Comparison of Contemporary Staging Systems for Oropharynx Cancer in a Surgically Treated Multi-Institutional Cohort

Running Title: Comparison of Oropharynx Cancer Staging Systems

Olgun Elicin, MD[1], Martina A. Broglie, MD[2,3], Niklaus Fankhauser[4], Sandro J. Stoeckli, MD[2], Philippe Pasche, MD[5], Antoine Reinhard, MD[5], Massimo Bongiovanni, MD[6], Gerhard F. Huber, MD, MSc[2,3], Grégoire B. Morand, MD[3], Alex Soltermann, MD[7], Andreas Arnold, MD[8], Matthias S. Dettmer, MD[9], André Arnoux, MD[10], Edouard Stauffer, MD[11], Vittoria Espeli, MD[12], Francesco Martucci, MD[13], Daniel M. Aebersold, MD[1], Roland Giger, MD[8]

1: Department of Radiation Oncology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland; 2: Department of Otorhinolaryngology, Head and Neck Surgery, Kantonsspital St. Gallen, St. Gallen, Switzerland; 3: Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Zurich, Zurich, Switzerland; 4: CTU Bern, and Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland; 5: Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital of Lausanne, Lausanne, Switzerland; 6: Service of Clinical Pathology, Lausanne University Hospital, Institute of Pathology, Lausanne, Switzerland; 7: Institute of Surgical Pathology, University Hospital of Zurich, Zurich, Switzerland; 8: Department of Otorhinolaryngology, Head and Neck Surgery, Inselspital, Bern University Hospital, Bern, Switzerland; 9: Institute of Pathology, University of Bern, Bern, Switzerland; 10: Department of Otorhinolaryngology, Head and Neck Surgery, Kantonsspital Aarau, Aarau, Switzerland; 11: Institute of Pathology, Promed SA, Fribourg, Switzerland; 12: Oncology Institute of Southern Switzerland (IOSI), Bellinzona, Switzerland; 13: Department of Radiooncology, Oncology Institute of Southern Switzerland (IOSI), Bellinzona, Switzerland

Corresponding author: Olgun Elicin, Department of Radiation Oncology, Inselspital, Bern University Hospital, University of Bern, Freiburgstrasse 4, 3010 Bern, Switzerland. Tel: +41 31 632 26 32. Email: olgun.elicin@insel.ch

Acknowledgements: The authors thank Diana Foerbs, Florian Lang, Laurence de Leval, Martin Müller, Mathieu Lutchmaya, Stefan Spreitzer, Thomas F. Pezier and Wolfram Jochum for their contribution to the initial part of this project which was previously published. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Keywords: head and neck cancer; oropharynx cancer; human papillomavirus; cancer staging; surgery

Published in final edited form as: *Head Neck*. 2019 May;41(5):1395-1402. doi: 10.1002/hed.25574

Abstract

Background: Between the publication of the Union of International Cancer Control staging system (UICC) 7th and 8th editions, other staging algorithms for oropharyngeal squamous cell carcinoma (OPSCC) were proposed from Radiation Therapy Oncology Group (RTOG), MD Anderson Cancer Center (MDACC), and Yale University.

Methods: With C-statistics, the above-mentioned 5 staging algorithms were compared for overall and relapse-free survival endpoints in a multi-institutional cohort of OPSCC cases (n=338) treated with primary surgery.

Results: Pathological UICC 8th ed yielded the highest C-indexes in the entire cohort and in the HPV- subset, whereas MDACC was superior for HPV+ OPSCC. RTOG was the simplest and holistic algorithm with a non-inferior discriminatory power.

Conclusion: UICC 8th ed, MDACC and RTOG offer moderate and comparable efficacy for staging in this OPSCC patient cohort undergoing surgical treatment. Notable discrepancy between clinical and pathological UICC 8th ed algorithms poses potential concerns in diagnosis, treatment, research and data management.

Published in final edited form as: *Head Neck*. 2019 May;41(5):1395-1402. doi: 10.1002/hed.25574

Introduction

Human papillomavirus (HPV) associated oropharyngeal squamous cell carcinoma (OPSCC) is known to have a better prognosis compared to its smoking and alcohol associated counterpart[1,2]. This prominent difference was shown in multiple series treated either with primary surgery[3,4] or primary (chemo)radiotherapy ((C)RT)[5,6]. In 2010, Ang *et al.* proposed a new staging system based on a recursive partitioning analysis (RPA) done on a prospective randomized Radiation Therapy Oncology Group (RTOG) 0129 trial population with specific eligibility criteria[5]. This was the first classification which was taking the HPV status into consideration. Since then, various staging systems including HPV status were proposed with the main aim of improving the risk stratification for a better estimation of prognosis compared to Union of International Cancer Control staging system (UICC) 7th ed[7].

The UICC 8th ed was implemented in January 2017[8]. The new OPSCC staging is separated based on HPV status. For the HPV associated patients, the International Collaboration on Oropharyngeal Cancer Network Staging (ICON-S) was chosen for the clinical staging (cUICC). This staging algorithm was first developed in Canada[9] and validated in multiple centres from North America and northern Europe[6]. In these training and validation cohorts, only 2% of the patients were treated surgically. On the other hand, the development of the pathological staging was based on a retrospective multicentre study on patients mainly treated in North America and United Kingdom[10]. Of note, this pathological staging was not systematically validated such as the clinical staging algorithm at the time of implementation into the UICC 8th ed. Furthermore, the classification of nodal staging for all head and neck cancer subsites except for HPV associated OPSCC was based on a large dataset of oral cavity tumours[8].

