

Predictive Classifier for Intensive Treatment of Head and Neck Cancer

Kaveh Zakeri MD, MAS,¹ Federico Rotolo PhD,^{2,3} Benjamin Lacas MS,^{2,3} Lucas K Vitzthum MD, MAS,⁴
Quynh-Thu Le MD,⁵ Vincent Gregoire MD, PhD,⁶ Jens Overgaard MD, DMSc,⁷ Allan Hackshaw MSc,⁸
Björn Zackrisson MD,⁹ Mahesh KB Parmar DPhil,¹⁰ Barbara Burtness MD,¹¹ Maria Grazia Ghi MD,¹²

Giuseppe Sanguineti MD,¹³ Brian O'Sullivan MD,¹⁴ Catherine Fortpied MSc,¹⁵

Jean Bourhis MD, PhD,¹⁶ Hanjie Shen MS,⁴ Jonathan Harris MS,¹⁷

Stefan Michiels PhD,^{2,3} Jean-Pierre Pignon MD, PhD,^{2,3} Loren K Mell MD⁴

on behalf of MARCH/MACH-NC Collaborative Group* and

the Head and Neck Cancer Intergroup (HNCIG)

*Complete lists of the members are provided in references 2 and 14

¹Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

²Ligue Nationale Contre le Cancer Meta-analysis Platform, Service de Biostatistique et d'Epidémiologie, Gustave Roussy, Univ Paris-Saclay, Villejuif, France

³CESP, Inserm U1018, Univ. Paris Sud, Univ. Paris-Saclay, Villejuif, France

⁴Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, CA, USA

⁵Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA, USA

⁶Radiation Oncology Department, Centre Léon Bérard, Lyon, France

⁷Department of Experimental Clinical Oncology, Aarhus University Hospital, Aarhus, Denmark

⁸Cancer Research UK & University College London Cancer Trials Centre, Cancer Institute, University College London Hospital, London, UK

⁹Department of Radiation Sciences - Oncology, Umeå University, Umeå, Sweden

¹⁰Medical Research Council Clinical Trials Unit at University College London, London, UK

¹¹Yale University, New Haven, Connecticut, USA

¹²Oncology Unit 2, Istituto Oncologico Veneto-IRCCS, Padua, Italy

¹³Department of Radiation Oncology, IRCCS Regina Elena National Cancer Institute, Rome, Italy

¹⁴Department of Radiation Oncology, Princess Margaret Cancer Centre, University of Toronto, Canada

¹⁵EORTC Headquarters, Brussels, Belgium

¹⁶Department of Radiotherapy, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

¹⁷NRG Oncology Statistics and Data Management Center, American College of Radiology, Philadelphia, PA, USA

Corresponding Author:

Loren K. Mell, M.D.

Professor and Vice Chair of Clinical and Translational Research

Department of Radiation Medicine and Applied Sciences

University of California San Diego

La Jolla, California

lmell@ucsd.edu

ABSTRACT

Background: We sought to test the hypothesis that the effectiveness of intensive treatment for locoregionally advanced head and neck cancer (LAHNC) depends on the proportion of patients' overall event risk attributable to cancer.

Methods: We analyzed 22,339 LAHNC patients treated on 81 randomized trials testing AFX (MARCH data set) or chemotherapy (MACH-NC dataset). We applied generalized competing event regression to the control arms in MARCH, stratifying patients by tertile according to ω score, which quantifies the relative hazard for cancer vs. competing events. We externally validated the classifier on the MACH-NC dataset. We tested for interactions between ω score and treatment effects on overall survival (OS).

Results: Factors associated with a higher ω score were younger age, better performance status, oral cavity site, higher T and N category, and p16 negative/unknown status. The effect of AFX on OS was greater in patients with high (hazard ratio [HR], 0.92; 95% confidence interval [CI], 0.85 to 0.99) and medium (HR, 0.91; 95% CI, 0.84 to 0.98) vs. low ω score (HR, 0.97; 95% CI, 0.90 to 1.05; $p=0.086$ for interaction). The effect of chemotherapy on OS was significantly greater in patients with high (HR, 0.81; 95% CI, 0.75 to 0.88) and medium (HR, 0.86; 95% CI, 0.78 to 0.88) vs. low ω score (HR, 0.96; 95% CI, 0.86 to 1.08; $p=0.011$ for interaction).

Conclusion: LAHNC patients with a higher risk of cancer progression relative to competing mortality – as reflected by a higher ω score – selectively benefit from more intensive treatment.

