Am J Cancer Res 2026;16(4):1540-1566
www.ajcr.us /ISSN:2156-6976/ajcr0172594

*Original Article*

# Concurrent chemoradiotherapy vs. total laryngectomy in locally advanced laryngeal squamous cell carcinoma: a propensity score-matched retrospective cohort study

Lili Chen, Jiawei Yu, Ruinan Sheng

*Department of Head and Neck Thyroid Surgery, Harbin Medical University Cancer Hospital, 150 Haping Road, Nangang District, Harbin 150081, Heilongjiang, China*

Received March 12, 2026; Accepted April 22, 2026; Epub April 25, 2026; Published April 30, 2026

**Abstract:** This study aimed to compare overall survival (OS) and disease-free survival (DFS) between concurrent chemoradiotherapy (CCRT) and total laryngectomy (TL) in patients with locally advanced laryngeal squamous cell carcinoma (LA-LSCC), using propensity score matching (PSM) to adjust for baseline imbalances. A total of 512 patients treated at Harbin Medical University Cancer Hospital between January 2015 and December 2022 were retrospectively analyzed. PSM was performed using 1:1 nearest-neighbor matching with a caliper of 0.2, incorporating age, Eastern Cooperative Oncology Group (ECOG) performance status, T stage, N stage, tumor subsite, vocal cord fixation, thyroid cartilage invasion, and airway obstruction, resulting in 220 matched pairs. Before matching, the TL group had higher proportions of stage IVA disease, vocal cord fixation, and thyroid cartilage invasion; after matching, baseline characteristics were well balanced. Both before and after PSM, TL was associated with significantly improved OS and DFS compared with CCRT. In the matched cohort, median OS and DFS were not reached in the TL group, whereas the CCRT group had a median OS of 49.0 months and a median DFS of 42.5 months (log-rank $P < 0.001$ for both). Multivariable Cox regression identified CCRT as an independent risk factor for OS (hazard ratio [HR] = 1.819, 95% confidence interval [CI] 1.386-2.386) and DFS (HR = 2.091, 95% CI 1.601-2.732). Subgroup analyses showed inferior outcomes with CCRT in most subgroups, with ECOG performance status and vocal cord involvement acting as significant effect modifiers. Eight sensitivity analyses consistently supported the primary findings (OS: HR range 1.50-1.89; DFS: HR range 1.68-2.13). In conclusion, in this real-world cohort of patients with LA-LSCC, TL was associated with better survival outcomes than CCRT, and treatment decisions should consider T/N stage, ECOG performance status, and vocal cord fixation to enable individualized selection and avoid potential over-preservation of the larynx.

**Keywords:** Locally advanced laryngeal squamous cell carcinoma, concurrent chemoradiotherapy, total laryngectomy, propensity score matching, overall survival, disease-free survival

## Introduction

Laryngeal cancer is one of the more common malignant tumors of the head and neck. Its persistent association with tobacco and alcohol use contributes to a substantial global public health burden. According to the latest data released by the International Agency for Research on Cancer, global cancer incidence and mortality continue to rise. Head and neck malignancies represent a significant health issue [2]. A study based on Global Burden of Disease data from 1990 to 2021 further demonstrated persistent geographic and socioeconomic disparities in the incidence, mortality, and disability-adjusted life-year burden of laryngeal cancer, underscoring the need for earlier detection and improved treatment strategies to narrow outcome gaps [2]. Surveillance data from the United States indicate a decreasing trend in laryngeal cancer incidence; although mortality has declined and survival has improved, these changes have not consistently paralleled shifts in treatment approaches [3]. Therefore, balancing organ preservation with durable oncologic control remains a key challenge in the management of locally advanced laryngeal squamous cell carcinoma (LA-LSCC).

https://doi.org/10.62347/ZYTL7298

LA-LSCC refers to stage III and IVA disease. Patients with these tumors typically present with a large, bulky mass that may obstruct laryngeal function, exhibit aggressive local invasion, and carry a risk of regional lymph node metastasis. Consequently, treatment aims to achieve two opposing goals: radical tumor control and functional preservation. According to the latest guidelines, multidisciplinary team (MDT) input and individualized treatment recommendations based on tumor stage, subsite, predicted laryngeal function, and patient comorbidities are essential [4, 5]. The National Comprehensive Cancer Network (NCCN) guidelines continue to recommend surgery-based strategies for certain high-risk scenarios, such as T4a tumors, while also endorsing concurrent chemoradiotherapy (CCRT) as a larynx-preservation option for selected patients [4, 5]. Similarly, the National Cancer Institute's Physician Data Query recognizes both surgery and chemoradiotherapy as important treatment pathways, with treatment selection largely driven by functional outcomes [6]. Recently, an international multidisciplinary Delphi consensus provided a refined stratification of organ-preservation eligibility for intermediate and advanced laryngeal and hypopharyngeal cancer. The consensus recommended against non-surgical preservation in patients with T4a disease, poor baseline swallowing or airway function, or significant comorbidity, and emphasized the importance of salvage strategies and quality-of-life (QoL) endpoints [7].

The two treatment strategies have distinct histories. For a long time, total laryngectomy (TL) was considered the standard curative approach for LA-LSCC. Although TL ensures reliable local control, it carries the drawbacks of irreversible tracheostomy and loss of voice. Beginning around the early 1990s, organ-preservation strategies began to gain traction. In 1991, the Veterans Affairs Laryngeal Cancer Study was the first randomized trial to demonstrate that induction chemotherapy followed by radiotherapy did not compromise overall survival (OS) and could preserve laryngeal function [8]. Long-term follow-up of the Radiation Therapy Oncology Group (RTOG) 91-11 trial established CCRT as superior to other non-surgical regimens for laryngeal preservation, establishing CCRT as a preferred organ-preservation pathway for locally advanced disease [9]. However, as larynx-preserving techniques became widely advocated and adopted, concerns regarding "over-preservation" emerged. Some patients who did not undergo TL experienced higher rates of local recurrence, chronic swallowing dysfunction, and complications from salvage surgery. More recent reviews and consensus statements have argued that organ preservation should not be conflated with "functional larynx preservation". The treatment goal should be redefined from anatomic preservation to usable function, emphasizing better patient selection and long-term toxicity management [7, 10].

The existing evidence base has important limitations. The landmark randomized trials included patient populations that underrepresent certain high-risk subgroups encountered in real-world practice, including those with T4a disease, cartilaginous invasion, severely compromised laryngeal function, or airway compromise. Caution should therefore be exercised when extrapolating trial findings to the entire locally advanced population due to this limited representativeness [10]. Several retrospective studies comparing TL with non-surgical organ preservation in routine practice are affected by significant selection bias. Patients undergoing TL often present with more advanced tumors and higher rates of cartilage or adjacent structural invasion, which can confound survival comparisons if not properly adjusted. Despite the recent report from the American College of Surgeons Oncology Group Z11 trial indicating that only 20%-25% of patients across different centers do not receive appropriate treatment, multicenter variation persists for T3/T4 disease [11]. Furthermore, the trade-offs between TL and non-surgical preservation with respect to oncologic and functional outcomes are complex, with the direction of benefit varying by stage and population subgroup [12]. Salvage surgery (e.g., salvage laryngectomy) plays an important role in the treatment of recurrent or residual disease but is not without complications and functional costs. Therefore, getting the first treatment decision right is critical [13]. In Chinese patient populations, disparities in stage distribution at presentation, treatment access, radiotherapy techniques, and peritreatment management necessitate large-sample studies with adequate confounding control to guide individualized treatment decision-making for LA-LSCC in the domestic clinical setting.

In this study, propensity score matching (PSM) was applied to optimize baseline balance. OS and disease-free survival (DFS) were compared between CCRT and TL in patients with LA-LSCC. The authors conducted various sensitivity analyses to assess the robustness of the findings, as well as subgroup and interaction analyses to explore heterogeneity. We aimed to generate more actionable evidence for risk stratification and treatment selection in the clinical setting where the tension between tumor control and functional preservation is most acute. Ultimately, this may help reduce both potential "over-preservation" and unnecessary radical surgery, while promoting patient-centered, precision-guided decision-making focused on long-term survival and usable function.

## Materials and methods

### *Study design and population*

This single-center, retrospective cohort study was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology statement. We retrospectively analyzed 512 patients with LA-LSCC treated at Harbin Medical University Cancer Hospital between January 2015 and December 2022. Of these, 244 received CCRT and 268 underwent TL. The study protocol was approved by the Ethics Committee of Harbin Medical University Cancer Hospital. Due to the retrospective design and anonymized data handling, the requirement for informed consent was waived.

### *Sample size estimation*

The primary endpoint was OS. The minimum number of events required to detect a difference in OS between the two groups was estimated using the Schoenfeld formula for the log-rank test. A two-sided $\alpha$ of 0.05 and a power (1-$\beta$) of 0.80 were specified. For the purpose of sample size estimation, an approximate 1:1 allocation ratio was assumed (allocation proportion = 0.5 for each group). As this was a retrospective cohort study, the actual number of patients in each group was determined by real-world clinical practice rather than pre-specified allocation. The assumed effect size was based on a published retrospective cohort of patients with LA-LSCC (T3-T4a), in which the multivariable Cox model yielded a hazard ratio (HR) of approximately 1.50 for CCRT versus surgery [14].

Substituting into the Schoenfeld formula with $Z(1-\alpha/2) = 1.96$, $Z(1-\beta) = 0.84$, HR = 1.50, and P = 0.5, the minimum required number of events was $D \approx 175$. Based on published approximately 55%-65% for LA-LSCC, the estimated 5-year death rates range from 35% to 45%. Using a conservative event rate of 40%, the minimum total sample was calculated as N = D/event rate = 175/0.40 = 438 patients. The authors further increased the required sample size to approximately 482 patients to account for an estimated 10% loss to follow-up and missing data. Thus, the 512 enrolled patients met this requirement, confirming sufficient statistical power.

### *Eligibility criteria*

Inclusion criteria: (1) Histopathologically confirmed squamous cell carcinoma of the larynx. (2) Clinically staged as locally advanced disease (stage III or IVA) according to the American Joint Committee on Cancer 8th edition [15]. (3) Initial curative-intent treatment was either CCRT or TL, with or without postoperative adjuvant radiotherapy or adjuvant chemoradiotherapy. (4) Complete clinical, imaging, and follow-up data available for primary endpoint assessment.

Exclusion criteria: (1) Distant metastasis (M1 disease). (2) Prior head and neck radiotherapy or chemotherapy, or recurrent laryngeal cancer receiving second-line treatment. (3) Concomitant active malignancy or history of another malignancy that could confound prognostic assessment. (4) Missing key baseline variables required for PSM.

### *Treatment protocols*

CCRT group: Patients received definitive CCRT using intensity-modulated radiation therapy. The primary tumor and involved nodal regions received 66-70 Gy in 33-35 fractions, and elective volumes received 50-56 Gy in 25-28 fractions. Concurrent chemotherapy was cisplatin-based, with a standard regimen of single-agent cisplatin (80-100 mg/m$^2$ every three weeks for 2-3 cycles); alternative regimens were used based on tolerability. All plans were determined

via MDT discussion. Radiotherapy completion rates and chemotherapy cycles were recorded.

TL group: Patients underwent TL with concurrent neck dissection as indicated. Postoperative adjuvant therapy was based on pathologic risk features. Adverse features (positive/close margins or extranodal extension) prompted postoperative CCRT. Intermediate-risk features (pT3-T4a, lymphovascular invasion, perineural invasion, or multiple nodal stations) prompted postoperative radiotherapy alone. Patients without these features underwent surveillance. Adjuvant decisions followed NCCN guidelines and were confirmed via MDT discussion. Proportions receiving postoperative radiotherapy, chemoradiotherapy, or surgery alone were recorded.

*Clinical data collection and measurement*

The following data were extracted from medical records and the follow-up database. Demographics and general status included age (dichotomized at 65 years), sex, body mass index (BMI) category (< 18.5, 18.5-23.9, or ≥ 24.0 kg/m$^2$), smoking history, alcohol history, and Eastern Cooperative Oncology Group (ECOG) performance status score (0, 1, or 2). Comorbidities recorded included hypertension, diabetes mellitus, coronary heart disease (CHD), and chronic obstructive pulmonary disease (COPD). Oncologic characteristics included T stage (T3 or T4a), N stage (N0, N1, or N2), clinical stage (III or IVA), tumor subsite (glottic or supraglottic), tumor differentiation grade (well, moderate, or poor), vocal cord fixation, airway obstruction, thyroid cartilage invasion, and pre-epiglottic space invasion. All staging was based on the American Joint Committee on Cancer 8th edition. Vocal cord fixation and airway status were assessed primarily by flexible laryngoscopy. Thyroid cartilage invasion and pre-epiglottic space invasion were determined from pretreatment contrast-enhanced computed tomography (CT) and/or magnetic resonance imaging reports, independently reviewed by two senior radiologists, with disagreements resolved by a third reviewer.

*Laboratory parameters*

Fasting venous blood samples were collected before treatment initiation: prior to the start of CCRT in the CCRT group, and within 7 days before surgery in the TL group. Hemoglobin (Hb) and neutrophil and lymphocyte counts were measured using an automated hematology analyzer (Sysmex XN-9000, Sysmex Corporation, Kobe, Japan). Serum albumin (Alb) was measured using an automated biochemistry analyzer (Roche cobas c702, Roche Diagnostics, Basel, Switzerland). The neutrophil-to-lymphocyte ratio (NLR) was calculated as the absolute neutrophil count divided by the absolute lymphocyte count. All assays were performed by the hospital laboratory following standard operating procedures, with routine participation in internal and external quality control programs.

*Follow-up protocol*

Follow-up data were retrospectively obtained from medical records, institutional databases, and telephone interviews. Although this was not a prospectively designed protocol, patients at Harbin Medical University Cancer Hospital are routinely followed after treatment according to standard clinical practice. The typical follow-up schedule includes visits every 3 months during the first 2 years, every 6 months from years 3 to 5, and annually thereafter. Follow-up evaluations generally comprise clinical history, physical examination, flexible laryngoscopy, and imaging studies (CT and/or magnetic resonance imaging), with positron emission tomography-CT performed when clinically indicated. The follow-up cutoff date was January 2026. The starting point for both OS and DFS was uniformly defined as the date of treatment initiation: the first session of radiotherapy or chemotherapy for the CCRT group, and the date of surgery for the TL group.

*PSM*

PSM was used to reduce treatment selection bias [16, 17]. A logistic regression model was fitted with treatment modality (CCRT vs. TL) as the dependent variable to estimate propensity scores. Covariates were selected based on three principles: (1) factors known to influence treatment decisions in clinical practice; (2) variables reported in the literature as important prognostic factors for LA-LSCC; and (3) baseline characteristics that differed or tended to differ between the two groups before matching. Following these principles, eight covariates were included in the model: age, ECOG perfor-

mance status, T stage, N stage, tumor subsite, vocal cord fixation, thyroid cartilage invasion, and airway obstruction. T stage, N stage, thyroid cartilage invasion, and vocal cord fixation directly reflect the extent of local tumor invasion and are critical when choosing between organ preservation and TL. Airway obstruction influences resistance to non-surgical treatment. ECOG performance status reflects overall functional status and treatment tolerability. Age and tumor subsite are common clinical factors affecting treatment choice. Because clinical stage is interdependent with T and N stages, introducing complex collinearity, it was not included in the matching model.

