A clinicogenomic model including GARD predicts outcome for radiation treated patients with HPV+ oropharyngeal squamous cell carcinoma

Emily Ho[1,2,†], Loris De Cecco[3,†], Stefano Cavalieri[4,5], Geoffrey Sedor[6], Frank Hoebers[7], Ruud H Brakenhoff[8], Kathrin Scheckenbach[9], Tito Poli[10], Kailin Yang[11], Jessica A. Scarborough[1,2,12], Shauna Campbell[11], Shlomo Koyfman[11], Steven A. Eschrich[13], Jimmy J. Caudell[14], Michael W. Kattan[15], Lisa Licitra[4,5,*], Javier F. Torres-Roca[13,14,*], and Jacob G. Scott[1,2,11,12,*]

[1]Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH
[2]School of Medicine, Case Western Reserve University, Cleveland, OH
[3]UO Molecular Mechanisms, Experimental Oncology Department, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[4]Head and Neck Medical Oncology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy
[5]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy
[6]Radiation Oncology Department, NYPH/Columbia University Vagelos College of Physicians and Surgeons, New York, NY
[7]Department of Radiation Oncology (Maastro), GROW School for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, The Netherlands
[8]Amsterdam UMC, Vrije Universiteit Amsterdam, Otolaryngology/Head and Neck Surgery, Cancer Center Amsterdam, the Netherlands
[9]Department of Otolaryngology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[10]Unit of Maxillofacial Surgery, Department of Medicine and Surgery, University of Parma-University Hospital of Parma, Parma, Italy
[11]Department of Radiation Oncology, Cleveland Clinic, Cleveland, OH
[12]Departments of Physics and Biology, Case Western Reserve University, Cleveland, OH
[13]Department of Biostatistics and Biomedical Informatics, Moffitt Cancer Center, Tampa, FL
[14]Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL
[15]Department of Quantitative Health Sciences, Cleveland Clinic, Clevelan OH
[*]scottj10@ccf.org, javier.torresroca@moffitt.org, lisa.licitra@institutotumori.mi.it
[†]contributed equally

ABSTRACT

Background: Treatment decision-making in oropharyngeal squamous cell carcinoma (OPSCC) includes clinical stage, HPV status, and smoking history. Despite improvements in staging with separation of HPV-positive and -negative OPSCC in AJCC 8th edition (AJCC8), patients are largely treated with a uniform approach, with recent efforts focused on de-intensification in low-risk patients. We have previously shown, in a pooled analysis, that the genomic adjusted radiation dose (GARD) is predictive of radiation treatment benefit and can be used to guide RT dose selection. We hypothesize that GARD can be used to predict overall survival (OS) in HPV-positive OPSCC patients treated with radiotherapy (RT).

Methods: Gene expression profiles (Affymetrix Clariom D) were analyzed for 234 formalin-fixed paraffin-embedded samples from HPV-positive OPSCC patients within an international, multi-institutional, prospective/retrospective observational study including patients with AJCC 7th edition stage III-IVb. GARD, a measure of the treatment effect of RT, was calculated for each patient as previously described. In total, 191 patients received primary RT definitive treatment (chemoradiation or RT alone, and 43 patients received post-operative RT. Two RT dose fractionations were utilized for primary RT cases (70 Gy in 35 fractions or 69.96 Gy in 33 fractions). Median RT dose was 70 Gy (range 50.88-74) for primary RT definitive cases and 66 Gy (range 44-70) for post-operative RT cases. The median follow up was 46.2 months (95% CI, 33.5-63.1). Cox proportional hazards analyses were performed with GARD as both a continuous and dichotomous variable and time-dependent ROC analyses compared the performance of GARD with the NRG clinical nomogram for overall survival.

Results: Despite uniform radiation dose utilization, GARD showed significant heterogeneity (range 30-110), reflecting the underlying genomic differences in the cohort. On multivariable analysis, each unit increase in GARD was associated with an improvement in OS (HR = 0.951 (0.911, 0.993), p = 0.023) compared to AJCC8 (HR = 1.999 (0.791, 5.047)), p = 0.143). ROC analysis for GARD at 36 months yielded an AUC of 80.6 (69.4, 91.9) compared with an AUC of 73.6 (55.4, 91.7) for the NRG clinical nomogram. GARD$\geq$64.2 was associated with improved OS (HR = 0.280 (0.100, 0.781), p = 0.015). In a virtual trial, GARD predicts that uniform RT dose de-escalation results in overall inferior OS but proposes two separate genomic strategies where selective RT dose de-escalation in GARD-selected populations results in clinical equipoise.

Conclusions: In this multi-institutional cohort of patients with HPV-positive OPSCC, GARD predicts OS as a continuous variable, outperforms the NRG nomogram and provides a novel genomic strategy to modern clinical trial design. We propose that GARD, which provides the first opportunity for genomic guided personalization of radiation dose, should be incorporated in the diagnostic workup of HPV-positive OPSCC patients.

Key Words: radiation therapy, genomics, mathematical model, personalized medicine, GARD

INTRODUCTION

Since the discovery that human papillomavirus (HPV) is an etiologic and strong prognostic factor in oropharyngeal squamous cell carcinoma, assessing this biomarker indirectly via p16 or directly via *in situ* hybridization has become standard of care in the diagnostic and staging work up of these patients.[1,2] Ang developed a three-group classification system based on clinical factors (HPV status, pack years of smoking, and T or N classification) that has informed the design of multiple clinical trials in the last decade.[3] As the low risk group in this classification had an overall survival (OS) of 93% at three years, there has been a significant clinical interest in the de-escalation of therapy for these patients. Initial efforts to uniformly de-escalate therapy as in the randomized phase II trial HN002 were promising. Of two potential de-escalated regimens tested, one of them (cisplatin + 60 Gy) met the predefined endpoint (PFS > 85%)[4]. This led to the development of HN005, a Phase 3 clinical trial testing the non-inferiority of uniform RT dose de-escalation. In this trial patients were randomized to one of three arms which included two RT dose de-escalated regimens(cisplatin + 60 Gy or nivolumab + 60 Gy) against the standard of care (cisplatin + 70 Gy). However, the interim analysis failed to demonstrate the non-inferiority of cisplatin + 60 Gy over standard of care. These early results suggest that clinical factors are not enough to provide therapeutic guidance with regards to optimization of RT and suggest that similar to many targeted and immunotherapy agents, RT dose optimization may need to be targeted to specific genomically-defined subpopulations.