In conclusion, it remains unclear, whether the UICC 8th ed can be applied in different areas of the world. Including this staging system, most recently proposed staging algorithms are heavily based on North American patient data mostly treated with primary (C)RT[5,6,10–13]. However, smoking, alcohol consumption and sexual habits and prevalence rates of HPV associated and non-HPV associated OPSCC are strongly dependent on geography. Even among developed countries, these rates vary substantially[14–18]. Therefore, the prognostic stratification power of a staging system developed in a population may vary elsewhere[19]. Despite the remaining importance and high prevalence rates of non-HPV associated OPSCC, the contrast between the abundance of literature about HPV associated OPSCC and the scarcity of published works about non-HPV associated OPSCC in the last decade is striking. Among many proposed staging algorithms for OPSCC, only the RTOG and Dutch systems covered both non-HPV and HPV associated OPSCC[5,13,20], whereas no new algorithm was proposed specifically for non-HPV associated OPSCC. Also in terms of differences in treatment modality, it is possible that a prognostic risk classification derived from a (C)RT cohort may perform worse than expected in a surgical cohort[10].

With this background, our aims were to (1) compare the prognostic qualities of 5 contemporary staging systems which were published until the official release of UICC 8th ed; (2) test their applicability in terms of risk prognostication; and (3) investigate the rates of inter- and intra-stage migration[21] among and within the c/pUICC 7th and 8th editions, respectively in our multi-institutional cohort treated by primary surgery.

Materials and Methods

A cohort of 338 patients diagnosed with OPSCC in 7 participating centres were included in the analysis. The ethics committee of St. Gallen centrally approved the study protocol for all

participating centres (reference number: 12/106/L), and the study was performed in accordance with the Declaration of Helsinki. Due to its retrospective design, it was not feasible and, therefore, not requested by the ethics committee to obtain written informed consent of each study participant. All patients were treated with primary surgery with or without adjuvant treatment based on the presence of pathological risk factors. Details of patient characteristics, diagnostic and treatment procedures were published in a previous paper[4]. To optimize the diagnostic accuracy for HPV status, both p16 and HPV DNA positivity was sought as reported in the literature[13,22]. A tumour was defined as HPV+ only if immunohistochemical p16 overexpression and HPV high risk type DNA by PCR were detected.

Proposed systems other than UICC were chosen, if they met the following criteria: published before 2018, clearly proposed as a staging algorithm by the authors, not being a preliminary version of a later modified algorithm from the same working group, relied on variables which were available in our cohort. Concerning the latter criteria, some models included variables like co-morbidity, non-standard definition for smoking, haemoglobin level, values of epidermal growth factor receptor overexpression, maximum ^{18}F- fluorodeoxyglucose standard uptake which were absent in our database[13,20,23–26]. The remaining eligible systems were tested: c/pUICC 7th and 8th editions, RTOG classification[5], MD Anderson Cancer Centre (MDACC)[11] and Yale University[12] staging systems.

All staging algorithms were compared with Harrell's Concordance-index (C)[27] for the co-primary endpoints of overall survival (OS) and recurrence-free survival (RFS). C-statistics were computed using the Inference of C function (Inf. Cval) of the survC1 package for the R language. The number of iterations of perturbation-resampling was set to the default of 1000. Additionally, hazard ratios (HR) for OS and RFS were computed using a cox proportional hazards regression model implemented in the coxph function of the survival package in the R language. The adjusted

model compensates for smoking above 10 pack years (py), age and extracapsular extension (ECE). The influences of those confounding variables for both endpoints were also reported separately. Due to missing data, alcohol consumption habits were not analysed as a confounding variable. The Student's t-test was used for mean comparison, the Wilcoxon rank-sum test for median comparison and the chi-squared test for categorical variables. Finally, the stage migration effect between the UICC 7th and 8th ed and the agreement among clinical and pathological ECE (cECE and pECE, respectively) were compared both descriptively and by means of Cohen's weighted kappa. All tests were two-tailed and p values < 0.05 were judged as statistically significant. The anonymised version of the data is available upon reasonable request to the corresponding author.

Results

Patient and disease characteristics as well as the corresponding differences between HPV+ and HPV- subgroups are provided in Table 1. Between patients with HPV+ and HPV- OPSCC, statistically significant discrepancies were observed in smoking and alcohol consumption habits, cECE and adjuvant treatment indication. After the post-operative pathological examination for pECE, the previously statistically significant difference in ECE between HPV+ and HPV- patients disappeared. Contrary to expected, there was no significant age difference among patients diagnosed with HPV+ and HPV- OPSCC.

Regardless of HPV status, concordance between cUICC and pUICC was higher within the 7th edition compared to the 8th ed (Table 2). In the 8th ed, pathology examination resulted in an up-staging rate of 19.4% in HPV-, and a down-staging rate of 45.4% in HPV+ patients.

Concerning the inter-staging discordance, cUICC from the 7th to the 8th edition yielded a modest up-staging rate of 8.6% in HPV-, and a high down-staging rate of 93.9% in HPV+ patients.

Similarly, pathological stage migration from 7th to the 8th ed manifested as up-staging in 21.1% of HPV-, and as down-staging in 92% in HPV+ patients (Table 3).

Cohen's weighted kappa coefficients for the inter-agreement of the above-mentioned parameters are found in the Supplementary Table S1.