We performed nearest-neighbor 1:1 matching without replacement. The caliper was set to 0.2 times the standard deviation of the logit of the propensity score. The standardized mean difference (SMD) was calculated for all baseline covariates after matching to assess balance, with an SMD below 0.1 indicating adequate balance. Matching quality was visualized using propensity score density plots and a Love plot.

*Outcome measures*

Primary endpoint: OS was defined as the time from the date of initial treatment to death from any cause or the date of last follow-up.

Secondary endpoint: DFS was defined as the time from the date of initial treatment to the first documented disease recurrence (local, regional, or distant), disease progression, or death from any cause, whichever occurred first, or the date of last follow-up.

One-year, 3-year, and 5-year OS and DFS rates were reported.

*Statistical analysis*

All statistical analyses were performed using R software (version 4.5.1) and SPSS (version 27.0). Continuous variables were expressed as mean ± standard deviation for normally distributed data or as median with interquartile range (IQR) for skewed data; between-group comparisons were performed using the independent-samples t-test or the Mann-Whitney U test, as appropriate. Categorical variables were expressed as counts (percentages) and compared using the $\chi^2$ test or Fisher's exact test.

Survival curves were estimated using the Kaplan-Meier method and compared using the log-rank test. Prognostic factor analysis was conducted using Cox proportional hazards (PH) regression models, and HRs with 95% confidence intervals were reported. Variables with $P < 0.10$ in univariate Cox analysis were entered into the multivariable model. Before fitting the multivariable model, collinearity among candidate variables was assessed using Spearman correlation analysis and the generalized variance inflation factor (GVIF). A GVIF^(1/(2×Df)) value below 2 was considered indicative of no serious multicollinearity [18].

The PH assumption for each multivariable Cox model was evaluated using Schoenfeld residual tests. If the global PH test was significant (global $P < 0.05$), variables that individually violated the PH assumption (single-variable Schoenfeld test $P < 0.05$) were removed stepwise until a final model satisfying the PH assumption was obtained. Schoenfeld residual plots over time were inspected to confirm the stability of residuals in the final model.

Subgroup analyses were pre-specified according to the following clinical variables: T stage (T3 vs. T4a), N stage (N0 vs. N1-2), age (< 65 vs. ≥ 65 years), ECOG performance status (0 vs. 1-2), tumor differentiation (well/moderate vs. poor), vocal cord fixation (present vs. absent), thyroid cartilage invasion (present vs. absent), and tumor subsite (glottic vs. supraglottic). Within each subgroup, the HR and 95% confidence interval for CCRT versus TL were calculated. Interaction tests (likelihood ratio tests) were used to evaluate effect modification between each subgroup variable and treatment modality.

To verify robustness, multiple sensitivity analyses were performed: (1) pre-match unadjusted univariate model; (2) pre-match 10-variable multivariable-adjusted model; (3) pre-match 6-variable (final model) multivariable-adjusted model; (4) post-match unadjusted univariate model; (5) post-match 6-variable multivariable-adjusted model (primary analysis); (6) inverse probability of treatment weighting (IPTW) with stabilized, truncated weights [19]; (7) propensity score stratification (stratified Cox regression by quintiles); and (8) propensity score covariate adjustment (propensity score entered as a continuous covariate in the Cox model).

An additional stratified sensitivity analysis was conducted within the TL group based on whether postoperative adjuvant therapy was administered, to assess the potential influence of adjuvant treatment on the main conclusions.

All tests were two-sided, and P < 0.05 was considered statistically significant.

## Results

*Pre-match baseline characteristics*

A total of 512 patients with LA-LSCC were included, comprising 244 in the CCRT group and 268 in the TL group. Before matching, the two groups differed significantly in clinical stage, vocal cord fixation, and thyroid cartilage invasion. Specifically, the TL group had a higher proportion of stage IVA patients than the CCRT group (P = 0.023), as well as higher rates of vocal cord fixation (P = 0.040) and thyroid cartilage invasion (P = 0.013). No significant differences were observed between the two groups in age (P = 0.140), sex (P = 0.620), BMI category (P = 0.747), smoking history (P = 0.502), alcohol history (P = 0.541), ECOG performance status (P = 0.462), hypertension (P = 0.606), diabetes (P = 0.524), CHD (P = 0.613), COPD (P = 0.605), T stage (P = 0.082), N stage (P = 0.315), tumor subsite (P = 0.643), tumor differentiation (P = 0.606), airway obstruction (P = 0.143), pre-epiglottic space invasion (P = 0.238), pretreatment Hb (P = 0.520), Alb (P = 0.974), or NLR (P = 0.829) (all P > 0.05) (**Table 1**).

*PSM and post-match baseline characteristics*

Based on the pre-match imbalances, propensity score matching (PSM) was performed using nearest-neighbor 1:1 matching with a caliper of 0.2. The following covariates were included: age, ECOG performance status, T stage, N stage, tumor subsite, vocal cord fixation, thyroid cartilage invasion, and airway obstruction. After matching, 220 patients remained in each group, and 72 patients were unmatched. Following matching, none of the baseline characteristics differed significantly between the two groups, including age ($P$ = 0.441), sex ($P$ = 0.623), BMI category ($P$ = 0.974), smoking history ($P$ = 0.565), alcohol history ($P$ = 0.633), ECOG performance status ($P$ = 0.911), hypertension ($P$ = 0.921), diabetes mellitus ($P$ = 0.427), CHD ($P$ = 0.195), COPD ($P$ = 0.522), T stage ($P$ = 0.695), N stage ($P$ = 0.900), clinical stage ($P$ = 0.391), tumor subsite ($P$ = 0.563), tumor differentiation ($P$ = 0.796), vocal cord fixation ($P$ = 0.775), airway obstruction ($P$ = 0.647), thyroid cartilage invasion ($P$ = 0.835), pre-epiglottic space invasion ($P$ = 0.238), pretreatment Hb ($P$ = 0.533), pretreatment Alb ($P$ = 0.854), and pretreatment NLR ($P$ = 0.914) (all $P$ > 0.05). These results indicated that baseline characteristics were well balanced after matching (**Table 2** and **Figure 1**).

*Follow-up and survival events*

Among the 512 patients in the pre-match cohort, the median follow-up was 50.0 months. The median follow-up was 56.0 months (IQR: 28.8-59.0) in the TL group and 48.0 months (IQR: 22.8-53.0) in the CCRT group. In the post-match cohort of 440 patients, the median follow-up was also 50.0 months. The TL group again had a longer follow-up duration, with a median of 56.0 months (IQR: 33.0-59.0) compared with 48.0 months (IQR: 21.0-53.0) in the CCRT group. During the observation period, a total of 220 deaths occurred, with fewer deaths in the TL group. Overall, 237 DFS events were recorded, and the TL group had a lower DFS event rate. In the post-match cohort, the 1-year, 3-year, and 5-year OS and DFS rates were all higher in the TL group than in the CCRT group. Waterfall plots of patient outcomes showed that death and recurrence/progression events were more densely clustered in the CCRT group and tended to occur during the mid-to-late follow-up period. Survival heatmaps revealed a widening gap in survival between the two groups over time, with the largest difference observed at the 5-year mark (**Figure 2**).

*Survival outcomes by treatment group*

Before matching, OS was significantly higher in the TL group than in the CCRT group (log-rank P = 0.002). Median OS was not reached in the TL group, whereas it was 54.0 months in the CCRT group. DFS was also significantly higher in the TL group (log-rank P < 0.001). Median DFS was not reached for TL patients but was 45.0 months for CCRT patients. Kaplan-Meier curves separated early during follow-up, with the gap widening progressively over time.

After matching, the survival difference between the two treatment groups remained significant. For OS, the median was not reached in the TL

**Table 1.** Baseline characteristics of patients before propensity score matching

| Variable | Total (N = 512) | CCRT (n = 244) | TL (n = 268) | Test statistic | *P* value |
|---|---|---|---|---|---|
| Age | | | | $\chi^2 = 2.180$ | 0.140 |
| < 65 years | 289 (56.45%) | 146 (59.84%) | 143 (53.36%) | | |
| ≥ 65 years | 223 (43.55%) | 98 (40.16%) | 125 (46.64%) | | |
| Sex | | | | $\chi^2 = 0.246$ | 0.620 |
| Male | 463 (90.43%) | 219 (89.75%) | 244 (91.04%) | | |
| Female | 49 (9.57%) | 25 (10.25%) | 24 (8.96%) | | |
| BMI category | | | | $\chi^2 = 0.584$ | 0.747 |
| < 18.5 (underweight) | 54 (10.55%) | 28 (11.48%) | 26 (9.70%) | | |
| 18.5-23.9 (normal) | 284 (55.47%) | 136 (55.74%) | 148 (55.22%) | | |
| ≥ 24.0 (overweight/obese) | 174 (33.98%) | 80 (32.79%) | 94 (35.07%) | | |
| Smoking history | | | | $\chi^2 = 0.451$ | 0.502 |
| Yes | 399 (77.93%) | 187 (76.64%) | 212 (79.10%) | | |
| No | 113 (22.07%) | 57 (23.36%) | 56 (20.90%) | | |
| Alcohol history | | | | $\chi^2 = 0.373$ | 0.541 |
| Yes | 280 (54.69%) | 130 (53.28%) | 150 (55.97%) | | |
| No | 232 (45.31%) | 114 (46.72%) | 118 (44.03%) | | |
| ECOG performance status | | | | $\chi^2 = 1.542$ | 0.462 |
| 0 | 215 (41.99%) | 108 (44.26%) | 107 (39.93%) | | |
| 1 | 221 (43.16%) | 104 (42.62%) | 117 (43.66%) | | |
| 2 | 76 (14.84%) | 32 (13.11%) | 44 (16.42%) | | |
| Hypertension | | | | $\chi^2 = 0.266$ | 0.606 |
| Yes | 180 (35.16%) | 83 (34.02%) | 97 (36.19%) | | |
| No | 332 (64.84%) | 161 (65.98%) | 171 (63.81%) | | |
| Diabetes mellitus | | | | $\chi^2 = 0.405$ | 0.524 |
| Yes | 109 (21.29%) | 49 (20.08%) | 60 (22.39%) | | |
| No | 403 (78.71%) | 195 (79.92%) | 208 (77.61%) | | |
| CHD | | | | $\chi^2 = 0.256$ | 0.613 |
| Yes | 67 (13.09%) | 30 (12.30%) | 37 (13.81%) | | |
| No | 445 (86.91%) | 214 (87.70%) | 231 (86.19%) | | |
| COPD | | | | $\chi^2 = 0.268$ | 0.605 |
| Yes | 80 (15.62%) | 36 (14.75%) | 44 (16.42%) | | |
| No | 432 (84.38%) | 208 (85.25%) | 224 (83.58%) | | |
| T stage | | | | $\chi^2 = 3.026$ | 0.082 |
| T3 | 305 (59.57%) | 155 (63.52%) | 150 (55.97%) | | |
| T4a | 207 (40.43%) | 89 (36.48%) | 118 (44.03%) | | |
| N stage | | | | $\chi^2 = 2.312$ | 0.315 |
| N0 | 213 (41.60%) | 108 (44.26%) | 105 (39.18%) | | |
| N1 | 180 (35.16%) | 86 (35.25%) | 94 (35.07%) | | |
| N2 | 119 (23.24%) | 50 (20.49%) | 69 (25.75%) | | |
| Clinical stage | | | | $\chi^2 = 5.182$ | 0.023 |
| III | 250 (48.83%) | 132 (54.10%) | 118 (44.03%) | | |
| IVA | 262 (51.17%) | 112 (45.90%) | 150 (55.97%) | | |
| Tumor subsite | | | | $\chi^2 = 0.214$ | 0.643 |
| Supraglottic | 217 (42.38%) | 106 (43.44%) | 111 (41.42%) | | |
| Glottic | 295 (57.62%) | 138 (56.56%) | 157 (58.58%) | | |
| Tumor differentiation | | | | $\chi^2 = 1.001$ | 0.606 |
| Well | 80 (15.62%) | 42 (17.21%) | 38 (14.18%) | | |
| Moderate | 296 (57.81%) | 140 (57.38%) | 156 (58.21%) | | |
| Poor | 136 (26.56%) | 62 (25.41%) | 74 (27.61%) | | |
| Vocal cord fixation | | | | $\chi^2 = 4.228$ | 0.040 |
| Yes | 251 (49.02%) | 108 (44.26%) | 143 (53.36%) | | |
| No | 261 (50.98%) | 136 (55.74%) | 125 (46.64%) | | |
| Airway obstruction | | | | $\chi^2 = 2.147$ | 0.143 |
| Yes | 124 (24.22%) | 52 (21.31%) | 72 (26.87%) | | |
| No | 388 (75.78%) | 192 (78.69%) | 196 (73.13%) | | |

| | | | | | |
|---|---|---|---|---|---|
| Thyroid cartilage invasion | | | | $\chi^2$ = 6.141 | 0.013 |
| Yes | 166 (32.42%) | 66 (27.05%) | 100 (37.31%) | | |
| No | 346 (67.58%) | 178 (72.95%) | 168 (62.69%) | | |
| Pre-epiglottic space invasion | | | | $\chi^2$ = 1.391 | 0.238 |
| Yes | 134 (26.17%) | 58 (23.77%) | 76 (28.36%) | | |
| No | 378 (73.83%) | 186 (76.23%) | 192 (71.64%) | | |
| Hb (g/L) | 130.00 [119.00, 143.00] | 129.00 [118.00, 143.25] | 131.00 [120.00, 142.25] | Z = 0.547 | 0.584 |
| Alb (g/L) | 38.35 [35.00, 41.60] | 38.30 [34.80, 42.10] | 38.40 [35.10, 41.23] | Z = 0.184 | 0.854 |
| NLR | 2.82 ± 1.31 | 2.83 ± 1.32 | 2.82 ± 1.31 | t = -0.109 | 0.914 |

Note: Data are presented as n (%), median [interquartile range (IQR)], or mean ± standard deviation (SD). CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; BMI, Body mass index; ECOG, Eastern Cooperative Oncology Group; CHD, Coronary heart disease; COPD, Chronic obstructive pulmonary disease; Hb, Hemoglobin; Alb, Albumin; NLR, Neutrophil-to-lymphocyte ratio. Continuous variables were compared using the independent-samples t-test (normal distribution) or Mann-Whitney U test (skewed distribution). Categorical variables were compared using the $\chi^2$ test. $P < 0.05$ is considered statistically significant.

