International Journal of Radiation Oncology biology • physics

www.redjournal.org

Clinical Investigation

Metabolic Tumor Volume as a Prognostic Imaging-Based Biomarker for Head-and-Neck Cancer: Pilot Results From Radiation Therapy Oncology Group Protocol 0522

David L. Schwartz, MD,[*] Jonathan Harris, MS,[†] Min Yao, MD, PhD,[‡] David I. Rosenthal, MD, FACR,[§] Adam Opanowski, CNMT, PET, NCT, RT (N),[||] Anthony Levering, RT (R) (CT) (MR), CIIP,[||] K. Kian Ang, MD, PhD,[§] Andy M. Trotti, MD,[¶] Adam S. Garden, MD, FASTRO,[§] Christopher U. Jones, MD, FACR,[#] Paul Harari, MD,[**] Robert Foote, MD, FASTRO,[††] John Holland, MD,[‡‡] Qiang Zhang, MD, PhD,[†] and Quynh-Thu Le, MD, FASTRO[§§]

Department of Radiation Oncology, University of Texas Southwestern School of Medicine, Dallas, Texas; [†]Radiation Therapy Oncology Group Statistical Center, Philadelphia, Pennsylvania; [‡]Department of Radiation Oncology, Case Western Reserve University School of Medicine, Cleveland, Ohio; [§]Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas; [||]American College of Radiology Imaging Network, Philadelphia, Pennsylvania; [¶]Department of Radiation Oncology, Moffitt Cancer Center, Tampa, Florida; [#]Sutter Medical Group, Sacramento, California; [**]Department of Human Oncology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; [††]Department of Radiation Oncology, Mayo Clinic, Rochester, Minnesota; [‡‡]Department of Radiation Medicine, Oregon Health & Science University, Portland, Oregon; and [§§]Department of Radiation Oncology, Stanford University School of Medicine, Palo Alto, California

Received Sep 8, 2014, and in revised form Dec 3, 2014. Accepted for publication Dec 9, 2014.

Summary

There is need to better match cancer treatment to

Purpose: To evaluate candidate fluorodeoxyglucose positron emission tomography/computed tomography (FDG-PET/CT) imaging biomarkers for head-and-neck chemoradiotherapy outcomes in the cooperative group trial setting.

Reprint requests to: David L. Schwartz, MD, Department of Radiation Oncology, U.T. Southwestern School of Medicine, Harold C. Simmons Comprehensive Cancer Center, 5801 Forest Park Rd, Dallas, TX 75390. Tel: (214) 645-8525; E-mail: david.schwartz@utsw.edu

Supported by Radiation Therapy Oncology Group grant U10 CA21661 and Community Clinical Oncology Program grant U10 CA37422 from the National Cancer Institute and Bristol-Myers Squibb. This manuscript's contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

Conflict of interest: R.F. reports grant support from Radiation Therapy Oncology Group, personal fees from the American Society for Radiation Oncology for editorial services, and personal fees from UpToDate, all outside the scope of the submitted work. D.I.R. reports personal fees from Bristol Myers during the conduct of the study.

Supplementary material for this article can be found at www.redjournal.org.

Int J Radiation Oncol Biol Phys, Vol. 91, No. 4, pp. 721–729, 2015
0360-3016/$ - see front matter
http://dx.doi.org/10.1016/j.ijrobp.2014.12.023

individual patient risk. Tumor imaging provides a means to define disease phenotype and treatment response. Predefined study objectives of the recently concluded RTOG 0522 trial included correlation of PET/CT findings with treatment outcomes. In this report we describe a potential prognostic signal for primary tumor FDG-PET–derived MTV in patients with locally advanced head-and-neck cancer. This finding remains preliminary and requires further confirmation.

Methods and Materials: Radiation Therapy Oncology Group (RTOG) protocol 0522 patients consenting to a secondary FDG-PET/CT substudy were serially imaged at baseline and 8 weeks after radiation. Maximum standardized uptake value (SUVmax), SUV peak (mean SUV within a 1-cm sphere centered on SUVmax), and metabolic tumor volume (MTV) using 40% of SUVmax as threshold were obtained from primary tumor and involved nodes.

Results: Of 940 patients entered onto RTOG 0522, 74 were analyzable for this substudy. Neither high baseline SUVmax nor SUVpeak from primary or nodal disease were associated with poor treatment outcomes. However, primary tumor MTV above the cohort median was associated with worse local-regional control (hazard ratio 4.01, 95% confidence interval 1.28-12.52, $P = .02$) and progression-free survival (hazard ratio 2.34, 95% confidence interval 1.02-5.37, $P = .05$). Although MTV and T stage seemed to correlate (mean MTV 6.4, 13.2, and 26.8 for T2, T3, and T4 tumors, respectively), MTV remained a strong independent prognostic factor for progression-free survival in bivariate analysis that included T stage. Primary MTV remained prognostic in p16-associated oropharyngeal cancer cases, although sample size was limited.

Conclusion: High baseline primary tumor MTV was associated with worse treatment outcomes in this limited patient subset of RTOG 0522. Additional confirmatory work will be required to validate primary tumor MTV as a prognostic imaging biomarker for patient stratification in future trials.