Funding: Ligue National Contre le Cancer, Institut National du Cancer (PHRC-K 10-02-58; PHRC-K-15-189), Programme d'Actions Intégrées de Recherche (2011-1-VADS-04), UC San Diego Head and Neck Cancer Center.

Key words: Head and Neck Cancer; Chemotherapy; Altered Fractionation; Competing Risks

Precis: The ω score measures the proportion of patients' overall event risk attributable to cancer. We created and externally validated an ω score using two large meta-analyses in head and neck cancer to predict the benefit of intensive treatment.

Manuscript word count: 3000

Pages: 15, **Tables:** 1, **Figures:** 5

Conflicts of Interest: KZ, BL, LKV, QTL, VG, JO, AH, BZ, MKBP, BB, MGG, GS, BO, CF, JB, HS, JH, SM, JPP, LKM – none; FR is an employee of Innate Pharma, Marseille (France)

Author contributions:

Kaveh Zakeri: conceptualization, data curation, formal analysis, investigation, methodology, original draft, review and editing

Federico Rotolo: data curation, formal analysis, investigation, methodology, original draft, review and editing

Benjamin Lacas: data curation, formal analysis, investigation, methodology, original draft, review and editing

Lucas K Vitzthum: data curation, formal analysis, investigation, methodology, original draft, review and editing

Quynh-Thu Le: review and editing

Vincent Gregoire: review and editing

Jens Overgaard: review and editing

Allan Hackshaw: review and editing

Björn Zackrisson: review and editing

Mahesh KB Parmar: review and editing

Barbara Burtness: review and editing

Maria Grazia Ghi: review and editing

Giuseppe Sanguineti: review and editing

Brian O'Sullivan: review and editing

Catherine Fortpied: review and editing

Jean Bourhis: review and editing

Hanjie Shen: review and editing

Jonathan Harris: review and editing

Stefan Michiels: data curation, formal analysis, investigation, methodology, original draft, review and editing

Jean-Pierre Pignon: data curation, formal analysis, funding acquisition, investigation, methodology, original draft, review and editing

Loren K Mell: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, original draft, review and editing

INTRODUCTION

Head and neck cancer (HNC) is a global health burden requiring a multi-disciplinary approach to individualize treatments and manage toxicities. Meta-analyses of randomized clinical trials have shown improved survival with the addition of chemotherapy to radiotherapy for locoregionally advanced HNC (LAHNC).^{1,2} However, a lack of benefit for chemotherapy has been observed in older patients and those with poor performance status, leading to controversy about the management of these patients.¹⁻⁴ As a result, chemotherapy utilization patterns vary widely between institutions and providers.⁵⁻⁸

Management of patients with LAHNC is uniquely complicated by their multiple comorbidities and variable tolerance for treatment.⁹ Many have argued that advanced age alone should not be a contraindication to chemotherapy.^{4,6} However, there is presently no validated risk prediction model for chemotherapy effectiveness. Nomograms and risk models that incorporate individual patient's overall disease and health status would be preferable to age thresholds.

We sought to develop and validate a model to predict the benefit of treatment intensification for patients with LAHNC, using a novel methodological approach called generalized competing event (GCE) regression,^{10,11} and to compare this to a standard modeling approach for risk-stratification. In contrast to models that predict patients' overall survival (OS) or progression-free survival (PFS), the GCE method is designed to stratify patients according to their hazard for primary events (e.g. cancer recurrence/progression) vs. competing events (e.g. death from non-cancer causes). Because the benefit of an intensive cancer treatment depends on its ability to reduce cancer progression to a greater extent than it increases competing mortality, we hypothesized that stratifying patients according to the ratio of the two events would improve selection of patients for treatment intensification strategies.

METHODS

Study Design, Population, and Sampling Methods

The study design was based on two individual patient-level meta-analyses of randomized controlled trials. Analysis and reporting conformed to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines, and followed a pre-specified protocol for the analysis.¹² We studied patients with previously untreated LAHNC who were candidates for standard radiation therapy (RT) with or without treatment intensification (i.e., altered fractionation (AFX) or chemotherapy). For model training, we used the control group from the first update of the Meta-Analysis of Radiotherapy in squamous cell Carcinomas of Head and neck (MARCH) dataset.^{13,14} The control group of MARCH was used to fit a model to stratify patients according to the ratio of the baseline cause-specific hazard for primary vs. competing events (ω ratio), without regard to the

effect of treatment (i.e., unsupervised classification). We tested the magnitude of effect of AFX according to different values of the ω ratio determined by the model, using both arms of MARCH.