**Table 2.** Baseline characteristics after propensity score matching

| Variable | Total (N = 440) | CCRT (n = 220) | TL (n = 220) | Test statistic | *P* value |
|---|---|---|---|---|---|
| Age | | | | $\chi^2$ = 0.594 | 0.441 |
| < 65 years | 252 (57.27%) | 130 (59.09%) | 122 (55.45%) | | |
| ≥ 65 years | 188 (42.73%) | 90 (40.91%) | 98 (44.55%) | | |
| Sex | | | | $\chi^2$ = 0.242 | 0.623 |
| Male | 399 (90.68%) | 198 (90.00%) | 201 (91.36%) | | |
| Female | 41 (9.32%) | 22 (10.00%) | 19 (8.64%) | | |
| BMI category | | | | $\chi^2$ = 0.053 | 0.974 |
| < 18.5 (underweight) | 49 (11.14%) | 25 (11.36%) | 24 (10.91%) | | |
| 18.5-23.9 (normal) | 249 (56.59%) | 125 (56.82%) | 124 (56.36%) | | |
| ≥ 24.0 (overweight/obese) | 142 (32.27%) | 70 (31.82%) | 72 (32.73%) | | |
| Smoking history | | | | $\chi^2$ = 0.331 | 0.565 |
| Yes | 343 (77.95%) | 169 (76.82%) | 174 (79.09%) | | |
| No | 97 (22.05%) | 51 (23.18%) | 46 (20.91%) | | |
| Alcohol history | | | | $\chi^2$ = 0.228 | 0.633 |
| Yes | 235 (53.41%) | 120 (54.55%) | 115 (52.27%) | | |
| No | 205 (46.59%) | 100 (45.45%) | 105 (47.73%) | | |
| ECOG performance status | | | | $\chi^2$ = 0.186 | 0.911 |
| 0 | 186 (42.27%) | 93 (42.27%) | 93 (42.27%) | | |
| 1 | 189 (42.95%) | 96 (43.64%) | 93 (42.27%) | | |
| 2 | 65 (14.77%) | 31 (14.09%) | 34 (15.45%) | | |
| Hypertension | | | | $\chi^2$ = 0.010 | 0.921 |
| Yes | 155 (35.23%) | 78 (35.45%) | 77 (35.00%) | | |
| No | 285 (64.77%) | 142 (64.55%) | 143 (65.00%) | | |
| Diabetes mellitus | | | | $\chi^2$ = 0.630 | 0.427 |
| Yes | 101 (22.95%) | 47 (21.36%) | 54 (24.55%) | | |
| No | 339 (77.05%) | 173 (78.64%) | 166 (75.45%) | | |
| CHD | | | | $\chi^2$ = 1.683 | 0.195 |
| Yes | 55 (12.50%) | 23 (10.45%) | 32 (14.55%) | | |
| No | 385 (87.50%) | 197 (89.55%) | 188 (85.45%) | | |
| COPD | | | | $\chi^2$ = 0.411 | 0.522 |
| Yes | 73 (16.59%) | 34 (15.45%) | 39 (17.73%) | | |
| No | 367 (83.41%) | 186 (84.55%) | 181 (82.27%) | | |
| T stage | | | | $\chi^2$ = 0.154 | 0.695 |
| T3 | 272 (61.82%) | 138 (62.73%) | 134 (60.91%) | | |
| T4a | 168 (38.18%) | 82 (37.27%) | 86 (39.09%) | | |
| N stage | | | | $\chi^2$ = 0.210 | 0.900 |
| N0 | 186 (42.27%) | 94 (42.73%) | 92 (41.82%) | | |
| N1 | 156 (35.45%) | 79 (35.91%) | 77 (35.00%) | | |
| N2 | 98 (22.27%) | 47 (21.36%) | 51 (23.18%) | | |

| | | | | | |
|---|---|---|---|---|---|
| Clinical stage | | | | $\chi^2$ = 0.737 | 0.391 |
| III | 227 (51.59%) | 118 (53.64%) | 109 (49.55%) | | |
| IVA | 213 (48.41%) | 102 (46.36%) | 111 (50.45%) | | |
| Tumor subsite | | | | $\chi^2$ = 0.335 | 0.563 |
| Supraglottic | 186 (42.27%) | 96 (43.64%) | 90 (40.91%) | | |
| Glottic | 254 (57.73%) | 124 (56.36%) | 130 (59.09%) | | |
| Tumor differentiation | | | | $\chi^2$ = 0.457 | 0.796 |
| Well | 69 (15.68%) | 36 (16.36%) | 33 (15.00%) | | |
| Moderate | 247 (56.14%) | 125 (56.82%) | 122 (55.45%) | | |
| Poor | 124 (28.18%) | 59 (26.82%) | 65 (29.55%) | | |
| Vocal cord fixation | | | | $\chi^2$ = 0.082 | 0.775 |
| Yes | 209 (47.50%) | 103 (46.82%) | 106 (48.18%) | | |
| No | 231 (52.50%) | 117 (53.18%) | 114 (51.82%) | | |
| Airway obstruction | | | | $\chi^2$ = 0.210 | 0.647 |
| Yes | 98 (22.27%) | 47 (21.36%) | 51 (23.18%) | | |
| No | 342 (77.73%) | 173 (78.64%) | 169 (76.82%) | | |
| Thyroid cartilage invasion | | | | $\chi^2$ = 0.043 | 0.835 |
| Yes | 132 (30.00%) | 65 (29.55%) | 67 (30.45%) | | |
| No | 308 (70.00%) | 155 (70.45%) | 153 (69.55%) | | |
| Pre-epiglottic space invasion | | | | $\chi^2$ = 1.394 | 0.238 |
| Yes | 119 (27.05%) | 54 (24.55%) | 65 (29.55%) | | |
| No | 321 (72.95%) | 166 (75.45%) | 155 (70.45%) | | |
| Hb (g/L) | 130.00 [119.00, 143.00] | 129.00 [118.00, 143.25] | 131.00 [120.00, 142.25] | Z = 0.547 | 0.584 |
| Alb (g/L) | 38.35 [35.00, 41.60] | 38.30 [34.80, 42.10] | 38.40 [35.10, 41.23] | Z = 0.184 | 0.854 |
| NLR | 2.82 ± 1.31 | 2.83 ± 1.32 | 2.82 ± 1.31 | t = -0.109 | 0.914 |

Note: Data are presented as n (%), mean ± standard deviation (SD), or median [interquartile range (IQR)]. CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; BMI, Body mass index; ECOG, Eastern Cooperative Oncology Group; CHD, Coronary heart disease; COPD, Chronic obstructive pulmonary disease; Hb, Hemoglobin; Alb, Albumin; NLR, Neutrophil-to-lymphocyte ratio. Continuous variables were compared using the independent-samples t-test (normal distribution) or Mann-Whitney U test (skewed distribution); categorical variables were compared using the $\chi^2$ test.

group, whereas the CCRT group had a median OS of 49.0 months (log-rank P < 0.001). For DFS, the median was not reached in the TL group, compared with 42.5 months in the CCRT group (log-rank P < 0.001). Notably, the survival difference between the two groups did not diminish after PSM; if anything, it became more pronounced after adjusting for confounders. Kaplan-Meier curves demonstrated early divergence and a progressively widening gap between the two groups, regardless of matching status (**Figure 3**).

*Univariate Cox regression analysis before and after matching*

In the pre-match univariate Cox analysis, CCRT was identified as a significant risk factor for both OS (P = 0.002) and DFS (P < 0.001). The following factors were significantly associated with worse OS and DFS: ECOG performance status score of 2 (OS: P < 0.001; DFS: P < 0.001), T4a stage (both P < 0.001), N2 stage (both P < 0.001), clinical stage IVA (both P < 0.001), thyroid cartilage invasion (OS: P = 0.002; DFS: P = 0.007), and NLR (both: P < 0.001; DFS: P = 0.001). BMI ≥ 24.0 reached significance only for DFS (P = 0.028), and poor differentiation was borderline significant for DFS alone (P = 0.048); neither was significantly associated with OS. The remaining variables did not reach statistical significance for either endpoint (all P > 0.05) (**Table 3**).

Post-match univariate results were broadly consistent with the pre-match findings. CCRT remained a significant risk factor for OS (P < 0.001) and DFS (P < 0.001), with slightly larger HR estimates than those observed before matching. ECOG score of 2, T4a stage, N2 stage, clinical stage IVA, thyroid cartilage invasion, and NLR continued to predict worse OS and DFS. A notable difference was that BMI ≥ 24.0 now reached significance for both OS (P = 0.030) and DFS (P = 0.010). Poor differentiation remained significant for DFS (P = 0.039) and showed a borderline trend for OS (P = 0.069). Other variables were not significant (all P > 0.05) (**Table 4**).

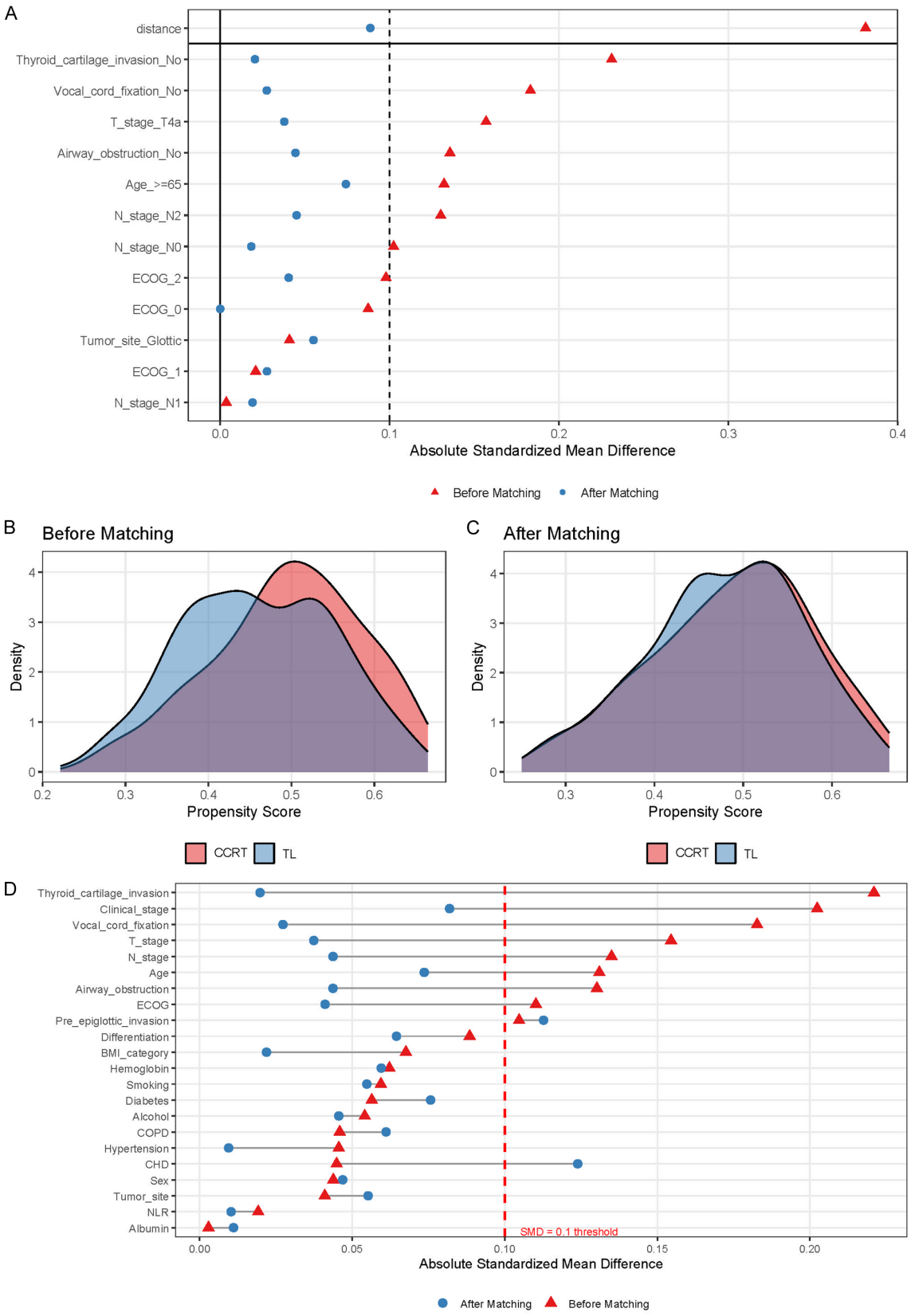


**Figure 1.** Balance assessment of propensity score matching. A. Love plot displaying the SMDs of all covariates before and after matching. An SMD below 0.1 indicates acceptable balance. B. Density distribution of propensity

scores in the CCRT and TL groups before matching. C. Density distribution of propensity scores in the CCRT and TL groups after matching. D. Dumbbell plot comparing the SMDs of baseline covariates before and after matching. Note: CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; ECOG, Eastern Cooperative Oncology Group; SMD, Standardized mean difference; COPD, Chronic obstructive pulmonary disease; BMI, Body mass index; CHD, Coronary heart disease; NLR, Neutrophil-to-lymphocyte ratio.

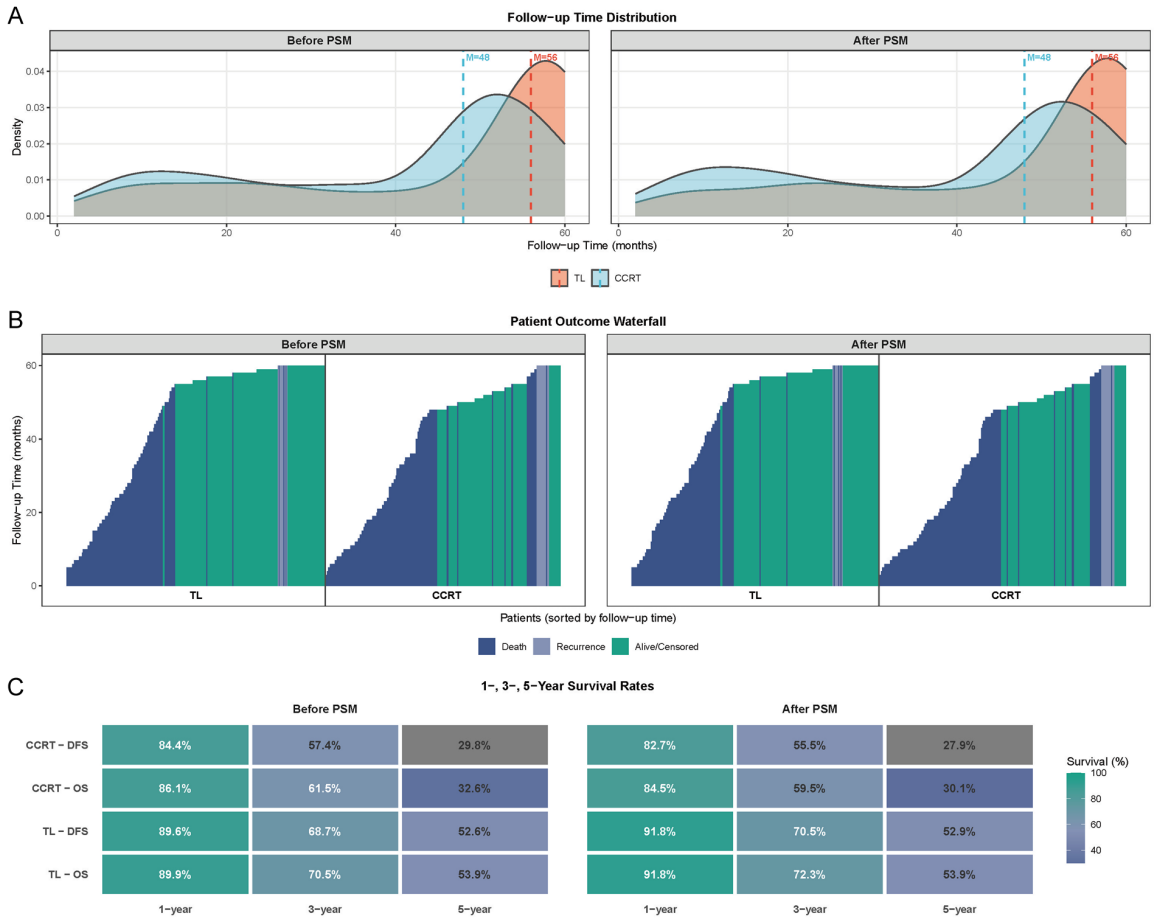


**Figure 2.** Follow-up and survival event overview before and after PSM. A. Density distribution of follow-up duration in the CCRT and TL groups before and after PSM. B. Waterfall plots of patient outcomes in the CCRT and TL groups before and after PSM. C. Heatmap summarizing 1-year, 3-year, and 5-year OS and DFS rates in the two groups before and after PSM. Note: CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; PSM, Propensity score matching; OS, Overall survival; DFS, Disease-free survival.

