Validation of the Lithuanian Version of the Glottal Function Index


Summary: Objective. To culturally adapt the Glottal Function Index (GFI) questionnaire to the Lithuanian language, and to validate it.

Methods. Psychometric analyses were performed on the translated Lithuanian version of the GFI (GFI-LT) as described by the Scientific Advisory Committee of the Medical Outcomes Trust. The GFI-LT was completed by 50 voice-disordered individuals and by 50 healthy subjects of the control group. Validity, reliability, reproducibility, sensitivity, and responsiveness to clinical change of the GFI-LT were evaluated. To assess the concurrent validity of the GFI-LT, all participants also completed the Voice Handicap Index (VHI) questionnaire.

Results. The GFI-LT showed a statistically significantly high reliability and internal consistency (Cronbach α = 0.8, r = 0.50), and moderate item-total correlation (r = 0.41–0.55). Cronbach α coefficients of the test-retest reliability were above the standard (≥0.9) for individuals testing. There was a statistically significant difference between the mean scores of the control and the voice-disordered groups (P < 0.001). The Receiver Operating Characteristic test indicated that the GFI-LT score of >3.0 was the optimal score distinguishing patients and healthy controls with the sensitivity of 88% and specificity of 84%. Statistically significant (P < 0.05) strong correlations were found between the GFI-LT and VHI scores. The GFI-LT was found to be a responsive measurement instrument to patients’ clinical statement; mean difference of the GFI-LT scores in the group of voice-disordered patients before and after surgical treatment was 5.7 (P < 0.001).

Conclusion. The GFI-LT is considered to be a valid and reliable tool for self-assessment of the severity of voice disorders in Lithuanian-speaking patients.

Key Words: Glottal function index—Voice—Dysphonia—Voice disorders—Questionnaires.

INTRODUCTION

Voice disorders affect approximately 3–9% of the population and have a profound impact on patients’ quality of life.1,2 Because voice disorders involve different aspects, clinical voice assessment should consist of several measurements: video laryngostroboscopy, perceptual voice assessment, objective instrumental measurements (acoustic voice analysis), and subjective patient’s self-evaluation of voice disorder.

Numerous patient-reported measures have been developed to quantify the impact of voice disorders on patient’s quality of life, and to evaluate treatment outcomes—Voice-Related Quality of Life (V-RQOL),3 Voice Outcome Survey,4 Voice activity and Participation Profile,5 Voice Symptom Scale,6 and Voice Handicap Index (VHI).7

The VHI introduced by Jacobson et al in 1997 was the first 30-item questionnaire specifically developed with the purpose of quantifying the impact perceived by an individual affected by a voice disorder in the spheres of voice function itself, voice-related physical ability, and emotions provoked by the dysphonia.1 Since then, it has been established that VHI is the most versatile instrument and one that contains the most relevant information about V-RQOL.8 The VHI has gained a wide popularity in both clinical practice and research and was validated in 12 different languages.9–20 Verdonck-de Leeuw et al21 evaluated the equivalence of the US VHI version and several translations into eight European languages. They established that the US VHI, and the translations appeared to be equivalent and that the results from the studies from various countries were comparable. Shortened versions of the VHI, which appear to be more clinically feasible, that is, 10-item VHI-1022 and 9-item VHI version23 well approximate the full version of VHI.

In 2005, Bach et al24 developed and validated the Glottal Function Index (GFI) questionnaire—easily self-administered and reliable 4-item battery that was designed to assess the presence and degree of vocal dysfunction in adults. According to the data of the authors, a strong correlation between GFI and VHI has been established. However, only two articles related to the GFI have appeared in the literature so far. Cohen et al25 presented results of the validation of the Hebrew version of the GFI and proved the utility of this tool to improve the evaluation of children with voice disorders—especially those with vocal fold lesions. Buckmire et al26 confirmed the usefulness of the GFI to determine the effectiveness of Gore-Tex medialization thyroplasty.

Because the patient’s individual perception of the voice disorder is not always related to the quality of the voice but is rather more related to professional and social vocal demands as well as to personal aspects, the GFI may have additional value in the self-examination of the patient’s voice.

This work presents the results of the validation process of the Lithuanian version of the GFI questionnaire (GFI-LT).
MATERIAL AND METHODS

Development of the GFI-LT

The validation process of the GFI-LT was performed during 2008–2009 at the Department Otolaryngology, Academy of Medicine of Lithuanian University of Health Sciences, Kaunas, Lithuania.