We then tested the model's ability to stratify patients by ω ratio and predict the effectiveness of chemotherapy, using the first update of the Meta-Analysis of Chemotherapy in Head and Neck Cancer (MACH-NC). This data set compares loco-regional treatment versus the same loco-regional treatment + chemotherapy; 29% of patients received concurrent chemotherapy.^{1,2} Trials with multiple experimental arms and 2x2 designs were treated as separate comparisons, with the control arm duplicated. There were 190 patients from 3 trials common to MARCH and MACH-NC. A flow diagram of the inclusions and exclusions is provided in **Figure 1**. Due to low frequency in other subsites, only patients with hypopharyngeal, oropharyngeal, laryngeal, and oral cavity cancers were included in this analysis.

Endpoints

OS was defined as time from randomization to death from any cause, or censoring. PFS was defined as time from randomization to a first primary cancer event (recurrence, progression, or death from primary cancer without previous recurrence or progression), or death from any cause, or censoring. Time to progression was defined as time from randomization to first cancer event. Time to competing mortality was defined as time from randomization to death in the absence of any cancer event, as defined above. Event times for patients who were alive without progression were censored at the last follow-up. The baseline ω value was defined as the cumulative hazard for a primary cancer event divided by the cumulative hazard for any PFS event at 3 years (*basehaz* function, R). If the cause of death was "cancer" (i.e. the index cancer), or the cause of death was unknown and the time to death was ≤ 5 years, this was considered a primary cancer event, consistent with prior methodology.¹⁴

Model Building and Specification

Models were initially trained using the patients in MARCH who were treated on a control arm (5,480 patients). Backward stepwise mixed effects multivariable Cox proportional hazards models (*coxme* package, R) were used to identify independent predictors of PFS, which is a strong surrogate endpoint for OS in HNC.¹⁵ Candidate covariates for inclusion were: age (continuous), sex, performance status (Eastern Cooperative Oncology Group (ECOG) status 0 vs. 1 vs. ≥ 2), disease site (oral cavity vs. oropharynx vs. larynx vs. hypopharynx), tumor stage (T0-1 vs. T2 vs. T3 vs. T4), nodal stage (N0 vs. N1 vs. N2 vs. N3), p16 status (positive vs. negative), smoking status (current vs. former/never smoker), and pack-years (≥ 10 vs. < 10). Postoperative status was treated as a trial-level covariate. Variables were retained for $p < 0.20$.

Individual risk scores for PFS were created by taking the inner product of the coefficients from the model with each patient's covariate vector (i.e., linear predictor). GCE models were then fit to estimate the effects of covariates on time to progression vs.

competing mortality (specifically, the ω ratio) respectively, using the same covariates as in the model for PFS with the *gcerisk* package in R.¹⁶ Details of the model are provided in the **Supplementary Methods**. Briefly, this modeling approach consisted of the following:

- PFS was treated as the composition of two mutually exclusive events: (1) any primary cancer event or (2) competing mortality in the absence of a primary event (i.e., competing event).
- A cause-specific Cox proportional hazards model was fit for time to progression and for competing mortality using the same (normalized) covariates.
- The coefficient estimates from both models were differenced to generate the respective GCE coefficient estimates.
- The sample was stratified by the GCE linear predictor (i.e., created by taking the inner product of the GCE coefficient vector and the covariate vector). This method separates patients according to their predicted ω (ω score), i.e., the product of the exponentiated GCE risk score and the predicted baseline ω for the sample.

Models were stratified by trial and included a random effects term for age grouped by trial (i.e., one effect per trial). The rationale for using age as a random effect was that age could play a different prognostic role across trials because of heterogeneous comorbidities and inclusion criteria, possibly impacting the risk of competing mortality. Retained covariates in the final models were age, performance status, tumor site, sex, tumor and nodal stage, p16 status, smoking, and pack years. p16 and smoking were available for 18% and 49% of patients in MARCH, respectively, and not available in MACH-NC (**Supplementary Table 1**), so we examined three potential model specifications: (1) excluding p16, smoking status, and pack-years from the model, (2) treating missing values as a separate category, (3) attributing missing p16, smoking status, and pack-years to the reference groups (p16 negative, non-smoker, < 10 pack-years, respectively). All three models performed similarly, without meaningful changes in the estimates for the other covariates (data not shown).

Model Validation, Sensitivity Testing, and Hypothesis Testing

The ability to stratify patients by ω ratio was tested independently in both the MARCH experimental group and the MACH-NC control group. We also compared cumulative incidences of primary cancer events and competing mortality within high, medium, and low risk groups defined by the GCE model (ω score) and the Cox model for PFS. Patients were divided into low, medium, and high risk groups according to tertiles of the PFS risk score and ω score, respectively. Forest plots were presented to show further gradation in effects according to ω score. Calibration was checked by plotting cumulative incidences of cause-specific events in the patients from MARCH who were treated on an experimental arm (5,694 patients).

