LARYNX PRESERVATION CLINICAL TRIAL DESIGN: KEY ISSUES AND RECOMMENDATIONS—A CONSENSUS PANEL SUMMARY

JEAN-LOUIS LEFEBVRE, M.D.,* AND K. KIAN ANG, M.D.,† ON BEHALF OF THE LARYNX PRESERVATION CONSENSUS PANEL

*Département de Cancérologie Cervico-Faciale, Centre Oscar Lambret, Lille, France; and †University of Texas MD Anderson Cancer Center, Houston, TX

Purpose: To develop guidelines for the conduct of Phase III clinical trials of larynx preservation in patients with locally advanced laryngeal and hypopharyngeal cancer.

Methods and Materials: A multidisciplinary international consensus panel developed recommendations after reviewing results from completed Phase III randomized trials, meta-analyses, and published clinical reports with updates available through November 2007. The guidelines were reviewed and approved by the panel.

Results: According to the recommendations, the trial population should include patients with T2 or T3 laryngeal or hypopharyngeal squamous cell carcinoma not considered for partial laryngectomy and exclude those with laryngeal dysfunction or age greater than 70 years. Functional assessments should include speech and swallowing. Voice should be routinely assessed with a simple, validated instrument. The primary endpoint should capture survival and function. The panel created a new endpoint: laryngo-esophageal dysfunction-free survival. Events are death, local relapse, total or partial laryngectomy, tracheotomy at 2 years or later, or feeding tube at 2 years or later. Recommended secondary endpoints are overall survival, progression-free survival, locoregional control, time to tracheotomy, time to laryngectomy, time to discontinuation of feeding tube, and quality of life/patient-reported outcomes. Correlative biomarker studies for near-term trials should include estimated glomerular filtration rate, excision repair cross-complementary-1 gene, E-cadherin and β-catenin, epiregulin and amphiregulin, and TP53 mutation.

Conclusions: Revised trial designs in several key areas are needed to advance the study of larynx preservation. With consistent methodologies, clinical trials can more effectively evaluate and quantify the therapeutic benefit of novel treatment options for patients with locally advanced laryngeal and hypopharyngeal cancer.

Head and neck carcinoma, Larynx preservation, Guidelines, Laryngeal carcinoma, Hypopharyngeal carcinoma.

INTRODUCTION

Until the 1980s, total laryngectomy, performed as initial treatment, was considered the most appropriate therapy for patients with locally advanced laryngeal and hypopharyngeal cancer. Although this strategy can provide disease control, it has a negative impact on patients’ quality of life because of the presence of a permanent tracheostomy and the loss of natural voice. Larynx preservation strategies were thus developed to avoid total laryngectomy and to preserve laryngeal function.

Since the early 1990s, evidence from large randomized trials has shown that larynx preservation strategies that use treatment with induction chemotherapy followed by definitive radiotherapy do not compromise survival when compared with initial total laryngectomy (1, 2). The subsequent randomized Radiation Oncology Therapy Group (RTOG) 91-11 trial demonstrated that when compared with induction chemotherapy with cisplatin-fluorouracil (PF) followed by radiotherapy, initial treatment with concurrent chemoradiotherapy yielded higher larynx preservation and local tumor control rates and similar laryngectomy-free survival (although both significantly better than radiotherapy alone) but did not improve survival (3–5). In presented in the manuscript.

Acknowledgments—The concept for this project originated with select members of the panel. Sanofi-Aventis provided support and funding for organizing the clinician-initiated panel meeting and the subsequent preparation of the manuscript by panel members assisted by a medical writer. Input of Jan Lewin, Ph.D., in methods of swallowing and voice evaluation as presented in the appendix is highly appreciated.

Received Aug 7, 2008, and in revised form Oct 30, 2008. Accepted for publication Oct 30, 2008.

To be published simultaneously in Head & Neck.
fact, with more effective induction chemotherapy regimens (e.g., TPF) (6), the emergence of novel molecular therapeutics (7), and careful patient selection (8), there is a prospect of improving survival with larynx preservation strategies. However, varying trial methodologies have prevented an accurate assessment of the benefits that patients receive through these strategies. The endpoints used to evaluate organ preservation have not, to date, been sufficiently refined or consistently applied across trials. In particular, because of the long-term effects of combined chemoradiotherapy on pharyngolarynx function in swallowing, functional assessments have not been appropriately defined or used. In addition, evidence from a retrospective analysis of RTOG 91-11 suggests salvage laryngectomy after a larynx preservation regimen may adversely affect survival compared with no salvage laryngectomy (9). Therefore, consistent collection of appropriate data is crucial to accurately assess the functional benefits and survival impact of larynx preservation strategies.

With these successes and limitations in mind, four goals for future clinical trials of larynx preservation become apparent: (1) to define the functional unit of a larynx, (2) to determine the best methods for assessment of function when evaluating organ preservation, (3) to accurately assess survival, and (4) using emerging modalities, to improve survival and organ preservation in patients with locally advanced laryngeal or hypopharyngeal cancer. To help achieve these goals, a workshop comprising members of an international consensus panel was convened to review randomized trials and clinical practice for achieving larynx preservation and to make recommendations for clinical trial methodologies that could be applied to future Phase III trials of larynx preservation. The recommended procedures in this document are not intended to be an exhaustive list of factors requiring attention in larynx preservation protocols but, rather, a selected list of recommendations related to specific aspects of these studies that the panel agreed should be addressed to improve the information obtained from future trials.