Forward and back-translation (English-Lithuanian-English) of the original GFI questionnaire, comparison of back-translations with the source version, review by lay panels, committee review, and psychometric testing of the translated version were performed as described by the Scientific Advisory Committee of the Medical Outcomes Trust.27

An original GFI questionnaire (Appendix A) was translated into Lithuanian by two bilingual speech-language pathologists, and the back-translation was done by an English teacher who had not participated in the previous stage of the procedure. All the translators were informed about the objective and the procedure of the research. After translation of the questionnaire, 20 persons from the medical staff of the Department of Otolaryngology reviewed the GFI-LT; they did not have any comments. After that, a committee of three voice specialists (the authors of this study: R.P., V.U., and A.V.) revised the final version of the questionnaire. To evaluate, the cultural and linguistic equivalency of the loan-translation, the option “not applicable” was introduced to each item of the questionnaire, which was then administered to the patients (N = 30, 14 males and 16 females, mean age 35.8 ± 14.4 years), who did not take part in the final application of the translated version of the GFI. None of the questions was shown to be invalid, and therefore, the GFI-LT was accepted (Appendix B).

Validation of the GFI-LT

Participants. The GFI-LT was completed by 50 voice-disordered individuals and by 50 healthy subjects of the control group. Characteristics of the patients and the controls are presented in Table 1. As it can be seen in Table 1, the studied groups were homogenous with respect to the individuals’ age, gender, and voice training.

The voice disorders group represented rather common, clinically discriminative groups of laryngeal diseases—mass lesions of vocal folds and unilateral laryngeal paralysis. Mass lesions of vocal folds included in the study consisted of noduli, polyps, cysts, papillomata, chronic laryngitis with keratosis, and Reinke edema. Fifty patients in the voice disorders group (28 males and 22 females; mean age 41.2 ± 5.7 years) were recruited from the consecutive patients who were diagnosed with the aforementioned laryngeal diseases. All the patients underwent clinical examination that included perceptual and acoustic voice assessments. The clinical diagnosis was based on typical clinical signs revealed during video laryngostroboscopy and direct microlaryngoscopy. All the patients with mass lesions of vocal folds underwent endolaryngeal microsurgical interventions; therefore, the final diagnosis was proven by the results of the histological examination of the removed tissue. The nodular lesions of the vocal folds (polyps, noduli) were the most common entities and comprised 60% (30) of the cases.

The control group—the subjects with unaffected voice—consisted of 50 randomly selected volunteer individuals (20 males and 30 females; mean age 39.7 ± 10.1 years) who considered their voice to be normal and who had no complaints concerning their voice, had no history of chronic laryngeal diseases or other long-lasting voice disorders, and had never seen an otolaryngologist for voice problems. The voice of this group of individuals was also evaluated as healthy voice by clinical voice specialists. No pathological alterations in the larynx of the subjects of the control group were found during video laryngostroboscopy. Acoustic voice-signal parameters of the control group subjects obtained using Tiger Electronics (Seattle, WA) Dr. Speech software (Voice Assessment, Version 3.0) were within the normal range.28

Methods. Each participant of the study (patients and controls) filled in the GFI-LT questionnaire at the baseline—at least one week before the treatment. Then, to determine the test-retest reproducibility of the GFI-LT, the voice-disordered patients were asked to fill in the GFI-LT questionnaire for a second time within the interval of one week. This interval was chosen according to the literature data where the most effective retest interval is usually recommended to be short enough so that not many changes in the answers would occur but long enough so that patients would not remember their previous answers. Typically, it is between 2 and 14 days after the first filling in of the questionnaires.29 Responsiveness of the GFI-LT to clinical change was evaluated comparing the questionnaire scores of pre- and posttreatment (1 month after).

The Lithuanian version of the VHI was used to establish the concurrent validity of the GFI-LT. The VHI is the only internationally recognized self-report instrument for voice-disordered patients translated into Lithuanian. The VHI contains detailed subscales of three possible voice handicap domains (physical, functional, and emotional).
To evaluate correlations between the GFI-LT and VHI, both questionnaires were filled in at the baseline and 1 month after the treatment. The subjects had to fill in the questionnaires independently (ie, without any assistance).

