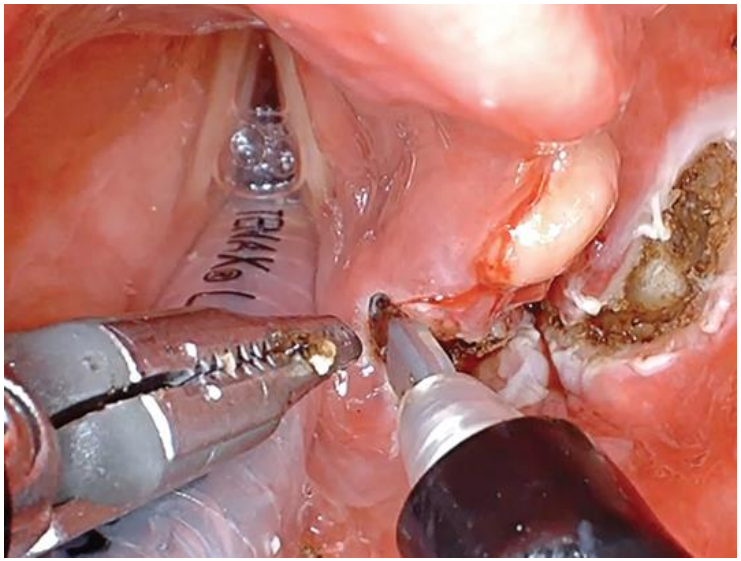
Head and Neck Cancer Survival Analysis: McGill FDG-PET/CT Database



DATA 623: Final Presentation

Marc McCoy (ID: 30136987)

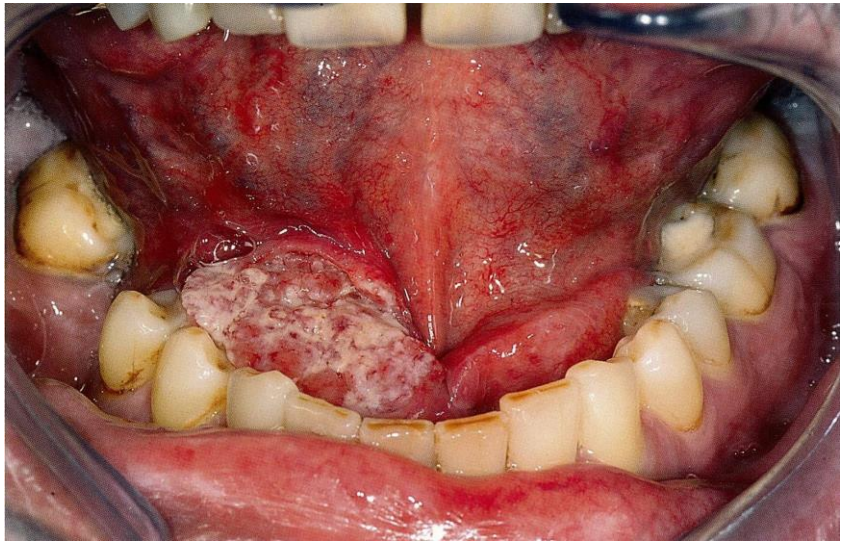
April 2022

Background/Context

What is Oral Cancer

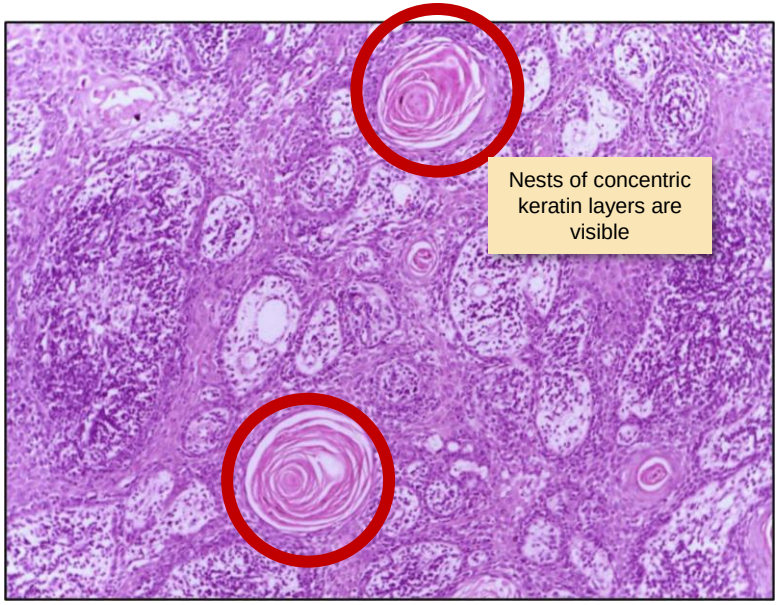
Clinical Presentation

- Oral cavity cancers refer to **malignant tumors** of the oral mucosa, tonsils, and salivary glands
- Predisposing factors include smoking, oral tobacco consumption, long term alcohol use, and human papilloma virus infection
- Oral cavity cancers usually present in males, aged 55–60 years, with clinical features like **pain, dysphagia, or a nonhealing ulcer on the tonsils, tongue, or oral mucosa**



Diagnosis

- Biopsy and histopathology of the lesion
- Panendoscopy: assessment of tumor extent
- HPV testing
- Chest x-ray, axial CT: assess tumor spread in solid organs; screen for lymph node and bone metastases
- PET-CT



Background/Context

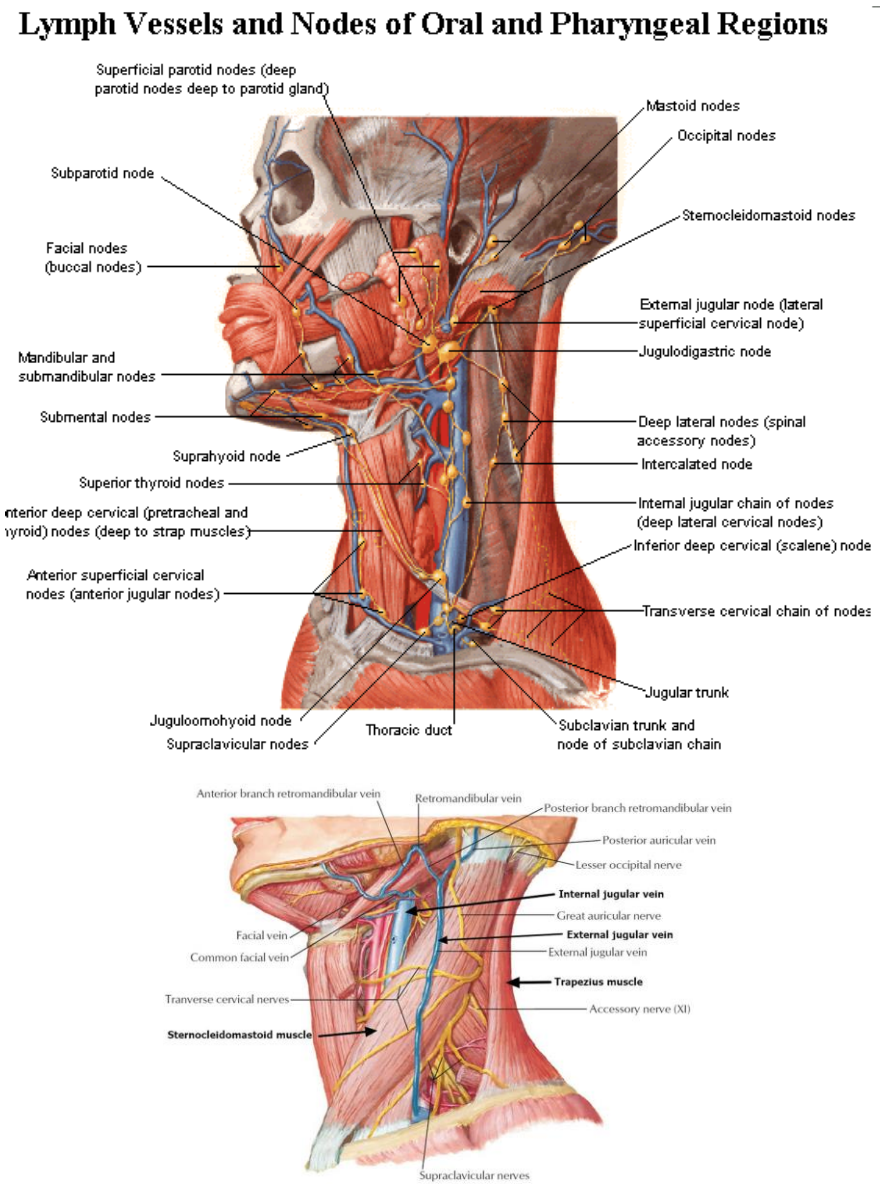
What is Oral Cancer

Epidemiology and Risk Factors

- Peak incidence: 55–60 years
- Sex: ♂ > ♀ 2:1
- Oral tobacco consumption and smoking
- Long-term alcohol consumption
- Poor oral hygiene, chronic mechanical irritation (e.g., badly positioned dentures)
- Human papillomavirus, particularly HPV 16, 18, and 31