Overall Survival

When the entire cohort was evaluated, the C-indexes revealed a higher prognostic accuracy for RTOG, cUICC and pUICC 8th ed in an increasing order, compared to UICC 7th ed (Table 4). Almost no difference was observed in the HPV- patients between 7th and 8th ed of both cUICC and pUICC. In HPV+ patients, pUICC 8th ed was superior to pUICC 7th ed, whereas no improvement in prognostic accuracy was observed in cUICC versions. Despite an improvement within the pUICC (from 7th to 8th ed), both pUICC editions fell short of both cUICC editions for HPV+ OPSCC. MDACC slightly outperformed the remaining staging algorithms for HPV+ OPSCC. The RTOG classification markedly lost its prognostic power when it was applied separately on HPV+ and HPV- groups (Table 4). Adjusted and unadjusted ordinal HRs with 95% confidence intervals (CI) for death are provided in Table 5. Corresponding categorical HRs can be found in the Supplementary Table S2. For the entire cohort, RTOG, cUICC 8th ed and pUICC 8th ed yielded statistically significant HRs for OS, with the highest observed in RTOG. For the HPV+ and HPV- OPSCC, no staging system showed any statistically significant risk stratification.

Recurrence-free Survival

In the entire cohort, RTOG, cUICC and pUICC 8th ed outperformed the cUICC and pUICC 7th ed based on C-statistics (Table 4). There was no difference between the RTOG and cUICC 8th ed, whereas pUICC 8th ed showed a marginal benefit. For the HPV- patients, no meaningful difference

was observed among any staging systems. In the HPV+ group, an increasing prognostic power in the following order was observed: cUICC 7th ed, pUICC 7th ed, Yale, cUICC 8th ed, pUICC 8th ed and MDACC. As observed for the OS endpoint, RTOG lost its prognostic power for RFS when separately applied on HPV+ and HPV- groups (Table 4). Adjusted and unadjusted ordinal HRs for recurrence or death are provided in Table 6. Corresponding categorical HRs are outlined in the Supplementary Table S3. For the whole cohort, RTOG, cUICC 8th ed and pUICC 8th ed yielded statistically significant HRs for RFS, with the highest observed in RTOG. In the HPV+ group, pUICC 8th ed, MDACC and Yale showed statistically significant HRs. However, only pUICC 8th ed remained statistically significant as the model was adjusted for risk factors. No staging system showed any statistically significant risk stratification in the HPV- group.

In terms of predefined potential confounding variables in the entire cohort, smoking > 10 py and cECE+ had a significant negative impact on both OS (HR: 2.33 [95% CI: 1.31 - 4.13], p<0.01 and HR: 1.67 [95% CI: 1.00 – 2.78], p=0.049, respectively) and RFS (HR: 1.76 [95% CI: 1.11 – 2.78], p=0.02; HR: 1.77 [95% CI: 1.13 – 2.76], p=0.01, respectively). For HPV+ OPSCC, smoking > 10 py and cECE+ were significantly associated with worse OS (HR: 2.67 [95% CI: 1.09 – 6.58], p=0.03) and RFS (HR: 2.15 [95% CI: 1.06 – 4.32], p=0.03), respectively. For HPV- OPSCC, cECE+ had a significant negative impact on OS (HR: 2.28 [95% CI: 1.18 – 4.40], p=0.01) and RFS (HR: 2.16 [95% CI: 1.18 – 3.94], p=0.01), whereas pECE+ was only significantly associated with worse OS (HR: 1.91 [95% CI: 1.12 – 3.26], p=0.02). The effects of possible confounding variables of age, smoking and ECE on OS and RFS are given in the Supplementary Table S4.

Published in final edited form as: *Head Neck*. 2019 May;41(5):1395-1402. doi: 10.1002/hed.25574

Discussion

Similar to non-surgical therapies, primary surgery demonstrates a superior outcome in HPV+ OPSCC compared to its HPV- counterpart[3,4,28–33]. Although no currently established single standard for the treatment of OPSCC exists, there are numerous ongoing trials about treatment de-intensification in HPV+ OPSCC[34] and comparison of surgery and radiotherapy in terms of functional outcome (EORTC Best-of trial NCT02984410), for which an accurate staging plays a key role for risk stratification. As already mentioned, the increasing academic interest in HPV+ OPSCC and diminishing interest in HPV- OPSCC can only be partially justified with the change in epidemiology mainly in the developed world. The majority of OPSCC cases worldwide are still HPV- and mainly attributed to tobacco and alcohol exposure.

To our knowledge, this is the first study comparing the recently proposed staging systems besides UICC in a cohort consisting of both HPV+ and HPV- OPSCC, all treated by surgery as the primary treatment modality. Recently, Mizumachi *et al.* applied the cUICC 8th to a cohort (n=195) of HPV+ and HPV- OPSCC patients treated with a heterogeneous treatment approach consisting of primary surgery or (C)RT[19]. They concluded that the 8th edition is more suitable for Japanese HPV-OPSCC patients. A similar study with a mixed cohort from Germany (n=561)[17] came to the opposite conclusion, indicating a better prognostic stage discrimination of the 8th edition for HPV+ OPSCC, which is in line with our findings.

Our results demonstrated a higher rate of ECE+ lymph node metastases of HPV+ OPSCC compared to HPV- cases. Moreover, the underestimation of cECE+ by imaging modalities was more prominent in HPV+ OPSCC. The specificity and sensitivity to accurately diagnose the presence of cECE+ with different imaging modalities vary between 70% and 90%[35]. Furthermore, a low intra- and inter-observer agreement can pose problems for clinical diagnosis[36].

