Validity and Reliability of the Glottal Function Index

Kevin K. Bach, MD; Peter C. Belafsky, MD, PhD; Kathleen Wasylik, MD; Gregory N. Postma, MD; Jamie A. Koufman, MD

Objective: To evaluate a symptom-focused vocal impairment instrument for the evaluation of patients with voice disorders.

Design: Prospective, nonrandomized study of patients with voice disorders undergoing treatment with validation of a new symptom index, the Glottal Function Index (GFI).

Setting: Voice disorders clinic at an academic tertiary care hospital.

Patients: Consecutive patients undergoing therapy for glottal insufficiency, adductor spasmodic dysphonia, nodules, and granuloma (40 patients in each group) and 40 control patients.

Interventions: The Pearson correlation coefficient was used to evaluate GFI reproducibility and to compare it with the Voice Handicap Index (VHI). The paired-samples t test was used to compare pretherapy and posttherapy GFI values.

Main Outcome Measures: Correlation of GFI with VHI; comparison of the GFI in normals, and in pretherapy and posttherapy GFI and VHI scores.

Results: The mean ± SD normative GFI score was 0.87 ±1.32. The correlation coefficient for GFI between independent pretherapy measurements was 0.56 (P<.001). The correlation coefficient between total GFI and total VHI scores was 0.61 (P<.001). The mean posttherapy GFI scores improved among all groups as follows: glottal insufficiency: presenting GFI score, 12.7±4.1; posttherapy GFI score, 6.8±5.4; nodules: presenting GFI score, 12.9±4.2; posttherapy GFI score, 8.9±4.6; adductor spasmodic dysphonia: presenting GFI score, 13.2±4.1; posttherapy GFI score, 8.9±4.9; and granuloma: presenting GFI score, 7.8±4.6; posttherapy GFI score, 3.8±2.1. Relative to controls, the GFI score at presentation was significantly elevated and demonstrated significant reduction following treatment across each of these entities (P<.05).

Conclusions: The GFI is a reliable, reproducible, 4-item, self-administered symptom index with excellent criterion-based and construct validity. Its advantages over existing indexes include brevity and ease of administration. The GFI is a useful adjunct in the evaluation and treatment of patients with glottal dysfunction.

Glottal Function Index (GFI) is a broad and lengthy battery designed to evaluate the impact of a voice disorder on a variety of levels, including emotional, functional, and physical. The Voice Outcome Survey, reported by Gliklich et al., is a shorter and more easily administered index but lacks specific focus on the symptoms of a voice disorder. The Voice-Related Quality of Life instrument was designed by Hogikyan et al. to assess dysphonia’s impact on perceived life quality. We have routinely employed the VHI and a less cumbersome, more specific dysphonia instrument was designed by Hogikyan et al. to assess dysphonia’s impact on perceived life quality. We have routinely employed the VHI and a less cumbersome, more focused, and easily self-administered 4-item battery, the Glottal Function Index (GFI), which is aimed specifically at identifying the presence and degree of symptoms of glottal dysfunction. The scale ranges from a minimum score of 0 (asymptomatic) to a maximum score of 20. This study is designed to evaluate the validity and reliability of the GFI and to assess its utility in evaluating patients presenting with dysphonia resulting from a variety of clinical entities and following the treatment thereof.

METHODS

The GFI has been employed at the Center for Voice Disorders of Wake Forest University for more than a decade and was initially conceived of as an instrument for evaluating glottal insufficiency and its response to therapy. This study was originally aimed at this group of disorders and included patients with vocal fold paralysis, paresis, and presbylaryngis. While critically evaluating its specificity, a potentially broader application was recognized, and other voice disorders were included for examination.

Forty consecutive patients undergoing laryngoplasty for glottal insufficiency from July 1, 2000, to June 30, 2001, were enrolled. At presentation, all study participants underwent a complete medical history and physical examination, including fiber optic laryngoscopy and videostroboscopy, and completed the GFI and VHI. All patients completed a second preoperative GFI on a subsequent visit. Laryngeal electromyography was selectively used to verify the clinical impression of subtle vocal fold paresis and to differentiate vocal fold paralysis from fixation. Acoustic and aerodynamic analysis with electroglottography was also selectively employed in subtle cases of glottal insufficiency, particularly those characterized by significant muscle tension patterns or hyperkinetic laryngeal behaviors.

These patients underwent unilateral or bilateral medialization laryngoplasty (ML) using either Silastic (Dow Corning, Midland, Mich) or Gore-Tex (W. L. Gore & Associates, Newark, Del) implantation (based on surgeons’ preference), with or without arytenoid adduction (AA), by 1 of the 2 senior authors (J.A.K. and G.N.P.). These techniques were developed by Ishikawa et al., and their modifications and applications at this institution have been detailed elsewhere. Following surgery, patients were examined using fiber optic laryngoscopy with stroboscopy and again completed the GFI.

