Post-chemoradiotherapy FDG PET with qualitative interpretation criteria for outcome stratification in esophageal squamous cell carcinoma

Yung-Cheng Huang[1], **Shau-Hsuan Li**[2], **Hung-I Lu**[3], **Chien-Chin Hsu**[1], **Yu-Ming Wang**[4], **Wei-Che Lin**[5], **Chao-Jung Chen**[6,7], **Kuo-Wei Ho**[8], **Nan-Tsing Chiu**[9]*

1 Department of Nuclear Medicine, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan, 2 Department of Hematology-Oncology, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan, 3 Department of Thoracic and Cardiovascular Surgery, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan, 4 Department of Radiation Oncology, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan, 5 Department of Diagnostic Radiology, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan, 6 Department of Nuclear Medicine, Yuan's General Hospital, Kaohsiung, Taiwan, 7 Department of Health Business Administration, Meiho University, Pingtung, Taiwan, 8 Department of Nuclear Medicine, Chang Gung Memorial Hospital, Chiayi, Taiwan, 9 Department of Nuclear Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

* ntchiu@mail.ncku.edu.tw

Abstract

Objectives

Post-chemoradiotherapy (CRT) FDG PET is a useful prognosticator of esophageal cancer. However, debate on the diverse criteria of previous publications preclude worldwide multi-center comparisons, and even a universal practice guide. We aimed to validate a simple qualitative interpretation criterion of post-CRT FDG PET for outcome stratification and compare it with other criteria.

Methods

The post-CRT FDG PET of 114 patients with esophageal squamous cell carcinoma (ESCC) were independently interpreted using a qualitative 4-point scale ($Qual_{4PS}$) that identified focal esophageal FDG uptake greater than liver uptake as residual tumor. Cohen's κ coefficient (κ) was used to measure interobserver agreement of $Qual_{4PS}$. The Kaplan-Meier method and Cox proportional hazards regression analyses were used for survival analysis. Other criteria included a different qualitative approach ($Qual_{BK}$), maximal standardized uptake values ($SUV_{max3.4}$, $SUV_{max2.5}$), relative change of SUV_{max} between pre- and post-CRT FDG PET (ΔSUV_{max}), mean standardized uptake values (SUV_{mean}), metabolic volume (MV) and total lesion glycolysis (TLG).

Results

Overall interobserver agreement on the $Qual_{4PS}$ criterion was excellent (κ: 0.95). Except the $Qual_{BK}$, $SUV_{max2.5}$, and TLG, all the other criteria were significant predictors for overall survival (OS). Multivariable analysis showed only $Qual_{4PS}$ (HR: 15.41; $P = 0.005$) and AJCC stage (HR: 2.47; $P = 0.007$) were significant independent variables. The 2-year OS rates of $Qual_{4PS}$(−) patients undergoing CRT alone (68.4%) and patients undergoing trimodality therapy (62.5%) were not significant different, but the 2-year OS rates of $Qual_{4PS}$(+) patients undergoing CRT alone (10.0%) were significantly lower than in patients undergoing trimodality therapy (42.1%).

Conclusions

The $Qual_{4PS}$ criterion is reproducible for assessing the response of ESCC to CRT, and valuable for predicting survival. It may add value to response-adapted treatment for ESCC patients, and help to decide whether surgery is warranted after CRT.

Introduction

Esophageal cancer is the sixth leading cause of cancer-related mortality worldwide, and the 5-year survival rate rarely exceeds 40% [1]. Most patients with esophageal cancer have advanced disease at the initial diagnosis, and are treated with neoadjuvant chemoradiotherapy (CRT) as the standard therapy [2]. A robust stratification of patient responses to CRT based on non-invasive tools has not yet been well developed. After neoadjuvant treatment, neither clinical parameters nor endoscopic ultrasonography or CT scans can reliably predict outcome. The post-CRT FDG PET, however, has emerged as a promising predictor of long-term survival, and it can be used to tailor individualized treatment for poor responders after neoadjuvant treatment [3]. Patients whose FDG PET results were a complete response might not benefit from added resection given their excellent outcomes without resection [4]. One study [5] reported that the pooled hazard ratio (HR) for a complete metabolic response (CMR) versus no response for OS was 0.51 (95% confidence interval [CI], 0.40–0.64) and for disease-free survival was 0.47 (95% CI, 0.38–0.57), respectively. Despite its utility for predicting outcomes, the lack of uniform and reliable criteria for post-CRT FDG PET interpretation appears to be the major drawback to using the reported criteria universally. Methods to improve the predictive value of PET include a qualitative approach, e.g., comparing the tumors with healthy surrounding tissue [6]; and quantitative approaches, e.g., comparing standardized uptake values (SUVs) with reported optimum SUV cut-off values, which vary from 2.5 to 4.5 [4, 7, 8], or comparing the relative reduction in SUV between pre- and post-CRT FDG PET (ΔSUV) with reported optimum cut-off values, which vary from 35% to 70% [9–11]. Wide ranges of sensitivities and specificities have been reported. The variations appear to depend upon the different sets of criteria—which are a matter of ongoing debate—used for FDG PET interpretation. Using a qualitative interpretative criterion for response assessment of FDG PET is well established and internationally recognized as the standard of care in FDG-avid lymphoma (referred to as Deauville criteria) [12] and useful in other malignancies including head and neck cancer [13], lung cancer [14] and cervical cancer [15]. Similar harmonization guidelines for interpretive criteria are needed for esophageal cancer to compare results from different studies, to perform multicenter trials, and to assist clinical practice in different sites. We developed a simple,

qualitative, interpretive criterion of FDG PET to assess esophageal cancer therapy. We validated its reader reproducibility, determined its value for predicting survival, and compared it with other visual-based and quantitative SUV-based assessment criteria.