Treatment

- Treatment depends upon the stage of the tumor and the extent of its spread, and may include **surgical resection (usually with neck dissection), radiation therapy, and/or chemotherapy**.
- Early diagnosis and treatment usually result in a good curative rate
- HPV-positive tumors have a good prognosis since they respond better to chemo- and/or radiotherapy



Background on the Problem

State of the Science

Prognostic Modeling is a Growing Field

- Prognosis is a huge issue because oral cancer has a high mortality rate
- **Should patients still be treated aggressively if they are likely to die due to their cancer?**
- What chemotherapy drugs should be used… there are thousands
- **Cancers are heterogenous** (same biomarkers does not mean same cancer behavior)
- Quality of life maximization
- Risk stratification tools can **guide the intensity of a treatment regimen, ensure accurate prognosis**
- Aid in patient communication about treatment options
- Other possibilities include optimized chemotherapy treatment and precision health development

JAMA Network Open

Original Investigation | Oncology

Development of a Machine Learning Model for Survival Risk Stratification of Patients With Advanced Oral Cancer

Yi-Ju Tseng, PhD; Hsin-Yao Wang, MD; Ting-Wei Lin, MD; Jang-Jih Lu, MD, PhD; Chia-Hsun Hsieh, MD, PhD; Chun-Ta Liao, MD

Abstract

IMPORTANCE A tool for precisely stratifying postoperative patients with advanced oral cancer is crucial for the treatment plan, such as intensifying or deintensifying the regimen to improve their quality of life and prognosis.

OBJECTIVE To develop and validate a machine learning–based algorithm that can provide survival risk stratification for patients with advanced oral cancer who have comprehensive clinicopathologic and genetic data.

DESIGN, SETTING, AND PARTICIPANTS In this prognostic cohort study, the elastic net penalized Cox proportional hazards regression–based risk stratification model was developed and validated using single-center data collected between January 1, 1996, and December 31, 2011. In total, comprehensive clinicopathologic and genetic data (including clinical, pathologic, and 44 cancer-related gene variant profiles) of 334 patients with stage III or IV oral squamous cell carcinoma were used to develop and validate the algorithm in this 15-year cohort study. Data analysis was conducted between February 1, 2018, and May 6, 2020.

MAIN OUTCOMES AND MEASURES The main outcomes were cancer-specific survival, distant metastasis–free survival, and locoregional recurrence–free survival. Model performance was compared in terms of the Akaike information criterion and the Harrell concordance index (C index).

RESULTS Complete data were available for 334 patients (315 men; median age at onset, 48 years [interquartile range, 42-56 years]). The predictive models using comprehensive clinicopathologic and genetic data outperformed those using clinicopathologic data alone. In the groups of postoperative patients receiving adjuvant concurrent chemoradiotherapy, the models demonstrated higher classification performance than those using clinicopathologic data alone in cancer-specific survival (mean [SD] C index, 0.689 [0.050] vs 0.673 [0.051]; $P = .02$) and locoregional recurrence–free survival (mean [SD] C index, 0.693 [0.039] vs 0.678 [0.035]; $P = .004$). The classification performance in distant metastasis–free survival was not different (mean [SD] C index, 0.702 [0.056] vs 0.688 [0.048]; $P = .09$).

Key Points

Question Can a machine learning model provide survival risk stratification for patients with advanced oral cancer who have comprehensive clinicopathologic and genetic data?

Findings In this 15-year cohort study of 334 patients, a risk stratification model using comprehensive clinicopathologic and genetic data accurately differentiated the high-risk group from the low-risk group in postoperative cancer-specific and locoregional recurrence–free survival for patients with advanced oral cancer.

Meaning The proposed model demonstrated good discrimination in stratifying patients with different risks of survival by using comprehensive clinicopathologic and genetic data, which can provide additional personalized information for postoperative management of patients with advanced oral squamous cancer.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Background/Context

What is the State of the Science

Metabolic Tumor Volume = **Gross Tumor Volume** × **SUV_{max}**

Significant (HR > 1) | Significant (HR > 1) | Not significant (HR=1)

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Validation of metabolic tumor volume as a prognostic factor for oral cavity squamous cell carcinoma treated with primary surgery ☆

Han Zhang [a], Hadi Seikaly [a], Nhu-Tram Nguyen [b], Jonathan T. Abele [c], Peter T. Dziegielewski [d], Jeffrey R. Harris [a], Daniel A. O'Connell [a,*]

[a] Division of Otolaryngology—Head and Neck Surgery, University of Alberta, Edmonton, Alberta, Canada
[b] Division of Radiation Oncology, McMaster University, Hamilton, Ontario, Canada
[c] Department of Radiology, University of Alberta, Edmonton, Alberta, Canada
[d] Department of Otolaryngology—Head and Neck Surgery, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Article history:
Received 17 December 2015
Received in revised form 14 March 2016
Accepted 15 March 2016
Available online 31 March 2016

Keywords:
Oral cavity squamous cell carcinoma
Prognostic marker
Metabolic tumor volume
Pet-ct
Imaging

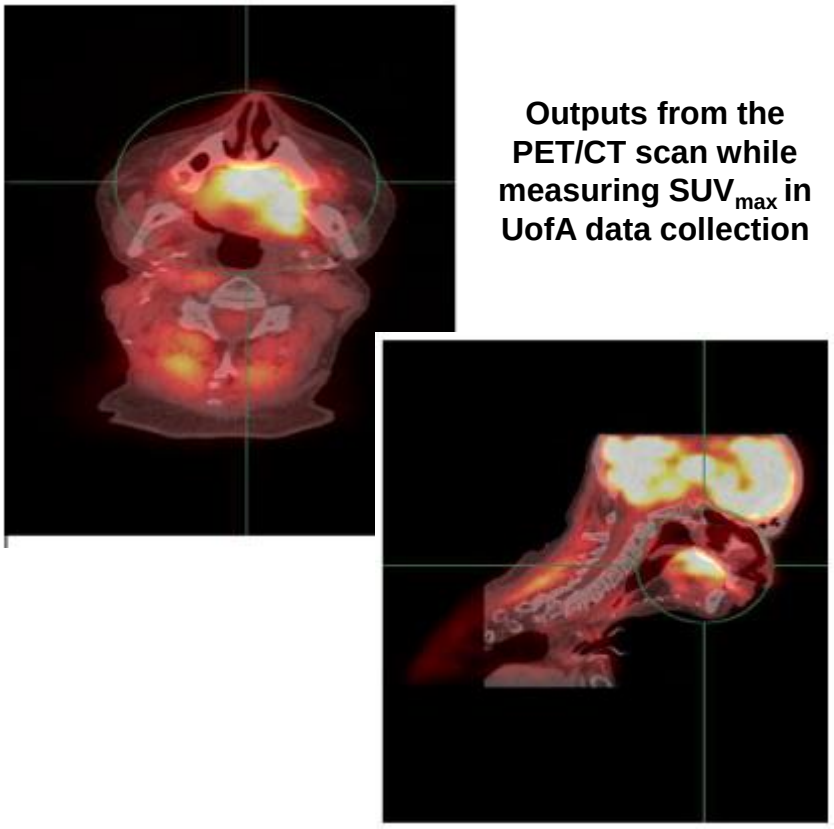
SUMMARY

Background: Despite the promise of metabolic tumor volume (MTV) as a risk-stratifying marker, the retrospective design of the initial study limits its generalizability. Therefore, this study sought to validate MTV as a prognostic factor for oral cavity squamous cell carcinoma (OCSCC) treated with primary surgery within an independent data set.