QUESTIONS

The following questions about larynx preservation trials were addressed by the panel with the goal of reaching consensus among members:

1. Which patients are considered for larynx preservation trials? Once selected, what are the stratification variables of highest importance to obtain the most valuable information from randomized trials?
2. What are the optimal assessments to conduct in patients enrolled in larynx preservation trials to assess the risks and benefits of the study treatment?
3. What are the optimal endpoints to use in larynx preservation trials? How are these endpoints defined?
4. What are the most promising translational research opportunities that should be explored? What clinical trial practices will foster translational research?

METHODS AND MATERIALS

The international consensus panel was composed of medical, radiation, and surgical oncologists with expertise in treating head-and-neck cancer (Appendix A). Seven countries in Europe (France, Germany, Switzerland, Italy, the Netherlands, Denmark, and Belgium) and the United States were represented by members of the panel. Panel members were chosen by identification of all groups that had conducted a randomized larynx preservation trial and appointment of up to 3 experts within their memberships.

Before the workshop, completed Phase III randomized trials (Table 1) (1, 2, 10–15) meta-analyses, and significant reports of clinical practice that evaluated larynx preservation were identified. Published trials and reports were located by searching a computerized online database (PubMed). Search terms included head-and-neck cancer, larynx preservation, laryngeal cancer, hypopharyngeal cancer, and chemotherapy. Abstracts of updated results were accessed through Web sites of oncology associations (American Society of Clinical Oncology, European Cancer Organization). During the workshop, the members reviewed and discussed the most recently updated results from these trials and reports and developed recommendations based on the four questions. The final manuscript was reviewed by all panel members and approved by each of the authors.

DISCUSSION AND RECOMMENDATIONS

Patient selection and stratification

Which patients are suitable for larynx preservation trials? Once selected, what are the stratification variables of highest importance to obtain the most valuable information from randomized trials?

DISCUSSION

Patient selection

Primary site. After the initial landmark trial by the Veterans Affairs Laryngeal Cancer Study Group (VALCSG) that demonstrated the success of an organ preservation approach for patients with squamous cell carcinoma of the larynx (1), the European Organization for the Research and Treatment of Cancer (EORTC) demonstrated the safety of this approach in the population of patients with hypopharyngeal cancer (2, 12). Since these two initial trials, two of the subsequent five randomized larynx preservation trials reviewed by the panel included only patients with laryngeal cancer (3, 13), and three trials included patients with either laryngeal or hypopharyngeal cancer (Table 1) (6, 14, 15).

T stage. Whether patients with T4 disease should be included in an organ preservation trial requires careful consideration, as these patients may experience worse outcomes with this approach. Analysis of data from the VALCSG study showed reduced tumor response to chemotherapy and more frequent salvage laryngectomy in patients with T4 tumors (11). The odds ratio of achieving a response to chemotherapy for patients with T1 to T3 vs. T4 disease was 5.6 (95% CI = 1.5–20.8; \( p = 0.0108 \)). In patients who responded to chemotherapy and received radiation therapy in an organ preservation approach, salvage laryngectomy was required in 56% of patients with T4 tumors compared with 28% of those with T1 to T3 tumors (\( p = 0.001 \)). In clinical practice, patients with T4...
disease, particularly when the tumor extends through the cartilage into neck soft tissue, generally undergo initial surgery and are not considered candidates for a larynx preservation approach. (Notably, tumors exhibiting minimal cartilage invasion remain classified as T3 and are therefore eligible for a conservative approach.)

At the other end of the spectrum, it should be remembered that partial laryngectomy is a form of larynx preservation. The issue then is whether patients with T2 disease who are candidates for partial laryngectomy should be included in larynx preservation trials. All patients eligible for partial laryngectomy were excluded from EORTC 24891 to prevent the need for total laryngectomy in the event of progression during induction chemotherapy (2). Similar criteria were used in Groupe Oncologie Radiotherapie Tete Et Cou (GORTEC) 2000-01 (14) and RTOG 91-11 (3). However, patients presenting with endophytic T2 tumor or with clinical lymphadenopathy (see below) might have extracapsular extension of nodal disease and are at high risk for regional and/or systemic failure. In these cases, based on the results of a recent combined analysis of adjuvant data of the EORTC and RTOG (16), these patients often receive postoperative radiotherapy with concurrent cisplatin. Because the functional outcome of such tri-modality therapy might be worse than combined chemoradiotherapy given as primary therapy, selected patients may be considered for a larynx preservation approach. However, this scenario represents a specific clinical situation, which should be studied separately.

