Outcomes Analysis of Voice and Quality of Life in Patients With Laryngeal Cancer

Michael G. Stewart, MD, MPH; Amy Y. Chen, MD; Carol B. Stach, MA, CCC-SLP

Objective: To assess relationships between voice satisfaction and global quality of life in patients who have been treated for laryngeal cancer.

Design: Cross-sectional survey study.

Setting: Veterans Affairs Medical Center.

Patients: Eighty patients who had completed treatment for laryngeal cancer with either total laryngectomy (n=17), radiotherapy (n=24), or both (n=39).

Main Outcome Measures: Subscale scores on a general health status instrument (the Medical Outcomes Study 36-item short-form health survey), and a validated voice-specific functional status instrument (the Voice Handicap Index).

Results: Self-rated global health did not correlate significantly with emotional, functional, or physical voice handicap, although some subscales on the 36-item short-form health survey correlated with voice handicap scores.

Global health status scores did not differ between patients who had undergone laryngectomy with a tracheoesophageal puncture and patients treated with radiotherapy only. Physical voice handicap scores did not differ significantly between those who underwent tracheoesophageal puncture and those who had radiotherapy, but emotional ($P=.07$) and functional ($P=.01$) handicap scores were lower in patients treated with radiotherapy. However, there was considerable overlap in voice handicap scores, with many patients who had had tracheoesophageal puncture showing less voice handicap than patients treated with radiotherapy.

Conclusions: These data demonstrate that health status is affected by other factors than voice handicap in patients with laryngeal cancer. In addition, there is a large amount of individual variation in voice handicap after treatment. These findings illustrate the need for prospective studies assessing voice handicap and quality of life after treatment for laryngeal cancer.


HEAD AND neck cancer strikes at some of the most basic human functions, including verbal communication, social interaction, eating, and breathing. Although the traditional treatment for laryngeal cancer has been surgical excision (total laryngectomy), there is increasing evidence that, for advanced tumors, radiation therapy (XRT) with or without systemic chemotherapy (followed by “salvage” laryngectomy, if necessary) offers equivalent short-term and long-term survival rates to treatment with laryngectomy.¹² Because of the widespread perception of significant decrements in overall functional status, ability to communicate, and quality of life (QOL) caused by total laryngectomy, these voice-sparing treatment protocols, which allow some patients to preserve their larynx, have become increasingly popular.

However, anatomic preservation of the larynx does not guarantee that its function will remain intact. In addition, contemporary voice rehabilitation following total laryngectomy has become fairly advanced, and many patients have intelligible speech after appropriate rehabilitation with a speech and language pathologist. Speech without a larynx (alaryngeal speech) is often intelligible and enables many patients who are treated with laryngectomy to communicate adequately using esophageal speech, an electrolarynx, or a tracheoesophageal puncture (TEP).³⁴

To better assess the relationships between voice-related functional status and global health status (QOL) in patients with laryngeal cancer, we performed a cross-sectional study in patients who had been treated with laryngectomy or radiotherapy, or both. Functional status refers to the ability to perform typical daily ac-
METHODS

We performed this cross-sectional study by requesting volunteers from eligible patients being seen at a head and neck tumor follow-up clinic at the Veterans Affairs Medical Center, Houston, Tex, between January 1996 and April 1996. Patients were eligible if they had been treated for laryngeal cancer with either total laryngectomy, partial laryngectomy, external beam XRT, chemotherapy, or any combination of treatments, and treatment had been completed at least 1 month prior to study enrollment. Patients currently undergoing treatment for laryngeal cancer were not eligible. Patients were excluded if they had recurrent or residual laryngeal cancer, but not excluded if they had lymph node or distant metastasis.

DATA COLLECTION

We used the VHI and a global health status instrument, the Medical Outcomes Study 36-item short-form health survey (SF-36).12 The VHI is divided into 3 subscales, each with a separate score; there is no summary or overall score for the VHI. Each subscale addresses a different construct (or dimension) of verbal communication handicap: functional handicap, emotional handicap, and physical handicap. Higher scores indicate higher levels of voice handicap, or worse functional status. The range of scores for each subscale—functional, emotional, and physical—is 0 to 40. Similarly, the SF-36 is divided into 8 subscales that address different constructs of global health status: bodily pain, general health, social functioning, physical functioning, physical role functioning, emotional role functioning, mental health, and vitality. There is no overall or summary score for the SF-36, and higher subscale scores on the SF-36 indicate better QOL. The range of scores for each subscale is 0 to 100.

