Functional outcomes after radiotherapy or laser surgery in early glottic carcinoma: A systematic review

Loon, Y. van; Sjogren, E.V.; Langeveld, T.P.M.; Jong, R.J.B. de; Schoones, J.W.; Rossum, M.A. van

Functional outcomes after radiotherapy or laser surgery in early glottic carcinoma: A systematic review

Yda van Loon, MD, MSc,[1] Elisabeth V. Sjögren, MD, PhD,[1] Ton P. M. Langeveld, MD, PhD,[1,3] Rob J. Baatenburg de Jong, MD, PhD,[3] Jan W. Schoones, MA,[2] Maya A. van Rossum, PhD[1]

[1]Department of ENT–Head and Neck Surgery, Leiden University Medical Center, Leiden, The Netherlands, [2]Walaeus Library, Leiden University Medical Center, Leiden, The Netherlands, [3]Department of Otorhinolaryngology–Head and Neck Surgery, Erasmus Medical Center, Rotterdam, The Netherlands.

Accepted 16 February 2011
Published online 30 September 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/hed.21783

ABSTRACT: *Background.* Early glottic carcinoma is treated with laser surgery or radiotherapy, but which treatment has better functional outcomes is unclear. This systematic review compared functional outcomes (voice, swallowing, quality of life [QOL]) in more extended T1a and limited T2 tumors (1) between treatments and (2) between greater and lesser laser resections.

Methods. A systematic literature search covered relevant databases from 1990 to 2009, combining all patient/problem, intervention, comparison, outcome (PICO) keyword variations.

Results. A total of 19 papers met the inclusion criteria, all of which were level IV evidence. Papers reported only voice and QOL. Heterogeneity of outcome measures prevented data pooling. Uncertainty about tumor comparability (depth, extent) between the 2 treatments, small subject numbers, and poor-quality reporting hindered interpretation.

Conclusions. To allow comparison of laser surgery versus radiotherapy, a standardized method is needed that accurately measures tumor extent and depth. Agreement on functional outcome measures is necessary to allow comparison of treatments and resection types. Multicenter studies should be encouraged to guarantee adequate subject numbers. *Head Neck* **34:** 1179–1189, 2012

KEY WORDS: **laryngeal cancer, radiotherapy, laser surgery, functional outcomes**

Radiotherapy and CO_2 endoscopic laser surgery (henceforth laser surgery) are the main treatment modalities for early glottic carcinoma. Key concerns remain disease control and posttreatment functional outcomes. Since the introduction of laser surgery in 1972,[1] oncologic and some, mostly voice-related, functional outcomes have been compared with radiotherapy (RT). For superficial midcord T1a glottic tumors, sufficient evidence exists to assume comparable local control rates, and the literature further agrees that voice outcome seems equal for both treatment modalities.[2–5] The voice is either normal or mildly dysphonic, but with very little perceived handicap by patients (as measured with the validated Voice Handicap Index [VHI]).[5–10] Therefore, assuming comparable disease and voice outcome, laser surgery is the treatment of choice for superficial T1a midcord lesions in the Dutch Guideline for Laryngeal Carcinoma 2008.[11] RT is still advocated for more extended T1 and for T2 lesions.

However, there is a trend to also perform laser surgery for more extended T1a, bilateral T1, and T2 glottic tumors, especially in countries such as Italy, France, and Germany, but also South American and Asian countries.[7,12–20] There is therefore no international consensus on preferred treatment, and, to our knowledge, no randomized trial exists to aid decision making. A Cochrane review concluded: ''There is currently insufficient evidence to guide management decisions on the most effective treatment.''[21] A recent meta-analysis that compared both therapies for early-stage glottic cancer showed no significant differences for local control and laryngectomy-free survival between the 2 modalities, but favored laser surgery for overall survival.[22] Currently, treatment choice primarily depends on the inclination of the clinician: progressive surgeons most probably prefer laser; cautious surgeons and radiotherapists may prefer RT. A recent survey among Dutch head and neck surgeons showed that concerns about functional outcomes after more extended resections still constituted an issue for many treating physicians.

To our knowledge, 1 meta-analysis exists on post-treatment voice quality comparing RT and laser surgery.[22] This meta-analysis showed no objective differences between the 2 modalities; however, these results are not informative enough because the included papers pooled different disease stages, from Tis to T2. To fill this gap, the aim of the present systematic review was to provide evidence-based insight into functional outcomes, such as voice-related aspects, swallowing, and weight, after treatment for more extended early glottic tumors.

***Corresponding author:** Y. van Loon, Department of ENT, Head and Neck Surgery, Leiden University Medical Center, Leiden, The Netherlands. E-mail: y.van_loon@lumc.nl

This article was awarded a grant by ZOLEON, Stichting Oncologie Holland West, Leiden, The Netherlands.

After treatment, RT may cause dryness and fibrosis that may affect vocal fold vibration and result in poorer voice. Dry mucosa, sticky mucus, and fibrosis may also affect swallowing.[8] On the other hand, laser surgery may affect voice because of scarring, which affects vibration, and loss of tissue and/or vocal muscle, leading to a glottal gap, the size of which is related to the extent of the resection.[6] However, there is no reason to expect that laser surgery would influence deglutition (causing dysphagia, intake, or weight loss).

We attempted to test the following hypotheses: (1) laser surgery for extended early glottic carcinoma yields poorer voice (outcomes), but RT results in more swallowing problems; (2) greater laser resections result in poorer voice (outcomes) than lesser laser resections.

MATERIALS AND METHODS

Search strategy

In cooperation with a trained librarian, a search strategy was composed. The following databases were searched: PubMed (1949 to August 2009), EMBASE (OVID-version, 1980 to August 2009), Web of Science (1945 to August 2009), Cochrane Library (1990 to August 2009), PsycINFO (EbscoHost-version, 1887 to August 2009), Academic Search Premier (EbscoHost-version, 1865 to August 2009), and CINAHL (EbscoHost-version, 1982 to August 2009). The search strategy consisted of the AND combination of 3 main concepts: early glottic cancer, cordectomy/laser therapy/radiation therapy, and functional outcome. For these 3 concepts, all relevant keyword variations were used, not only keyword variations in the controlled vocabularies of the various databases, but the free text word variations of these concepts as well. In general, the search consisted of the combination of the following terms: (1) early glottic carcinoma, laryngeal cancer; (2) cordectomy, chordectomy, laser surgery, laser therapy, radiotherapy; and (3) functional outcome, voice, speech, acoustic, talking, communication, intelligibility, spectrography, speech perception, hoarseness, breathiness, GRBAS (= grade, roughness, breathiness, asthenia, strain scale [voice disorder assessment]), jitter, shimmer, aerodynamic measurement, video stroboscopy, swallowing, swallowing disorders, swallowing dysfunction, dysphagia, weight change, diet, nutrition, intake, quality of life (QOL), pain.