Specificity of the GFI was analyzed by administering the index to 120 patients (40 in each group) with voice disorders that were not necessarily characterized by glottic insufficiency, specifically vocal nodules, adductor spasmodic dysphonia, and granuloma. The index was administered at initial presentation and following a course of therapy directed at their specific clinical entity, which included standard medical and/or operative treatment, alone or in combination. Normative data were established by administering the GFI to 40 asymptomatic control subjects, each without a known history of voice disorder, laryngopharyngeal reflux, or neck surgery.

The Pearson product-moment correlation coefficient was used to evaluate the linear association among GFI measures. The paired-samples t test was used to compare pretherapy and posttherapy GFI values and to compare these values against controls. The GFI was compared with the VHI using the Pearson correlation coefficient for both total and subscale scores. All data were coded and recorded using SPSS statistical software, version 6.1.1 for the Macintosh computer (SPSS Inc, Chicago, Ill). Data are presented as mean ± SD.

RESULTS

The median age of the normative control group was 39 years; 50% were female. The mean GFI score was 0.87 ± 1.3 (0.80 at the first administration and 0.94 at the second, 2 weeks later). This value was significantly lower than both the pretherapy and posttherapy GFI scores for the study cohort (P < .05).

The Pearson correlation coefficient between GFI and total VHI scores was 0.61 (P < .001). For the subscales, Pearson correlation coefficients were as follows: physical, 0.60; functional, 0.58; and emotional, 0.55 (P < .001 for all). Patients with a GFI score greater than 5 were 3.5 times more likely to have a total VHI score greater than 30 (95% confidence interval, 1.65–7.41).

The median age of the laryngoplasty cohort was 49 years, and 63% were female. The Table gives the data for the cause of glottal insufficiency. The initial preoperative GFI score was 12.7 ± 4.1, and the mean second preoperative GFI score was 12.5 ± 3.9. The mean ± SD interval between these administrations was 3.0 ± 0.76 months (range, 2.1–6.2 months), and the correlation coefficient for the total GFI score between these administrations was 0.56 (P < .001). The single-item coefficients ranged from 0.40 to 0.66. The mean glottal insufficiency GFI im-
proved to 6.8±5.4 (mean follow-up, 3.5 months; P<.001) after laryngoplasty. The GFI values for other voice disorders were as follows: nodules: presenting GFI score, 12.9±4.2; posttherapy GFI score, 8.9±4.6; adductor spasmodic dysphonia: presenting GFI score, 13.2±4.1; posttherapy GFI score, 8.9±4.9; and granuloma: presenting GFI score, 7.8±4.6; posttherapy GFI score, 3.8±2.1. The mean course of therapy between GFI measurements was 4.1 months, with a range of 3.1 to 6.7 months. Relative to normal controls, the GFI score at presentation was elevated, and its reduction following treatment across each of these entities was statistically significant (P<.05).

The use of statistically validated and reliable outcome instruments provides objective insight into the patients’ impressions of their initial disability and their perceived benefit following surgery. Presently, several such voice-specific instruments exist. The VHI is a 30-item self-administered battery designed to quantify various aspects of impairment resulting from a diverse group of voice disorders. The VHI is divided into 3 subscales: functional, to assess the social and economic impact of the patient’s disability; emotional, to assess the affective impact of the disability; and physical, to assess the symptomatic component of the disability. In our experience, it requires significant time and patience to complete and often goes unfinished by the patient in a busy clinical environment already rife with administrative paperwork for the patient to complete. A shorter battery, the Voice Outcome Survey, was designed as a disease-specific instrument to evaluate patients with unilateral vocal fold immobility. It contains 5 items, but only 2 are directed specifically at the symptoms of glottal insufficiency: effortful phonation (or vocal strain) and aspiration. Of the remaining 3 items, 1 asks patients to grade the severity of their dysphonia (nonspecific), and 2 relate to the psychosocial aspects of their disability.

The GFI (Figure) contains only 4 items, each aimed at assessing symptoms of glottal dysfunction. We have routinely administered it to patients who present with dysphonia and have found it to be useful in detecting the presence and severity of glottal dysfunction, in addition to assessing the response to treatment in a variety of voice disorders. For the patient, the GFI is brief, easily understood, and readily completed. If information regarding other facets of vocal disability is desired, lengthier formats such as the VHI, the Voice-Related Quality of Life instrument, or general quality-of-life instruments such as the 36-Item Short Form Health Survey may be employed.