All patients completed a data sheet and both the VHI and SF-36. No patient entered the study more than once. In most instances, patients completed the data sheet and questionnaires directly on the question sheet, but in a few cases where the patient had visual loss or a low literacy level, the instruments were given by interview. The patient data sheet contained questions about the patient’s employment status, marital status, residency status (eg, lives alone or lives with family), highest level of education, and self-perceived overall health status. The medical records were then reviewed for pertinent clinical data, such as age, date of treatment, and original TNM tumor stage. In addition, the speech pathology records from the Veterans Affairs speech pathology service were reviewed to retrieve the pretreatment assessment by speech pathologists of the potential for successful voice rehabilitation after treatment.

STATISTICAL ANALYSIS

Data were entered into a spreadsheet in SPSS version 7.0 (SPSS Inc, Chicago, Ill). To assess relationships between item responses and subscale scores, Spearman correlation coefficient (the nonparametric analog to the Pearson correlation coefficient) was used. To compare means between continuous variables, the Student t test or 1-way analysis of variance was used. To compare medians for interval data, the Mann-Whitney U test or Kruskal-Wallis test was used. For dichotomous categorical variables, the χ² test was used to assess statistical significance. The level of statistical significance was set at P<.05 for all tests. This study was approved by the Baylor College of Medicine Institutional Review Board, and the Veterans Affairs Research and Development Committee (study number 97B02.H).

RESULTS

We surveyed 80 patients in this study, and all were men. The mean age was 65.3 years (range, 44-86 years). Race distribution was as follows: white, 71%; black, 21%; and Hispanic, 8%. Patients had been treated with either total laryngectomy (n=17), XRT (n=24), or both (n=39); no patient had been treated with chemotherapy alone, and none had undergone partial laryngectomy. The mean time since treatment was 4.9 years (range, 6 months to 22 years). Original tumor sites and staging are listed in

Table 1. Six patients (7.5%) were treated with adjuvant chemotherapy, and only 1 patient (1.3%) had metastatic disease at the time of survey completion.

The mean VHI subscale scores for groups of patients with different social and demographic characteristics are shown in Table 2. We found that patients who lived alone had lower functional voice handicap (P=.01), but they demonstrated no significant differences in emotional or physical voice handicap. Lack of employment or lower level of education did not predict significantly worse voice handicap scores. Younger patients tended to report higher levels of voice handicap, but the differences were not statistically significant. We assessed the relationship between length of time since treatment and VHI subscale scores, and the correlations were very low (r=0.01 to 0.11, P=.36 to .98), indicating that more recent treatment did not predict better or worse outcome. Similarly, we assessed the relationship between pretreatment prediction of eventual voice outcome made by speech pathologists and VHI subscale scores, and the correlation coefficients were very low (r=-0.07 to 0.21, P=.22 to .69), indicating that the speech pathologist’s pretreatment estimation of outcome was a poor predictor of actual self-rated voice handicap after treatment.
Self-rated global health did not correlate significantly with emotional (P = .69), functional (P = .99), or physical (P = .45) voice handicap scores on the VHI. However, some subscale scores on the SF-36 correlated with voice handicap scores (Table 3). Correlation coefficients were all negative because the SF-36 and VHI use different reference values (high scores on the VHI indicate poor status, and high scores on the SF-36 indicate good status). Overall, mean subscale scores on the SF-36 in our population were significantly lower than age- and sex-matched population norms for healthy subjects. In fact, mean subscale scores approximated those of patients with congestive heart failure. Many of the patients in our series had other illnesses, so their SF-36 scores may reflect the effects of comorbid medical conditions in addition to the effects of treatment for laryngeal cancer.

