TILs in oropharyngeal carcinoma: prognostic value and evaluation of a standardized method

A standardized method for TIL assessment in OPSCC

Stijn De Keukeleire¹, Tijl Vermassen¹,², Astrid De Meulenaere¹, Philippe Deron³, Wouter Huvenne³, Frédéric Duprez⁴, David Creytens⁵,⁶, Jo Van Dorpe⁵,⁶, Sylvie Rottey¹,²,⁶ and Liesbeth Ferdinande⁵,⁶

¹Ghent University Hospital, Department of Medical Oncology, Ghent, Belgium
²Ghent University Hospital, Drug Research Unit Ghent, Ghent, Belgium
³Ghent University Hospital, Department of Head and Neck Surgery, Ghent, Belgium
⁴Ghent University Hospital, Department of Radiation Oncology, Ghent, Belgium
⁵Ghent University Hospital, Department of Pathology, Ghent, Belgium
⁶Cancer Research Institute Ghent (CRIG), Ghent, Belgium

word count: 3652

Corresponding author
Stijn De Keukeleire, MD
Department of Medical Oncology – 2B2
Ghent University Hospital
De Pintelaan 185
9000 Ghent
Belgium
Telephone: +32 (0)9 332 12 11
Fax: +32 (0)9 332 62 85
Mail: Stijn.DeKeukeleire@UZGent.be
1. Introduction

The tumor microenvironment (TME) is a complex dynamic bidirectional system of cells shaped during cancer development. Essentially, it consists of tumor cells and host cells, an extracellular matrix (ECM), endothelial cells, fibroblasts and an immune infiltrate consisting of lymphocytes, macrophages, dendritic cells and granulocytes [1-3]. Tumor infiltrating lymphocytes (TILs) are an important component of this immune infiltrate and include lymphocytes recruited from both the adaptive and innate immune system such as B cells, CD4+ regulatory T cells, CD8+ cytotoxic T cells, natural killer cells and γδ T cells. These subsets of TILs have been thoroughly investigated and correlated with outcome, survival and therapy sensitivity (e.g. to immune checkpoint inhibition) in several malignancies, including squamous cell cancer of the head and neck (SCCHN) [3-10].

SCCHN is the sixth most common malignancy in the world with the oropharynx being the most affected anatomical subsite [11,12]. The majority of patients with SCCHN has a history of tobacco and alcohol abuse, though HPV infection has also been involved in the pathogenesis of oropharyngeal squamous cell carcinoma (OPSCC) [13-15], having a prevalence rate ranging between 23-73 %, depending on the geographical region and time period [16]. Curative treatment strategies mainly comprise a combination of surgery, radiation therapy and platin-based systemic therapy. Unfortunately, as recurrence and metastasis are common events, SCCHN has a high mortality rate ranging up to 50% with a median overall survival (mOS) of 6-8 months [17-20]. Recently, immune checkpoint inhibition (ICI) was successfully implemented as a first- and / or second line treatment in recurrent or metastatic SCCHN improving mOS up to 14.5 months when administered in monotherapy. Nonetheless, overall response rate of ICI remains relatively low reaching up to a mere 18% [12,21].

Aforementioned data have incited scientists to further unravel the TME and investigate potential biomarkers or targets. These may aid oncologists to tailor use of therapy in which TILs may play their own part. TIL evaluation is mostly assessed via microscopic evaluation of tumor tissue sections. Identification of TILs is based on morphological assessment on hematoxylin-eosin (HE) staining or immunohistochemical labelling of different subtypes of TILs. Quantification of TILs can be done semi-quantitatively by categorical scoring or quantitatively using digital imaging analysis. Although many studies report a prognostic value for TILs in head and neck cancer, TIL scoring has not yet found its way to daily clinical practice [22-26]. The different reasons for this are reviewed elsewhere [8] but the main problem resides in the lack of uniform scoring of TILs. Different methodologies, often study group-specific, lead
to different and poorly comparable results, eventually leaving the question about clinical utility unanswered.

