findings of the CFA used in this study. This is despite the lack of individual-item consistency which is due to the NDI use of ordinal scaling—where each question's response is ordered or ranked in terms of how strongly the chosen answer applies to the individual patient rather than each response having a mathematically equal

Table 4 Factor loadings from multi-group analyses

Item	Standardized regression weights			
	WAD[a]	Non-WAD[a]	Males[b]	Females[b]
NDI 1	0.702***	0.509***	0.631***	0.634***
NDI 2	0.667***	0.506***	0.576***	0.601***
NDI 3	0.686***	0.580***	0.642***	0.630***
NDI 4	0.720***	0.613***	0.630***	0.671***
NDI 5	0.629***	0.399***	0.509***	0.517***
NDI 6	0.720***	0.622***	0.664***	0.671***
NDI 7	0.797[c]	0.691[c]	0.718[c]	0.768[c]
NDI 8	0.732***	0.353***	0.609***	0.370***
NDI 9	0.665***	0.575***	0.617***	0.641***
NDI 10	0.812***	0.654***	0.743***	0.738***
	WAD males[b]	WAD females[b]	Non-WAD males[b]	Non-WAD females[b]
NDI 1	0.688***	0.704***	0.514***	0.495***
NDI 2	0.656***	0.662***	0.490***	0.512***
NDI 3	0.677***	0.674***	0.597***	0.571***
NDI 4	0.678***	0.728***	0.594***	0.615***
NDI 5	0.609***	0.622***	0.392***	0.382***
NDI 6	0.720***	0.715***	0.614***	0.616***
NDI 7	0.750[c]	0.812[c]	0.672[c]	0.711[c]
NDI 8	0.763***	0.710***	0.540***	0.283***
NDI 9	0.671***	0.673***	0.564***	0.589***
NDI 10	0.820***	0.805***	0.662***	0.658***

NDI 1–10 refer to the NDI item questions numbers one through ten

Note None of the models allowed item errors to covary

*** $p < 0.001$

[a] Regression weights obtained from Unconstrained Model

[b] Regression weights obtained from Metric Model

[c] Regression weight manually set to 1

value. This inconsistency between the total score and the value that each individual question contributes to this total was shown to be present through other statistical processes, specifically with Rasch analysis [28–30]. Consequently, researchers might want to consider this if using the NDI for more rigorous and precise analyses. A further consideration is that, though the factor loadings between males and females with and without WAD are both significantly different, the clinical implications are unlikely to be either practically meaningful or clinically significant.

Is a two-factor solution appropriate?

Our results demonstrated that the two-factor model also fits the data but is statistically an inferior solution. However, this structure may serve specific purposes in certain research groups and populations or other applications. This would be consistent with the recommendations of previous authors. This includes Sterling et al. [49] who suggest that "co-relationships between biological and psychological factors [of neck pain] be disentangled if clinicians are to gain from the outcomes of research and…adequately address all aspects of the patient's condition". Nieto [50] recommended that coping and catastrophizing were independently related to disability and depression, while Young et al. [24] recommended separation of "activity limitations and participation restrictions might give a truer picture of disability associated with neck pain for patients with psychological distress". Furthermore, the Rasch analysis findings clearly delineate into separate constructs [25, 28, 30].

Limitations and strengths

The study limitations include the lack of assessment of the different subgroups by clinical characteristics, psychological and socio-demographic status. This was a consequence of the restricted variables available from the original contributing researchers which was only from physical therapy outpatient clinics, and consequently, may not reflect the general WAD and non-WAD neck problematic population. The study strengths include the size of the samples being compliant with the minimum requirements for CFA.

Implications for future research

Further larger studies are required that investigate the NDI concurrently with other PRO measures that are summated to a single score. This should include recent developments such as PROs derived from both classic test theory and item response theory [25, 26], PROs that consider the whole-spine not just the neck alone and PRO derived through computerized adaptive testing [51] and computerized decision support software systems [52]. Such studies should concurrently consider Rasch analysis and CFA along with psychological assessment tools so that underlying influential psychological traits may be targeted [24, 49].