Introduction

Effective patient selection drives successful clinical cancer trial design. Tissue-based biomarkers have been used toward this end, with a promise to increase study power and reduce treatment development costs. However, tumor tissue collection is expensive and burdensome. Tumor imaging provides an alternative means to noninvasively define disease phenotype and treatment response.

Functional imaging delivers quantitative characterization of tumor and host tissue physiology. Fluorodeoxyglucose positron emission tomography (FDG-PET), with or without coregistered computed tomography (CT), serves as the traditional "work horse" for functional head-and-neck imaging. Published experience with FDG-PET–defined staging is mature and has been summarized by meta-analyses (1), expert consensus reports (2, 3), and comparative effectiveness studies (4). Imaging by FDG-PET incrementally improves staging accuracy and treatment response assessment over anatomic imaging (5-8), particularly if guided by complementary clinical features (9, 10). Some institutional series suggest that certain FDG-PET parameters, such as the maximum or peak standardized uptake values (SUVs), may serve as quantifiable imaging biomarkers for radiation therapy outcomes (11-13). However, conflicting reports refute the predictive value of SUV (14), and quantitative head-and-neck FDG-PET outcome measures remain untested in the cooperative group trial setting.

Radiation Therapy Oncology Group (RTOG) protocol 0522 recently completed enrollment of 940 patients diagnosed with locally advanced head-and-neck cancer. Study subjects received concurrent radiation therapy and cisplatin, with or without the addition of cetuximab. All RTOG 0522 subjects with N2-3 disease (with the exception of N2c with both sides $\leq$ N1) were eligible for baseline and posttreatment coregistered PET/CT imaging analysis. Predefined study objectives included correlation of pre- and posttreatment PET/CT scan findings with histologic findings of neck dissection specimens and treatment outcomes. Given interval publication of encouraging institutional pilot findings for use of FDG-PET–defined metabolic tumor volume (MTV) as an imaging biomarker for radiation therapy treatment response and tumor control outcomes (15-17), MTV was included as a secondary post hoc study objective in the present report.

Methods and Materials

Study population

Patients enrolled to RTOG 0522 with N2a, N2b, N2c (with right and/or left side N2a-N2b), or N3 disease, who agreed to participate in the PET/CT study, and for whom at least 1 PET image set was available for central review, were included in this analysis.

PET/CT image and scanner compatibility requirements

All centers participating in this imaging study had to provide 1 test case to the American College of Radiology Imaging Network (ACRIN) PET Core Laboratory before start of enrollment to credential their file transfer capabilities and image quality. The PET Core Laboratory provided software for imaging facilities to collect, deidentify, and submit image sets either from a PET/CT scanner or a picture archiving and communication system (PACS) to the ACRIN image archive. All imaging had to be performed on a combined PET/CT instrument with full ring PET and 4-

slice or greater multidetector CT operating in high-sensitivity 2-dimensional mode, if available. To simplify multi-institutional participation, centers were not required to use uniform PET/CT software. Three-dimensional mode was permissible for patients imaged on combined PET/CT scanners without a 2-dimensional mode.

Patient preparation and FDG injection

Participating centers were instructed to record patient height and weight before each PET scan, to have patients observe a 4- to 6-hour fasting period before FDG injection, and to measure serum glucose concentration before scanning. A serum glucose value $<$200 mg/dL was necessary to proceed to imaging. Centers were instructed to inject a dose of 10 to 20 mCi of FDG intravenously and to begin imaging within 50 to 70 minutes after tracer administration.

Baseline and posttreatment PET/CT imaging

Imaging was required to encompass the vertex of the head down through the entire pelvis. The recommended imaging protocol incorporated 2 discrete phases: during the first phase, head-and-neck scanning was performed with arms resting at sides and full neck extension using a 120 keV/300 mA, 0.5-second detector rotation time ("high mA") CT scan with intravenous contrast (100 mL contrast bolus administered at 1.5 mL/s, with a 50-second scan delay and with the scan started inferiorly, moving cranially), followed by a 120 keV/80 mA, 0.8-second detector rotation time ("low mA") CT scan for PET attenuation correction, followed last by PET scanning. Standard manufacturer recommendations for specific low and high mA CT scanning parameters could be substituted. Two fields of view (approximately 15 cm) were required for full head-and-neck PET imaging from manubrium to vertex. Centers were instructed to allow patients to rest their neck for 1 to 2 minutes. For the second phase of imaging, the neck was shifted into neutral position, and the remainder of the body was surveyed per routine local institutional protocol with arms raised above the head to allow for thoracic and abdominal imaging. At least 4 to 5 PET fields of view were to be used for this phase. The PET data were corrected for dead time, scatter, random coincidence events, and attenuation, and ultimately reconstructed into images via the filtering algorithm provided by the scanner manufacturer. A posttreatment FDG-PET/CT scan was recommended 8 to 9 weeks after completion of treatment (in addition to mandatory CT or magnetic resonance posttreatment imaging). Centers were requested to perform posttreatment PET/CT imaging in this time frame and on the same scanner, if possible.