Materials and methods

Inclusion and exclusion criteria

Inclusion criteria for the study were (a) histopathology-confirmed ESCC between January 2011 through December 2014, and (b) having undergone a post-therapy assessment FDG PET after the patient had completed CRT at our hospital. Exclusion criteria were (a) prior treatment for ESCC, (b) a history of other malignancies, or (c) post-therapy assessment FDG PET done more than 6 months after the patient had completed CRT. The Institutional Review Board approved this retrospective study (IRB #: 201700267B0).

Treatment and follow-up

The CRT consisted primarily of two cycles of 5-fluorouracil/cisplatin-based chemotherapy and thoracic radiation (42–66 Gy). Trimodality therapy included a post-CRT esophagectomy, which was usually scheduled 2–4 months after the patient had completed CRT. The 7th American Joint Committee on Cancer (AJCC) staging system was used to evaluate all patients, and all were followed-up until September 2016 or until death.

FDG PET imaging and analysis

FDG PET scans (Discovery ST PET/CT system; GE Healthcare, Waukesha, WI, USA) were principally begun one hour after the patients, who had fasted for at least 6 hours, had been injected with 370–555 MBq of FDG. Unenhanced CT scans were acquired first for attenuation correction and imaging fusion, and then PET scans (5 min/bed) from the skull to the mid-thigh were done. The PET images were reconstructed to a resolution of $5.47 \times 5.47 \times 3.27$ mm using an ordered subsets expectation maximization algorithm. The reconstructed images were displayed in transaxial, coronal, and sagittal planes, and as a maximum intensity projection for interpretation.

For each PET dataset, the SUV_{max} was defined as the highest SUV within hypermetabolic tumor boundaries. The SUV_{max} reduction rate, i.e., the percentage reduction of the primary tumors' SUV_{max} from pre-CRT FDG PET to post-CRT FDG PET, was calculated using the formula:

$$\Delta SUV_{max}(\%) = \frac{(pre - CRT\ SUV_{max} - post - CRT\ SUV_{max})}{pre - CRT\ SUV_{max}} \times 100$$

Qualitatively, the post-CRT FDG PET was scored using the $Qual_{4PS}$ qualitative 4-point scale on the esophageal tumor.

- Score 1: No detectable focal uptake.

- Score 2: Focal FDG uptake greater than that in the surrounding tissue or in mediastinal blood pool, but not greater than that of the liver.

- Score 3: Diffuse FDG uptake greater than that in the mediastinal blood pool up to marginally greater than that of the liver.

- Score 4: Focal FDG uptake substantially greater than that of liver.

Based on the Qual$_{4PS}$, the post-CRT FDG PET tumor assessments were grouped as positive (+) or negative (−). Scores 1 (CMR), 2 (likely CMR), and 3 (likely post-radiation inflammation) were considered negative for a residual tumor [Qual$_{4PS}$(−)]. Score 4 (likely viable tumor) was considered positive for a residual tumor [Qual$_{4PS}$(+)]. Representative examples are shown in Fig 1.
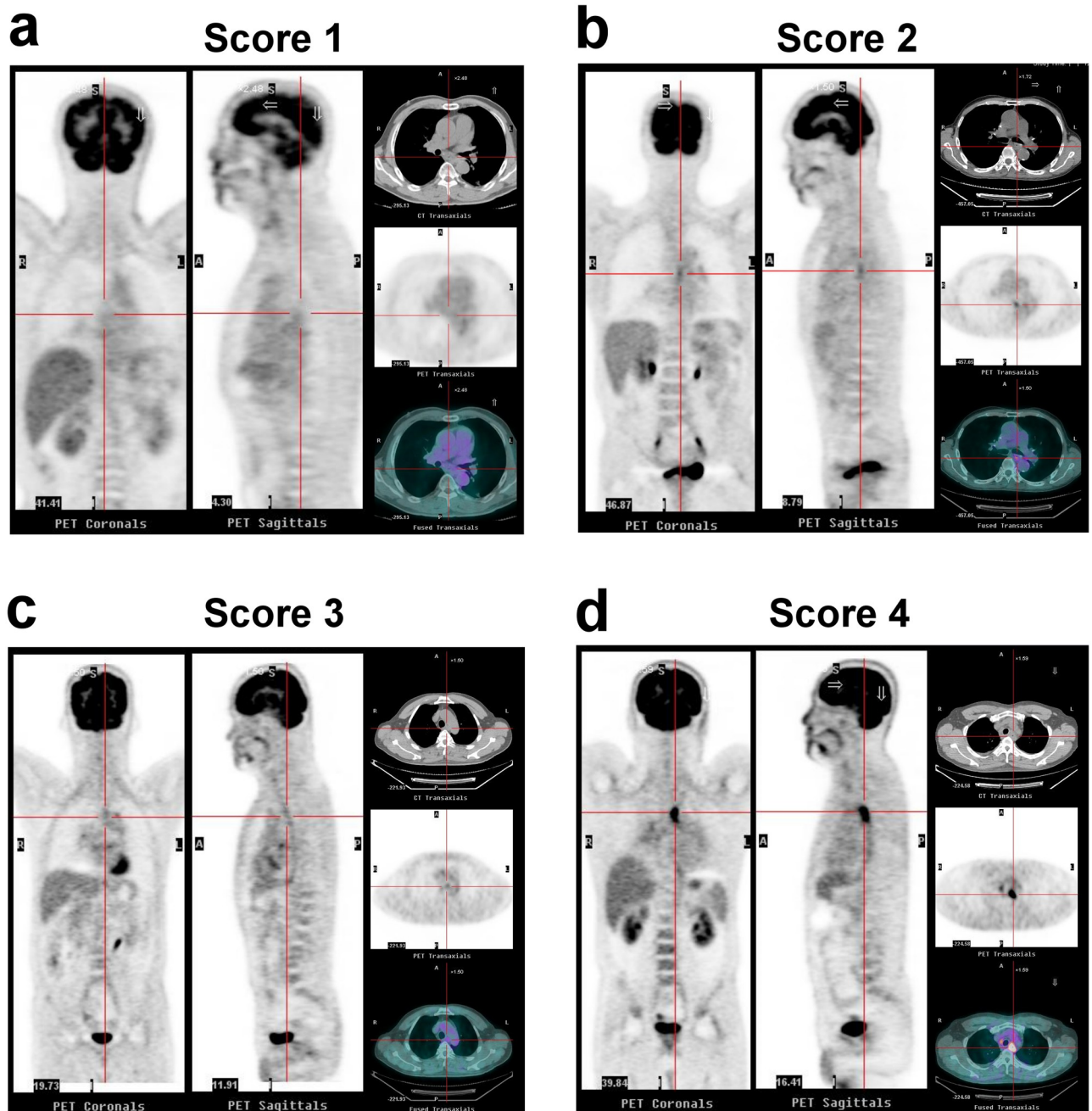