**STATISTICAL ANALYSIS**

Statistical analysis was performed using the statistical package SPSS, Version 13.0 (2006; SPSS, Inc, Chicago, IL). The significance level of 0.05 was chosen for testing statistical hypotheses. The Kolmogorov-Smirnov test was used to test the normality of the distribution of the GFI-LT scores. Floor and ceiling effects were assessed by calculating the frequency of minimal and maximal values of GFI-LT in patients and controls and were expressed as the percentage of all minimum (0) and maximum (5) responses. The internal consistency of the GFI-LT was evaluated by Cronbach $\alpha$ and by Pearson correlations within the scale. Concurrent validity of the test was assessed by calculating mean correlation coefficients of the GFI-LT and VHI. The test-retest reliability was measured by Spearman correlations, Cronbach $\alpha$, kappa agreement, and intraclass correlation coefficients. Independent sample $t$ test was used to determine the discriminatory power of the GFI-LT between the patients’ and the control groups. Paired sample $t$ test was used to assess the responsiveness of the GFI-LT to clinical change. To test the accuracy of the GFI-LT and the limiting score on which the sensitivity and specificity curves cross when identifying voice disorders, Receiver Operating Characteristic (ROC) curve analysis was used. Graphical trade-off between false-negative (sensitivity) and false-positive (specificity) rates for every possible cutoff was presented. The test accuracy was calculated to explain the proportion of true results (both true positives and true negatives) in the study sample. The crude odds ratio and its 95% Confidence Interval (CI) were calculated to explore the probability of the study sample. The crude odds ratio and its 95% Confidence Interval (CI) were calculated to explore the probability of the study sample. The crude odds ratio and its 95% Confidence Interval (CI) were calculated to explore the probability of the study sample. The crude odds ratio and its 95% Confidence Interval (CI) were calculated to explore the probability of the study sample.

**RESULTS**

**Psychometric properties of GFI-LT**

Table 2 represents descriptive statistics of GFI-LT scores, including mean, standard deviation, median, range, and the percent scoring at the lowest possible value (floor) and the highest possible value (ceiling). Floor and ceiling effects were 2.0% and 4.0% in voice-disordered patients group, and 56.0% and 0.0% in healthy controls. The Kolmogorov-Smirnov test confirmed normal distribution of the GFI-LT scores in patients but not in controls.

Good internal consistency of the questionnaire satisfied the requirement for positive and significant correlation between the items (Tables 3 and 4). Cronbach $\alpha$ coefficients of internal reliability were above the standard ($\geq 0.7$ for groups) in patients and controls. No higher correlations than 0.50 were received in the GFI-LT correlation matrix (Table 4). It suggests that there were no redundant items in the questionnaire. Correlations among the responses were higher in the patients’ group as compared with the healthy control group; however, these differences were not statistically significant. Frequency analysis in the responses categories showed that more than half of healthy controls scored zero (no voice disorder symptoms), eight persons scored 1 in one item and only few of them scored 3 or 4 to one or two questions of the questionnaire. That served as a reason of lower correlations inside the GFI-LT. In voice-disordered patients group response categories were more consistent inside the GFI-LT.

Good test-retest reliability of the GFI-LT was confirmed by Spearman correlation coefficients ranging from 0.86 in item 1 to 0.94 in item 2 ($P < 0.001$) (Table 5). Cronbach $\alpha$ coefficients of the test-retest reliability were above the standard ($0.9$) for the tested individuals. The intraclass correlations of 0.85–0.93 confirmed high reproducibility of the GFI-LT ($P < 0.0001$).
Cohen Kappa coefficient showed substantial agreement between test and retest ratings of 0.66–0.80, *P* < 0.0001. The discriminatory power of the GFI-LT is demonstrated in Table 6. The mean score of the GFI-LT was found to be higher in patients with vocal disorders before and after surgery compared with healthy controls (mean difference, −7.8 and −2.1 points, respectively; *P* < 0.01).

Concurrent validity of the GFI-LT
Statistically significant (*P* < 0.05) strong correlations were found between the GFI-LT and VHI total scores both before (*r* = 0.72) and 1 month after the surgery (*r* = 0.86). Mean, minimal, and maximum values of all possible correlations between the scores of the GFI-LT and VHI scales, and total VHI before surgery are shown in Table 7. The highest concurrent correlation was revealed between the scores of the GFI-LT and the functional scale of the VHI.

Responsiveness of the GFI-LT to clinical change
Responsiveness of the GFI-LT to clinical change was found to be statistically significant. The difference of the GFI-LT mean values in the group of voice-disordered patients before and after surgical treatment was 5.7 (*P* < 0.001). The same tendency with the VHI was established—the mean VHI value after surgery decreased statistically significantly (Table 8).

