

# Voice Handicap Index Severity Grading: Do We Need to Recalibrate it for the Indian Population?

Unnikrishnan Menon<sup>1</sup>, Shreya Venugopal<sup>2</sup>, Deepu Chirayil<sup>3</sup>, Renjitha Bhaskaran<sup>4</sup>

## ABSTRACT

**Background:** The voice handicap index (VHI 30), used to assess patients' perception of the severity of their voice disorder, is a standard evaluation parameter in dysphonia cases. The score has conventionally set cut-offs to grade the severity. The latter will be effective only if there are appropriate responses from the patient. If not, then there may be a need to re-set the cut-offs.

**Objective:** The objective of this study is to compare and correlate laryngeal findings and voice analysis parameters with the VHI scores, to see if there is a need to recalibrate the normatives.

**Materials and methods:** Document analysis of VHI forms of patients who have visited and undergone treatment at the voice clinic at Amrita Institute of Medical Sciences. The VHI scores were correlated with the laryngeal findings and voice analysis parameters.

**Results:** As much as 64% of patients were in the normal and mild grades of the VHI score. No statistically significant correlation could be found between VHI scores and Jitter and Shimmer, by any of the attempted methods, although there was a clinical correlation in some cases. The validity of the present cut-offs of VHI severity could not be assessed.

**Conclusion:** There is a need to look at VHI scoring patterns in other population groups in India, as also to consider better statistical methods to approach the problem.

**Keywords:** Statistical analysis, Voice handicap index, Voice analysis.

*International Journal of Phonosurgery and Laryngology* (2018): 10.5005/jp-journals-10023-1160

## INTRODUCTION

The VHI 30, a 30-question tool, has been used over the years as a method to assess the degree of subjective difficulty experienced by a patient with dysphonia, in various aspects of his/her life. Normatives have been established to classify the handicap into mild, moderate, and severe, depending on the score from the questionnaire. This is then used to compare the pre- and posttreatment status of the patient, and is considered one of the standard tools to check the success of treatment, be it surgery or rehabilitation. Hence, it stands to reason that the usefulness of this tool depends on the accuracy of responses to the questionnaire. cursory observation of the VHI 30 responses in a voice clinic raised doubts in this regard.

In this study, the attempt has been to check the reliability of the VHI scoring system by comparing with objective voice analysis parameters, and to see if the normatives need to be reset for the local population.

## MATERIALS AND METHODS

Study type: Retrospective document analyses.

Study setting: Voice clinic in a tertiary care center.

VHI 30's validated Malayalam version is administered to patients attending the weekly voice clinic, initially at the first visit, followed by a repeat at reasonable posttreatment intervals. The protocol, in the case of phonosurgery for benign vocal fold lesions, is usually 1-month postoperative. In cases managed by voice therapy alone (whether voice misuse spectrum or neurologic), it would be as per the speech-language pathologist (SLP) input. Both cases would also be judged based on the videolaryngostroboscopy (VLSS) findings, with respect to posttreatment healing and/or status of mucosal waves. Dr Speech is the voice analysis software used. Out of the many parameters generated, the ones finally chosen for this study were frequency perturbation (Jitter) and intensity perturbation (Shimmer). This was after a discussion with the SLPs.

<sup>1-3</sup>Department of ENT, Amrita Institute of Medical Sciences and Research Centre, Amrita Vishwa Vidyapeetham and Amrita University, Kochi, Kerala, India

<sup>4</sup>Department of Biostatistics, Amrita Institute of Medical Sciences and Research Centre, Amrita Vishwa Vidyapeetham and Amrita University, Kochi, Kerala, India

**Corresponding Author:** Unnikrishnan Menon, Department of ENT, Amrita Institute of Medical Sciences and Research Centre, Amrita Vishwa Vidyapeetham and Amrita University, Kochi, Kerala, India, Phone: +91 9447831755, e-mail: unnikrishnanmenon8@gmail.com

**How to cite this article:** Menon U, Venugopal S, *et al.* Voice Handicap Index Severity Grading: Do We Need to Recalibrate it for the Indian Population? *Int J Phonosurg Laryngol* 2018;8(2):71-73.