Due to major differences in the c/pN and c/pUICC of the 8th edition, we observed a high rate of post-surgical stage migration within the 8th edition compared to the 7th edition of UICC. This stage migration phenomenon is an expected result, and has been recently reported as down-staging in HPV+ OPSCC only cohorts[19,37]. On the other hand, the more than double surgical up-staging rate within the 8th vs. 7th edition (19.4% vs. 8.6%) in our HPV- patients is mainly caused by the incorporation of ECE status into the nodal staging system of UICC 8th ed.

The inter-stage migrations from the 7th to 8th editions in HPV+ and HPV- OPSCC to opposite directions in our series confirm the intention to widen the prognostic gap between two entirely different diseases. HPV+ cases underwent a dramatic down-staging both in clinical (93.9%) and pathological (92%) stages, whereas HPV- cases showed a rather modest up-staging pattern in clinical (8.6%) and pathological (21.1%) classifications. Except for one patient, all inter-stage migrations in HPV+ and HPV- subsets were unidirectional. The similarity between our inter-stage migration rates to a National Cancer Database study (n=6465)[38] is striking: in HPV+ OPSCC, clinical 93.9% / 93.9% and pathological 92% / 91.7% down-staging, respectively.

Except for the ICON-S (i.e. cUICC 8th ed) which partially contained patients from northern Europe, all remaining staging algorithms were developed on cohorts diagnosed and treated in North America. Moreover, it is worth to note that the UICC 8th ed for HPV- OPSCC is based on a North American oral cavity data[8]. The comparison of 5 different staging algorithms in our cohort confirmed the applicability and moderate superiority of pUICC 8th ed for the OS estimation of HPV- OPSCC patients. For the estimation of RFS, RTOG and cUICC 8th ed demonstrated comparable prognostic accuracy, and pUICC 8th ed only showed a marginal benefit over them. When our cohort was separated by HPV status, only a slight improvement was revealed by the UICC 8th ed for HPV+ OPSCC, as the cUICC and pUICC 8th ed showed a moderately better accuracy to its previous edition and Yale, respectively. On the other hand, all of them were

outperformed by MDACC. For the HPV- OPSCC, no substantial improvement was observed with the newer systems.

The prognostic power of the RTOG deserves a careful interpretation. Since it is a three-tiered system which is largely separated by HPV and smoking status, two important problems arise when it is used in a purely HPV+ or HPV- cohort. First, it gets curbed down to a two-tiered system. Then, an imbalanced patient distribution (and therefore a lack of events) occurs between the intermediate and low (if HPV+) or high (HPV-) risk groups. Therefore, RTOG seems to be only useful when used as a holistic staging system. In terms of RFS estimation, RTOG offers a more practical solution than the more complicated c/pUICC 8th ed, which is even not consistent within itself (discordance in clinical and pathological classifications). If, on the other hand, a higher prognostic accuracy is deemed as a more important factor than practicability, MDACC offers an even higher accuracy than UICC with a less complicated and one-stop solution. However, it is worth to note that none of RTOG, MDACC or pUICC 8th ed was validated until the UICC 8th ed was released. Moreover, investigators from John Hopkins Hospital recently compared the cUICC 8th ed and MDACC in a cohort of HPV+ OPSCC where 38% of the patients were treated with primary surgery. Contrary to our findings, they showed a superior risk stratification of the cUICC 8th ed over MDACC both for OS and RFS[39]. This difference to our findings may be caused by geographical factors, difference in primary treatment modalities, as well as due to the fact that the cUICC 8th ed and MDACC were developed on cohorts treated with (C)RT.

We also evaluated the prognostic impact of age, smoking status and ECE as well-established risk factors in the literature. Age was not a statistically significant prognostic factor. In the entire cohort, smoking above 10 py revealed a statistically significant impact on OS and RFS. This effect remained statistically significant for OS in the HPV+ patients. The loss of its significance for the HPV- OPSCC can be explained by 92% of our HPV- OPSCC cases being heavy smokers. In the

literature, there are contradicting reports about the prognostic impact of tobacco exposure on HPV+ OPSCC[5,17].

In our series, a negative prognostic impact of pECE+ on OS was only observed in HPV- OPSCC cases. This finding is in agreement with some of the previously published literature[35,40] showing the lack of prognostic value of ECE status in HPV+ OPSCC, whereas contradicting evidence also exists[37,38,41]. Interestingly, cECE demonstrated a statistically significant impact on both endpoints of OS and RFS except for the HPV+ OPSCC patients' OS. This may be a result of either a type I error, or an unconventional hypothesis-generating finding worth to be investigated further: The incidence of ECE+ is correlated with lymph node size[42]. Another fact is the prognostic impact of volumetric burden of the disease[43,44]. Therefore, it is possible that the volumetric nodal burden of the disease actually plays a more important role than the 'true' pECE, and ECE is a rather surrogate marker for nodal volume, and not necessarily vice versa.