Methods: The validation data set consisted of 42 patients diagnosed with OCSCC between 2008 and 2012. The original cohort consisted of 80 patients. MTV and SUVmax were calculated for the primary tumor and nodal metastasis separately, as well as combined. Before statistical analysis, MTV and SUVmax values were divided into intertertile thirds to allow for intergroup survival analysis. Validation analysis was conducted on the validation data set alone. Data from both cohorts were then combined ($n = 122$) to increase statistical power.

Results: An increase in combined MTV of 17.5 cm^3 was associated with statistically significant increase in risk of disease recurrence (HR = 19.2, $p < 0.001$) and death (HR = 9.2, $p < 0.05$). Combined SUVmax failed to predict overall (HR = 1.0, $p > 0.05$) and disease-free survival (HR = 1.0, $p > 0.05$). Increase in the MTV of the primary tumor was associated with an increase in the risk of disease recurrence (HR = 21.7, $p = 0.0001$) and risk of death (HR = 7.0, $p = 0.0001$), while increase in the MTV of the locoregional neck metastasis was not ($p > 0.05$). An MTV cutoff value of greater than 10.2 cm^3 was found to significantly affect survival.

Conclusion: Due to the reproducibility of MTV findings, this study validates MTV as an independent prognostic factor for OCSCC treated with primary surgery.



Outputs from the PET/CT scan while measuring SUV_{max} in UofA data collection

Research Question

Two Primary Analysis Topics

Research Question

- **Stratified survival analysis of oropharynx, nasopharynx, hypopharynx and larynx** with the following primary outcomes:
  - Death
- **Stratified survival analysis of HPV positive vs. HPV negative oropharynx, nasopharynx, and hypopharynx cancer** with the following primary outcomes:
  - Death
  - Locoregional metastasis
  - Distant metastasis

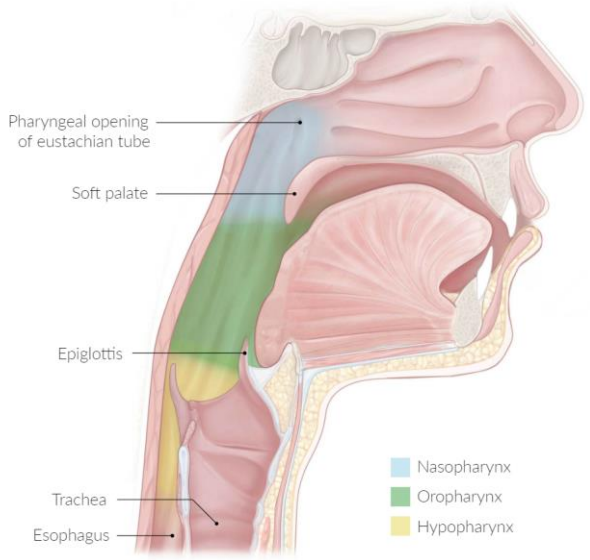


Table 1. Tumor–Node–Metastasis Classification of Human Papillomavirus (HPV)–Positive and HPV-Negative Oropharyngeal Cancer.*

Classification	HPV-Positive Oropharyngeal Cancer	HPV-Negative Oropharyngeal Cancer
Tumor		
TX	Primary tumor cannot be assessed	Primary tumor cannot be assessed
Tis	Carcinoma in situ	Carcinoma in situ
T0	No tumor identified	No tumor identified
T1	Tumor <2 cm in greatest dimension	Tumor <2 cm in greatest dimension
T2	Tumor >2 cm but <4 cm in greatest dimension	Tumor >2 cm but <4 cm in greatest dimension
T3	Tumor >4 cm in greatest dimension or extension to lingual surface of epiglottis	Tumor >4 cm in greatest dimension or extension to lingual surface of epiglottis
T4	Moderately advanced local disease; tumor invades larynx, extrinsic muscle of tongue, medial pterygoid muscle, hard palate or mandible, or beyond†	
T4a		Moderately advanced local disease; tumor invades larynx, extrinsic muscle of tongue, medial pterygoid muscle, hard palate, or mandible†
T4b		Very advanced local disease; tumor invades lateral pterygoid muscle, pterygoid plates, lateral nasopharynx, or skull base or encases carotid artery
Node		
Nx	Regional lymph nodes cannot be assessed	Regional lymph nodes cannot be assessed
N0	No regional lymph-node metastases	No regional lymph-node metastases
N1	Metastases to 1 or more ipsilateral lymph nodes, none >6 cm in greatest dimension	Metastasis to a single ipsilateral lymph node, ≤3 cm in greatest dimension, without extranodal extension
N2	Metastases to contralateral or bilateral lymph nodes, none >6 cm in greatest dimension	
N2a		Metastasis to a single ipsilateral node, >3 cm but <6 cm in greatest dimension, without extranodal extension
N2b		Metastases to multiple ipsilateral lymph nodes, none >6 cm in greatest dimension, without extranodal extension
N2c		Metastases to bilateral or contralateral lymph nodes, none >6 cm in greatest dimension, without extranodal extension
N3	Metastases to one or more lymph nodes, >6 cm in greatest dimension	
N3a		Metastasis to a lymph node, >6 cm in greatest dimension, without extranodal extension
N3b		Metastases to one or more lymph nodes, with clinically overt extranodal extension
Metastasis		
M0	No distant metastases	No distant metastases
M1	Distant metastases	Distant metastases

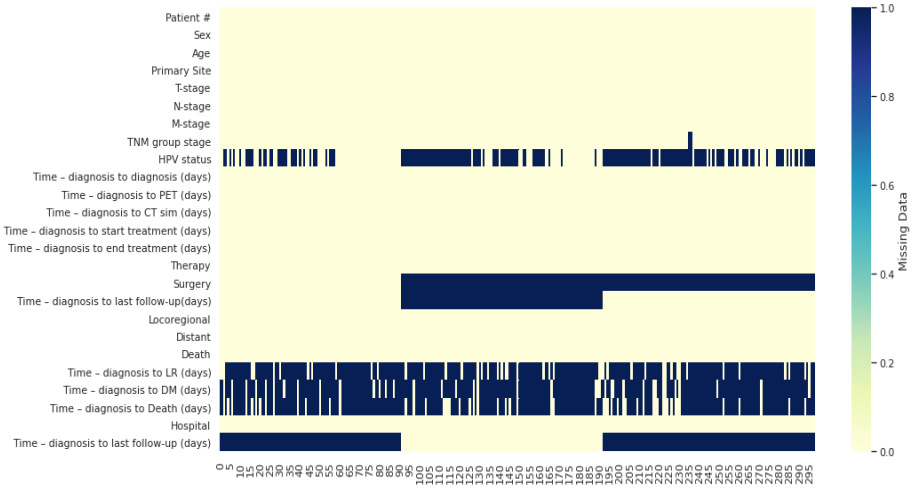
* Shown is the tumor–node–metastasis (TNM) classification of oropharyngeal tumors issued by the American Joint Commission on Cancer and the Union for International Cancer Control, 8th edition.[9,13]
† Mucosal extension of primary tumors of the base of the tongue and vallecula to the lingual surface of the epiglottis does not constitute invasion of the larynx.

