Int. J. Radiation Oncology Biol. Phys., Vol. 78, No. 4, pp. 1057–1063, 2010

0360-3016/$–see front matter

ELSEVIER

doi:10.1016/j.ijrobp.2009.09.022

CLINICAL INVESTIGATION **Esophagus**

COMPARISON OF STANDARDIZED UPTAKE VALUE–BASED POSITRON EMISSION TOMOGRAPHY AND COMPUTED TOMOGRAPHY TARGET VOLUMES IN ESOPHAGEAL CANCER PATIENTS UNDERGOING RADIOTHERAPY

FAISAL S. VALI, M.SC., M.D.,* SUNEEL NAGDA, M.D.,* WILLIAM HALL, B.SC.,* JAMES SINACORE, PH.D.,† MINGCHENG GAO, PH.D.,* STEVEN H. LEE, D.O.,‡ ROBERT HONG, M.D.,§ MARGARET SHOUP, M.D., F.A.C.S.,‖ AND BAHMAN EMAMI, M.D., F.A.C.R., F.A.S.T.R.O.*

*Department of Radiation Oncology, †Department of Preventive Medicine and Epidemiology, and ‖Division of Surgical Oncology, Department of Surgery, Loyola University Medical Center, Maywood, IL; ‡Department of Radiology, Section of Nuclear Medicine, Orlando VA Medical Center, Orlando, FL; and §Department of Radiation Oncology, Virginia Hospital Center, Arlington, VA

Purpose: To study various standardized uptake value (SUV)-based approaches to ascertain the best strategy for delineating metabolic tumor volumes (MTV).

Methods and Materials: Twenty-two consecutive previously treated esophageal cancer patients with positron emission tomography (PET) imaging and computed tomography (CT)-based radiotherapy plans were studied. At the level of the tumor epicenter, MTVs were delineated at 11 different thresholds: SUV ≥2, ≥2.5, ≥3, ≥3.5 (SUV_n); ≥40%, ≥45%, and ≥50% of the maximum ($SUV_{n\%}$); and mean liver SUV + 1, 2, 3, and 4 standard deviations ($SUV_{Ln\sigma}$). The volume ratio and conformality index were determined between MTVs, and the corresponding CT/endoscopic ultrasound-based gross tumor volume (GTV) at the epicenter. Means were analyzed by one-way analysis of variance for repeated measures and further compared using a paired *t* test for repeated measures.

Results: The mean conformality indices ranged from 0.33 to 0.48, being significantly ($p < 0.05$) closest to 1 at $SUV_{2.5}$ (0.47 ± 0.03) and $SUV_{L4\sigma}$ (0.48 ± 0.03). The mean volume ratios ranged from 0.39 to 2.82, being significantly closest to 1 at $SUV_{2.5}$ (1.18 ± 0.36) and $SUV_{L4\sigma}$ (1.09 ± 0.15). The mean value of the SUVs calculated using the $SUV_{L4\sigma}$ approach was 2.4.

Conclusions: Regardless of the SUV thresholding method used (i.e., absolute or relative to liver mean), a threshold of approximately 2.5 yields the highest conformality index and best approximates the CT-based GTV at the epicenter. These findings may ultimately aid radiation oncologists in the delineation of the entire GTV in esophageal cancer patients.

Esophageal cancer, Positron emission tomography, Radiotherapy planning, Target volume, Standardized uptake value.

INTRODUCTION

Locally advanced esophageal cancer continues to be a lethal disease, with an estimated 5-year survival rate of 20% (1). The majority of such patients are treated neoadjuvantly or definitively with a combination of radiotherapy (RT) and chemotherapy (2). The accurate delineation of tumor volume can be difficult in the nonsurgical setting and is vital to maximizing the therapeutic ratio from radiotherapy.

Computed tomography (CT) is the standard imaging modality used for tumor volume delineation during RT planning for esophageal cancer (3, 4). Computed tomography, combined with information from endoscopic ultrasound (EUS), is deemed sufficiently accurate for delineating the radial extent of the tumor (5–7). However, CT is often inadequate in defining the tumor's longitudinal extent (8). Other modalities, such as esophagoduodenoscopy, may aid in longitudinal delineation of tumor but cannot be directly imported in RT planning systems and are therefore limiting.

The value of ^{18}F-fluorodeoxy-2-D-glucose positron emission tomography (^{18}F-FDG PET) in the staging of esophageal cancer has been demonstrated, with reported sensitivities and specificities for detecting metastatic disease in the 80–90% range (5). Furthermore, much interest has been expressed in incorporating metabolic data from ^{18}F-FDG PET scans to improve the accuracy of delineating tumor for RT planning by identifying a metabolic tumor volume (MTV) (9, 10). Standardized uptake value (SUV)-based approaches that identify a threshold SUV in the region of

Reprint requests to: Suneel N. Nagda, M.D., Department of Radiation Oncology, Loyola University Medical Center, 2160 S. First Ave., Bldg. 105-2936, Maywood, IL 60153. Tel: (708) 216-2562; Fax: (708) 216-6076; E-mail: snagda@gmail.com

Presented at the 50th Annual Meeting of the American Society for Therapeutic Radiology and Oncology, September 21–25, 2008, Boston, MA.