In previous studies we developed the genomic adjusted radiation dose (GARD), a radiation-specific metric that quantifies the RT treatment effect in a given patient as a function of their RT dose and tumor genomics. GARD has proposed that the treatment effect of a uniform dose of RT (e.g. 70 Gy) is biologically highly heterogeneous. In a recent pooled analysis of 1,615 patients in seven different disease sites, we demonstrated that GARD was associated with overall survival and recurrence risk as a continuous variable and predicted RT treatment benefit for each individual patient. We proposed that GARD could serve as a decision support tool to optimize and inform the RT dose for each patient.[5,6]

One limitation of this pan-cancer analysis was that many of the individual cohorts were underpowered and that GARD's association with clinical outcome as a continuous variable was only realized when all cohorts were pooled. Since RT benefit is one of the critical factors defining clinical outcome in HPV+ HNC patients, we hypothesized that with a larger, modern cohort of radiation treated patients with HNSCC, GARD would be associated with outcome. Further, we hypothesize that the addition of genomic information can provide an improvement to outcome prediction compared to clinical factors alone. Importantly, any improvement in outcome prediction utilizing GARD can also fundamentally be used to make quantitative predictions for differential outcome given specific RT dose adjustments,

thus providing a critical tool for the definition of sub-populations for clinical trial design.

Here we describe an analysis of patients treated with radiation therapy with HPV+ HSNCC as part of the Big Data to Decide project.[7] We use modern methods to assess individual patient radiation sensitivity indices (RSI)[8] from gene expression data derived from formalin fixed tumor specimens, and use radiation-dosing information for each patient to calculate GARD. We then use continuous Cox proportional hazards regression to determine the relationship between GARD and outcome, and present a discrete analysis both with the median as a cutpoint, and at two cutpoints identified *post hoc* to suggest optimal stratification. We demonstrate that as a prognostic factor, GARD resolves prognosis of HPV+ patients with similar magnitude (e.g. effect size) as HPV does in HNC and outperforms the existing NRG nomogram of clinical outcome. Furthermore, we demonstrate that GARD identifies genomic subpopulations with differential RT therapeutic benefit and incorporating this knowledge into clinical trial design predicts that uniform RT dose de-escalation (cisplatin + 60 Gy) would result in a worse 3 year overall survival than standard of care (cisplatin + 70 Gy). Finally, GARD proposes that through a selective genomic-based approach, RT dose de-escalation is feasible for most HPV+ patients without detriment in clinical outcome (e.g. overall survival).

METHODS

Patient Cohort. Patients in this analysis were part of the Big Data to Decide project (BD2Decide, NCT028322102), a collaboration of seven European centers to develop a clinico-genomic database of head and neck cancer patients. The details of this project have been previously described.[7] Briefly, BD2Decide enrolled a total of 1,537 patients (1,086 retrospectively and 451 prospectively) with loco-regional advanced head and neck cancer (Stage III-IVa, IVb, AJCC 7th edition) treated with curative intent, including 377 patients with HPV+ oropharyngeal cancer. After excluding patients with inadequate tissue sample (n=85), poor RNA quality (n=6) or HPV DNA-negative (n=14), patients with locally-advanced disease that underwent single-modality treatment (n=37) and 1 patient treated with surgery alone (no GARD could be calculated), a final study population of 234 patients remained. The study was approved by institutional review boards of each of the participating institutions and when possible patients consented to enrollment or a waiver for consent was approved. Patients were treated between 2008 and 2017, and follow-up closed in September 2019.

HPV testing was performed with p16 immunohistochemistry and confirmed by HPV DNA testing following positive staining. In total, 191 patients received definitive RT primary treatment (chemoradiation (n=172) or RT alone (n=19)), and 43 patients received post-operative RT (post-op chemoradiation (n=29) or post-op RT alone (n=14)). Two RT dose fractionations were prescribed for definitive RT cases (70 Gy in 35 fractions or 69.96 Gy in 33 fractions). Median RT dose was 70 Gy (range 50.88-74) for primary RT cases and 66 Gy (range 44-70) for

post-operative cases. The median follow up was 46.2 months (IQR 33.5-63.1).

Bioinformatic and Statistical Analysis. All patient tumors underwent gene expression profiling using Affymetrix Clariom D on their formalin fixed samples and RSI values were generated using a 10-gene signature as previously described.[9,10]. RSI has been previously clinically validated in multiple cohorts.[11–13] A patient-specific genomic parameter, α_g, was subsequently calculated using the linear-quadratic model to estimate patient radiosensitivity, the derivation of which we have previously described, yielding the relation:

$$\alpha_g = -\frac{\ln \mathrm{RSI}}{nd} - \beta d, \tag{1}$$

where dose d is 2 Gy, and the number of fractions, n is 1, as this moves from a genomic measure to the familiar Surviving Fraction after 2Gy (SF2). We further make the simplifying assumption that β is a constant at 0.05/Gy^2. This genomic parameter, α_g, is then used together with each patients specific radiation dosing to calculate their clinical GARD value (GARD$_c$).

$$GARD_c = n_c d_c(\alpha_g + \beta d_c), \tag{2}$$

where n_c is the number of fractions and d_c is the dose per fraction per the clinically delivered radiation plan to each patient. These values were calculated without information about clinical outcome, using the described, previously specified model.