We divided patients into 2 groups: those who had a laryngectomy regardless of the method of voice rehabilitation (n = 56), and those who had not had a laryngectomy—all of whom had undergone external beam XRT (n = 24). The comparison between VHI subscale scores is shown in (Table 4); emotional and functional voice handicap scores were significantly higher in patients who underwent laryngectomy. We also compared the original tumor stage of patients treated with and without laryngectomy; as expected, patients treated without laryngectomy had smaller tumors (median stage, I) than laryngectomy patients (median stage, III), so an analysis of outcomes by tumor stage would be confounded by the different treatments received. In fact, 16 (89%) of 18 patients with stage I disease received XRT only, and 33 (89%) of 37 patients with stage III or IV disease underwent a laryngectomy, so when attempting to control for tumor stage, there were too few cases left to perform an analysis of treatment effects on VHI scores. Considering only the subgroup of patients who had undergone laryngectomy (n = 56), there were no significant differences in VHI subscale scores between those who had undergone adjuvant XRT (n = 39) and those who had not (n = 17). Global health status subscale scores on the SF-36 did not differ significantly between patients treated with laryngectomy (n = 56) or XRT alone (n = 24); however, median self-rated overall health was higher in patients who had undergone laryngectomy than in patients treated with XRT (P = .005).

Among patients who had undergone laryngectomy, we evaluated the subset who had undergone a TEP...
for alaryngeal speech (n=23) and compared them with patients treated with XRT only (n=24). A comparison between VHI subscale scores in patients receiving TEP and XRT is shown in Table 5. Physical voice handicap scores did not differ significantly, but emotional and functional handicap scores were lower in patients receiving XRT. The Figure shows the distribution of VHI subscale scores in patients receiving TEP and XRT. Of note, there was considerable overlap in voice handicap scores, with many laryngectomy patients who used a TEP demonstrating lower levels of voice handicap than patients with an intact larynx who were treated with XRT. Global health status subscale scores from the SF-36 were also not significantly different between those undergoing TEP and XRT, although self-rated global health was higher in patients receiving TEP than in those receiving XRT (P=.001).

**COMMENT**

One of the first studies to address QOL issues in patients with laryngeal cancer was the so-called fireman study by McNeil et al.14 In this study, 37 healthy volunteers (12 firefighters and 25 executives) were asked to choose treatment options based on perceived changes in QOL after treatment with a laryngectomy. Subjects were also asked questions to determine how averse they were to risks. The time-tradeoff technique of utility analysis was then used to assess the subjects’ preference for survival vs voice preservation. In their group of subjects, the authors calculated that the utility of 10 years of survival with normal speech was equivalent to 7 years of survival with artificial speech. Furthermore, the authors estimated that 20% of subjects would choose a treatment to preserve their larynx, even if it meant potentially shorter survival, probably because of the perceived decrement in QOL after laryngectomy.

This landmark article14 was the first to address QOL issues faced by patients who underwent laryngectomy. However, there are several shortcomings of this study that should be considered. First, although subjects were played tapes of patients using esophageal speech, subjects were not educated about voice rehabilitation using a TEP or an electrolarynx. Second, subjects were asked to assume that post-XRT voices would be normal, although clinical and experimental experience suggests that many patients experience a decrement in voice quality and overall QOL after XRT.15,16 Third, the study was flawed by the use of healthy subjects. The expected value, or utility, of a given health state may be considered differently by healthy subjects and subjects with disease, or by patients before and after a disease is diagnosed.17,18 For instance, the change in QOL caused by hemodialysis treatment may be perceived differently by a healthy subject than by a patient with renal failure for whom dialysis represents a treatment to prolong life. In addition, the way questions are phrased has a significant impact on the decisions made by healthy volunteers. For instance, if identical probabilities are proposed to subjects but asked in 2 different ways (chance of survival and chance of death), subjects usually make different decisions even though the probabilities are identical.17 Therefore, the decisions on health status and QOL made by healthy volunteers should...