Conflict of interest: none.

Received April 25, 2009, and in revised form Aug 26, 2009. Accepted for publication Sep 10, 2009.

the tumor above which all cells are deemed to be neoplastic have received much attention (11–16). The challenges with these approaches have included the interobserver and intraobserver deviations in SUV-based MTV delineation. These deviations are compounded by the variability in the dose of radioactive glucose administered and the time at which it is administered—both parameters that can alter SUV-based MTVs dramatically (17). In the setting of non–small-cell lung cancer, Nestle *et al*. (18) document a 3.6-fold difference in mean MTV, depending on the SUV-based thresholding mechanism used. Although none of the thresholding approaches have been standardized, the most common ones are broadly categorized as (*1*) absolute SUV (SUV_{Abs}), (*2*) SUV relative to the maximum tumor SUV ($SUV_{\%}$), and (*3*) SUV relative to the mean liver SUV ($SUV_{L\sigma}$) (8, 13, 18, 19). It is unclear which general category to use, and more importantly, what value to use within that category.

To discern the optimal SUV-based thresholding strategy to use for delineating the entire MTV (including longitudinal extent) in esophageal cancer patients, we systematically compared the radial extent of the corresponding MTVs against the radial extent of the least ambiguous portion of the tumor epicenter as established on CT and EUS. In doing so, we assumed that the strategy that would provide the best match of the easily identified radial margins on CT/EUS would also provide the best match of the less easily identified longitudinal margins of the tumor.

METHODS AND MATERIALS

Study population

This study was approved by Loyola University Medical Center's institutional review board. Twenty-two consecutive patients with locally advanced esophageal cancer treated in the Department of Radiation Oncology were retrospectively reviewed. All patients had ^{18}F-FDG PET imaging performed as part of the initial staging workup. Patients were deemed eligible for the study as long as their primary esophageal tumor was metabolically active and distinct from other metabolic activity. A single patient with an inflammatory pneumonitic infiltrate adjacent to the primary tumor was excluded. Hence the final study population comprised 21 patients.

PET acquisition

Patients were scanned with a dedicated whole-body Philips Allegro PET scanner (Philips Medical Systems, Bothell, WA) from the mid-skull to the mid-thigh, after being injected intravenously with FDG. All patients had their height and weight determined at the time of study. Serum resting glucose values were measured and determined to be within acceptable ranges (65–200 mg/dL) before FDG administration. Fluorodeoxyglucose quantity was based on weight and ranged between 10 and 15 mCi (370–555 MBq). With gentle intravenous fluid hydration, FDG was injected intravenously, and isotope was distributed over 90 min in a resting state. The patient was positioned supine, and scan acquisition occurred over approximately 45 min. Cesium-137 transmission was used for attenuation correction. Images were reconstructed using an iterative three-dimensional algorithm (row-action maximum likelihood algorithm) with 4-mm voxels and stored using a Digital Imaging and Communications in Medicine (DICOM) format. All studies were read on the same day of acquisition by an experienced American Board of Nuclear Medicine–certified physician. The maximum SUV in the region of the tumor was calculated and recorded. The mean SUV of the liver was retrospectively obtained for the purposes of this study (Hermes Medical Solutions, Stockholm, Sweden).

Planning CT acquisition

Patients were simulated supine with arms overhead using a Philips AcQSim CT scanner (Philips Medical Systems) acquiring images with a slice thickness and table index of 5 mm (512 × 512 matrix, 0.94-mm pixel size) from the second cervical vertebral body through to the liver. Custom cradle immobilization was used (Smithers Medical Products, North Canton, OH), and 1 tablespoon of oral barium paste and 100 mL of intravenous contrast were administered to each patient immediately before the scan.

PET/CT coregistration and fusion

For the purposes of this study, the diagnostic PET-DICOM data were transferred to a RT planning workstation (Focal v4.31; CMS, St. Louis, MO) and fused to the original CT used for RT planning. A combination of manual and automated coregistration was used. Manual coregistration was carried out by visualizing the PET images using arbitrary window and level values that adequately displayed normal metabolically active organs, such as the kidneys and heart, and fusing these organs with their CT counterparts.

CT/EUS-based target volume delineation

To ultimately compare the metabolic and CT volumes, we defined a reference CT/EUS-based target volume, CT-GTV_E. The CT-GTV_E consisted of a 1-cm subsection of the original CT/EUS-based gross tumor volume (used for RT planning), at the level of the tumor epicenter. This was derived by reducing the original GTV superiorly and inferiorly around the axial CT slice containing the greatest tumor burden and the least amount of discrepancy in delineating the radial extent. No alterations were made in the radial direction. The tumor epicenter was determined by a single radiation oncologist (S.N.) and verified by another (F.S.V.). Endoscopic information was also used for verification. Figure 1 depicts the axial and sagittal view of one such CT-GTV_E in a patient.

