January 2016

Refining The Staging Scheme For Human Papillomavirus-Related Oropharyngeal Carcinoma

Tiange Chen
Yale University, chentiange.pku@gmail.com

Yale University School of Public Health

Refining the Staging Scheme for Human Papillomavirus-Related Oropharyngeal Carcinoma

A Thesis
Submitted in partial fulfillment of the requirements
For the degree of
Master of Public Health

By
Tiange Chen

Thesis Advisor: Zuoheng Wang
Second Reader: Anthony Kim

May 2016

ABSTRACT

Current TNM staging system derived empirically from human papillomavirus (HPV) unrelated oropharyngeal cancer (OPC) has been shown inadequate to predict survival for HPV-related OPC. This study used three recursive partitioning algorithms, Classification Trees (CART), Conditional Inference Trees (CTree) and Model-based Recursive Partitioning (MOB) to derive a new staging scheme based on data from the National Cancer Data Base (NCDB). The derived staging systems were compared to the current system using the criteria such as hazard consistency within staging groups, hazard discrimination between groups, predictive ability and balance of distribution across groups. A total of 5,712 patients were included in the analysis. The staging system derived using the model-based recursive partitioning (MOB) has the best predictive ability and overall performance. It separates patients into four stages: Stage I (T1-2N0-2a), Stage II (T1-2N2b-3), Stage III (T3), and Stage IV (T4). Stage V is reserved for metastatic patients (M1). The theoretical advantages for the MOB algorithm of fitting the local parametric model in each node and adjusting for covariates affecting survival were confirmed with empirical analysis. Thus MOB algorithm is recommended for future TNM cancer staging studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Oropharyngeal carcinoma (OPC) is a type of head and neck cancer with a rising incidence in the United States (Chaturvedi et al, 2011). An increasing proportion of OPC is related with human papillomavirus (HPV) infection (Ernster, 2007). OPC staging system that defines homogeneous populations is essential for selecting treatment, assessing prognosis and interpreting outcomes (Sobin, Gospodarowicz and Wittekind, 2010).

Currently, the American Joint Committee on Cancer (AJCC)/ Union for International Cancer Control (UICC) TNM staging system is widely accepted for OPC patients (Greene and Sobin, 2002). The extent of the tumor (T), the extent of spread to the lymph nodes (N), and the presence of metastasis (M) categories combine to create staging groups from I to IV that stratify patients according to survival outcomes. The seventh edition AJCC system was derived empirically from smoking-related (i.e. human papillomavirus (HPV) unrelated) OPC outcomes (Edge, Byrd and Compton, 2010). Researchers have shown that this staging system is not adequate to predict survival for HPV-related OPC (Ward et al., 2015; Huang et al., 2015; Dahlstrom et al., 2013), thus a separate staging system is needed.

Recently, Huang et al. (2015) proposed a new staging system for HPV-related OPC using recursive partitioning analysis (RPA). Dahlstrom et al. (2016) and O'sulllivan et al. (2016) externally validate Huang's RPA staging system based on HPC-related OPC patients treated at their institution and developed new staging systems with RPA as well.

Recursive partitioning is a tree-based regression modeling technique introduced by Morgan and Sonquist (1963). The implementations of such algorithm includes Classification and Regression

Trees (CART), Multivariate Adaptive Regression Splines (MARS), and their variants and extensions (Zhang and Singer, 2010). The algorithm consists of two stages: first partition the observations by the covariate showing the best split that maximized an information measure of node impurity in a recursive way, and second fit a regression model in each node of the resulting partition. Gordon and Olshen (1985) first adapted the method of recursive partitioning to censored outcomes. The idea behind their algorithm was to force each node to be homogeneous as measured by a distance metric between the within-node Kaplan-Meier survival function and a survival function that has mass on at most one finite point (Bou-Hamad et al., 2011). Other splitting criteria using the logrank statistic (Ciampi et al., 1986; Segal, 1988) or likelihood ratio statistic (Davis and Anderson, 1989; LeBlanc and Crowley, 1992; Ciampi et al., 1995) was also suggested.

Recursive partitioning has been widely used in TNM staging since the Radiation Therapy Oncology Group (RTOG) applied RPA to their head and neck database and create homogeneous groups based on anatomic and demographic factors (Cooper et al., 1996). Prognostic factors (Shepherd et al., 1993; Roach et al., 2000; Chansky et al., 2009; Huang et al., 2015), anatomic factors such as clinical T, N and M categories (Mountain, 1997; Rice et al., 2003; Huang et al., 2015; Pan et al., 2016) and genetic characteristics (Zhang et al., 2001) are three common types of partitioning covariates that the algorithm split nodes on. Using different splitting criteria and pruning methods often results in different variable selection and splitting values, thus distinct decision trees for staging would be created. But most researchers did not specify the splitting criteria and pruning method being used for RPA, which makes the results unable to replicate and makes it impossible to compare the effectiveness of different recursive partitioning models. In this article various recursive partitioning algorithms, including Classification Trees (CART),

Conditional Inference Trees (CTree) and Model-based Recursive Partitioning (MOB) were applied to HPV-related OPC patients, comparative evaluation for the derived staging systems was performed. The goal of this study was to (1) compare the performance of different recursive partitioning algorithms in cancer staging. (2) propose an alternative staging system for HPV-related OPC that separate the patients into homogeneous groups with respect to survival.

Theory of Recursive Partitioning Analysis for Censored Data

Recursive partitioning is a tree-based regression model introduced by Morgan and Sonquist (1963). The algorithm consists of two stages: first partition the observations by the covariate showing the best split in a recursive way, and second fit a regression model in each node of the resulting partition. Gordon and Olshen (1985) first adapted the idea of recursive partitioning to censored outcomes by forcing each node to be homogeneous. Theories for three recursive partitioning algorithms available for censored data are introduced in the article.

Classification and Regression Trees

Classification and Regression Trees (CART) was introduced in 1984 by Breiman, Friedman, Olshen and Stone as an umbrella term of classification trees and regression trees. They employed a generalization of the binomial variance called the Gini index as the measure of node impurity and split a node by exhaustively searching over all covariates that minimize the total impurity of its two daughter nodes. The process is applied recursively until the relative decrease in total impurity is below a pre-specified threshold.

The measure of node impurity for censored outcome was firstly developed by Gordon and Olshen (1985). They regarded a node as pure if all failures in the node occurred at the same time and defined *P* as the collection of all such Kaplan-Meier curves. The distance between within-node Kaplan-Meier curve and any of the curves in *P* can be used to measure node impurity.

