Multimodal Evaluation of Voice Outcome in Early Glottic Cancers Treated With Definitive Radiotherapy

MARIANNA TRIGNANI1, ANGELO DI PILLA1, CONSUELO ROSA1,2, MARZIA BORGL1, DAVID FASCIOL1, LUCREZIA GASPARINI1, FIORELLA DI GUGLIELMO1, ALBINA ALLAJBEJ1, MARTA DI FRANCESCO3, GIANNICLA FALCONE1, FRANCESCA VITULLO4, ADELCHI CROCE1, DOMENICO GENOVESI1,2 and LUCIANA CARAVATTA1

1Department of Radiation Oncology, SS. Annunziata Hospital, “G. D’Annunzio” University of Chieti, Chieti, Italy; 2Department of Neuroscience, Imaging and Clinical Sciences, “G. D’Annunzio” University of Chieti, Chieti, Italy; 3Speech Rehabilitation and Phoniatrics, “Sant’Agostino” Centre, Fondazione Papa Paolo VI, Chieti, Italy; 4Department of Otolarinology, SS. Annunziata Hospital, Chieti, Italy

Abstract. Background/Aim: We employed a multimodal evaluation of voice outcome (MEVO) model to assess long-term voice outcome in early glottic cancer (EGC) patients treated with primary radiotherapy (RT). The model consisted of objective and subjective vocal evaluation during follow-up, by a dedicated Speech Pathologist and Speech Therapist. Patients and Methods: MEVO methodology includes Self-perception Voice Handicap Index (VHI-30), evaluation of parameters Grade (G), Roughness (R), Breathiness (B), Asthenia (A) and Strain (S) according to GRBAS scale, objective analysis and aerodynamics using the PRAAT software and laryngeal evaluation with videostroboscope (VS). Results: The MEVO methodology was described and tested on a sample of 10 EGCs submitted to definitive RT (total dose 66-70 Gy). Mean follow-up was 48.9 months (range=9-115). VHI was mild-moderate in 90% of patients; overall voice function (GRBAS) was normal-mildly impaired in 70% of patients; VS evaluation showed normal vocal cord motion in 90% of patients, but complete glottic closure in 60%. PRAAT scores confirmed these findings. Conclusion: A multidimensional voice evaluation is time consuming, but useful to objectify vocal impact of radiotherapy. The MEVO model allowed to quantify vocal dysfunction, showing a good objective vocal outcome.

Laryngeal carcinomas represent the most common head and neck tumors; approximately 75% of cases involve the glottis and 75-80% of these glottic carcinomas are diagnosed at an early stage (1,2). The main treatment modalities in early glottic carcinoma (EGC) are radiotherapy (RT), Laser Surgery (LS) or open partial laryngectomy. These therapeutic options offer comparable oncological and survival outcomes thus suggesting that therapeutic choice could be based on quality-of-life and vocal function outcomes (3-5).

Commonly, voice quality is believed to be better after radiotherapy and worse after surgical laryngeal procedures (5-10). However, long-term radiation-induced effects including fibrosis, chronic edema, laryngeal stenosis and xerostomia, could impact patients’ voice outcomes and quality of life (6). Several prospective and retrospective cohort studies have compared vocal outcomes after LS and RT for EGC, with a wide range and controversial results (5, 7-10). The main reason for this controversy is lack of standardization in post-treatment dysphonia evaluation, with great variability both in terms of assessment methodology of voice function and follow-up duration (3-11).

Most published studies report voice quality data within 2 years from the end of the treatments and to date there is limited experience using videostroboscopy (VS) to assess the impact of RT and LS upon functional voice outcomes (12-13). An appropriate and standardized methodology for qualitative voice assessment is an essential pre-requisite for understanding the effects of different therapeutic modalities on voice function. Therefore, a multidimensional methodology
to evaluate post-treatment voice outcome in EGC patients, was implemented in our hospital by a multidisciplinary team involving Radiation Oncologist, Otolaryngologist, Speech Pathologist and Speech Therapist.

The aim of our study was to describe the methodology of our newly implemented multimodal evaluation voice outcome (MEVO) model and to report voice outcome results of the MEVO model on a sample of EGC following RT.

Patients and Methods

Multidimensional set of voice analysis. A multimodal model to evaluate voice outcome was implemented by a multidisciplinary team involving Radiation Oncologist, Otolaryngologist, Speech Pathologist and Speech Therapist. MEVO model avails of 4 modalities for voice analysis: self-perception using Voice Handicap Index (VHI-30) questionnaire, perceptual dysphonia rating using the GRBAS scale, objective phonetic speech analysis and aerodynamics using PRAAT software performed by speech therapist, and laryngeal and cordal evaluation using VS performed by speech pathologist.