Cox proportional hazards regression was used to assess the association between GARD as a continuous variable and overall survival (OS). OS was defined as the time between primary diagnosis and death or last follow-up. While the continuous analysis is statistically the most rigorous, to make clinical translation simpler, we also performed discrete analyses to mimic the discrete dose levels commonly used by clinicians. Discrete analyses with log-rank statistics were performed with median GARD and also, for hypothesis generation, using an algorithm to minimize χ^2 to derive optimal groups in three dose levels.

***In silico* simulation of HN005.** We performed an *in silico* trial designed after HN005 (NCT03952585), a prospective randomized phase 3 clinical trial testing the non-inferiority of RT de-intensification (from 70Gy to 60Gy) in early stage HPV-positive oropharyngeal cancer. Each virtual patient was randomly selected from the RSI distribution of the complete cohort. The virtual patients were randomized to receive 70Gy in 35 fractions or 60Gy in 30 fractions. The estimated enrollment for HN005 was 590, so we used 200 patients per arm (the original trial design had three arms, but this is simulating two of them). GARD was calculated for each virtual patient, and an OS curve was

predicted based on the GARD level achieved, using the optimized three dose GARD level approach (GARD $\leq$ 46 = low, 46< GARD $\leq$ 65 = intermediate, GARD $>$ 65 = high). Each of these survival curves was modeled with a Weibull curve fit to the Kaplan-Meier estimate of that GARD level. The weighted average of individual patients' survival curves represented the overall survival estimate for each trial arm. This entire process was repeated a total of 20 times to replicate the variability between different groups of patients.

To determine whether GARD could identify a successful de-intensification strategy, we also performed a variation of this trial with selective de-intensification based on GARD. In this approach, rather than de-intensifying all patients, only patients who would remain in either GARD dose level high (GARD > 65= high) or intermediate (46< GARD $\leq$ 65 = intermediate) with either 60 or 70 Gy were assigned to 60 Gy. All patients that achieve GARD dose level low (GARD $\leq$ 46 = low) were excluded from selective de-escalation. In addition, patients who drop from GARD high to GARD intermediate when de-escalated were also excluded from selective de-escalation.

Creation of a clinical nomogram. After performing standard analyses to determine the association of GARD with outcome, we incorporated GARD to the established NRG nomogram of outcome (stage, pack-years smoking, ECOG performance status, HPV status, education) to determine whether GARD improved its prognostic performance. In addition, we also integrated the three-cluster gene expression model[9] following standard methods.[14] The nomogram was evaluated by comparison of time-dependent receiver operating characteristic (ROC) curve analysis with the `timeROC` package in R.[15]

RESULTS

GARD reveals underlying genomic heterogeneity in RT effect. We have previously shown that GARD reveals underlying heterogeneity in radiation treatment effect within groups presumed to have been treated uniformly (with approximately equivalent physical dose).[5,6,16,17] In this cohort we demonstrate again that GARD reveals wide heterogeneity in predicted RT effect in spite of relatively uniform RT dose prescribed. As shown in **Figure 1 (Left)**, delivered GARD ranged from 31.7 to 108.9 (IQR 56.4-71.9) with a median for the whole cohort of 63.5. Plotted along the edges of the jointplot between GARD and EQD2 are kernel density estimates for the entire cohort, revealing wide heterogeneity in delivered GARD (std = 13.8) in the setting of near homogeneity in RT dose (std = 3.1). The difference between GARD and EQD2 is best exemplified by the patients who received the whole course of 'standard' radiation dose – with EQD2 measures between 69-71, see **Figure 1 (Right)**. The range of GARD for those patients ranged was 31.7-108.9 (IQR 56.5-71.8) even though they all were treated to (approximately) the same RT dose (EQD2), highlighting the wide differential in the predicted effect of our uniform clinical dosing strategies.

GARD is continuously associated with OS in RT-treated HPV-positive OPSCC patients. Previously, we demonstrated that GARD was associated with overall survival, recurrence risk and was predictive of RT benefit in a pooled pan-cancer analysis including 1,615 patients.[6] Since RT therapeutic benefit is a critical factor impacting clinical outcome in HPV+ patients, we hypothesized that GARD would be associated with clinical outcome in this analysis of HPV+ oropharyngeal squamous cell carcinoma patients collected through the B2DECIDE project.[9] To test this, we performed a Cox proportional hazards analysis of GARD and OS in both the entire cohort, and also the subset that were treated with primary RT. As shown in **Supplemental Figure 1**, GARD is associated with OS as a continuous variable in both groups. Analyzing the entire cohort, we found that for each unit increase in GARD there is an improvement in OS (HR (95% CI) = 0.968 (0.938, 0.999) *per unit* GARD, p = 0.042). This association of GARD with OS was also significant when including only patients treated with primary definitive RT (HR (95% CI) = 0.956 (0.917, 0.996) per unit GARD, p = 0.033). Examining only patients who received the standard-of-care range 69-71 Gy EQD2 within this group, the HR was 0.942 (0.901, 0.984), p = 0.008. While we found significant differences throughout the entire cohort, in order to keep the cohort population uniform, we focus the remainder of the analysis on definitive primary RT patients (191 total).