Methods/Techniques

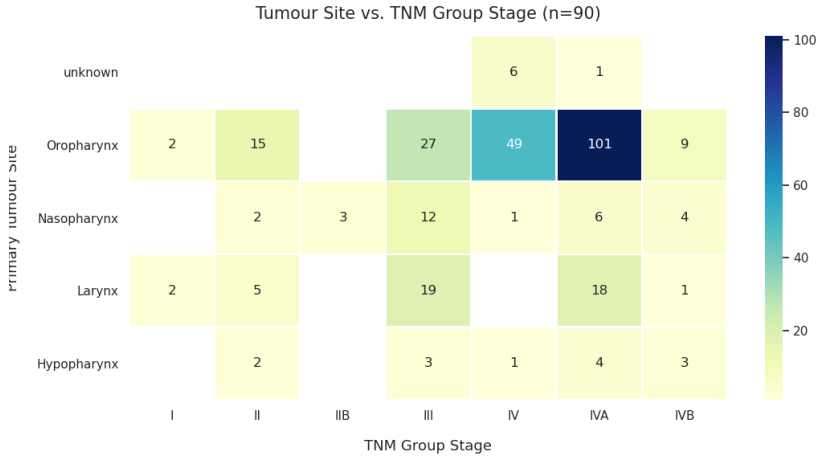
The Data was not Clean or Well-Distributed

- There was considerable missing data, as data was collected across four different hospitals in Quebec
- Depending on if patients had recorded locoregional metastasis, distant metastasis, or if they died, there was potentially **missing data for time to a diagnosis that they did not have**
- **A significant number of patients were "right-censored"** and I placed the "study end date" at the seven year mark (*this is not a "technically" perfect way of doing this)
- All time periods were measured in days

Considerable Missing Data



Sample Distribution (n=298) Heatmap

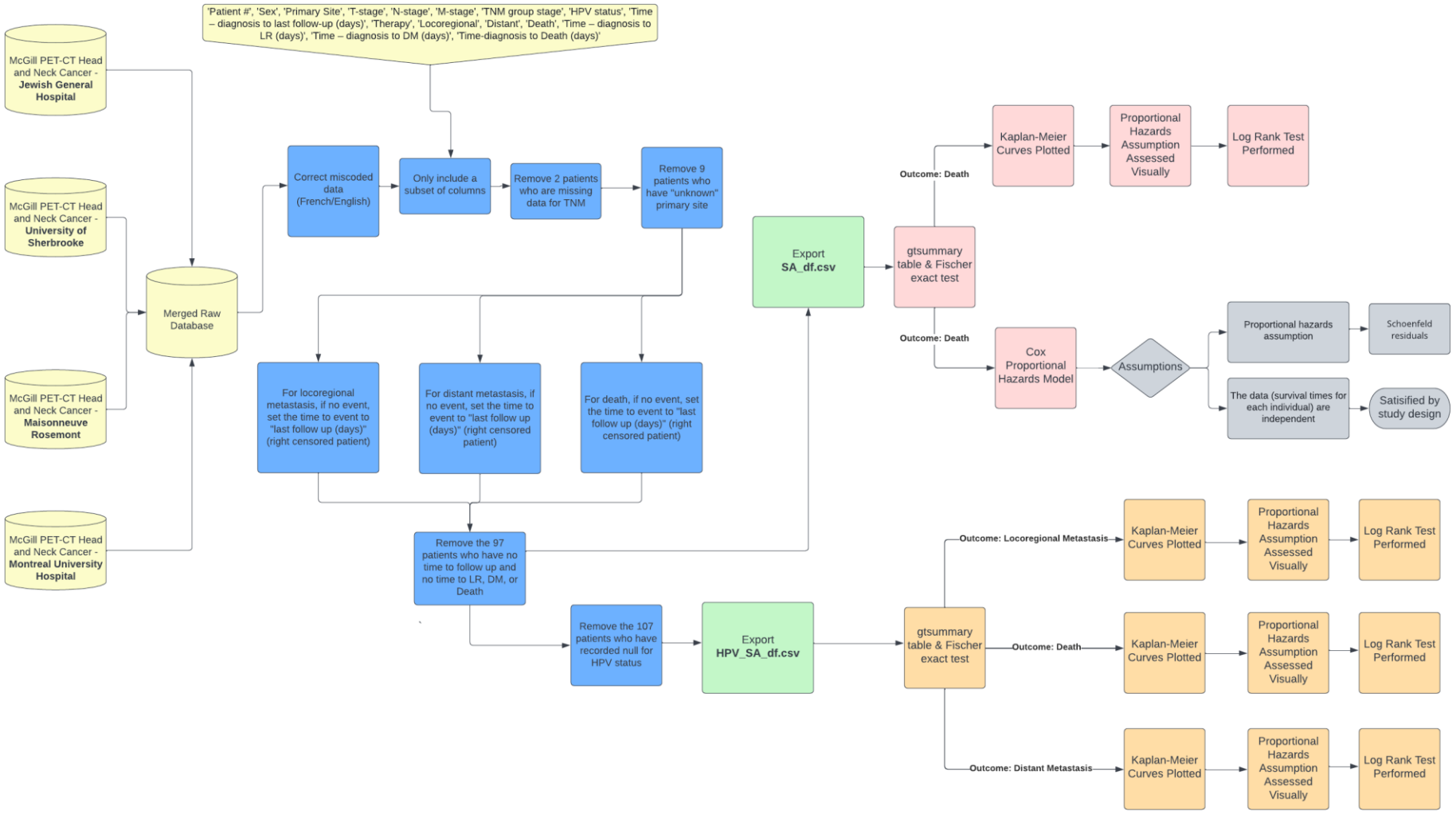


- Very few patients (n=11) have low grade cancer (stage I or stage II based on the TNM staging guide)
- Most patients have stage IVA oropharynx cancer (n=101)
- Any cancer that is stage III or stage IV is considered to be "*late-stage*" and is frequently analyzed together in the current medical literature for OCSCC
- I made the decision to only use late stage tumours in my analysis because of the results of a Fischer's exact test

Methods/Techniques

Flowchart

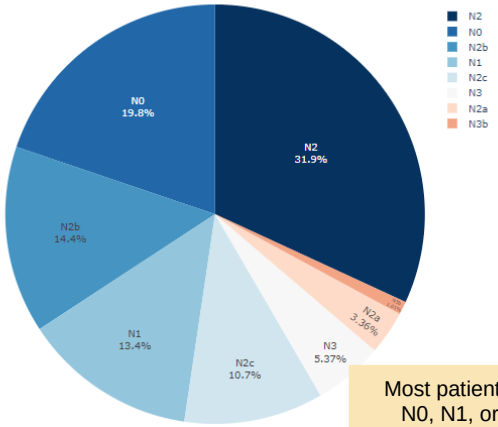
Analysis Steps



Results

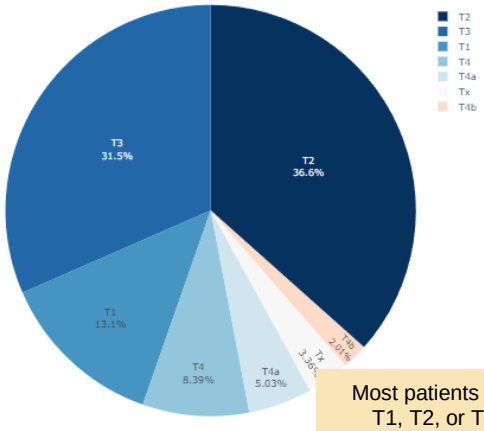
Descriptive

N-Stage (n=298)



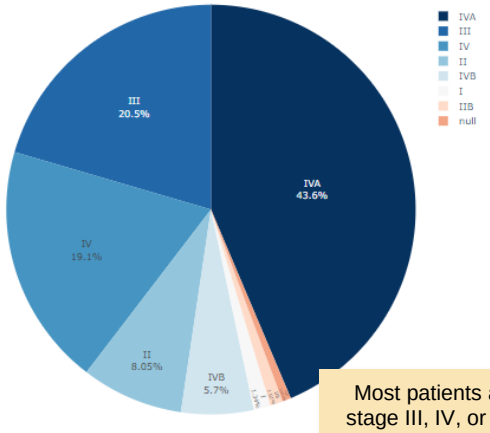
Most patients are N0, N1, or N2

T-Stage (n=298)