The pruning process for CART is conducted by choosing a best value for complexity parameter using cross-validation. A cost-complexity of tree *T* is defined as

$$R_\alpha(T) = R(T) + \alpha\left|\tilde{T}\right| \qquad [1]$$

where α is the complexity parameter, $|\tilde{T}|$ is the number of terminal nodes in *T*, and $R(T)$ is the sum of the costs over all terminal nodes. According to Breiman et al. (1984), for any value of the complexity parameter α, there is a unique smallest subtree of the initial tree T_o that minimized the cost-complexity. We first derive *m* typical values for complexity parameter spanning from 0 to infinity. Then fit a full model on the reduced training set with (*n-n/s*) observations and determine the subtrees for each complexity parameter. Under each of the *m* models, predict the outcome for each observation in the test set and sum over the cost of *s* subsets. The subtree derived by the complexity parameter with the smallest cost is chosen as the best pruned tree. (Therneau and Atkinson, 2015)

Conditional Inference Trees

Exhaustive search over all possible splits that maximize an information measure of node impurity often leads to a selection bias towards covariates with many possible splits.

Hothorn, Hornik and Zeileis (2006) thus proposed Conditional Inference Trees (CTree), a non-parametric class of regression trees embedding tree-structured regression models into the theory of permutation tests (Strasser and Weber, 1999), to fix this problem.

The CTree algorithm can be formulated using non-negative integer case weights $w = (\omega_1, \ldots, \omega_n)$. Each node is represented by a vector of case weights when the corresponding observations are within this node. For case weights *w* test the global null hypothesis of independence between any of the *m* covariates and the response variable· The association between *Y* and X_j was measured by a linear statistics of the form

$$T_j(L_n, w) = vec(\sum_{i=1}^{n} w_i g_j(X_{ji}) h(Y_i, (Y_1, \ldots, Y_n))^T) \quad [2]$$

where g_j is a non-random transformation of the covariate X_j, h is the influence function depending on the response $(Y_1, \ldots, Y_n)$ in a permutation symmetric way. Under the null hypothesis one can dispose the dependency of $T_j(L_n, w)$ on the joint distribution of *Y* and X_j by fixing the covariates and conditioning on all possible permutations of *Y*. Then standardize the linear statistic and take the maximum of the absolute value and derive the P-value for the conditional distribution of this new test statistic. Next split the node over the covariate X_{j^*} with strongest association to *Y* (i.e. minimum P-value). Recursively repeat this process until the null hypothesis cannot be rejected at a pre-specified level α.

For censored regression the influence function *h* may be chosen as Logrank or Savage scores (Segal, 1988) and one can proceed as for univariate continuous regression. Alternatively, one can use the weighting scheme suggested by Molinaro, Dudoit, and van der Laan (2004) and take the weighted Kaplan-Meier curve for the case weights *w(x)* as prediction.

Model-Based Recursive Partitioning

Motivated by the fact that constant fits in each node tend to produce large and hard-to-interpret trees (Chan and Loh, 2004), the incorporation of parametric models into recursive partitioning has been of increased interest in the last decade (Zeileis et al., 2008). Inspired by algorithms of GUIDE (Loh, 2002), CRUISE (Kim and Loh, 2001) and LOTUS (Chan and Loh, 2004) that attached parametric models to terminal nodes, and maximum likelihood trees developed by Su

wt al. (2004), Zeileis, Hothorn and Hornik (2008) introduced a framework that embeds recursive partitioning into statistical model estimation and variable selection.

Consider a parametric model $M\ (Y, \theta)$ observations Y and a k-dimensional vector of parameters θ. In many situations a single global model will not fit all n observations well. Then it might be possible to partition the observations using another set of covariates $Z_1, \cdots, Z_l$ such that the model can be well-fitted locally in each node.

The Model-based Recursive Partitioning (MOB) algorithm is used to find such a partition adaptively using a greedy forward search. Firstly, fit the model to observations in the current node by minimization of some objective function Ψ. If there is some overall instability in the parameter estimates with respect to any of the partitioning variables Z_j, split the node over the variable Z_j associated with the highest parameter instability. To assess whether the parameter estimates are stable, the general class of score-based fluctuation test for parameter instability (Zaileis and Hornik, 2007) is performed. The idea is to check whether the scores $\psi(Y, \theta) = \frac{d\Psi(Y,\theta)}{d\theta}$ fluctuate randomly around their mean 0 or exhibit systematic deviations from 0 over Z_j. These deviations can be formulated as

$$\mathrm{W}_j(t) = \hat{J}^{-1/2} n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \hat{\psi}_{\sigma(Z_{ij})} \qquad [3]$$

where $\sigma(Z_{ij})$ is the ordering permutation which gives the anti-rank of the observation Z_{ij} in the vector Z_j, and $\hat{J}$ is a suitable estimate of the covariate matric $\mathrm{cov}(\psi(Y, \theta))$. A test statistic can be derived by applying a scalar function that captures the fluctuation in the empirical process to the

fluctuation process. Next compute the split point that locally optimize the Ψ. When no more significant instabilities can be found, the recursion stops. Post-pruning can be applied by first growing a large tree and then pruning back splits that did not improve the model based on information criteria such as AIC or BIC (Su, Wang, and Fan, 2004).

METHODS

Study population

This project was a retrospective study that included patients with HPV-related OPC from National Cancer Data Base (NCDB). The NCDB is a joint project of the Commission on Cancer (CoC) of the American College of Surgeons (ACS) and the American Cancer Society. It draws from more than 1,500 hospital registries and captures approximately 70% of all newly diagnosed cancer cases in the U.S. (Bilimoria et al. 2008) The database recorded patient demographics, socioeconomic status, stage, tumor characteristics, comorbidity score, and treatment information.

Patients diagnosed with squamous cell carcinoma of the oropharynx between 2010-2012 were enrolled in the study. Patients with unknown HPV status or HPV-negative status were excluded. Patients with metastatic disease (M1) were also excluded from the analysis of RPA stage derivation because they had distinct survival regardless of clinical T and N stages and only comprised 1.5% of study population. The analysis was eventually conducted on 5,626 cases meeting clinical and pathological inclusion criteria.

Statistical Analysis

Three variants of recursive partitioning algorithms for censored outcomes including CART, CTree and MOB were used to determine the new staging system for HPV-related OPC. The underlying statistical theories of the algorithms were elaborated in section 2. **Table 1** summarizes the differences in splitting criteria, pruning method and program of implementation of the three algorithms. Ordinal clinical T stage (T1/T2/T3/T4a/T4b) and clinical N stage (N0/N1/N2a/N2b/N2c/N3) were entered into the model as possible partitioning variables. Age,

Charlson-Deyo score and treatment were used to construct the parametric model in MOB algorithm.

