

Cross-cultural Adaption and Validation of the Danish Voice Handicap Index

*†Jesper Roed Sorensen, *Trine Printz, *Camilla Slot Mehlum, ‡Christian Hamilton Heidemann, *Aagot Moeller Groentved, and *Christian Godballe, *†Odense C, and ‡Vejele, Denmark

Summary: Objectives: We aimed to assess psychometric properties, including internal consistency, reliability, and clinical validity of the Danish version of the Voice Handicap Index (VHI).

Study design: A cross-sectional survey study was carried out.

Methods: For validation, the existing nonvalidated Danish version of the VHI was used. Data from 208 patients with voice disorders of different etiology (neurogenic, functional, and structural) and a control group of 85 vocally healthy individuals were included. A test-retest reliability analysis of 42 patients and 45 control persons was performed. The internal consistency, test-retest reliability, and clinical validity of the questionnaire were assessed.

Results: Internal consistency was high with a Cronbach $\alpha > 0.90$ for both the patient and control group. Test-retest reliability measured as intraclass correlation coefficient was good with 0.93 (95% confidence interval [95% confidence interval]: 0.87–0.96) for patients and 0.78 (95% confidence interval: 0.63–0.87) for the control group which indicates sufficient reliability of the questionnaire. The Danish VHI has good clinical validity as it has a strong correlation between patient's perception of the severity of their voice disorder and the VHI score from the Spearman correlation of 0.69.

Conclusion: The existing Danish version of the VHI has been thoroughly validated and found to be in line with the original VHI from Jacobsen et al. It showed good internal consistency, test-retest reliability, and clinical validity. It is suitable for use in daily practice and in research projects as it is able to assess patients' perception of their voice disorder severity.

Key Words: Voice Handicap Index—Validity—Reliability—Dysphonia—Danish.

INTRODUCTION

Voice disorders can considerably impact Health-Related Quality of Life (HRQoL). Their effect on HRQoL is similar to or even greater than that of many other conditions such as angina, sciatica, and chronic sinusitis.^{1,2} The self-perceived severity of a voice disorder offers valuable insight to clinicians in daily clinical practice³ and is frequently assessed by either the Voice-Related Quality of Life or the Voice Handicap Index (VHI) questionnaires.^{4,5} The VHI questionnaire was developed in 1997 by Jacobsen et al and had shown strong internal consistency, reliability, and test-retest reliability in the assessment of voice disorders.⁴ Since its development, the VHI has undergone extensive *post hoc* validation, showing satisfactory scale content, good discrimination ability, and the ability to measure treatment effect.^{1,2,6,7} However, the correlation between the VHI and other more objective measurements of voice quality is often insufficient, indicating that multidimensional voice examination including an assessment of daily voice use is necessary to truly reflect the severity of a voice disorder.^{2,3,8–11}

The original VHI has been translated into many languages and validated, and all versions have shown good to excellent internal consistency, reliability, and test-retest reliability.^{7,12–19} Although the VHI is used extensively in Denmark,^{20,21} the existing version has not been validated. The aim of this study was to assess the psychometric properties of the existing version of the Danish VHI.

MATERIALS AND METHODS

Scale content and translation

The VHI consists of 30 statements covering various aspects of voice disorders. The statements are divided into three subscales (physical, emotional, and functional) each consisting of 10 statements. The statements are rated on a 5-point Likert scale from 0 (never) to 4 (always). The existing Danish version of the VHI was established in 2006 by translation from English into Danish by an expert group of speech-language pathologists (SPLs) and laryngologists proficient in English.²² This translated version was translated back into English by a bilingual SPL. The new English version of the questionnaire was compared with the original version by the expert group and no inconsistencies were found. Lastly, the Danish version of the VHI was tested in a pilot study and amended based on suggestions from patients.²²

Participants

The VHI was administered prospectively to patients with voice disorders referred to the outpatient clinic of the Department of Otolaryngology Head & Neck Surgery of a tertiary university hospital. A control group of vocally

Accepted for publication January 10, 2018.

Conflict of interest: The authors have no conflicts of interest. The study has been funded by the Odense University Hospital, the main employer of the authors.