To address this issue, the International Immuno-Oncology Biomarkers working group (IBWG) proposed a standardized methodology for TIL assessment in solid tumors [27]. This method is affordable, accessible, easily and widely applicable, and has demonstrated to be reproducible, reliable and of prognostic and predictive value in invasive breast carcinoma. The clinical validity and utility of this standardized approach of TIL quantification was already explored in other solid tumors such as melanoma and lung cancer but remains to be investigated in SCCHN. The purpose of this retrospective study was to apply the standardized methodology for the scoring of TILs, as proposed by the IBWG, in a population of patients diagnosed with OPSCC. The study has two main objectives: first, to assess the analytical validity of the scoring system by identifying confounding factors and determining inter-observer variability and second, to assess its clinical usefulness by analysing its prognostic value in this patient population.
2. Materials and methods

2.1 Study population

Ninety-two treatment-naive patients with biopsy-proven OPSCC (2004–2013) were selected from the archival database of the Department of Pathology, Ghent University Hospital. The study was approved by the local ethics committee of the Ghent University Hospital (Belgian registration number B670201420039). Tissue sections were derived from biopsies (n = 74) or resections (n = 18). Immunohistochemistry (P16, CD8) was performed and scored as previously described [28].

The majority of patients were male (n = 76) and older than 60 years (n = 55). Patients were staged applying the 8th edition of TNM classification, with P16 immunohistochemistry as surrogate marker for HPV* tumors [29]. Patients were classified as clinical stage I/II (n = 16), III (n = 14) or IV (n = 62). We adapted the topographic classification of OPSCC suggested by Tham et al. [30] in our manuscript. In this meta-analysis OPSCC was divided into: (1) tonsil related area (TRA) comprising of the base of tongue, lingual tonsil, tonsillar fossa and tonsillar Pillar, and tonsil not otherwise specified, (2) non-tonsil related area (nTRA) comprising of soft palate, uvula, lateral and posterior oropharyngeal wall, and (3) overlapping area (OA) comprising of overlapping lesions of OPSCC or OPSCC not otherwise specified. The most commonly affected anatomical region was TRA (n = 56), followed by nTRA (n = 25) and OA (n = 11). OPSCCs were classified as well (n = 7), moderate (n = 57), poor (n = 21), or basaloid differentiated (n = 7). Eight poorly differentiated OPSCCs had the morphological aspect of a lymphoepithelioma-like carcinoma (LELC). HPV positivity was determined via P16 CINtec® P16 Histology on a BenchMark XT automated stainer (both from Ventana Medical Systems, Inc., Arizona, USA) under ISO15189:2012. We identified 28 cases (30%) that were P16*. An additional HPV CISH was performed retrospectively on this cohort of older paraffin blocks. However, as the results were variable and therefore possibly false negative (only 17 of 28 P16* samples was HPV positive), P16 was considered the most reliable discriminator for HPV status in this study population [31-35].

All patients received primary therapy based on surgery (n = 49), radiotherapy (n = 22), or radio- and chemotherapy (cisplatin min. dose 200mg/m²) (n = 21). Forty-four patients that were primarily treated with surgery received adjuvant therapy based on radio(chemo)therapy of which only one patient did not reach the full described dose of radiotherapy. Of patients that were prescribed primary treatment with radio- and chemotherapy, four patients were ineligible for platinum: two received cetuximab (250mg/m²) during radiotherapy and two were unfit for cisplatinum and were treated with primary radiotherapy. Three patients did not complete the
full dose of radio-and chemotherapy as treatment was withheld early due to diagnosis of a second primary tumor and two patients had rapid disease progression with eventually, death.

2.2 TIL scoring
TILs were evaluated following the protocol from the IBWG (Supplementary table 1). Briefly, HE stained sections containing invasive squamous cell carcinoma were scored for mononuclear inflammatory cells, including lymphocytes, plasma cells and macrophages in stromal and intratumoral tissue, while granulocytes were excluded from evaluation. TILs were reported as a continuous variable (%) for the stromal and intratumoral compartment separately. The percentage of stromal TILs (TILstr) is the area of stromal tissue occupied by mononuclear inflammatory cells over the total stromal area. For intratumoral TILs (TILtum), the area covered by mononuclear inflammatory cells over the amount of tumoral cells covering the area had to be assessed [2,27,36]. In the current study, TIL scoring was performed by four independent observers, three of which were senior pathologists and one PhD fellow. All observers were blinded from each other’s assessment and patient outcomes. Figure 1 depicts illustrative OPSCC HE sections used for TIL assessment (Figure 1A-D). Per patient, only one HE section with representative tumor tissue was selected and assessed. Attention was paid to possible confounding factors that may complicate reproducible TIL scoring. These factors include site-specific characteristics such as presence of pre-existing lymphoid stroma, superficial ulcerations or erosions (Figure 1F) or tissue-specific factors such as insufficient intertumoral stroma that may hamper adequate TILstr evaluation in biopsy specimens. Particularly, pre-existing lymphoid tissue may serve as an issue in TIL assessment as it is commonly present in TRA of OPSCC. Therefore, this may impede proper differentiation between lymphoid infiltrate and native tissue for which the IBWG suggested to focus primarily on tumoral TILs [27].