PET isocontour delineation

We studied 11 different SUV-based MTVs for each patient. Standardized uptake value thresholds were used to generate these metabolic volumes, each containing voxels with SUVs greater than or equal to the threshold value. Again, the thresholded values fell into three broad categories: SUV_{Abs}, $SUV_{\%}$, and $SUV_{L\sigma}$ (see Introduction). Within SUV_{Abs} we studied SUV $\geq$2.0, $\geq$2.5, $\geq$3.0, and $\geq$3.5 ($SUV_{2.0}$, $SUV_{2.5}$, $SUV_{3.0}$, and $SUV_{3.5}$, respectively). Within $SUV_{\%}$ we studied SUV $\geq$40%, $\geq$45%, and $\geq$50% of maximum ($SUV_{40\%}$, $SUV_{45\%}$, and $SUV_{50\%}$, respectively). Within $SUV_{L\sigma}$ we studied SUV $\geq$ mean liver SUV (L) + 1 standard deviation (SD), $\geq$ L + 2SD, $\geq$ L + 3SD, and $\geq$ L + 4SD ($SUV_{L1\sigma}$, $SUV_{L2\sigma}$,$SUV_{L3\sigma}$, and $SUV_{L4\sigma}$, respectively) (13, 18, 20). The window/level values on the RT planning system were adjusted to display the appropriate SUV threshold-based volume using previously published methods (13). The resulting metabolic volume was manually contoured and was modified to exclude any overlap with the heart, but any overlap with the liver or lung was left unmodified. We anatomically excluded the heart because its muscle is usually at least as FDG avid as tumor (if not more) and would have resulted in clinically insignificant MTVs. The normal liver and lungs were not necessarily

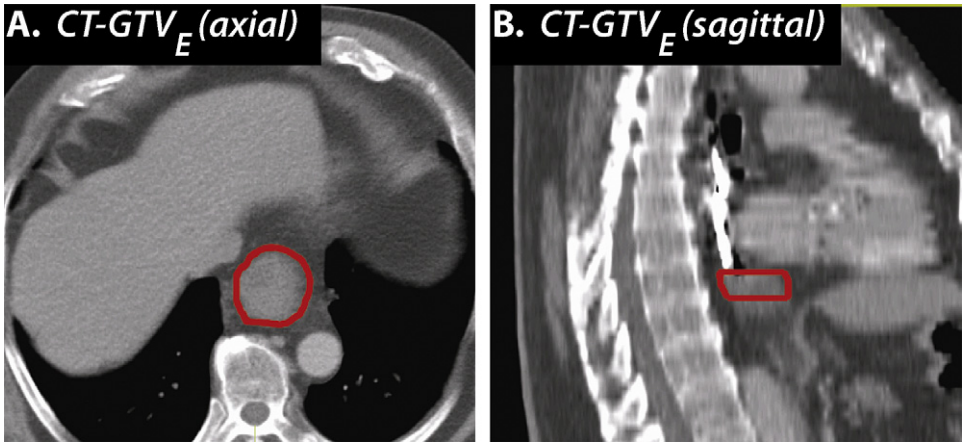

Fig. 1. Example of the reference computed tomography/endoscopic ultrasound–based gross tumor volume (CT-GTV$_E$) (red). The CT-GTV$_E$ represents a 1-cm-long subsection of the original GTV that was used for initial treatment. In deriving the CT-GTV$_E$ the original GTV was reduced in the superior and inferior direction, while being centered at the tumor epicenter. The radial contours of the original GTV were not altered.

excluded because they typically have only low-level FDG activity. Therefore, any significant FDG activity in the lung/liver was included as tumor because we believed that other causes for increased FDG activity in these organs could not be distinguished reliably. For each patient we generated 12 tumor volumes at the level of the tumor epicenter: 1 CT/EUS based (CT-GTV$_E$) and 11 metabolic based. Figure 2 demonstrates four such MTVs (SUV$_{2.0}$, SUV$_{2.5}$, SUV$_{3.0}$, and SUV$_{40\%}$) in relation to the CT-GTV$_E$, in the same patient on the same axial slice.

Volume ratio and conformality index

To compare the metabolic volumes with the CT-GTV$_E$ we used two measures: volume ratio (VR) and conformality index (CI). The VR is simply the ratio of two volumes. The CI describes the spatial relationship between two volumes and is calculated using the formula $CI = \frac{V1 \cap V2}{V1 \cup V2}$, where V1 and V2 are the two volumes of interest, $\cap$ is the symbol for their intersection, and $\cup$ symbolizes their union. A normalized CI was calculated by identifying the maximum CI for a particular patient and then dividing each of their 11 CIs by that maximum.

Statistical analysis

The mean CI, normalized CI, and VR for each of the 11 SUV thresholds were calculated. The overall difference among the means for each of the thresholds was tested for significance with a repeated-measures analysis of variance. Then, the SUV threshold resulting in a mean CI closest to 1 was compared against the mean CIs produced by the other SUV thresholds using a paired *t* test for related samples. In addition, if any of the mean indices for the SUV thresholds were within an arbitrarily picked 10% of the ideal value of 1, a one-sampled *t* test was conducted to ascertain whether any of the 11 SUV thresholds pointed to a theoretical SUV threshold that would consistently result in the ideal value of 1, suggesting that the thresholding strategy was no worse than some ideal threshold. This entire process was then repeated using the mean normalized CIs and VRs (SPSS v16; SPSS, Chicago, IL).