From the *Department of ORL Head & Neck Surgery, Odense University Hospital, Odense C, Denmark; †OPEN, Odense Patient Data Exploratory Network, Odense University Hospital, Odense C, Denmark; and the ‡Department of Otorhinolaryngology, Lillebaelt Hospital, Vejle, Denmark.

Address correspondence and reprint requests to Jesper Roed Sorensen, Department of ORL Head & Neck Surgery, Odense University Hospital, J.B. Winslows vej 4, 5000 Odense C, Denmark. E-mail: jesper.roed.sorensen@rsyd.dk

Journal of Voice, Vol. 33, No. 4, pp. 441–444

0892-1997

© 2018 The Voice Foundation. Published by Elsevier B.V. All rights reserved.

<https://doi.org/10.1016/j.jvoice.2018.01.010>

healthy individuals was recruited from hospital staff, patients referred with conditions other than voice disorders, and relatives of both patients and staff. Control persons had neither previous voice disorders requiring therapy or treatment, self-reported voice changes, nor obvious voice disorder according to audio-perceptual assessment by an SPL or an ENT resident with focus on laryngology. Patients with voice disorders were divided into three groups based on etiology: group 1: “neurogenic disorders,” including recurrent laryngeal nerve paralyses/paresis, Parkinson disease, damage to the superior laryngeal nerve, or spasmodic dysphonia; group 2: “functional voice disorders,” including muscle tension dysphonia and hypokinetic dysphonia; and group 3: “structural disorders,” including polyps, cysts, nodules in the vocal cords, phono trauma, laryngitis, Reinke edema, leukoplakia, and early laryngeal cancers. The self-reported subjective severity of voice complaints was assessed by a 7-point Likert scale ranging from no complaints (1 point) to severe complaints (7 points). These scores were pooled into 3 levels of voice complaints: no- or mild complaints (1–3 points), moderate voice complaints (4–5 points), and severe voice complaints (6–7 points).

To estimate the test-retest reliability of the VHI, 42 patients and 45 control persons completed the questionnaire a second time approximately 2 weeks after the initial completion of the questionnaire. In this interval, no substantial change was expected in the patient’s health conditions. The study was approved by the local ethics committee. All participants signed an informed consent declaration before participation.

Statistical analyses

The internal consistency of the Danish version of the VHI was measured by the Cronbach α coefficient. The numerical results were interpreted as sufficient if larger than >0.7 , good with a value of >0.8 , and excellent with a value of >0.9 .²³

Clinical validity was assessed by the Spearman rank correlation coefficient evaluating the association between self-reported voice complaints and the VHI score. Values of 0.60 – 0.80 were considered a strong correlation and values >0.80 a very strong relation between voice complaints and VHI score.

Test-retest reliability was tested by the intraclass correlation coefficient (ICC) measuring the proportion of the

variability from between-subject differences out of the total variability. In general, ICC values less than 0.50 were considered poor reliability, values of 0.51 – 0.75 moderate reliability, and values above 0.75 good reliability.²⁴ The questionnaire’s ability to discriminate between control individuals and patients with voice disorders was evaluated with the Kruskal-Wallis and the Conover-Iman pairwise comparisons test comparing the three etiological subgroups of patients with voice disorders and the control group. Bonferroni correction was used to adjust for multiple comparisons. A $P < 0.008$ was required to reach significance in this test. Multiple linear regression analysis was used for evaluating the impact of age, sex, and having a voice disorder on the VHI score.

RESULTS

The VHI was completed by 208 patients with voice disorders (mean age 60 ± 12 years with 37% female patients) (Table 1). The three subgroups consisted of patients with “neurogenic disorders” ($n = 18$ [9%]), “functional disorders” ($n = 11$, [5%]), and “structural disorders” ($n = 179$ [86%]). Eighty-five individuals were recruited as control group (mean age 44 ± 15 years, 65% female individuals). Patients with voice disorders reported mild or no voice complaints (33%), moderate complaints (40%), or severe complaints (23%) (Table 2). Of the control persons, 4 of 85 individuals reported moderately to severe voice complaints (5%) despite normal audio-perceptual assessment. Neither age ($P = 0.27$) nor sex ($P = 0.64$) influenced the VHI score, but having a voice disorder significantly impacted the VHI score when using multiple regression analyses ($P < 0.001$).