Table 1 Comparison of CART, CTree and MOB algorithms

	Type of Model	Splitting Criteria	Pruning Method	Implementation Program
CART	Non-parametric	Minimizing total node impurity measured by the difference from Kaplan-Meier curve to standard curves	Choosing smallest complexity parameter using cross-validation	Rpart() function from R package Rpart
CTree	Non-parametric	Partitioning over the covariate with strongest association to response	None	Ctree() function from R package partykit
MOB	Parametric	Partitioning over the covariate associated with the highest parameter instability	Pruning back the splits with no improvement on AIC	Mob() function from R package partykit

3-year overall survival (OS) was calculated for the 7th edition AJCC staging groups, the proposed CART-derived groups, the CTree-derived groups, and the MOB-derived groups using Kaplan-Meier method. Pairwise log-rank tests were used to detect differences in survival between staging groups. Adjusted hazard ratio (AHR) with 95% confidence interval was derived using cox proportional hazard models. Variables that had significant effect on survival in univariate analysis, including age, Charlson-Deyo score and treatment were included in the final multivariate model. All tests were two sided, and a P-value of < 0.05 was considered statistically significant. Bonferroni correction was used for subgroup analysis and pairwise comparison.

The four staging systems were then evaluated using the criteria proposed by Groome et al. (2001): (1) hazard consistency measuring whether observations within a staging group have similar survival rate. A weighted average of the survival deviation between each staging group and the TNM subgroups that comprise this staging group was used. (2) hazard discrimination measuring whether the survival rates differ between staging groups. The average of a measure of

evenness of the curves and the span of the curves was used. (3) outcome prediction is high. We use Brier score (Gerds and Schumacher, 2006) and concordance probability estimate (CPE) (Gönen and Heller, 2005) to measure the predictive and discriminative ability of the models. (4) balance in the distribution of cases. As in the original Groome study, the first three criteria were given a weight of 2 and balance was given a weight of 1. Different weights assignments were also discussed. The overall score was then calculated. Bootstrap with replacement was performed for internal validation.

R version 3.2.3 was used for all statistical analyses.

RESULTS

Study Cohort Descriptive Analysis

A total of 5,712 patients with HPV-related OPC entered into the statistical analysis. The median follow-up was 28.45 months (95% CI: 28.09 to 28.91), estimated using reverse Kaplan-Meier method (Schemper and Smith, 1996). Demographic and clinical characteristics for the 5,626 patients are provided in **Table 2**. The median age was 58. Among those patients, 86% were men, 84% had no comorbid conditions recorded, 56% received primary radiation therapy and 41% received primary surgical therapy. The distribution of clinical T and N categories is also listed in the table.

Table 2 Demographic and Clinical Characteristics of 5,626 HPV-related OPC Patients

Characteristic	No. (%) of Patients	Hazard Ratio (P-value)
Age		
Mean (Standard deviation)	58.43 (9.33)	1.049 (< .001)
Median (Quartiles)	58 (52, 64)	
Sex		
Male	4,911 (85.98)	0.953 (0.67)
Female	801 (14.02)	
Charlson-Deyo score		
0	4,809 (84.19)	
1	726 (12.71)	1.967 (< .001)
2+	177 (3.1)	1.023 (0.83)
Clinical T category		
T1	1,732 (30.32)	
T2	2,480 (43.42)	4.082 (< .001)
T3	895 (15.67)	0.906 (0.42)
T4a	491 (8.6)	0.955 (0.63)
T4b	114 (19.96)	1.022 (0.8)
Clinical N category		
N0	814 (14.25)	
N1	1,113 (19.49)	1.963 (< .001)
N2a	640 (11.2)	2.204 (< .001)
N2b	2,099 (36.75)	0.735 (0.005)
N2c	822 (14.39)	0.854 (0.14)
N3	224 (3.92)	1.212 (0.104)

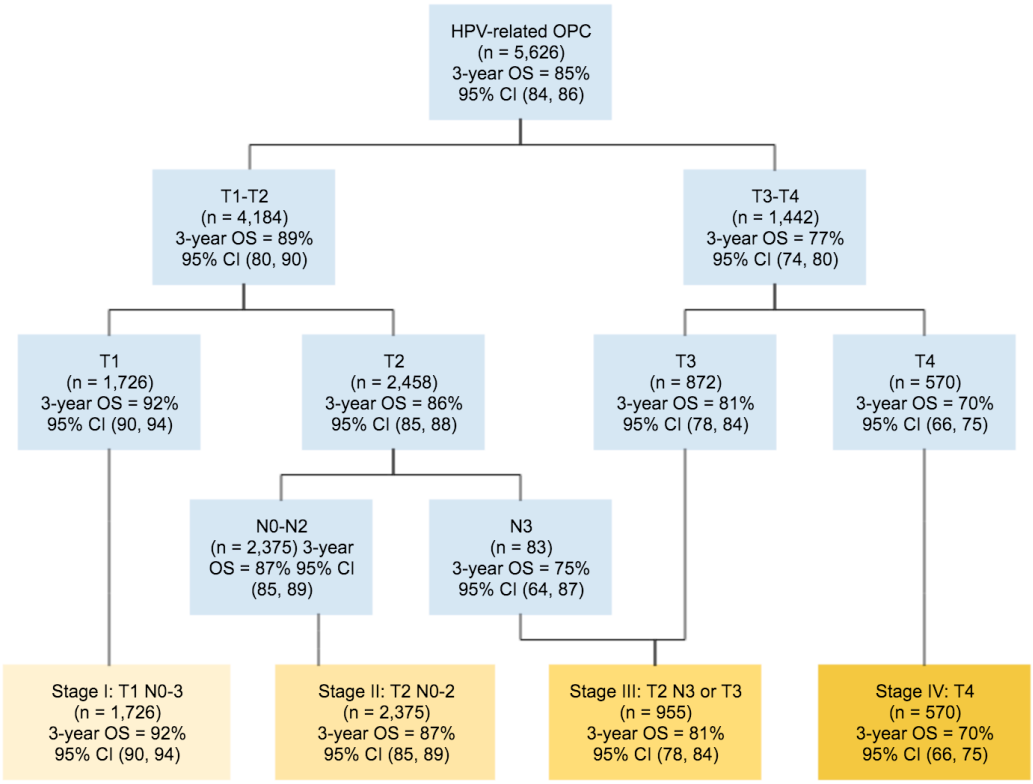
Treatment		
Primary radiation	3,222 (56.41)	0.698 (< .001)
Primary surgery	2,361 (41.33)	0.365 (< .001)
Other	129 (2.26)	

Recursive Partitioning Analysis

Utilizing the recursive partitioning algorithms of CART, CTree and MOB, three TNM staging systems were derived respectively. **Figure 1** shows the tree-based staging groups derived using CART with 3-year OS estimates, and Figure 1 B shows the combination of clinical T and N categories for each staging group. The tree and table representation of CTree derived staging groups (

Figure 2), and MOB derived staging groups (**Figure 3**) are also provided. For all three staging systems, patients with metastatic tumors (M1) are grouped in to a separate stage. Both MOB derived and CART derived staging systems eventually have five staging groups. Stage IV for MOB and CART staging systems is the same (T4). Ctree and CART staging systems have the same stage I (T1).