For sensitivity analyses, we also tested a model in which unknown cause of death for any reason was considered a competing event. Model calibration was tested by comparing predicted vs. observed ω values (i.e., ω plots). We tested the effect of chemotherapy on overall survival in all patients based on the pre-defined risk groups as well as by stratifying into tertiles with even numbers of events within each subgroup. We also compared the GCE risk strata to age > 70 threshold for the purpose of predicting the effects of treatment intensification, since previous studies^{2,14} have noted interactions between age and these treatment effects. Survival curves stratified on trial were computed for control and experimental groups using Peto's method and were used to calculate absolute benefit at 5 years.¹⁷

Effects of treatment intensification (i.e., AFX or chemotherapy) on OS were estimated using mixed effects Cox models (coxme package, R) stratified by trial, with a random treatment by trial interaction (age effect not included). Stratified survival curves were generated to compare event probabilities by treatment and risk group.¹⁷ They were calculated with an annual hazard ratio (HR) estimated using a fixed effect meta-analysis model. HRs were based on the Peto estimator, estimated with the logrank test (without random effect). Interactions between treatment effects and ω score were tested both on the continuous scale and across risk strata (i.e., trend test). Two-sided P values < 0.05 were considered statistically significant. Analyses were by intention-to-treat.

RESULTS

Risk Model Results and Comparison

Patient characteristics for MARCH (30 trials; 11,174 patients) and MACH-NC (54 trials; 11,355 patients) are provided in **Supplementary Table 1**, and have been reported elsewhere.^{1,2,13,14,18} In the MARCH control group, factors associated with higher ω score - indicating a higher hazard for cancer recurrence/progression relative to competing mortality - were younger age, better performance status, oral cavity tumors, higher tumor and nodal stage, and female sex, whereas laryngeal site, p16 positive status, current smoking, and higher pack-years were associated with a lower score (**Table 1**). For example, increasing age was associated with an increased HR for PFS, but because age increased the HR for competing mortality more than the HR for a cancer event, the *relative* HR was lower in the GCE model. While sex, laryngeal site, and smoking information were not statistically significantly associated with ω score at $p < 0.05$, they were retained in the stepwise model according to the pre-specified analysis plan. The values of the GCE risk score that define the tertile risk groups were: < -25, -25 to 25, and > 25, respectively, with corresponding ω scores of < 0.80, 0.80 to 0.87, and > 0.87, respectively (**Supplementary Methods**).

The Cox model to predict PFS stratified patients satisfactorily according to primary cancer events, but not competing mortality events; in contrast, the GCE model stratified patients well according to both primary cancer events and competing events, increasing the separation between these events within risk strata (**Figure 2**). These patterns were replicated on external validation in

the MACH-NC control group (**Figure 2**), indicating good model performance in both the training and validation data set. With both models, the risk for primary events significantly increased with higher risk score ($p < 0.001$ for trend tests). However, with the GCE model, the risk for competing events significantly decreased with higher risk score, whereas with the standard PFS model the risk for competing events significantly increased with higher risk score ($p < 0.001$ for trend tests), which highlights the principal difference in the two modeling approaches. The distribution of covariates by GCE risk group are provided for the MACH-NC control group in **Supplementary Table 2**. The distribution of event types by GCE risk group for MARCH and MACH-NC are provided in **Supplementary Table 3**.

Effects of AFX and Chemotherapy

In the MARCH data, the effect of AFX on OS was greater in patients with high (HR=0.92 [0.85-0.99]) and medium (HR [95% CI=0.91[0.84-0.98]]) vs. low ω scores (HR=0.97 [0.90-1.05]) – indicating patients with higher ω ratios appear to preferentially benefit from AFX, according to model predictions (**Supplementary Figure 1**). Tests for interaction between AFX effect and ω score were not statistically significant ($p=0.086$; trend test $p=0.15$), except in patients with ω score above the 85th percentile ($p=0.021$). The effect of hyperfractionation on OS was also greater with increasing ω score (HR=0.86 [0.74-1.00; p for interaction 0.054). In contrast, the Cox model to predict PFS did not identify (p for interaction 0.48) differences in the effect of AFX on OS according to risk group (low: HR=0.94, 95% CI 0.86-1.03; medium: HR=0.92, 95% CI 0.86-1.00; high: HR=0.94, 95% CI 0.87-1.01).