2.3 Statistical analysis
TIL density was assessed as a continuous variable and was dichotomized as a categorical variable (‘high’ versus ‘low’ density) at the cut-off value determined through ROC curve analysis. Missing values (e.g. not evaluable tumor tissue section by at least one observer) were replaced by the mean value of all scoring observers. The inter-observer validity in TILstr and TILtum was firstly determined by means of the intraclass correlation coefficient (ICC) using TILs as continuous variable. Next, the inter-observer validity was assessed via the overall percentage agreement (OPA), the inter-rater reliability (IRR) [37] and Fleiss k. Associations between TILs and clinicopathological characteristics were determined via Chi-square test or Fisher’s exact tests when necessary. Univariate hazard ratio (HR) of each baseline and clinical parameter on OS was determined by a log-rank (Mantel–Cox) test. OS was calculated from time of (biopsy-proven) diagnosis until date of death or final follow-up. The survival curves
were plotted using the Kaplan–Meier method. Multivariate analyses were conducted using the Cox regression model. Decision curve analysis was performed to evaluate the added prognostic value of TIL assessment to standard clinicopathological parameters. All tests were calculated as two-sides and a P value reached statistical significance if <0.05.
3. Results

Ninety-two OPSCC tissue samples were evaluable for TIL assessment. TIL density was assessed for both stromal and intratumoral compartment as a continuous variable (%). The scoring of each observer was therefore quite arbitrary when commencing quantification, as the observers applied different thresholds for assessment (lowest score being <5% or <1%) or using intervals (5-10%; 10-15%; etc.). This made statistical analysis more complex, thus the lowest cut-off value for TIL$_{str}$ and TIL$_{tum}$ was set at <5% and <1%, respectively. Intervals were assessed as median values for each observer. To perform survival analysis and interpretation, the TIL percentages of all scoring observers for each sample were calculated as median percentages.

An overview of all scores (TIL$_{str}$ and TIL$_{tum}$) per case for each observer is shown in Figure 2. We observed that TIL$_{str}$ positively correlated to TIL$_{tum}$ ($R^2 = 0.46; P < 0.001$), independently from P16 status; Supplementary figure 1).

3.1 Inter-observer validity

Inter-observer variance was firstly determined for TILs as a continuous variable through the ICC by means of absolute agreement. Concordance was observed as excellent between all observers for TIL$_{str}$ (ICC = 0.90, $P < 0.001$) and TIL$_{tum}$ (ICC = 0.89, $P < 0.001$). Presence of any confounding factor did only slightly reduce inter-observer concordance (TIL$_{str}$ ICC = 0.88, $P < 0.001$; TIL$_{tum}$ = 0.86, $P < 0.001$). Of all confounding factors, ICC was considerably lower in samples with restricted intertumoral stroma (TIL$_{str}$ ICC = 0.75; TIL$_{tum}$ = 0.76). All ICCs are given in Supplementary table 2.