Fig 1. FDG PET/CT transaxial and sagittal images of four representative patients with Qual$_{4PS}$ scores of 1–4. (a) Score 1: no detectable focal uptake; (b) Score 2: focal FDG uptake greater than that in the surrounding tissue or in the mediastinal blood pool, but not greater than that of the liver; (c) Score 3: diffuse FDG uptake greater than that in the mediastinal blood pool up to marginally greater than that of the liver, and suggestive of esophagitis; (d) Score 4: focal FDG uptake substantially greater than that of liver.

The FDG PET/CT images were retrieved from a picture archiving and communication system and were read by five experienced—25 years (NTC), 11 years (CCH), 10 years (YCH), 8 years (CJC), and 5 years (KWH)—board-certified nuclear medicine physicians at four hospitals. Blinded to patient histories and outcomes, the reviewers scored the scans independently to determine whether the reporting system would be reproducible between observers across different institutions. The final consensus result of negative or positive for a residual tumor was assigned at least 3 reviewers achieved a common agreement.

To compare the predictive value of each set of criteria, eight other criteria were also used. $Qual_{BK}$ used the same 4-point scale, but the cut-off level was changed to surrounding background uptake [6] to make Scores 1 and 3 negative and to make Scores 2 and 4 positive. $SUV_{max3.4}$ [16] and $SUV_{max2.5}$ [7] were different SUV_{max} cut-offs in the primary tumor uptake on post-CRT FDG PET scans. $\Delta SUV_{max71.6\%}$ and $\Delta SUV_{max50\%}$ [11] used relative reduction of SUV_{max} cut-off values in the primary tumor between pre- and post-CRT FDG PET scans. The mean standardized uptake values (SUV_{mean}), metabolic volume (MV), and total lesion glycolysis (TLG = $SUV_{mean} \times MV$) were also extracted for each primary lesion. The cut-off value of ΔSUV_{max} 71.6%, SUV_{mean} 2.4, MV 2.2, and TLG 4.99 was the median data of the primary tumor in our patients. The 9 criteria stratified patients into good-responders and poor-responders.

Association between survival outcome and FDG PET/CT imaging with negative [($Qual_{4PS}(-)$] or positive [($Qual_{4PS}(+)$] for residual primary tumor and negative (PET-CR) or positive (PET-nonCR) for malignant disease (i.e., including metastatic lesions) were analyzed to test the ability to tailor treatment for selective surgical resection. The final consensus result of $Qual_{4PS}(+)$ was documented for PET-nonCR. Unexplained FDG-avid foci in lymph nodes or distant organs were reported positive for metastases and also documented for PET-nonCR. Exceptions included mediastinal nodal tracer uptake with calcification or high attenuation (>70 household units [HU]), or characterized by symmetric low-to-intermediate intensity FDG uptake in both pulmonary hilar regions with or without extending into the subcarinal and paratracheal nodal regions. All metastatic lesions had pathologically proved or followed-up information.

Statistical analysis

Categorical variables were expressed as frequencies (%), and continuous variables as means ± SD or medians (IQR). The levels of agreement between five reviewers were analyzed using Cohen's κ. OS was defined as the period from the pathologically verified ESCC to the date of the last follow-up or death of the patient from any cause. The Kaplan-Meier method was used for survival analysis, and the difference between survival curves was analyzed using a log-rank test. Univariable and multivariable Cox proportional hazards regression analyses were used to identify independent predictors of OS. SPSS 17 for Windows (SPSS Inc., Chicago, IL, USA) was used for all statistical analyses. Significance was set at $P < 0.05$.