In the MACH-NC data, the effect of chemotherapy on OS was statistically significant in patients with high (HR=0.81, 95% CI 0.75-0.88) and medium (HR=0.86, 95% CI 0.78-0.93) ω scores, but not in patients with a lower ω score (HR=0.96 [0.86-1.08]). Tests for interaction between chemotherapy effect and ω score were statistically significant ($p=0.011$; trend test $p=0.0047$). Conversely, the effect of chemotherapy on OS was statistically significant in all risk groups defined by the Cox model to predict PFS (low: HR=0.79, 95% CI 0.69-0.91; medium: HR=0.82, 95% CI 0.74-0.90; high: HR=0.86, 95% CI 0.78-0.93), indicating the inability of this model (p for interaction 0.22) to select patients that benefit more or less from chemotherapy (**Figures 3-4**). Note in **Figure 4C** that the effect of chemotherapy was greater in patients with the most favorable PFS (lowest risk score group). When we restricted analysis only to trials testing cisplatin-based or concurrent chemotherapy, results of interaction tests with ω score were similar (cisplatin-based: $p=0.06$; concurrent: $p=0.04$).

Model Validation, Sensitivity Analysis, Exploratory Analyses

Calibration plots showed high correspondence between the observed and predicted ω ratios at varying risk score cutoffs (**Supplementary Figure 2**). Results of sensitivity analysis showed similar model estimates regardless of how unknown cause of death

was handled. We conducted exploratory analyses to test for treatment interactions with ω score in older patients, because there is significant ongoing controversy in the management of this population. Patients older than 70 comprised 13.8% (n=1540) of the MARCH dataset and 8.0% (n=905) of the MACH-NC dataset. On the whole, patients older than 70 appeared not to benefit from AFX (HR=1.00; 95% CI 0.89-1.12; p=0.14 for interaction) or chemotherapy (HR=0.97; 95% CI 0.82-1.14; p=0.03 for interaction), as previously reported.^{1,2,13} In patients older than 70, the effect of AFX did not differ significantly with respect to ω score ≥ 0.80 vs. < 0.80 (HR=1.11 vs 1.00, interaction p=0.44) (**Supplementary Figure 3**). However, the effect of chemotherapy on OS in patients older than 70 with ω score ≥ 0.80 (HR=0.80; 95% CI 0.58-1.12), appeared similar to the effectiveness of chemotherapy in the whole population (HR=0.84; 95% CI 0.80-0.89), in contrast to those with ω score < 0.80 (HR=1.02; 95% CI 0.83-1.24; interaction p=0.19) (**Figure 5**).

DISCUSSION

In this study, we found that stratifying LAHNC patients according to their risk of cancer event (i.e., recurrence/progression) *relative to competing mortality* is useful in identifying patients who selectively benefit from intensive treatment. In particular, LAHNC patients with a ω score < 0.80 did not benefit from chemotherapy, in contrast to patients with a ω score ≥ 0.80 . The effect of AFX on OS, albeit weaker than chemotherapy, also appeared greater in patients with higher ω score. Note that the model we developed to predict ω scores (independent of treatment) performed well across two separate populations and treatment types, which is a more stringent test of model validity than had we developed separate models for each data set. Moreover, models to stratify patients by PFS did not show significantly different predicted treatment effects across strata. The ω score is simple to calculate and could influence clinical practice in situations where there is uncertainty about whether or not treatment intensification would be beneficial.

A remarkable finding was that patients with better outcomes (in terms of better PFS) were no less likely to benefit from intensive therapy (**Figure 4**), counter to prevailing risk-stratification schemes, which typically steer patients with better expected PFS toward de-intensification. A problem with using expected PFS and OS to define risk is that these endpoints treat primary and competing events interchangeably, placing patients with a high risk for non-cancer death into the same risk pool as those with a high risk for cancer recurrence and mortality. The strength of the GCE model, in contrast, stems from its ability to select patients at higher risk of cancer event relative to competing mortality. Even small variations in hazard for competing events can meaningfully impact this ratio, thus the effect of treatment, which in turn affects the efficiency of randomized trials.^{19,20}

Using ω scores may be preferable to age thresholds for determining the appropriate intensiveness of therapy in LAHNC and other populations with competing risks. Age thresholds are often used in treatment selection for simplicity, but are generally not

avored since they do not consider a patient's full health status. ω scores have the advantage of incorporating multiple prognostic characteristics, including age, to provide a fuller assessment of a patient's overall risk for cancer event relative to competing mortality. Incorporating measures of comorbidity, such as the Charlson or ACE-27 indices, and more comprehensive geriatric assessments that measure frailty, psychosocial status, and cognitive function, into future models could improve treatment decisions for older patients.²¹⁻²³ Unfortunately, meta-analyses including these variables are presently lacking.