The mean VHI score was 43.5 ± 25.6 for patients with voice disorders (maximum score 120) in comparison with 6.5 ± 8.3 for the control group (Table 2) ($P < 0.001$). The mean scores for the three subscales for the patient group were functional: 12.1 ± 9.4 , physical: 19.2 ± 10.2 , and emotional: 11.9 ± 9.3 for the patient group (maximum score 40) compared with functional: 2.6 ± 3.4 , physical: 2.9 ± 3.8 , and emotional: 1.0 ± 2.2 for the control group (all subscales: $P < 0.001$).

The Cronbach α of the VHI questionnaire was 0.97 for the patient group and 0.92 for the control group, indicating a good to excellent internal consistency. The internal consistency of the subscales ranged from 0.91 to 0.93 for the patient group and 0.81 to 0.86 for the control group

TABLE 1.
Sociodemographic Characteristics for Patients With Voice Disorders ($n = 208$) and the Control Group ($n = 85$)

Etiology	Number of Participants	Female Sex (%)	Age, Mean \pm SD	Occupational Voice Users (%)
Neurogenic	18	56%	61 ± 15	28%
Functional	11	55%	56 ± 22	36%
Structural	179	35%	60 ± 12	23%
Overall	208	37%	60 ± 12	24%
Control persons	85	65%	44 ± 15	39%

Abbreviation: SD, standard deviation.

TABLE 2.
Mean Values of the Danish VHI According to Self-reported Severity of Voice Complaints

VHI Scale	Control Group (n=85)	Level of Voice Complaints (Mean ± SD)			Overall (n=199)
		N/Mild (n=68)	Moderate (n=83)	Severe (n=48)	
Functional	2.6 ± 3.4	5.5 ± 5.9	12.9 ± 7.8	21.2 ± 7.8	12.1 ± 9.4
Physical	2.9 ± 3.8	10.8 ± 7.8	21.6 ± 7.4	28.5 ± 6.4	19.2 ± 10.2
Emotional	1.0 ± 2.2	4.6 ± 5.1	14.0 ± 8.1	19.1 ± 8.9	11.9 ± 9.3
Total	6.5 ± 8.3	20.7 ± 16.1	49.1 ± 21.0	67.0 ± 20.0	43.5 ± 26.6

Nine patients with voice disorders were not included as they did not complete the self-reported voice severity scale.

Abbreviations: N, no voice complaints; SD, standard deviation; VHI, Voice Handicap Index.

(Table 3), also indicating good to excellent internal consistency of each subscale.

The clinical validity of the Danish VHI was 0.69 using the Spearman rank correlation coefficient, indicating strong correlation between self-reported voice complaints and the VHI score. The test-retest reliability was good with ICC scores of 0.93 (95% confidence interval: 0.87–0.96) for patients and 0.78 (95% confidence interval: 0.63–0.87) for the control group, with only minor differences of 2.2 ± 9.8 for patients and 0.6 ± 5.5 for control persons between initial examinations and the 2-week re-examination.

Kruskal-Wallis and the *post hoc* Conover-Imam testing found that the three etiological subgroups of patients with voice disorders scored differently from the control group (6 ± 9 points) on total VHI scores and individual subscale scores [$P < 0.001$] (Figure 1). Further, patients with “neurogenic disorders” had higher VHI scores (60 ± 27 points) compared with patients with “structural disorders” (41 ± 26 points, $P = 0.001$) but not “functional disorders” (50 ± 21 points, $P = 0.21$). Patients with “neurogenic disorders” had significantly higher functional ($P < 0.001$) and physical ($P = 0.005$) subscale scores but not emotional subscale scores ($P = 0.01$), compared with patients with “structural disorders.” Patients with “functional disorders” did not significantly differ from patients with either “neurogenic disorders” or “structural disorders” on total VHI score or subscale scores.