Limiting value of GFI-LT
ROC test indicated that the GFI-LT score of >3.0 was the optimal score (limiting value) distinguishing patients and healthy controls (Figure 1). Of 50 patients with vocal disorders, the GFI-LT correctly identified 44 individuals (sensitivity, 88%), while 42 disease-free cases were correctly classified as negative (specificity 84%). When using the GFI-LT limiting value (>3.0), classification accuracy of 86% was detected. Odds ratio for limiting value of the GFI-LT was 38.5 (95% CI, 12.32–120.36).

DISCUSSION
In recent decades, research has proven the importance and usefulness of the use of patient-centered measurements of voice disorders along with the objective and perceptual measures for the assessment of dysphonia and treatment outcomes. Several patient-reported outcome measures have been developed to quantify the impact of voice disorder as well as to evaluate the outcomes from the patient’s attitude and to guide informed therapeutic decisions.3–7

The GFI questionnaire has been used in 2005 at the Center for Voice Disorders of Wake Forest University (Winston-Salem, NC) and was initially conceived as an instrument for evaluating glottal insufficiency and its response to therapy. After administering the GFI to 120 patients with various voice disorders (vocal nodules, adductor spasmodic dysphonia, and granuloma), the authors have found the GFI to be useful for detecting the presence and the severity of voice dysfunction and the response to treatment in a variety of voice disorders.24 Our study showed that the mean scores of the GFI-LT of consecutive patients with various voice disorders before the treatment...
(9.1 ± 5.2) were statistically significantly higher (P < 0.001) than those in the control group of healthy persons (1.3 ± 1.9). It should be noted that there was no statistically significant difference (P = 0.26) between the mean GFI-LT score (1.3 ± 1.9) of the control group in our study and that mentioned by Bach (0.9 ± 1.3).

Based on the normative data, Bach et al.24 considered the GFI score higher than 4.0 (mean ± 2 SDs) to be abnormal. Accordingly, the ROC test in our study indicated that GFI-LT score higher than 3.0 was found to be the limiting value distinguishing patients and healthy controls with the sensitivity of 88% and specificity of 84%, respectively. The same “optimum” score of 3.0 using ROC curves was revealed by Cohen et al.25 during the validation of the GFI in children. Of the 54 patients who had vocal disorders (the most common disorder was vocal fold nodules), the GFI index identified 38 patients (70%). The sensitivity of the GFI was 70% and the specificity—72%.

The results of our study showed a statistically significant (P < 0.05) strong correlation between the GFI-LT and the VHI, confirming a good concurrent validity of the GFI-LT. Pearson correlation coefficient between the GFI-LT and the VHI in voice-disordered patient group at the baseline was r = 0.72 and 0.86—one month after the phonamicrosurgical treatment. Similar results were presented by Bach et al.24 demonstrating a mild correlation 0.61 (P < 0.001) between the GFI and the VHI. Recent study shows a moderate correlation (r = 0.71) between the GFI and the V-RQOL scores.26

The GFI-LT revealed good responsiveness to clinical change as well. Statistically significant difference (P < 0.01) between the mean GFI-LT scores before (9.1 ± 5.2) and after surgery (3.4 ± 4.5) were found in voice-disordered patient group. Consequently, the mean difference of the GFI-LT scores in the group of voice-disordered patients before and after surgical treatment was 5.7 (P < 0.001). Good responsiveness of the GFI was demonstrated in a study by Buckmire et al.26 when assessing surgical effectiveness (mean follow-up time of 7.8 months). Statistically significant (P < 0.01) difference between pre- and postoperative GFI was found.

According to the results of the present study, the patients see GFI-LT as a brief, easily understandable, symptom-focused, and easy to fill form for the identification and self-assessment of voice disorders. Moreover, the linguistically validated GF-LT embodies culturally relevant patient perspectives. This is of great importance for the 3.3 million-population of Lithuania because—according to the survey Europeans and their Languages in 2005—only 32% of the population in Lithuania speak English.30 Moreover, there are about one million Lithuanian speakers living abroad31 and—according to the American Community Survey in 2006—more than 722.8 thousands of Lithuanian-speaking individuals in the USA.32 Presumably, the incidence of voice disorders in Lithuania corresponds to the international data.12

Good psychometric properties of the Lithuanian version of the GFI and its responsiveness to clinical change suggest that it is a valid, sensitive, and suitable instrument for the evaluation of patients’ self-perception of voice disorders and is considered to be a good follow-up assessment tool in clinical practice for Lithuanian-speaking individuals.

REFERENCES