Similarly, the inter-observer validity for TIL density was evaluated as a dichotomized variable. Cut-off was set at 30% for TIL$_{str}$ and 5% for TIL$_{tum}$ as these were the thresholds retrieved by ROC curve analysis. Other cut-offs were equally assayed for determining the highest inter-observer concordance in TIL$_{str}$ (10% – 20% – 30% – 40%) and TIL$_{tum}$ (1% – 5% – 10% – 20%). For TIL$_{str}$, inter-observer concordance was moderate when based on Fleiss $\kappa$ for all thresholds (< 20%, $\kappa = 0.450$; < 30%, $\kappa = 0.484$; < 40%, $\kappa = 0.479$) except < 10% ($\kappa = 0.314$). Moderate concordance was observed for TIL$_{tum}$ regarding classes < 10 % ($\kappa = 0.446$) and < 20% ($\kappa = 0.460$), though only fair and poor agreement was noted for < 5% ($\kappa = 0.353$) and < 1% ($\kappa = 0.114$), respectively. The OPA was ranged between 51.1% and 59.8% for all TIL$_{str}$ thresholds between 43.5 and 84.8 % for all TIL$_{tum}$ classes, of which lowest cut-offs (5% and 10%) had the poorest agreement (43.5% each). The IRR was excellent for both TIL$_{str}$ and TIL$_{tum}$, ranging from 83.7 to 87.8 % for TIL$_{str}$ and 78.5 to 91.3% for TIL$_{tum}$. All inter-observer validating parameters seemed to improve when only using samples with no confounding factor present
in comparison to samples with presence of any confounding factor. No differences in inter-observer variability were detected when comparing TIL scoring on biopsy or resection material. An overview of the inter-rater agreement for the dichotomized variables is shown in Supplementary table 3 to 7.

3.2 TIL density and clinicopathological characteristics in OPSCC
An overview of all clinicopathological features is given in Table 1. A correlation was observed between TIL_\text{tum} and tumor topography: TRA OPSCC had higher TIL_{str} and TIL_{tum} infiltration than other sites. Similarly, low T-stage (T1/T2) tumors had higher TIL_{tum} infiltration. A positive P16 status correlated with augmented TIL_{str} and TIL_{tum} infiltration. Regarding differentiation grade, basaloid and poorly differentiated (particularly LELC-variant) OPSCC had higher TIL infiltration compared to well and moderately differentiated OPSCC.

3.3 Survival analysis
OS analysis was done using thresholds of 30% and 5% for TIL_{str} and TIL_{tum}, respectively. Median OS was 1.90 years for low TIL_{str} and 3.04 years for high TIL_{tum}, while being 9.75 years for both high TIL_{str} and high TIL_{tum}. Kaplan Meier plotted OS curves clearly indicate a survival benefit for high TIL_{str} and high TIL_{tum} versus low TIL_{str} and TIL_{tum} (Figure 3).

On univariate analysis, high TIL_{str} (P < 0.0001) and TIL_{tum} (P = 0.0022) correlated with longer survival. The same applied for a positive P16 status (P = 0.0060) and a high stromal CD8 count (P = 0.0058) which also correlated with better OS (Supplementary figure 2).

Multivariate analysis to determine which type of TIL count should be used for OS proved that high TIL_{str} achieved the best OS benefit (Supplementary table 8). Using TIL_{str} in a multivariate model next to P16 status, showed that high TIL_{str} was an independent prognostic predictor for OS (P < 0.0001), being superior to P16 status. In addition, we replaced TIL_{str} by TIL_{tum} for the subset of patients where pre-existing lymphoid tissue was present (n = 26). This however exerted no effect on the observed OS benefit (P < 0.0001: Supplementary figure 2).

All univariate and multivariate HRs are given in Table 2.

The impact of this observation that TIL_{str} is an independent prognosticator for OS was further demonstrated. Firstly, high TIL count was always associated with improved OS, when stratifying for P16 status. This was significantly associated for TIL_{str} but not for TIL_{tum} (Supplementary figure 3). Based on this observation, P16 status was combined with TIL_{str} to evaluate the effect on OS outcome. Herein, it is noticed that the highest survival curves reached, are those for P16^+ / high TIL_{str} and for P16^- / high TIL_{str} (P < 0.0001; Figure 4).
In addition, a decision curve analysis was performed to evaluate the added value of TIL$_{str}$ in the current clinical model. An experimental model, which includes TIL$_{str}$ next to all other clinicopathological parameters clearly showed an increase in net benefit for prognostic use in OPSCC (Supplementary Figure 4).
4. Discussion and conclusion

To our knowledge, this is the first study in which the standardized TIL scoring methodology proposed by the IBWG was applied in a cohort of patients diagnosed with OPSCC, assessing both its analytical validity and clinical utility.