*Collinearity diagnostics for the multivariable Cox model*

To ensure the reliability of the multivariable Cox regression results, collinearity was assessed among variables that met the P < 0.1 threshold in univariate analysis. Because clinical stage is jointly determined by T and N stages, it was excluded a priori. The remaining 10 variables entered into the diagnostics were treatment modality, BMI category, ECOG performance status, T stage, N stage, tumor differentiation, vocal cord fixation, thyroid cartilage invasion, pre-epiglottic space invasion, and NLR.

Spearman correlation analysis showed that most pairwise correlation coefficients were below 0.2 in absolute value before matching. The only strong negative correlation was observed between moderate and poor differentiation (r = -0.704), which is an expected structural artifact of dummy-coding a single categorical variable. N1 and N2 stages also showed a moderate negative correlation (r = -0.405), another struc-

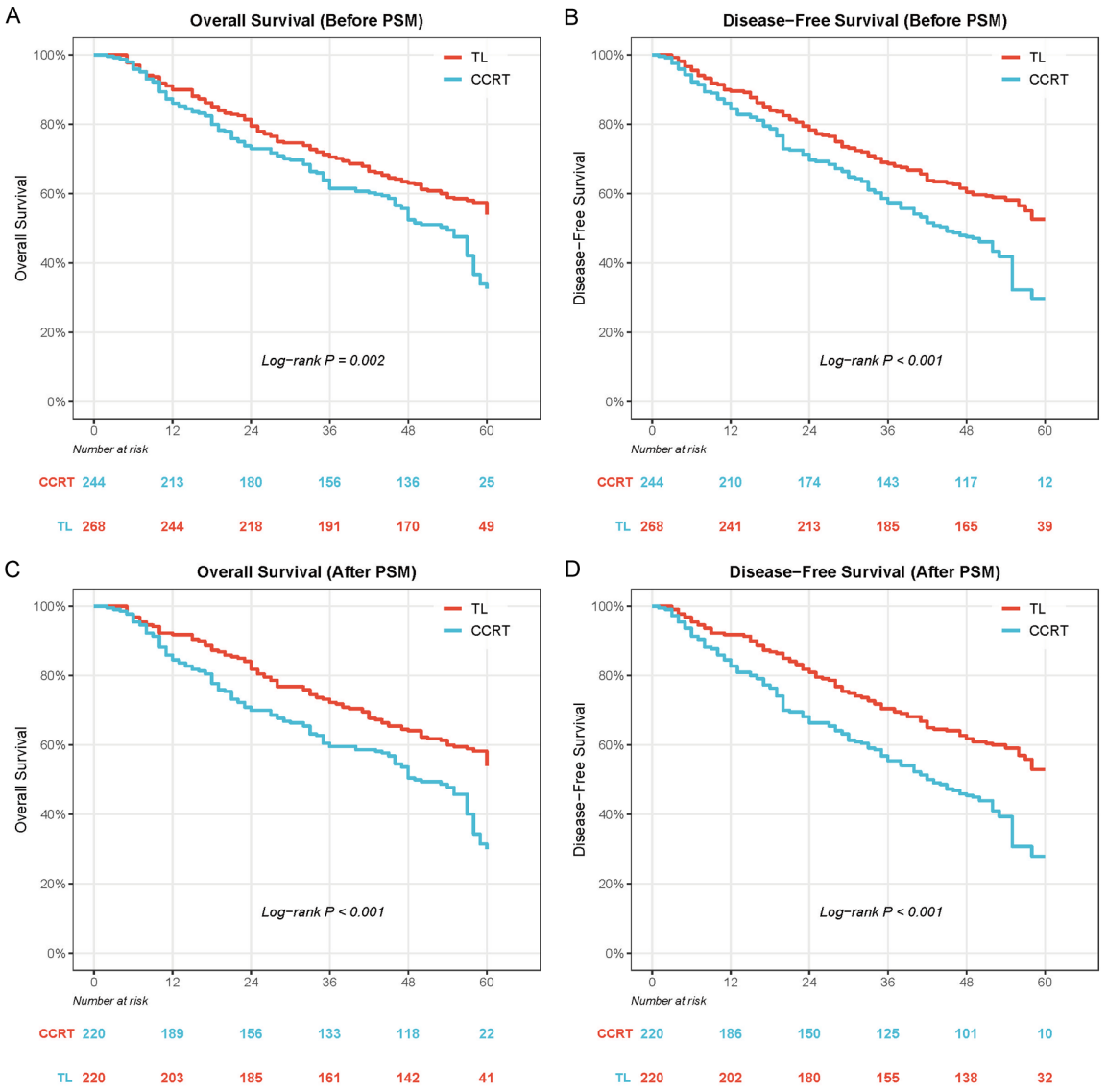


**Figure 3.** Kaplan-Meier survival curves before and after PSM. A. OS curves for the CCRT and TL groups before PSM. B. DFS curves for the CCRT and TL groups before PSM. C. OS curves for the CCRT and TL groups after PSM. D. DFS curves for the CCRT and TL groups after PSM. Note: CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; PSM, Propensity score matching; OS, Overall survival; DFS, Disease-free survival.

tural relationship inherent to the same categorical variable. The correlation pattern remained essentially unchanged after matching.

GVIF analysis demonstrated that all variables had GVIF^(1/(2×Df)) values well below the threshold of 2 before matching, with the highest being 1.054 for thyroid cartilage invasion. After matching, all values remained well below 2, with the maximum again observed for thyroid cartilage invasion (1.074). GVIF estimates were stable between the pre-match and post-match datasets, with no notable changes. These results confirmed the absence of serious multicollinearity and supported the simultaneous inclusion of all 10 variables in the multivariable Cox model (**Figures 4** and **5**).

*PH assumption testing*

Before fitting the multivariable Cox models, the proportional hazards (PH) assumption was assessed for the full 10-variable model using Schoenfeld residuals. In all four settings, the global test indicated violation of the PH assumption, with global *P* values of 0.004 for pre-

**Table 3.** Univariate Cox regression analysis for OS and DFS before propensity score matching

| Variable | OS | | | DFS | | |
|---|---|---|---|---|---|---|
| | β | *P* value | HR_(95% CI) | β | *P* value | HR_(95% CI) |
| Treatment | | | | | | |
| TL (Ref) | | | | | | |
| CCRT | 0.404 | 0.002 | 1.497 (1.165-1.924) | 0.519 | < 0.001 | 1.680 (1.315-2.148) |
| Age | | | | | | |
| < 65 years (Ref) | | | | | | |
| ≥ 65 years | 0.023 | 0.854 | 1.024 (0.797-1.315) | 0.015 | 0.900 | 1.016 (0.798-1.293) |
| Sex | | | | | | |
| Male (Ref) | | | | | | |
| Female | 0.042 | 0.843 | 1.043 (0.690-1.576) | 0.063 | 0.757 | 1.065 (0.715-1.585) |
| BMI category | | | | | | |
| 18.5-23.9 (Ref) | | | | | | |
| < 18.5 | 0.268 | 0.188 | 1.307 (0.877-1.946) | 0.243 | 0.223 | 1.275 (0.863-1.884) |
| ≥ 24.0 | 0.250 | 0.067 | 1.283 (0.982-1.676) | 0.287 | 0.028 | 1.333 (1.031-1.723) |
| Smoking history | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.194 | 0.186 | 0.824 (0.618-1.098) | -0.207 | 0.141 | 0.813 (0.616-1.071) |
| Alcohol history | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.014 | 0.911 | 1.014 (0.791-1.301) | 0.041 | 0.736 | 1.042 (0.819-1.325) |
| ECOG performance status | | | | | | |
| 0 (Ref) | | | | | | |
| 1 | 0.235 | 0.098 | 1.265 (0.958-1.670) | 0.209 | 0.123 | 1.233 (0.945-1.609) |
| 2 | 0.772 | < 0.001 | 2.163 (1.532-3.054) | 0.694 | < 0.001 | 2.002 (1.430-2.803) |
| Hypertension | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.141 | 0.280 | 1.151 (0.892-1.486) | 0.106 | 0.399 | 1.112 (0.869-1.424) |
| Diabetes mellitus | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.098 | 0.531 | 0.906 (0.666-1.233) | -0.025 | 0.866 | 0.975 (0.729-1.304) |
| CHD | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.079 | 0.670 | 1.082 (0.754-1.553) | 0.096 | 0.590 | 1.100 (0.778-1.557) |
| COPD | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.113 | 0.529 | 0.893 (0.627-1.271) | -0.148 | 0.399 | 0.863 (0.612-1.216) |
| T stage | | | | | | |
| T3 (Ref) | | | | | | |
| T4a | 0.584 | < 0.001 | 1.793 (1.399-2.299) | 0.529 | < 0.001 | 1.697 (1.335-2.157) |
| N stage | | | | | | |
| N0 (Ref) | | | | | | |
| N1 | 0.035 | 0.825 | 1.036 (0.760-1.411) | 0.041 | 0.786 | 1.041 (0.777-1.395) |
| N2 | 1.013 | < 0.001 | 2.753 (2.046-3.706) | 0.922 | < 0.001 | 2.515 (1.883-3.358) |
| Clinical stage | | | | | | |
| III (Ref) | | | | | | |
| IVA | 0.974 | < 0.001 | 2.649 (2.036-3.447) | 0.781 | < 0.001 | 2.183 (1.705-2.794) |
| Tumor subsite | | | | | | |
| Glottic (Ref) | | | | | | |
| Supraglottic | 0.066 | 0.603 | 1.069 (0.832-1.372) | -0.011 | 0.931 | 0.989 (0.776-1.261) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Tumor differentiation | | | | | | |
| Well (Ref) | | | | | | |
| Moderate | 0.147 | 0.442 | 1.158 (0.796-1.684) | 0.209 | 0.265 | 1.232 (0.854-1.779) |
| Poor | 0.358 | 0.084 | 1.430 (0.953-2.144) | 0.401 | 0.048 | 1.493 (1.003-2.221) |
| Vocal cord fixation | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.218 | 0.085 | 1.244 (0.970-1.594) | 0.188 | 0.124 | 1.207 (0.950-1.533) |
| Airway obstruction | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.052 | 0.725 | 1.053 (0.790-1.403) | 0.097 | 0.486 | 1.102 (0.838-1.449) |
| Thyroid cartilage invasion | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.412 | 0.002 | 1.509 (1.168-1.951) | 0.341 | 0.007 | 1.407 (1.096-1.806) |
| Pre-epiglottic space invasion | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.171 | 0.246 | 0.843 (0.632-1.125) | -0.125 | 0.372 | 0.882 (0.670-1.161) |
| Pretreatment Hb (g/L) | -0.003 | 0.386 | 0.997 (0.990-1.004) | -0.003 | 0.421 | 0.997 (0.990-1.004) |
| Pretreatment Alb (g/L) | 0.000 | 0.970 | 1.000 (0.975-1.026) | -0.004 | 0.763 | 0.996 (0.972-1.021) |
| Pretreatment NLR | 0.227 | < 0.001 | 1.255 (1.143-1.379) | 0.152 | 0.001 | 1.164 (1.063-1.275) |

Note: HR, Hazard ratio; CI, Confidence interval; Ref, Reference category; OS, Overall survival; DFS, Disease-free survival; CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; BMI, Body mass index; ECOG, Eastern Cooperative Oncology Group; CHD, Coronary heart disease; COPD, Chronic obstructive pulmonary disease; NLR, Neutrophil-to-lymphocyte ratio. β represents the regression coefficient. P < 0.05 is considered statistically significant.

**Table 4.** Univariate Cox regression analysis for OS and DFS after propensity score matching

| Variable | OS | | | DFS | | |
|---|---|---|---|---|---|---|
| | β | *P* value | HR (95% CI) | β | *P* value | HR (95% CI) |
| Treatment | | | | | | |
| TL (Ref) | | | | | | |
| CCRT | 0.514 | < 0.001 | 1.672 (1.276-2.190) | 0.631 | < 0.001 | 1.879 (1.443-2.448) |
| Age | | | | | | |
| < 65 years (Ref) | | | | | | |
| ≥ 65 years | -0.037 | 0.789 | 0.964 (0.737-1.261) | -0.061 | 0.643 | 0.941 (0.726-1.219) |
| Sex | | | | | | |
| Male (Ref) | | | | | | |
| Female | 0.152 | 0.491 | 1.164 (0.756-1.793) | 0.186 | 0.379 | 1.205 (0.796-1.823) |
| BMI category | | | | | | |
| 18.5-23.9 (Ref) | | | | | | |
| < 18.5 | 0.298 | 0.159 | 1.347 (0.890-2.040) | 0.277 | 0.181 | 1.319 (0.879-1.980) |
| ≥ 24.0 | 0.316 | 0.030 | 1.372 (1.030-1.828) | 0.362 | 0.010 | 1.437 (1.092-1.892) |
| Smoking history | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.175 | 0.264 | 0.839 (0.618-1.141) | -0.203 | 0.177 | 0.816 (0.608-1.096) |
| Alcohol history | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.008 | 0.956 | 0.993 (0.762-1.293) | 0.049 | 0.706 | 1.050 (0.813-1.356) |
| ECOG performance status | | | | | | |
| 0 (Ref) | | | | | | |
| 1 | 0.220 | 0.142 | 1.247 (0.929-1.674) | 0.195 | 0.175 | 1.216 (0.917-1.611) |
| 2 | 0.682 | < 0.001 | 1.978 (1.365-2.868) | 0.608 | < 0.001 | 1.838 (1.279-2.640) |
| Hypertension | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.116 | 0.404 | 1.123 (0.855-1.475) | 0.074 | 0.585 | 1.076 (0.826-1.402) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Diabetes mellitus | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.177 | 0.285 | 0.838 (0.605-1.159) | -0.092 | 0.556 | 0.912 (0.672-1.238) |
| CHD | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.055 | 0.785 | 1.057 (0.711-1.571) | 0.087 | 0.653 | 1.090 (0.747-1.591) |
| COPD | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.100 | 0.592 | 0.905 (0.628-1.304) | -0.132 | 0.465 | 0.876 (0.615-1.249) |
| T stage | | | | | | |
| T3 (Ref) | | | | | | |
| T4a | 0.593 | < 0.001 | 1.810 (1.388-2.359) | 0.538 | < 0.001 | 1.713 (1.326-2.213) |
| N stage | | | | | | |
| N0 (Ref) | | | | | | |
| N1 | 0.024 | 0.886 | 1.024 (0.739-1.419) | 0.031 | 0.841 | 1.032 (0.758-1.405) |
| N2 | 1.009 | < 0.001 | 2.743 (1.996-3.769) | 0.923 | < 0.001 | 2.516 (1.845-3.431) |
| Clinical stage | | | | | | |
| III (Ref) | | | | | | |
| IVA | 1.007 | < 0.001 | 2.737 (2.075-3.609) | 0.826 | < 0.001 | 2.284 (1.760-2.964) |
| Tumor subsite | | | | | | |
| Glottic (Ref) | | | | | | |
| Supraglottic | 0.027 | 0.842 | 1.028 (0.786-1.343) | -0.043 | 0.744 | 0.958 (0.739-1.241) |
| Tumor differentiation | | | | | | |
| Well (Ref) | | | | | | |
| Moderate | 0.190 | 0.358 | 1.209 (0.806-1.814) | 0.249 | 0.218 | 1.283 (0.863-1.908) |
| Poor | 0.401 | 0.069 | 1.494 (0.969-2.303) | 0.447 | 0.039 | 1.563 (1.024-2.388) |
| Vocal cord fixation | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.206 | 0.127 | 1.229 (0.943-1.601) | 0.180 | 0.166 | 1.197 (0.928-1.545) |
| Airway obstruction | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.032 | 0.841 | 1.033 (0.753-1.416) | 0.086 | 0.575 | 1.090 (0.807-1.470) |
| Thyroid cartilage invasion | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.500 | < 0.001 | 1.648 (1.251-2.172) | 0.443 | 0.001 | 1.557 (1.191-2.035) |
| Pre-epiglottic space invasion | | | | | | |
| No (Ref) | | | | | | |
| Yes | -0.234 | 0.135 | 0.791 (0.582-1.076) | -0.196 | 0.189 | 0.822 (0.613-1.101) |
| Pretreatment Hb (g/L) | -0.002 | 0.545 | 0.998 (0.990-1.005) | -0.002 | 0.617 | 0.998 (0.991-1.005) |
| Pretreatment Alb (g/L) | -0.003 | 0.806 | 0.997 (0.970-1.024) | -0.006 | 0.671 | 0.994 (0.969-1.020) |
| Pretreatment NLR | 0.228 | < 0.001 | 1.256 (1.137-1.386) | 0.150 | 0.002 | 1.161 (1.055-1.278) |