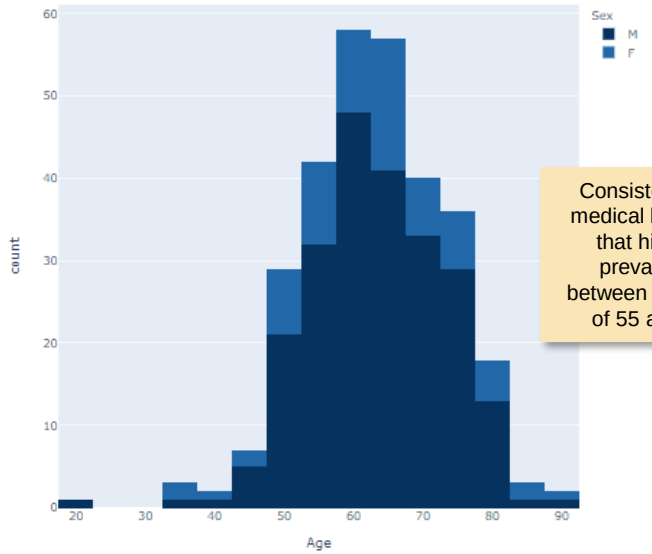
Most patients are T1, T2, or T3

TNM Stage (n=298)



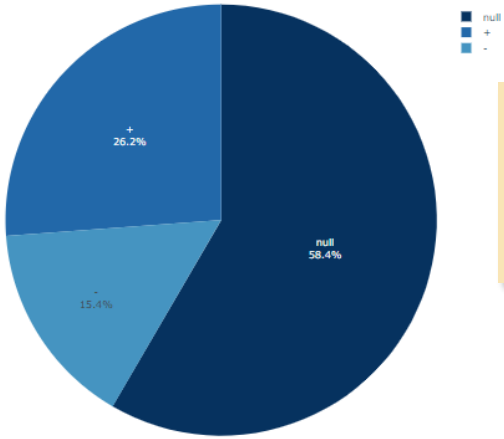
Most patients are stage III, IV, or IVA

Histogram of Age (Segregated by Sex) (n=298)



Consistent with medical literature that highest prevalence between the ages of 55 and 60

HPV Status (n=298) – Not Recorded At One Center

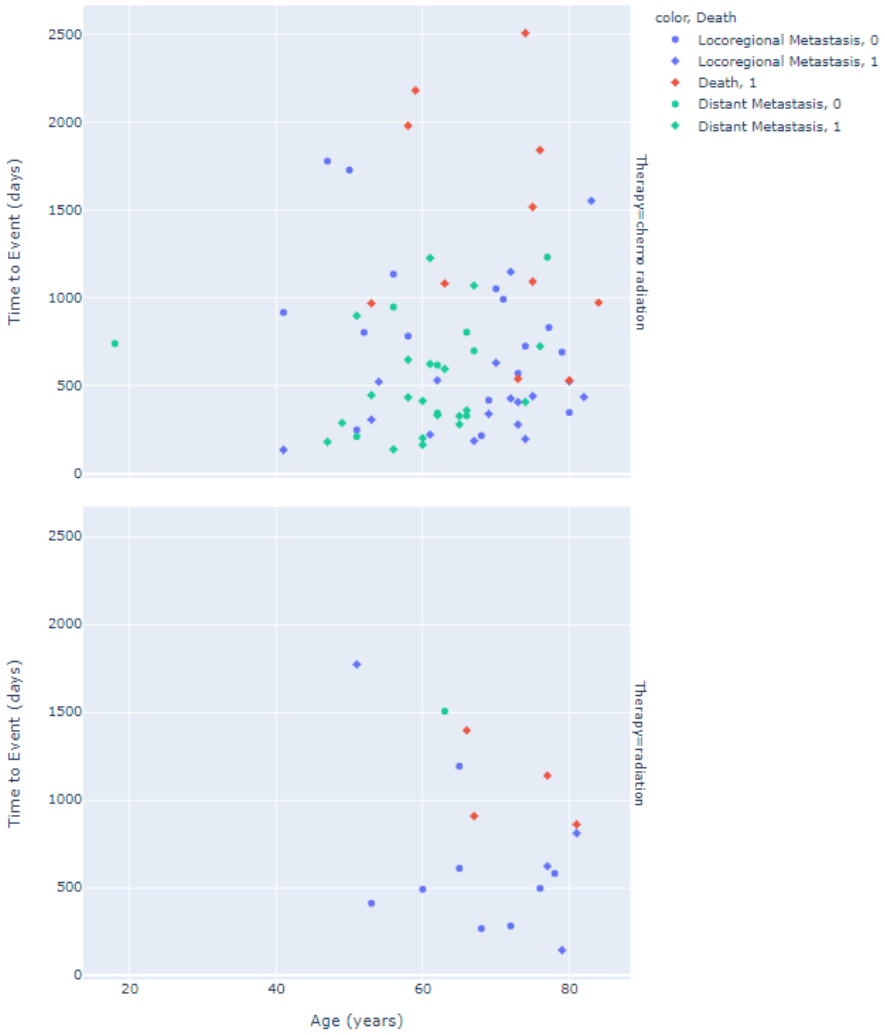


Large number of missing patients in terms of HPV status led to a smaller sample size in the secondary analysis (n=83)

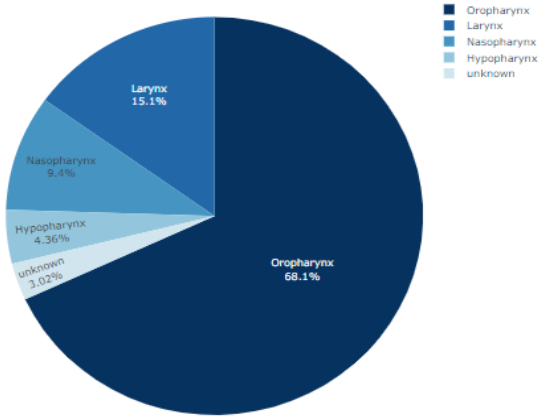
Results

Descriptive

Scatterplot (Time to Event vs. Age) by Therapy

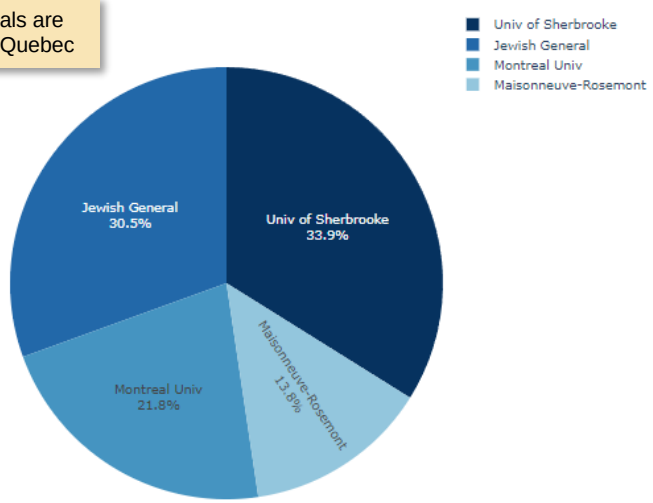


Anatomic Location of Cancerous Lesion (n=298)



Hospital Where Data Was Collected (n=298)

All hospitals are located in Quebec



Results

Statistical Tests

Fisher's Exact Test

- Analogous to the chi-square test, the Fisher exact test is a nonparametric test for categorical data but can be used in situations in which the chi-square test cannot, such as with small sample sizes
- Determine if there are non-random associations between two categorical variables
  - H0: the true odds ratio is equal to 1
  - HA: the true odds ratio is not equal to 1