Results

Patient characteristics

One hundred fourteen patients (mean age: 55.2 years; range: 32–80; 3 women and 111 men) were included in the study. The median follow-up was 33.2 months for living patients (range: 20.3–69.2 months). Most patients had an Eastern Cooperative Oncology Group (ECOG) performance status score of 1 (n = 98 [86.0%]), and most were in AJCC stage III (n = 93 [81.6%]). The mean SUV_{max} of residual tumor uptake on post-CRT FDG PET was 3.8 ± 2.4. In 68 patients (60%) with available pre-CRT FDG PET scans, the mean SUV_{max} of the pre-treatment

tumor was 12.8 ± 6.7, and the evaluable median ΔSUV_{max} was 71.6% (IQR: 47.9–82.7). Forty-three patients (37.7%) underwent trimodality therapy, and 71 (62.3%) underwent CRT alone (dCRT), including 10 with a salvage esophagectomy between 190 and 595 days post-CRT. The median interval between the date of the post-CRT FDG PET and the esophagectomy was 27 days (IQR: 21–37) in the trimodality group. The time between the injection of FDG and the acquisition of PET images was 58.7 ± 6.6 minutes (range: 43–80 minutes). The demographic features of the patients are summarized in Table 1.

Agreement among reviewers

In the 114 post-CRT FDG PET scans, the rates of residual tumors categorized as positive (Score 4) by the five reviewers ranged from 33.3% to 35.1%, and was 34.2% in the final consensus. The agreement of $Qual_{4PS}$ between paired reviewers for negative versus positive results was "excellent" (Cohen's κ: 0.923–0.961, Table 2). The overall reviewer agreement measured using Randolph's free marginal multirater kappa was 0.95 (95%CI: 0.91–0.99). Discordant classification occurred in only 6 patients (5.3%, 3 for 3:2 and 3 for 4:1) and their tumor SUV_{max} ranged from 2.9 to 3.8. Nineteen of all study patients (16.7%) had an SUV_{max} between 2.9 and 3.8, and the discordant classification accounts for 6 of these 19 tumors (31.6%) (Fig 2).

Table 1. Demographic and clinical characteristics of patients.

Characteristic	
Age, years, mean ± SD	55.2 ± 7.9
Male sex [a]	111 (97.4%)
ECOG performance status [a]	
0	3 (2.6%)
1	98 (86.0%)
2	13 (11.4%)
AJCC 7th stage [a]	
II	6 (5.3%)
III	93 (81.6%)
IV	15 (13.2%)
Radiation dose [a]	
≤ 50 Gy	42 (36.8%)
> 50 Gy	72 (63.2%)
Post-CRT PET/CT parameter	
$Qual_{4PS}$ (+) [a]	39 (34.2%)
SUV_{max} [b]	2.8 (2.3–4.4)
ΔSUV_{max} (%) [b]	71.6 (47.9–82.7)
SUV_{mean} [b]	2.4 (1.9–3.0)
MV [b]	2.2 (1.2–4.4)
TLG [b]	4.99 (2.53–13.67)
CRT to PET/CT interval, days [b]	41 (33–53)

SD: standard deviation; dCRT: treated with CRT alone; Trimodality: CRT plus a post-CRT esophagectomy; ECOG: Eastern Cooperative Oncology Group; AJCC: American Joint Committee on Cancer; $Qual_{4PS}$: qualitative 4-point scale with liver uptake as cut-off; SUV_{max}: maximal standardized uptake value; ΔSUV_{max}: percentage reduction of SUV_{max} of the primary tumor from pre- to post-CRT PET/CT; SUV_{mean}: mean standardized uptake value; MV: metabolic volume; TLG: total lesion glycolysis.

[a] Data are n (%);

[b] Data are median (IQR).

https://doi.org/10.1371/journal.pone.0210055.t001

Table 2. Agreement between pairs of reviewers with respect to positive versus negative PET/CT by Qual$_{4PS}$ using Cohen's k.

	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	Reviewer 5
Reviewer 1	1	0.942	0.923	0.943	0.961
Reviewer 2	0.942	1	0.942	0.923	0.941
Reviewer 3	0.923	0.942	1	0.943	0.961
Reviewer 4	0.943	0.923	0.943	1	0.942
Reviewer 5	0.961	0.941	0.961	0.942	1

Qual$_{4PS}$: qualitative 4-point scale with liver uptake as cut-off

https://doi.org/10.1371/journal.pone.0210055.t002

Outcomes

The median OS was 22.4 months and the 2-year OS rate was 48.1%. The 1- and 2-month post-operative mortality rates were 9.3% (4/43) and 14.0% (6/43), respectively. The Kaplan-Meier survival method and Cox proportional hazards regression analyses showed significant differences in OS between patients classified according to post-CRT FDG PET using all but the Qual$_{BK}$, SUV$_{max2.5}$, and TLG$_{4.99}$ criteria (Figs 3 and 4). Univariable Cox regression analysis identified AJCC stage, and post-CRT FDG PET using the Qual$_{4PS}$, SUV$_{max3.4}$, ΔSUV$_{max71.6\%}$, ΔSUV$_{max50\%}$, SUV$_{mean2.4}$, and MV$_{2.2}$, cut-offs as significant predictors of OS (Table 3). Multivariable analysis identified only the AJCC stage (HR = 2.47; P = 0.007) and the post-CRT FDG PET using the Qual$_{4PS}$ cut-off (HR = 15.41; P = 0.005) as significant independent variables correlated with OS.