N stage. The nodal status of patients enrolled in completed randomized larynx preservation trials has varied, in particular with respect to N2c and N3 disease, and the issue is complex. At present, data are insufficient to determine the best approach to treat patients with advanced nodal status. Of consideration is the increased risk of distant metastasis in patients with N2 to N3 disease (17). Generally, studies of induction chemotherapy have demonstrated reduced rates of distant relapse compared with the nonchemotherapy arms (1–3) Brockstein et al. demonstrated a reduction in distant relapse in patients who received induction chemotherapy (13%) vs. no chemotherapy (22%; p = .03) before chemoradiotherapy in an analysis of five Phase II trials (17). Furthermore, TAXotere (TAX 324) showed a decrease in distant metastasis in patients receiving TPF vs. PF (5% vs. 9% respectively; p = 0.14) as induction therapy before chemoradiotherapy (6). As such, evaluating chemoradiotherapy, particularly novel regimens, in patients with advanced nodal stage is of interest. The risk of extracapsular spread also increases with higher N stage (18), resulting in an increased likelihood of the ultimate use of adjuvant chemotherapy with radiation in patients who receive initial treatment with surgery. These issues can be further explored if future larynx preservation trials do not exclude patients on the basis of advanced nodal status, on condition that their disease is resectable.

Laryngeal dysfunction. Indicators of baseline laryngeal dysfunction include a tracheotomy, gastric tube, and recent history of pneumonia. Randomized trials have varied with respect to inclusion of patients with baseline tracheotomy. Such patients made up 25% of all patients entered in the VALCSG study and were not excluded from RTOG 91-11 but were excluded from EORTC 24654 (1, 3, 15). Reports of whether the presence of a pretreatment tracheotomy negatively affects recurrence have been conflicting (19, 20). The circumstances and treatment of patients requiring a pretreatment tracheotomy differ between laryngeal and hypopharyngeal cancers and among patients with varying acceptances of total laryngectomy.

In addition to signifying baseline laryngeal dysfunction, recurring pneumonia is an important toxicity to monitor in post-treatment follow-up. Including patients with a recent history of recurring pneumonia requiring hospitalization would confound toxicity data. Furthermore, pneumonia represents a severe morbidity for patients with significant chronic obstructive pulmonary disease (COPD). Patients with COPD requiring hospitalization within the previous 12 months were excluded from TAX 324 (6).

Patient age. Whether patients of advanced age should be excluded is controversial, largely because patients are treated based on factors such as performance status rather than age alone. In a subset analysis of the Meta-Analysis of Chemotherapy to radiotherapy on Head and Neck Cancer (MACH-NC), patients greater than 70 years of age did not benefit from the addition of chemotherapy to radiotherapy (21). However, the reasons are unclear (e.g., was dose intensity similar in older patients vs. younger patients?). In clinical practice, patient selection based on age frequently occurs; therefore, limiting enrollment based on age represents a pragmatic design.

Stratification

Tumor subsite. Although ideally the tumor subsites of supraglottis, glottis, transglottis, epilarynx (the suprahypoid epiglottis, aryepiglottic folds, and arytenoids) and hypopharynx, would be strata, limiting the number of stratification factors for the purposes of maintaining a manageable sample size is essential. Tumors of the glottis, supraglottis, and hypopharynx differ from each other in the radiation portal (because of variable patterns of lymphatic spread), risk of distant metastasis, and natural history. Hypopharyngeal tumors also appear to respond differently to treatment. In the MACH-NC, of three larynx preservation trials (VALCSG, Groupe d’Etude des Tumeurs de la Tete et du Cou (GETTEC), and EORTC 24891), a differential effect of induction chemotherapy was suggested between sites with a potentially beneficial effect on disease-free survival (DFS) noted in hypopharyngeal tumors (HR 0.9) vs. a negative effect in laryngeal tumors (hazard ratio = 1.4), in particular, glottic subsite (21). In contrast, epilarynx tumors are similar to hypopharyngeal tumors with respect to risk of distant metastasis, natural history, and radiosensitivity (22) and therefore can be grouped together. Transglottic tumors do not occur with sufficient frequency to warrant stratification.