Surgery of the laryngeal framework for glottal incompetence has become widely accepted and extensively described since the first report by Isshiki in 1974. Reports of improved postoperative glottal competence and function by objective as well as subjective measures are prevalent. Injection augmentation techniques are diversely and successfully employed for similar indications. Although many investigators recognize the importance of patients’ perceptions regarding the resolution of their symptoms and the improvement in their functional status following laryngoplasty, relatively few formal evaluations of these perceptions have been undertaken. Alternatively, postoperative assessment has included acoustic and aerodynamic voice laboratory data as well as evaluation by videostroscopy and even magnetic resonance imaging. Blinded evaluation of voice samples by voice professionals has also been used. While these methods may be objective, their ability to completely characterize the patient’s postoperative experience may be limited. Although one would expect improvements in these measures to correlate with improved outcomes, severity of disease as gauged by these tests does not necessarily measure the impact of disability on a given individual (eg, the professional vocalist who may remain devastated by a miniscule residual decline in dynamic range following surgery). Moreover, these methods are generally costly and cumbersome to apply in the routine evaluation of these patients.

Improvement in aerodynamic and acoustic function following laryngoplasty supports the premise that improving glottal competence improves the efficiency and quality of the voice and may substantiate the subjective reduction in symptoms reported on indices such as the GFI. Several studies have demonstrated significant improvements in these parameters, including a decrease in the mean flow rate and an increase in the maximal phonation time following surgery, whereas effects on other measures have been highly variable. Spector et al identified an increase in the vocal intensity and range of these patients postoperatively, suggesting that appreciable dynamic vocal control is regained following surgery. Notably, many preoperative aerodynamic and acoustic parameters in patients with known glottal incompetence (eg, unilateral true vocal fold paralysis) are within the normal range despite patients’ subjective dysfunction, and this may limit the utility of these measures in assessing postoperative outcomes. Normal parameters in this group may result from compensatory or hyperkinetic laryngeal behaviors that exist preoperatively and/or reflect the broad range of “normal” values for these tests.

It is important to recognize that while incompletely compensated glottal insufficiency is often evidenced by typical symptoms, glottal insufficiency that is well compensated by abnormal laryngeal muscle tension patterns may be without these same symptoms. Spector et al again noted a significant decrease in the fundamental frequency of glottal insufficiency in patients who had undergone laryngoplasty and proposed that this may result from the resolution of compensatory muscle tension patterns or hyperkinetic laryngeal behaviors following the surgical restoration of glottic competence. The physician and clinician should be aware that the finding of abnormal muscle tension patterns on examination, even in the absence of typical symptoms, may indicate an underlying glottal closure problem.

Patients undergoing ML, both with and without AA, were included for study on the basis of the procedures’ common surgical goals, specifically, closure of the glottic gap. Arytenoid adduction is typically used only for cases of unilateral vocal fold paralysis in which there is
a significant posterior component to the glottic gap. The decision to perform AA is usually based on the intraoperative evaluation of phonation and observation of a residual posterior glottic gap following ML. We believed that prospectively studying only one of these groups (ie, ML with AA or ML alone) would not only be technically difficult, given the inability to reliably predict the operative procedure in advance, but would also be of limited significance in evaluating postoperative outcomes as both are aimed at restoring glottal competence.

The GFI demonstrates excellent test-retest reliability, comparable to that of the VHI. It also demonstrates good criterion-based validity because it is consistently and significantly lower following both traditional medical and surgical therapy for the voice disorders evaluated in this study. Based on our normative data, we consider a GFI score greater than 4 (mean + 2 SDs) to be abnormal. The GFI shows good correlation with the VHI, giving it good construct validity. Although previously available surveys are comparable, the GFI provides significant practical advantages over them in that it is brief, symptom focused, and easily completed. As with other vocal impairment batteries, the GFI is not specific for any single clinical entity and shows utility for a variety of voice disorders.

The GFI should serve only as an adjunct to other historical, acoustic, and physical examination findings. It appears to be most useful as an evaluative instrument before and after therapy aimed at improving glottal function and is perhaps less so as an outright diagnostic tool. A vocal professional with “normal” acoustic parameters and stroboscopic findings may present with a GFI score of 18, whereas a laryngectomee with a newly fashioned voice prosthesis may have a GFI score of only 3. We wish to emphasize, however, the clinical utility of this pragmatic, symptom-oriented survey in the comprehensive evaluation of abnormalities of glottal dysfunction.

In conclusion, the GFI is a 4-item, self-administered outcome instrument with good reproducibility and excellent criterion-based and construct validity. It is brief, easily administered, focused, and reliable. It is a useful adjunct in the initial evaluation and follow-up of patients with voice disorders and accurately documents improvement following therapy.

Submitted for Publication: February 22, 2005; final revision received May 27, 2005; accepted May 31, 2005. Correspondence: Jamie A. Koufman, MD, Center for Voice Disorders, Department of Otolaryngology, Wake Forest University, Winston-Salem, NC 27157-1034 (jkoufman@wfubmc.edu).

Financial Disclosure: None.

Previous Presentation: This study was presented at the annual meeting of the Western Section of the Triological Society; February 3, 2002; Pasadena, Calif.

REFERENCES