This search strategy was optimized for all consulted databases, taking into account the differences of the various controlled vocabularies as well as the differences of database-specific technical variations (such as the use of quotation marks). Results were limited to articles written in English, French, German, and Dutch.

Inclusion criteria. This review included studies published from January 1990 to August 2009. Studies had to be research-based and published in a peer-reviewed journal. Studies had to report on functional outcomes after radiotherapy or laser surgery for early glottic carcinoma. Title and abstract were screened for the following inclusion criteria: (1) T1–T2 *glottic* or early *glottic* carcinoma (or tumor); if T classification or location (subglottic, glottic, supraglottic) was not explicitly stated in the abstract, the paper was screened for this information; (2) intervention investigated in a study had to be laser surgery, or radiotherapy, or both (but only 1 modality per patient); (3) one or more functional outcomes had to be mentioned.

If the title or abstract did not provide sufficient information on inclusion, the paper was screened and included if the criteria were met.

Exclusion criteria. Excluded were: (1) studies reporting on laryngeal cancer in general without specifying the location of the tumor; (2) studies reporting only on oncologic outcomes, or chemoradiation; (3) descriptive reviews, meeting or conference abstracts, short surveys, or letters to the editor.

Selection process

The search yielded 3813 abstracts using the specified keywords described earlier; 2425 abstracts remained after eliminating duplicates (PubMed 1145, EMBASE 1334 with 565 duplicates, Web of Science 919 with 618 duplicates, Cochrane 44 with 35 duplicates, PsycINFO 35 with 34 duplicates, Academic Search Premier 224 with 69 duplicates, and CINAHL 112 with 67 duplicates).

After screening the titles and abstracts, 67 studies remained. The references of the selected papers were handsearched for relevant citations. Relevant citations were obtained and checked in accord with the same inclusion criteria used for the main search. The flowchart in Figure 1 illustrates the selection process.

Evaluation procedure. Two observers (Y.V.L., M.V.R.) evaluated the papers independently. If criteria had been evaluated differently, this was resolved through discussion.

Relevance criteria

Table 1 gives specific criteria relevant for this review. Disease characteristics, intervention, and functional outcomes determined selection of relevant papers. A crucial criterion was whether we could extract relevant functional outcome data for extended T1 and/or T2 tumors or, in the case of laser surgery, for greater resections (resections extending into the vocal muscle, ie, equivalent to European Laryngological Society (ELS) type III and further). Because we were interested in the functional outcomes of early glottic tumors larger than T1a superficial midcord, papers that combined results for T classification (eg, T1–T4); resection types (eg, ELS I–V); supraglottic, subglottic, and glottic tumors; or only investigated Tis or T1a superficial midcord tumors, were excluded. In a paper comparing results before and after reconstruction,[1] the results after laser surgery but before reconstruction were included. One author presented similar data in separate German and English papers; we included the English version.[23,24] In total, 19 papers answered the relevance criteria and were included in the review.[6,7,9,16–20,23,25–34] The references of the articles not included (the rest of the 67 that were initially selected after screening title and abstract) are available on request. An overview of the included studies with the number of patients per tumor stage and per treatment is presented in Table 3.

Voice measurement is complex and largely subjective. To date, no standard, validated voice measurement protocol exists. However, in 2002 the ELS proposed a basic multidimensional protocol to evaluate voice, consisting of a number of components.[35] The voice data in this review are presented broadly in accord with the ELS protocol and

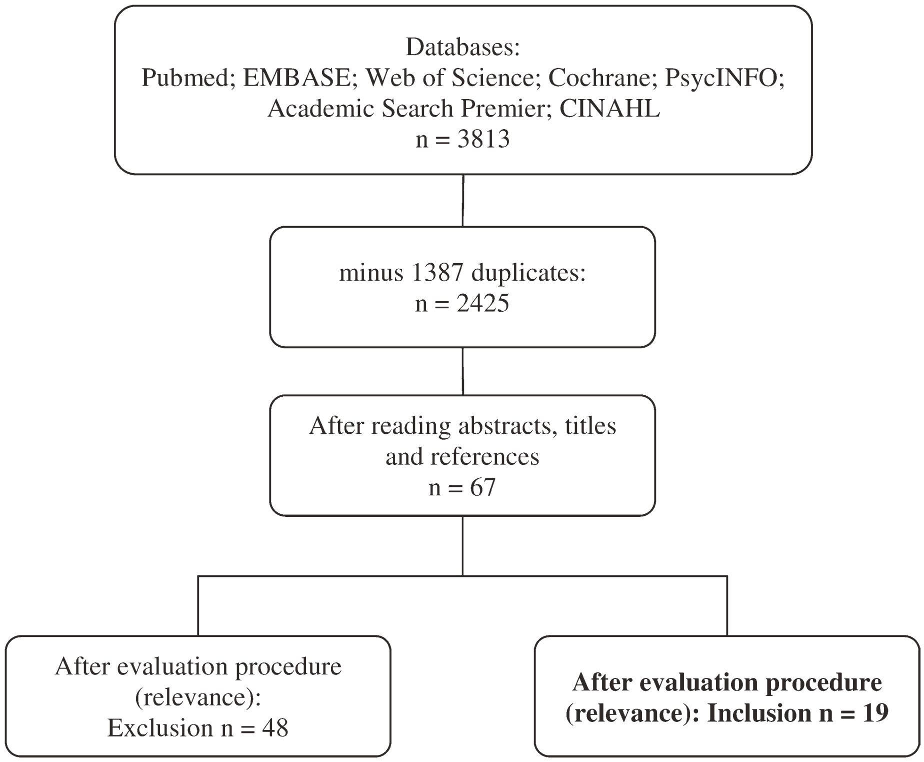

FIGURE 1. Schematic overview of the selection process for studies on functional outcomes.

in the same order as Table 1: voice quality (auditory perception and acoustic analysis), voice function (aerodynamics, stroboscopy, voice profile), and voice performance (voice-specific questionnaires). Because terms such as ''function'' or ''quality'' were used interchangeably, we reclassified some papers' results: outcomes that might have been called ''voice function'' are now described under ''voice quality'' or vice versa.

Quality criteria

Table 2 gives the criteria we used to evaluate the quality of the papers. For all the different sections of a paper, the main concerns were clarity and completeness of the information.

The quality of the assessment tools used to measure the functional outcomes was also evaluated. For validated instruments, appropriate validation studies had to be mentioned. Well-established, recommended, internationally accepted, but not validated assessment tools were judged as standardized (eg, perturbation measures such as jitter and shimmer, which are, for example, recommended by the ELS, or auditory-perceptual rating systems such as the GRBAS[35] or Consensus Auditory Perceptual Evaluation–Voice (Cape-V).[36] If a reference was provided, but the assessment tool was not validated, or established/accepted internationally, it was evaluated as referenced only. If assessment tools were designed for the study in question, without a proper validation procedure and no references to other validating publications, it was judged as created for study. Descriptive refers to 1 or 2 sentences in the Results section without explicit mention of an assessment tool in the Methods section (eg, laryngologists reporting that *x* number of patients had poor voice quality).