Note: HR, Hazard ratio; CI, Confidence interval; Ref, Reference category; OS, Overall survival; DFS, Disease-free survival; CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; BMI, Body mass index; ECOG, Eastern Cooperative Oncology Group; CHD, Coronary heart disease; COPD, Chronic obstructive pulmonary disease; NLR, Neutrophil-to-lymphocyte ratio. β represents the regression coefficient. P < 0.05 is considered statistically significant.

match OS, < 0.001 for pre-match DFS, 0.021 for post-match OS, and < 0.001 for post-match DFS. At the individual-variable level, ECOG performance status, thyroid cartilage invasion, and NLR consistently showed evidence of non-proportionality across multiple models. In addition, T stage and N stage also violated the PH assumption in the pre-match DFS model, and N stage violated the assumption in the post-match DFS model. Therefore, ECOG performance status, thyroid cartilage invasion, pre-epiglottic space invasion, and NLR were removed, yielding a final model including six variables: treatment modality, BMI category, T

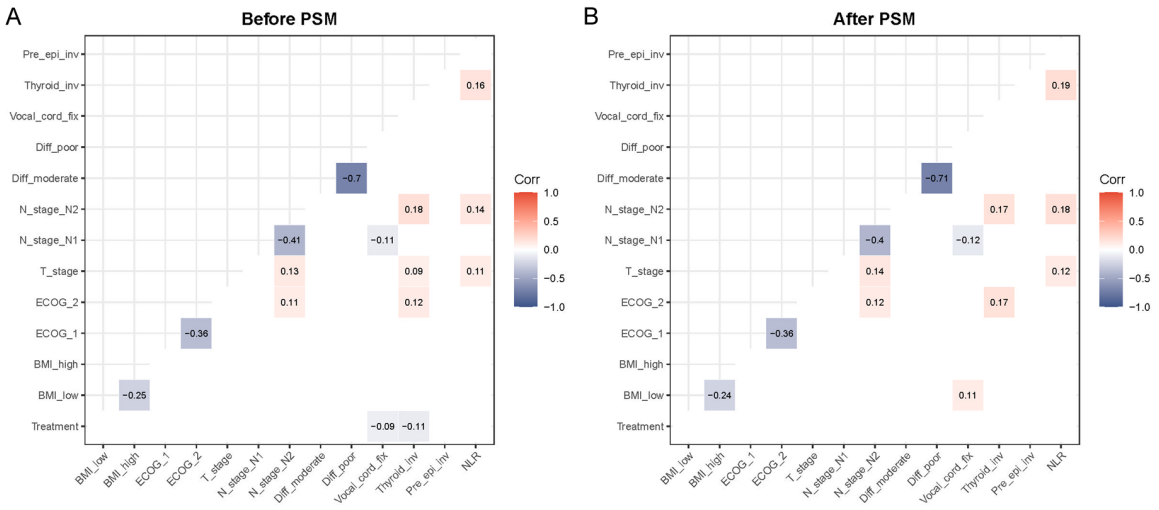


**Figure 4.** Spearman correlation heatmaps of variables with univariate significance (P < 0.1) before and after PSM. A. Correlation matrix before PSM. B. Correlation matrix after PSM. Note: Color intensity and numerical values represent Spearman's rank correlation coefficients (r). Red/blue shades indicate positive/negative correlations, respectively. BMI, Body mass index; ECOG, Eastern Cooperative Oncology Group; NLR, Neutrophil-to-lymphocyte ratio; PSM, Propensity score matching.

stage, N stage, tumor differentiation, and vocal cord fixation. In the final six-variable model, the global PH assumption was satisfied for pre-match OS (global P = 0.248), post-match OS (global P = 0.329), and post-match DFS (global P = 0.055), whereas the pre-match DFS model remained borderline (P = 0.032). At the individual-variable level, N stage showed mild deviation in the pre-match OS model (P = 0.048), T stage and N stage showed mild deviation in the pre-match DFS model (P = 0.028 and P = 0.008, respectively), and N stage showed mild deviation in the post-match DFS model (P = 0.014). However, given their clinical relevance and the limited degree of global violation, these variables were retained in the final analyses. Overall, the final six-variable model was considered acceptable for multivariable Cox regression analysis (**Figure 6**).

*Multivariable Cox regression before and after matching*

In the pre-match multivariable Cox analysis, after adjusting for BMI category, T stage, N stage, tumor differentiation, and vocal cord fixation, CCRT was confirmed as an independent risk factor for both OS (P < 0.001) and DFS (P < 0.001). T4a stage (OS: P < 0.001; DFS: P < 0.001) and N2 stage (OS: P < 0.001; DFS: P < 0.001) were also independent risk factors for both endpoints. N1 stage was not significant for either OS (P = 0.833) or DFS (P = 0.809). BMI categories (underweight: OS P = 0.224, DFS P = 0.299; overweight/obese: OS P = 0.293, DFS P = 0.119), tumor differentiation levels (moderate: OS P = 0.920, DFS P = 0.499; poor: OS P = 0.269, DFS P = 0.105), and vocal cord fixation (OS P = 0.165, DFS P = 0.192) all failed to reach statistical significance after adjustment (**Table 5**).

Post-match results were highly consistent with the pre-match findings. CCRT remained an independent risk factor for OS (P < 0.001) and DFS (P < 0.001), with HR estimates slightly higher than those observed before matching. T4a stage (OS: P < 0.001; DFS: P < 0.001) and N2 stage (OS: P < 0.001; DFS: P < 0.001) continued to independently predict worse OS and DFS. One difference from the pre-match model was that BMI ≥ 24.0 reached significance for DFS (P = 0.048) in the post-match analysis, although it remained non-significant for OS (P = 0.165). N1 stage (OS P = 0.928, DFS P = 0.919), underweight BMI (OS P = 0.143, DFS P = 0.175), tumor differentiation levels (moderate: OS P = 0.855, DFS P = 0.469; poor: OS P = 0.285, DFS P = 0.110), and vocal cord fixation (OS P = 0.312, DFS P = 0.400) were all non-significant in the post-match model (**Table 6**).

*Subgroup analysis and interaction tests*

To further investigate whether treatment effectiveness varied across clinical subgroups, the

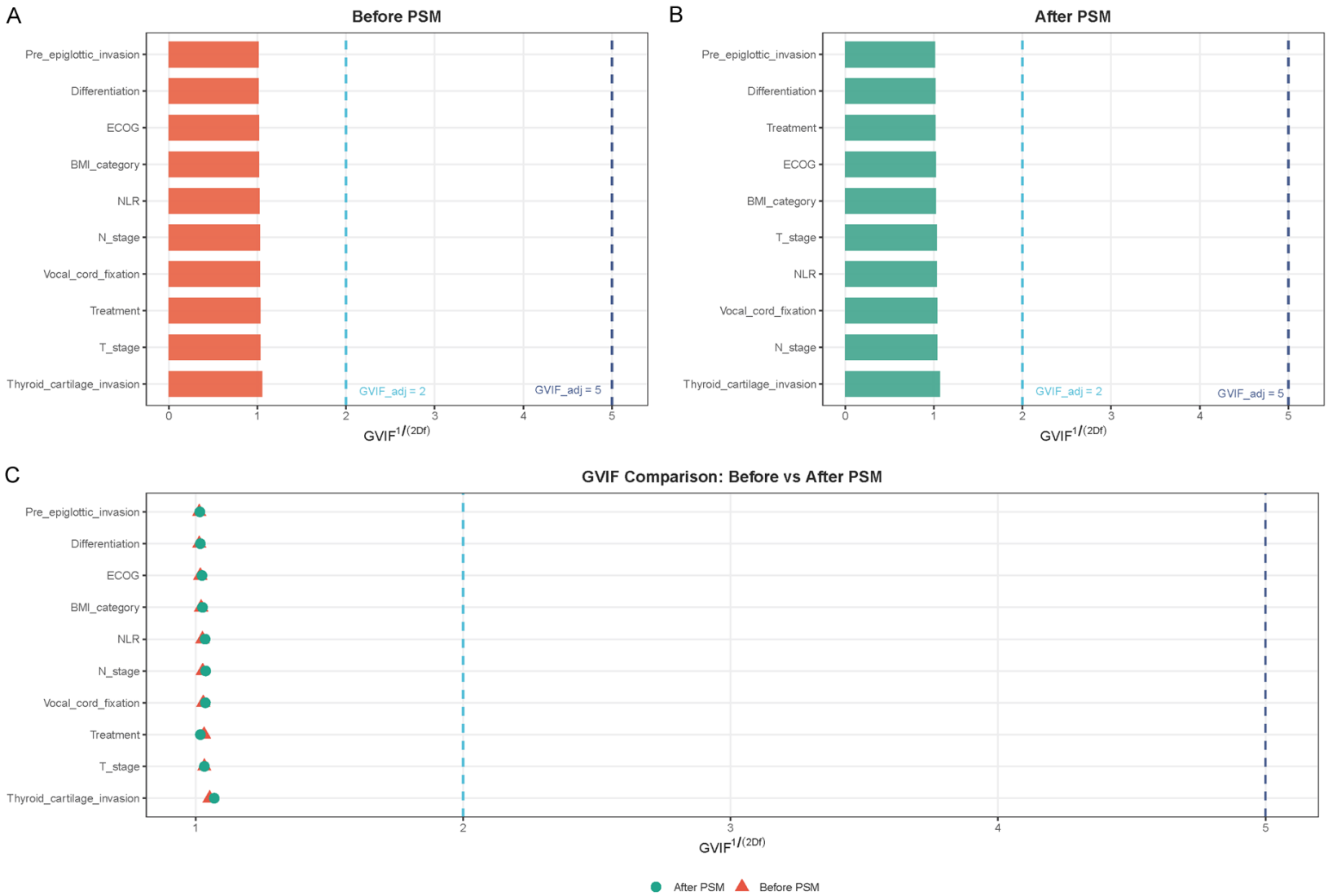


**Figure 5.** Variance inflation factor analysis of variables included in the multivariable Cox model before and after PSM. A. GVIF^(1/(2×Df)) values for each variable before PSM. The dashed vertical line indicates the threshold of 2, above which multicollinearity may be concerning. B. GVIF^(1/(2×Df)) values for each variable after PSM. C. Dumbbell plot comparing GVIF^(1/(2×Df)) values before and after PSM, with points connected by lines to visualize changes in collinearity. Note: GVIF, Generalized variance inflation factor; PSM, Propensity score matching; BMI, Body mass index; ECOG, Eastern Cooperative Oncology Group; NLR, Neutrophil-to-lymphocyte ratio. GVIF^(1/(2×Df)) values < 2 indicate no serious multicollinearity.