Maximum SUV (SUVmax), Peak SUV (SUVpeak), and MTV

The SUV normalized by specific injected dose and patient weight was calculated on centralized review by 2 head-and-neck radiation oncologists with specific expertise in PET/CT interpretation (D.L.S. and M.Y.) using commercial image analysis software (MIM Software, v.5.2, Cleveland, OH). Detection of primary and nodal disease by FDG-PET/CT was determined qualitatively as FDG uptake greater than surrounding normal soft tissue within a CT-delineated anatomic (primary disease or nodal) abnormality. The SUVmax was defined as (tissue activity) (μCi/mL)/(injected dose) (mCi)/(patient weight) (kg) within the voxel having the highest activity within a given region of interest (ROI). These values were determined separately for ROIs within primary tumor and involved cervical lymph nodes. Each ROI had to encompass the entire FDG-avid lesion of interest, with boundaries guided by CT delineation. The SUVpeak for primary and nodal disease was measured from a 1-cm^3 sphere centered on the voxel with the highest mean SUV value. The location of this sphere was manually checked to ensure reasonable position within disease. Primary tumor MTV contours were generated from voxels that were $\geq$40% of primary SUVmax. A specific tumor can have a much higher SUV than another tumor, but MTV calculations and volume generation will remain consistent across tumors because the analysis tool performs identical calculations of proportionality regardless of baseline SUVmax value. Nodal MTV was measured separately for each side of the neck, inclusive of all metabolically active nodes (including cystic regions within these nodes), and defined as the nodal tumor volume above 40% of the nodal SUVmax. For bilateral neck disease, the maximum of right and left sides was used for SUV and the sum for MTV. Total MTV, defined as primary MTV plus nodal MTV, was calculated for patients with both measures.

Data analysis and statistics

The following pre- and posttreatment PET values were obtained from central review: SUVmax, SUVpeak, and MTV of the primary tumor, right nodal disease, and left nodal disease. The protocol analysis plan prespecified the analysis of SUV measures. Evaluation of MTV represents post hoc secondary analysis and must be formally considered exploratory in nature. The study cohort was dichotomized into equal-size comparator groups by median PET values to maximize statistical power for analysis. Patients with at least one readable value from either reader were included. Consistency between readers was measured by intraclass correlation coefficient.

Clinical endpoints specified by the trial protocol were local relapse (including salvage surgery for primary site with tumor present/unknown), regional relapse (including neck dissection $>$15 weeks after the end of radiation therapy with tumor present/unknown), local−regional relapse, distant metastasis, progression-free survival, and overall survival. Only death was considered a competing risk for local relapse, regional relapse, local−regional relapse, or distant metastasis. Progression-free survival

Table 1 Patient characteristics

Characteristic	Included in PET/CT study (n=74)	Eligible for PET/CT study but excluded (n=577)
Assigned treatment, $P=.26$*		
RT + cisplatin	42 (56.8)	287 (49.7)
RT + cisplatin + cetuximab	32 (43.2)	290 (50.3)
Age (y), $P=1.00$†		
Mean (SD)	56.8 (6.67)	56.7 (8.22)
Median (range)	56 (42-73)	57 (34-79)
Gender, $P=.68$*		
Male	65 (87.8)	516 (89.4)
Female	9 (12.2)	61 (10.6)
Zubrod performance status, $P=.03$*		
0	58 (78.4)	380 (65.9)
1	16 (21.6)	197 (34.1)
Smoking history: pack-years, $P=.03$†	(n=54)	(n=512)
Mean (SD)	20.8 (29.88)	26.1 (26.96)
Median (range)	8.75 (0-135)	21 (0-162)
Primary site, $P=.72$*		
Oropharynx	58 (78.4)	449 (77.8)
Hypopharynx	7 (9.5)	43 (7.5)
Larynx	9 (12.2)	85 (14.7)
p16 status, oropharynx only, $P=.98$*	(n=33)	(n=229)
Negative	8 (24.2)	55 (24.0)
Positive	25 (75.8)	174 (76.0)
T stage, $P=.04$†		
T2	29 (39.2)	303 (52.5)
T3	26 (35.1)	157 (27.2)
T4	19 (25.7)	117 (20.3)
N stage, $P=.23$†		
N2a	5 (6.8)	73 (12.7)
N2b	34 (45.9)	259 (44.9)
N2c	29 (39.2)	207 (35.9)
N3	6 (8.1)	38 (6.6)

Abbreviations: CT = computed tomography; PET = positron emission tomography; RT = radiation therapy.
Values are number (percentage) unless otherwise noted.
* Pearson χ^2 test.
† Wilcoxon rank-sum test.

events were defined as local or regional relapse, distant metastasis, or death due to any cause. All endpoints were measured from the date of randomization. Rates for local, regional, and local–regional relapse, and distant metastasis were estimated by the cumulative incidence method (18), and rates for progression-free and overall survival were estimated by the Kaplan-Meier method (19). Hazard ratios were estimated by Cox proportional hazards models (20). To address whether negative (SUVmax <3, per Yao et al [21]) posttreatment PET/CT in patients who achieve a clinical complete nodal response (CR, defined as no evidence of nodal disease on physical examination and/or imaging) predicts for a low nodal relapse rate (<10%) at 2 years, the 2-year regional relapse rates in 4 groups were calculated: (*1*) clinical CR, negative PET; (*2*) clinical < CR, negative PET; (*3*) clinical CR, positive PET; and (*4*) clinical < CR, positive PET.