Level of evidence was graded using definitions that Wasserman et al[37] adapted for otolaryngology. These definitions are based on standards advocated by the Oxford Centre of Evidence-based Medicine,[38] and largely depend on study design. The levels range from 1 (highest level of evidence) to 5 (lowest level of evidence).

RESULTS

Factors that prohibited the systematic evaluation and pooling of data

The following factors complicated analysis and interpretation of data.

Uncertainty about Tumors Investigated. With regard to studies investigating both treatment groups (radiotherapy and laser), it has to be noted that none of these papers could guarantee that tumor extent/depth was comparable between the groups. This hindered fair comparison between the 2 treatments.

Heterogeneity in (reporting of) outcome measures. Different papers reported on different outcomes; none reported on deglutition, and thus the present review consists of voice and QOL outcomes only.

Furthermore, most papers reported only 1 or 2 components, and the tests used to measure the components differed greatly, which prevented comparison among studies. For instance, rating scales used to evaluate perceptual voice quality were incompatible. Similarly, papers

TABLE 1. Relevance criteria used to evaluate papers.

Disease characteristics	❒ glottic ❒ supraglottic	❒ T1a N0M0 ❒ T1b N0M0 ❒ T1 N0M0	❒ T2a N0M0 ❒ T2 N0M0 ❒ T1–T2 N0M0
Number of patients	n =		
Patient characteristics	Sex Age Smoking Alcohol	❒ M ❒ F … median yr / … mean yr ❒ Yes ❒ No ❒ Yes ❒ No	
Intervention/Comparison	❒ Laser ❒ RT ❒ Laser vs RT		
Outcomes	Voice quality	❒ perceptual/auditory →	❒ GRBAS ❒ arts rating ❒ CAPE-V ❒ VAS
		❒ acoustic →	❒ jitter ❒ subjective ❒ shimmer ❒ other, …. ❒ harmonics/noise ratio ❒ spectrographic rating
	Voice function	❒ aerodynamics →	❒ Maximum Phonation Time ❒ Phonation Quotient ❒ subglottal pressure ❒ airflow
		❒ stroboscopy	
		❒ voice profile / phonetography → ❒ intensity ❒ F0	❒ F0 range ❒ loudness range ❒ dB mean ❒ maximum intensity ❒ F0 mean ❒ maximum F0
	Voice performance	❒ VHI ❒ Vocal Performance Questionnaire ❒ VoiSS ❒ VRQOL	
	Swallowing	❒ FEES ❒ videofluoroscopy ❒ questionnaire	❒ subjective rating patient ❒ subjective rating clinician
	Weight changes	❒ BMI ❒ % of change in weight	
	Eating / diet / intake	❒ questions from EORTC	
	QoL	❒ EORTC QLQ-C30 ❒ EORTC QLQ-H&N35 ❒ COOP/ WONCA ❒ SF-36	❒ HAD ❒ UW-QoL ❒ FACT

evaluated the acoustic signal using a variety of parameters and measurement tools. Although all used 1 or more traditional perturbation measures such as jitter or shimmer, these were generally not defined properly.

Also, no single standard aerodynamic measure exists in the literature. Measures such as maximum phonation time (MPT), phonation quotient (vital capacity/MPT = PQ), and airway resistance were combined or used interchangeably to evaluate glottal (in)sufficiency.

Not all studies that mentioned stroboscopy used it as an outcome measure, and not all studies reported on all their stroboscopy findings, since the examinations were sometimes inadequate for evaluation.[6,28]

Some studies would report values (eg, Hertz [Hz]), in terms of the mean or the median or the range, whereas other papers would report number or percentages of subjects that presented with abnormal values (although different papers referred to different norms).

Combination and classification of resection types.

Papers that combined ELS resection types did not always combine the same resections: some combined ELS type II and type III, others II–IV, and so forth. Roh et al[8] divided the ELS resections into 3 groups, of which 2 groups were considered greater resections: ELS type III–IV and ELS type Va plus bilateral type II, including anterior commissure. One author specified the location of the tumor on the vocal cord but not the depth of the excision.[9] Second, in studies from before 2000, authors used their own classification system.[6,18,23,28,34] The resections of McGuirt et al[6,28] were defined as lesser and greater

TABLE 2. Quality criteria used to evaluate papers.

Research question	❐clear, focused, (providing testable hypothesis)	
Study population	❐ in/exclusion criteria explicit and detailed ❐ implicit	
Study design	❐ randomized → ❐ systematic review → ❐ cohort → ❐ case series →	LoE* 1❐ LoE 1❐ ❐ prospective ❐ LoE 2; ❐retrospective ❐LoE 3 ❐ rationale for matching criteria & method, confounding factors accounted for ❐ baseline comparability provided ❐ LoE 4 ❐ pre-post ❐ post ❐ reconstruction (pre-post)
Outcomes	❐clearly defined primary and secondary ❐ quality of assessment →	 ❐ validated ❐ standardized ❐ referenced ❐ created for study ❐ descriptive
Replication potential	❐ methods adequately described ❐ reliability described	 ❐ inter-rater reliability ❐ intra-rater reliability
Statistical methods	❐clear description	❐ separate heading ❐ separate paragraph ❐ described: sentence in text
Results	❐ tables/graphs self-explanatory→ ❐CI, P, R-coefficient reported ❐ results match methods ❐drop-outs/missing accounted → Follow-up	❐ all data ❐ part of data ❐ only text Drop-outs: *n* = Missing data rep: ❐ Y ❐ N ❐ median : ….. yr ❐ mean: ….. yr ❐ Post measurement ≤ 3 mo ❐ Post measurement > 3 mo
Conclusions	❐ supported by results ❐ study limitations reported and taken into account	

Abbreviation: LoE, level of evidence.

resections. However, the lesser resections reached the superficial part of the vocalis muscle and consisted of ELS I, II, and superficial type III resections, whereas the greater resections reached one third to one half of the depth of the vocalis muscle; thus they were classified as type III for this review. We could not include the lesser resections because of the heterogeneity of this group.

Consequence. As a result of the factors mentioned earlier, a direct comparison between the papers' outcomes was not feasible. Only overall results and differences and trends between certain (sub)groups given within papers were therefore included in the present review, and the only way in which we could present the results was as a descriptive narrative.

General findings

An overview of the 19 papers included in this review is provided in Table 3, discussing the T classification and number of patients per resection group.

Because of the heterogeneity of the used tools and presented outcome measures of these papers, unfortunately it proved impossible to create a clear table covering the outcomes. Nevertheless, the general findings of the 19 papers are discussed in the following text.