---

**Table 5. Subscale Scores on the Voice Handicap Index (VHI) in Patients Who Underwent Laryngectomy and Use a Tracheoesophageal Puncture (TEP) for Alaryngeal Speech and Patients Who Underwent External Beam Radiation Therapy (XRT) Only**

<table>
<thead>
<tr>
<th>VHI</th>
<th>TEP (n=23)</th>
<th>XRT (n=24)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional</td>
<td>9.6±6.7</td>
<td>5.8±7.1</td>
<td>.07</td>
</tr>
<tr>
<td>Functional</td>
<td>14.3±7.4</td>
<td>8.2±6.3</td>
<td>.01</td>
</tr>
<tr>
<td>Physical</td>
<td>13.3±7.5</td>
<td>11.0±9.1</td>
<td>.37</td>
</tr>
</tbody>
</table>

*Values are mean (±SD).
not be assumed to coincide with the values and decisions of patients with laryngeal cancer, who usually have decreased functional status and QOL at the time of decision making.

In addition, there is evidence that the patient’s perception of change in overall QOL after treatment differs from the physician’s perspective of change in overall QOL. Mohide et al. performed a retrospective study comparing the relative importance of various QOL dimensions as ranked by patients after undergoing laryngectomy and health care professionals. The authors found that the relative importance of dimensions differed significantly between the 2 groups. Health care professionals ranked communication impairment and self-image and self-esteem as the most important dimensions, but patients identified communication as only the third most important dimension and self-image and self-esteem as the seventh most important. After undergoing laryngectomy patients ranked physical consequences and interference with social activities as the 2 most important QOL dimensions. Therefore, physicians were inaccurate in assessing the most important dimensions of QOL in a group of patients who underwent laryngectomy.

Schuller et al. used a nonvalidated questionnaire to assess overall satisfaction with treatment and post-treatment voice in a group of 75 patients with laryngeal cancer who had been treated with combinations of surgery and XRT. The patients’ relatives were also surveyed concerning the patients’ posttreatment voice. The authors found that 88% of patients expressed satisfaction with their posttreatment voice, and 96% of patients’ relatives also expressed satisfaction with the posttreatment voice. The major limitations of this study were the retrospective nature, and that the authors did not use validated outcome instruments.

List et al. prospectively studied performance status and QOL using several validated instruments, which measure functional and performance status in patients with head and neck cancer (the PSS-HN, the Karnofsky performance status scale, and the FACT-HN), in a group of patients treated for laryngeal cancer. The authors found that performance status scores showed significant improvement after treatment. The QOL scores showed a trend toward improvement after treatment, but statistical significance was not achieved. The authors found that both ability to speak and ability to eat were not correlated with overall QOL.

Currently, there are several potential techniques for voice rehabilitation after total laryngectomy. Speech without a larynx (alaryngeal speech) is often intelligible and enables many laryngectomy patients to communicate adequately. Options for rehabilitation include esophageal speech, use of an electrolarynx, or a TEP. Esophageal speech or an electrolarynx speech is usually intelligible, and both are successfully used by many patients. Creation of a TEP is another technique for alaryngeal voice rehabilitation. A prosthetic valve is placed in a surgically created shunt to direct air into the esophagus; the most commonly used is the Blom-Singer prosthesis. Acceptable speech is achieved in most patients using a tracheoesophageal prosthesis. In one study, postlaryngectomy patients using a tracheoesophageal prosthesis reported an improvement in speech intelligibility over their preoperative voice.

In our study, we found that demographic factors usually did not predict voice handicap. We expected that patients who lived alone, were unemployed, or were older would have lower levels of perceived voice handicap because those patients have fewer communication demands placed on them and have lower expectations for their voice. We found that most subscale scores on the VHI were not significantly different, although patients living alone had a significantly lower level of functional voice handicap, and older patients showed a trend toward lower levels of voice handicap.

Comparisons of subscale scores on the SF-36 and the VHI demonstrated the construct validity of the VHI. Constructs from the SF-36 that were not represented on the VHI, such as bodily pain and general health, showed no significant correlations with VHI subscale scores. However, other constructs from the SF-36 that were represented on the VHI, such as social functioning and role functioning—emotional and physical—demonstrated significant correlations with VHI subscale scores. In addition, the SF-36 mental health subscale correlated with the VHI emotional subscale, and the SF-36 physical functioning subscale correlated with the VHI physical subscale.