Country or region. Stratification by country or region is recommended because of variability in treatment. In particular, radiation technique and selection of patients for partial laryngectomy differ among sites.
<table>
<thead>
<tr>
<th>Study (Ref)</th>
<th>N</th>
<th>Patients</th>
<th>Stratification</th>
<th>Efficacy endpoints</th>
<th>Treatment arms</th>
<th>Overall survival</th>
<th>Larynx preservation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>VA (1, 10, 11)</strong></td>
<td>332</td>
<td>Stage III/IV Laryngeal SCC No T1N1</td>
<td>KPS Stage No/1 vs. N2/3 Glottic vs. supraglottic</td>
<td>Response OS DFS</td>
<td>Surgery + RT</td>
<td>68</td>
<td>2 y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EORTC 24891 (2, 12)</strong></td>
<td>202</td>
<td>Hypopharyngeal SCC T2-T4 N0-N3 (no N2c)</td>
<td>Institution OS (primary) DFS Pyriform sinus vs. aryepiglottic fold</td>
<td>OS DFS FLS*</td>
<td>Surgery + RT</td>
<td>43</td>
<td>3 y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GETTEC (13)</strong></td>
<td>68</td>
<td>Laryngeal SCC T3 N0-N2b</td>
<td>None</td>
<td>OS DFS LP</td>
<td>Surgery + RT</td>
<td>69</td>
<td>2 y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GORTEC 2000-01 14</strong></td>
<td>220</td>
<td>Laryngeal and hypopharyngeal SCC</td>
<td>Not reported</td>
<td>3-year LP rate (primary)</td>
<td>PF</td>
<td>Not reported</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TAX 324 (6)</strong></td>
<td>501</td>
<td>Stage III/IV SCC of head and neck- unresectable or candidates for organ preservation</td>
<td>Tumor site N0/1 vs. N2/3 Institution</td>
<td>OS (primary) PFS Response</td>
<td>PF→CRT (carboplatin)</td>
<td>48</td>
<td>3 y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>RTOG 91-11 (3, 4)</strong></td>
<td>547</td>
<td>Laryngeal SCC Stage III/IV No T1 or large-volume T4</td>
<td>Glottic vs. supraglottic N0/1 vs. N2/3 T2 vs. T3 fixed vs. T3 not fixed vs. T4</td>
<td>LP (primary) OS DFS Local/locoregional control TTDM LFS</td>
<td>CRT (cisplatin)</td>
<td>75</td>
<td>2 y</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EORTC 24954-22950 (15)</strong></td>
<td>450</td>
<td>T3-4 laryngeal SCC</td>
<td>Not reported</td>
<td>FLS* (primary)</td>
<td>Sequential PF→RT</td>
<td>48</td>
<td>5 y</td>
</tr>
</tbody>
</table>

(Continued)
Biomarkers. Lack of assessment for and stratification according to biomarkers has hindered the ability of past trials to provide valuable biologic insight. For example, stratification by estimated glomerular filtration rate (EGFR) status as defined by immunohistochemistry is desirable. However, the panel agreed that presently no biomarker can be identified that would supersede stratification factors as identified above.

Recommendations

- Patients eligible should have T2 or T3 laryngeal (glottic or supraglottic) or hypopharyngeal squamous cell carcinoma not considered for partial laryngectomy.
- Exclusion criteria should include laryngeal dysfunction (defined as pretreatment tracheotomy, tumor-related dysphagia requiring feeding tube, or recurring pneumonia within preceding 12 months requiring hospitalization). Age greater than 70 years should also be considered.
- Stratification factors should include the primary tumor subsite (glottis, supraglottis [except epilarynx], or hypopharynx/epilarynx), N stage (N0, N1 vs. N2, N3), and country or region.

ASSESSMENTS

What are the optimal assessments to conduct in patients enrolled in larynx preservation trials to assess the risks and benefits of the study treatment?

Discussion

Baseline assessment. The panel discussed assessment of vocal cord fixation. Variability occurs between the US and European Union as to whether patients with vocal cord fixation are included in larynx preservation trials. Generally, these patients are included in US trials, and the overall larynx preservation rates have been high (84% in RTOG 91-11) (3). However, the rate of functional larynx preservation is not known, and 5-year laryngectomy-free survival was <50%, even with induction chemotherapy and chemoradiotherapy (4, 5). Findings from the GETTEC study, in which all patients had vocal cord fixation, suggest that these patients have worse outcomes with larynx preservation strategies (13). In that study, survival and disease-free survival were significantly shorter ($p = .006$ and $p = .02$, respectively) in patients who underwent induction chemotherapy compared with immediate surgery. Clear recording of vocal cord fixation present at baseline and restoration of mobility after induction is useful to assess the prognostic significance of mobility restoration.

Additional discussion surrounded use of baseline PET. This practice was not deemed mandatory but was considered to be useful in selective cases to assess response to induction treatment.

On-treatment assessment. Assessment of patients during induction therapy is not well established and has been arbitrary in prior study designs. A consistent and almost universal observation in neoadjuvant chemotherapy trials, however, has been the favorable prognosis associated with response to induction chemotherapy and subsequent favorable
response to radiation. Investigators at the University of Michigan demonstrated that assessing response after one cycle correctly predicted overall response 90% of the time, an association that was statistically significant ($p = 0.018$) (23). These data supported a novel approach using early response assessment to select patients for concurrent chemoradiation as opposed to primary surgery. Three-year cause-specific survival rates were 87% in an unselected group of 97 patients with advanced disease, including those with T4 primary tumors. Although this represents a favorable approach for patients with T4 disease, such patients are not recommended by the panel for inclusion in larynx preservation protocols; thus early evaluation is not recommended. Ultimately, the majority of panel members concurred that, apart from a diagnosis of early progression, formal response assessment should occur after both the second and third cycles. However, second-cycle data would be used for later correlative analysis rather than decision making (unless the patient experiences disease progression after Cycle 2). Specific criteria to determine continued participation in the trial were discussed. The panel concurred that a partial response of 50% to induction chemotherapy was sufficient to continue with a larynx preservation protocol. The panel was divided with respect to continuing larynx preservation in the event of continued vocal cord fixation after induction chemotherapy, and no specific recommendation was made.