Relevance Criteria. Thirteen papers investigated laser surgery,[7,16–20,23,26–29,33,34] 5 papers compared laser surgery with radiotherapy,[6,9,30–32] and 1 paper investigated radiotherapy.[25]

Staging. Nine studies reported on Tis data,[7,16–20,23,26,31] and all studies reported on T1 and 7 studies on T2 tumors.[16,20,23,25,29,31,34] Outcome data were separately reported per T classification in 8 studies,[6,9,25,27,28,30,32] whereas 10 studies combined different T classifications.[7,16–20,23,29,31,33,34]

Twelve of 19 papers reported outcomes in accord with ELS resection type.[35] Six reported outcomes per ELS resection; 6 combined ELS resection types.

Resection types. In this review, cordectomies were grouped into lesser or greater resections to simplify discussion.

Lesser resections were defined as subepithelial or subligamental. Greater resections included at least part of the vocalis muscle as well as more extended cordectomies

TABLE 3. Overview of studies included in the review.

Author (year)	Total no. of patients in paper	No. of patients per Tis/T1/T2	No. of patients in review	No. of patients per lesser/greater resection group
Laser				
Zeitels (2002)	32	0/28/4	32	23/9‡
McGuirt (1992)	22	0/12/0	22	10/12
Xu (2007)	140	ns	60	30/60
Sittel (1998)	80	3/38/39	72	34/38†
Remacle (1997)	74	14/26/0	ns	39/32
Vilaseca (2008)	42	2/40/0	42	17/25§
Peretti (2003, p 174)	101	15/86/0	69§	18/51
Peretti (2003, p 759)	89	8/63/18	51*	41/48
Kennedy (2007)	53	0/19/34	19	6/13
Roh (2007)	85	16/69/0	75*	46/39
Ledda (2006)	141	20/106/15	133*	44/89
Crevier-Buchman (2007)	12	ns	12	ns
Motta (2008)	140	0/83/52	135	0/135

Author (year)	Laser/RT	No. of patients per Tis/T1/T2	Laser/RT	No. of patients per lesser/greater resection group
Laser and RT				
Peeters (2004)	56/46	0/56/0	52/40	52/0
McGuirt (1994)	11/13	0/24/0	11/13	0/11
Rydell (1995)	18/18	0/36/0	18/18	18/0‡
Policarpo (2004)	63/20	20/38/2	20/20	0/20¶
Tamura (2003)	14/6	0/20/0	14/6	0/14
RT				
Benninger (1994)	63	0/48/15	5	na

Abbreviations: na, not applicable; ns, not specified; RT, radiotherapy; ELS, European Laryngological Society.
* T classification not reported for patients included in this review.
† Resection adapted to ELS classification.
‡ Resection extent is not specified.
§ Outcomes for lesser (ELS I and II) and greater resections (ELS III) were combined.
¶ Three laser patients previously had RT.

encompassing surrounding structures, such as the ventricular fold or the contralateral vocal cord.

Voice outcomes. The majority of the studies evaluated voice quality: 12 using perceptual[6,9,16,17,19,23,25–27,33,34] and 15 using acoustic analysis.[6,7,9,16,17,19,20,25,26,28,29,31–34] Regarding voice function, 13 studies evaluated aerodynamics,[6,7,9,13,16,18–20,28,29,32–34] 7 studies stroboscopy,[6,16,19,20,28,31,34] 6 intensity,[6,18,28,29,32,34] and 12 studies fundamental frequency (F0).[6,9,18–20,25,26,28,31–34] Four studies investigated voice performance, using the VHI.[7,19,26,30]

Quality of life. Quality of life was evaluated in 2 studies using either the European Organization for Research and Treatment of Cancer–Head and Neck (EORTC H&N35) questionnaire[19] or the Dartmouth Coop Functional Health Assessment Chart/World Organization of Family Doctors (COOP/Wonca) questionnaire.[30]

Deglutition. None of the studies evaluated swallowing, diet, or weight.

Quality Criteria. Only 2 papers clearly formulated research questions such that readers could deduct the hypothesis that the study investigated.[19,33] See Table 4.

Inclusion criteria were explicitly described in 5 papers.[6,7,16,28,30] Three papers comparing the 2 treatments did not adequately describe or define inclusion criteria. One study did not define tumors adequately in the radiotherapy group; we therefore used only the results of the laser surgery group.[31] In the 2 other studies, more superficial lesions received laser surgery, thus inducing selection bias,[9,30] and we included only radiotherapy results.

All included papers were assigned to level 4 evidence. The majority of the articles, 13 of 19, were case series[9,16–18,20,23,25,27–32]; 6 were cohort studies, but of poor quality, and therefore had to be assigned to a lower evidence level.[6,9,26,30–32] These 6 cohort studies lacked baseline comparisons of the treatment groups and information on how confounding factors were controlled for (either in the design or the analysis).

The replication potential of a study depends on the method. Methods were adequately described in 12 studies[6,7,9,17,19,20,23,26,27,29,30,32]; 1 study referred to a study protocol published in another paper.[34]

Primary and secondary outcomes were reasonably clearly defined in 4 papers.[19,26,27,33]

Four papers[17,19,26,30] relied on validated instruments: VHI, EORTC, and COOP/Wonca.

Reliability was mentioned in only 1 study.[9] Statistical methods were clearly stated (separate heading or paragraph) in 7 papers.[9,16,19,20,25,30,33] One study stated that the numbers prohibited statistical analysis.[26]

Statistical results were adequately described in 4 studies. The *p* values were given in these studies, but no

TABLE 4. General findings on relevance criteria for 19 studies included in the final analysis.

Review criteria	Results (*x/n*)	Review criteria	Results (*x/n*)
Site			
Glottic	19/19	Voice quality	
		Perception	12/19
		Acoustic	15/19
Treatment		Voice function	
Laser	13/19	Aerodynamics	13/19
RT	1/19	Stroboscopy	7/19
Laser vs RT	5/19	Intensity (mean)	6/19
		F0 (mean)	11/19
Total number of patients	1276	Voice performance	
n used in analysis	934	VHI	4/19
		Other	0/19
T classification		Swallowing	0/19
Tis	6/19	Weight changes	0/19
T1	18/19*	Eating / diet / intake	0/19
T2	7/19		
ELS resection type		Quality of life	
I	6/19	EORTC H&N35	1/19
II	9/19	COOP/WONCA	1/19
III	16/19	Other	0/19
IV	7/19		
V	8/19		

Abbreviations: RT, radiotherapy; F0, fundamental frequency; VHI, Voice Handicap Index; ELS, European Laryngological Society; EORTC H&N35, European Organization for Research and Treatment of Cancer. Head and Neck Module; COOP/Wonca, Dartmouth Coop Functional Health Assessment Chart/World Organization of Family Doctors.
* One study did not report T classification for the patients (Crevier-Buchman, 2007).

other scores (confidence interval [CI], Z, F, or R value) were taken into account.[9,20,29,32] Most of the studies, 14 of 19, presented their results in accord with the methods described.[7,9,16,19,20,25–32,34]

Reports on drop-outs or missing data were given in 8 of 19 studies.[7,16,17,25,27,28,32,34] All studies stated their follow-up time, mostly giving the mean; in 3 studies the exact moments of measurement for all patients could be extracted.[9,26,29] Only 1 study took biases and limitations adequately into account when drawing conclusions.[6] See Table 5.