**Source of support:** Nil

**Conflict of interest:** None

Statistical analysis was performed using the IBM SPSS version 20.0. To test the statistical significance of the correlation of Jitter and Shimmer with the VHI score, the Pearson correlation coefficient was used. To find the cut-off values for Jitter and Shimmer with respect to the VHI score, the ROC curve analysis was used.

## RESULTS

Sample size: 106 (M 76, F 30).

The break-up of diagnoses was as follows:

- Nodules: 44
- Cysts: 22
- Polyps: 14
- Muscle tension dysphonia (MTD): 11
- Sulcus vocalis: 8
- Unilateral palsy: 7

The association of diagnosis with VHI score is shown in Table 1. Nearly a quarter of the patients scored themselves in the normal range. As much as 64% of patients fell in the normal and mild categories. The highest percent of moderate severity was in the sulcus vocalis patients. In the mild category, unilateral vocal fold palsy had the biggest presence, followed closely by polyps and nodules. In the largest diagnosis group (vocal nodules), nearly half the cases were in the mild severity category.

The correlation of Jitter and Shimmer with the VHI score is given in Table 2. It is seen that with increasing VHI scores, Jitter and Shimmer scores are also seen to be increasing. This can be considered to be clinically significant. However, there is no statistical significance.

The cases were then divided into two groups, based on diagnosis (laryngeal findings).

Group I: Vocal cord cysts and polyps ( $n = 36$ ).

Group II: Other vocal cord pathologies ( $n = 70$ ).

A comparison of correlation of VHI and Jitter and Shimmer scores between these two groups (Table 3) showed a similar pattern in the first group, i.e., clinically significant correlation but not statistical. However, in the second group, Jitter showed a negative correlation, again without any statistical significance. Between the two groups, the correlation coefficients of both Jitter and Shimmer were higher in group I compared to group II.

Receiver operating characteristic (ROC) curve analysis of Jitter and Shimmer with the VHI score (Table 4) was checked. Here, the area under the receiver operating characteristic (AUROC) curve of the Shimmer scores (0.562, 95% confidence interval 0.343–0.780) was higher than that of the Jitter scores (0.475, 95% confidence interval 0.250–0.699) in group I. Similarly, in group II, the AUROC of the Shimmer scores (0.557, 95% confidence interval 0.384–0.731)

was higher than that of the Jitter scores (0.475, 95% confidence interval 0.310–0.657). So, the Shimmer had a slightly better predictive value for the abnormal VHI score, but with inadequate sensitivity and specificity.

## DISCUSSION

VHI is a measure of subjective disability perceived by the patient with dysphonia. It was proposed by Jacobson et al.<sup>1</sup> It is a self-assessment tool consisting of 10 items (questions) in three different domains—physical, functional, and emotional. The former contains questions that relate to the patient’s perception of laryngeal discomfort or the voice output characteristics. The functional segment includes questions to analyze the impact of a person’s voice on his daily activities. The emotional segment indicates how the voice disorder has an impact on the emotional state of the patient. The VHI has been designed to assess all types of voice disorders. For each of the 30 questions, a score of 0–4 is given. This generates a possible range of response scores from 0 to 120. Based on this score, a classification system has been described to grade the severity of the voice handicap. Accordingly, 0–30 is taken as the normal, 31–60 is taken as the mild, 61–90 is taken as the moderate, and 91–120 is taken as the severe handicap. From literature search, there is no statistical validation for this classification. However, this has been conventionally accepted and is used in all the related reportage.

The importance of the above is by way of being a marker for the success of treatment, including reporting and documentation aspects. Unlike with other diseases and conditions, dysphonia (hoarseness) does not have clear-cut objective investigation markers. VLSS, which is the “gold standard” tool, does not yield objective, measurable results. The other tool is the voice analysis software which generates quite a few parameters (numbers), but with poor clinical correlation. In this setting, the VHI score assumes great relevance. A good study has even reported the singular use of the VHI score in prognostication of unilateral vocal fold paralysis.<sup>2</sup> Needless to say, it is expected that patient response to the questionnaire will match the laryngeal pathology, and the resultant voice problem. However, this may not always be the case. This is due to two reasons:

- Subjectivity of the voice problem, which may bear no relation to the causative lesion.
- Sincerity of responding to a questionnaire.