DISCUSSION

The Danish version of the VHI questionnaire showed good internal consistency, test-retest reliability, and discrimination

TABLE 3.
Internal Consistency of the Danish VHI subscales

VHI scale	Number of Items	Cronbach Alpha (α)	
		Patients	Control Group
Functional	10	0.92	0.83
Physical	10	0.93	0.86
Emotional	10	0.91	0.82
Total	30	0.97	0.92

Abbreviation: VHI, Voice Handicap Index.

ability very similar to the original VHI questionnaire developed by Jacobson et al.⁴ The internal consistency of the Danish VHI questionnaire was good to excellent, indicating that the present version fully covers various aspects of the voice complaints. In addition, the test-retest reliability was excellent for patients and good for control persons. The clinical validity of the questionnaire was robust, with a strong correlation between self-reported severity of voice disorder and the VHI score. However, with a value of 0.69, the VHI questionnaire cannot be used to discriminate between patients and control persons as it was not developed for this purpose.⁴

Patients had higher scores on the physical VHI domains than on the functional and emotional domains. This is comparable with other translated versions of the VHI^{14–16} and could reflect that patients are more aware of the physical symptoms of the voice disorders. However, the presence of three individual subscales (emotional, functional, and physical) has been questioned in a study applying Rasch statistics.⁶ This study identified only two individual subscales in

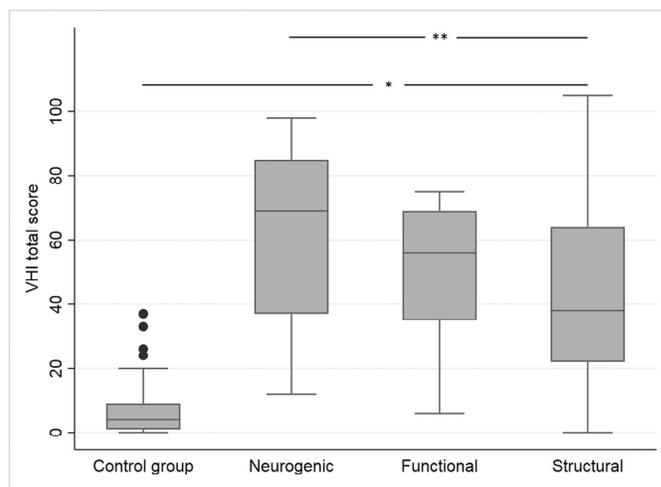


FIGURE 1. Box plot of the VHI scores for the individuals with voice disorders, subdivided by etiology, and the control group. (*) The control group was significantly different from the three groups of patients with voice disorders ($P < 0.001$). (**) Patients with a neurogenic voice disorder had a significantly higher score compared with patients with structural voice disorders ($P < 0.001$). VHI, Voice Handicap Index.

the VHI, that is, a psychosocial scale and a physical-functional scale, as many of the individual items were correlated.⁶ Lastly, the VHI scores of patients differed significantly in subgroups of patients according to the etiology of voice disorder. Patients with “neurogenic disorders” had higher VHI scores and functional and physical subscale scores compared with patients with “structural disorders.” This is consistent across VHI questionnaires^{13–16} and could reflect a real difference in impact of voice disorder on HRQoL. However, it could also indicate that patients with voice disorders of different etiology score differently on the questionnaire as the VHI would seem to be an ordinal scale and not a linear scale from the presence of a misfitting item in the questionnaire.⁶

The existing Danish version of the VHI is widely used in Denmark.^{20,21} However, this is the first study to assess psychometric properties of this VHI version. The results indicate that the existing Danish version of VHI has acceptable quality for continued use in populations of patients with voice disorders. It has shown to be a well-functioning questionnaire able to assess patients’ perception of the severity of their voice disorder.

A few limitations need to be addressed. Firstly, the patient cohort reflects the inclusion base of our hospital, with a large focus on malignancy and therefore includes more patients with structural changes on the vocal cords than seen in most places. This does hamper the generalizability of our results. Secondly, the control group was significantly younger than the patient group, which may confound the analyses. There was a high standard deviation of the VHI scores of both the patient and control group. The standard deviation of the patient group is explained by the various diagnoses of voice disorders ranging from minor voice disorders to bilateral recurrent laryngeal nerve paralyses. However, the variability in the control group was higher than expected. This is explained by the four patients (5%) in the control groups reporting moderate to severe voice complaints despite normal auditory-perceptual assessment. No videostroboscopic examination was performed in the control group; accordingly, it is possible that a subgroup of these patients had an underlying, unknown voice disorder. For future studies, it is relevant to assess how well the Danish VHI functions in the evaluation of treatment effects, that is, effectiveness/responsiveness.