The four observers of this study agreed that this method of TIL quantification is quite challenging in OPSCC. This was mainly due to the large proportion of samples (62%) with one or more confounding factors such as pre-existing lymphoid tissue (28%), restricted intertumoral stroma (19%) and presence of ulcerations or erosions (35%). These characteristics are specific to OPSCC tumor samples and are not commonly encountered in breast cancer, the tumor type for which this scoring method was initially developed. Biopsies often displayed a restricted amount of intertumoral stroma (22% versus 0% in resection material), making adequate scoring of TIL\(_{\text{str}}\) sometimes difficult. All observers concurred that lack of intertumoral stromal tissue was subjectively the most disturbing problem to adequately assess TILs. One may therefore suggest resection material to be superior for TIL quantification. However, surgical management is not always an option in OPSCC patients, who are alternatively treated with radio(chemo)therapy, with biopsy material being the only available tumor tissue for additional analyses or testing. In resection specimens, heterogeneity in TILs distribution may complicate the overall assessment of TIL density (Fig. 1 E). Despite these difficulties, we noticed that the inter-observer variance was only slightly affected by the limited intertumoral stroma and remained considerably well for other confounders.

The current IBWG-designed method states that TIL quantification should be assessed as a continuous variable. This methodology could however be simplified and made more practicable for the pathologist by quantifying TILs as a non-continuous variable. Assessing and validating the ideal threshold to differentiate between high or low TIL infiltration is therefore needed as most reported cut-offs are study-dependent. Here we have assessed the inter-observer variability for categorical variables (Fleiss k, IRR) when implementing various cut-off values. The inter-observer variability proved moderate to excellent, depending on the employed statistical parameters. The variability can of course be attributed to the fact that results were based on the observers’ scores as a continuous variable and not as a dichotomous variable as every observer has scaling differences or uses different cut-offs. Further exploration, harmonization and standardisation is therefore needed before considering binary scoring in the future.

In line with other studies, our results showed that stromal and intratumoral count of TILs, P16 status and stromal CD8 infiltration were correlated to longer survival when performing
univariate analysis. Multivariate analysis however showed stromal TIL count to be the only independent prognosticator for mortality in OPSCC. Moreover, adding TIL\textsubscript{str} to a clinical model of currently known clinicopathological parameters clearly results in an increased net benefit for OS prognosis (cfr. Supplementary material fig. 4). As we showed with this study that the IBWG method is reproducible for future research in OPSCC, standardized TIL quantification could be an additional diagnostic modality to be used for routine pathological assessment.

It is well known that P16/HPV\textsuperscript{+} OPSCC is correlated with a better response to radio(chemo)therapy, a favourable OS and lower likelihood of relapse. It is characterized by a highly cellular TME due to augmented expression of foreign virus-related antigens, thus provoking a higher inflammatory response than P16/HPV\textsuperscript{-} SCCHN with higher TIL counts. This is considered to contribute to the more favourable outcome of this type of tumors [38-46]. As expected, our analysis found P16\textsuperscript{+} status to be correlated with an increased survival benefit (cfr. Supplementary figure 3), and a significantly higher infiltration of TILs (Figure 4). However, in multivariate analysis, P16 was not withheld as an independent prognostic factor in our cohort. In fact, TILs scoring turned out to be the most important prognostic factor, resulting in survival benefit of P16\textsuperscript{-} tumors with high TILs infiltration compared to p16\textsuperscript{+} tumors with low TILs infiltration. This suggests that the better prognosis of HPV\textsuperscript{+} tumors is mainly determined by the TILs infiltration that is induced by HPV infection, and that HPV\textsuperscript{+} tumors without high TILs infiltration lose a significant portion of this survival benefit (Figure 4).

As mentioned above, pre-existing background lymphoid stroma may be abundantly present in OPSCC, especially TRA, and complicate TIL assessment. Indeed, 33\% of TRA-samples contained lymphoid tissue, and TRA-subsites were characterized by both higher TIL infiltration (both stromal and intratumoral) and better prognosis than SCCHN at other anatomical regions. However, no significant correlation was observed in the presence of pre-existing lymphoid tissue and topography (P = 0.1052) in our cohort. We performed an additional analysis by replacing TIL\textsubscript{str} with TIL\textsubscript{tum}, as recommended by the IBWG, in all patients with pre-existing lymphoid tissue. Eventually, this did not inflict any changes to our main conclusion about TILs as a prognostic factor in OPSCC (supplementary figure 2). Our data suggest that, regardless of the fact that the higher TIL counts are tumor-related or site-related, high TILs seem to contribute to a better prognosis compared to other OPSCC subtypes.