A

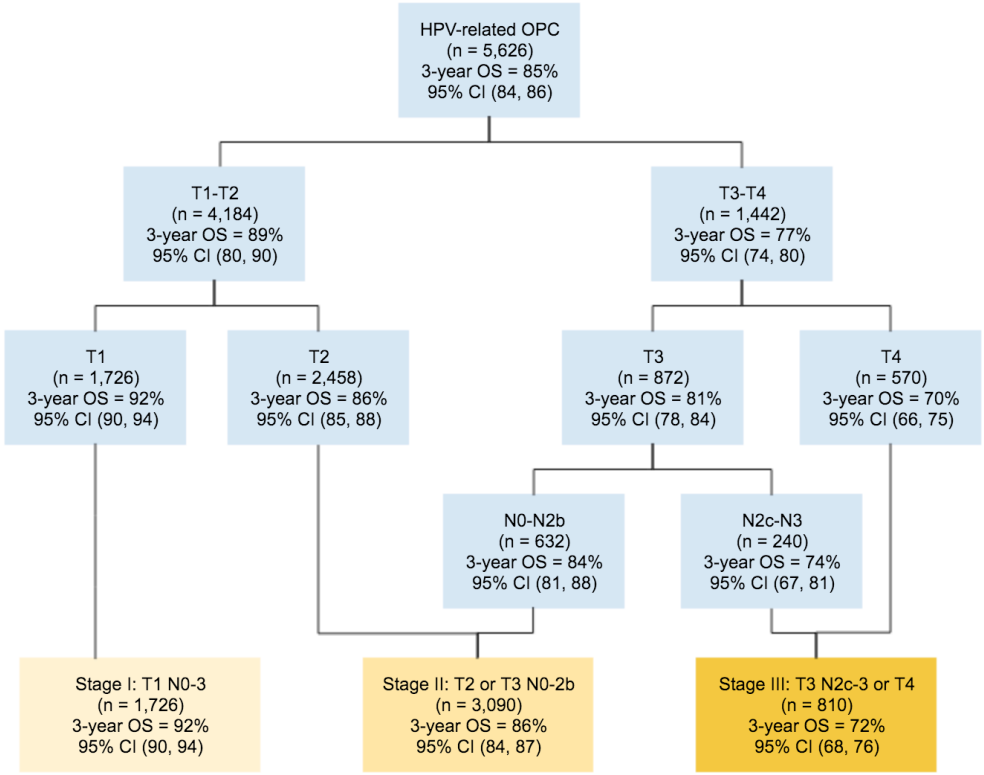


B

	T1	T2	T3	T4A	T4B
N0	I	II	III	IV	IV
N1	I	II	III	IV	IV
N2A	I	II	III	IV	IV
N2B	I	II	III	IV	IV
N2C	I	II	III	IV	IV
N3	I	III	III	IV	IV

Figure 1 Staging groups derived using CART algorithm (A) Staging groups and 3-year OS (B) Clinical T and N categories for each staging group

A

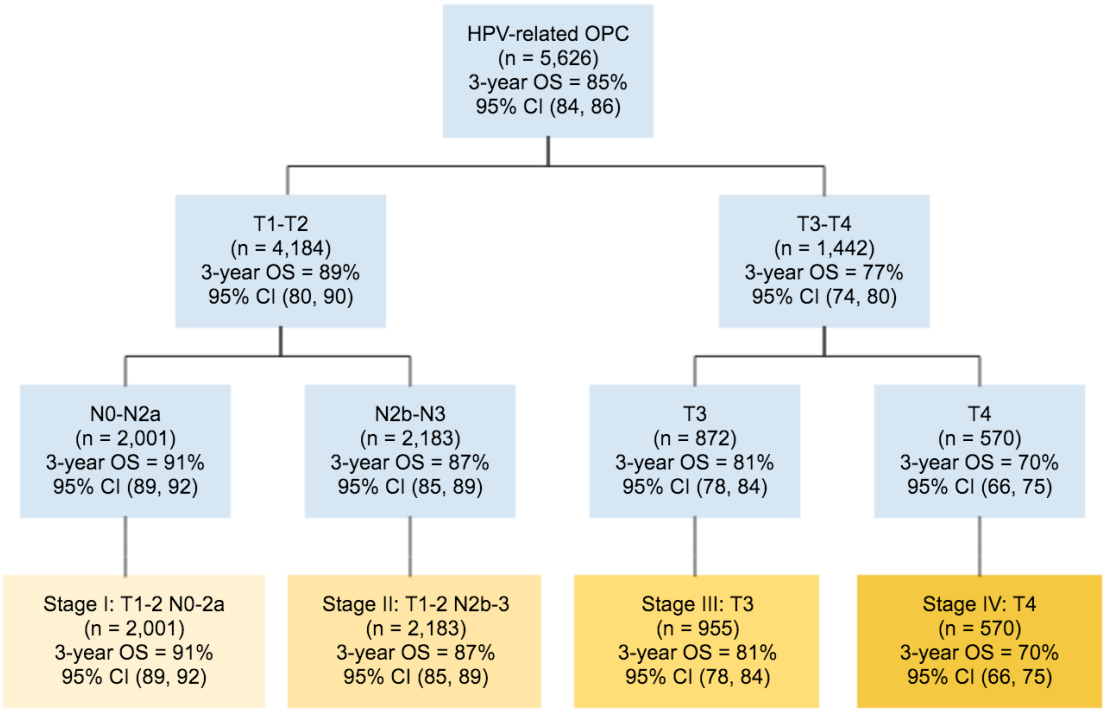


B

	T1	T2	T3	T4A	T4B
N0	I	II	II	III	III
N1	I	II	II	III	III
N2A	I	II	II	III	III
N2B	I	II	II	III	III
N2C	I	II	III	III	III
N3	I	II	III	III	III