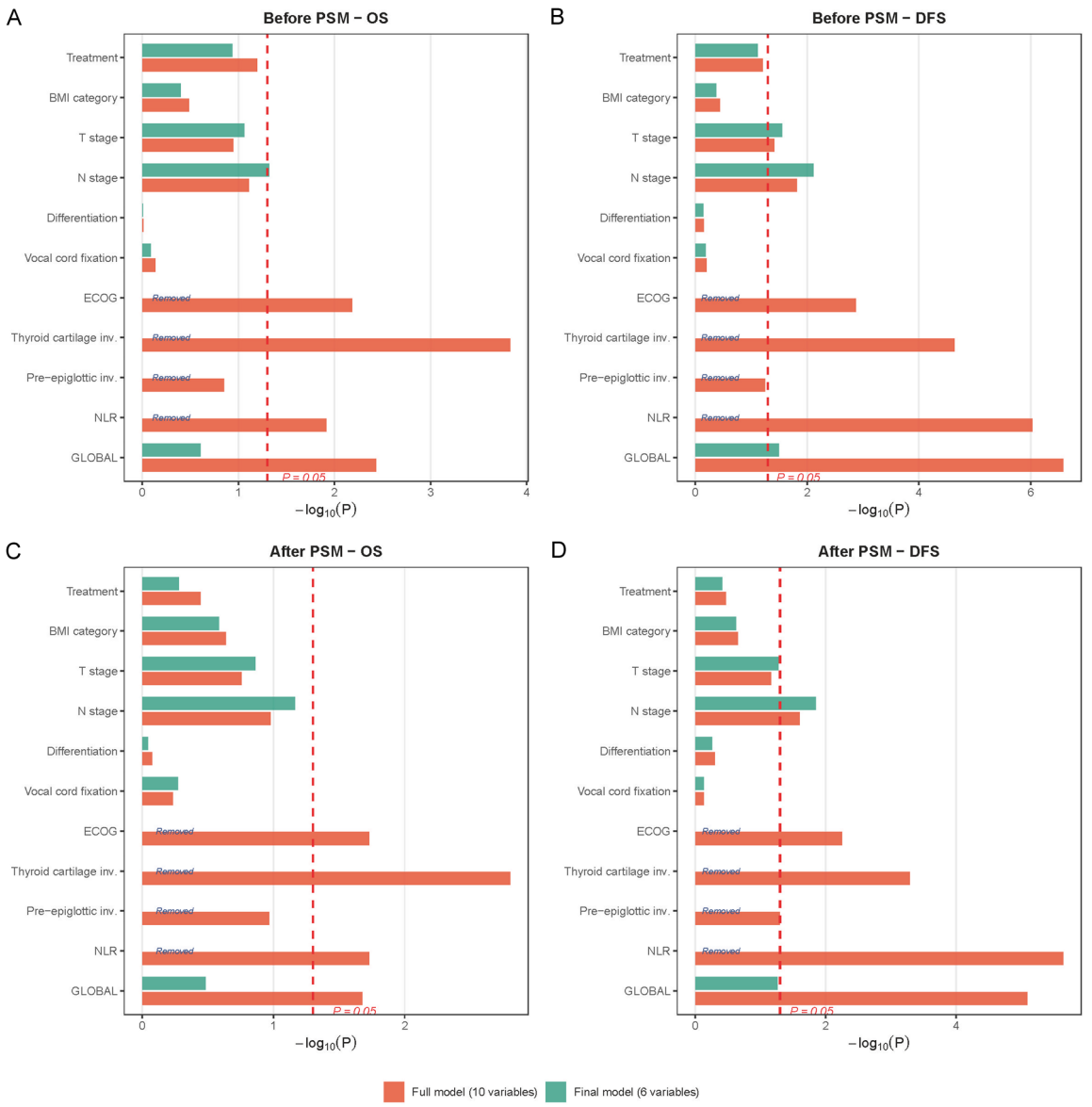


**Figure 6.** Comparison of PH assumption test *P* values between the full model (10 variables) and the final model (6 variables). A. PH test *P* values for the full and final models for pre-match OS. B. PH test *P* values for the full and final models for pre-match DFS. C. PH test *P* values for the full and final models for post-match OS. D. PH test *P* values for the full and final models for post-match DFS. Note: The dashed horizontal line indicates the significance threshold (P = 0.05). Values below the line indicate violation of the PH assumption. The final model (6 variables: treatment, BMI category, T stage, N stage, tumor differentiation, and vocal cord fixation) showed improved PH assumption adherence compared with the full model (10 variables), which included ECOG performance status, thyroid cartilage invasion, pre-epiglottic space invasion, and NLR. PH, Proportional hazards; OS, Overall survival; DFS, Disease-free survival; PSM, Propensity score matching; ECOG, Eastern Cooperative Oncology Group; NLR, Neutrophil-to-lymphocyte ratio; BMI, Body mass index.

440 matched patients were stratified by T stage, N stage, age, ECOG performance status, tumor differentiation, vocal cord fixation, thyroid cartilage invasion, and tumor subsite.

For OS, CCRT was associated with significantly worse outcomes in most subgroups, including T3 (P = 0.002), T4a (P = 0.018), N1-2 (P < 0.001), age <65 years (P = 0.002), age ≥ 65 years (P = 0.043), ECOG 1-2 (P < 0.001), well/moderately differentiated tumors (P = 0.007), poorly differentiated tumors (P = 0.004), no vocal cord fixation (P < 0.001), no thyroid cartilage invasion (P = 0.013), thyroid cartilage inva-

**Table 5.** Multivariable Cox regression analysis for OS and DFS before propensity score matching

| Variable | OS | | | DFS | | |
|---|---|---|---|---|---|---|
| | β | *P* value | HR (95% CI) | β | *P* value | HR (95% CI) |
| Treatment | | | | | | |
| TL (Ref) | | | | | | |
| CCRT | 0.564 | < 0.001 | 1.758 (1.363-2.266) | 0.692 | < 0.001 | 1.998 (1.555-2.566) |
| BMI category | | | | | | |
| 18.5-23.9 (Ref) | | | | | | |
| < 18.5 | 0.251 | 0.224 | 1.285 (0.858-1.924) | 0.210 | 0.299 | 1.234 (0.830-1.832) |
| ≥ 24.0 | 0.145 | 0.293 | 1.156 (0.882-1.513) | 0.206 | 0.119 | 1.228 (0.948-1.591) |
| T stage | | | | | | |
| T3 (Ref) | | | | | | |
| T4a | 0.541 | < 0.001 | 1.718 (1.331-2.217) | 0.500 | < 0.001 | 1.648 (1.289-2.108) |
| N stage | | | | | | |
| N0 (Ref) | | | | | | |
| N1 | 0.034 | 0.833 | 1.034 (0.757-1.413) | 0.036 | 0.809 | 1.037 (0.772-1.393) |
| N2 | 0.964 | < 0.001 | 2.622 (1.932-3.557) | 0.891 | < 0.001 | 2.438 (1.811-3.282) |
| Tumor differentiation | | | | | | |
| Well (Ref) | | | | | | |
| Moderate | 0.019 | 0.920 | 1.020 (0.698-1.490) | 0.128 | 0.499 | 1.137 (0.784-1.648) |
| Poor | 0.232 | 0.269 | 1.261 (0.835-1.904) | 0.334 | 0.105 | 1.397 (0.933-2.090) |
| Vocal cord fixation | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.179 | 0.165 | 1.196 (0.929-1.539) | 0.161 | 0.192 | 1.175 (0.922-1.498) |

Note: HR, Hazard ratio; CI, Confidence interval; Ref, Reference category; OS, Overall survival; DFS, Disease-free survival; CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; BMI, Body mass index. β represents the regression coefficient. P < 0.05 is considered statistically significant.

sion present (P = 0.002), glottic subsite (P = 0.006), and supraglottic subsite (P = 0.012). No significant difference in OS between CCRT and TL was observed in the N0 subgroup (P = 0.127), the ECOG 0 subgroup (P = 0.385), or the vocal cord fixation-present subgroup (P = 0.234). ECOG performance status and vocal cord fixation were identified as significant effect modifiers. Specifically, the OS disadvantage of CCRT compared with TL was similar among patients with ECOG score 1-2 and those without vocal cord fixation.

For DFS, CCRT was significantly inferior to TL across all subgroups (all P < 0.05). A significant interaction was observed between treatment and vocal cord fixation (P_interaction = 0.018): the DFS disadvantage of CCRT was more prominent in patients without vocal cord fixation, whereas in the subgroup with fixation, the difference, although present, was not significant (P = 0.086). All other subgroup variables showed non-significant interaction effects (P_interaction > 0.05). The subgroup results before matching were largely consistent with the post-match findings (**Figure 7**).

*Sensitivity analyses*

To validate the robustness of the survival difference between CCRT and TL, eight statistical approaches were performed: (1) pre-match unadjusted model; (2) pre-match 10-variable adjusted model; (3) pre-match 6-variable adjusted model; (4) post-match unadjusted model; (5) post-match 6-variable adjusted model (primary analysis); (6) IPTW; (7) propensity score stratification by quintiles; and (8) propensity score covariate adjustment. For OS, the HR estimates across the eight methods ranged from 1.50 to 1.89. All approaches demonstrated significantly worse OS in the CCRT group compared with the TL group (all P < 0.05), with the primary analysis yielding an HR of 1.82. For DFS, the HR estimates ranged from 1.68 to 2.13, and all methods consistently showed significantly inferior DFS in the CCRT group (all P < 0.001), with the primary analysis

**Table 6.** Multivariable Cox regression analysis for OS and DFS after propensity score matching

| Variable | OS | | | DFS | | |
|---|---|---|---|---|---|---|
| | β | *P* value | HR (95% CI) | β | *P* value | HR (95% CI) |
| Treatment | | | | | | |
| TL (Ref) | | | | | | |
| CCRT | 0.598 | < 0.001 | 1.819 (1.386-2.386) | 0.738 | < 0.001 | 2.091 (1.601-2.732) |
| BMI category | | | | | | |
| 18.5-23.9 (Ref) | | | | | | |
| < 18.5 | 0.316 | 0.143 | 1.372 (0.898-2.094) | 0.287 | 0.175 | 1.332 (0.880-2.016) |
| ≥ 24.0 | 0.205 | 0.165 | 1.228 (0.919-1.640) | 0.281 | 0.048 | 1.324 (1.003-1.749) |
| T stage | | | | | | |
| T3 (Ref) | | | | | | |
| T4a | 0.518 | < 0.001 | 1.678 (1.276-2.207) | 0.475 | < 0.001 | 1.607 (1.234-2.094) |
| N stage | | | | | | |
| N0 (Ref) | | | | | | |
| N1 | 0.015 | 0.928 | 1.015 (0.730-1.412) | 0.016 | 0.919 | 1.016 (0.744-1.388) |
| N2 | 0.919 | < 0.001 | 2.508 (1.807-3.482) | 0.847 | < 0.001 | 2.332 (1.693-3.213) |
| Tumor differentiation | | | | | | |
| Well (Ref) | | | | | | |
| Moderate | 0.038 | 0.855 | 1.039 (0.689-1.568) | 0.148 | 0.469 | 1.160 (0.776-1.734) |
| Poor | 0.241 | 0.285 | 1.272 (0.818-1.979) | 0.352 | 0.110 | 1.422 (0.924-2.190) |
| Vocal cord fixation | | | | | | |
| No (Ref) | | | | | | |
| Yes | 0.139 | 0.312 | 1.149 (0.878-1.505) | 0.111 | 0.400 | 1.118 (0.862-1.449) |

Note: Ref, Reference category; OS, Overall survival; DFS, Disease-free survival; CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; BMI, Body mass index; HR, Hazard ratio; CI, Confidence interval. β represents the regression coefficient. P < 0.05 is considered statistically significant.

HR equal to 2.09. Most confounding-control strategies adopted resulted in HRs for CCRT versus TL greater than 1 and statistically significant. There was agreement in both the direction and magnitude of the HR estimates across methods. Together, these findings consistently indicate that CCRT is associated with worse survival outcomes than TL across various analytical approaches (**Figure 8**).

*Supplementary analyses of treatment regimens and postoperative adjuvant therapy*

To further characterize treatment heterogeneity between the two groups, treatment regimens in the CCRT group were summarized, and postoperative adjuvant therapy in the TL group was illustrated. Table S1 shows that among CCRT patients, the majority received the standard cisplatin regimen (80-100 mg/m$^2$ every three weeks), accounting for 80.3% and 80.0% of patients before and after PSM, respectively. The remaining patients received cisplatin-containing alternative regimens, including carboplatin, nedaplatin, or reduced-dose cisplatin. Most patients completed three cycles of chemotherapy (72.1% before PSM and 73.2% after PSM), and over 89% completed the prescribed radiotherapy dose (≥ 66 Gy).

The distribution of postoperative adjuvant therapy in the TL group is presented in Table S2. The majority of TL patients received postoperative radiotherapy or chemotherapy, whereas a small proportion underwent surgery alone (6.7% before PSM and 7.3% after PSM).

To evaluate whether postoperative adjuvant therapy in the TL group influenced the main findings, we conducted a sensitivity analysis restricted to TL patients treated with surgery alone and compared them with CCRT patients. As shown in Table S3, the results remained consistent with the main analysis. CCRT was associated with significantly worse OS and DFS compared with TL even under this restricted comparison, both before and after PSM.

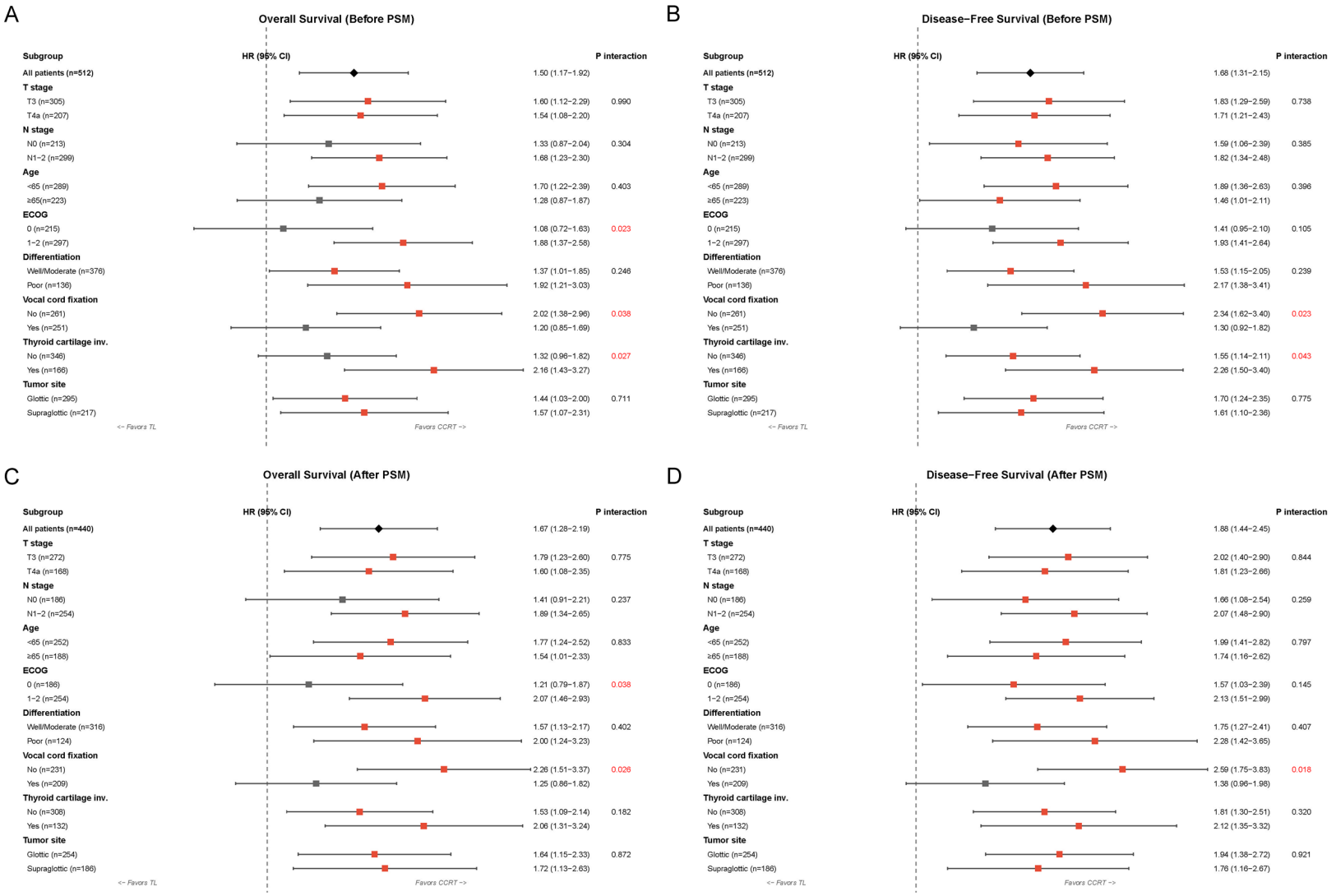


**Figure 7.** Forest plots of subgroup analysis for OS and DFS before and after PSM. A. Forest plot of OS HRs and 95% CIs by subgroup before PSM. B. Forest plot of DFS HRs and 95% CIs by subgroup before PSM. C. Forest plot of OS HRs and 95% CIs by subgroup after PSM. D. Forest plot of DFS HRs and 95% CIs by subgroup after PSM. Note: OS, Overall survival; DFS, Disease-free survival; PSM, Propensity score matching; CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; HR, Hazard ratio; CI, Confidence interval; ECOG, Eastern Cooperative Oncology Group.