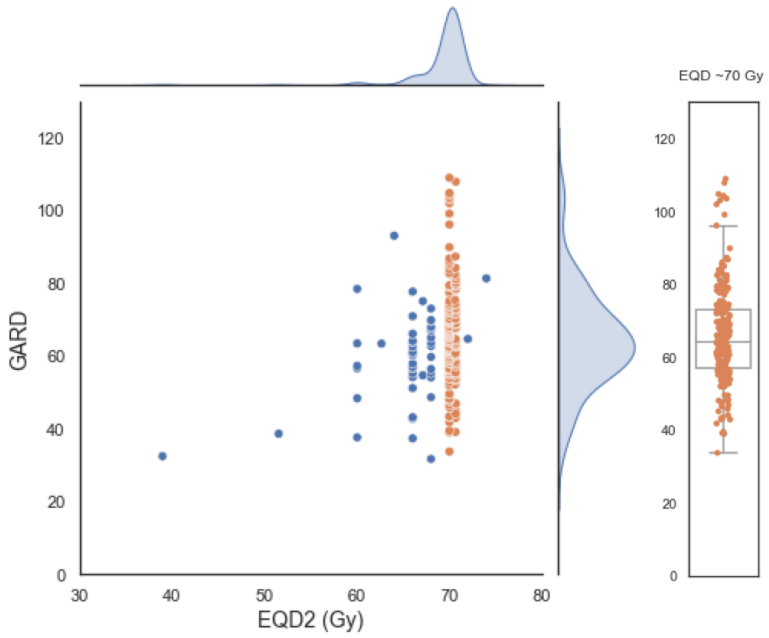


Figure 1. GARD exhibits large underlying genomic heterogeneity in radiation effect compared to radiation dose alone. Left: EQD2 (median 70.0, std 3.2) is plotted against associated GARD (median 63.5, std 13.8) for each patient in the whole cohort. Kernel density estimates are plotted on each edge to show the distributions of the individual variables. Patients are plotted with their GARD on the y-axis and EQD2 on the x-axis as blue dots, with who received an EQD2 of 69-71 Gy only (standard dosing) are colored orange. **Right:** The GARD value of all patients receiving an EQD2 of 69-71Gy (standard of care) are plotted to highlight GARD's ability to stratify patients by their genomic heterogeneity. Data points are overlaid with a box-whisker plot with box representing quartiles and whiskers extending to 1.5 times IQR.

On multivariate analysis, GARD was the only variable statistically associated with OS both as a continuous (HR $= 0.951$ $(0.911, 0.993)$, p $= 0.023$) and discrete variable (HR 0.258 $(0.091, 0.735)$, p $= 0.011$). Smoking (pack-years) was associated with OS as a continuous variable but not as a discrete variable. These results are summarized in **Table 1**.

While significant as a continuous variable, to further evaluate GARD's prognostic ability in a more clinically

	HR (95% CI)	p
GARD (continuous)	0.951 (0.911, 0.993)	**0.023**
stage = high	1.999 (0.791, 5.047))	0.143
ECOG = 1 or 2	0.811 (0.231, 2.851)	0.744
Pack-years (continuous)	1.024 (1.005, 1.043)	**0.012**
GARD (discrete)	0.258 (0.091, 0.735)	**0.011**
stage = high	1.862 (0.742, 4.677)	0.186
ECOG = 1 or 2	0.816 (0.233, 2.860)	0.751
Pack-years (discrete)	2.378 (0.922, 6.131)	**0.073**

Table 1. Multivariable analysis of primary RT patients. GARD is associated with overall survival both as a continuous (p = 0.023) and discrete (p = 0.011) variable.

relevant, discrete dosing paradigm, we performed an analysis using the GARD median as a cut-point (GARD = 64.2) in the patients treated with primary definitive RT. We found that GARD high vs. low significantly associated with OS ($p = 0.01$) (**Figure 2**): GARD-high patients had a 99.0% and 94.6% 3yr and 5 yr-OS rate whereas GARD-low patients achieved a 89.2% and 79.2% 3 yr and 5-yr OS rate, respectively. This dichotomization resulted in a HR of 0.280 (0.100, 0.781), $p = 0.015$.

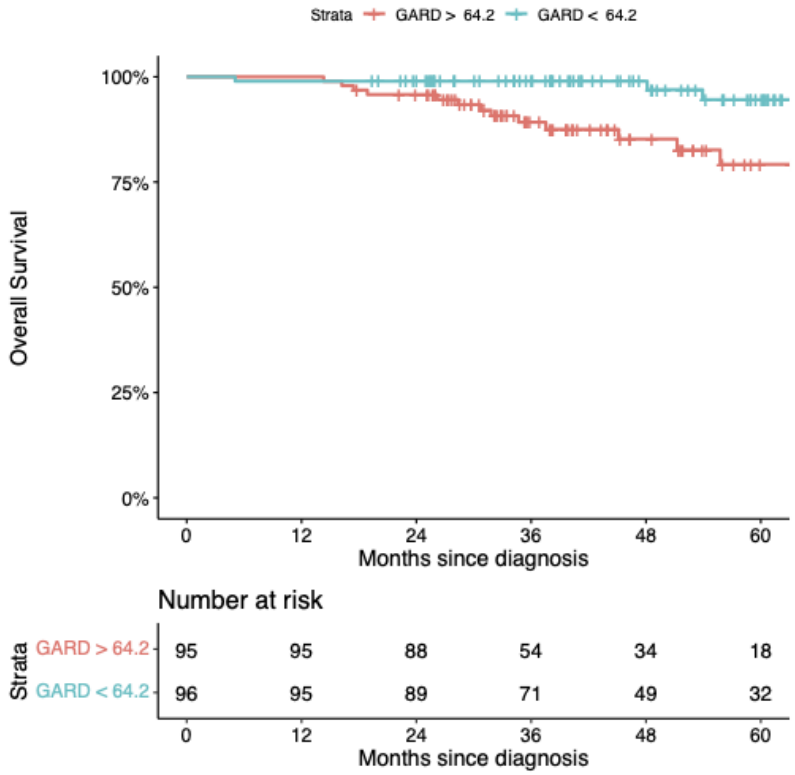


Figure 2. Median GARD cutpoint demonstrate survival differences between groups. Stratifying patients by median GARD shows a significant difference in OS at 5 years by the log-rank test.