Based on FDG PET classification according to the Qual$_{4PS}$ interpretation criteria [Qual$_{4PS}$(−) or Qual$_{4PS}$(+)] and therapeutic management (dCRT or trimodality), our patients could be divided into four distinct subgroups with different OS rate. The 2-year OS rates were 68.4% for the Qual$_{4PS}$(−)/dCRT group, 62.5% for the Qual$_{4PS}$(−)/trimodality group, 42.1% for the Qual$_{4PS}$(+)/trimodality group, and 10.0% for the Qual$_{4PS}$(+)/dCRT group. The Qual$_{4PS}$(−)/dCRT and Qual$_{4PS}$(−)/trimodality had equivalent OS rates. The Qual$_{4PS}$(−)/trimodality group had a nonsignificantly higher survival rate than did the Qual$_{4PS}$(+)/trimodality group. The Qual$_{4PS}$(+)/dCRT group had a significantly lower survival rate than did the other three groups (Fig 5a). These data indicated that dCRT or trimodality resulted in no significant difference of 2-year OS for patients with Qual$_{4PS}$(−) while trimodality might be a better choice for patients with Qual$_{4PS}$(+). Using PET-CR based on Qual$_{4PS}$ for FDG PET scan to subgrouping patients, the 2-year OS rates were 74.8% for the PET-CR/dCRT group, 65.2% for the PET-CR/trimodality group, 40.0% for the PET-nonCR/trimodality group, and 14.8% for the PET-nonCR/dCRT group (Fig 5b). After excluding patients with stage IV, similar differences in survival between subgroups were also obtained (S1 Fig).

Discussion

We found that using the proposed Qual$_{4PS}$ qualitative interpretation criterion for post-CRT FDG PET to assess treatment response of ESCC provided good predictive value for survival outcome and yielded excellent interobserver agreement between reviewers from different hospitals. Additionally, it might offer a guide for deciding on post-CRT surgery. FDG PET has been evaluated to optimize monitoring therapeutic response of ESCC and other malignancies. Widely available and easy-to-use SUV is the method of choice in most studies. However, the reliability of SUV measurement is affected by many factors, such as inter-scanner variability, calibration errors, image acquisition and reconstruction parameters, attenuation correction, scatter correction, respiratory motion, and the partial volume effect, all of which make proper
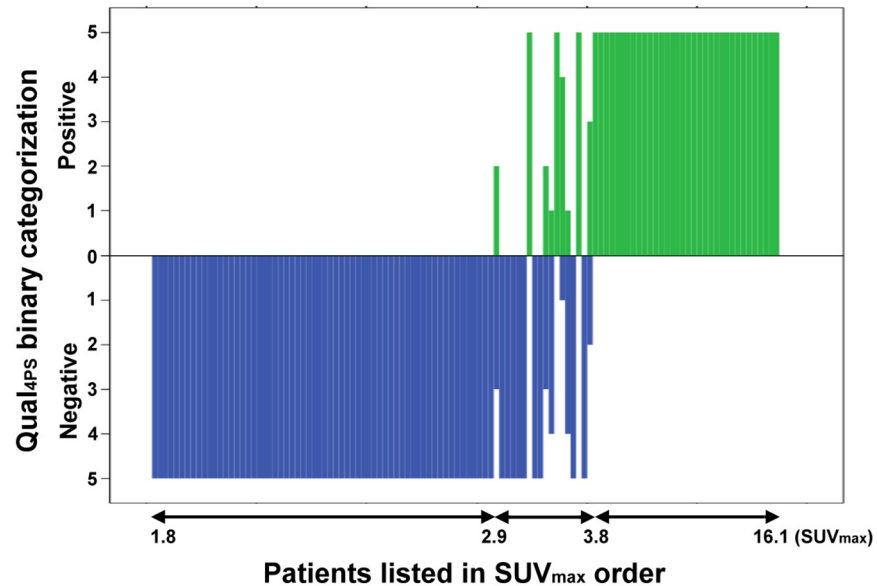
Fig 2. Patients listed by order of SUV$_{max}$ plots against the number of binary categorization (positive versus negative) by five reviewers using the Qual$_{4PS}$ criterion for post-CRT FDG PET interpretation.

https://doi.org/10.1371/journal.pone.0210055.g002

comparisons between different cohorts problematic [17]. One aim of this study was to determine whether a qualitative scoring system is practical and sufficiently robust to enable standardization of reporting across different hospitals. The major problems of a qualitative visual interpretation are the necessity of suitable criteria for interpretation and of reproducibility between different observers. Using the liver cut-off as a reference target seemed appropriate for post-therapy FDG PET interpretation. After they had studied a serial example of the scoring system (Fig 1), five reviewers from four different hospitals showed excellent agreement (Randolph's kappa: 0.95).