In addition to confirming the construct validity of the VHI, the SF-36 scores demonstrate that there were several aspects of QOL that were not affected by verbal communication handicap. These data confirm the findings of other investigators that voice handicap is not the most important or the only dimension affecting QOL after treatment for laryngeal cancer. In fact, general health and vitality were not correlated with voice handicap. Furthermore, even though SF-36 subscales that demonstrated statistically significant correlations with VHI subscale scores had fairly low levels of actual correlation (about 0.33-0.42). Correlation coefficients less than 0.50 indicate a large amount of unexplained variance. In other words, differences in voice handicap explain only a small portion of the differences in QOL between different patients; much of the difference in QOL is therefore caused by other (unmeasured) factors.

As expected, the group of patients who had undergone XRT demonstrated lower levels of voice handicap than the group of patients who underwent total laryngectomy. In addition, the subset of patients who had undergone laryngectomy were rehabilitated with a TEP also demonstrated higher levels of voice handicap than patients who had received XRT. However, the Figure demonstrates a significant degree of overlap of voice handicap scores between patients who had TEP and those given XRT. Despite the different group mean values, several patients who had undergone laryngectomy using a TEP reported lower levels of functional, emotional, and physical voice handicap than patients with an intact larynx who had undergone XRT. This individual variation in voice outcome after treatment is not surprising to clinicians who regularly treat patients with laryngeal cancer. Many pa-
patients have excellent communication skills after a total laryngectomy, and other patients have a significant amount of voice handicap after XRT. The wide degree of individual variation in voice handicap after treatment also demonstrates that group data should probably not be used to predict individual results to patients making decisions about treatment. Rather, the multifactorial and subjective nature of health status and satisfaction with voice after treatment should probably be stressed during discussion of outcomes with patients.

Global QOL subscale scores were not significantly different between patients who did and did not undergo laryngectomy. Therefore, differences in voice handicap scores between groups are not the result of differences in overall QOL. In addition, the similarity of global QOL scores is interesting because the 2 groups had significant differences in pretreatment tumor stage and underwent significantly different treatments. In this series, patients who had undergone a laryngectomy actually reported better overall health on a single-item scale than patients treated with XRT.

We also noted that pretreatment prediction by experienced speech pathologists was a poor predictor of eventual voice outcome in this study. These findings support the findings of other investigators, that the perception of changes in functional status and QOL by health care professionals is a poor predictor of the changes reported by patients with laryngeal cancer themselves.

These data demonstrate that some aspects of voice handicap are not necessarily worse in patients who have undergone laryngectomy, and they refute the common perception that sparing the larynx during treatment will result in improvements in verbal communication. There are other disease-specific issues that are important when considering outcomes after larynx-sparing treatment protocols, such as swallowing and airway protection, and this study addresses only voice production.

The 2 major shortcomings of this study are its retrospective nature, and the lack of female patients. In this retrospective study, we do not have data on either the pretreatment voice handicap or QOL for any patients, or the pretreatment expectations of voice outcome and QOL. In addition, neither treatment group nor type of voice rehabilitation were assigned in a randomized manner, so the possibility for selection bias exists. Therefore, these data should not be used to compare voice handicap and QOL outcomes after different treatments. Prospective studies are needed to further examine these issues.

In this cross-sectional study of patients who had completed treatment for laryngeal cancer, we found that overall QOL did not differ significantly between patients treated with a laryngectomy and patients treated with XRT. Some aspects of the patients’ QOL showed a moderate correlation with verbal communication handicap, but dimensions other than voice significantly affected global QOL in these patients. Although the group mean scores demonstrated significantly greater verbal communication handicap among patients who underwent laryngec-

tomy than among those who underwent XRT, there was a large amount of individual variation, and group mean data were not a good predictor of an individual patient’s outcome.

Accepted for publication September 12, 1997.


Reprints: Michael G. Stewart, MD, MPH, Bobby R. Alford Department of Otorhinolaryngology and Communication Sciences, Baylor College of Medicine, Houston, TX 77030 (e-mail: mgstew@bcm.tmc.edu).

REFERENCES