This study was strengthened by external validation using robust meta-analyses of randomized clinical trials in LAHNC that tested different types of treatment. The effectiveness of AFX and chemotherapy are already known, which allowed us to determine whether subgroups of patients with higher ω scores selectively benefitted from these treatments. The MARCH and MACH-NC meta-analyses also contained important tumor and patient characteristics as well as competing mortality information (i.e., death in the absence of a primary cancer event) that facilitated our approach to risk modeling. However, our study was limited by low numbers of older patients and lack of some known predictors, including p16 status, smoking history, and comorbidity; inclusion of these and other covariates would likely enhance the discriminatory power of future ω score classifiers.²⁴ In addition, RT techniques and supportive care have changed considerably over the study period, which could have influenced our results.

Despite such limitations, the ω score validated in our study outperformed models of PFS for predicting the effectiveness of AFX and chemotherapy. It should be noted that our sample represents patients selected for participation in randomized trials; thus, how these results will translate into more general populations is not entirely clear. However, previous work, including population-based studies, has shown the validity of this modeling approach in varying contexts.^{10,11,25-27} As seen in **Supplementary Table 2**, younger, female patients with good performance status may preferentially benefit from intensified therapy due to their higher ω score, which is consistent with findings from other sources.²⁶

In conclusion, we found that LAHNC patients with a higher risk of cancer progression relative to competing mortality - as reflected by a higher ω score - selectively benefit from more intensive treatment. Inclusion of additional prognostic variables could further improve the selection of patients for optimal therapeutic intensity.

ACKNOWLEDGMENT

We thank the investigators and the associated research groups of the MACH-NC and MARCH project, and the European Organisation for Research and Treatment of Cancer (EORTC) for providing study data. The Meta-analysis Platform is supported by grants from Ligue National Contre le Cancer and the MACH-NC and MARCH project by grants from Institut National du Cancer (PHRC-K 10-02-58; PHRC-K-15-189), the Programme d'Actions Intégrées de Recherche (2011-1-VADS-04), the La Jolla Center for Precision Radiation Medicine, and the University of California San Diego Head and Neck Cancer Center.

REFERENCES

1. Pignon JP, Bourhis J, Domenge C, Designé L. Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: three meta-analyses of updated individual data. MACH-NC Collaborative Group. Meta-Analysis of Chemotherapy on Head and Neck Cancer. *Lancet*. 2000; 355: 949-55.
2. Pignon JP, le Maitre A, Maillard E, Bourhis J. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17 346 patients. *Radiother Oncol*. 2009; 92: 4–14.
3. Bonner JA, Harari PM, Giralt J, Cohen RB, Jones CU, Sur RK, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *Lancet Oncol*. 2010;11(1):21-8.
4. Porceddu SV, Haddad RI. Management of elderly patients with locoregionally confined head and neck cancer. *Lancet Oncol*. 2017;18(5):e274-e283.
5. Amini A, Jones BL, McDermott JD, et al. Survival outcomes with concurrent chemoradiation for elderly patients with locally advanced head and neck cancer according to the National Cancer Data Base. *Cancer*. 2016; 122: 1533–43.
6. Juarez JE, Choi J, St John M, Abemayor E, TenNapel M, Chen AM. Patterns of Care for Elderly Patients With Locally Advanced Head and Neck Cancer. *Int J Radiat Oncol Biol Phys*. 2017;98(4):767-774.
7. VanderWalde NA, Meyer AM, Liu H, Tyree SD, Zullig LL, Carpenter WR, et al. Patterns of care in older patients with squamous cell carcinoma of the head and neck: a surveillance, epidemiology, and end results-medicare analysis. *J Geriatr Oncol*. 2013;4(3):262-70.
8. VanderWalde NA, Meyer AM, Deal AM, et al. Effectiveness of chemoradiation for head and neck cancer in an older patient population. *Int J Radiat Oncol Biol Phys*. 2014;89: 30–37.
9. Machtay M, Moughan J, Trotti A, et al. Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: an RTOG analysis. *J Clin Oncol*. 2008;26:3582–89.