Conclusion

This study investigated and determined the factor structure of the NDI using CFA in an adequately large sample in both WAD and non-WAD patients and in gender specific subgroups. The findings indicated that the NDI remains a valid PRO in WAD populations and that it provides acceptable measurement of neck status that is appropriate for basic functional assessment across genders. However, in WAD patients the NDI remains an ambiguous PRO that was best characterized as one-dimensional from both the perspectives of statistical analysis and that of parsimony and clinical relevance. It is recommended that both clinicians and researchers initiate a graduated transition towards more rigorous, stable and less ambiguous PROs that measure the neck either alone or preferably with regional PROS that consider the neck as part of the single kinetic chain of the whole-spine. This measurement process could also include computerized adaptive testing (CAT) and computerized decision support systems (CDSS). This could be achieved in the future through a graduated movement toward other PROs in further research. Such a process should be initiated with a concurrent investigation of different PROs and the NDI in different groups in both clinical and research settings to establish ways to convert existing scored data prior to the introduction and standardized use of the suggested regional, CAT and CDSS PROs. Consequently it can be stated that the NDI is a valid assessment instrument in whiplash and non-whiplash populations regardless of their gender. Clinicians may use the NDI as a valid instrument when investigating patients with neck problems.

Acknowledgments We would like to thank all participating researchers and patients who contributed data to the study. This included the University of the Sunshine Coast, Queensland, Australia, Faculty of Medicine Malaga University, Spain, University of Queensland, Mobile Spine and Rehabilitation Centre USA, The Melbourne Whiplash Centre staff and patients, Victoria, Australia and Complejo Hospitalraio Ciudad de Jaen, Spain. The study was a secondary analysis of existing de-identified data that had Human Research Ethics Committee Approval from each of the contributing researchers' institution.

Compliance with ethical standards

Conflict of interest None.

References

1. Jensen TS, Kasch H, Bach FW, Bendix T, Kongsted A (2010) Definition, classification and epidemiology of whiplash—article in Danish. Ugeskr Laeger 172:1812–1814
2. Angst F, Gantenbein AR, Lehmann S, Gysi-Klaus F, Aeschlimann A, Michel BA, Hegemann F (2014) Multidimensional associative factors for improvement in pain, function, and working capacity after rehabilitation of whiplash associated disorder: a prognostic, prospective outcome study. BMC Musculoskelet Disord 15:130. doi:10.1186/1471-2474-15-130
3. de Vet HCW, Terwee CB, Mokkink LB (2011) Measurement in medicine a practical guide. Cambridge Press, London
4. Vernon H, Mior S (1991) The neck disability index: a study of reliability and validity. J Manipulative Physiol Ther 14:409–415
5. Gabel CP, Burkett B, Neller A, Yelland M (2008) Can long term impairment in general practitioner whiplash patients be predicted using screening and patient report outcomes? Int J Rehabil Res 31:79–80
6. Wenzel HG, Vasseljen O, Mykletun A, Nilsen TI (2012) Pre-injury health-related factors in relation to self-reported whiplash: longitudinal data from the HUNT study, Norway. Eur Spine J 21:1528–1535. doi:10.1007/s00586-012-2186-2
7. Wenzel HG, Mykletun A, Nilsen TI (2009) Symptom profile of persons self-reporting whiplash: a Norwegian population-based study (HUNT 2). Eur Spine J 18:1363–1370. doi:10.1007/s00586-009-1106-6
8. Hoy D, March L, Woolf A, Blyth F, Brooks P, Smith E, Vos T, Barendregt J, Blore J, Murray C, Burstein R, Buchbinder R (2014) The global burden of neck pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 73:1309–1315. doi:10.1136/annrheumdis-2013-204431
9. Sterling M (2006) Balancing the 'bio' with the psychosocial in whiplash associated disorders. Man Ther 11:180–181
10. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, Goldsmith CH, MacDermid JC (2009) Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther 39:334–350
11. Rebbeck T, Maher C, Refshauge KM (2006) Evaluating two implementation strategies for whiplash guidelines in physiotherapy: a cluster randomised trial. Aust J Physiother 52:165–174
12. Pinfold M, Niere KR, O'Leary EF, Hoving JL, Green S, Buchbinder R (2004) Validity and internal consistency of a whiplash-specific disability measure. Spine 29:263–268
13. Stupar M, Côté P, Beaton DE, Boyle E, Cassidy JD (2015) Structural and construct validity of the whiplash disability questionnaire in adults with acute whiplash-associated disorders. Spine J. doi:10.1016/j.spinee.2015.07.006