GARD outperforms the prognostic ability of the NRG clinical nomogram. In a previous study, NRG developed a clinical nomogram to predict overall survival in oropharyngeal cancer patients.[18] The overall survival nomogram is based on age (<50 vs. >50), smoking ($\leq$ 10-pack-years vs. $>$ 10 pack-years), age x pack-years interaction, performance status (0 vs. 1), education (high school or less vs. others), anemia (defined as hgb$\leq$ 13.5 in men and $\leq$ 12.5 in women) (yes vs. no), p16, T stage (T2-3 vs. T4), N stage (8th edition, N0-1 vs. N2-3). When tested in an independent validation cohort, the nomogram achieved a c-index of 0.76. We evaluated the performance of this nomogram (without the education and anemia parameters which were not available) in the

BD2DECIDE cohort and tested whether the integration of GARD impacted its performance. As shown in **Figure 3**, the NRG nomogram achieves an AUC of 73.6 whereas GARD alone achieves AUC (3 yr OS) 80.6. Integrating GARD into the nomogram improves the prognostic ability of the model with AUC 82.6. Integrating the previously developed 3-cluster model also improves the prognostic ability of the existing NRG nomogram to AUC 84.7.

GARD identifies a sub-population of HPV+ patients with poor prognosis. Although HPV+ patients have excellent prognosis, the interim analysis of HN005 have emphasized the importance of developing clinical tools to identify patient subsets with higher risk of clinical failure. We hypothesized that GARD could identify a sub-population of HPV+ patients at higher risk of failure who may not be good candidates for unselected treatment de-escalation. Thus, we performed an exploratory discrete analysis based on an optimized two cut-point analysis. Minimizing the χ^2 statistic at two discrete values reveals three groups with maximally different outcomes, as shown in **Figure 4A**. This analysis revealed two cutpoints at GARD 65 and 46 which optimally stratified patients.

Patients that achieved the highest GARD (GARD $>$ 65) had a 3yr-OS of 100% compared with 91.3% (85.3, 97.7) for the GARD intermediate group (46$<$ GARD $<$ 65) and 62.5% (33.6, 100) for the group that achieves the lowest GARD (GARD $\leq$ 46). These differences are statistically significant with p $<$ 0.001, though this analysis should be interpreted carefully as it was performed for hypothesis generation and the groups were chosen by maximizing differences *post hoc*.

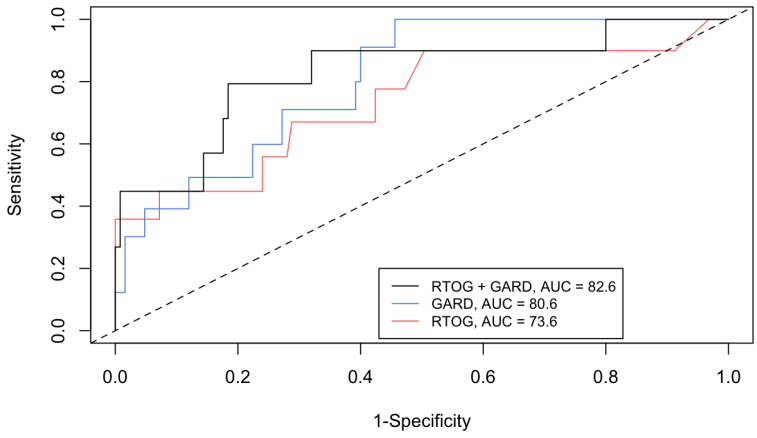


Figure 3. AUC analysis shows GARD outperforms standard nomograms. In this AUC analysis we compare the clinical nomogram (red) to GARD alone (blue) and a combined model using all factors (black). The combined model shows dramatic improvement compared to RTOG alone. This model included all 191 definitive primary RT patients, and analysed outcome at 3 years.

GARD predicts that empiric dose de-escalation would result in inferior clinical outcome. Recently, NRG announced that an interim analyis revealed that the dose de-escalation arm in HN-005 (chemo-RT to 60 Gy) had failed to achieve the non-inferiority threshold defined by the statistical analysis plan. One possible explanation for these results is that empiric dose de-escalation results in a small number of patients falling from the GARD intermediate cohort (46 $<$ GARD $<$ 65) to the GARD low cohort (GARD $\leq$ 46) leading to an inferior result for empiric dose de-escalation. To test this hypothesis, we performed an *in silico* clinical trial

to evaluate GARD-based predictions of clinical outcome for empiric dose de-escalation to 60 Gy (with concurrent chemotherapy) as in HN-005. We found that GARD predicts that empiric (unselected) dose de-escalation would result in an inferior clinical outcome. The predicted 3 yr OS for patients modeled at 70 Gy is 94.2% compared with 90.2% for patients modeled at 60 Gy **Figure 4C**. Empiric, unselected dose de-intensification is predicted to increase the proportion of patients in the GARD low group while decreasing the proportion of patients in the GARD high group. The 70Gy *in silico* arm had an average of 13, 90, and 97 patients in the low, intermediate, and high GARD groups, while the 60Gy *in silico* arm had 36, 123, and 41 patients in those groups. Next, we determined whether we could use GARD to develop a clinical trial strategy that would predict equivalent outcome at 70 or 60 Gy. In one approach, GARD can identify patients that would remain above the GARD-high cutpoint (65) at 70 or 60 Gy. This approach would select patients with RSI < 0.115, which comprise 18% of the total HPV+ population. In another approach, GARD-high and intermediate patients at 70 Gy that remain in the same group at 60 Gy, would also be predicted to achieve equipoise with selective dose de-escalation. Approximately 55% of HPV+ patients would be eligible for this approach. It should be noted that both approaches exclude GARD-low patients (as these patients are predicted to require *dose intensification*) and patients that fall from GARD-high to GARD-intermediate or GARD-intermediate to GARD-low at 60Gy. The predicted OS curve for the second approach to de-escalation is shown in **Figure 4D**.