Functional Outcomes

Voice quality

Auditory perception. Different rating systems are available to judge the voice quality; some used the GRBAS,[35] others the Oates Russel[39] or Sittel rating scale,[7,23,24] and some created their own rating scales. Some used more than 1 of the above-cited rating systems.

Laser surgery. Nine papers looked at perceptual voice quality.[7,16,17,19,23,26,27,33,34] One study found significantly poorer voice quality after greater resections than after lesser resections.[19] Five papers reported poorer voice quality after greater resections than after lesser resections, although this difference was not tested statistically.[7,16,17,23,34] Another study found that voice quality correlated with resection type, but no correlation coefficients were provided.[33] Another paper also showed poorer voice quality with increasing resection type, although outcomes for ELS types II–III were combined.[26] Similarly, 1 study showed that the number of subjects with moderate-to-severe dysphonia increased with greater resections.[27]

Laser versus radiotherapy. Two studies compared perceptual voice quality.[6,9] No difference was found in the study by McGuirt[6], whereas in the study by Rydell[9] voices were rated as poorer after laser than after radiotherapy.

Radiotherapy. One study reported perceptual voice quality after radiotherapy, based on reviews of clinic notes.[25] Two-thirds had normal or near-normal voices after treatment, as judged by both the surgeon and the patient.

Acoustic analysis. Some studies also included evaluation of the spectrogram, but because this is a subjective visual rating this result was not included.[28,31]

Laser surgery. Ten papers evaluated the acoustic signal[7,16,17,19,20,26,28,29,33,34] using 1 or more perturbation measures. In 2 studies, perturbation was significantly worse in greater resection types when compared with controls, whereas lesser resection types did not differ significantly from controls.[7,16] In 2 studies lesser resections also differed significantly from the control groups, but the difference between the greater resections and the controls was much larger.[17,20] In another study, approximately half the patients with lesser resections showed abnormally high perturbation measures, whereas almost all patients that had received a greater resection (equivalent to ELS III) had abnormally high perturbation measures.[34] One study reported significant differences among 3 groups (the first group equivalent to ELS I and II, the second to ELS III and IV, the third to ELS V), but no post hoc analysis was reported, so it is unclear which group differed from which; the general trend, however, was the more extensive the resection, the higher the perturbation.[19] One study stated that the greater the resection, the higher the number of parameters that differed from the control group, but presented the combined results for ELS types I–III (and ELS V separately), so that their statement could not be verified.[33] In another study, the groups were combined according to type of voicing (eg, glottal, inferior mixed, etc.).[29] Type III and most of the type IV resections that were grouped as ''glottal voicing'' showed significantly better harmonic-to-noise ratio than the type V resections, but all resection types (III–V) also had significantly poorer perturbation measures when compared with the normal control group.

Another study reported that the acoustic parameters for ELS type V could not be given since the F0 could not be detected. Oddly, perturbation tended to be higher after type I than that after II–III resections in this study, but the difference was not tested statistically.[26] Finally, 1 study reported results for 2 groups, 1 group consisting of resections including less than a third of the vocalis muscle, the other group more than a third; this study also found more abnormal values in the lesser resection group than in the greater resection group.[28]

Laser versus radiotherapy. Four papers compared acoustic measures between laser and radiotherapy.[6,9,31,32] McGuirt[6] found poorer perturbation measures after radiotherapy than after laser. Tamura[32] found no difference in perturbation measures between type III resections and radiotherapy. Rydell et al[9] mentioned that the tumors in the radiotherapy group were probably more extensive; no mention

TABLE 5. General findings on quality criteria for 19 studies included in the final analysis.

Review criteria	Results (x/n)
Research question	
Clear, focused	2/19
Study population	
Explicit inclusion criteria	5/19
Level of evidence	
IV (poor cohorts, case series without internal control group)	19/19
Study design	
Randomized	0/19
Systematic review/meta analysis	0/19
Cohort	6/19
Rationale for matching criteria and baseline comparability given	0/5
Case series	13/19
Pre-post	2/13
Post	11/13
Outcomes	
Clearly defined outcomes	3/19
Validated assessment tools*	4/19
Replication potential	
Reliability described	1/19
Methods adequately described	13/19
Statistics	
Clear description	7/19
Results	
Values reported (CI, *p*, R-coefficient)	4/19
Results match methods	14/19
Drop-outs / missing data reported	8/19
Follow-up described	19/19
Moment of measurement reported	2/19
Conclusions	
Biases and limitations taken into consideration	1/19

Abbreviation: CI, confidence interval.
* The quality of assessment is given in the text, separately per outcome.

is made of the type of resection, but the laser group had significantly poorer perturbation measures compared with radiotherapy. Policarpo[31] also found poorer perturbation measures for type III and type IV resections when compared with radiotherapy.

Radiotherapy. One study reported tumor extent and perturbation measures for 5 individual patients.[25] Two of these patients had bilateral vocal fold involvement and 1 patient had T2 tumor staging. In these 3 patients 1 or more perturbation measures were abnormally high, whereas all perturbation measures were normal in the other 2 patients with T1a tumors.

Voice Function

Aerodynamics

Laser surgery. Ten papers reported aerodynamic results.[7,16,18–20,26,28,29,33,34] In 4 studies no significant differences were found between greater and lesser resections for any of the aerodynamic measures,[18–20,28] although in 3 studies both greater and lesser resections showed significantly poorer results when compared with normals.[18,20,28] Similarly, in another study the number of subjects with abnormal MPT was comparable for the different resection groups.[34] In contrast, in 1 study the greater resection types (III, IV, V) showed a significantly poorer PQ when compared with normals, whereas the lesser resection types (I, II) did not.[7] A poorer MPT was found for extended resections (V) compared with normal controls, but lesser resections also including type III showed no significant difference for MPT with normal controls.[33]

In another study, the greater the resection, the poorer the MPT; unfortunately, differences between lesser and greater resections were not tested statistically.[16] The MPT improved after surgery for ELS types I and II–III but deteriorated after ELS type V resection.[26] One study revealed a significantly better MPT in a glottal voicing group (ELS type III and majority of type IV resections) versus other voicing groups (ELS type V resections).[29]

Laser versus radiotherapy. Three studies reported on aerodynamics.[6,9,32] One study found no difference in MPT between laser and radiotherapy, but laser was significantly poorer when compared with normals.[32] McGuirt found that laryngeal airway resistance after radiotherapy was 4.5-fold the norm, whereas for laser it was 3-fold the norm.[6,9] Another study looked at duration of reading time and number of breaths (as indirect measure of voicing efficiency) and found significantly poorer results for laser.[9]

Radiotherapy. No papers reported aerodynamics.