Exploratory post hoc subgroup analysis was performed to selectively evaluate PET measures in patients with p16-positive oropharyngeal cancer. Missing p16 values were imputed 20 times using conditional model specification for multivariate imputation via Gibbs sampling (22), and the resulting datasets were combined using Rubin's formula (23), and sensitivity analyses were conducted to validate the robustness of the imputation procedure. Akaike information criterion (24) was used to measure the relative quality of our statistical models.

Results

Imaging study cohort

One hundred sixteen patients agreed to participate in the PET/CT study. Forty-two subjects did not follow through with posttreatment imaging or had unreadable image files, yielding 74 patients from 19 centers with both pre- and posttreatment PET/CT imaging available for this

Table 2 FDG-PET/CT outcome measures

Measurement	Value	Intraclass correlation coefficient (95% CI)
Pretreatment primary SUVmax*	(n=68)	1.00 (N/A)
Mean (SD)	15.84 (6.68)	
Median (range)	15.07 (3.45-40.02)	
Posttreatment primary SUVmax*,†	(n=67)	0.93 (0.89-0.96)
Mean (SD)	4.65 (2.06)	
Median (range)	4.31 (1.75-14.48)	
Pretreatment primary SUVpeak*	(n=68)	0.99 (0.98-0.99)
Mean (SD)	12.68 (6.19)	
Median (range)	11.23 (2.54-37.36)	
Posttreatment primary SUVpeak*,†	(n=67)	0.83 (0.73-0.89)
Mean (SD)	3.23 (1.53)	
Median (range)	3.18 (1.28-11.78)	
Pretreatment primary MTV*	(n=68)	0.99 (0.98-0.99)
Mean (SD)	13.80 (12.55)	
Median (range)	8.76 (2.22-64.45)	
Pretreatment nodal SUVmax*,‡	(n=65)	R: 0.99 (0.98-0.99)
Mean (SD)	11.62 (6.81)	L: 1.00 (N/A)
Median (range)	10.55 (2.08-42.96)	
Posttreatment nodal SUVmax*,†,‡	(n=62)	R: 0.82 (0.69-0.90)
Mean (SD)	2.93 (1.20)	L: 0.87 (0.77-0.93)
Median (range)	2.66 (1.03-6.91)	
Pretreatment nodal SUVpeak*,‡	(n=65)	R: 0.99 (0.98-0.99)
Mean (SD)	8.47 (5.00)	L: 1.00 (N/A)
Median (range)	7.73 (1.34-27.93)	
Posttreatment nodal SUVpeak*,†,‡	(n=62)	R: 0.59 (0.34-0.76)
Mean (SD)	1.97 (0.56)	L: 0.51 (0.23-0.71)
Median (range)	1.95 (0.73-3.90)	
Pretreatment nodal MTV*,§	(n=61)	R: 0.99 (0.98-0.99)
Mean (SD)	14.38 (18.25)	L: 0.95 (0.91-0.97)
Median (range)	10.10 (1.46-141.01)	
Pretreatment total MTV*,‖	(n=60)	
Mean (SD)	27.56 (21.89)	
Median (range)	21.77 (3.85-149.64)	

Abbreviations: CI = confidence interval; FDG = fluorodeoxyglucose; L = left neck; MTV = metabolic tumor volume; N/A = not applicable; R=right neck; SUV = standardized uptake value. Other abbreviations as in Table 1.

* If both reader values available, mean was used.

† Patients that progressed before posttreatment PET were excluded.

‡ If bilateral, maximum of right and left sides was used.

§ If bilateral, sum of right and left sides was used.

‖ Sum of primary and nodal MTV.

analysis. In addition, 535 patients were eligible for the PET/CT study according to N stage but chose not to participate, for a total of 577 patients not included in this analysis. Analyzed patients enjoyed better performance status and shorter pack-year tobacco exposure but presented at a more advanced T stage than excluded patients (Table 1). At the time of analysis, median follow-up for surviving patients (58 of 74) was 4.2 years (range, 3.1-6.2 years).

PET summary and reviewer agreement

Summary FDG-PET/CT outcome measures are listed in Table 2. Reader agreement was excellent (intraclass correlation coefficient $\geq$0.80) for all measurements except for posttreatment nodal SUVpeak values (0.51-0.59). Low posttreatment SUV values made thresholding for MTV calculation difficult to reproduce; thus, posttreatment MTVs were not tabulated.

Correlation of baseline SUV values with treatment outcomes

Elevated baseline primary/nodal FDG SUVmax or SUVpeak values above median value of the study cohort were not associated with poor clinical outcomes (Table 3). This also held true when either value was alternatively analyzed as a continuous variable (data not shown).