Discussion

The development of prognostic models to more accurately classify oncologic outcomes is a central goal of personalized medicine. In this paper, we show that GARD, a previously described model of the treatment effect of RT is associated with overall survival in HPV+ oropharyngeal cancer patients treated with RT both as a continuous and dichotomous variable. Furthermore, using time-dependent ROC analysis, we show that GARD outperforms the current NRG clinical nomogram for prediction of overall survival of these patients. Finally, we show that GARD predicts that uniform RT dose de-escalation would result in an inferior overall survival over standard RT dose, similar to the results recently announced for HN005. However, GARD creates opportunities for future trial strategies to selective RT dose de-escalation that it predicts would achieve clinical equipoise.

Since its confirmation as a biomarker of outcome in HNC, HPV status has been incorporated into the diagnostic algorithm of the disease.[1,2] In a landmark paper, Ang et al. demonstrated that HPV+ HNC patients had 58% reduction in the risk of death when compared with HPV- patients. This effect translated into a 25-point absolute difference in 3 year OS distinguished by HPV status (82.4% vs. 57.1%).[3] In this analysis, we demonstrate that

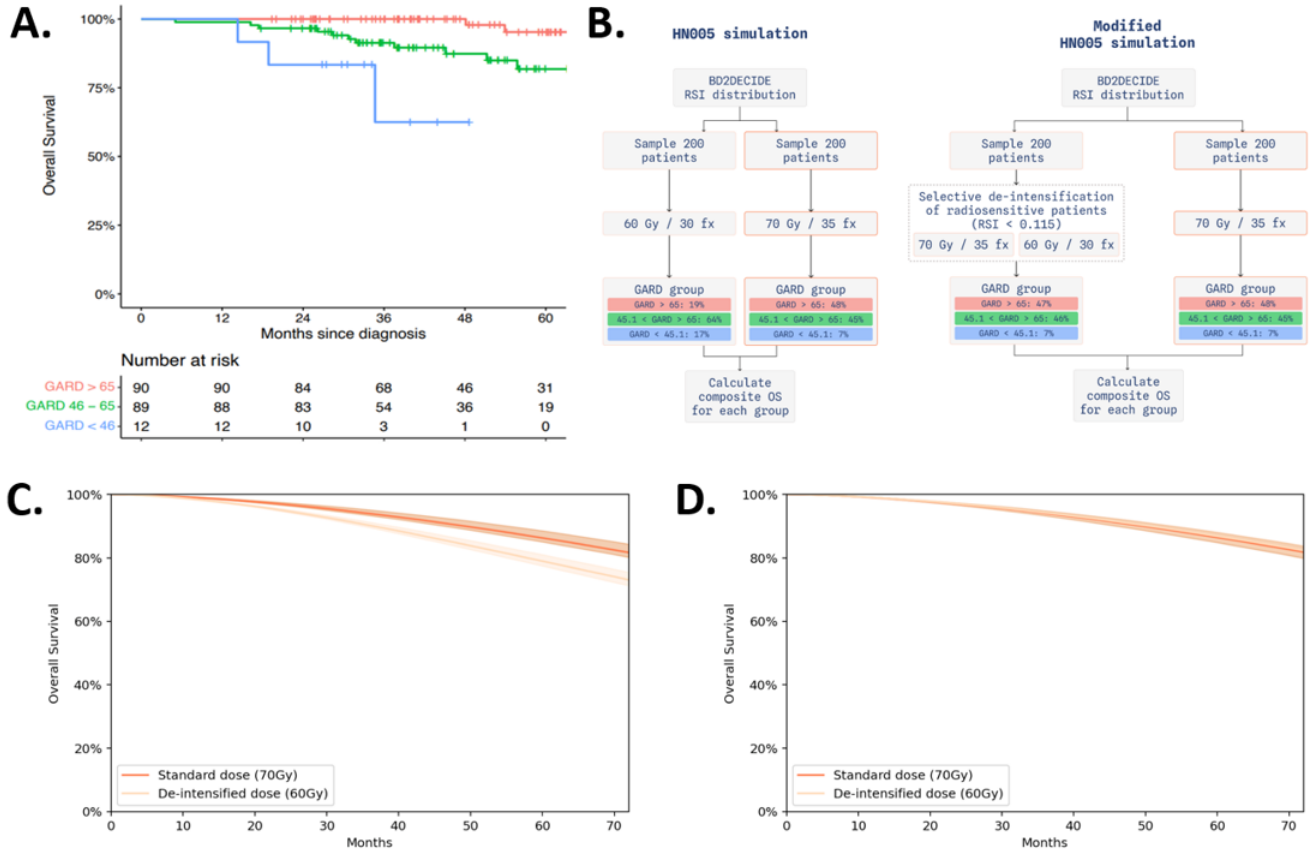


Figure 4. GARD predicts that uniform RT dose de-escalation in HPV+ patients would result in an inferior clinical outcome compared to standard of care. A. GARD identifies a sub-group of HPV+ patients with poor prognosis. An exploratory analysis identified two cut-points which group patients in three risk levels. Patients that achieve the lowest GARD (<46) have a high risk of recurrence (3 year OS = 62.5%). **B.** In-silico clinical trial. We simulate two of the arms of HN005 (cisplatin + 60 Gy vs cisplatin + 70 Gy). We utilize the RSI distribution of the B2Decide cohort to generate GARD for 400 virtual patients randomized to either 60 or 70 Gy. We also simulate a GARD-selected clinical trial where only patients with RSI < 0.115 are eligible for randomization. **C.** Unselected RT dose de-escalation (cisplatin + 60 Gy) results in inferior OS compared to standard of care (cisplatin + 70 Gy). The unselected *in silico* clinical trial predicts that patients treated with RT dose descalation experience a worse overall survival when compared to standard of care (3-yr OS 90.3% vs 94.0%) **D:** Selective de-intensification produces similar OS even when approximately two-thirds of patients in the second arm receive the lower RT dose.