VHI-30. The VHI is a validated self-assessment tool introduced by Jacobson et al. and translated in different languages (14). It is a questionnaire of 30 questions to quantify functional, physical and emotional impacts of voice disability on a patient’s quality of life; 5-point rating scale with a total score for the 30 questions from 0 to 120 with a lower score indicating a less severe patient-reported voice-related handicap. The translated and validated Italian questionnaire was administered by speech therapist or filled by patients themselves.

GRBAS Scale. The GRBAS, developed by the Committee for Phonatory Function Tests of the Japan Society of Logopedics and Phoniatrics, is a scale for perceptual analysis of voice quality performed by speech therapists, consisting of Grade (G), Roughness (R), Breathiness (B), Asthenia (A) and Strain (S) parameters. Each parameter is graded as 0=normal, 1=mild, 2=moderate and 3=severe impairment. Higher score corresponds to a more dysphonic voice. The perceptual evaluation of voice was performed by speech therapist on prolonged a/a: / for as long as possible after maximal inspiration, and at a spontaneous, comfortable pitch and loudness (15).

Objective acoustic analysis and aerodynamics. Acoustic analysis provides objective and non-invasive measures of vocal function. It was performed using PRAAT software, easily accessible and user-friendly software available for free and developed by P. Boersma and D. Weenink from the Phonetics Department of the University of Amsterdam (www.Praat.org). Praat’s voice analysis output consists of a waveform of the analyzed signal, a spectrogram, and a voice report evaluating the following acoustic parameters: fundamental frequency (F0, Hz), Jitter or frequency variation (%), Shimmer or amplitude variation (%) and harmonics to noise ratio (HNR, dB). An omnidirectional microphone with a G&BL condenser (Frequency Response 20 Hz - 16 kHz, Impedance 2.0 KΩ), suitable for all sound cards, positioned at a distance of 20 cm from the lips and with an axial tilt of 45°, preventing disruption related to the airflow. The recording was conducted by the same speech therapist in the very same conditions, in the same room, with a sampling rate of 22 050 Hz with a background noise of less than 50 dB.

Participants were asked to pronounce the vowel “a” with an intensity of voice similar to a normal conversation, with no changes in intensity and frequency, for at least 5 s. The test was repeated three times; then, further recording with Praat and analyzing the three middle seconds of each spectrogram was performed. For each subject, we evaluated the mean of the values obtained in the three recordings.

According to the Società Italiana Foniatria e Logopedia (SIFEL) protocol, we also carried out the maximum phonation time (MPT) evaluation, asking our subject to repeat “a” three times and taking into account only the maximum value obtained from the three tests. MPT was assessed by speech therapist.

F0 is defined as the number of vocal cord vibration per second, expressed in Hz. It varies with sex, age and professional use of voice (16). Typical values of F0 are 120 Hz for men and 210 Hz for women. The standard range recommended by P. Boersma and D. Weenink is from 75 Hz to 500 Hz. As advised, for our analysis, we set two different pitch ranges: one specifically for female voices (100-500 Hz) and the other specifically for male voices (75-300 Hz) (17). Jitter is defined as parameter of frequency variation from cycle to cycle and it is affected mainly by lack of control vibration of the cord. Jitter local, jitter local absolute, jitter rap, jitter ppq5, jitter ddp were the domains evaluated. Threshold value for Jitter Local is <0.356% for females and <0.508% for males, higher percentages are considered pathological. Shimmer relates to the amplitude variation of the sound wave and it changes with reduction of glottal resistance and lesions on the vocal cord. The domains assessed were shimmer local, shimmer local dB, shimmer apq3, shimmer apq5, shimmer apq11 and shimmer ddp (18).

Normal Shimmer Local values are <5.187% for female and <5.248% for male adults. HNR is an assessment of the ratio between periodic components and non-periodic component comprising a segment of voice speech (19). A value of HNR less than 7 dB is considered pathological (20).

Laryngeal function evaluation. Laryngeal videostroboscopy (VS) is the gold standard for evaluation of dysphonia and laryngeal function (21). VS is an endoscopic tool that uses synchronized pulsed light at a frequency allowing the examiner to observe normal and pathologic vocal fold (VF) vibration and movement during phonation. Four stroboscopic parameters were assessed: VF mobility, VF vibration, mucosal wave (erythema/edema) and glottic closure (21).