10. Carmona R, Zakeri K, Green G, Hwang L, Gulaya S, Xu B, et al. Improved Method to Stratify Elderly Patients with Cancer at Risk for Competing Events. *J Clin Oncol*. 2016;34(11):1270-7.
11. Carmona R, Gulaya S, Murphy JD, Rose BS, Wu J, Noticewala S, et al. Validated competing event model for the stage I-II endometrial cancer population. *Int J Radiat Oncol Biol Phys*. 2014;89(4):888-98.
12. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Br J Cancer*. 2015;112(2):251-9.
13. Bourhis J, Overgaard J, Audry H, et al, for the Meta-Analysis of Radiotherapy in Carcinomas of Head and Neck (MARCH) Collaborative Group. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet* 2006; 368: 843–54.
14. Lacas B, Bourhis J, Overgaard J, Zhang Q, Grégoire V, Nankivell M, et al. Role of radiotherapy fractionation in head and neck cancers (MARCH): an updated meta-analysis. *Lancet Oncol*. 2017;18(9):1221-1237.
15. Michiels S, Le Maître A, Buyse M, Burzykowski T, Maillard E, Bogaerts J, et al. Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analyses of individual patient data. *Lancet Oncol*. 2009;10(4):341-50.
16. <https://cran.r-project.org/web/packages/gcerisk/gcerisk.pdf>, accessed March 5, 2019.
17. Lueza B, Rotolo F, Bonastre J, Pignon JP, Michiels S. Bias and precision of methods for estimating the difference in restricted mean survival time from an individual patient data meta-analysis. *BMC Med Res Methodol*. 2016;16:37.
18. Lassen P, Lacas B, Pignon JP, Trotti A, Zackrisson B, Zhang Q, Overgaard J, Blanchard P; MARCH Collaborative Group. Prognostic impact of HPV-associated p16-expression and smoking status on outcomes following radiotherapy for oropharyngeal cancer: The MARCH-HPV project. *Radiother Oncol*. 2018;126(1):107-115.
19. Zakeri K, Rose BS, Gulaya S, D'Amico AV, Mell LK. Competing event risk stratification may improve the design and efficiency of clinical trials: secondary analysis of SWOG 8794. *Contemp Clin Trials*. 2013;34(1):74-9.
20. Zakeri K, Rose BS, D'Amico AV, Jeong JH, Mell LK. Competing events and costs of clinical trials: Analysis of a randomized trial in prostate cancer. *Radiother Oncol*. 2015;115(1):114-9.
21. Hamaker ME, Jonker JM, de Rooij SE, Vos AG, Smorenburg CH, van Munster BC. Frailty screening methods for predicting outcome of a comprehensive geriatric assessment in elderly patients with cancer: a systematic review. *Lancet Oncol*. 2012;13(10):e437-44.
22. Puts MT, Hardt J, Monette J, Girre V, Springall E, Alibhai SM. Use of geriatric assessment for older adults in the oncology setting: a systematic review. *J Natl Cancer Inst*. 2012;104(15):1133-63.

23. VanderWalde NA, Deal AM, Comitz E, Stravers L, Muss H, Reeve BB, et al. Geriatric Assessment as a Predictor of Tolerance, Quality of Life, and Outcomes in Older Patients with Head and Neck Cancers and Lung Cancers Receiving Radiation Therapy. *Int J Radiat Oncol Biol Phys.* 2017;98(4):850-857.
24. Reid BC, Alberg AJ, Klassen AC, et al. Comorbidity and survival of elderly head and neck carcinoma patients. *Cancer.* 2001; 92: 2109–16.
25. Mell LK, Shen H, Nguyen-Tân PF, Rosenthal DI, Zakeri K, Vitzthum LK, et al. Nomogram to predict the benefit of intensive treatment for locoregionally advanced head and neck cancer. *Clin Cancer Res.* 2019;25:7078-7088.
26. Park A, Alabaster A, Shen H, Mell LK, Katzel JA. Undertreatment of women with locoregionally advanced head and neck cancer. *Cancer.* 2019;125(17):3033-3039.
27. Vitzthum LK, Park H, Zakeri K, Bryant AK, Feng C, Shen H, et al. Selection of head and neck cancer patients for intensive therapy. *Int J Radiat Oncol Biol Phys.* 2020;106:157-166.

FIGURE LEGENDS

Figure 1. Flow diagram of the datasets used in this analysis. The overall number of patient and trials is 22,339 and 81 respectively as some multiple arm trials are included in both meta-analyses. †: Other than nasopharynx, larynx, hypopharynx, oral cavity, oropharynx; *: No smoking status or p16 data in the MACH-NC dataset; ‡: 190 patients included in both meta-analyses: 1 trial + 1 arm from 2 three-arm trials.