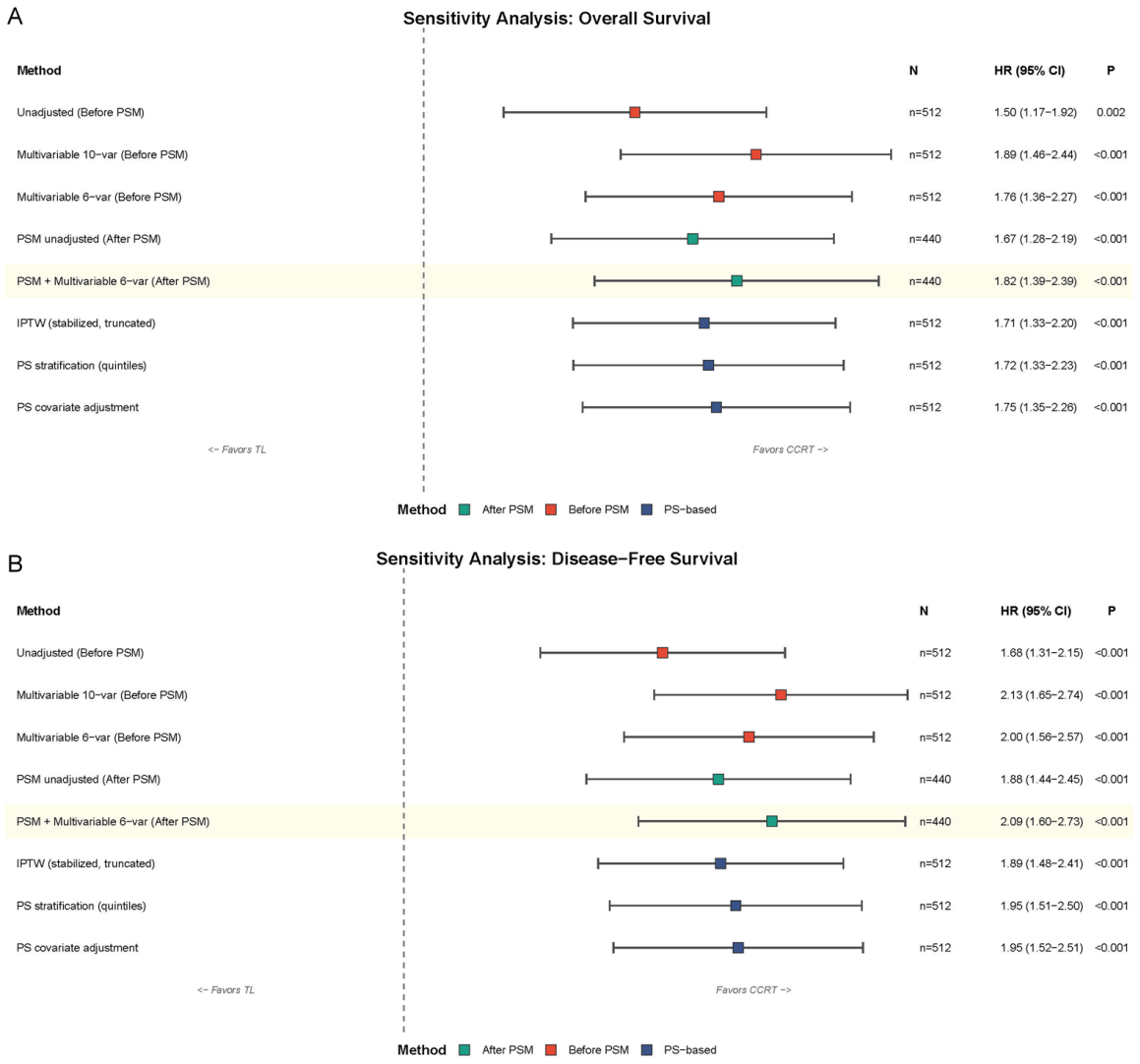


**Figure 8.** Forest plots of sensitivity analysis for OS and DFS using multiple statistical methods. A. Comparison of OS HRs and 95% CIs for CCRT versus TL across eight statistical methods. B. Comparison of DFS HRs and 95% CIs for CCRT versus TL across eight statistical methods. The yellow-highlighted row indicates the primary analysis. Note: OS, Overall survival; DFS, Disease-free survival; PSM, Propensity score matching; CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; IPTW, Inverse probability of treatment weighting; PS, Propensity score; HR, Hazard ratio; CI, Confidence interval.

## Discussion

The present study makes several methodological contributions to the ongoing debate over treatment selection for LA-LSCC. By combining PSM with eight complementary sensitivity approaches - including IPTW, propensity score stratification, and propensity score covariate adjustment - we achieved a level of confounding control that is uncommon in single-institution retrospective studies of this condition. The high concordance of effect estimates across all analytical strategies strengthens confidence in the direction and approximate magnitude of the observed survival difference between TL and CCRT. To our knowledge, PSM-based head-to-head comparisons of CCRT versus TL specifically addressing the high-risk LA-LSCC population remain limited in the domestic literature, and the present work provides complementary real-world evidence on this clinically consequential question.

The historical context of organ preservation must be considered when interpreting our findings. The Veterans Affairs Laryngeal Cancer

Study and the long-term results of RTOG 91-11 established CCRT as a mainstay of larynx-preservation therapy, the their impact on clinical practice is unquestionable. Nonetheless, both trials predominantly enrolled patients with T3 disease and systematically excluded those with severe laryngeal dysfunction, airway compromise, or poor performance status. These patients represent the high-risk cohort that dominates the real-world locally advanced population. The clinical repercussions of this limited representativeness are significant, as any survival benefit demonstrated in these trials may not be applicable to LA-LSCC encountered in daily practice. Consequently, recent analyses have urged caution in applying findings from key trials to higher-risk groups. An international multidisciplinary Delphi consensus specifically advised against non-surgical organ preservation for patients with T4a disease, poor baseline swallowing or airway function, or significant comorbidity burden. The current study cohort included more patients with T4a tumors, vocal cord fixation, and thyroid cartilage invasion; thus, it is more representative of patients for whom the decision to preserve laryngeal functions poses the greatest difficulty. The survival disadvantage of CCRT observed here therefore has direct clinical applicability for this previously underrepresented population.

The survival advantage of TL over CCRT observed in our study is generally supported by large retrospective analyses, although differences across studies highlight the importance of patient selection and methodological rigor. According to Patel et al. [20], patients with T4 stage disease who underwent TL demonstrated significantly improved OS compared with non-T4 patients with low nodal burden in the National Cancer Database - a pattern similar to our own subgroup analyses. Moreover, their findings reinforce that the oncologic benefit of TL is not uniform across the locally advanced spectrum. Fu and co-authors [21] conducted a PSM-based National Cancer Database analysis restricted to cT4a patients, as previously mentioned, and reported similar results. This analysis involved 452 matched pairs and also favored surgery, but yielded a more modest HR than ours. This discrepancy likely reflects dilution of effects due to multi-institutional heterogeneity in administrative databases, as well as restriction to a single T stage. Merdad et al. [22] performed a systematic review and meta-analysis of T3 laryngeal cancer and found that TL was associated with better DFS and OS compared with organ preservation. Although the absolute differences were modest, these findings support individualized treatment decisions for this intermediate-risk group. In contrast, a meta-analysis by Rao et al. [23], which included over 10,000 T3 patients, did not find any significant difference in OS between TL and CCRT. Similarly, García-Cabo et al. [24] reported no difference in OS or disease-specific survival in a matched-pair study of hypopharyngeal cancers. These discrepancies can likely be explained by differences in inclusion criteria, follow-up duration, and, importantly, control for confounding by indication. Through comprehensive application of PSM and additional sensitivity analyses, we are able to reach a consistent conclusion. Although the methodological strengths of our study do not allow for causal inference, each analytical approach lends support to a similar directional effect favoring TL.

A methodologically instructive finding is that the survival advantage of TL became more pronounced rather than diminished after PSM. This trend has important implications for interpreting the existing unmatched literature. Before matching, the TL group was enriched for patients with stage IVA disease, vocal cord fixation, and thyroid cartilage invasion - characteristics that simultaneously drove the choice of surgery and predicted poor prognosis. This classic “indication bias” created negative confounding, which compressed the true survival difference between the two treatment options. After PSM rectified these structural imbalances, the treatment effect under assessment became more estimable. The key takeaway is that surgical versus non-surgical comparisons in LA-LSCC that are inadequately adjusted - or left unadjusted - are likely to underestimate the oncologic benefit of TL, particularly in series where surgeons selectively operated on more advanced cases. Although PSM addresses certain confounding factors (e.g., chemoradiotherapy completion rates, radiation dose intensity, and institutional surgical volume), residual unmeasured confounding cannot be entirely excluded. Therefore, multicenter prospective studies are needed to further refine the effect estimates.

The subgroup and interaction analyses suggest that the findings are exploratory in nature and

may have clinical relevance, albeit with caution. These exploratory findings are hypothesis-generating and have not been adjusted for multiplicity. The notable interaction between ECOG performance status and treatment suggests that patients with poor functional capacity derive disproportionally less benefit and/or experience disproportionably more harm from CCRT compared with TL. Biologically and clinically, patients with lower functional reserve are generally less able to tolerate the acute and cumulative toxicities of cisplatin-based chemoradiotherapy; they may also have a higher frequency of treatment interruptions. A combined analysis of RTOG 9501 and the European Organisation for Research and Treatment of Cancer 22931 reported that although postoperative CCRT reduced tumor-specific mortality, it was associated with significantly higher non-cancer mortality [25]. This trade-off would be particularly unfavorable in patients already impaired at baseline. The relationship between vocal cord fixation and treatment benefit with respect to OS is more intricate and should not be overinterpreted. The observed paradox - that CCRT performed relatively better in the fixation-present subgroup - likely reflects residual selection bias, as CCRT-treated patients with vocal cord fixation were probably more stringently screened by the MDT. According to Nobacht et al. [11], oncologic outcomes between modalities begin to converge when treatment selection is based on expected laryngeal function rather than on T stage alone. This suggests that functional assessment must become an integral part of the selection process rather than an afterthought. Findings from all subgroups should be regarded as hypothesis-generating and require independent validation in adequately powered prospective cohorts.

Our results are consistent with, and provide real-world empirical support for, the current international redefinition of organ-preservation eligibility. The Delphi consensus published by Ferrari et al. [7], recommended shifting the paradigm from preservation of laryngeal anatomy to preservation of usable function - a clinically relevant distinction that is often underrecognized. A non-phonating, chronically aspirating, or permanently feeding tube-dependent anatomically intact larynx should not considered a successful functional preservation outcome. Patients undergoing organ-preserving treatment exhibited a high rate of subsequent TL requirement. Moreover, those with recurrent disease who did not undergo salvage surgery had poorer disease-specific survival than their counterparts who underwent organ sacrifice, as reported by Victor et al. [26]. Lee et al. [27] found that approximately 40% of patients who failed organ preservation were not eligible for salvage surgery, thereby completely closing the therapeutic window. Among those who did undergo salvage TL, prior radiotherapy substantially increased surgical morbidity: a multicenter study reported a pharyngocutaneous fistula rate of 25.4% after salvage laryngectomy [29], and lymphovascular invasion together with positive margins have been identified as independent adverse prognostic factors following salvage surgery [30]. The complication burden associated with salvage TL after prior radiotherapy has been well characterized [28], and the oncologic outcomes after salvage are substantially inferior to those achieved with upfront surgery in comparable patients. Collectively, these data suggest that non-surgical organ preservation in high-risk LA-LSCC patients carries compounding risk: a higher probability of local failure, a narrower salvage window, and greater surgical morbidity if salvage is attempted. The role of salvage laryngectomy in managing residual or recurrent disease after chemoradiotherapy has been reviewed in detail by Vander Poorten et al. [13], who highlighted that oncologic outcomes and prognostic factors after salvage surgery are markedly different from those in the primary surgical setting. Our study adds population-level survival evidence that should inform the initial risk stratification process within MDT deliberations.

Based on the interaction analyses and broader evidence synthesis, we propose a preliminary individualized decision-making framework articulated according to the mechanistic rationale for each risk stratum. Patients with T4a disease and thyroid cartilage invasion have a significantly increased risk of gross cartilage destruction, which reduces the likelihood of durable local control with radiation alone. When this is combined with N2 nodal disease or an ECOG performance status of 1 or greater, the cumulative disease burden and reduced treatment tolerance converge to make TL-based strategies the more defensible curative approach. The intermediate-risk category includes pa-

tients with T3 tumors with N1 disease or vocal cord fixation. Substantial uncertainty remains regarding the balance between oncologic control and functional preservation in this group. Therefore, appropriate MDT consideration, including laryngeal function assessment, is warranted. As suggested by Nobacht et al. [11] and Merdad et al. [22], functional prognosis - rather than T stage alone - should inform this decision. Patients with T3, N0, ECOG 0 disease and no cartilage invasion form the subgroup best suited for CCRT to achieve durable local control while preserving good laryngeal function, provided that careful monitoring and feasible salvage options are available. This concept of precision-guided treatment selection reflects a commitment to patient-centered care and is largely consistent with current NCCN guidelines [4, 5] and the National Cancer Institute's Physician Data Query recommendations [6], though not exclusively.

Although the present study primarily analyzed OS and DFS, a full assessment of treatment trade-offs in LA-LSCC requires consideration of QoL outcomes, which were not captured in our data. According to Murariu et al. [31], patients who underwent TL had significantly lower QoL scores at 12 months than those treated with CCRT. Thus, although surgery may confer survival benefits, it may also lead to significant reductions in patient-reported QoL, particularly in the domains of communication, social functioning, and body image. This trade-off does not negate the survival advantage of TL, but it reinforces that treatment decisions must be made jointly with the patient, based on realistic information regarding both oncologic prognosis and functional consequences. Preserving the larynx is a worthwhile goal for all patients who have a reasonable chance of achieving successful organ preservation without compromising oncologic outcomes. The danger lies in attempting preservation in patients for whom a good outcome is unlikely, or for whom the preserved larynx would be nonfunctional. Global data on cancer burden [1, 2] highlight the need for more optimized individualized treatments and reduced outcome disparities. Future studies investigating treatment selection for laryngeal squamous cell carcinoma should prioritize the incorporation of patient-reported outcome measures and more formal assessment of laryngeal function.

Several limitations warrant consideration. First, owing to the single-center retrospective design, PSM cannot account for unmeasured confounders such as patient treatment preferences (e.g., surgical vs. CCRT), surgeon experience, and variability in dose intensity or treatment completion rates of CCRT. Multi-center validation is necessary to establish external generalizability, particularly given differences in staging practices, peri-treatment management, and radiotherapy techniques across institutions. Second, the absence of a validated assessment scale for laryngeal function and patient-reported outcomes limits our ability to interpret survival data in the context of functional preservation. Third, both the CCRT and TL groups exhibited treatment heterogeneity, and the limited size of individual subgroups restricted the statistical power of the interaction analyses. Fourth, in the TL group, a small subset of patients underwent surgery alone without postoperative adjuvant therapy. These patients had highly selected pathological features, and all analyses involving this subgroup should be interpreted with caution regarding selection bias. Fifth, the ECOG performance status, thyroid cartilage invasion, and NLR were excluded from the multivariable model because they violated the PH assumption. Future studies employing time-varying coefficient Cox models, restricted mean survival time analysis, or piecewise Cox regression could recover the prognostic information contained in these variables across different follow-up periods. Sixth, our study period (2015-2022) predates the integration of immune checkpoint inhibitor-based combination strategies into the management of LA-LSCC. Consequently, the comparative merits of CCRT versus surgery are likely to evolve. Nevertheless, it remains difficult to predict how assertively these strategies will be adopted in this disease setting.