It is important to emphasize that a C-index of 0.5 means that a prognostic model does not have a discriminatory power better than coincidence, and an index of 1 indicates an ideal model. In our current findings, the C-indexes vary between 0.45 and 0.64 which do not suggest a convincingly high prognostic accuracy. Nevertheless, it is interesting to witness the proclamation of marginal benefits for prognostic ability in the magnitude of 0.003 (C-index) as "improvements" in the literature[45]. As we see from cross-validation studies, even an externally validated system may not yield the same value in another cohort and can be outperformed by another staging algorithm[12,26]. As the UICC staging committee members correctly pointed out, not only the prognostic accuracy, but also the practicality, physician compliance and worldwide applicability belong to the important aspects of an ideal staging system[8]. Although we value and respect the effort put forward by the UICC committee, in the light of the recently available data and based on our everyday experience, we fail to recognise those principles implemented in the UICC 8th ed. As reproduced in our study,

both clinical and pathological staging algorithms of UICC 8th ed have a higher or at least non-inferior prognostic value compared to other proposed and previously existing staging systems. As demonstrated in the literature and confirmed by our data, there is a ≥90% down-staging of HPV+ OPSCC with UICC 8th compared to 7th ed. Although we would like to de-intensify the treatment of those patients, the phase III treatment de-intensification trials for HPV+ OPSCC are ongoing, and the current treatment indications are still not based on HPV or any other risk-stratification introduced by the UICC 8th edition, which is based on the analyses of retrospective data of patients who were treated with conventional algorithms and guidelines. Currently, it is questionable, what kind of a practical consequence except for a more complex staging experience the UICC 8th ed, especially the pUICC will bring in everyday practice. The discrepancy between clinical and pathological staging algorithms is a known issue since the earlier versions of UICC[46]. However, with the UICC 8th ed, this gap due to intra-stage migration became vivid. In other words, even without gaining any new information through surgery (i.e. same post-operative pathological tumour and nodal features as clinically identified via imaging, endoscopy and manual examination), different clinical and pathological stages can be assigned to the very same patient. It is quite unconventional and counterintuitive to see a head and neck cancer patient not only differently staged but also differently processed for staging, only due to the chosen treatment modality. This poses potential concerns in diagnosis[41], treatment, research and data management.

Our study has its limitations. Despite being quite homogenous, the cohort carries the intrinsic potential source of bias due to its retrospective nature. Furthermore, during the development of the compared staging systems, different definitions for HPV status concerning p16 IHC and/or HPV DNA had been used, impairing the possibility to compare them.

In conclusion, with each of their strengths and weaknesses, the UICC 8th ed, MDACC and RTOG systems offer moderate efficacy in similar magnitudes as staging algorithms for patients diagnosed

with OPSCC undergoing surgical treatment. Building complex models only to achieve a marginal benefit is a futile effort. Beyond an acceptable risk stratification accuracy, pursuing practical utility should be the primary goal of developing future algorithms of the UICC, as the internationally accepted authority for cancer staging.

References

1. Bhatia A, Burtness B. Human Papillomavirus-Associated Oropharyngeal Cancer: Defining Risk Groups and Clinical Trials. *J. Clin. Oncol.* 2015;33(29):3243-50.

2. Dillon MT, Harrington KJ. Human Papillomavirus-Negative Pharyngeal Cancer. *J. Clin. Oncol.* 2015;33(29):3251-61.

3. Haughey BH, Sinha P. Prognostic factors and survival unique to surgically treated p16+ oropharyngeal cancer. *Laryngoscope* 2012;122 Suppl(SUPPL. 2):S13-33.

4. Broglie MA, Stoeckli SJ, Sauter R, et al. Impact of human papillomavirus on outcome in patients with oropharyngeal cancer treated with primary surgery. *Head Neck* 2017;39(10):2004-2015.

5. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N. Engl. J. Med.* 2010;363(1):24-35.

6. O'Sullivan B, Huang SH, Su J, et al. Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study. *Lancet. Oncol.* 2016;17(4):440-51.

7. Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A, eds. *AJCC Cancer Staging Manual*. 7th ed. New York: Springer; 2010.

8. Lydiatt WM, Patel SG, O'Sullivan B, et al. Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA. Cancer J.*

Clin. 2017;67(2):122-137.

9. Huang SH, Xu W, Waldron J, et al. Refining American joint committee on cancer/union for international cancer control TNM stage and prognostic groups for human papillomavirus-related oropharyngeal carcinomas. *J. Clin. Oncol.* 2015;33(8):836-845.

10. Haughey BH, Sinha P, Kallogjeri D, et al. Pathology-based staging for HPV-positive squamous carcinoma of the oropharynx. *Oral Oncol.* 2016;62:11-19.

11. Dahlstrom KR, Garden AS, William WJN, Lim MY, Sturgis EM. Proposed staging system for patients with hpv-related oropharyngeal cancer based on nasopharyngeal cancer n categories. *J. Clin. Oncol.* 2016;34(16):1848-1854.

12. Husain ZA, Chen T, Corso CD, et al. A Comparison of Prognostic Ability of Staging Systems for Human Papillomavirus-Related Oropharyngeal Squamous Cell Carcinoma. *JAMA Oncol.* 2017;3(3):358-365.

13. Rietbergen MM, Witte BI, Velazquez ER, et al. Different prognostic models for different patient populations: validation of a new prognostic model for patients with oropharyngeal cancer in Western Europe. *Br. J. Cancer* 2015;112(11):1733-1736.

14. Anantharaman D, Abedi-Ardekani B, Beachler DC, et al. Geographic heterogeneity in the prevalence of human papillomavirus in head and neck cancer. *Int. J. cancer* 2017;140(9):1968-1975.

15. Lucas-Roxburgh R, Benschop J, Lockett B, van den Heever U, Williams R, Howe L. The prevalence of human papillomavirus in oropharyngeal cancer in a New Zealand population. Tornesello ML, ed. *PLoS One* 2017;12(10):e0186424.

16. Hama T, Tokumaru Y, Fujii M, et al. Prevalence of human papillomavirus in oropharyngeal cancer: a multicenter study in Japan. *Oncology* 2014;87(3):173-82.