Figure 2 Staging groups derived using CTree algorithm (A) Staging groups and 3-year OS (B) Clinical T and N categories for each staging group

A



B

	T1	T2	T3	T4A	T4B
N0	I	I	III	IV	IV
N1	I	I	III	IV	IV
N2A	I	I	III	IV	IV
N2B	II	II	III	IV	IV
N2C	II	II	III	IV	IV
N3	II	II	III	IV	IV

Figure 3 Staging groups derived using MOB algorithm (A) Staging groups and 3-year OS (B) Clinical T and N categories for each staging group

Survival Analysis using AJCC and Recursive Partitioning Derived Staging groups

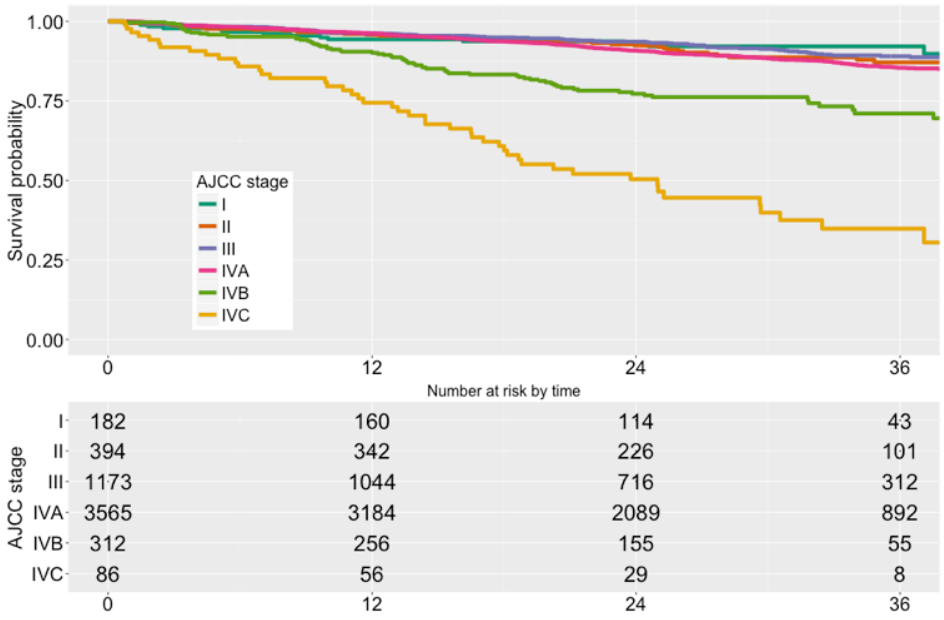
Kaplan-Meier survival curves with number-at-risk table for the AJCC, CART, CTree and MOB staging groups appear in

Figure 4. Adjusted hazard ratios and 3-year overall survival with their 95% CI are provided in

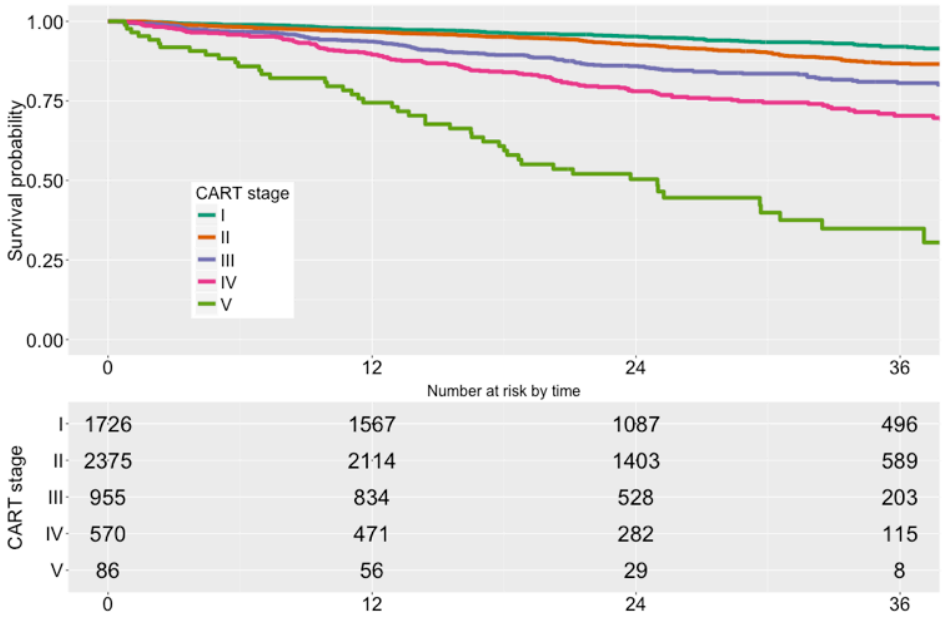
Table 3.

In all four cases, overall log-rank test demonstrated statistically significant difference in survival rate across staging groups ($P < 0.01$). In three staging systems based on recursive partitioning, a monotonic reduction in 3-year OS according to higher TNM stages can be seen. In AJCC staging system, however, pairwise log-rank tests showed that no significant difference exist between Stage I and Stage II ($P = 0.48$), Stage II and Stage III ($P = 0.53$) and Stage III and Stage IVA ($P = 0.011$). Table 2 also suggests relatively indistinguishable 3-year OS for AJCC staging group I to group IVA (92%, 87%, 89%, 85% respectively).