RESULTS

Twenty of the 21 patients had American Joint Committee on Cancer 2002 Stage III esophageal cancer. Positron emission tomography scans were obtained an average of 5.8 $\pm$ 2.9 days before the planning CT, with a range of 43 days before planning CT to 28 days after.

SUV characteristics

The maximum SUV of the tumors ranged from 4.1 to 25.4 (mean $\pm$ SE, 10.6 $\pm$ 1.2). Table 1 lists the mean SUVs calculated using the respective thresholding strategies. The means ranged from 2.0 to 5.3. Of note, the mean SUV associated with the SUV$_{L4\sigma}$ strategy was 2.4 (Table 1).

Mean CI, normalized CI, and VR

As listed in Table 2, SUV$_{L4\sigma}$ resulted in the most desirable (closest to 1) mean CI (0.48 $\pm$ 0.03), normalized CI (0.85 $\pm$ 0.04), and VR (1.09 $\pm$ 0.15). With regard to mean CI, after SUV$_{L4\sigma}$ the next most desirable means resulted from SUV$_{L3\sigma}$ (0.47 $\pm$ 0.03) and SUV$_{2.5}$ (0.47 $\pm$ 0.03). With regard to mean VR, after SUV$_{L4\sigma}$ the next most desirable means resulted from SUV$_{2.5}$ (1.18 $\pm$ 0.35) and SUV$_{L3\sigma}$ (1.24 $\pm$ 0.16).

In general, the mean CIs ranged from 0.325 (SUV$_{50\%}$) to 0.478 (SUV$_{L4\sigma}$) (F = 3.28, p = 0.02), and the mean normalized CIs ranged from 0.575 (SUV$_{50\%}$) to 0.846 (SUV$_{L4\sigma}$) (F = 3.59, p = 0.01). The mean VRs ranged from 0.393 (SUV$_{50\%}$) to 2.875 (SUV$_{2.0}$) (F = 6.6, p = 0.01) (Table 2). The trend in mean CI, as it relates to the thresholding strategies and their calculated SUVs, is depicted in Fig. 3. Similarly, the trend in mean VR is shown in Fig. 4. Both figures indicate that SUV$_{L4\sigma}$ results in mean CIs and VRs closest to 1, with SUV$_{L3\sigma}$ and SUV$_{2.5}$ not far behind.

Comparison with SUV$_{L4\sigma}$

When SUV$_{L4\sigma}$ (being the threshold strategy that resulted in volumes most similar to our reference CT/EUS-based volume) was compared against each of the other thresholding approaches, significant differences in the mean CI were seen with SUV$_{3.0}$, SUV$_{3.5}$, SUV$_{40\%}$, SUV$_{45\%}$, and SUV$_{50\%}$ (Table 3). No differences were seen with SUV$_{2.5}$ and