Figure 2. Cumulative incidences of primary cancer events (solid lines) and competing mortality (dotted lines) according to risk group, illustrating that the GCE model better maximizes the ratio of primary events to competing events. (A) Risk groups defined by progression-free survival (PFS) in the MARCH data (control arm only). (B) Risk groups defined by ω score in the MARCH data (control arm only). (C) Risk groups defined by PFS in the MARCH data (experimental arm only). (D) Risk groups defined by ω score in the MARCH data (experimental arm only). (E) Risk groups defined by PFS in the MACH-NC data (control arm only). (F) Risk groups defined by ω score in the MACH-NC data (control arm only). High Risk = green; Intermediate Risk = red; Low Risk = blue.

Figure 3. Survival curves in patients undergoing chemotherapy (squares) compared to controls (circles) in MACH-NC showing that the GCE model is predictive of treatment effect while the PFS model is prognostic. (A) High-risk group, GCE Model. (B) Medium-risk group, GCE Model. (C) Low-risk group, GCE Model. (D) High-risk group, Cox Model for PFS. (E) Medium-risk group, Cox Model for PFS. (F) Low-risk group, Cox Model for PFS.

Figure 4. Forest plots of effectiveness of altered fractionation (A & B) and chemotherapy (C & D) on overall survival in the MARCH and MACH-NC datasets, respectively, for the Cox model of PFS (A & C) compared to the GCE model (B & D), by risk tertile. Hazard ratios less than 1.0 favor treatment intensification.

Figure 5. Survival curves by age group, ω score, and treatment with chemotherapy. (A) Age < 70, ω score < 0.80 (1403 patients) (B) Age < 70, ω score \geq 0.80 (9047 patients) (C) Age \geq 70, ω score < 0.80 (636 patients) (D) Age \geq 70, ω score \geq 0.80 (269 patients).

TABLES

Table 1. Cox model of progression-free survival vs. GCE regression model for patients in the control arm (i.e., standard radiotherapy) of the MARCH dataset (N= 5,480).

Covariate	Cox Model		GCE Model	
	HR (95% CI)	p	RHR (95% CI)	p
Age (normalized)	1.08 (1.02-1.14)	<0.01	0.68 (0.63-0.74)	<0.001
Female sex	0.95 (0.92-0.98)	<0.001	1.04 (0.97-1.12)	0.28
Performance status	Reference		Reference	
0	Reference		Reference	
1	1.13 (1.10-1.17)	<0.001	0.86 (0.80-0.93)	<0.001
2	1.13 (1.10-1.16)	<0.001	0.91 (0.86-0.97)	<0.01
Anatomic subsite	Reference		Reference	
Hypopharynx	Reference		Reference	
Oral cavity	1.02 (0.98-1.06)	0.38	1.19 (1.07-1.33)	<0.01
Oropharynx	0.88 (0.84-0.93)	<0.001	1.04 (0.92-1.17)	0.57
Larynx	0.87 (0.83-0.92)	<0.001	0.91 (0.81-1.03)	0.14
T stage	Reference		Reference	
T0/T1	Reference		Reference	
T2	1.12 (1.06-1.20)	<0.001	1.23 (1.09-1.39)	<0.001
T3	1.29 (1.21-1.38)	<0.001	1.22 (1.08-1.38)	<0.01
T4	1.40 (1.31-1.49)	<0.001	1.25 (1.11-1.42)	<0.001
N stage	Reference		Reference	
N0	Reference		Reference	
N1	1.10 (1.07-1.14)	<0.001	1.06 (0.97-1.15)	0.20
N2	1.19 (1.15-1.24)	<0.001	1.13 (1.04-1.24)	<0.01
N3	1.18 (1.15-1.22)	<0.001	1.12 (1.02-1.21)	0.011
p16 status	Reference		Reference	
Negative / Unknown	Reference		Reference	
Positive	0.87 (0.83-0.90)	<0.001	0.88 (0.81-0.95)	<0.01
Current Smoker	Reference		Reference	
No or unknown	Reference		Reference	
Yes	1.11 (1.06-1.16)	<0.001	0.99 (0.92-1.08)	0.93
≥10 Pack-Years	Reference		Reference	
No or unknown	Reference		Reference	
Yes	1.06 (1.01-1.11)	0.018	0.93 (0.86-0.99)	0.048

Abbreviations: CI= confidence interval; MARCH=Meta-Analysis of Radiotherapy in Carcinomas of Head and neck; GCE=generalized competing event; HR=hazard ratio for progression-free survival; RHR=relative hazard ratio for progression vs. competing mortality.