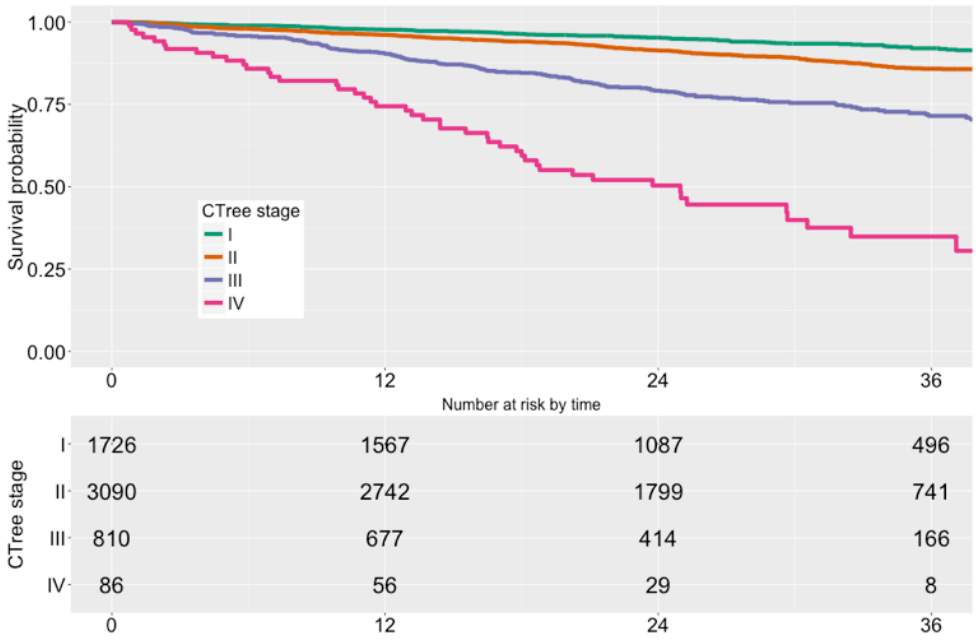
A



B



C



D

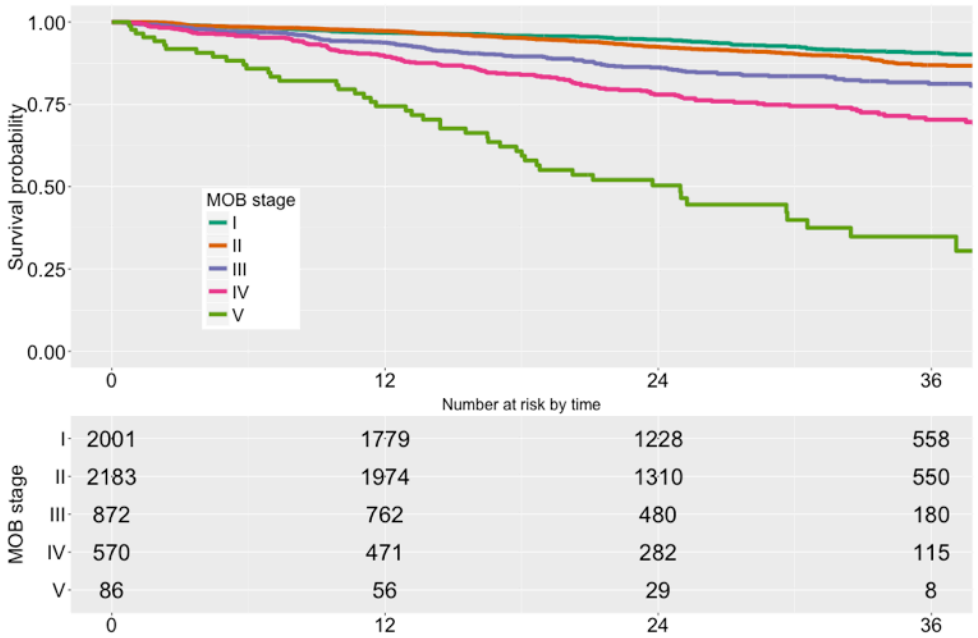


Figure 4 Overall Kaplan-Meier survival curve by (A) AJCC stages (B) CART stages (C) CTree stages (D) MOB stages

Table 3 Three-year OS, AHR with 95% CIs for AJCC, CART, CTree and MOB staging systems

Stage	No. of patients	3-year Overall Survival (95% Confidence Interval)	Hazard Ratio	p-value for HR
AJCC				
I	182	92.11% (88.03%, 96.38%)	Ref	
II	394	87.12% (82.98%, 91.46%)	1.290 (0.696, 2.392)	0.418
III	1173	89.01% (86.78%, 91.29%)	1.416, (0.805, 2.491)	0.228
IVA	3565	85.38% (83.90%, 86.88%)	1.990 (1.156, 3.425)	0.013 *
IVB	312	71.02% (64.80%, 77.83%)	4.794 (2.675, 8.591)	< .001 ***
IVC	86	34.84% (24.29%, 49.96%)	10.632 ((5.841, 19.353)	< .001 ***
CART				
I	1726	92.02% (90.43%, 93.65%)	Ref	
II	2375	86.73% (84.95%, 88.55%)	1.383 (1.101, 1.737)	0.00534 **
III	955	80.59% (77.54%, 83.77%)	2.477 (1.933, 3.175)	< .001 ***
IV	570	70.34% (65.75%, 75.25%)	3.726 (2.882, 4.819)	< .001 ***
V	86	34.84% (24.29%, 49.96%)	9.710 (6.884, 13.695)	< .001 ***
CTree				
I	1726	92.02% (90.43%, 93.65%)	Ref	
II	3090	85.83% (84.25%, 87.44%)	1.553 (1.251, 1.929)	< .001 ***
III	810	71.49% (67.68%, 75.50%)	3.613 (2.838, 4.60)	< .001 ***
IV	86	34.84% (24.29%, 49.96%)	9.682 (6.866, 13.655)	< .001 ***
MOB				
I	2001	90.65% (89.06%, 92.28%)	Ref	
II	2183	86.91% (85.06%, 88.81%)	1.622 (1.305, 2.015)	< .001 ***
III	872	81.23% (78.10%, 84.49%)	2.456 (1.932, 3.122)	< .001 ***
IV	570	70.34% (65.75%, 75.25%)	3.823 (2.999, 4.872)	< .001 ***
V	86	34.84% (24.29%, 49.96%)	9.801 (7.0413, 13.642)	< .001 ***

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Performance Evaluation of the AJCC and Recursive Partitioning Derived Staging systems

Comparing the seventh edition AJCC staging system and the three recursive partitioning derived staging systems using Groome's four criteria, the MOB stage performed best overall, followed by CART and CTree stage. The AJCC stage performed least well. The comparative result was also validated by bootstrap. (**Table 4**)

CART stage achieved best hazard consistency score, indicating that patients within each CART staging group have relatively consistent survival experience. With the worst hazard consistency score, AJCC stage has heterogeneous patients in each group with regard to survival. Hazard discrimination measures the evenness of the distribution of survival curves across groups and the span of the curves. With the patients of M1 disease excluded, CTree stage distributed most evenly across group (absolute differences in survival rates are at least 2%, 5%, 6%, and 4% for AJCC, CART, CTree and MOB respectively). All four staging systems have similar span of the curves (differences in survival rates between the first and the last stage are 21%, 22%, 21% and 20% respectively) according to Table 2. With the greatest number of subgroups, the AJCC staging scheme outperformed other systems with regard to hazard discrimination. The MOB stage derived by parametric model-based recursive partitioning has a Brier score of 13.54%, indicating a greater predictive power than other systems, whereas the AJCC stage has worst discriminative ability. The three recursive partitioning derived staging systems are well-balanced with respect to the number of patients in each staging group, while the sample size distribution is unbalanced for AJCC stages.