F0 and intensity: average and ranges. Twelve studies reported the average fundamental frequency (F0).[6,9,16,19,20,25,26,28,31–34] Some papers investigated different vowels, but only the vowel /a/ was considered in the present review. Given that average F0 is gender-dependent, it was surprising that only 1 of 5 papers[16,19,25,31,34] presented results separately.[34] Seven papers reported data for men only.[6,9,20,26,28,32,33]

Average intensity was measured in 5 studies.[6,18,28,29,32]

Laser surgery. Six studies compared lesser and greater resections.[16,18–20,29,33] One study found significantly higher F0 for greater resections.[19] Three studies found significantly higher F0 for all the resection groups compared with normal controls, with the greater resections also showing a trend toward a higher average F0 than the lesser resections.[16,20,33] One study showed the same trend, although this was not statistically tested (McGuirt). Zeitels[34] reported 6/18 lesser resections and 1/8 greater resections with abnormally high F0.

F0 range was measured in 2 papers, but was presented in Hz instead of semitones, which prevents meaningful comparisons.[33,34]

Average intensity was determined in 3 studies.[18,28,29] In Motta et al,[29] who combined the groups according to type of vocal compensation (eg, glottal, inferior mixed, etc.), type III and most of the type IV resections that were grouped as ''glottal voicing'' showed significantly higher intensity than that of type IV and type V (nonglottal voicing). Furthermore, all resection types (III–V) also had significantly lower intensity compared with that of the normal control group. Both Remacle and McGuirt reported comparable values for lesser and greater resections, although this was not tested statistically. One study reported a postoperative rise in F0 for type I and a ''normalization'' of the F0 for types II–III, but could not detect F0 for ELS type V resections.[26]

Laser versus radiotherapy. Four studies reported F0.[6,9,31,32] Two studies reported lower F0 after radiotherapy,[6,9] of which 1[9] reported that this difference was significant. Results between groups were similar in the other studies,[31,32] but both groups had elevated F0 (compared with normals).

Three studies mentioned average intensity,[6,32,34] of which 1 reported values[6]; intensity was similar between the treatment groups, but lower when compared with ''normal'' values.

Radiotherapy. One study reported F0.[25] Two of 5 patients (1 T1b and 1 T2) had lower F0 than normal, and frequency ranges for both T1b tumors were smaller than ''norm'' values.

Videostroboscopy. The results for structural abnormalities (eg, scarring, granuloma, stenosis, anterior commissure web, or synechia) are reported separately from functional deficits (eg, glottal gap, vocal fold immobility, mucosal wave, asymmetry, supraglottic involvement).

Laser surgery. Structural abnormalities: 3 studies reported on structural abnormalities,[19,20,31] but 1 study did not report separately for different resection types.[31] Scar tissue, granulomas, stenosis, anterior commissure web, and synechia were more common after greater resections than after lesser resections,[19,20] but tested statistically in 1 study.[19]

Functional abnormalities: 5 studies compared lesser and greater resections[16,19,20,28,29] and found a higher number of functional abnormalities such as altered or absent mucosal wave, incomplete glottic closure, vocal fold immobility, and supraglottic hyperfunction. These alterations increased with the extent of the resections. One study reported a significant higher incidence of glottal gaps, scarring, and mucosal waves for greater resections.[19]

Motta et al[29] classified stroboscopic findings according to type of voicing. All type III and the majority of type IV resections resulted in glottal voicing, whereas type V or greater resulted in nonglottal voicing.[29]

Radiotherapy versus laser. Two studies looked at stroboscopic outcomes,[6,31] but no statistics were reported.

Structural abnormalities: Policarpo et al[31] mentioned structural abnormalities only for greater laser resections. A scar neocord was present in 30% to 100% and a fixed homolateral arytenoid at the place of intervention was seen in 15% of the greater resections. The other study did not report structural abnormalities.[6]

Functional abnormalities: McGuirt et al[6] reported on 5/13 irradiated and 11/11 lasered patients. Glottic closure was not affected after radiotherapy, but was affected in 20% of patients after laser surgery. In both treatment groups mucosal wave was equally diminished and muscle tension dysphonia was equally present. Policarpo et al[31] did not describe mucosal wave or glottic closure for the laser surgery group, but after radiotherapy ($n = 20$) mucosal wave was decreased in 81%, and 25% showed severe glottic insufficiency.

Radiotherapy. No stroboscopy results were reported.

Voice Performance. Studies used the VHI to evaluate voice impact. The VHI is a 30-item validated questionnaire measuring impact of voice problems on daily life; the higher the score, the greater the perceived voice handicap.

Laser surgery. Three studies reported VHI scores. In 1 study the VHI was significantly higher for the greater resections than that for the lesser resections.[19] Another study reported improvement for type I and type II–III combined, but deterioration for type V.[26] In the third study, ''low VHI-score categories'' contained more patients with a lesser than patients with a greater resection, but differences among the VHI-score categories were not tested statistically.[17]

Radiotherapy versus laser surgery. Only 1 study evaluated voice handicap. The mean VHI score of 18 for a group of 40 irradiated patients was significantly higher than the mean VHI of 12 for 52 laser-treated patients. However, deeper invading tumors were treated with radiotherapy.[30]

Quality of life. Two studies that reported on quality of life used validated instruments, EORTC H&N35 and COOP/Wonca.[19,30]

Laser surgery. One study reported that speech and social contact items were significantly less affected after lesser resections than after greater resections, using the EORTC H&N35. Compared with pretreatment, posttreatment QOL scores improved significantly after lesser resections but not after greater resections.[19]

Radiotherapy versus laser surgery. No statistical differences were found between laser surgery and radiotherapy using the COOP/Wonca questionnaire.[30] However, more invasive tumors were irradiated.

Deglutition. None of the studies evaluated swallowing outcomes, weight, or deglutition.

DISCUSSION

In this systematic review, we investigated 2 hypotheses: (1) in extended T1 and limited T2 glottic tumors, laser surgery results in poorer voice quality, function, and performance, but better swallowing function than those in radiotherapy; (2) greater laser resections result in poorer voice quality, function, and performance than those in lesser laser resections.