Some trials of concurrent chemoradiotherapy have evaluated patients after 40 to 50 Gy with the purpose of discontinuing therapy for inadequate response. However, in practice, very few cases have been interrupted. The panel is not convinced that 40 to 50 Gy is sufficient to assess biologic effect or to predict response to treatment, nor is there evidence to support the predictive value of response to chemoradiotherapy. Furthermore, patients may be resistant to stopping radiation once they have committed to the full course of treatment. This issue represents a major difference between concurrent chemoradiotherapy and induction chemotherapy approaches; that is, induction chemotherapy can be interrupted based on poor response. However, interruption occurs in only about 15% to 20% of patients.

**Initial post-radiotherapy treatment assessment.** Allowing planned neck dissections on protocol was discussed by the panel. It is a controversial subject on which surgeons disagree. Issues addressed by the panel include the fact that planned neck dissection may induce severe late toxicity (24) but that enrollment in the US would dramatically decline if not allowed. Because of differing opinions, some patients will undergo planned neck dissection, and others will not, depending upon local policy. Including planned neck dissections as a stratification factor would not be feasible as dissections are generally considered mandatory in the absence of complete response (CR). Furthermore, whether patients with planned dissections were balanced between arms would be irrelevant because the probability of CR differs by treatment arm. Protocols should specify methods for collecting data to adequately assess late toxicity associated with planned neck dissections. Specifically, within each arm of the trial, among patients who achieve a nodal CR, a comparison of regional recurrence rates and complications between patients who undergo dissection and those who do not should be performed.

**Long-term follow-up.** Examples of swallowing and voice evaluations are described in Appendix B. An important long-term outcome not routinely captured in previously conducted larynx preservation trials is voice quality. Validated measures of voice quality and intelligible speech exist such as the Voice Handicap Index-10 (VHI-10) (25) and Voice-Related Quality of Life (V-RQOL) (26), but have not been consistently applied in past trials. Clearly, long-term measures of voice and speech must be conducted and applied to all patients enrolled in the trial. Several criteria should be considered when choosing the correct tool. Two straightforward criteria are (1) validation, and (2) translation into multiple languages. A less obvious criterion is balance between sufficient simplicity to ensure compliance. For example, it is important to include measures of social communication, such as speaking in public or in a restaurant, which require adequate projection. However, ultimately, it is important to keep the tool sufficiently simple to ensure use and to promote collection of such essential data (27).

Other key issues pertaining to follow-up assessments include:

- Swallowing evaluations should be conducted at 1 and 2 years because sufficient recovery after 1 year still occurs.
- The recording period for pneumonia can be terminated at 2 years to maintain consistency with the timeline of other assessment parameters.
- Esophageal dilation should be recorded with follow-up swallowing evaluations.
- Hearing should be assessed as part of the neurologic examination, and a deficiency would be detected within 6 months after treatment with a platinum agent.
- Patient Reported Outcomes (PRO) instruments are not yet validated in head-and-neck cancer, but represent an opportunity for new research.
- Appropriate tools are not available to adequately characterize the late toxic effects of therapy. Existing scales should be improved to better capture the magnitude and frequency of these toxicities.

**Recommendations**

- Baseline assessment for speech and swallowing function (e.g., a barium esophagram) may be useful for longitudinal comparison. Baseline assessment of vocal cord fixation should be performed. It is recommended that imaging (computed tomography [CT] or magnetic resonance imaging [MRI]) be performed before endoscopy when possible. When positron emission tomography–CT (PET-CT) is used for imaging instead of head-and-neck CT or MRI, the CT component should be performed with contrast and read by both a CT radiologist and a nuclear radiologist.
- Recommendations for on-treatment assessments of efficacy were developed by the panel. In the induction chemotheraapy arm, if any, response and restoration of vocal cord mobility (if initially fixed) should be assessed after two
cycles and documented for later correlation with outcomes. Unless tumor progression occurs after two cycles, further treatment decisions should be made based on response assessment after the last cycle of chemotherapy. The recommended definition of partial response is $\geq 50\%$ decrease under baseline in the sum of the products of perpendicular diameters of all measurable lesions with no progression of evaluable disease and no new lesions.

• Recommendations for initial assessment after radiotherapy were developed by the panel. The assessment should occur between 2 and 3 months after the last day of radiotherapy. Post-treatment assessment by endoscopy and comparative imaging is mandatory. Routine biopsy is not recommended. When the decision to perform salvage local surgery is made because of the presence of persistent primary tumor, total laryngectomy is preferred, but partial laryngectomy can be considered (according to local expertise). Management of the neck should be performed according to findings. Neck dissection is not recommended for nodal complete response but is not prohibited and is left to the participating center; potential impact on late toxicity should be assessed. If residual nodal disease without evidence of a persistent primary tumor is detected, a selective neck dissection is the preferred method, as is standard of care.