Table 3 Correlation of baseline SUV values with treatment outcomes

Variable	Endpoint	Events/total	Hazard ratio (95% CI)	*P*
Primary SUVmax (> vs ≤ median)	Local relapse	0/34 vs 8/34	Cannot estimate	
	Local−regional relapse	4/34 vs 12/34	0.31 (0.10-0.97)	.04
	Distant metastasis	2/34 vs 6/34	0.32 (0.06-1.57)	.16
	Progression-free survival	6/34 vs 18/34	0.30 (0.12-0.75)	.01
	Overall survival	4/34 vs 11/34	0.38 (0.12-1.20)	.10
Nodal SUVmax (> vs ≤ median)	Regional relapse	6/32 vs 7/33	0.89 (0.30-2.66)	.84
	Local−regional relapse	6/32 vs 9/33	0.72 (0.26-2.05)	.54
	Distant metastasis	2/32 vs 5/33	0.40 (0.08-2.05)	.27
	Progression-free survival	8/32 vs 15/33	0.56 (0.24-1.33)	.19
	Overall survival	4/32 vs 10/33	0.44 (0.14-1.39)	.16
Primary SUVpeak (> vs ≤ median)	Local relapse	1/34 vs 7/34	0.15 (0.02-1.23)	.08
	Local−regional relapse	6/34 vs 10/34	0.60 (0.22-1.65)	.32
	Distant metastasis	3/34 vs 5/34	0.62 (0.15-2.62)	.52
	Progression-free survival	9/34 vs 15/34	0.62 (0.27-1.42)	.25
	Overall survival	5/34 vs 10/34	0.55 (0.19-1.62)	.28
Nodal SUVpeak (> vs ≤ median)	Regional relapse	6/32 vs 7/33	0.93 (0.31-2.77)	.90
	Local−regional relapse	6/32 vs 9/33	0.75 (0.27-2.13)	.59
	Distant metastasis	3/32 vs 4/33	0.81 (0.18-3.63)	.79
	Progression-free survival	9/32 vs 14/33	0.74 (0.32-1.72)	.49
	Overall survival	5/32 vs 9/33	0.61 (0.20-1.82)	.37

Abbreviations as in Table 2.
Hazard ratios estimated from Cox models.

Correlation of posttreatment SUV values with nodal response

The 2-year nodal relapse rates were 3.9% (95% confidence interval [CI] 0.0-11.4) in patients with clinical CR, negative PET; 33.3% (95% CI 8.4-58.2) in clinical < CR, negative PET; 16.7% (95% CI 0.0-38.8) in clinical CR, positive PET; and 11.1% (95% CI 0.0-33.1) in clinical < CR, positive PET. There were only 10 nodal relapses across these 4 groups, and differences were not statistically significant ($P = .15$).

Correlation of baseline MTV with treatment outcomes

Pretreatment primary MTV above the study cohort median was associated with local−regional relapse (hazard ratio 4.01, 95% CI 1.28-12.52, $P = .02$) and poorer progression-free survival (hazard ratio 2.34, 95% CI 1.02-5.37, $P = .05$) (Fig. 1, Table 4). Elevated baseline nodal MTV values, on the other hand, were nonprognostic. Results for combined primary and nodal MTV were similar to primary MTV alone (hazard ratio for progression-free survival 2.37, 95% CI 0.99-5.69, $P = .05$). Primary MTV and total MTV were highly correlated (Spearman correlation 0.78 on the raw values; φ coefficient 0.57 for measures dichotomized by group median).

Table E1 (available online at www.redjournal.com) presents summary statistics for primary MTV by T stage. The mean primary MTV roughly doubled with each upstage: T2 6.39, T3 13.24, T4, 26.79. Primary MTV was a stronger prognostic factor than T stage for local−regional relapse and progression-free survival (Table 5). The hazard ratios for primary MTV changed little when T stage was added to a bivariate model, but T stage lost much of its prognostic value for local−regional relapse, and all for progression-free survival, when primary MTV was added to the model. Akaike information criterion is lowest when primary MTV remained as a single variable in the model. Collinearity was not an issue, with the variance inflation factor limited to 1.27.

Subgroup analysis for p16-positive oropharynx cases

Primary MTV and total MTV remained prognostic in patients with p16-associated oropharyngeal cancer (Table E2, available online at www.redjournal.com). This ability was associated exclusively with MTV; SUVmax and SUVpeak remained nonprognostic.