GARD identifies a 72% reduction in the risk of death in GARD-high HPV+ patients. This translates into an absolute 10% and 16% difference in 3 and 5 year OS between GARD-high and GARD-low patients. In addition GARD alone outperformed the established NRG nomogram for overall survival. Finally, a model integrating GARD and the NRG nomogram achieved the highest prognostic ability. Thus, GARD can resolve prognosis for HPV+ patients with the same magnitude that HPV did for head and neck cancer patients.

A central clinical question for HPV+ oropharyngeal cancer patients is whether their treatment can be de-intensified while preserving their excellent prognosis. While there has been significant enthusiasm for this approach, NRG recently announced that in the early interim analysis one of the dose de-escalated arms in HN-005 (cisplatin + 60 Gy) did not meet the pre-defined statistical threshold for non-inferiority. Thus, clinical factors alone may not enough to identify patients where radiation dose de-intensification can be performed without clinical outcome

detriment.

While we demonstrate that GARD outperforms standard clinical variables as a prognostic factor in HPV+ oropharyngeal cancer, GARD can further improve the ability to define appropriate sub-populations for treatment de-intensification. For example in an exploratory analysis, we show that GARD identifies a small group of HPV+ patients with poor prognosis (GARD < 46) who achieve a 3-yr OS of 62.5%. In addition, GARD is also an actionable model that can provide guidance on RT dosing for genomically-defined subpopulations. This can inform the design of the next generation of clinical trials for these patients. We demonstrate proof of principle for GARD-based clinical trial design by showing that GARD predicts that empiric dose de-escalation as performed in HN-005 would result in an inferior 3 year OS for the patients treated to the lower dose. Finally, we propose at least two designs (of many possible) that GARD-based modeling predicts would result in clinical equipoise between 70 and 60 Gy with appropriately chosen patients for de-escalation. In the first design, only patients in the top 18% of the GARD distribution would be eligible, while in the second design approximately 55% of HPV+ patients would be eligible. A key observation is that a small subset of patients may need dose intensification and should not be eligible for these trials. Further work on chemotherapy sensitivity[19] and balancing toxicity with tumor control[20] could also inform future iterations of these trials, most of which include cisplatin as a radiosensitizer.[21]

In conclusion, we demonstrate that GARD outperforms the NRG clinical nomogram of outcome as a prognostic biomarker in HPV+ OPSCC patients and defines prognostic groups that can inform clinical trial design. While HPV is a classic biomarker in that its result is fixed and cannot be changed, the GARD value for a patient can be optimized by adjusting the RT dose. This supports the hypothesis that GARD could be used to optimize clinical outcome for HPV+ oropharyngeal SCC patients by the personalization of RT dose. Even without the use for dose personalization, however, the strong improvement (quantitatively equivalent to the seminal findings of HPV positivity itself) in outcome prognostication suggests that obtaining GARD should be considered for all HPV+ oropharyngeal cancer patients as a clinical decision support tool.

References

1. De Felice, F. *et al.* Advances in the management of hpv-related oropharyngeal cancer. *Journal of oncology* **2019** (2019).

2. Machczyński, P., Majchrzak, E., Niewinski, P., Marchlewska, J. & Golusiński, W. A review of the 8th edition of the ajcc staging system for oropharyngeal cancer according to hpv status. *European Archives of Oto-Rhino-Laryngology* **277**, 2407–2412 (2020).

3. Ang, K. K. *et al.* Human papillomavirus and survival of patients with oropharyngeal cancer. *New England Journal of Medicine* **363**, 24–35 (2010). DOI 10.1056/nejmoa0912217.

4. Yom, S. S. *et al.* Reduced-dose radiation therapy for hpv-associated oropharyngeal carcinoma (nrg oncology hn002). *Journal of Clinical Oncology* **39**, 956–965 (2021). DOI 10.1200/jco.20.03128.

5. Scott, J. G. *et al.* A genome-based model for adjusting radiotherapy dose (gard): a retrospective, cohort-based study. *The Lancet Oncology* **18**, 202–211 (2017).

6. Scott, J. G. *et al.* Pan-cancer prediction of radiotherapy benefit using genomic-adjusted radiation dose (gard): a cohort-based pooled analysis. *The Lancet Oncology* **22**, 1221–1229 (2021).

7. Cavalieri, S. *et al.* Development of a multiomics database for personalized prognostic forecasting in head and neck cancer: The big data to decide eu project. *Head & Neck* **43**, 601–612 (2020). DOI 10.1002/hed.26515.

8. Eschrich, S. *et al.* Systems biology modeling of the radiation sensitivity network: a biomarker discovery platform. *International Journal of Radiation Oncology* Biology* Physics* **75**, 497–505 (2009).

9. Cavalieri, S. *et al.* Clinical validity of a prognostic gene expression cluster-based model in human papillomavirus–positive oropharyngeal carcinoma. *JCO Precision Oncology* 1666–1676 (2021). DOI 10.1200/po.21.00094.

10. Eschrich, S. A. *et al.* A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. *International Journal of Radiation Oncology* Biology* Physics* **75**, 489–496 (2009).

11. Strom, T. *et al.* Radiosensitivity index predicts for survival with adjuvant radiation in resectable pancreatic cancer. *Radiotherapy and Oncology* **117**, 159–164 (2015).

12. Ahmed, K. A. *et al.* The radiosensitivity index predicts for overall survival in glioblastoma. *Oncotarget* **6**, 34414–34422 (2015).

13. Ahmed, K. A. *et al.* Differences between colon cancer primaries and metastases using a molecular assay for tumor radiation sensitivity suggest implications for potential oligometastatic sbrt patient selection. *Int J Radiat Oncol Biol Phys* (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/25838188. DOI 10.1016/j.ijrobp.2015.01.036.