• Recommendations for follow-up assessments related to function and long-term toxicities include the following: a barium esophagram both at 1 year and at 2 years. Use of a feeding tube should be recorded. Episodes of pneumonia should be recorded through 2 years post-treatment (with the ultimate goal of reporting the occurrence of 3 or more episodes within 2 years). The presence of esophageal stricture requiring dilation or other surgical intervention should be recorded. Vocal cord mobility should be assessed at the end of treatment, and at 1 and 2 years. Assessment of voice should be done with a simple, validated instrument (e.g., VHI-10 or V-RQOL) at 1 and 2 years. (Note that the assessment time point of 2 years is arbitrary; choosing a consistent time point for all long-term measures facilitates clinical trial conduct and helps to ensure that assessments are conducted as specified in the protocol.) Hearing should be assessed at 6 months, and renal dysfunction requiring intervention at 6 months should be recorded.

ENDPOINTS

What are the optimal endpoints to use in larynx preservation trials? How are these endpoints defined?

Discussion

The primary goals of head-and-neck cancer treatment are cure and survival (28). An additional important objective for patients with laryngeal or hypopharyngeal cancer is the presence of a functional larynx. A composite endpoint that measures each of these important treatment goals would be the most clinically relevant primary endpoint. Neither overall survival, local regional control, laryngectomy-free survival, nor larynx preservation alone as an endpoint captures the overall true goal of treatment, which is survival with controlled disease and a functional larynx and esophagus. Events for this composite endpoint would be: death, local relapse, total laryngectomy, tracheotomy at 2 years or later, or feeding tube at 2 years or later. By defining a time point for presence of a tracheotomy or feeding tube, a temporary placement of these devices (such as may occur during radiation, particularly when combined with chemotherapy) would not be calculated as failures. Although a composite endpoint would not distinguish certain events separately, such as death with or without a larynx, secondary endpoints would capture these statistics. A preliminary test of a comparable composite endpoint comes from the EORTC studies (2, 12, 15) and GORTEC 2000-01, the latter of which captured the number of patients alive and with a functional larynx at 3 years (14).

The definition of failure was death, laryngectomy, prolonged tracheotomy, or permanent feeding tube. Data for this endpoint from GORTEC 2000-01 appear to effectively differentiate the arms (TPF arm, 53% vs. PF arm, 37%). This study will also assist in the determination of sample size based on this new composite endpoint.

The endpoint of time to tracheotomy (TTT) will provide a measure separate from loss of voice and capture an event that is dreadful to many patients. In addition, TTT will assist in measuring the occurrence of salvage partial laryngectomies that may be performed at the surgeon’s discretion. Note that candidacy for a partial laryngectomy is considered by the panel to be a contraindication to enrollment in larynx preservation trials (except in North America) unless in the presence of N2 to N3 nodes, which often require adjuvant treatment with concurrent chemoradiation. However, after induction chemotherapy, a patient may be considered suitable for this procedure, particularly because initial determination of suitability was subjective. Importantly, if a patient undergoes a partial laryngectomy as the salvage procedure, the event must be counted as a failure for the primary endpoint. Likewise, feeding tube placement is a major determinant of quality of life in head-and-neck cancer patients (29, 30) and actuarial length of time of feeding tube placement could be a valuable endpoint.

Assessing the survival impact of salvage laryngectomy is an important aspect of determining the true risk, if any, of using larynx preservation approaches. A secondary analysis of patients in RTOG 91-11 who required salvage laryngectomy demonstrated a decrement in survival compared with those who did not require a laryngectomy ($p = .02$) (9). In an analysis of the National Cancer Center Database, Hoffman et al. identified a trend toward diminished survival among patients treated for laryngeal cancer from the mid-1980s to mid-1990s (31). During that interval, initial treatment trends involved an increase in chemoradiation and a decrease in surgery. Although other factors may explain the worsening mortality demonstrated by Hoffman et al., the issue remains unresolved, and only diligent reporting of outcomes associated with salvage laryngectomy can improve knowledge about this aspect of larynx preservation strategies.
**Recommendations**

- The primary endpoint should combine assessment of survival and function. The panel created a new endpoint for this purpose: laryngo-esophageal dysfunction-free survival. This endpoint would be measured as the time from randomization, and events would include: death, local relapse, total or partial laryngectomy, tracheotomy at 2 years or later, or feeding tube at 2 years or later.
- Recommended secondary endpoints include overall survival, progression-free survival, locoregional control, time to tracheotomy, time to laryngectomy, time to discontinuation of feeding tube, and quality of life/patient reported outcomes.
- Outcomes (including survival) and characteristics of patients who fail organ preservation and require a salvage laryngectomy should be recorded and reported.

**TISSUE BANKING AND BIOMARKER ASSESSMENT**

*What are the most promising translational research opportunities that should be explored? What clinical trial practices will foster translational research?*

**Discussion**

Translational research is a key priority in head-and-neck cancer. Ideally, future studies should include a comprehensive evaluation of the head-and-neck cancer molecular signature for correlation with treatment response, toxicity, and survival. To facilitate this goal, centralized collection of tumor specimens should be pursued. The panel agreed that uniform biomarker testing is not currently considered feasible, particularly in the European Union where regulatory barriers exist for shipping tumor samples across international borders for analysis in designated laboratories.