Various analytic models have been published for predicting outcome and optimal discrimination between responders and nonresponders by defining (a) a cut-off level for the residual tumor FDG uptake on the post-therapy scan, or (b) a percentage decrease of SUV level between pre- and post-therapy scans. There is, however, no consensus about which post-therapy FDG PET cut-off criteria are the best predictors of the outcomes of ESCC patients or which most accurately identify patients who benefitted from surgery. Jeong et al. [6] qualitatively defined PET-CR as a decreased FDG uptake to a level indistinguishable from that of the surrounding normal tissue. Moreover, SUV$_{max}$ levels ≤ 3.0 [4] and < 2.5 [7] have been used quantitatively for PET-CR. All of them verified that post-therapy PET-CR was a significant independent predictor of improved outcomes for CRT. There has been growing interest in using ΔSUV as a robust method of measuring metabolic response for predicting good and poor outcomes [18]. A large part of these studies chose the median value of their cohort data for the cut-off. Thus, a wide range of SUV reduction cut-offs, from 35% to 70%, have been reported [9–11]. The variance suggests a lack of standardization and might be explained by factors such as the spectrum of disease severity, and differences in the clinical features and therapy of each selected patient group. In this study, we also tested (a) the adjusted cut-off of SUV$_{max3.4}$, because of its optimal ability to detect post-CRT viable residual tumors in our institution [16], and (b) the adjusted cut-off of ΔSUV$_{max}$ 71.6%, which was the median reduction value in this cohort. As expected, the adjusted cut-off of SUV$_{max3.4}$ was better than the reported
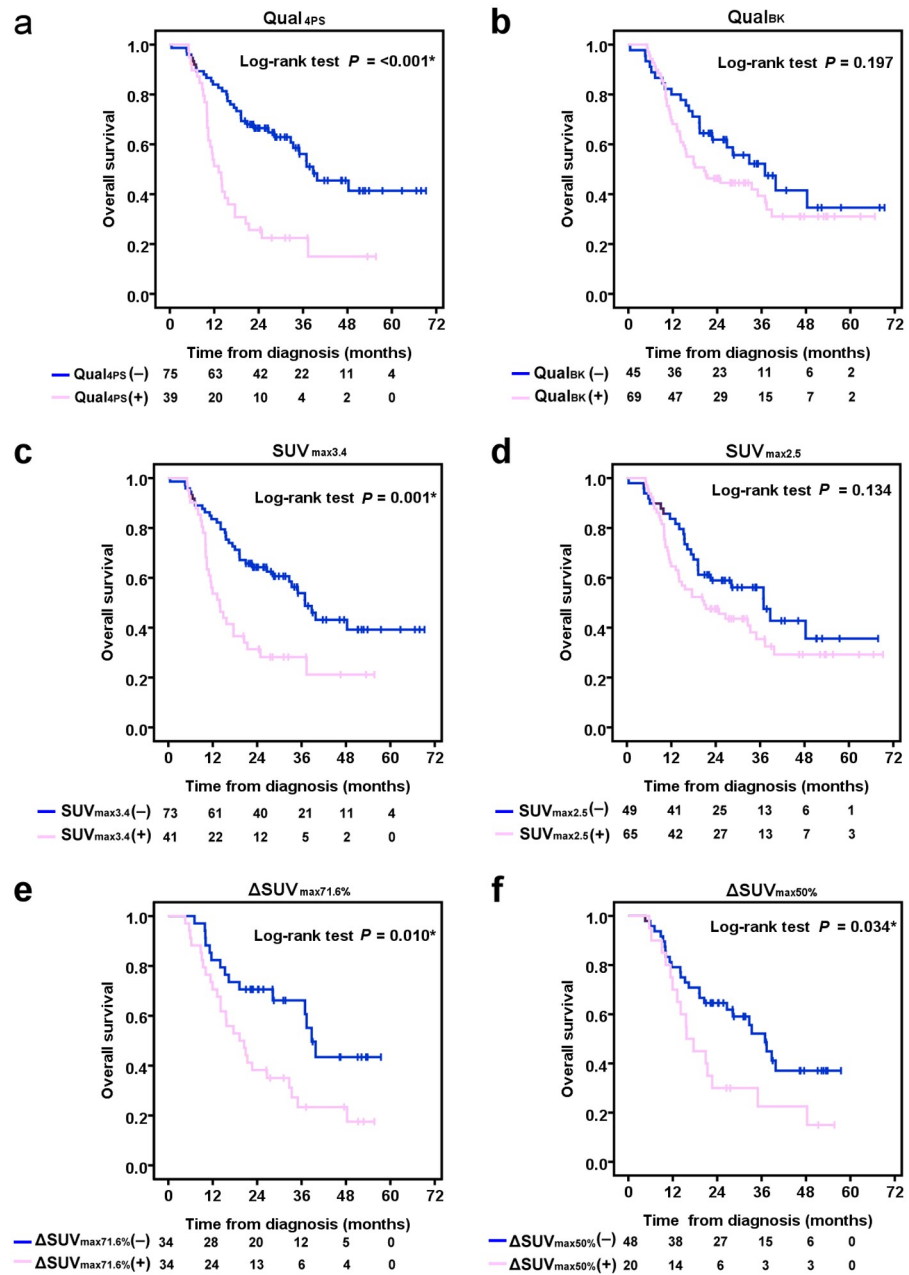
Fig 3. Kaplan-Meier plots of overall survival based on post-CRT FDG PET scans using (a) Qual$_{4PS}$, (b) Qual$_{BK}$, (c) SUV$_{max3.4}$, (d) SUV$_{max2.5}$, (e) ΔSUV$_{max71.6\%}$, and (f) ΔSUV$_{max50\%}$ cut-offs.

https://doi.org/10.1371/journal.pone.0210055.g003

SUV$_{max2.5}$, and the adjusted cut-off of ΔSUV$_{max71.6\%}$ was better than the reported ΔSUV$_{max50\%}$ for predicting good and poor outcomes.