We could not find conclusive evidence to support our hypotheses, because we could report results only as a descriptive narrative. These descriptive findings provided us with very few insights when comparing radiotherapy and laser surgery. With regard to voice quality (perceptual and acoustic), studies seemed to contradict each other. With respect to voice function, aerodynamic and stroboscopy data also appear inconclusive,[6,9,28,31,32] but the average fundamental frequency seems higher after laser surgery. One study that investigated voice performance (voice handicap) and quality of life was affected by selection bias.[30] It is noteworthy that the perceived voice handicap after radiotherapy in 1 study[30] was comparable to the perceived voice handicap of greater laser resections (ELS types III–IV) in another study.[19]

Concerning our second hypothesis, our descriptive findings seem to indicate that greater resections show more abnormalities on stroboscopy and poorer voice quality than those in lesser resections. Only 2 studies on voice quality reported statistical differences: 1 on perceptual data[19] and 1 study on acoustics.[29] Regarding voice function, the results for aerodynamics are inconclusive. Voice intensity tends to decrease as the extent of the resection increases,[18,28,29] and the average fundamental frequency tends to be higher for more extended resections, although this difference was statistically tested in only 1 study.[19] Regarding stroboscopic evaluation, there was a greater incidence of structural abnormalities (scar tissue, granuloma, anterior commissure web) and functional problems (incomplete closure, abnormal mucosal wave, vocal fold immobility) in greater resections than in lesser resections. As for voice performance, speech and social contact items of the EORTC H&N35 were significantly worse for the greater resections compared with lesser resections.[19]

A comparison with earlier literature reviews is not viable. A recent meta-analysis showed a trend toward better objective measures of voice outcome for radiotherapy. However, the included studies pooled different tumor classifications (Tis–T2) and the authors already mentioned the limitation of the small sample size.[22]

Of 2 other reviews that considered T1(a) glottic cancer 1 paper did not report sufficient information on the outcomes measures.[40] Although the other review, which included a meta-analysis, showed an overall lower voice handicap after laser surgery than that after radiotherapy, lesser and greater resection types were combined, so that interpretation of results for the purpose of this review was not possible.[2]

A number of limitations affected this systematic review, some preventing pooling of data or quantitative reporting of results.

First, uncertainty about tumor comparability: in studies that compared the treatments, the chance that smaller and more superficial lesions were selected for laser surgery and more extended lesions subjected to radiotherapy, leading to treatment bias, was mentioned in at least 2 papers.[9,30] To counteract this problem, new techniques are needed that more accurately define extent and depth of glottic tumor, so that treatment groups are comparable.

A second factor was heterogeneity of the outcome measures, such as incompatible rating scales to evaluate perceptual voice quality, as well as different acoustic, aerodynamic, and stroboscopic measures. A plausible explanation might be that very few validated or internationally standardized voice assessment instruments exist. The only validated instruments, quality of life questionnaires and the VHI, were used in a minority of the studies. Until more reliable assessment tools have been developed and validated, the protocol proposed in 2001 by the European Laryngological Society[35] might be adhered to, because this guarantees some uniformity, allowing a meta-analysis of pooled data. The ELS protocol covers only voice assessment, whereas it has been reported that some subjects suffer from dry mouth and sticky mucus after narrow-field irradiation for early glottic carcinoma.[8] Possible effects of radiotherapy on deglutition have been largely ignored in the literature, perhaps because side effects are deemed negligible. However, a systematic study of all likely short- and long-term side effects is necessary to allow a fair comparison between radiotherapy and laser surgery.

Third, papers on laser surgery did not uniformly categorize resection types. We divided the data into 2 quite general groups: lesser and greater resections, where the latter included all resections cutting through the vocalis muscle and further. Exactly which resection types might result in unacceptable voice quality or function is uncertain. Furthermore, the definition of the term ''unacceptable voice'' may vary considerably between patients but also between surgeons. Future studies that systematically compare the different resection types on all aspects of voice are essential.

There were also other limitations that affected the quality of this systematic review. Small sample sizes affected the statistical power of the studies. Study designs consisted of case studies or poor quality cohort studies. Such limitations influenced the evidence level of included studies, all of which were assigned to level 4. This level, however, is in accord with otolaryngology articles generally: of 1924 clinical research articles, 63% was assigned to level 4, with papers on therapy being assigned to the lowest levels.[37]

Cohort studies included in this review were of poor quality (and therefore assigned a lower evidence level), because no baseline comparability or criteria for matching were reported, and possible confounding factors were not taken into account. Larger cohort studies are feasible if institutions favoring laser surgery and institutions favoring radiotherapy standardize their assessment protocol and pool their data. Although tumor matching is a challenge, the option of a standardized method describing and measuring tumor depth and extent during diagnostic panendoscopy should at least be explored.

With respect to the quality of reporting, there is ample room for improvement. Basic features such as a clear, focused research question, well-defined primary and secondary outcomes, mention of biases, and limitations were neglected. Proper reporting of p values, confidence intervals, or correlation coefficients would also allow the reader to better judge the results and compare treatment effects. We believe authors could improve the quality and structure of study reports if they used accessible (online) tools created specifically for this purpose, such as STROBE (STrengthening the Reporting of OBservational studies in Epidemiology), or checklists provided by the Centre for Evidence-Based Medicine.[41]

In conclusion, we will be able to test our hypotheses only when we have adequately addressed a number of problems. First, a standardized method is needed that accurately measures tumor extent and depth, allowing us to create matching treatment groups and thus fair comparison of radiotherapy versus laser surgery. Second, developing uniform measurement of functional outcomes is another prerequisite to compare treatments and resection types. Finally, multicenter studies should be encouraged to guarantee adequate subject numbers.

REFERENCES

1. Strong MS, Jako GJ. Laser surgery in the larynx. Early clinical experience with continuous CO_2 laser. *Ann Otol Rhinol Laryngol* 1972;81:791–798.
2. Cohen SM, Garrett CG, Dupont WD, Ossoff RH, Courey MS. Voice-related quality of life in T1 glottic cancer: irradiation versus endoscopic excision. *Ann Otol Rhinol Laryngol* 2006;115: 581–586.