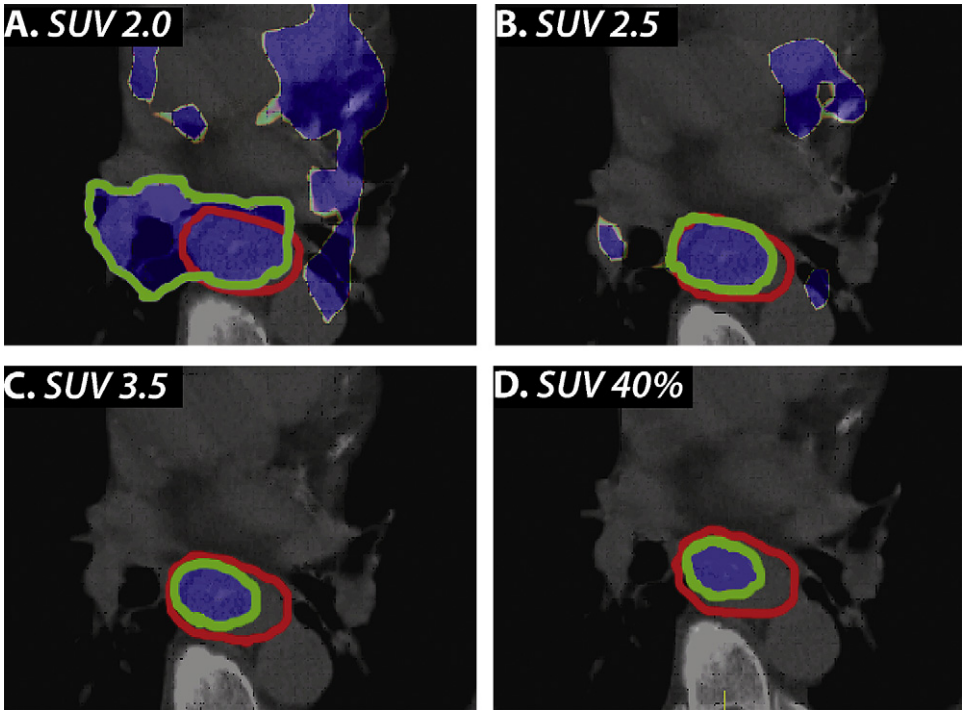

Fig. 2. Isocontours of the metabolic tumor volumes (MTV, green) using a sample of the various thresholding approaches studied. The computed tomography/endoscopic ultrasound–based gross tumor volume is indicated in red. All images are from the same patient. The blue islands within each image represent metabolic activity with the appropriate standardized uptake value (SUV) threshold applied. Images A, B, C, and D demonstrate thresholds of SUV ≥2.0, ≥2.5, ≥3.5, and 40% of maximum, respectively. On the basis of the SUV threshold used, different MTVs were delineated. In addition, the islands of SUV-based activity in the heart were not included in the MTV.

$SUV_{L3\sigma}$. With regard to VR, when compared with $SUV_{L4\sigma}$ significant differences were observed with all SUV thresholds except for $SUV_{2.0}$, $SUV_{2.5}$, and $SUV_{40\%}$ (Table 3).

Comparison with the hypothetical ideal threshold

When all 11 of the thresholding strategies were compared against some hypothetical ideal threshold (with a mean VR of exactly 1), only $SUV_{2.5}$, $SUV_{L4\sigma}$, and $SUV_{L3\sigma}$ were noted to be significantly similar to the ideal, with regard to VR (Table 4). In addition, among the three strategies, $SUV_{2.5}$ ($p = 0.63$) and $SUV_{L4\sigma}$ ($p = 0.56$) were more strongly similar to the ideal threshold than $SUV_{L3\sigma}$ ($p = 0.15$).

Table 1. Mean SUVs corresponding to each of the 11 SUV thresholding strategies

SUV threshold	Mean SUV ± SE*
SUV2.0	2.0
SUVL1σ	2.007 ± 0.090
SUVL2σ	2.138 ± 0.095
SUVL3σ	2.270 ± 0.101
SUVL4σ	2.401 ± 0.108
SUV2.5	2.5
SUV3.0	3.0
SUV3.5	3.5
SUV40%	4.238 ± 0.493
SUV45%	4.768 ± 0.554
SUV50%	5.298 ± 0.616

Abbreviations: SUV = standardized uptake value; SE = standard error; SUVn = SUV of n; SUVLnσ = mean liver SUV + n standard deviations; SUVn% = n% of maximum tumor SUV.

* Because SUV 2.0, 2.5, 3.0, and 3.5 directly specify their numeric threshold value (instead of being relative to the mean liver SUV or the maximum SUV), they have no variation in their mean values

DISCUSSION

This study is unique in that it not only evaluates all three commonly used strategies in the setting of esophageal cancer,

Table 2. Mean CI, NCI, and VR corresponding to each of the 11 SUV thresholding strategies

SUV threshold	Mean CI ± SE*	Mean NCI ± SE*	Mean VR ± SE*
SUV2.0	0.454 ± 0.044	0.790 ± 0.064	2.875 ± 0.893
SUV2.5	0.467 ± 0.034	0.832 ± 0.046	1.176 ± 0.355
SUV3.0	0.409 ± 0.039	0.723 ± 0.058	0.645 ± 0.094
SUV3.5	0.362 ± 0.044	0.640 ± 0.069	0.464 ± 0.069
SUV40%	0.399 ± 0.035	0.703 ± 0.051	0.735 ± 0.130
SUV45%	0.373 ± 0.038	0.658 ± 0.059	0.514 ± 0.083
SUV50%	0.326 ± 0.037	0.575 ± 0.059	0.393 ± 0.059
SUVL1σ	0.416 ± 0.038	0.740 ± 0.056	1.990 ± 0.256
SUVL2σ	0.457 ± 0.033	0.813 ± 0.043	1.510 ± 0.196
SUVL3σ	0.474 ± 0.030	0.839 ± 0.032	1.243 ± 0.163
SUVL4σ	0.478 ± 0.031	0.846 ± 0.036	1.091 ± 0.152

Abbreviations: CI = conformality index; NCI = normalized CI; VR = volume ratio. Other abbreviations as in Table 1.

* Values closer to 1 point toward a better thresholding strategy.