Discussion

A mixed picture has emerged regarding utility of FDG-PET−derived imaging biomarkers in head-and-neck cancer. Although older series suggest poor outcomes in patients presenting with highly elevated SUV values (11-13), prospective institutional data (14) suggest that baseline SUVmax and SUVpeak values are not prognostic, either in unselected patients or in patients with HPV-associated oropharyngeal disease. In fact, our present results from RTOG 0522 alternatively suggest better

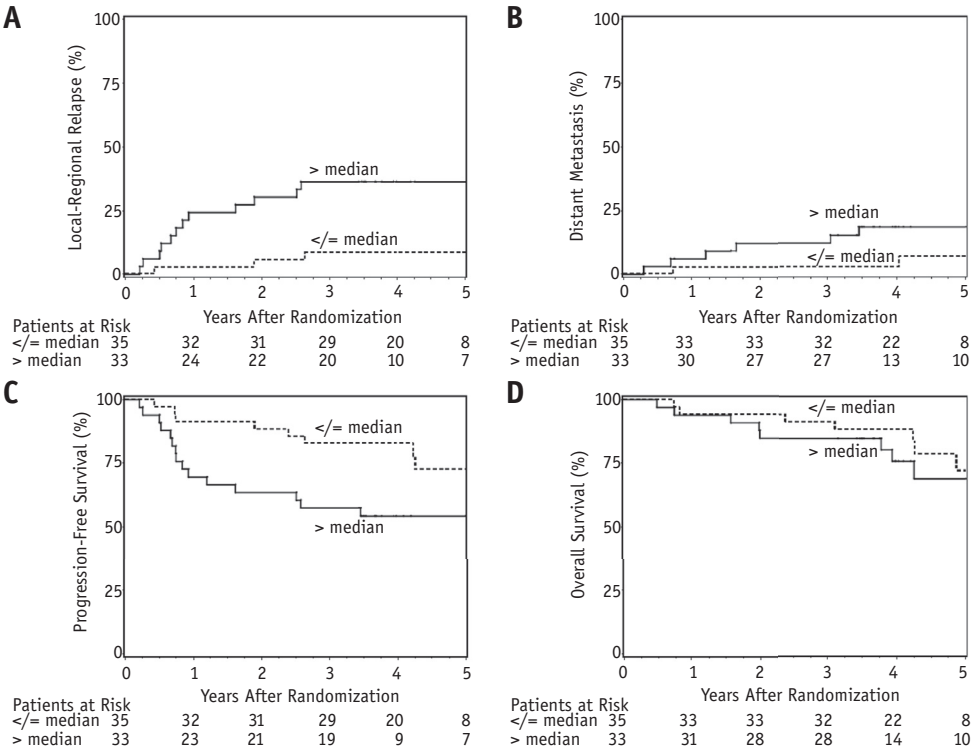

Fig. 1. Cumulative incidence estimates of local–regional relapse (A) and distant metastasis (B) and Kaplan-Meier estimates of progression-free survival (C) and overall survival (D) for patients with baseline primary metabolic tumor volume (MTV) ≤median or > median. Two-year local–regional relapse rates were 5.7% (95% confidence interval [CI] 0-13.5) for patients with MTV ≤ median and 30.3% (95% CI 14.3-46.3) for patients with MTV > median. Two-year distant metastases rates were 2.9% (95% CI 0-8.5) for patients with MTV ≤ median and 12.1% (95% CI 0.8-23.4) for patients with MTV > median. Two-year progression-free survival rates were 88.6% (95% CI 78.0-99.1) for patients with MTV ≤ median and 63.6% (95% CI 47.2-80.0) for patients with MTV > median. Two-year overall survival rates were 94.3% (95% CI 86.6-100) for patients with MTV ≤ median and 84.9% (95% CI 72.6-97.1) for patients with MTV > median.

locoregional control and progression-free survival in patients presenting with primary SUVmax above the median value for the study cohort (Table 3). Whether this reflects statistical noise from a small sample or true limitations of SUV-based measures, it is clear that validation of a more robust FDG-PET–defined biomarker could be of considerable clinical value, complementing the incrementally improved staging accuracy of PET/CT.

Metabolic tumor volume is a novel PET-specific measure, which incorporates complementary information relevant to both disease burden (akin to clinical stage and CT-derived tumor volume, which are prognostic for outcomes [25, 26]) and tumor metabolic activity. Although the size of our dataset is limited, our findings add to an emerging literature, which suggests that high primary MTV is associated with local–regional relapse and reduced progression-free survival after radiation therapy.

Despite expected correlation of primary MTV with T stage (15), earlier reports suggest that MTV dominates univariate and multivariate models of treatment outcome, independent of T stage or human papillomavirus (HPV) infection status (16, 17). Interestingly, these same reports also suggest lack of utility for SUV values and nodal MTV. Although formal multivariate analysis was not deemed appropriate owing to limited events, the effect of primary MTV was independent of age, gender, Zubrod performance status, pack-years, T stage, and N stage in bivariate analysis. Additionally, although elevated primary MTV was strongly associated with regional and distant disease failure in our series, nodal MTV remained nonprognostic. Although tempting to speculate that biology and tissue architecture specific to nodal disease (such as cystic adenopathy seen in HPV-associated disease) confounds productive use of MTV in regional disease sites, this requires follow-up study for definitive characterization. Along this vein, closer correlation of primary MTV with locoregional/distant disease failure than with local failure (Table 4) is in keeping with prior reported data (17) and suggests need for complementary work to mechanistically explain the exact biological significance of primary MTV.