3. Loughran S, Calder N, MacGregor FB, Carding P, MacKenzie K. Quality of life and voice following endoscopic resection or radiotherapy for early glottic cancer. *Clin Otolaryngol* 2005;30:42–47.
4. Sjogren EV, Langeveld TPM, Baatenburg de Jong RJ, et al. Clinical outcome of T1 glottic carcinoma since the introduction of endoscopic CO_2 laser surgery as treatment option. *Head Neck* 2008;30:1167–1174.
5. Wedman J, Heimdal JH, Elstad I, Olofsson J. Voice results in patients with T1a glottic cancer treated by radiotherapy or endoscopic measures. *Eur Arch Otorhinolaryngol* 2002;259:547–550.
6. McGuirt WF, Blalock D, Koufman JA, et al. Comparative voice results after laser resection or irradiation of T1 vocal cord carcinoma. *Arch Otolaryngol Head Neck Surg* 1994;120:951–955.
7. Peretti G, Piazza C, Balzanelli C, Cantarella G, Nicolai P. Vocal outcome after endoscopic cordectomies for Tis and T1 glottic carcinomas. *Ann Otol Rhinol Laryngol* 2003;112:174–179.
8. Roh JL, Kim AY, Cho MJ. Xerostomia following radiotherapy of the head and neck affects vocal function. *J Clin Oncol* 2005; 23:3016–3023.
9. Rydell R, Schalen L, Fex S, Elner A. Voice evaluation before and after laser excision vs. radiotherapy of T1A glottic carcinoma. *Acta Otolaryngol* 1995;115:560–565.
10. Sjogren EV, van Rossum MA, Langeveld TPM, et al. Voice outcome in T1a midcord glottic carcinoma: laser surgery vs radiotherapy. *Arch Otolaryngol Head Neck Surg* 2008;134:965–972.
11. Nederlandse Werkgroep Hoofd Halstumoren (NWHHT). Richtlijn Larynxcarcinoom Kwaliteitsintistuut voor de Gezondheidszorg CBO, Version 3.0. Utrecht, The Netherlands: Vereniging van Intergrale Kankercentra;2010.
12. Ansarin M, Santoro L, Cattaneo A, et al. Laser surgery for early glottic cancer: impact of margin status on local control and organ preservation. *Arch Otolaryngol Head Neck Surg* 2009;135: 385–390.
13. Gallo A, de Vincentiis M, Manciocco V, Simonelli M, Fiorella ML, Shah JP. CO_2 laser cordectomy for early-stage glottic carcinoma: a long-term follow-up of 156 cases. *Laryngoscope* 2002; 112:370–374.
14. Haddad L, Abrahao M, Cervantes O, et al. Vocal assessment in patients submitted to CO_2 laser cordectomy. *Braz J Otorhinolaryngol* 2006;72: 295–301.
15. Hartl DM, de Mones E, Hans S, Janot F, Brasnu D. Treatment of early-stage glottic cancer by transoral laser resection. *Ann Otol Rhinol Laryngol* 2007;116:832–836.
16. Ledda GP, Grover N, Pundir V, Masala E, Puxeddu R. Functional outcomes after CO_2 laser treatment of early glottic carcinoma. *Laryngoscope* 2006;116:1007–1011.
17. Peretti G, Piazza C, Balzanelli C, Mensi MC, Rossini M, Antonelli AR. Preoperative and postoperative voice in Tis-T1 glottic cancer treated by endoscopic cordectomy: an additional issue for patient counseling. *Ann Otol Rhinol Laryngol* 2003;112:759–763.
18. Remacle M, Lawson G, Jamart J, Minet M, Watelet JB, Delos M. CO_2 laser in the diagnosis and treatment of early cancer of the vocal fold. *Eur Arch Otorhinolaryngol* 1997;254:169–176.
19. Roh JL, Kim DH, Kim SY, Park CI. Quality of life and voice in patients after laser cordectomy for Tis and T1 glottic carcinomas. *Head Neck* 2007; 29:1010–1016.
20. Xu W, Han D, Hou L, Zhang L, Yu Z, Huang Z. Voice function following CO_2 laser microsurgery for precancerous and early-stage glottic carcinoma. *Acta Otolaryngol* 2007;127:637–641.
21. Dey P, Arnold D, Wight R, MacKenzie K, Kelly C, Wilson J. Radiotherapy versus open surgery versus endolaryngeal surgery (with or without laser) for early laryngeal squamous cell cancer. Cochrane Database Syst Rev 2002;2:CD002027.
22. Higgins KM, Shah MD, Ogaick MJ, Enepekides D. Treatment of early-stage glottic cancer: meta-analysis comparison of laser excision versus radiotherapy. *J Otolaryngol Head Neck Surg* 2009;38:603–612.
23. Sittel C, Eckel HE, Eschenburg C. Phonatory results after laser surgery for glottic carcinoma. *Otolaryngol Head Neck Surg* 1998;119:418–424.
24. Sittel C, Eckel HE, Eschenburg C, Vossing M, Pototschnig C, Zorowka P. Voice quality after partial laser laryngectomy. *Laryngorhinootologie* 1998;77:219–225.
25. Benninger MS, Gillen J, Thieme P, Jacobson B, Dragovich J. Factors associated with recurrence and voice quality following radiation therapy for T1 and T2 glottic carcinomas. *Laryngoscope* 1994;104:294–298.
26. Crevier-Buchman L, Colomb I, Hans S, Vaissiere J, Brasnu D. Evolution of voice assessment and quality of life after endoscopic laser cordectomy. *Rev Laryngol Otol Rhinol (Bord)* 2007;128: 315–320.
27. Kennedy JT, Paddle PM, Cook BJ, Chapman P, Iseli TA. Voice outcomes following transoral laser microsurgery for early glottic squamous cell carcinoma. *J Laryngol Otol* 2007;121:1184–1188.
28. McGuirt WF, Blalock D, Koufman JA, Feehs RS. Voice analysis of patients with endoscopically treated early laryngeal carcinoma. *Ann Otol Rhinol Laryngol* 1992;101:142–146.
29. Motta S, Cesari U, Mesolella M, Motta G. Functional vocal results after CO_2 laser endoscopic surgery for glottic tumours. *J Laryngol Otol* 2008; 122:948–951.
30. Peeters AJ, van Gogh CD, Goor KM, Verdonck-de LI, Langendijk JA, Mahieu HF. Health status and voice outcome after treatment for T1a glottic carcinoma. *Eur Arch Otorhinolaryngol* 2004;261:534–540.
31. Policarpo M, Aluffi P, Brovelli F, Borello G, Pia F. Oncological and functional results of CO_2 laser cordectomy. *Acta Otorhinolaryngol Ital* 2004; 24:267–274.
32. Tamura E, Kitahara S, Ogura M, Kohno N. Voice quality after laser surgery or radiotherapy for T1a glottic carcinoma. *Laryngoscope* 2003;113: 910–914.
33. Vilaseca I, Huerta P, Blanch JL, Fernandez-Planas AM, Jimenez C, Bernal-Sprekelsen M. Voice quality after CO_2 laser cordectomy: what can we really expect? *Head Neck* 2008;30:43–49.
34. Zeitels SM, Hillman RE, Franco RA, Bunting GW. Voice and treatment outcome from phonosurgical management of early glottic cancer. *Ann Otol Rhinol Laryngol Suppl* 2002;190:3–20.
35. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol* 2001;258:77–82.
36. Kempster GB, Gerratt BR, Verdolini AK, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol* 2009;18: 124–132.
37. Wasserman JM, Wynn R, Bash TS, Rosenfeld RM. Levels of evidence in otolaryngology journals. *Otolaryngol Head Neck Surg* 2006;134: 717–723.
38. Phillips B, Ball C, Sackett D, et al. Levels of evidence and grades of recommendation. Oxford Centre for Evidence-based Medicine. URL: http://www.cebm.net/index.aspx?o=1025.
39. Oates J, Russell A. Learning voice analysis using an interactive multimedia package: development and preliminary evaluation. *J Voice* 1998;12: 500–512.
40. Luscher MS, Pedersen U, Johansen LV. Treatment outcome after laser excision of early glottic squamous cell carcinoma: a literature survey. *Acta Oncol* 2001;40:796–800.
41. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4:e296.