The first factor is well known and accepted as an inherent attribute of VHI. However, the latter aspect is more intuitive than proven. Especially, in our Indian population, the conscientiousness of answering a questionnaire is often not up to the mark. And, in the case of the VHI, it has been a personal observation that in many cases, the severity of disability of a patient seen on physical examination and voice analysis parameters is not accurately reflected in the VHI. Now, for most routine clinical work, this may not be an issue, but when it comes to reporting posttreatment success rates, this becomes an obvious glitch. A patient with good recovery after microphonosurgery for a benign vocal fold lesion would end up with VHI scores in the same range before and after surgery. This set in motion a thought process regarding the possible need to reset the present cut-off values of severity grading of VHI.

The only possible way to do this would be to see if the severity grading correlated statistically with any of the voice quality

**Table 1:** Association of diagnosis with VHI score

Diagnosis (n)	VHI score grading			
	Normal n (%)	Mild n (%)	Moderate n (%)	Severe n (%)
Sulcus vocalis (8)	1 (12.5)	1 (12.5)	6 (75)	0 (0)
Cysts (22)	6 (27.3)	7 (31.8)	7 (31.8)	2 (9.1)
Nodules (44)	8 (18.2)	21 (47.7)	12 (27.3)	3 (6.8)
Unilateral palsy (7)	2 (28.6)	4 (57.1)	1 (14.3)	0 (0)
Polyps (14)	5 (35.7)	7 (50.0)	1 (7.1)	1 (7.1)
MTD (11)	2 (18.2)	4 (36.4)	4 (36.4)	1 (9.1)
Total (106)	26 (22.6)	44 (41.5)	31 (29.2)	7 (6.6)

**Table 2:** Correlation of Jitter and Shimmer with VHI score

Variables	VHI score	
	Correlation coefficient	p value
Jitter	0.001	0.994
Shimmer	0.091	0.355

**Table 3:** Comparison of correlation of VHI and Jitter and Shimmer scores between two groups of diagnoses

Groups	Variables	VHI score	
		Correlation coefficient	p value
1	Jitter	0.012	0.946
	Shimmer	0.084	0.625
2	Jitter	−0.108	0.372
	Shimmer	0.029	0.814

**Table 4:** AUROC of Jitter and Shimmer with VHI score

Group	Variables	Area	Cut-off value	Sensitivity (%)	Specificity (%)
1	Jitter	0.475	0.26	72	36.4
	Shimmer	0.562	2.28	76	36.4
2	Jitter	0.484	0.37	64.9	38.5
	Shimmer	0.557	2.65	71.9	30.8

parameters. In general, the lack of significant correlation between subjective parameters such as the VHI and the objective voice analysis software has been reported by many studies.<sup>3,4</sup> In contrast, there has been at least one study reporting a good correlation.<sup>5</sup> However, the attempt here would be to look for a numerical correlation between individual VHI scores and one appropriate voice quality measure. Dr Speech software generated the following parameters: habitual pitch, frequency perturbation (Jitter), amplitude perturbation (Shimmer), tremor, signal-to-noise ratio (SNR), harmonic to noise ratio (HNR), and normalized noise error (NNE). Discussion with the voice pathologists indicated SNR and HNR as the most representative of voice quality. However, the statistical requirement of classifying the values into “normal” and “abnormal” became an insurmountable issue, as the normatives were not adequately standardized. So, that left jitter and shimmer as the parameters which could be checked for correlation with the VHI scores.