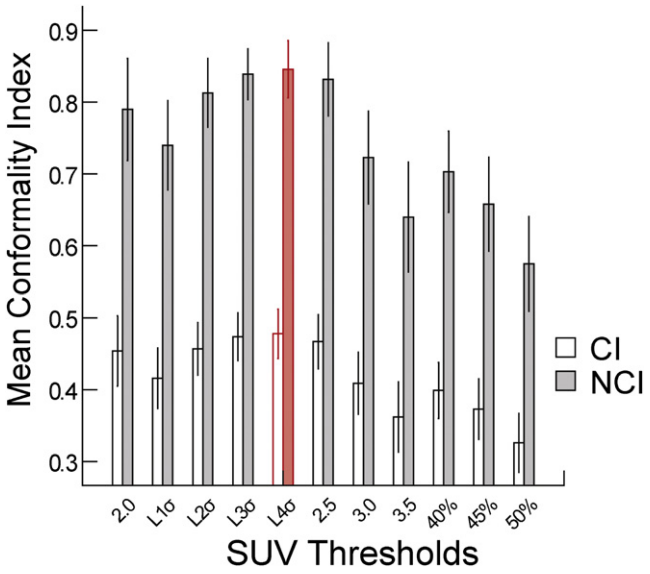

Fig. 3. Mean conformality index (CI) and mean normalized conformality index (NCI) for all 11 standard uptake value (SUV) thresholding approaches. The error bars represent the standard error. The thresholds are ordered from left to right according to their mean SUV value (refer to Table 1). Higher CIs suggest that the thresholding approach results in more conformal volumes. Lnσ = mean liver SUV + n standard deviations; n% = n% of maximum tumor SUV.

but does so by using objective indices of similarity, including CIs and VRs. Although we focused our comparisons on the radial extent of the tumor, our results suggest that using either an absolute SUV threshold of 2.5 or an SUV threshold that is 4 SDs above the mean liver SUV might be accurate enough for contouring MTVs in esophageal cancer. The results of this study do not contradict data presented elsewhere (11, 20, 21).

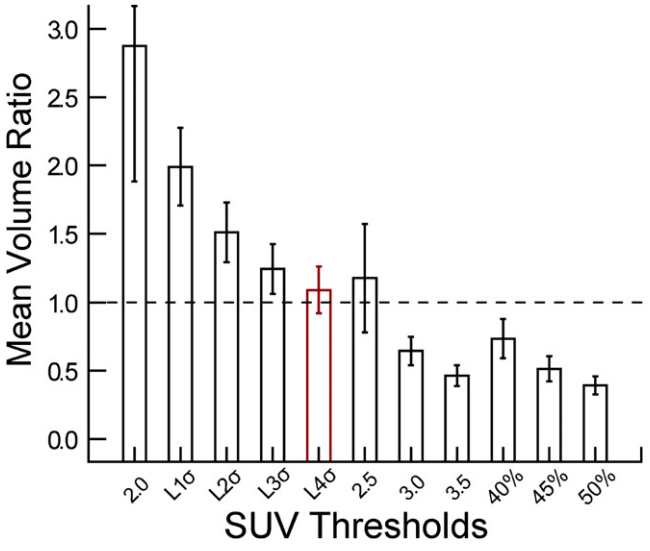

Fig. 4. Mean volume ratios for all 11 standard uptake value (SUV) thresholding approaches. The error bars represent the standard error. The thresholds are ordered from left to right according to their mean SUV value (refer to Table 1). Volume ratios closest to 1 suggest that the thresholding approach results in more similar volumes. Lnσ = mean liver SUV + n standard deviations; n% = n% of maximum tumor SUV.

Table 3. Comparison of all the other SUV thresholding strategies against the SUVL4σ strategy

SUV threshold	CI* (p)	NCI* (p)	VR* (p)
SUV2.0	0.55	0.45	0.06
SUV2.5	0.72	0.77	0.82
SUV3.0	0.04	0.03	0.01
SUV3.5	<0.01	<0.01	<0.01
SUV40%	0.05	0.04	0.10
SUV45%	0.02	0.02	<0.01
SUV50%	<0.01	<0.01	<0.01
SUVL1σ	0.08	0.09	<0.01
SUVL2σ	0.34	0.40	<0.01
SUVL3σ	0.61	0.67	<0.01

Abbreviations as in Tables 1 and 2.

* A strategy is as good as SUVL4σ with regard to the similarity index if the p value is >0.05.

For our comparison volume, we used a CT/EUS-based volume at the level of the tumor epicenter. Because 95% of our patients had tumor invading the adventitia (Stage III) at the epicenter, and because CT, combined with information from the EUS, has been deemed quite accurate in determining the radial extent of the tumor, we believed this was an acceptable reference volume (3, 4, 6, 7). Other studies have used pathologic tumor measurements as a reference comparison (19, 21). This is not feasible in patients with locally advanced esophageal cancer, who typically undergo neoadjuvant or definitive concurrent chemoradiotherapy. Additionally, the accurate quantification of the GTV and tumor length on the pathology specimen can be quite challenging with the constriction of tissue that takes place after esophagectomy (21).

The utility of a CI to compare PET and CT volumes has been described before (11). A CI captures the spatial relationship between two volumes and describes the overlap between them. Whereas an index of 1 suggests that there is 100%

Table 4. Comparison of all 11 thresholding strategies vs. some hypothetical ideal thresholding strategy, with regard to VR

SUV threshold	p*
SUV2.0	0.05
SUV2.5	0.63
SUV3.0	<0.01
SUV3.5	<0.01
SUV40%	0.05
SUV45%	<0.01
SUV50%	<0.01
SUVL1σ	<0.01
SUVL2σ	0.02
SUVL3σ	0.15
SUVL4σ	0.56

Abbreviations as in Tables 1 and 2.