14. Schmitt MA, Stenneberg MS, Schrama PP, van Meeteren NL, Helders PJ, Schröder CD (2013) Measurement of clinically relevant functional health perceptions in patients with whiplash-associated disorders: the development of the whiplash specific activity and participation list (WAL). Eur Spine J 22:2097–2104
15. Feise RJ, Menke JM (2001) Functional Rating Index. A new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. Spine 26:78–86
16. Gabel CP, Melloh M, Burkett B, Michener LA (2014) The Spine Functional Index: development and clinimetric validation of a new whole-spine functional outcome measure. Spine J. doi:10.1016/j.spinee.2013.09.055
17. Liang MH, Jette AM (1981) Measuring functional ability in chronic arthritis: a critical review. Arthritis Rheum 24:80–86
18. Gabel CP, Burkett B, Yelland M (2009) Balancing fidelity and practicality in short version musculoskeletal outcome measures. Phys Ther Rev 14:221–225
19. Ware JEJ, Gandek B, Sinclair SJ, Bjorner JB (2005) Item response theory and computerised adaptive testing: implications for outcome measurement in rehabilitation. Rehabil Psychol 50:71–78
20. Tulsky DS, Carlozzi NE, Cella D (2011) Advances in outcomes measurement in rehabilitation medicine: current initiatives from the National Institutes of Health and the National Institute on Disability and Rehabilitation Research. Arch Phys Med Rehabil 92:S1–S6. doi:10.1016/j.apmr.2011.07.202
21. Nieto R, Miró J, Huguet A (2008) Disability in subacute whiplash patients: usefulness of the neck disability index. Spine 33:E630–E635
22. Gabel CP, Melloh M, Cuesta-Vargas AI (2011) Letter to the Editor Re Pickering PM, Osmotherly PG, Attia JR et al. An examination of outcome measures for pain and dysfunction in the cervical spine: a factor analysis. Spine 2011;36(7):581–588. Spine 36:1815
23. Hains F, Waalen J, Mior S (1998) Psychometric properties of the neck disability index. J Manipulative Physiol Ther 21:75–80
24. Young SB, Aprill C, Braswell J, Ogard W, Richards JS, McCarthy JP (2009) Psychological factors and domains of neck pain disability. Pain Med 10:310–318
25. van der Velde G, Beaton D, Hogg-Johnston S, Hurwitz E, Tennant A (2009) Rasch analysis provides new insights into the measurement properties of the neck disability index. Arthritis Rheum 61:544–551
26. Gabel CP, Cuesta-Vargas AI, Osborne JO, Burkett B, Melloh M (2014) Confirmatory factory analysis of the neck disability index indicates a one-factor model. Spine J 14:1410–1416. doi:10.1016/j.spinee.2013.08.026
27. Osborne J (2007) Best practices in quantitative methods. Sage Publications Thousand Oaks, California
28. Walton DM, MacDermid JC (2013) A brief 5-item version of the neck disability index shows good psychometric properties. Health Qual Life Outcomes. doi:10.1186/1477-7525-11-108
29. Swanenburg J, Humphreys K, Langenfeld A, Brunner F, Wirth B (2013) Validity and reliability of a German version of the neck disability index (NDI-G). Man Ther. doi:10.1016/j.math.2013.07.004
30. Johansen JB, Andelic N, Bakke E, Holter EB, Mengshoel AM, Røe C (2013) Measurement properties of the Norwegian version of the neck disability index in chronic neck pain. Spine (Phila Pa 1976) 38:851–856. doi:10.1097/BRS.0b013e31827fc3e9
31. Cuesta-Vargas AI, Gabel CP (2014) Validation of a Spanish version of the spine functional index. Health Qual Life Outcomes 12:96. http://www.hqlo.com/content/12/1/96
32. Takasaki H, Johnston V, Treleaven JM, Jull GA (2012) The Neck Pain Driving Index (NPDI) for chronic whiplash-associated disorders: development, reliability, and validity assessment. Spine J 12(10):912–920