However, pragmatic recommendations for near-term studies can be made based on early evidence with certain markers that have demonstrated prognostic value and/or a potential role for guiding treatment decisions. Overexpression of EGFR is well documented in head-and-neck cancer and has been shown to correlate with survival, DFS, and local-regional relapse (32, 33). Overexpression of the most common mutant form of EGFR, EGFRvIII, was demonstrated in 42% of 33 head-and-neck squamous cell (HNSCC) tumor samples. Transfection of EGFRvIII into HNSCC cells led to reduced response to cisplatin and cetuximab, suggesting an important role for targeting this signaling pathway (34). Data from 107 patients who received cisplatin-based chemotherapy for locally advanced HNSCC demonstrated increased benefit from chemotherapy and a lower risk of cancer-related death in patients whose tumors expressed lower levels of ERCC1 (35). In addition to EGFR (total, p-EGFR, and EGFRvIII) and ERCC1, recommended exploratory correlative biomarker studies include E-cadherin and β-catenin (markers for epithelial-to-mesenchymal transition) (36), epiregulin and amphiregulin (markers for response to EGFR antagonists in colorectal carcinoma) (37), and TP53 mutation (38).

Of additional importance is the utility of consistent sample collection across trials. Guidelines outlining which specimens should be collected before treatment will foster the advancement of translational research.

**Recommendations**

- Recommended proof-of-principle correlative biomarker studies for near-term trials include EGFR (total, p-EGFR, and EGFRvIII) defined by IHC, excision repair cross-complementary-1 gene, E-cadherin and β-catenin, epiregulin and amphiregulin, and TP53 mutation.
- Recommended samples to collect pretreatment include fresh-frozen and formalin-fixed tumor specimens, plasma and serum, and saliva.

**CONCLUSION**

Review of completed Phase III trials of larynx preservation strategies elucidated a great number of methodologic inconsistencies, making result application confusing and difficult. These observations highlighted the need to revise trial designs in several key areas in order to advance this field of study. By focusing on the three treatment goals of greatest importance to patients—survival, disease control, and laryngeal-esophageal function—clinical trials can more effectively evaluate and quantify the therapeutic benefit of novel treatment options for patients with locally advanced laryngeal and hypopharyngeal cancer.

**REFERENCES**


APPENDIX A: LARYNX PRESERVATION CONSENSUS PANEL MEMBERS

Members are listed in alphabetical order:

Ang, K. Kian, M.D. Anderson Cancer Center/radiotherapy, Houston, TX, USA

Bardet, Etienne, Centre Rene Gauducheau/radiotherapy, Nantes-Saint-Herblain, France

Barry, Beatrix, Hopital Claude Bernard Bichat/head and neck surgery, Paris, France

Bernier, Jacques, Genolier Swiss Medical Network/radiotherapy, Genolier, Switzerland

Bourhis, Jean, Institut Gustave Roussy/radiotherapy, Villejuif, France

Budach, Volker, Universitatsklinikum Charite /radiotherapy, Berlin, Germany

Calais, Gilles, Centre Henry S Kaplan/radiotherapy, Tours, France

Conley, Barbara, (Michigan State University/medical oncology, East Lansing, MI, USA

Forastiere, Arlene, Johns Hopkins University/medical oncology, Baltimore, MD, USA

Knecht, Rainald, University Medical Center Hamburg-Eppendorf/head and neck surgery, Hamburg, Germany

Langendijk, Johannes, University Medical Center Groningen /radiotherapy, Groningen, the Netherlands

Lefebvre, Jean-Louis, Centre Oscar Lambret/head and neck surgery, Lille, France

Leemans, René, Free University Hospital/surgery, the Netherlands

Licitra, Lisa, Istituto Nazionale Tumori/medical oncology, Milan, Italy

Posner, Marshall, Dana Farber Cancer Institute/medical oncology, Boston, MA, USA

de Raucourt, Dominique, Centre Francois Baclesse/head and neck surgery, Caen, France

Rolland, Frederic, Centre Rene Gauducheau/medical oncology, Nantes, France

Temam, Stephane, Institut Gustave-Roussy/head and neck surgery, Villejuif, France

Vermorken, Jan Baptist, University Hospital Antwerp/medical oncology, Edegem, Belgium

Vokes, Everett, University of Chicago Medical Center/medical oncology, Chicago, IL, USA

Weber, Randal S., M.D. Anderson Cancer Center/head and neck surgery, Houston, TX, USA

Wolf, Gregory, University of Michigan Health System/head and neck surgery, Ann Arbor, MI, USA

APPENDIX B: SWALLOWING AND VOICE EVALUATIONS MODIFIED BARIUM SWALLOW

The modified barium swallow (MBS) is considered to be the gold standard for the assessment of swallowing because of its ability to radiographically evaluate the four phases of swallowing. A typical protocol for performing MBS is described here. Studies are performed using standard radiographic systems with videofluoroscopic capabilities and the image is stored on digital videodisc (DVD). A video counter imprints a time code (accurate to 0.001 s) on the DVD. Videofluoroscopic imaging is completed in the lateral and anterior–posterior (AP) planes. The fluoroscopic camera is focused on the patient’s lips anteriorly, the posterior pharyngeal wall posteriorly, the hard palate superiorly, and the upper esophageal segment inferiorly. Fluoroscopy continues for 3 s after each swallow to allow observation of penetration or aspiration after the swallow and the patient’s reaction to it. The order of bolus presentation is as follows: two 5-ml Varibar thin liquid boluses, two 10-ml Varibar thin liquid boluses, two 20-ml Varibar thin liquid boluses, two cup sips of Varibar thin liquid, two pureed/Varibar pudding boluses, two solid boluses consisting of one fourth of a shortbread cookie or cracker coated with Varibar pudding, and two trials of the most difficult consistency in the A-P plane.