	Column 1	Column 2	Total
Row 1	a	b	a+b
Row 2	c	d	c+d
Total	a+c	b+d	N = a+b+c+d

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!}$$

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

Results

Variable	N	Early Stage, N = 11[1]	Late Stage, N = 179[1]	p-value[2]	
Sex	190			0.5	OR = 1
F		1 (9.1%)	38 (21%)		
M		10 (91%)	141 (79%)		
Primary.Site	190			0.058	OR = 1
Hypopharynx		1 (9.1%)	11 (6.1%)		
Larynx		0 (0%)	23 (13%)		
Nasopharynx		4 (36%)	19 (11%)		
Oropharynx		6 (55%)	126 (70%)		
T.stage	190			0.021	OR ≠ 1
T1		5 (45%)	25 (14%)		
T2		6 (55%)	60 (34%)		
T3		0 (0%)	64 (36%)		
T4		0 (0%)	23 (13%)		
T4a		0 (0%)	3 (1.7%)		
T4b		0 (0%)	3 (1.7%)		
Tx		0 (0%)	1 (0.6%)		
HPV.status	190			>0.9	OR = 1
		6 (55%)	101 (56%)		
-		2 (18%)	29 (16%)		
+		3 (27%)	49 (27%)		
Therapy	190			<0.001	OR ≠ 1
chemo radiation		4 (36%)	169 (94%)		
radiation		7 (64%)	10 (5.6%)		

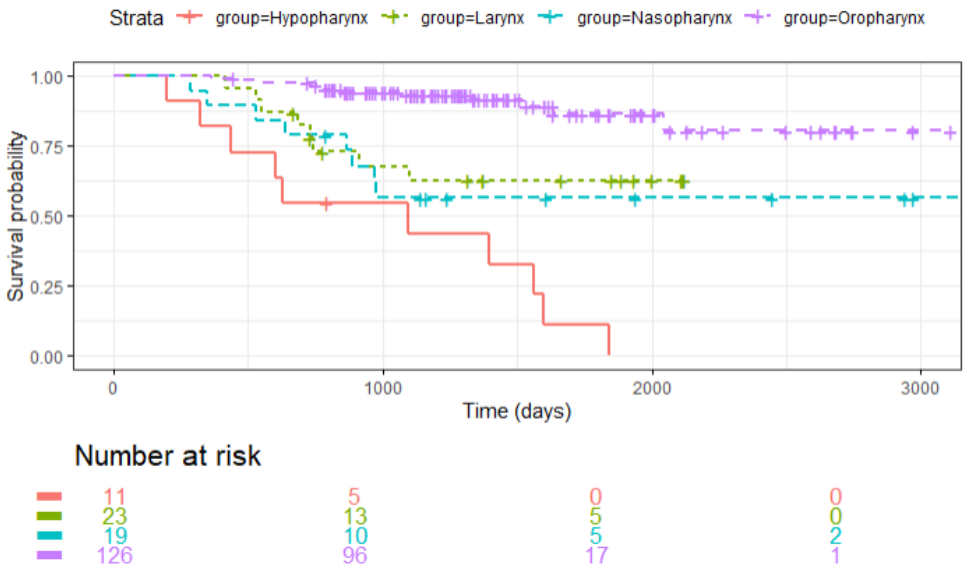
[1] n (%)

[2] Fisher's exact test

Results

Primary Analysis

Survival Analysis



- H_0: the two survival distributions are identical
- H_A: the two survival distributions are not identical

Proportional Hazards Assumption

- For all values of t, $S1(t) = S2(t)^k$
  - H_0: k = 1
  - H_A: k ≠ 1

- visual inspection of the two curves (do the survival curves cross or not?)

R-Outputs

```
Call:
survdiff(formula = Surv(time, status) ~ group, data = LR_df)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
group=Hypopharynx  11       10     1.23      62.4      66.5
group=Oropharynx  126       12    20.77       3.7      66.5

 Chisq= 66.5  on 1 degrees of freedom, p= 3e-16
```

```
Call:
survdiff(formula = Surv(time, status) ~ group, data = LR_df)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
group=Hypopharynx 11       10     5.49      3.71      5.41
group=Nasopharynx 19        8    12.51      1.63      5.41

 Chisq= 5.4  on 1 degrees of freedom, p= 0.02
```

```
Call:
survdiff(formula = Surv(time, status) ~ group, data = LR_df)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
group=Larynx       23        8      2.9      8.97      10.6
group=Oropharynx 126       12     17.1      1.52      10.6

 Chisq= 10.6  on 1 degrees of freedom, p= 0.001
```

```
Call:
survdiff(formula = Surv(time, status) ~ group, data = LR_df)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
group=Hypopharynx 11       10     4.68      6.06      8.29
group=Larynx       23        8    13.32      2.13      8.29

 Chisq= 8.3  on 1 degrees of freedom, p= 0.004
```

```
Call:
survdiff(formula = Surv(time, status) ~ group, data = LR_df)

                    N Observed Expected (O-E)^2/E (O-E)^2/V
group=Nasopharynx  19        8     2.52     11.91      13.8
group=Oropharynx  126       12    17.48      1.72      13.8

 Chisq= 13.8  on 1 degrees of freedom, p= 2e-04
```

Results

Primary Analysis

Cox Proportional Hazards Model

$$h(t) = h_0(t) \cdot e^{\beta_1 x}$$

$$HR = \frac{h(t|x=1)}{h(t|x=0)} = \frac{h_0(t) \cdot e^{\beta_1}}{h_0(t)} = e^{\beta_1}$$