14. Harrell Jr, F. E., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* **15**, 361–387 (1996).

15. Blanche, P., Dartigues, J.-F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine* **32**, 5381–5397 (2013). URL http://onlinelibrary.wiley.com/doi/10.1002/sim.5958/full.

16. Ahmed, K. A. *et al.* Radiosensitivity of lung metastases by primary histology and implications for stereotactic body radiation therapy using the genomically adjusted radiation dose. *J Thorac Oncol* **13**, 1121–1127 (2018). URL https://www.ncbi.nlm.nih.gov/pubmed/29733909. DOI 10.1016/j.jtho.2018.04.027.

17. Ahmed, K. A. *et al.* Utilizing the genomically adjusted radiation dose (gard) to personalize adjuvant radiotherapy in triple negative breast cancer management. *EBioMedicine* **47**, 163–169 (2019).

18. Fakhry, C. *et al.* Development and validation of nomograms predictive of overall and progression-free survival in patients with oropharyngeal cancer. *Journal of Clinical Oncology* **35**, 4057 (2017).

19. Scarborough, J. A., Eschrich, S. A., Torres-Roca, J., Dhawan, A. & Scott, J. G. Exploiting convergent phenotypes to derive a pan-cancer cisplatin response gene expression signature. *npj Precision Oncology* **7**, 38 (2023).

20. Scott, J. G. *et al.* Personalizing radiotherapy prescription dose using genomic markers of radiosensitivity and normal tissue toxicity in non-small cell lung cancer. *Journal of Thoracic Oncology* (2020).

21. Song, B. *et al.* Ct radiomic signature predicts survival and chemotherapy benefit in stage i and ii hpv-associated oropharyngeal carcinoma. *NPJ Precision Oncology* **7**, 53 (2023).

Supplemental Information

Code and data availability

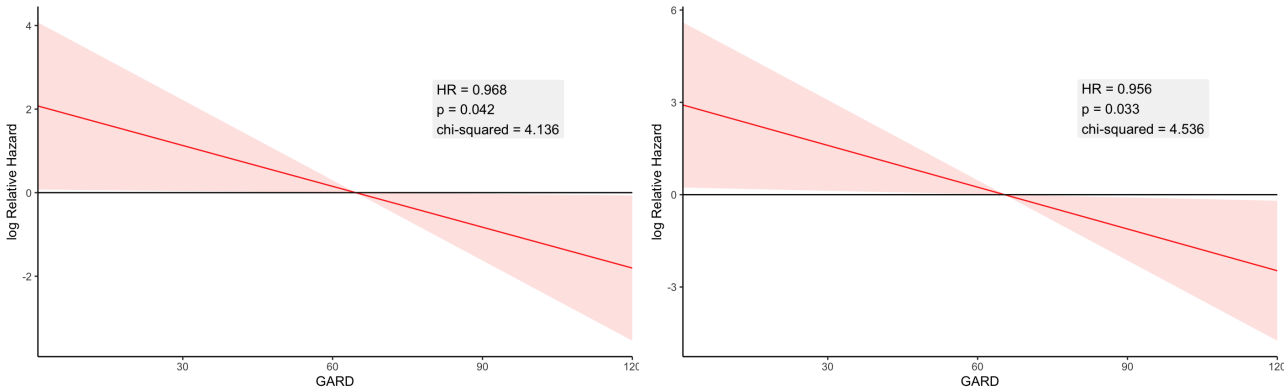
All statistical analyses were conducted using Python v3.9.0 and the `rms` package in R v4.2.0. Code and data for all analyses are available by request.

Funding statement

JGS would like to thank the NCI for their support through the Cleveland Clinic/Emory ROBIN center, U54-CA274513, Project 2. In addition, this work was supported in part by European Union Horizon 2020 Framework Programme, Grant/Award Number: 689715 (LL), AIRC (ID 23573 projects -LDC) and ERA-NET ERA PerMed JTC2019/FRRB project SuPerTreat (Supporting Personalized Treatment Decisions in Head and Neck Cancer through Big Data) (LDC).

Cox Analyses

Performing a Cox regression analysis for GARD as a continuous variable revealed statistically significant associations with OS for the entire cohort (**Supplemental Figure 1, Left**), which remained significant when completed for patients treated with RT alone (**Supplemental Figure 1, Right**).



(a) Entire BD2D cohort

(b) Patients treated with primary RT

Supplemental Figure 1. GARD is a continuous predictor of OS in radiation treated patients with HPV+ OPSCC. (a) Cox proportional hazards analysis demonstrates significant continuous association between GARD and OS for the entire cohort (p = 0.042, HR = 0.968 (0.938, 0.999) *per unit* GARD). **(b)** Cox proportional hazards analysis demonstrates significant continuous association between GARD and OS for the subset of patients treated with primary definitive RT (p = 0.033, HR = 0.956 (0.917, 0.996) *per unit* GARD).

AUCs for various predictors

Comparison of the 3 most significant predictors individually via ROC analysis showed the greatest AUC for GARD alone (80.6), while the nomogram combining GARD, AJCC8, and the molecular clusters produced the highest AUC (82.6); values are listed in **Supplemental Table 1** and displayed. Of note, if RSI is compared here with the same cohort, an almost identical score to GARD is achieved of 80.6 (95% CI: 70.7 to 90.5) as the dose range in this cohort is so narrow making RSI and GARD functionally equivalent.

	3-yr AUC (95% CI)
Clusters	72.8 (59.0, 86.7)
RTOG	73.6 (55.4, 91.7)
GARD	80.6 (69.4, 91.9)
RTOG + GARD	82.6 (67.8, 97.4)
RTOG + GARD + Clusters	84.7 (70.5, 98.9)

Supplemental Table 1. The time-dependent AUC and 95% CI at 3 years are shown for the predictors individually as well as for the combined nomogram.