An optimal treatment strategy should balance improved survival with minimized therapy-related morbidity, mortality, and quality-of-life deterioration. The necessity of surgical resection after CRT remains controversial. We found equivalent survival for patients in the CRT-alone and trimodality groups, which was consistent with the results of other randomized trials in which most patients had ESCC [19, 20]. Post-CRT esophagectomies are associated with an
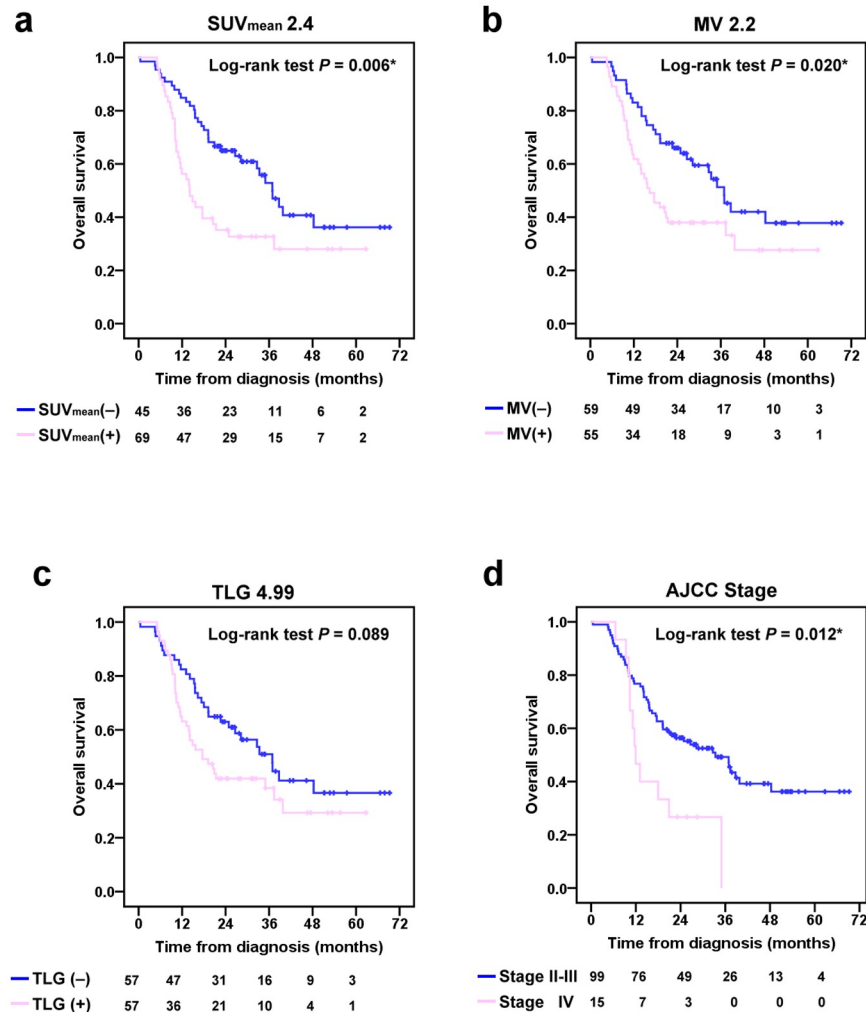
Fig 4. Kaplan-Meier plots of overall survival based on post-CRT FDG PET scans using (a) $SUV_{mean2.4}$, (b) $MV_{2.2}$, (c) $TLG_{4.99}$ cut-offs, and (d) AJCC stage.

https://doi.org/10.1371/journal.pone.0210055.g004

approximately 50% postoperative morbidity rate and a 10% postoperative mortality rate [21–23]. As in this cohort, the 1- and 2-month postoperative mortality rates were 9.3% and 14.0%, respectively. A non-invasive surrogate marker after CRT is needed to indicate that additional surgery can be delayed, or even omitted, or that is can be requested. An endoscopic biopsy for pathologic responses might not be the best predictor of outcomes after CRT in esophageal cancer; the association of PET-CR with outcomes is believed to be more clinically relevant [24, 25]. Retrospective studies which evaluated the potential of PET response-adapted strategy to identify patients for whom surgery might be avoided reported that the OS of PET-CR patients treated with CRT alone were equivalent to those treated with trimodality therapy [4, 26]. In a prospective multicenter study of 43 patients, tailoring treatment based on post-CRT FDG PET scans for selective surgical resection showed promising efficacy [27]. In our study, the additional post-CRT esophagectomy significantly improved the OS of post-CRT $Qual_{4PS}$(+) patients, but it did not significantly improve the OS of post-CRT $Qual_{4PS}$(−) patients. After CRT, using FDG PET with the $Qual_{4PS}$ interpretation criterion might be useful for determining the need for additional surgery. Large randomized multicenter studies to further evaluate

Table 3. Prognostic factors for overall survival by univariable and multivariable Cox proportional hazards regression analysis.