Table 4 Performance evaluation of AJCC, CART, CTree and MOB staging systems

Evaluation Criteria	AJCC Stage	CART Stage	CTree Stage	MOB Stage
Performance evaluation for the study cohort				
% Hazard consistency	2.51	1.27	1.49	1.56
Score	1.00	0.00	0.18	0.23
Rank	4	1	2	3
Hazard discrimination	0.13	0.43	0.44	0.35
Score	0.00	0.95	1.00	0.71
Rank	1	3	4	2
Outcome prediction (% variance explained)	11.51	12.29	12.31	13.54
Score	1.00	0.61	0.61	0.00
Rank	4	3	2	1
Balance	0.88	0.46	0.43	0.49
Score	1.00	0.06	0.00	0.12
Rank	4	2	1	3
Overall score	0.71	0.45	0.51	0.29

Overall rank	4	2	3	1
Performance evaluation using bootstrap (n=1000)				
% Hazard consistency	2.51	1.71	1.70	1.71
Score	0.95	0.18	0.16	0.17
Rank	3.89	2.22	1.93	2.08
Hazard discrimination	0.17	0.25	0.35	0.25
Score	0.02	0.42	1.00	0.42
Rank	1.11	2.48	3.98	2.43
Outcome prediction (% variance explained)	11.85	12.53	12.51	13.85
Score	0.75	0.53	0.52	0.08
Rank	3.17	2.66	2.69	1.47
Balance	0.88	0.46	0.43	0.49
Score	1.00	0.06	0.00	0.12
Rank	4.00	1.93	1.07	3.00
Overall score	0.60	0.45	0.47	0.34
Overall rank	4	2	3	1
% Rank=1	0.04	0.20	0.31	0.45
% Rank=2	0.07	0.29	0.25	0.39

DISCUSSION

The seventh edition AJCC staging system was derived empirically from HPV-unrelated OPC and has been shown to be inadequate to predict survival for HPV-related OPC. Hence, this article used three recursive partitioning algorithms, CART, CTree and MOB to derive a new staging scheme based on NCDB patients. In addition to proposing a new staging system valid for HPV-related OPC, the other goal of this study was to compare the recursive partitioning algorithms with respect to the performance of the staging systems derived by them, and suggest the best algorithm for cancer staging.

The importance of cancer staging lies in its application in planning treatment, assessing prognosis, stratifying patients for therapeutic studies, evaluating treatment outcome, and supporting cancer control. Therefore, unlike model validation which merely uses measures of predictive power or goodness of fit as assessing criteria, the evaluation for staging schemes need to take their applicability in real life into account. Groome et al. (2001) identified four characteristic for useful staging systems: similar survival outcomes within each group; heterogeneous survival between groups; high predictive ability; and balanced distribution across groups. These four criteria were then used for evaluation of the standard AJCC and three derived RPA staging schemes.

According to evaluation analysis based on the study cohort and bootstrap validation (Table 4), MOB derived staging system has best predictive ability and overall performance. Thus it became the stage scheme for HPV-related OPC recommended by this article. Patients were grouped into four stages: Stage I (T1-2N0-2a), Stage II (T1-2N2b-3), Stage III (T3), and Stage IV (T4). Stage V is reserved for metastatic patients (M1).

MOB is a parametric model-based recursive partitioning algorithm developed to address the issue of constant fits in nodes. Rather than fitting a single global model using one set of covariates, MOB examines the possibility to partition the observations using another set of variables and construct locally well-fitted models in each node. Since a convention of OPC staging is to group patients by the combination of clinical T, N and M categories, other covariates affecting survival e.g. age, gender and treatment cannot be entered into the model if we use non-parametric recursive partitioning algorithms such as CART and CTree. The benefit of MOB is then very clear; it adjusts for covariates affecting survival by fitting multivariate cox models with those covariates locally and splitting a node over T, N or M if significant parameter instability is observed.

Based on empirical analysis in this article and theoretical advantage of MOB algorithm, MOB is recommended as a default method for future TNM cancer staging.

There are limitations of this study that worth discussion. The median follow-up is 28.5 months for the study cohort. This only allows an extrapolation to three-year overall survival in analysis, making it impossible to observe late distant metastases occurring 3 years or more after treatment that has been described for HPV-related disease (Huang et al., 2013). In addition, NCDB does not provide information on smoking history, a strong predictor of increased risk of failure (Gillison et al., 2011). Even though internal validation was performed with bootstrap, studies using independent datasets for external validation are needed to confirm the recommended recursive partitioning algorithm as well as the staging scheme.

REFERENCES

Bilimoria, K. Y., Stewart, A. K., Winchester, D. P., & Ko, C. Y. (2008). The National Cancer Data Base: a powerful initiative to improve cancer care in the United States. Annals of surgical oncology, 15(3), 683-690.

Bou-Hamad, I., Larocque, D., & Ben-Ameur, H. (2011). A review of survival trees. Statistics Surveys, 5, 44-71.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

Chan, K. Y., & Loh, W. Y. (2004). LOTUS: An algorithm for building accurate and comprehensible logistic regression trees. Journal of Computational and Graphical Statistics, 13(4), 826-852.

Chansky, K., Sculier, J. P., Crowley, J. J., Giroux, D., Van Meerbeeck, J., & Goldstraw, P. (2009). The International Association for the Study of Lung Cancer Staging Project: prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. Journal of Thoracic Oncology, 4(7), 792-801.

Chaturvedi, A. K., Engels, E. A., Pfeiffer, R. M., Hernandez, B. Y., Xiao, W., Kim, E., ... & Liu, L. (2011). Human papillomavirus and rising oropharyngeal cancer incidence in the United States. Journal of Clinical Oncology, 29(32), 4294-4301.

Ciampi, A., Thiffault, J., Nakache, J. P., & Asselain, B. (1986). Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates.Computational statistics & data analysis, 4(3), 185-204.

Ciampi, A., Negassa, A., & Lou, Z. (1995). Tree-structured prediction for censored survival data and the Cox model. Journal of clinical epidemiology,48(5), 675-689.

Cooper, J. S., Scott, C. B., Asbell, S. O., Rotman, M., Marcial, V., Fu, K. K., ... & Emami, B. (1996). Comparison of RPA- derived staging and AJCC staging in head and neck cancers based on RTOG data. Radiation Oncology Investigations, 4(3), 140-144.