17. Würdemann N, Wagner S, Sharma SJ, et al. Prognostic Impact of AJCC/UICC 8th Edition New Staging Rules in Oropharyngeal Squamous Cell Carcinoma. *Front. Oncol.* 2017;7(June):129.

18. Combes JD, Chen AA, Franceschi S. Prevalence of human papillomavirus in cancer of the oropharynx by gender. *Cancer Epidemiol. Biomarkers Prev.* 2014;23(12):2954-2958.

19. Mizumachi T, Homma A, Sakashita T, Kano S, Hatakeyama H, Fukuda S. Confirmation of the eighth edition of the AJCC/UICC TNM staging system for HPV-mediated oropharyngeal cancer in Japan. *Int. J. Clin. Oncol.* 2017;22(4):682-689.

20. Rios Velazquez E, Hoebers F, Aerts HJWL, et al. Externally validated HPV-based prognostic nomogram for oropharyngeal carcinoma patients yields more accurate predictions than TNM staging. *Radiother. Oncol.* 2014;113(3):324-30.

21. Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N. Engl. J. Med.* 1985;312(25):1604-8.

22. Garnaes E, Frederiksen K, Kiss K, et al. Double positivity for HPV DNA/p16 in tonsillar and base of tongue cancer improves prognostication: Insights from a large population-based study. *Int. J. cancer* 2016;139(11):2598-605.

23. Rietbergen MM, Brakenhoff RH, Bloemena E, et al. Human papillomavirus detection and comorbidity: Critical issues in selection of patients with oropharyngeal cancer for treatment

De-escalation trials. *Ann. Oncol.* 2013;24(11):2740-2745.

24. Vainshtein JM, Spector ME, McHugh JB, et al. Refining risk stratification for locoregional failure after chemoradiotherapy in human papillomavirus-associated oropharyngeal cancer. *Oral Oncol.* 2014;50(5):513-519.

25. Spector ME, Gallagher KK, Bellile E, et al. Patterns of nodal metastasis and prognosis in human papillomavirus-positive oropharyngeal squamous cell carcinoma. *Head Neck* 2014;36(9):1233-40.

26. Fakhry C, Zhang Q, Nguyen-Tân PF, et al. Development and Validation of Nomograms Predictive of Overall and Progression-Free Survival in Patients With Oropharyngeal Cancer. *J. Clin. Oncol.* 2017;35(36):4057-4065.

27. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* 1996;15(4):361-87.

28. Maxwell JH, Ferris RL, Gooding W, et al. Extracapsular spread in head and neck carcinoma: impact of site and human papillomavirus status. *Cancer* 2013;119(18):3302-8.

29. Iyer NG, Dogan S, Palmer F, et al. Detailed Analysis of Clinicopathologic Factors Demonstrate Distinct Difference in Outcome and Prognostic Factors Between Surgically Treated HPV-Positive and Negative Oropharyngeal Cancer. *Ann. Surg. Oncol.* 2015;22(13):4411-21.

30. Sinha P, Lewis JS, Piccirillo JF, Kallogjeri D, Haughey BH. Extracapsular spread and adjuvant therapy in human papillomavirus-related, p16-positive oropharyngeal carcinoma.

Cancer 2012;118(14):3519-30.

31. Klozar J, Koslabova E, Kratochvil V, Salakova M, Tachezy R. Nodal status is not a prognostic factor in patients with HPV-positive oral/oropharyngeal tumors. *J. Surg. Oncol.* 2013;107(6):625-33.

32. Kaczmar JM, Tan KS, Heitjan DF, et al. HPV-related oropharyngeal cancer: Risk factors for treatment failure in patients managed with primary transoral robotic surgery. *Head Neck* 2016;38(1):59-65.

33. Lindel K, Beer KT, Laissue J, Greiner RH, Aebersold DM. Human papillomavirus positive squamous cell carcinoma of the oropharynx: a radiosensitive subgroup of head and neck carcinoma. *Cancer* 2001;92(4):805-13.

34. Chera BS, Amdur RJ. Current Status and Future Directions of Treatment Deintensification in Human Papilloma Virus-associated Oropharyngeal Squamous Cell Carcinoma. *Semin. Radiat. Oncol.* 2018;28(1):27-34.

35. Mermod M, Tolstonog G, Simon C, Monnier Y. Extracapsular spread in head and neck squamous cell carcinoma: A systematic review and meta-analysis. *Oral Oncol.* 2016;62:60-71.

36. van den Brekel MWM, Lodder WL, Stel H V, Bloemena E, Leemans CR, van der Waal I. Observer variation in the histopathologic assessment of extranodal tumor spread in lymph node metastases in the neck. *Head Neck* 2012;34(6):840-5.

37. Zhan KY, Eskander A, Kang SY, et al. Appraisal of the AJCC 8th edition pathologic staging modifications for HPV-positive oropharyngeal cancer, a study of the National Cancer Data

Base. *Oral Oncol.* 2017;73(September 2017):152-159.

38. Cramer JD, Hicks KE, Rademaker AW, Patel UA, Samant S. Validation of the eighth edition American Joint Committee on Cancer staging system for human papillomavirus-associated oropharyngeal cancer. *Head Neck* 2017;(September):1-10.

39. Malm I-J, Fan CJ, Yin LX, et al. Evaluation of proposed staging systems for human papillomavirus-related oropharyngeal squamous cell carcinoma. *Cancer* 2017;123(10):1768-1777.

40. Lohaus F, Linge A, Tinhofer I, et al. HPV16 DNA status is a strong prognosticator of loco-regional control after postoperative radiochemotherapy of locally advanced oropharyngeal carcinoma: results from a multicentre explorative study of the German Cancer Consortium Radiation Oncology Group (. *Radiother. Oncol.* 2014;113(3):317-23.