Three instruments, described below, can be used by the speech pathologist when analyzing the MBS. None can be performed in the absence of the MBS.

Penetration–aspiration Scale

The penetration–aspiration scale (PAS) is a clinician-rated eight-point, interval scale used to describe penetration and aspiration events (39). Scores are determined by depth of bolus
invasion into the airway and the patient’s response. Scores are ordinal and a higher score is assumed to be a more severe sign of dysphagia. Intra- and interjudge reliability has been established in studies of head-and-neck cancer patients.

**Oropharyngeal swallow efficiency**

Oropharyngeal swallow efficiency (OPSE) is a global measure of swallow function defined as the ratio of the percent swallowed into the esophagus divided by oropharyngeal transit time. Thus, a higher OPSE score indicates a safer and more efficient oropharyngeal swallow. The numeric score has been found to correlate with the degree of oropharyngeal dysphagia in head and neck cancer patients (40). This instrument is used during analysis of the MBS for 1-ml liquid boluses, pudding boluses, and cracker.

**National Institutes of Health Swallowing Safety Scale**

The National Institutes of Health Swallowing Safety Scale (NIH-SSS) provides a numeric score to quantify swallowing safety using seven dysphagia symptoms: residue, penetration, aspiration, response to aspiration, esophageal entry, regurgitation, and multiple swallows. The scale demonstrates high reliability (intra- and interrater intraclass correlation coefficient >0.95) and validity in dysphagic patients (41).

**Performance Status Scale for Head and Neck Cancer Patients**

The Performance Status Scale for Head and Neck Cancer Patients (PSS-H&N) is a clinician-rated instrument consisting of three subscales: normalcy of diet, public eating, and understandability of speech (42). The understandability of speech subscale is a relatively simple five-point scale. The PSS-H&N is typically completed by a research nurse or speech-language pathologist and can be administered via telephone interview should a patient miss a scheduled appointment.

**MD Anderson Dysphagia Inventory**

The MD Anderson Dysphagia Inventory (MDADI) measures swallowing-related quality of life (QOL) in patients with swallowing dysfunction. It evaluates the patient’s physical (P), emotional (E), and functional (F) perceptions of swallowing dysfunction (43). This instrument has been validated at the MD Anderson Cancer Center in head-and-neck cancer patients.

**Voice Handicap Index**

The Voice Handicap Index (VHI-10) is a patient self-assessment instrument of vocal function (25). It is an abbreviated assessment that quantifies patients’ perception of their voice handicap. It evaluates patient’s physical (P), emotional (E), and functional (F) perceptions of voice and has shown to be highly reliable for internal consistency and test-retest stability. The VHI-10 uses a 10-item questionnaire in which the patient circles the response that most accurately reflects his or her own experience on a linear scale (from “never” to “always”).

**Voice-Related Quality of Life**

The Voice-Related Quality of Life (V-RQOL) 10-item patient self-assessment instrument was developed to measure patients’ voice-related quality of life (V-RQOL) in two domains: social–emotional and physical functioning (26). The instrument is reliable and valid. Patients rate each voice-related problem on a five-point Likert-like scale (from “not a problem” to “problem as bad as it can be”).

**Consensus Auditory-Perceptual Evaluation of Voice**

The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) quantifies perceptual vocal attributes during sustained phonation, sentence production, and spontaneous speech. Overall severity, roughness, breathiness, strain, pitch, and loudness are scored from 0 to 100 using a visual analog scale. Severity ratings (mild, moderate, severe) are determined from the attribute scores (44).

**MD Anderson Symptom Inventory–Head and Neck**

The MD Anderson Symptom Inventory–Head and Neck (MDASI-HN) is a simple, reliable, and validated instrument to measure head-and-neck cancer symptom burden and the impact symptoms have on major aspects of a patient’s daily life (45). The instrument consists of 13 items included in the core MDASI with an additional 9 head-and-neck cancer–specific items: mouth sores, problem tasting food, constipation, teeth/gum problems, skin pain, voice/speech difficulties, choking/coughing, chewing/swallowing problems, and mucus.

**EORTC Quality of Life Questionnaire–Head and Neck 35**

The EORTC Quality of Life Questionnaire–Head and Neck 35 (EORTC QLQ-H&N 35) is a 35-item questionnaire that includes assessment of swallowing (four items) and speech (three items) (46). However, this instrument may not capture certain problems such as speaking in a noisy environment (47).