The first striking thing on checking the results is the fact that as much as 23% of patients rated themselves as having no voice handicap, i.e., their score fell in the range of normal VHI! It merits pointing out that they all needed treatment (surgical or voice therapy), and although specific data could not be included here, they had an improvement in their VHI scores. In the mild severity group, unilateral vocal fold palsy had the biggest incidence, followed closely by polyp and nodules cases. The single biggest diagnosis group (vocal nodules) had a majority in the mild category. Sulcus vocalis cases had a majority in the moderate category. Several studies have reported different types of association between laryngeal conditions and VHI scores.<sup>6–8</sup>

The attempt at checking the correlation between the VHI scores and the selected voice analysis parameters showed merely a corresponding trend. However, there was no statistical significance. This was, in a way, a “neither here, nor there” situation, in the context of the hypothesis behind this study. The idea was to show that the VHI scores did not correlate in a linear manner with voice analysis parameters, and so, the severity cut-offs were faulty. This could not be proved or disproved by the simple Pearson correlation check. The next consideration was whether the underreporting of the VHI score would be more evident in those patients with vocal fold cysts and polyps. This arose from the empiric observation that those with nodules and MTD were more from the professional voice user category, and so tended to score higher on their VHIs. Similarly, the sulcus vocalis patients also marked their VHI forms in the severe grade. Hence, the further correlation checks of the two groups of diagnoses (cysts and polyps in one, the rest in the second). Here again, there was no statistically significant finding, apart from the fact that both correlations (Jitter and Shimmer) were slightly higher in group I as compared to group II. This too was insufficient to come to any conclusion. As per the statistician’s suggestion, we then attempted an AUROC. The ROC is usually the favored statistical method to check or establish cut-offs. But the

sensitivity and specificity for both the voice analysis values were too low to be of any help.

This study seems to have given rise to more questions than answers, mainly from the statistical methodology standpoint. The authors invite readers to contribute their suggestions to this end. On an empirical basis, we would like to suggest looking at recalibrating the normal level of VHI score from 0 to 20 and change the other severity cutoffs accordingly.

## CONCLUSION

Laryngeal findings and VHI score association seem to show a clinically disproportionate trend of a normal and mild severity grade.

It has proved difficult to establish correlations between VHI and the various software-generated voice analysis parameters.

Jitter and Shimmer (as also HNP and SNR) are all not adequately indicative of voice quality, so as to be used in objective measurements.

## RECOMMENDATIONS

Similar studies to be attempted at other institutions with voice clinic facilities.

Trial recalibration of VHI score grading, empirically with 0–20 as the normal, 21–40 as the mild, and 41–70 as moderate.

## REFERENCES

- Jacobson BH, Johnson A, et al. The voice handicap index (VHI): development and validation. *Am J Speech Lang Pathol* 1997 Aug 1;6(3):66–70. DOI: 10.1044/1058-0360.0603.66.
- Joshi AA, Singh V, et al. A prospective study to evaluate the etiologies and parameters of voice assessment in patients of vocal cord paralysis. *Otolaryngol Head Neck Surg* 2017 Sep 22;3(4):962–967. DOI: 10.18203/issn.2454-5929.ijohns20174315.
- Hsiung MW, Pai L, et al. Correlation between voice handicap index and voice laboratory measurements in dysphonic patients. *Eur Arch Otorhinolaryngol* 2002 Feb 1;259(2):97–99. DOI: 10.1007/s004050100405.
- Woisard V, Bodin S, et al. The voice handicap index: correlation between subjective patient response and quantitative assessment of voice. *J Voice* 2007 Sep 1;21(5):623–631. DOI: 10.1016/j.jvoice.2006.04.005.
- Niebudek-Bogusz E, Woznicka E, et al. Correlation between acoustic parameters and Voice Handicap Index in dysphonic teachers. *Folia Phoniatr Logop* 2010;62(1–2):55–60. DOI: 10.1159/000239064.
- Rosen CA, Murry T. Voice handicap index in singers. *J Voice* 2000 Sep 1;14(3):370–377. DOI: 10.1016/S0892-1997(00)80082-X.
- Amir O, Ashkenazi O, et al. Applying the Voice Handicap Index (VHI) to dysphonic and nondysphonic Hebrew speakers. *J Voice* 2006 Jun 1;20(2):318–324. DOI: 10.1016/j.jvoice.2005.08.006.
- Karlsen T, Sandvik L, et al. Acoustic Voice Analysis and Maximum Phonation Time in Relation to Voice Handicap Index Score and Larynx Disease. *J Voice* 2018 Aug 6. DOI: 10.1016/j.jvoice.2018.07.002.