Pretreatment imaging biomarkers can potentially direct personalized treatment intensity. Our findings also suggest that intraobserver PET reading consistency is best in the pretreatment setting. Both issues are of relevance to HPV-associated oropharyngeal disease, which typically responds

Table 4 Correlation of baseline primary MTV with treatment outcomes

Variable	Endpoint	Events/total	Hazard ratio (95% CI)	*P*
Primary MTV (continuous)	Local relapse	8/68	1.05 (1.00-1.09)	.06
	Local–regional relapse	16/68	1.05 (1.02-1.08)	<.01
	Distant metastasis	8/68	1.04 (1.01-1.08)	.02
	Progression-free survival	24/68	1.05 (1.02-1.07)	<.01
	Overall survival	15/68	1.03 (1.00-1.06)	.08
Primary MTV (> vs ≤ median)	Local relapse	5/33 vs 3/35	1.96 (0.47-8.23)	.36
	Local–regional relapse	12/33 vs 4/35	4.01 (1.28-12.52)	.02
	Distant metastasis	6/33 vs 2/35	3.62 (0.73-18.04)	.12
	Progression-free survival	15/33 vs 9/35	2.34 (1.02-5.37)	.05
	Overall survival	8/33 vs 7/35	1.40 (0.51-3.86)	.52

Abbreviations as in Table 2.
Hazard ratios estimated from Cox models.

briskly to treatment and may benefit from pretreatment PET-guided deintensification strategies to limit toxicity. Seven of 8 oropharyngeal cancer relapses in the PET imaging analysis cohort were p16-positive. These 7 patients suffered recurrences in varied locations: 1 tumor recurred in the primary site, 4 recurred in the neck, 1 recurred locoregionally, and 1 recurred both locoregionally and distantly. Six of these 7 patients had discordant above-median baseline primary MTV values in the face of below-median baseline SUVmax. This spotlights the candidacy of MTV as an imaging-based identifier of "intermediate-risk" p16-positive disease (27, 28) (ie HPV-associated oropharynx patients with bulky primary tumor and/or traditional risk-factor exposure history who suffer inferior treatment outcomes and are not appropriate for deintensified therapy). Nonetheless, this remains an early hypothesis-generating result subject to the potential biases and study power limitations of unplanned subgroup analysis. Formal testing in future multicenter trials is mandatory for validation.

Presence of a strong prognostic signal for MTV in our small, heterogeneous study cohort is encouraging. Nonetheless, MTV has limitations as a candidate imaging biomarker. No standard definition for MTV currently exists. Earlier series have used variable, arbitrarily defined percent primary and nodal SUVmax cutoffs to delineate MTV. Accordingly, the relative prognostic performance of specific MTV cutoff values is difficult to compare across published series. In addition, practical use of MTV after treatment is challenging given return of responding tumor sites to background tracer uptake levels. Although prior pilot work supports the use of a post-treatment MTV threshold of SUV = 2.0 (29), this could not be feasibly reproduced in this study with commercially available software. Future confirmatory work remains necessary.

Another key limitation of our findings is that formal participation of patients enrolled onto the RTOG 0522 trial was limited to 116 patients, and that only 74 of these patients had a full complement of serial imaging available for the analysis specified by protocol. Although reasons for low participation remain speculative, the resulting small sample size suffers from limited statistical power and potential provider-dependent selection biases. Fortunately, the clinical characteristics of the subgroup match reasonably with those of the eligible population as a whole (Table 1), suggesting broader applicability of our findings to the general head-and-neck radiation therapy patient population. Similar study power issues also impact our secondary subgroup analysis of p16-positive patients.

Table 5 Primary MTV versus T stage

Model	AIC	Covariate(s)	Hazard ratio (95% CI)	*P*
Local–regional relapse				
1	120.28	Primary MTV (> vs ≤ median)	4.01 (1.28-12.52)	.02
2	124.63	T stage (T4 vs T2-3)	2.34 (0.83-6.59)	.11
3	121.98	Primary MTV (> vs ≤ median)	3.59 (1.07-12.11)	.04
		T stage (T4 vs T2-3)	1.36 (0.45-4.11)	.58
Progression-free survival				
1	183.20	Primary MTV (> vs ≤ median)	2.34 (1.02-5.37)	.05
2	186.52	T stage (T4 vs T2-3)	1.54 (0.63-3.74)	.34
3	185.19	Primary MTV (> vs ≤ median)	2.31 (0.94-5.70)	.07
		T stage (T4 vs T2-3)	1.03 (0.39-2.71)	.95

Abbreviation: AIC = Akaike information criterion. Other abbreviations as in Table 2.
Hazard ratios estimated from Cox models.

To conclude, we found a unique prognostic signal for primary tumor FDG-PET–derived MTV in patients with locally advanced head-and-neck cancer. This finding remains preliminary and requires technical refinement and clinical confirmation. There is growing need to better match resource-intensive treatment to individual patient risk. Validation of MTV, a straightforward, noninvasive measure derived from routine workup imaging, could be an important step toward accomplishing this goal.