In this cohort, poorly differentiated OPSCC had increased stromal and intratumoral TIL infiltration in comparison to well and/or moderately differentiated OPSCC, suggesting a prognostic benefit for this subset. This contradicts the fact that dedifferentiation is an indicator of poor prognosis, which prompted us to re-classify the poorly differentiated group in conventional poorly differentiated squamous cell carcinomas and tumors with LELC
morphology. LELC is a non-keratinizing undifferentiated carcinoma that most commonly arises in the nasopharyngeal region as an Epstein-Barr virus (EBV)-related carcinoma, characterized by a dense lymphoplasmocytic cell infiltrate. This type of malignancy may also develop in other mucosal sites of the head and neck, preferably the oropharynx. In case of the latter, they are considered a variant of HPV-associated OPSCC and are mostly P16/HPV⁺. The prognosis of LELC is considered good, comparable to other HPV-associated OPSCC. In our cohort of poorly differentiated OPSCCs, eight out of thirteen poorly differentiated carcinomas turned out to be LELC. In this subgroup, five were HPV⁺ and two were EBV⁺. One patient was HPV⁻ and EBV-status could not be assessed due to technical issues. The foregoing explains the higher TIL scores and the better prognosis in the group of poorly differentiated carcinomas in our cohort [47,48].

Finally, we should comment on the limitations of the current study. These are mainly the retrospective nature of our research and the rather limited cohort size of 92 patients. The fact that the majority of samples were biopsy-based tissue samples (n=75) while the IBWG guidelines recommend the use of resection material as this ensures a better average assessment of TILs, may be considered a limitation as well. However, this disproportionate distribution of biopsy versus resection material reflects the real-life samples of OPSCC that are available for the pathologist and this should be therefore also regarded as a strength of this study.

We may conclude that, despite specific confounding factors characteristic of OPSCC, the IBWG proposed method is a reproducible way of assessing stromal and intratumoral TILs in OPSCC with significant prognostic value. This method for quantification of TILs could be implemented in future research in SCCHN for further validation. This should preferably be covered in prospective trials exploring its association with other prognosticators such as PD-L1, smoking and viral status, its impact on survival, disease relapse, therapy responsiveness (especially for immune-modulating-based therapies), and its role in daily clinical practice.
5. Additional information

Acknowledgments
We thank all clinicians that referred the patients to our hospital.