- **Ties handled with the Efron method**

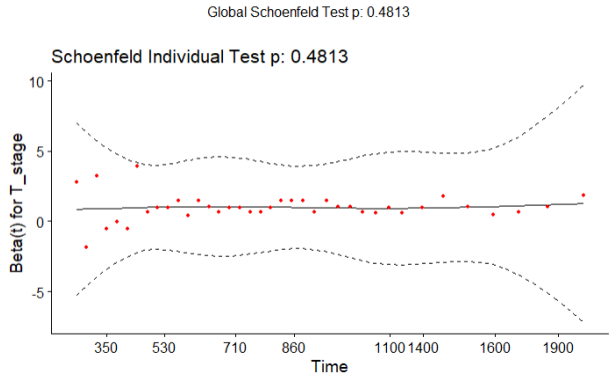
R-Outputs

```
Call:
coxph(formula = survobj ~ T_stage, data = final_df, method = "efron")

  n= 179, number of events= 38

               coef  exp(coef)   se(coef)      z Pr(>|z|)
T_stageT2  2.014e-01  1.223e+00  3.841e-01  0.524   0.6001
T_stageT3  5.671e-01  1.763e+00  3.405e-01  1.665   0.0959 .
T_stageT4  1.054e+00  2.868e+00  4.206e-01  2.505   0.0122 *
T_stageT4a 1.412e+00  4.104e+00  7.304e-01  1.933   0.0532 .
T_stageT4b 5.007e+00  1.495e+02  7.807e-01  6.414 1.42e-10 ***
T_stageTx -1.375e+01  1.067e-06  2.426e+03 -0.006   0.9955
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model Assumptions

1. The data (survival times for each individual) are independent

2. The hazard ratio is constant over time (proportional hazards assumption)



Interpretation

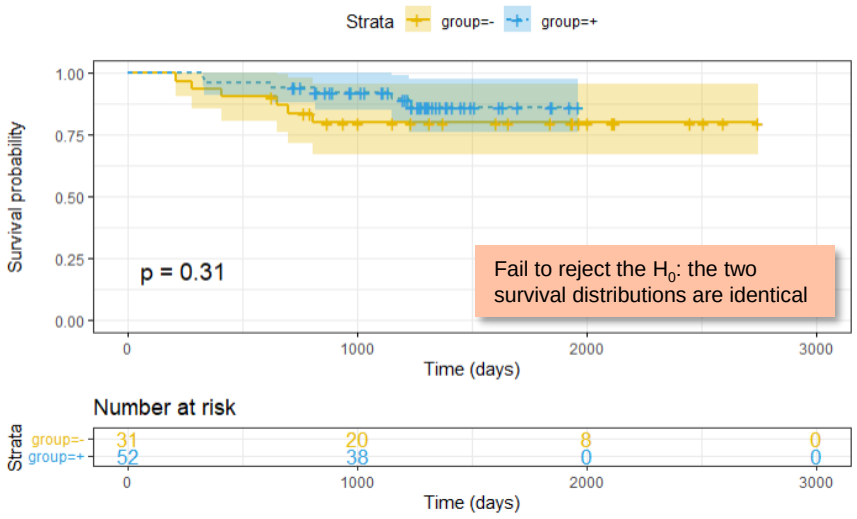
- 1.054 is the estimated change in the log hazard ratio for patients with stage T4 vs. patients with stage T1
- 2.868 is the hazard ratio for patients with stage T4 vs. patients with stage T1
- The hazard of death in patients with stage T4 is 2.868x the hazard of death for those with stage T1
- Aids in **patient communication** is the key takeaway

Stage	Explanation
	Primary Tumor (T)
Tx	Primary tumor cannot be assessed
T0	No evidence of tumor
Tis	Carcinoma in situ
T1	Tumor ≤ 2 cm in greatest dimension
T2	Tumor > 2 cm but < 4 cm in greatest dimension
T3	Tumor > 4 cm in greatest dimension
T4	Tumor invades adjacent structures (mandible, tongue musculature, maxillary sinus, skin)

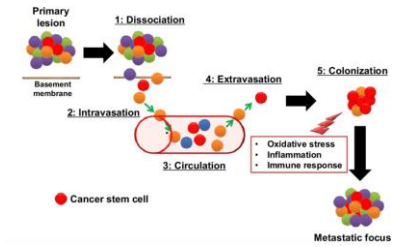
Results

Secondary Analysis

Distant Metastasis

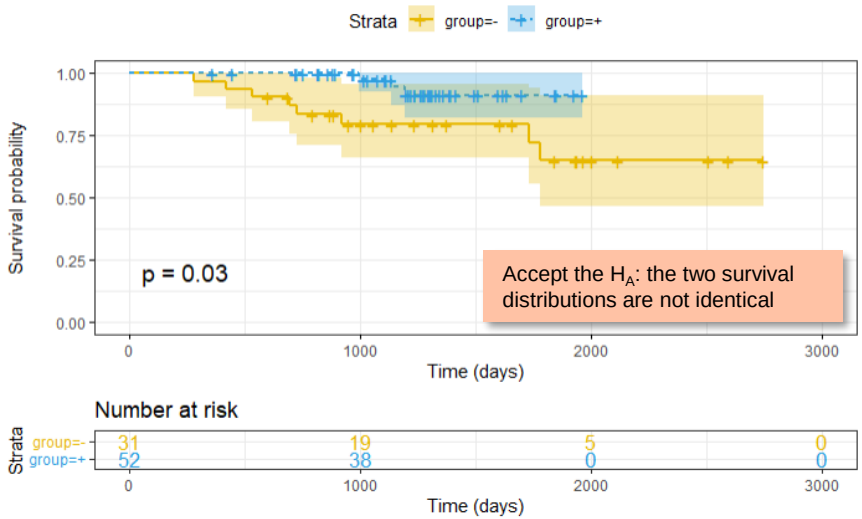



```
Call:
survdiff(formula = Surv(time, status) ~ group, data = df_final)

          N Observed Expected (O-E)^2/E (O-E)^2/V
group=- 31        6     4.31     0.664      1.04
group=+ 52        6     7.69     0.372      1.04

 Chisq= 1  on 1 degrees of freedom, p= 0.3
```

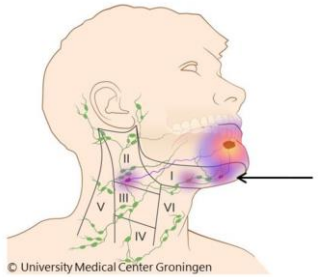


Very poor prognostic factor, likely to treat less aggressively

Locoregional Metastasis




```
Call:
survdiff(formula = Surv(time, status) ~ group, data = df_final)

          N Observed Expected (O-E)^2/E (O-E)^2/V
group=- 31        8     4.61      2.5       4.7
group=+ 52        3     6.39      1.8       4.7

 Chisq= 4.7  on 1 degrees of freedom, p= 0.03
```

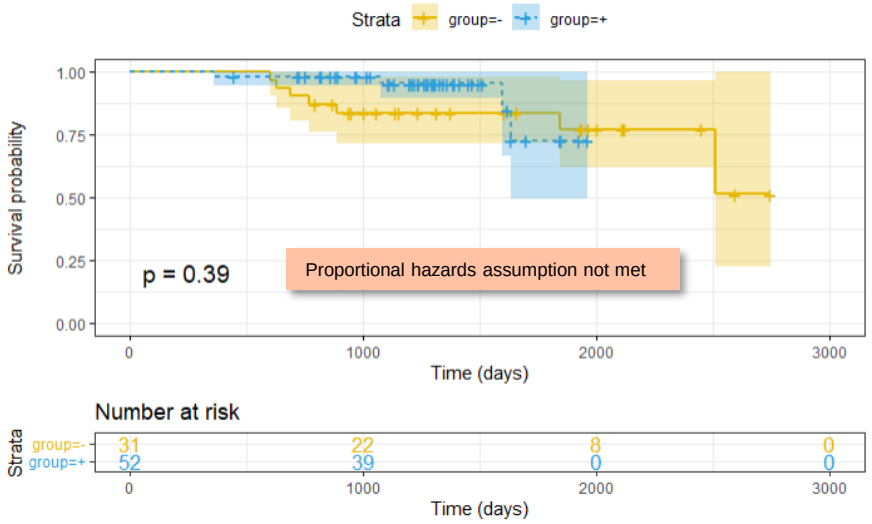


Head and neck dissection followed by adjuvant chemotherapy and radiotherapy (depending on the stage)

Results

Secondary Analysis

Death



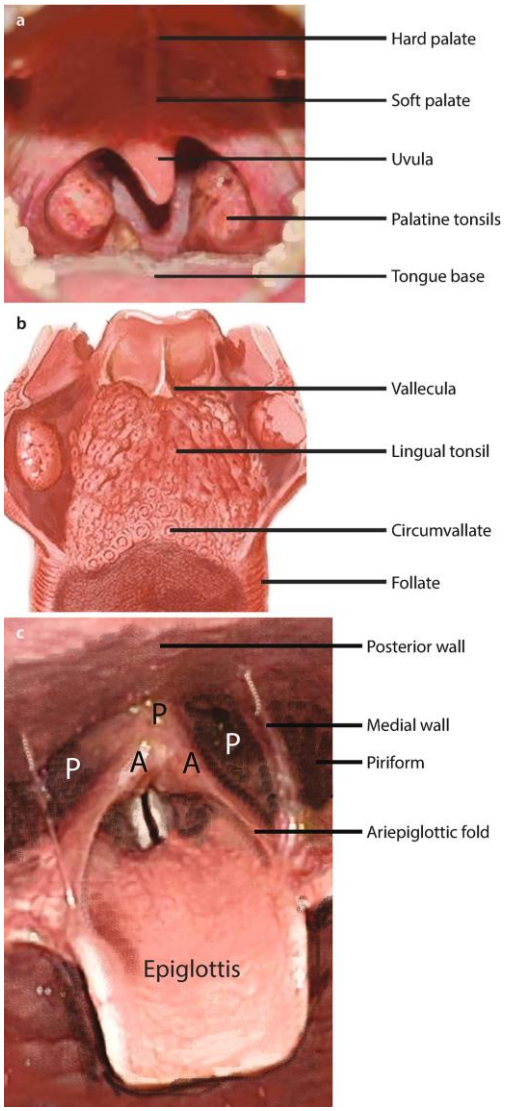