References

1. Kyzas PA, Evangelou E, Denaxa-Kyza D, et al. 18F-fluorodeoxyglucose positron emission tomography to evaluate cervical node metastases in patients with head and neck squamous cell carcinoma: A meta-analysis. *J Natl Cancer Inst* 2008;100:712-720.
2. Antoch G, Saoudi N, Kuehl H, et al. Accuracy of whole-body dual-modality fluorine-18-2-fluoro-2-deoxy-D-glucose positron emission tomography and computed tomography (FDG-PET/CT) for tumor staging in solid tumors: Comparison with CT and PET. *J Clin Oncol* 2004;22:4357-4368.
3. Czernin J, Allen-Auerbach M, Schelbert HR. Improvements in cancer staging with PET/CT: Literature-based evidence as of September 2006. *J Nucl Med* 2007;48(Suppl 1):78S-88S.
4. Facey K, Bradbury I, Laking G, et al. Overview of the clinical effectiveness of positron emission tomography imaging in selected cancers. *Health Technol Assess* 2007;11. iii-iv, xi-267.
5. Syed R, Bomanji JB, Nagabhushan N, et al. Impact of combined (18) F-FDG PET/CT in head and neck tumours. *Br J Cancer* 2005;92:1046-1050.
6. Schoder H, Yeung HW, Gonen M, et al. Head and neck cancer: Clinical usefulness and accuracy of PET/CT image fusion. *Radiology* 2004;231:65-72.
7. Schwartz DL, Ford E, Rajendran J, et al. FDG-PET/CT imaging for preradiotherapy staging of head-and-neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys* 2005;61:129-136.
8. Veit-Haibach P, Luczak C, Wanke I, et al. TNM staging with FDG-PET/CT in patients with primary head and neck cancer. *Eur J Nucl Med Mol Imaging* 2007;34:1953-1962.
9. Gregoire V. Is there any future in radiotherapy planning without the use of PET: Unraveling the myth. *Radiother Oncol* 2004;73:261-263.
10. Moeller BJ, Rana V, Cannon BA, et al. Prospective imaging assessment of mortality risk after head-and-neck radiotherapy. *Int J Radiat Oncol Biol Phys* 2010;78:667-674.
11. Halfpenny W, Hain SF, Biassoni L, et al. FDG-PET. A possible prognostic factor in head and neck cancer. *Br J Cancer* 2002;86:512-516.
12. Allal AS, Dulguerov P, Allaoua M, et al. Standardized uptake value of 2-[(18)F] fluoro-2-deoxy-D-glucose in predicting outcome in head and neck carcinomas treated by radiotherapy with or without chemotherapy. *J Clin Oncol* 2002;20:1398-1404.
13. Schwartz DL, Rajendran J, Yueh B, et al. FDG-PET prediction of head and neck squamous cell cancer outcomes. *Arch Otolaryngol Head Neck Surg* 2004;130:1361-1367.
14. Moeller BJ, Rana V, Cannon BA, et al. Prospective risk-adjusted [18F] fluorodeoxyglucose positron emission tomography and computed tomography assessment of radiation response in head and neck cancer. *J Clin Oncol* 2009;27:2509-2515.
15. Chung MK, Jeong HS, Park SG, et al. Metabolic tumor volume of [18F]-fluorodeoxyglucose positron emission tomography/computed tomography predicts short-term outcome to radiotherapy with or without chemotherapy in pharyngeal cancer. *Clin Cancer Res* 2009;15:5861-5868.
16. La TH, Filion EJ, Turnbull BB, et al. Metabolic tumor volume predicts for recurrence and death in head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2009;74:1335-1341.
17. Tang C, Murphy JD, Khong B, et al. Validation that metabolic tumor volume predicts outcome in head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2012;83:1514-1520.
18. Kalbfleish J, Prentice R. The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons; 1980.
19. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457-481.
20. Cox DR. Regression models and life tables. *J R Stat Soc* 1972;34:187-229.
21. Yao M, Graham MM, Hoffman HT, et al. The role of post-radiation therapy FDG PET in prediction of necessity for post-radiation therapy neck dissection in locally advanced head-and-neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys* 2004;59:1001-1010.
22. Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, et al. Fully conditional specification in multivariate imputation. *J Stat Comput Sim* 2006;76:1049-1064.
23. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons; 1987.
24. Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control* 1974;19:716-723.
25. Hermans R, Op de beeck K, Van den Bogaert W, et al. The relation of CT-determined tumor parameters and local and regional outcome of tonsillar cancer after definitive radiation treatment. *Int J Radiat Oncol Biol Phys* 2001;50:37-45.
26. Lodder WL, Pameijer FA, Rasch CR, et al. Prognostic significance of radiologically determined neck node volume in head and neck cancer: A systematic review. *Oral Oncol* 2012;48:298-302.
27. Weinberger PM, Yu Z, Haffty BG, et al. Molecular classification identifies a subset of human papillomavirus–associated oropharyngeal cancers with favorable prognosis. *J Clin Oncol* 2006;24:736-747.
28. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* 2010;363:24-35.
29. Murphy JD, La TH, Chu K, et al. Postradiation metabolic tumor volume predicts outcome in head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2011;80:514-521.