CONCLUSIONS

The existing Danish version of the VHI has been thoroughly validated and found to be in line with the original VHI from Jacobsen et al. It has shown good internal consistency, test-retest reliability, and clinical validity. It is satisfactory for further use in daily practice and in research projects as it can assess patients’ perception of the severity of their voice disorder.

REFERENCES

- Benninger MS, Ahuja AS, Gardner G, et al. Assessing outcomes for dysphonic patients. *J Voice*. 1998;12:540–550.
- Behrman A, Sulica L, He T. Factors predicting patient perception of dysphonia caused by benign vocal fold lesions. *Laryngoscope*. 2004;114:1693–1700.
- Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol*. 2001;258:77–82.
- Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI) Development and Validation. *Am J Speech Lang Pathol*. 1997;6:66–70.
- Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice*. 1999;13:557–569.
- Bogaardt HC, Hakkesteegt MM, Grolman W, et al. Validation of the Voice Handicap Index using Rasch analysis. *J Voice*. 2007;21:337–344.
- Behlau M, Alves Dos Santos Lde M, Oliveira G. Cross-cultural adaptation and validation of the Voice Handicap Index into Brazilian Portuguese. *J Voice*. 2011;25:354–359.
- Cheng J, Woo P. Correlation between the Voice Handicap Index and voice laboratory measurements after phonosurgery. *Ear Nose Throat J*. 2010;89:183–188.
- Schindler A, Mozzanica F, Vedrody M, et al. Correlation between the Voice Handicap Index and voice measurements in four groups of patients with dysphonia. *Otolaryngol Head Neck Surg*. 2009;141:762–769.
- Wheeler KM, Collins SP, Sapienza CM. The relationship between VHI scores and specific acoustic measures of mildly disordered voice production. *J Voice*. 2006;20:308–317.
- Niebudek-Bogusz E, Woznicka E, Zamyslowska-Szmytko E, et al. Correlation between acoustic parameters and Voice Handicap Index in dysphonic teachers. *Folia Phoniatr Logop*. 2010;62:55–60.
- Malki KH, Mesallam TA, Farahat M, et al. Validation and cultural modification of Arabic Voice Handicap Index. *Eur Arch Otorhinolaryngol*. 2010;267:1743–1751.
- Helidoni ME, Murry T, Moschandreas J, et al. Cross-cultural adaptation and validation of the Voice Handicap Index into Greek. *J Voice*. 2010;24:221–227.
- Schindler A, Ottaviani F, Mozzanica F, et al. Cross-cultural adaptation and validation of the Voice Handicap Index into Italian. *J Voice*. 2010;24:708–714.
- Taguchi A, Mise K, Nishikubo K, et al. Japanese version of Voice Handicap Index for subjective evaluation of voice disorder. *J Voice*. 2012;26:e615–e669.
- Moradi N, Pourshahbaz A, Soltani M, et al. Cross-cultural equivalence and evaluation of psychometric properties of Voice Handicap Index into Persian. *J Voice*. 2013;27:e215–e258 e222.
- Verdonck-de Leeuw IM, Kuik DJ, De Bodt M, et al. Validation of the Voice Handicap Index by assessing equivalence of European translations. *Folia Phoniatr Logop*. 2008;60:173–178.
- Guimaraes I, Abberton E. An investigation of the Voice Handicap Index with speakers of Portuguese: preliminary data. *J Voice*. 2004;18:71–82.
- Attieh AY, Searl J, Shahaltough NH, et al. Voice restoration following total laryngectomy by tracheoesophageal prosthesis: effect on patients’ quality of life and voice handicap in Jordan. *Health Qual Life Outcomes*. 2008;6:26.
- Printz T, Sorensen JR, Godballe C, et al. Test-retest reliability of the dual-microphone voice range profile. *J Voice*. 2017. <https://doi.org/10.1016/j.jvoice.2017.03.019>.
- Grontved AM, Faber CE, Jakobsen J. [Assessment of thyroplasty for vocal fold paralysis]. *Ugeskr Laeger*. 2009;171:117–121.
- Fibiger S. Voice Handicap Index og effektuevaluering af logopædisk stemmebehandling. *Stud Audiol*. 2006;6:1–17.
- Bland JM, Altman DG. Cronbach’s alpha. *BMJ*. 1997;314:572.
- Portney LGW. *MP Foundations of Clinical Research: Applications to Practice*. New Jersey: Prentice Hall; 2000.