33. Chien A, Eliav E, Sterling M (2010) The development of sensory hypoesthesia after whiplash injury. Clin J Pain 26:722–728
34. Sterling M, Jull G, Kenardy J (2006) Physical and psychological factors maintain long-term predictive capacity post-whiplash injury. Pain 122:102–108
35. Pickering PM, Osmotherly PG, Attia JR, McElduff P (2011) An examination of outcome measures for pain and dysfunction in the cervical spine: a factor analysis. Spine 36:581–588
36. Vernon H (2008) The neck disability index: state-of-the-art, 1991–2008. J Manipulative Physiol Ther 31:491–502. doi:10.1016/j.jmpt.2008.08.006
37. Sterling M, Kenardy J, Jull G, Vicenzino B (2003) The development of psychological changes following whiplash injury. Pain 106:481–489
38. Byrne BM (2010) Structural equation modeling with AMOS: basic concepts, applications, and programming. Routledge, New York
39. Meade AW, Johnson EC, Brady PW (2008) Power and sensitivity of alternative fit indices in tests of measurement invariance. J Appl Psychol 93:568–592. doi:10.1037/0021-9010.93.3.568
40. Cheung GW, Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model 9:233–255
41. Akaike H (1974) A new look at the statistical model identification. Autom Control, IEEE Trans 19:716–723. doi:10.1109/TAC.1974.1100705
42. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R (2002) Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 324:1417
43. Meads D, Doward L, McKenna S, Fisk J, Twiss J, Eckert B (2009) The development and validation of the Unidimensional Fatigue Impact Scale (U-FIS). Mult Scler 15:1228–1238
44. Stenneberg MS, Schmitt MA, van Trijffel E, Schröder CD, Lindeboom R (2015) Validation of a new questionnaire to assess the impact of whiplash associated disorders: the whiplash activity and participation list (WAL). Man Ther 20:84–89. doi:10.1016/j.math.2014.07.008
45. Doward LC, McKenna SP (2004) Defining patient-reported outcomes. Value Health 7:S4–S8
46. Costello AB, Osborne J (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval 10:1–9
47. Hertzum-Larsen R, Petersen H, Kasch H, Bendix T (2014) Do X-ray-occult fractures play a role in chronic pain following a whiplash injury? Eur Spine J 23:1673–1679. doi:10.1007/s00586-014-3362-3
48. Southerst D, Stupar M, Côté P, Mior S, Stern P (2013) The reliability of measuring pain distribution and location using body pain diagrams in patients with acute whiplash-associated disorders. J Manipulative Physiol Ther 36:395–402. doi:10.1016/j.jmpt.2013.05.023
49. Sterling M, Jull G, Vicenzino B, Kenardy J, Darnell R (2005) Physical and psychological factors predict outcome following whiplash injury. Pain 114:141–148
50. Nieto R, Miró J, Huguet A, Saldaña C (2011) Are coping and catastrophising independently related to disability and depression in patients with whiplash associated disorders? Disabil Rehabil 33:389–398. doi:10.3109/09638288.2010.491576
51. Hung M, Stuart AR, Higgins TF, Saltzman CL, Kubiak EN (2013) Computerized adaptive testing using the PROMIS physical function item bank reduces test burden with less ceiling effects compared to the short musculoskeletal function assessment in orthopaedic trauma patients. J Orthop Trauma 28:439–443. doi:10.1097/BOT.0000000000000059
52. Lobach D, Sanders GD, Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux R, Samsa G, Hasselblad V, Williams JW, Wing L, Musty M, Kendrick AS (2012) Enabling health care decisionmaking through clinical decision support and knowledge management. Evid Rep Technol Assess 203:1–784