Patient and treatment characteristics. Thirty patients with early glottic cancer (Tis-T1-T2N0M0 based on TNM staging) treated with primary RT between 2008 and 2017 were selected for evaluation. All patients were asked to participate to the study and only 10 accepted to be subjected to voice analysis according to our MEVO model. The MEVO model was applied in collaboration with speech pathologist and speech therapist. Speech samples were recorded and analyzed. All patients had biopsy-proven glottis tumor. The decision between radiation and surgery was mainly by patient first choice, after a combined evaluation by both radiation oncologist and otolaryngologist. Patients with previous surgery were excluded. Patient characteristics including sex, age at diagnosis and tumor stage were recorded, as shown in Table I. Radiotherapy prescription adopted in our center was 66 Gy in 33 fractions (2 Gy/fraction) for Tis and T1a cancers and 70 Gy in 35 fractions (2 Gy/fraction) for T1b and T2 cancers, delivered with 3D conformal radiotherapy technique. Follow-up duration was calculated from the date of the end of RT to the last follow-up visit.
Results

We evaluated 30 EGC patients submitted to primary RT between 2008 and 2017. Baseline characteristics of early glottic cancer patients and radiotherapy schedule are reported in Table I.

The overall mean follow-up was 48.88 months (range=9-115 months). OS and DFS were 96.6% and 96.6%; one patient experienced disease failure 24 months after RT. Radiotherapy was generally well tolerated. All patients were weekly evaluated by radiation oncologist during treatment. No acute and late radiation-induced toxicity grade >2 (skin, dysphonia, xerostomia and dysphagia) was observed. No emergency or prophylactic tracheotomy was performed during RT. From the original sample of 30 patients, in the context of ongoing voice quality analysis 10 patients were evaluated.

MEVO model results. Results obtained from the VHI-30 questionnaire in the functional, physical and emotional scales are reported in Table II. The mean VHI was 25.5±21.17. Six (60%) patients were classified as mild, 3 (30%) as moderate and 1 (10%) as severe dysphonia. According to GIRBAS scale, Table III summarizes the distribution of dysphonia severity. The overall voice function tends to be normal in 30% of patients; voice dysfunction was mild in 40%,
preservation, oncologic and survival rates. RT is generally optimal therapeutic approach with comparable laryngeal analysis including VS evaluation was 35 minutes.

The mean time required for performing MEVO amplitude vibration in 50%, normal mucosal wave in 80% and complete glottic closure in 50% of them. For 1 patient of overall grade, roughness, breathiness and asthenia showed moderate in 20% and severe in 10% of patients. Mean scores of overall grade, roughness, breathiness and asthenia showed a mild profile whereas means of strain tend to be normal.

Objective scores generated by the PRAAT software also confirmed the subjective findings. Average values of acoustic parameters are summarized in Table IV and the characteristics of each patient who underwent voice analysis evaluation are shown in Table V. One patient presented diplonia. The patients presented normal-mild impairment of acoustic parameters. Videostroboscopy evaluation showed normal vocal cord motion in 80% of patients, normal amplitude vibration in 50%, normal mucosal wave in 80% and complete glottic closure in 50% of them. For 1 patient laryngeal examination was not possible due to vagovagal reflex. The mean time required for performing MEVO analysis including VS evaluation was 35 minutes.

Discussion

The goal of treatment of EGC is tumor eradication with vocal function preservation. Both radiotherapy and endolaryngeal laser resection are well established as the optimal therapeutic approach with comparable laryngeal preservation, oncologic and survival rates. RT is generally well tolerated, with a mild acute toxicity profile and a low (<5%) long-term complication rate (22). Previous studies have demonstrated that RT negatively affects voice quality (23-27); however, these studies have examined limited voice outcome measures and well-structured multimodal voice analyses are lacking in literature.

There are numerous prospective and retrospective cohort studies investigating LS and RT for EGC using heterogeneous outcome measures, with a wide range and controversial results regarding voice quality (perceptual and acoustic). The only randomized study conducted by Aaltonen et al. compared both treatments evaluating voice quality using GRBAS scale and VS. It showed similar overall voice quality after the treatments, with less breathy voice, better glottic closure and less inconvenience in daily lives due to voice quality for the patients of RT group (9).