Dahlstrom, K. R., Calzada, G., Hanby, J. D., Garden, A. S., Glisson, B. S., Li, G., ... & Sturgis, E. M. (2013). An evolution in demographics, treatment, and outcomes of oropharyngeal cancer at a major cancer center. Cancer, 119(1), 81-89.

Dahlstrom, K. R., Garden, A. S., William, W. N., Lim, M. Y., & Sturgis, E. M. (2016). Proposed Staging System for Patients With HPV-Related Oropharyngeal Cancer Based on Nasopharyngeal Cancer N Categories.Journal of Clinical Oncology, JCO646448.

Davis, R. B., & Anderson, J. R. (1989). Exponential survival trees. Statistics in Medicine, 8(8), 947-961.

Edge, S. B., & Compton, C. C. (2010). The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Annals of surgical oncology, 17(6), 1471-1474.

Ernster, J. A., Sciotto, C. G., O'Brien, M. M., Finch, J. L., Robinson, L. J., Willson, T., & Mathews, M. (2007). Rising incidence of oropharyngeal cancer and the role of oncogenic human papilloma virus. The Laryngoscope,117(12), 2115-2128.

Gerds, T. A., & Schumacher, M. (2006). Consistent Estimation of the Expected Brier Score in General Survival Models with Right- Censored Event Times. Biometrical Journal, 48(6), 1029-1040.

Gillison, M. L., Zhang, Q., Jordan, R., Xiao, W., Westra, W. H., Trotti, A., ... & Ang, K. K. (2012). Tobacco smoking and increased risk of death and progression for patients with p16-positive and p16-negative oropharyngeal cancer. Journal of clinical oncology, JCO-2011.

Gönen, M., & Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. Biometrika, 92(4), 965-970.

Gordon, L., & Olshen, R. A. (1985). Tree-structured survival analysis. Cancer treatment reports, 69(10), 1065-1069.

Greene, F. L., & Sobin, L. H. (2002). The TNM system: our language for cancer care. Journal of surgical oncology, 80(3), 119-120.

Groome, P. A., Schulze, K., Boysen, M., Hall, S. F., & Mackillop, W. J. (2001). A comparison of published head and neck stage groupings in carcinomas of the oral cavity. Head & neck, 23(8), 613-624.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics, 15(3), 651-674.

Huang, S. H., Perez-Ordonez, B., Weinreb, I., Hope, A., Massey, C., Waldron, J. N., ... & Ringash, J. (2013). Natural course of distant metastases following radiotherapy or chemoradiotherapy in HPV-related oropharyngeal cancer. Oral oncology, 49(1), 79-85.

Huang, S. H., Xu, W., Waldron, J., Siu, L., Shen, X., Tong, L., ... & Cho, J. (2015). Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus–related oropharyngeal carcinomas. Journal of Clinical Oncology, 33(8), 836-845.

LeBlanc, M., & Crowley, J. (1992). Relative risk trees for censored survival data. Biometrics, 411-425.

Molinaro, A. M., Dudoit, S., & Van der Laan, M. J. (2004). Tree-based multivariate regression and density estimation with right-censored data.Journal of Multivariate Analysis, 90(1), 154-177.

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. Journal of the American statistical association, 58(302), 415-434.

Mountain, C. F. (1997). Revisions in the international system for staging lung cancer. Chest Journal, 111(6), 1710-1717.

O'Sullivan, B., Huang, S. H., Su, J., Garden, A. S., Sturgis, E. M., Dahlstrom, K., ... & Adelstein, D. (2016). 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. The Lancet Oncology, 17(4), 440-451.

Pan, J. J., Ng, W. T., Zong, J. F., Chan, L. L., O'Sullivan, B., Lin, S. J., ... & Kan, W. K. (2015). Proposal for the 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity- modulated radiotherapy. Cancer.

Rice, T. W., Blackstone, E. H., Rybicki, L. A., Adelstein, D. J., Murthy, S. C., DeCamp, M. M., & Goldblum, J. R. (2003). Refining esophageal cancer staging. The Journal of thoracic and cardiovascular surgery, 125(5), 1103-1113.

Roach, M., Lu, J., Pilepich, M. V., Asbell, S. O., Mohuidden, M., Terry, R., & Grignon, D. (2000). Four prognostic groups predict long-term survival from prostate cancer following radiotherapy alone on Radiation Therapy Oncology Group clinical trials. International Journal of Radiation Oncology* Biology* Physics, 47(3), 609-615.

Schemper, M., & Smith, T. L. (1996). A note on quantifying follow-up in studies of failure time. Controlled clinical trials, 17(4), 343-346.

Segal, M. R. (1988). Regression trees for censored data. Biometrics, 35-47.

Shepherd, F. A., Ginsberg, R. J., Haddad, R., Feld, R., Sagman, U., Evans, W. K., ... & Maki, E. (1993). Importance of clinical staging in limited small-cell lung cancer: a valuable system to separate prognostic subgroups. The University of Toronto Lung Oncology Group. Journal of clinical oncology,11(8), 1592-1597.

Sobin, L. H., Gospodarowicz, M. K., & Wittekind, C. (Eds.). (2011). TNM classification of malignant tumours. John Wiley & Sons.

Strasser, H., & Weber, C. (1999). On the asymptotic theory of permutation statistics.

Su, X., Wang, M., & Fan, J. (2004). Maximum likelihood regression trees.Journal of Computational and Graphical Statistics, 13(3), 586-598.

Therneau, T. M., & Atkinson, E. J. (2015). An introduction to recursive partitioning using the RPART routines. Rochester: Mayo Foundation.

Ward, M. J., Mellows, T., Harris, S., Webb, A., Patel, N. N., Cox, H. J., ... & King, E. V. (2015). Staging and treatment of oropharyngeal cancer in the human papillomavirus era. Head & neck, 37(7), 1002-1013.

Zeileis, A., & Hornik, K. (2007). Generalized M- fluctuation tests for parameter instability. Statistica Neerlandica, 61(4), 488-508.

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. Journal of Computational and Graphical Statistics, 17(2), 492-514.

Zeileis, A., & Hothorn, T. Parties (2012), Models, Mobsters: A New Implementation of Model-Based Recursive Partitioning in R.

Zhang, H., Yu, C. Y., Singer, B., & Xiong, M. (2001). Recursive partitioning for tumor classification with gene expression microarray data. Proceedings of the National Academy of Sciences, 98(12), 6730-6735.

Zhang, H., & Singer, B. (2010). Recursive partitioning and applications. Springer Science & Business Media.