* A p value of >0.05 suggests, with regard to VR, that the thresholding strategy is as good as some hypothetical ideal thresholding strategy that results in a mean VR of exactly 1.

overlap between the two volumes, with no portion of either volume outside the other, a CI of 0.5 reflects an overlap of 67% between two equally sized volumes, leaving 33% of each volume outside the other. In our study, in addition to measuring the CI, we also calculated the normalized CI. We believed that normalizing the indices for each patient would compensate for some of the uncertainties introduced by coregistration and fusion of the PET scan with the CT.

In deciding which thresholding strategies were most desirable, we considered their mean CIs and VRs. With regard to CI, the three best means resulted from (*1*) $SUV_{L4\sigma}$, (*2*) $SUV_{L3\sigma}$, and (*3*) $SUV_{2.5}$. Although $SUV_{L4\sigma}$ was significantly better than the other eight strategies, none of the top three strategies were significantly better than the other, with regard to CI. The same three strategies resulted in the best VRs, but in a slightly different order: (*1*) $SUV_{L4\sigma}$, (*2*) $SUV_{2.5}$, and (*3*) $SUV_{L3\sigma}$. Although there was no difference of significance between $SUV_{L4\sigma}$ and $SUV_{2.5}$, with regard to VR, $SUV_{L3\sigma}$ was noted to be significantly inferior. In addition, when compared with the hypothetical ideal threshold, only $SUV_{L4\sigma}$, $SUV_{2.5}$, and $SUV_{L3\sigma}$ were significantly similar, with regard to VR. However, when comparing the strength of its similarity (to the ideal threshold) with that of $SUV_{L4\sigma}$ and $SUV_{2.5}$, once again $SUV_{L3\sigma}$ seemed to be inferior. Thus, we believed that $SUV_{L4\sigma}$ and $SUV_{2.5}$ were generally superior to $SUV_{L3\sigma}$ and the rest of the thresholding strategies.

It is also interesting to note that as the mean SUV values obtained from the various SUV thresholding strategies moved further away from $SUV_{L4\sigma}$ (mean SUV of 2.4) in either direction, the mean VRs moved consistently further away from 1. Thus, even though we did not specifically study $SUV_{L5\sigma}$ (and greater), the aforementioned trend in VRs suggests that their study may not have resulted in a better VR.

Our results also indicate that a thresholding strategy that relies on the maximum SUV ($SUV_{\%}$) is significantly inferior to SUV_{Abs} and $SUV_{L\sigma}$, in terms of mean CI. In addition, the $SUV_{\%}$ strategy generally results in smaller VRs than the other strategies. Although this is mostly consistent with the findings of Nestle *et al*. (18), it is important to note that $SUV_{40\%}$ results in mean VRs that are not significantly different than $SUV_{L4\sigma}$. Yet we think that the consistently low mean VRs, combined with the significantly poorer mean CIs, suggest that the $SUV_{\%}$ strategy is an inferior strategy to use in general when delineating the MTV in esophageal cancer. This is in contrast to the findings in the setting of lung cancer (22), which could be a consequence of methodologic differences.

Because both $SUV_{L4\sigma}$ and $SUV_{2.5}$ result in mean VRs and CIs that are closest to the ideal values, and because neither results in values that are significantly different from the other, we conclude that either strategy is equally efficacious. It is interesting to note that by either method the mean SUV threshold resulting in the highest CI and VR was between 2.4 and 2.5. Although there may be theoretical advantages to using the $SUV_{L\sigma}$ strategy, because it individualizes the threshold according to background liver uptake, at certain institutions it may be more practical to use an absolute SUV threshold of 2.5 in delineating the MTV. Our findings, along with the findings of Zhong *et al*. (21), support the use of $SUV_{2.5}$.

One may note that although the CIs of $SUV_{2.5}$ and $SUV_{L4\sigma}$ were among the highest, they were still far from ideal. Certain limitations in our study may account for this. First, we did not have access to an integrated PET/CT scanner. Our coregistrations and fusions may have introduced more uncertainty than if they had been done with an integrated scanner. Yet the mean CIs resulting from $SUV_{2.5}$ (0.454) and $SUV_{L4\sigma}$ (0.478) were similar to the mean CI (0.464) obtained by Gondi *et al*. (11), who did use an integrated scanner. This should be interpreted in the context of their thresholding strategy, which was much more subjective. Second, respiratory motion of the tumor during the much longer acquisition time of the PET scan may have decreased accuracy during coregistration. Four-dimensional CT simulation was not used in this study. Whether the comparisons of the volumes should be adjusted for these uncertainties is an area of future investigations.

Finally, because of being unable to identify an accurate reference volume for studying the length of the tumor (CT and esophagoduodenoscopy are limited in this regard), it was not possible to determine how well our thresholding strategies truly represented tumor length. Because we focused our comparisons on the radial extent of the tumor in extrapolating our results to include the longitudinal extent of the tumor, one assumes that the least metabolically active neoplastic cells within the tumor epicenter might have a similar metabolic index to the least metabolically active neoplastic cells at the longitudinal edges of the tumor. Although we believed this to be a reasonable assumption, this area warrants further study.