Results of meta-analysis conducted by Greulich et al. suggest no clinically significant difference between subjective voice outcomes (VHI-10 and VHI-30 scores) following RT versus surgery for treatment of T1 glottic carcinoma, with a trend toward slightly better scores in the RT group (8, 28-31).

Considering that the preservation of adequate phonation is a noteworthy issue in EGC patients, the VHI proves to be a very important tool in obtaining information regarding
patients treated with exclusive radiotherapy. Patients were retrospectively analyzed and compared to a group of similar patients treated with CO₂ laser cordectomy. Subjective and objective tools such as the Voice Handicap Index-10 (VHI-10) and the Multidimensional Voice Program (MDVP™) software were employed, showing a lower deterioration rate after RT but without statistically significant difference, even when analyzing different domains (35).

In our study, voice outcome in ECG submitted to primary RT was studied by application of multimodal voice evaluation in order to obtain a comprehensive vocal function analysis. Patients were submitted to the MEVO analysis, involving 4 modalities for voice analysis: self-perception (VHI-30), dysphonia rating (GRBAS scale), objective phonetic speech analysis and aerodynamics (PRAAT software) performed by speech therapist, and laryngeal and cordal evaluation (VS performed) by speech pathologist. The VHI results were in agreement with the GRBAS scores with normal overall voice function in 30% of patients; voice dysfunction was mild in 40%, moderate in 20% and severe in 10% of patients.

Regarding objective acoustic analysis, EGC patients of our sample presented normal-mild impairment of acoustic parameters. VS findings showed normal vocal cord motion in 80% of patients, normal amplitude vibration in 50%, normal mucosal wave in 80% and complete glottic closure in 50% of them.

A previous study that assessed perceptual voice quality after radiotherapy reported normal or near-normal voices after treatment in two-thirds of patients, 2 of 5 patients (1 T1b and 1 T2) presented a fundamental frequency (F₀, Hz) lower than normal, and frequency ranges for both T1b tumors were smaller than “norm” values (33). In patients with bilateral vocal fold involvement and T2 tumor staging, one or more perturbation measures were abnormally high, whereas all perturbation measures were normal in patients with T1a tumors. No stroboscopy and aerodynamics results were reported.

To date, no standard, validated voice measurement protocol exists. However, in 2001 the ELS proposed a basic multidimensional protocol to evaluate voice, consisting of a number of components: 1) acoustic (jitter, shimmer, fundamental frequency range, and softest intensity); 2) aerodynamics (phonation quotient); 3) perceptual analysis (grade, roughness, and breathiness); 4) videostroboscopy (closure, regularity of vibration, mucosal valve, and symmetry); and 5) subjective scores (VHI) (34). ELS protocol covers only voice assessment without any concern about radiation induced toxicity. The multidimensional voice assessment protocol adopted in our study was similar to that proposed by ELS.

In our study, we presented the multidisciplinary methodology of a protocol aimed to a qualitative and quantitative voice analysis in EGC patients. The protocol was applied on a sample of EGC patients submitted to primary RT and pre-RT voice assessments were not required. Similarly to our study, Potenza et al. in 2015 reported clinical outcomes and voice quality analysis of 55 EGC patients treated with exclusive radiotherapy. Patients were under voice function impact deriving from RT or LS is lacking; therefore, applying a complex and standardized methodology is required in order to quantify post-treatment changes and to choose the better therapy in terms of both oncological outcome and vocal/life quality. Additional factors such as tobacco smoking, age-related changes, gastroesophageal reflux or occupational voice usage should also be considered because they contribute to deteriorate phonation (36).

In conclusion, our study showed a voice function mildly impaired in EGC patients after primary radiotherapy. To obtain an optimal evaluation of the vocal function a comprehensive methodology which includes both subjective and objective parameters, such as with the MEVO applied is our study, is desirable. These analyses can require a long-time (>30 minutes) and a certain patient compliance, however in our experience the methodology applied was feasible and well accepted by patients. Therefore, multimodal voice evaluations should be performed and comparative studies aimed to evaluate voice function following both RT and laser microsurgery using a multidisciplinary and multimodal methodology, similarly to our reported protocol, should be promoted.

Conflicts of Interest

The Authors report no conflicts of interest.

Authors’ Contributions

MT, ADP and AA contributed to the concept and design of the work. AA, CR, MB, DF, LG, FDG, MDF, GF, FV and AC provided data and performed main data analysis and provided pictures elaboration. AA, MT, LC drafted the manuscripts. All author revised the article and approved the version to be published. Each author has sufficiently participated in the work to take public responsibility for appropriate portions of the content.

References