## Conclusion

This study provides evidence from a large real-world cohort that, after PSM and multiple robustness verification methods, TL was associated with improved OS and DFS compared with CCRT in patients with LA-LSCC. The results suggest that ECOG performance status and vocal cord fixation may influence the treatment effect. When making treatment decisions, careful consideration of essential patient

and tumor characteristics is necessary to select the most appropriate therapy and to prevent over-preservation of the larynx.

## Disclosure of conflict of interest

None.

**Address correspondence to:** Lili Chen, Department of Head and Neck Thyroid Surgery, Harbin Medical University Cancer Hospital, 150 Haping Road, Nangang District, Harbin 150081, Heilongjiang, China. E-mail: llchen@hrbmu.edu.cn

## References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74: 229-263.

[2] Jiang J, Xia Z and Yao W. Global, regional, and national larynx cancer burden and health inequality analysis from 1990 to 2021 with a prediction from 2022 to 2040. Front Oncol 2025; 15: 1617613.

[3] Mousavi SE, Ilaghi M, Aslani A, Najafi M, Yekta Z and Nejadghaderi SA. Laryngeal cancer incidence trends in the United States over 2000-2020: a population-based analysis. Arch Public Health 2024; 82: 106.

[4] Colevas AD, Cmelak AJ, Pfister DG, Spencer S, Adkins D, Birkeland AC, Brizel DM, Busse PM, Caudell JJ, Durm G, Fakhry C, Galloway T, Geiger JL, Gillison ML, Glastonbury C, Haddad RI, Hicks WL, Hitchcock YJ, Jimeno A, Juloori A, Kase M, Leizman D, Maghami E, Mell LK, Mittal BB, Pinto HA, Price K, Rocco JW, Rodriguez CP, Schwartz D, Shah JP, Sher D, John MS, Wang H, Weinstein G, Worden F, Bruce JY, Yom SS, Zhen W, Montgomery S and Darlow SD. NCCN guidelines® insights: head and neck cancers, version 2.2025. J Natl Compr Canc Netw 2025; 23: 2-11.

[5] Caudell JJ, Gillison ML, Maghami E, Spencer S, Pfister DG, Adkins D, Birkeland AC, Brizel DM, Busse PM, Cmelak AJ, Colevas AD, Eisele DW, Galloway T, Geiger JL, Haddad RI, Hicks WL, Hitchcock YJ, Jimeno A, Leizman D, Mell LK, Mittal BB, Pinto HA, Rocco JW, Rodriguez CP, Savvides PS, Schwartz D, Shah JP, Sher D, St John M, Weber RS, Weinstein G, Worden F, Yang Bruce J, Yom SS, Zhen W, Burns JL and Darlow SD. NCCN guidelines® insights: head and neck cancers, version 1.2022. J Natl Compr Canc Netw 2022; 20: 224-234.

[6] PDQ Adult Treatment Editorial Board. Laryngeal cancer treatment (PDQ®): health professional version. PDQ cancer information summaries. Bethesda (MD): National Cancer Institute (US); 2002.

[7] Ferrari M, Mularoni F, Smussi D, Gaudioso P, Bonomo P, Friborg J, Ghi MG, Gregoire V, Harrington K, Hunter K, Maroldi R, Martino R, Mesia R, Peretti G, Psyrri A, Schindler A, Succo G, Szturz P, Vilaseca I, Nicolai P and Bossi P. International consensus on laryngeal preservation strategies in laryngeal and hypopharyngeal cancer. Lancet Oncol 2025; 26: e264-e281.

[8] Wolf GT, Fisher SG, Hong WK, Hillman R, Spaulding M, Laramore GE, Endicott JW, McClatchey K and Henderson WG. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. N Engl J Med 1991; 324: 1685-1690.

[9] Forastiere AA, Zhang Q, Weber RS, Maor MH, Goepfert H, Pajak TF, Morrison W, Glisson B, Trotti A, Ridge JA, Thorstad W, Wagner H, Ensley JF and Cooper JS. Long-term results of RTOG 91-11: a comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. J Clin Oncol 2013; 31: 845-852.

[10] Ramsey T, Tikhtman R and Tang AL. Laryngeal preservation strategies. Surg Oncol Clin N Am 2024; 33: 761-773.

[11] Nobacht A, Meijer TWH, Oosting SF, van der Vegt B, Wedman J, Halmos GB and Plaat BEC. Oncological and functional outcomes in T3 and T4 laryngeal cancer patients: choice for larynx preservation or total laryngectomy based on expected laryngeal function. J Laryngol Otol 2024; 138: 672-678.

[12] Pfuetzenreiter EG Jr, Ferreron GF, Sadka JZ, Souza ABP, Matos LL, Kowalski LP and Dedivitis RA. Total laryngectomy vs. non-surgical organ preservation in advanced laryngeal cancer: a metanalysis. Braz J Otorhinolaryngol 2024; 90: 101404.

[13] Vander Poorten V, Meulemans J, Beitler JJ, Piazza C, Kowalski LP, Mäkitie AA, Paleri V, Rinaldo A, Robbins KT, Rodrigo JP, Silver CE, Sjögren EV, Strojan P, Takes RP and Ferlito A. Salvage surgery for residual or recurrent laryngeal squamous cell carcinoma after (Chemo)radiotherapy: oncological outcomes and prognostic factors. Eur J Surg Oncol 2021; 47: 2711-2721.

[14] Shelan M, Anschuetz L, Schubert A, Bojaxhiu B, Aebersold DM, Elicin O and Giger R. Superior loco-regional control after primary surgery compared to chemo-radiotherapy for advanced stage laryngeal cancer. Front Oncol 2023; 13: 1132486.

[15] Perrier ND, Brierley JD and Tuttle RM. Differentiated and anaplastic thyroid carcinoma: major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin 2018; 68: 55-63.
[16] Wang J and Marion-Gallois R. Propensity score matching and stratification using multiparty data without pooling. Pharm Stat 2023; 22: 4-19.
[17] Ségalas C, Leyrat C, Carpenter JR and Williamson E. Propensity score matching after multiple imputation when a confounder has missing data. Stat Med 2023; 42: 1082-1095.
[18] Shi J. Ambient ammonium exposure is associated with physical dysfunction in older adults in China. Sci Rep 2025; 15: 19162.
[19] Bettega F, Mendelson M, Leyrat C and Bailly S. Use and reporting of inverse-probability-of-treatment weighting for multicategory treatments in medical research: a systematic review. J Clin Epidemiol 2024; 170: 111338.
[20] Patel SA, Qureshi MM, Dyer MA, Jalisi S, Grillone G and Truong MT. Comparing surgical and nonsurgical larynx-preserving treatments with total laryngectomy for locally advanced laryngeal cancer. Cancer 2019; 125: 3367-3377.
[21] Fu BJ, Potter AL, Pipkorn P, Kraimer KL, Chen MM, Rajasekaran K, Yang CJ and Lee JJ. Differences in survival following surgery versus chemoradiotherapy for clinical stage T4a laryngeal squamous cell carcinoma: a propensity score-matched analysis. Head Neck 2026; 48: 1319-1331.
[22] Merdad M, Mogharbel A, Alghamdi AS, Almansouri OS, Alahmari AF and Alqutub A. Comparative effectiveness of total laryngectomy versus organ preservation in T3 laryngeal cancer: a systematic review and meta-analysis. Eur Arch Otorhinolaryngol 2025.
[23] Rao KN, Pai PS, Dange P, Kowalski LP, Strojan P, Mäkitie AA, Guntinas-Lichius O, Robbins KT, Rodrigo JP, Eisbruch A, Takes RP, de Bree R, Coca-Pelaz A, Piazza C, Chiesa-Estomba C, López F, Saba NF, Rinaldo A and Ferlito A. Survival outcomes in T3 laryngeal cancers: primary total laryngectomy vs. concurrent chemoradiation or radiation therapy-a meta-analysis. Biomedicines 2023; 11: 2128.
[24] García-Cabo P, López F, Sánchez-Canteli M, Fernández-Vañes L, Álvarez-Marcos C, Llorente JL, Rúa M, Blay P and Rodrigo JP. Matched-pair analysis of survival in the patients with advanced laryngeal and hypopharyngeal squamous cell carcinoma treated with induction chemotherapy plus chemo-radiation or total laryngectomy. Cancers (Basel) 2021; 13: 1735.
[25] Zumsteg ZS, Luu M, Fortpied C, Jang JK, Chen MM, Mallen-St Clair J, Walgama E, Le QT, Machtay M, Tribius S, Forastiere A, Wong S, Ozsahin EM, Gregoire V, Vermorken JB, Ho AS and Yom SS. Re-examining post-operative chemoradiotherapy in head and neck cancer: an updated long-term combined analysis of RTOG 9501/EORTC 22931. Ann Oncol 2025; 36: 1379-1388.
[26] Victor MT, Faraji F, Voora R, Kalavacherla S, Mell LK, Rose BS and Guo TW. Factors associated with total laryngectomy following organ-preserving treatment of laryngeal SCC. Laryngoscope Investig Otolaryngol 2024; 9: e1317.
[27] Lee MY, Belfiglio M, Zeng J, Fleming CW, Koyfman S, Joshi NP, Lamarre E, Prendes B, Scharpf J, Lorenz RR, Woody NM, Adelstein DJ, Geiger JL, Chute DJ and Ku JA. Primary total laryngectomy versus organ preservation for locally advanced T3/T4a laryngeal cancer. Laryngoscope 2023; 133: 1122-1131.
[28] Higashino M, Aihara T, Terada T and Kawata R. Influence of preoperative radiation therapy on the occurrence of pharyngocutaneous fistula after total laryngectomy. Cureus 2021; 13: e13797.
[29] Šifrer R, Strojan P, Tancer I, Dolenc M, Fugina S, Zore SB and Aničin A. The incidence and the risk factors for pharyngocutaneous fistula following primary and salvage total laryngectomy. Cancers (Basel) 2023; 15: 2246.
[30] Moreno MA, Wax MK, Gardner JR, Cannady SB, Graboyes EM, Bewley AF, Dziegielewski PT, Khaja SF, Bayon R, Ryan J, Al-Khudari S, El-Deiry MW, Ghanem TA, Huang A, Patel R, Higgins KM, Jackson RS and Patel UA. Reconstruction for salvage laryngectomy with limited pharyngectomy. JAMA Otolaryngol Head Neck Surg 2024; 150: 492-499.
[31] Murariu MO, Boia ER, Sitaru AM, Mot CI, Negru MC, Brici AC, Zahoi DE and Balica NC. Long-term quality of life and functional outcomes in patients with total laryngectomy. Cancers (Basel) 2025; 17: 1011.

**Table S1.** Detailed chemoradiotherapy regimens and treatment completion in the CCRT group before and after PSM

| Cohort | Variable | Category | n (%) |
|---|---|---|---|
| Before PSM (n = 244) | Chemotherapy regimen | Standard cisplatin (80-100 mg/m$^2$ q3w) | 196 (80.3) |
| | | Cisplatin-containing alternative regimen | 48 (19.7) |
| | Alternative regimen subtype (n = 48) | Carboplatin | 20 (41.7) |
| | | Nedaplatin | 18 (37.5) |
| | | Reduced-dose cisplatin (< 80 mg/m$^2$) | 10 (20.8) |
| | Chemotherapy cycles completed | 2 cycles | 68 (27.9) |
| | | 3 cycles | 176 (72.1) |
| | Radiotherapy completion | Completed prescribed dose (≥ 66 Gy) | 220 (90.2) |
| | | Incomplete | 24 (9.8) |
| After PSM (n = 220) | Chemotherapy regimen | Standard cisplatin (80-100 mg/m$^2$ q3w) | 176 (80.0) |
| | | Cisplatin-containing alternative regimen | 44 (20.0) |
| | Alternative regimen subtype (n = 44) | Carboplatin | 18 (40.9) |
| | | Nedaplatin | 17 (38.6) |
| | | Reduced-dose cisplatin (< 80 mg/m$^2$) | 9 (20.5) |
| | Chemotherapy cycles completed | 2 cycles | 59 (26.8) |
| | | 3 cycles | 161 (73.2) |
| | Radiotherapy completion | Completed prescribed dose (≥ 66 Gy) | 196 (89.1) |
| | | Incomplete | 24 (10.9) |

Note: PSM, Propensity score matching; CCRT, Concurrent chemoradiotherapy; q3w, Every 3 weeks.

**Table S2.** Distribution of postoperative adjuvant therapy in the TL group before and after PSM

| Cohort | Postoperative adjuvant therapy | n (%) |
|---|---|---|
| Before PSM (n = 268) | Surgery alone | 18 (6.7) |
| | Postoperative radiotherapy alone | 142 (53.0) |
| | Postoperative chemotherapy alone | 108 (40.3) |
| | Postoperative chemoradiotherapy | 0 (0) |
| After PSM (n = 220) | Surgery alone | 16 (7.3) |
| | Postoperative radiotherapy alone | 114 (51.8) |
| | Postoperative chemotherapy alone | 90 (40.9) |
| | Postoperative chemoradiotherapy | 0 (0) |

Note: PSM, Propensity score matching; TL, Total laryngectomy.

**Table S3.** Sensitivity analysis comparing CCRT with TL patients treated with surgery alone only

| Endpoint | Cohort | Model | N | Events | TL (surgery alone, n) | CCRT (n) | HR (95% CI) | *P* value |
|---|---|---|---|---|---|---|---|---|
| OS | Before PSM | Unadjusted Cox regression | 262 | 137 | 18 | 244 | 5.607 (1.778-17.685) | 0.003 |
| | Before PSM | Adjusted Cox regression | 262 | 137 | 18 | 244 | 4.221 (1.285-13.861) | 0.018 |
| | After PSM | Unadjusted Cox regression | 236 | 129 | 16 | 220 | 5.266 (1.668-16.624) | 0.005 |
| | After PSM | Adjusted Cox regression | 236 | 129 | 16 | 220 | 3.686 (1.120-12.126) | 0.032 |
| DFS | Before PSM | Unadjusted Cox regression | 262 | 150 | 18 | 244 | 5.120 (1.879-13.953) | 0.001 |
| | Before PSM | Adjusted Cox regression | 262 | 150 | 18 | 244 | 3.768 (1.330-10.676) | 0.013 |
| | After PSM | Unadjusted Cox regression | 236 | 141 | 16 | 220 | 4.760 (1.745-12.988) | 0.002 |
| | After PSM | Adjusted Cox regression | 236 | 141 | 16 | 220 | 3.240 (1.140-9.204) | 0.027 |

Note: OS, Overall survival; DFS, Disease-free survival; PSM, Propensity score matching; CCRT, Concurrent chemoradiotherapy; TL, Total laryngectomy; HR, Hazard ratio; CI, Confidence interval.