Ethics approval
The study was approved by the local ethics committee of the Ghent University Hospital (Belgian registration number B670201420039). This study was performed in accordance with the declaration of Helsinki. The necessity of obtaining informed consent from the included patients was waived by our ethics committee as the majority of patients had deceased at the moment the study was initiated.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest statement
All authors declare no conflict of interest.
### Table 1: Association between TILs and clinicopathological variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N</th>
<th>TILstr (%) Median (IQR)</th>
<th>P-value</th>
<th>TILtum (%) Median (IQR)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 60 years</td>
<td>37</td>
<td>30.0 (9.7 – 50.0)</td>
<td>0.0885</td>
<td>3.0 (0.5 – 7.5)</td>
<td>0.2953</td>
</tr>
<tr>
<td>&gt; 60 years</td>
<td>55</td>
<td>15.0 (7.5 – 39.4)</td>
<td></td>
<td>1.0 (0.5 – 3.5)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>16</td>
<td>21.3 (10.0 – 43.8)</td>
<td>0.9835</td>
<td>1.0 (0.5 – 4.0)</td>
<td>0.7066</td>
</tr>
<tr>
<td>Male</td>
<td>76</td>
<td>21.3 (7.5 – 45.0)</td>
<td></td>
<td>1.5 (0.5 – 5.0)</td>
<td></td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nTRA</td>
<td>25</td>
<td>22.5 (9.7 – 45.0)</td>
<td></td>
<td>1.0 (0.5 – 3.1)</td>
<td></td>
</tr>
<tr>
<td>OA</td>
<td>11</td>
<td>7.5 (5.4 – 17.5)</td>
<td>0.0328</td>
<td>0.5 (0.1 – 1.0)</td>
<td>0.0014</td>
</tr>
<tr>
<td>TRA</td>
<td>56</td>
<td>25 (7.5 – 45.0) a</td>
<td></td>
<td>3.0 (0.8 – 6.9) b</td>
<td></td>
</tr>
<tr>
<td>( T )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>27.5 (11.3 – 47.5)</td>
<td>0.3374</td>
<td>3.5 (1.3 – 8.8) c</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>17.5 (8.1 – 65.0)</td>
<td></td>
<td>3.0 (0.5 – 6.0)</td>
<td>0.0150</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>25.0 (5.8 – 45.0)</td>
<td></td>
<td>0.8 (0.5 – 1.5)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>10.0 (7.5 – 30.0)</td>
<td></td>
<td>1.0 (0.5 – 2.9) c</td>
<td></td>
</tr>
<tr>
<td>( N )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>27</td>
<td>20.0 (10.6 – 30.0)</td>
<td>0.4597</td>
<td>1.0 (0.6 – 3.4)</td>
<td>0.4766</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>10.0 (6.9 – 31.9)</td>
<td></td>
<td>1.0 (0.5 – 3.5)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>25.0 (7.5 – 50.0)</td>
<td></td>
<td>3.0 (0.5 – 7.5)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>22.5 (8.8 – 27.9)</td>
<td></td>
<td>1.0 (0.5 – 3.0)</td>
<td></td>
</tr>
<tr>
<td>Prognostic stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I / II</td>
<td>16</td>
<td>16.3 (11.3 – 27.5)</td>
<td>0.8633</td>
<td>1.3 (0.5 – 3.3)</td>
<td>0.7566</td>
</tr>
<tr>
<td>III</td>
<td>14</td>
<td>21.3 (7.5 – 37.5)</td>
<td></td>
<td>1.0 (0.5 – 5.0)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>62</td>
<td>25.0 (7.5 – 45.0)</td>
<td></td>
<td>1.3 (0.5 – 5.0)</td>
<td></td>
</tr>
<tr>
<td>Grade of differentiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>7</td>
<td>10.0 (6.2 – 14.4) d</td>
<td></td>
<td>0.5 (0.0 – 0.9) g</td>
<td></td>
</tr>
<tr>
<td>Moderately</td>
<td>57</td>
<td>17.5 (7.5 – 32.5) e</td>
<td></td>
<td>1.0 (0.5 – 3.0) f</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>13</td>
<td>20.0 (6.0 – 45.0) f</td>
<td>0.009</td>
<td>3.0 (0.5 – 10.6)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Poor - LELC</td>
<td>8</td>
<td>60.0 (47.5 – 72.5)</td>
<td></td>
<td>10.0 (5.7 – 20.0) g</td>
<td></td>
</tr>
<tr>
<td>Basaloid</td>
<td>7</td>
<td>30.0 (22.5 – 45.0)</td>
<td></td>
<td>3.5 (2.3 – 10.6) h</td>
<td></td>
</tr>
<tr>
<td>( P16 ) status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>63</td>
<td>15.0 (7.5 – 30.0)</td>
<td>0.0002</td>
<td>1.0 (0.5 – 3.0)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Positive</td>
<td>29</td>
<td>37.5 (20.0 – 58.8)</td>
<td></td>
<td>5.0 (1.8 – 12.5)</td>
<td></td>
</tr>
<tr>
<td>Therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>49</td>
<td>20.0 (7.5 – 38.8)</td>
<td></td>
<td>1.5 (0.5 – 5.3)</td>
<td></td>
</tr>
<tr>
<td>Chemotherapy + radiotherapy</td>
<td>21</td>
<td>12.5 (7.5 – 29.1)</td>
<td>0.1235</td>
<td>1.5 (0.5 – 5.0)</td>
<td>0.7777</td>
</tr>
<tr>
<td>Radiotherapy</td>
<td>22</td>
<td>37.5 (12.5 – 65.0)</td>
<td></td>
<td>1.0 (0.5 – 3.1)</td>
<td></td>
</tr>
</tbody>
</table>