```
Call:
survdiff(formula = Surv(time, status) ~ group, data = df_final)

         N Observed Expected (O-E)^2/E (O-E)^2/V
group=- 31        7      5.7     0.297     0.754
group=+ 52        4      5.3     0.319     0.754

 Chisq= 0.8  on 1 degrees of freedom, p= 0.4
```

HPV-positive tumors have a good prognosis since they respond better to chemo- and/or radiotherapy.

Sim CQ. Cancers of the Oral Mucosa. In: Elston DM *Cancers of the Oral Mucosa*. New York, NY: http://emedicine.medscape.com/article/1075729. June 28, 2017.

Discussion & Impact

What the Results Mean



Discussion Points

GENERAL:

- Context and literature is important to guide variable inclusion: **intuition before statistical or model complexity**

PRIMARY ANALYSIS:

- Oropharynx, hypopharynx, parapharyngeal space have different cell types and microbiome, very different pathophysiology of the cancer progression (general rule → lower down the upper digestive tract is more severe)

- Cox proportional hazards model and the resultant Hazard Ratios provide a tool that the surgeon can utilize to communicate with the patient ("you have stage T3" vs. "there is a x higher chance of death in this case vs. the T1 case")

- Patients and family members need to make difficult decisions about treatment intensity so communication is vital to the therapeutic relationship you have with a patient (evidence-based medicine)

SECONDARY ANALYSIS:

- Supporting evidence for HPV-positive oropharyngeal cancer having a reduction in the risk of locoregional metastasis and reduction in the risk disease recurrence. Secondary primary tumour (SPT) in patients with HPV-positive cancer is very rare, and has improved better survival rate compared to patients with HPV negative tumours:

- Chu A, Genden E, Posner M, et al. A patient centered approach to counseling patients with head and neck cancer undergoing human papillomavirus testing: a clinician's guide. *Oncologist.* 2013;18:180–189.

References

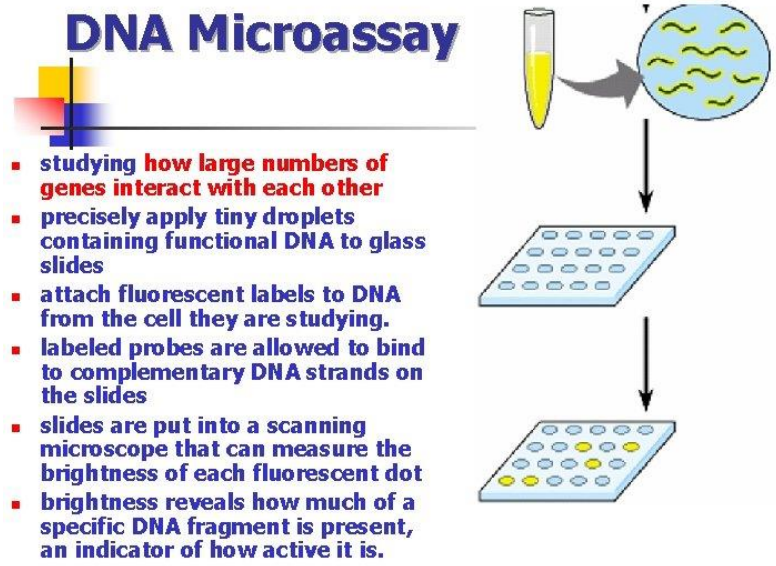
1. Zhang, H., Seikaly, H., Abele, J.T. et al. Metabolic tumour volume as a prognostic factor for oral cavity squamous cell carcinoma treated with primary surgery. J of Otolaryngol - Head & Neck Surg 43, 33 (2014). https://doi.org/10.1186/s40463-014-0033-7
2. Zhang H, Seikaly H, Nguyen NT, et al. Validation of metabolic tumor volume as a prognostic factor for oral cavity squamous cell carcinoma treated with primary surgery. Oral Oncol. 2016;57:6-14. doi:10.1016/j.oraloncology.2016.03.013
3. Simon N, Friedman J, Hastie T, Tibshirani R (2011). “Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent.” Journal of Statistical Software, 39(5), 1–13.
4. Tay, J. Kenneth et al. Elastic Net Regularization Paths for All Generalized Linear Models. (2021).
5. Romesser PB, Qureshi MM, Shah BA, Chatburn LT, Jalisi S, Devaiah AK, Subramaniam RM, Truong MT: Superior prognostic utility of gross and metabolic tumor volume compared to standardized uptake value using PET/CT in head and neck squamous cell carcinoma patients treated with intensity-modulated radiotherapy. Annal Nucl Med. 2012, 26: 527-534. 10.1007/s12149-012-0604-5.
6. Moeller BJ, Rana V, Cannon BA, et al. Prospective risk-adjusted [18F] Fluorodeoxyglucose positron emission tomography and computed tomography assessment of radiation response in head and neck cancer. J Clin Oncol 2009;27:2509–15.
7. Seikaly H, Biron VL, Zhang H, et al. The role of primary surgery in the treatment of advanced oropharyngeal cancer. Head Neck 2015.
8. La TH, Filion EJ, Turnbull BB, et al. Metabolic tumor volume predicts for recurrence and death in head-and-neck cancer. Int J Radiat Oncol Biol Phys 2009;74:1335–41.
9. Fan WL, Yang LY, Hsieh JC, Lin TC, Lu MJ, Liao CT. Prognostic Genetic Biomarkers Based on Oncogenic Signaling Pathways for Outcome Prediction in Patients with Oral Cavity Squamous Cell Carcinoma. Cancers (Basel). 2021;13(11):2709. Published 2021 May 30. doi:10.3390/cancers13112709
10. Su SC, Lin CW, Liu YF, et al. Exome Sequencing of Oral Squamous Cell Carcinoma Reveals Molecular Subgroups and Novel Therapeutic Opportunities. Theranostics. 2017;7(5):1088-1099. Published 2017 Feb 26. doi:10.7150/thno.18551
11. Kleinbaum, D.G. and Klein, M. (2012) Survival Analysis: A Self-Learning Text. 3rd Edition, Springer, New York.
12. Tabe-Bordbar, S., Emad, A., Zhao, S.D. et al. A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models. Sci Rep 8, 6620 (2018). https://doi.org/10.1038/s41598-018-24937-4
13. Vallières, M. *et al.* Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep* **7,** 10117 (2017). doi: 10.1038/s41598-017-10371-5

Appendix

My Ongoing Research

Research Method

- Elastic net penalized cox proportional hazards models to measure the association between clinicopathologic, genetic features, and patient survival (overall and disease-free) will be fit for two groups:
  - patients who underwent surgical resection +/− reconstruction followed by adjuvant radiotherapy
  - patients who underwent surgical resection +/− reconstruction followed by combined adjuvant chemotherapy and radiotherapy.



Modeling Methodology

- *Univariate* Cox proportional hazards regression models will be constructed for age, gender, clinical T-stage, N-Stage, ECOG performance status, CCI scores, and every included clinicopathologic and genetic feature [for which there is data]
  - **p16, p53, Bcl-xL, EGFR, Ki67, pancytokeratin, and DAPI**

$$(\hat{\beta}_0, \hat{\beta}) = \underset{(\beta_0,\beta)\in\mathbb{R}^{p+1}}{\operatorname{argmin}} \left[\underbrace{\frac{1}{2n}\sum_{i=1}^{n}(y_i - \beta_0 - x_i^T\beta)^2}_{\text{OLS}} + \underbrace{\lambda\left(\frac{1-\alpha}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1\right)}_{\text{Elastic Net}} \right]$$

- The elastic net introduces LASSO (L1) and ridge (L2) regularization penalty terms, but at the outset has a low probability of finding the correct subset of predictors, which is why the univariate model step must be taken.
- Confirmed via a Wald test, or the equivalent likelihood ratio test, the features that are significantly associated with the primary outcomes of interest ($\alpha = 0.05$) in the univariate analysis will be used in the development of the multivariate Cox proportional hazards regression model.

Appendix

My Ongoing Research