CONCLUSIONS

We conclude that regardless of the SUV thresholding method used (i.e., absolute or relative to liver mean), a threshold of approximately 2.5 yields the highest CI and best approximates the radial extent of the gross esophageal tumor at the level of the epicenter. These findings may ultimately aid radiation oncologists in the delineation of the entire GTV in esophageal cancer patients.

REFERENCES

1. Jemal A, Siegel R, Ward E, *et al*. Cancer statistics, 2008. *CA Cancer J Clin* 2008;58:71–96.
2. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology, esophageal cancer (v.1.2009). Fort Washington, PA: NCCN; 2009. p. 48.
3. Bradley J, Mutic S, *et al*. Carcinoma of the esophagus, target volume definition. In: Brady LW, editor. Technical basis of radiation therapy: Practical clinical applications. New York: Springer; 2008. 518.
4. Czito B, Denittis A, Willett C, *et al*. Esophageal cancer, radiation therapy techniques. In: Halperin EC, editor. Perez and Brady's principles and practice of radiation oncology. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2007. p. 1139.

5. Kato H, Kuwano H, Nakajima M, *et al*. Comparison between positron emission tomography and computed tomography in the use of the assessment of esophageal carcinoma. *Cancer* 2002;94:921–928.
6. Komaki R, Liao Z, Forster K, *et al*. Target definition and contouring in carcinoma of the lung and esophagus. *Rays* 2003; 28:225–236.
7. Gore R, Yaghmai V. Esophageal cancer, comparison of imaging modalities. In: Bragg D, Rubin P, Hricak H, editors. Oncologic Imaging. 2nd ed. Philadelphia: Saunders; 2002. p. 379.
8. Leong T, Everitt C, Yuen K, *et al*. A prospective study to evaluate the impact of FDG-PET on CT-based radiotherapy treatment planning for oesophageal cancer. *Radiother Oncol* 2006; 78:254–261.
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. Vrieze O, Haustermans K, De Wever W, *et al*. Is there a role for FGD-PET in radiotherapy planning in esophageal carcinoma? *Radiother Oncol* 2004;73:269–275.
11. Gondi V, Bradley K, Mehta M, *et al*. Impact of hybrid fluorodeoxyglucose positron-emission tomography/computed tomography on radiotherapy planning in esophageal and non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2007;67:187–195.
12. Grills IS, Yan D, Black QC, *et al*. Clinical implications of defining the gross tumor volume with combination of CT and 18FDG-positron emission tomography in non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2007;67:709–719.
13. Hong R, Halama J, Bova D, *et al*. Correlation of PET standard uptake value and CT window-level thresholds for target delineation in CT-based radiation treatment planning. *Int J Radiat Oncol Biol Phys* 2007;67:720–726.
14. Moureau-Zabotto L, Touboul E, Lerouge D, *et al*. Impact of CT and 18F-deoxyglucose positron emission tomography image fusion for conformal radiotherapy in esophageal carcinoma. *Int J Radiat Oncol Biol Phys* 2005;63:340–345.
15. Paulino AC, Johnstone PA. FDG-PET in radiotherapy treatment planning: Pandora's box? *Int J Radiat Oncol Biol Phys* 2004;59: 4–5.
16. Zheng XK, Chen LH, Wang QS, *et al*. Influence of FDG-PET on computed tomography-based radiotherapy planning for locally recurrent nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys* 2007;69:1381–1388.
17. MacManus M, Hicks R, Bayne M, *et al*. In regard to Paulino and Johnstone: Use of PET and CT imaging data in radiation therapy planning (*Int J Radiat Oncol Biol Phys* 2004;59:4-5). *Int J Radiat Oncol Biol Phys* 2004;60:1005–1006. author reply 1006.
18. Nestle U, Kremp S, Schaefer-Schuler A, *et al*. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. *J Nucl Med* 2005;46:1342–1348.
19. Mamede M, El Fakhri G, Abreu-e-Lima P, *et al*. Pre-operative estimation of esophageal tumor metabolic length in FDG-PET images with surgical pathology confirmation. *Ann Nucl Med* 2007;21:553–562.
20. Mamede M, Abreu-E-Lima P, Oliva MR, *et al*. FDG-PET/CT tumor segmentation-derived indices of metabolic activity to assess response to neoadjuvant therapy and progression-free survival in esophageal cancer: Correlation with histopathology results. *Am J Clin Oncol* 2007;30:377–388.
21. Zhong X, Yu J, Zhang B, *et al*. Using 18F-fluorodeoxyglucose positron emission tomography to estimate the length of gross tumor in patients with squamous cell carcinoma of the esophagus. *Int J Radiat Oncol Biol Phys* 2009;73:136–141.
22. Bradley J, Thorstad WL, Mutic S, *et al*. Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer. *Int J Radiat Oncol Biol Phys* 2004;59:78–86.