*a-i* Letters indicated which subgroups are statistically significantly different

IQR, interquartile range; str, stromal; TIL, tumor infiltrating lymphocyte; tum, tumoral; LELC, lympho-epithelial like carcinoma
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Median survival (95% CI)</th>
<th>Univariate HR (95% CI)</th>
<th>P-value</th>
<th>Multivariate HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TIL_{tum}</strong></td>
<td>Low 3.04 (1.72 – 4.39)</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High 9.75 (3.80 – 9.75)</td>
<td>0.46 (0.28 – 0.75)</td>
<td>0.0022</td>
<td>1.60 (0.69 – 3.71)</td>
<td>0.2772</td>
</tr>
<tr>
<td><strong>TIL_{str}</strong></td>
<td>Low 1.90 (1.57 – 4.09)</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High 9.75 (4.31 – 9.75)</td>
<td>0.33 (0.20 – 0.53)</td>
<td>&lt; 0.0001</td>
<td>0.31 (0.17 – 0.56)</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>CD8_{tum}</strong></td>
<td>Low 4.09 (1.59 – 5.08)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High 6.35 (2.45 – 9.75)</td>
<td>0.60 (0.35 – 1.06)</td>
<td>0.0778</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CD8_{str}</strong></td>
<td>Low 2.45 (1.58 – 4.39)</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High 9.75 (4.31 – 9.75)</td>
<td>0.46 (0.27 – 0.80)</td>
<td>0.0058</td>
<td>0.66 (0.29 – 1.50)</td>
<td>0.3189</td>
</tr>
<tr>
<td><strong>p16 status</strong></td>
<td>Negative 2.58 (1.72 – 4.44)</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive Not reached</td>
<td>0.44 (0.27 – 0.72)</td>
<td>0.0013</td>
<td>0.52 (0.27 – 1.00)</td>
<td>0.0504</td>
</tr>
</tbody>
</table>

CI, confidence interval; HR, hazard ratio; OPSCC, oropharyngeal squamous cell cancer; OS, overall survival; str, stromal; TIL, tumor infiltrating lymphocyte; tum, tumoral.
1: Microscopic evaluation of OPSCC (tonsil).
High TIL density versus low TIL density is seen in (magnification 100x: A-B; magnification 400x: C-D). Heterogenous dissemination of high (arrow) and low (arrowhead) TIL infiltration can be present (E). Presence of polymorphonuclear cells (arrow) due to ulceration and/or erosion can make TIL assessment more complex. (F). OPSCC, oropharyngeal squamous cell cancer; TIL, tumor infiltrating lymphocytes.
Figure 2: TIL score distribution plot for all observers.
X-axis depicts the case number; Y-axis depicts the score (%) for each observer (coloured dot). Scoring has been depicted for TIL\textsubscript{str} (A) and TIL\textsubscript{lum} (B). Trend lines have been depicted for each observer. One may notice that both observers 3 and 4 tend to underscore TIL\textsubscript{str}, in comparison to observers 1 and 2. The same applies for observer 4 regarding TIL\textsubscript{lum}. str, stromal; TIL, tumor infiltrating lymphocyte; tum, tumoral.
Figure 3. Effect of TIL scores on OS in OPSCC patients.

X-axis depicts survival time (in years), Y-axis depicts the cumulative OS probability. OS has been depicted for (A) TIL\textsubscript{str} (9.75 years versus 1.90 years; HR = 0.33 [0.20 – 0.53]; P < 0.0001) and (B) TIL\textsubscript{tum} (9.75 years versus 3.04 years; HR = 0.46 [0.28 – 0.75]; P = 0.0022). TIL scores were dichotomized between high (dashed line) and low scores (full line). Number at risk for each group (high versus low infiltration) has been indicated beneath each OS curve. OPSCC, oropharyngeal squamous cell cancer; OS, overall survival; str, stromal; TIL, tumor infiltrating lymphocyte; tum, tumoral.
Figure 4. Effect of TIL infiltration (TIL_{str}) on OS in OPSCC according to P16 status.

High TIL_{str} infiltration in P16^+ and P16^- status has better OS than low TIL_{str} infiltration in P16^+ and P16^- status
References