41. Boscolo-Rizzo P, Dietz A. The AJCC/UICC eighth edition for staging head and neck cancers: Is it wise to de-escalate treatment regimens in p16-positive oropharyngeal cancer patients? *Int. J. cancer* 2017;141(7):1490-1491.

42. Ghadjar P, Schreiber-Facklam H, Gräter R, et al. Quantitative analysis of extracapsular extension of metastatic lymph nodes and its significance in radiotherapy planning in head and neck squamous cell carcinoma. *Int. J. Radiat. Oncol. Biol. Phys.* 2010;76(4):1127-32.

43. Studer G, Glanzmann C. Volumetric stratification of cT4 stage head and neck cancer. *Strahlentherapie und Onkol.* 2013;189(10):867-873.

44. Elicin O, Schmucking M, Bromme J, et al. Volumetric regression ratio of the primary tumor and metastatic lymph nodes after induction chemotherapy predicts overall survival in head

and neck squamous cell carcinoma: a retrospective analysis. *J. BUON*. 2016;21(1):175-81. Available at: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L71624629.

45. Ho AS, Kim S, Tighiouart M, et al. Metastatic Lymph Node Burden and Survival in Oral Cavity Cancer. *J. Clin. Oncol.* 2017;35(31):JCO.2016.71.117.

46. Koch WM, Ridge JA, Forastiere A, Manola J. Comparison of clinical and pathological staging in head and neck squamous cell carcinoma: results from Intergroup Study ECOG 4393/RTOG 9614. *Arch. Otolaryngol. Head. Neck Surg.* 2009;135(9):851-8.

Published in final edited form as: *Head Neck*. 2019 May;41(5):1395-1402. doi: 10.1002/hed.25574

Table 1: Patient and disease characteristics

	Total	**HPV-**	**HPV+**	**p-value**
n (%)	338	175 (51.8)	163 (48.2)	
Gender: female (%)	95 (28.1)	44 (25.1)	51 (31.3)	0.256
Mean age (standard deviation)	68.4 (9.8)	67.8 (8.9)	69.2 (10.7)	0.196
Median age (interquartile range)	69 (62, 75)	69 (62, 72.5)	68 (62, 76)	0.341
Subsite (%)				<0.001
- Base of tongue	60 (17.8)	37 (21.1)	23 (14.1)	
- Pharyngeal wall	16 (4.7)	15 (8.6)	1 (0.6)	
- Soft palate	27 (8)	24 (13.7)	3 (1.8)	
- Tonsil	235 (69.5)	99 (56.6)	136 (83.4)	
Adjuvant treatment (%)	223 (66)	99 (56.6)	124 (76.1)	<0.001
Smoking > 10 py (%)	240 (71)	161 (92)	79 (48.5)	<0.001
Alcohol ≥3 units* (%)				<0.001
- Missing	38 (11.2)	14 (8)	24 (14.7)	
- No	177 (52.4)	72 (41.1)	105 (64.4)	
- Yes	123 (36.4)	89 (50.9)	34 (20.9)	
cECE+ (%)	55 (16.3)	21 (12)	34 (20.9)	0.040
pECE+ (%)	98 (29)	44 (25.1)	54 (33.1)	0.134

*: 3 dL beer or 1 dL wine or 4 cL spirit per day
cECE: clinical extracapsular extension; pECE: pathological extracapsular extension; py: pack-years

Table 2: Agreement of the clinical and pathological classifications within each UICC edition

7th edition	**Concordant**	**%**	**Up-staging**	**%**	**Down-staging**	**%**	**N total**
All	285	84.3	27	8.0	26	7.7	338
HPV+	137	84.0	12	7.4	14	8.6	163
HPV-	148	84.6	15	8.6	12	6.9	175
8th edition							
All	213	63.0	38	11.2	87	25.7	338
HPV+	85	52.1	4	2.5	74	45.4	163
HPV-	128	73.1	34	19.4	13	7.4	175

Table 3: Rates of stage migration from UICC 7th to 8th edition

Clinical	**Concordant**	**%**	**Up-staging**	**%**	**Down-staging**	**%**	**N total**
All	169	50.0	15	4.4	154	45.6	338
HPV+	10	6.1	0	0.0	153	93.9	163
HPV-	159	90.9	15	8.6	1	0.6	175
Pathological							
All	151	44.7	37	10.9	150	44.4	338
HPV+	13	8.0	0	0.0	150	92.0	163
HPV-	138	78.9	37	21.1	0	0.0	175

Table 4: C-statistics for overall and recurrence-free survival

Staging	Full cohort	HPV+	HPV-
Overall Survival			
cUICC 7th ed	0.45 (0.34 to 0.56)	0.58 (0.33 to 0.82)	0.56 (0.40 to 0.72)
pUICC 7th ed	0.46 (0.37 to 0.56)	0.48 (0.33 to 0.63)	0.59 (0.39 to 0.78)
cUICC 8th ed	0.62 (0.52 to 0.72)	0.58 (0.29 to 0.87)	0.57 (0.41 to 0.73)
pUICC 8th ed	0.64 (0.55 to 0.74)	0.55 (0.41 to 0.68)	0.60 (0.41 to 0.79)
RTOG	0.61 (0.52 to 0.69)	0.50 (0.44 to 0.56)	0.52 (0.48 to 0.57)
MDACC		0.61 (0.42 to 0.80)	
Yale		0.53 (0.37 to 0.69)	
Relapse-free Survival			
cUICC 7th ed	0.50 (0.43 to 0.57)	0.52 (0.36 to 0.68)	0.52 (0.42 to 0.63)
pUICC 7th ed	0.51 (0.43 to 0.59)	0.53 (0.37 to 0.70)	0.53 (0.42 to 0.63)
cUICC 8th ed	0.58 (0.52 to 0.64)	0.55 (0.37 to 0.74)	0.53 (0.41 to 0.66)
pUICC 8th ed	0.59 (0.53 to 0.65)	0.56 (0.49 to 0.64)	0.54 (0.41 to 0.67)
RTOG	0.58 (0.52 to 0.64)	0.49 (0.45 to 0.53)	0.51 (0.47 to 0.55)
MDACC		0.60 (0.46 to 0.73)	
Yale		0.54 (0.42 to 0.66)	

The 95% confidence intervals are shown in parentheses.

Table 5: Unadjusted and adjusted ordinal hazard ratios for death

Staging	Hazard ratio	p	Adjusted hazard ratio	p
Full Cohort				
cUICC 7th ed	0.99 (0.81 to 1.21)	0.956	0.98 (0.79 to 1.20)	0.820
pUICC 7th ed	0.99 (0.82 to 1.20)	0.920	0.95 (0.76 to 1.18)	0.627
cUICC 8th ed	1.42 (1.19 to 1.68)	<0.001	1.29 (1.07 to 1.55)	0.008
pUICC 8th ed	1.40 (1.22 to 1.61)	<0.001	1.33 (1.13 to 1.56)	<0.001
RTOG	1.68 (1.30 to 2.17)	<0.001	1.79 (1.26 to 2.55)	0.001
HPV+				
cUICC 7th ed	1.15 (0.71 to 1.87)	0.562	1.14 (0.69 to 1.87)	0.608
pUICC 7th ed	0.93 (0.61 to 1.44)	0.754	0.99 (0.61 to 1.61)	0.963
cUICC 8th ed	1.20 (0.57 to 2.52)	0.638	1.08 (0.52 to 2.21)	0.842
pUICC 8th ed	1.59 (0.70 to 3.59)	0.266	1.66 (0.72 to 3.85)	0.237
RTOG	1.32 (0.48 to 3.63)	0.592	0.85 (0.28 to 2.55)	0.773
MDACC	1.47 (0.89 to 2.42)	0.128	1.30 (0.77 to 2.17)	0.324
Yale	1.24 (0.76 to 2.02)	0.381	1.22 (0.76 to 1.94)	0.408
HPV-				
cUICC 7th ed	1.12 (0.89 to 1.40)	0.344	0.99 (0.77 to 1.26)	0.914
pUICC 7th ed	1.15 (0.93 to 1.43)	0.198	1.01 (0.79 to 1.31)	0.921
cUICC 8th ed	1.15 (0.93 to 1.42)	0.188	1.15 (0.93 to 1.42)	0.197
pUICC 8th ed	1.19 (0.98 to 1.43)	0.075	1.03 (0.79 to 1.34)	0.829
RTOG	0.57 (0.26 to 1.25)	0.163	0.55 (0.25 to 1.21)	0.138

The 95% confidence intervals are shown in parentheses.

Table 6: Unadjusted and adjusted ordinal hazard ratios for recurrence or death

Staging	Hazard ratio	p	Adjusted hazard ratio	p
Full Cohort				
cUICC 7th ed	0.93 (0.78 to 1.10)	0.393	0.90 (0.75 to 1.07)	0.221
pUICC 7th ed	0.90 (0.76 to 1.06)	0.202	0.85 (0.70 to 1.03)	0.092
cUICC 8th ed	1.29 (1.11 to 1.50)	0.001	1.19 (1.01 to 1.39)	0.035
pUICC 8th ed	1.27 (1.12 to 1.43)	<0.001	1.22 (1.06 to 1.41)	0.005
RTOG	1.51 (1.22 to 1.87)	<0.001	1.66 (1.23 to 2.25)	0.001
HPV+				
cUICC 7th ed	1.05 (0.73 to 1.50)	0.797	1.03 (0.72 to 1.48)	0.865
pUICC 7th ed	0.88 (0.63 to 1.22)	0.435	0.85 (0.59 to 1.22)	0.381
cUICC 8th ed	1.51 (0.82 to 2.77)	0.189	1.37 (0.76 to 2.46)	0.301
pUICC 8th ed	1.99 (1.11 to 3.56)	0.020	1.99 (1.09 to 3.64)	0.025
RTOG	1.04 (0.45 to 2.40)	0.923	0.88 (0.34 to 2.27)	0.792
MDACC	1.57 (1.05 to 2.37)	0.030	1.44 (0.95 to 2.18)	0.082
Yale	1.50 (1.03 to 2.19)	0.036	1.42 (0.98 to 2.06)	0.062
HPV-				
cUICC 7th ed	1.00 (0.82 to 1.23)	0.968	0.89 (0.71 to 1.11)	0.299
pUICC 7th ed	1.00 (0.83 to 1.22)	0.959	0.90 (0.72 to 1.14)	0.379
cUICC 8th ed	1.06 (0.87 to 1.28)	0.571	1.06 (0.88 to 1.28)	0.551
pUICC 8th ed	1.05 (0.89 to 1.25)	0.547	0.92 (0.72 to 1.16)	0.470
RTOG	0.79 (0.36 to 1.73)	0.561	0.76 (0.35 to 1.66)	0.497

The 95% confidence intervals are shown in parentheses.