Characteristic	Univariable					Multivariable			
						n = 68[a]		n = 114	
	1 years (%)	2 years (%)	MCLR *P*	HR (95%CI)	*P*	HR (95%CI)	*P*	HR (95%CI)	*P*
Age									
< 60 years (n = 82)	72.0	51.1		1					
≥ 60 years (n = 32)	75.0	56.1	0.791	1.07 (0.63–1.82)	0.791				
Gender									
Female (n = 3)	100	66.7		1					
Male (n = 111)	72.1	52.1	0.368	2.41 (0.33–17.36)	0.384				
ECOG score									
0 (n = 3)	100	66.7		1	0.817				
1 (n = 98)	71.4	51.9		1.24 (0.30–5.11)	0.765				
2 (n = 13)	76.9	53.8	0.816	0.98 (0.20–4.75)	0.982				
BMI									
< 23.5 (n = 77)	71.4	50.4		1					
≥ 23.5 (n = 37)	75.7	56.8	0.554	1.17 (0.69–1.98)	0.555				
AJCC 7th stage									
II-III (n = 99)	76.8	56.4		1		1		1	
IV (n = 15)	46.7	26.7	0.012	2.23 (1.18–4.23)	0.014*	2.44 (1.03–5.73)	0.042*	2.47 (1.28–4.77)	0.007*
Radiation dose									
≤ 50 Gy (n = 42)	69.0	52.2		1					
> 50 Gy (n = 72)	75.0	52.8	0.657	0.98 (0.58–1.67)	0.951				
Post-CRT PET									
• $Qual_{4PS}$(-), Score 1,2,3 (n = 75)	84.0	66.5		1		1		1	
$Qual_{4PS}$(+), Score 4 (n = 39)	51.3	25.6	<0.001	2.89 (1.76–4.73)	<0.001*	6.55 (0.55–77.46)	0.136	15.41 (2.07–114.99)	0.005*
• $Qual_{BK}$(-), Score 1 & 3 (n = 45)	80.0	61.9		1					
$Qual_{BK}$(+), Score 2 & 4 (n = 69)	68.1	46.3	0.197	1.39 (0.84–2.32)	0.199				
• SUV_{max} ≤ 3.4 (n = 73)	83.6	64.3		1		1		1	
SUV_{max} > 3.4 (n = 41)	53.7	31.4	0.001	2.30 (1.40–3.75)	0.001*	0.54 (0.09–3.29)	0.506	0.39 (0.10–1.57)	0.186
• SUV_{max} ≤ 2.5 (n = 49)	83.7	59.0		1					
SUV_{max} > 2.5 (n = 65)	64.6	47.6	0.134	1.46 (0.89–2.40)	0.136				
• ΔSUV_{max} >71.6% (n = 34)	82.4	70.6		1		1			
ΔSUV_{max} ≤ 71.6% (n = 34)	70.6	38.2	0.010	2.27 (1.20–4.30)	0.012*	1.73 (0.88–3.41)	0.111		
• ΔSUV_{max} > 50% (n = 48)	79.2	64.6		1					
ΔSUV_{max} ≤ 50% (n = 20)	70.0	30.0	0.034	1.96 (1.04–3.68)	0.037*				
• SUV_{mean} ≤ 2.4 (n = 66)	84.8	65.0		1		1		1	
SUV_{mean} > 2.4 (n = 48)	56.2	35.2	0.006	1.96 (1.21–3.19)	0.007*	0.65 (0.07–5.64)	0.693	0.42 (0.09–1.92)	0.264
• MV ≤ 2.2 (n = 59)	83.1	66.0		1		1		1	
MV > 2.2 (n = 55)	61.8	37.9	0.020	1.77 (1.09–2.88)	0.021*	0.78 (0.34–1.80)	0.565	1.50 (0.81–2.81)	0.200
• TLG ≤ 4.99 (n = 57)	82.5	63.0		1					
TLG > 4.99 (n = 57)	63.2	41.9	0.089	1.52 (0.93–2.47)	0.091				
Treatment protocol									
dCRT (n = 71)	73.2	52.0		1					
Trimodality (n = 43)	72.1	53.4	0.946	0.98 (0.60–1.62)	0.946				

MCLR: Mantel-Cox log-rank; HR: hazard ratio; CI: confidence interval; ECOG: Eastern Coope0.946rative Oncology Group; BMI: body mass index; AJCC: American Joint Committee on Cancer; $Qual_{4PS}$: qualitative 4-point scale with liver uptake as cut-off; $Qual_{BK}$: qualitative 4-point scale with surrounding background uptake as cut-off; SUV_{max}: maximal standardized uptake value; ΔSUV_{max}: percentage reduction of SUV_{max} of the primary tumor from pre- to post-CRT PET; SUV_{mean}: mean standardized uptake value; MV: metabolic volume; TLG: total lesion glycolysis; dCRT: treated with CRT alone; Trimodality: CRT plus a post-CRT esophagectomy.
[a] Sixty-eight of total 114 patients had pre-CRT FDG PET scans and evaluable ΔSUV_{max}.
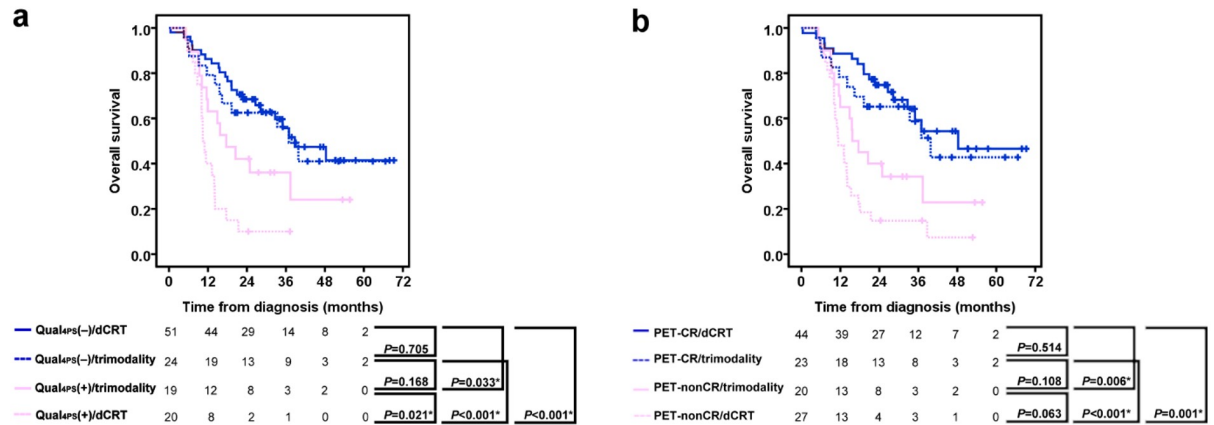
Fig 5. Kaplan-Meier plots of overall survival for subgrouping patients according to therapeutic management: dCRT versus trimodality and post-CRT FDG PET results based on (a) Qual$_{4PS}$ and (b) PET-CR.

https://doi.org/10.1371/journal.pone.0210055.g005

this organ-preserving approach of individualized therapy are still required, however. We believe that the harmonious interpretation criterion "Qual$_{4PS}$" presented here is suitable for post-CRT FDG PET scans of future multicenter trials to identify patients who benefit from CRT and, therefore, have a favorable outcome.

The study was retrospective and thus prone to a selection bias. Our results are not sufficient enough to change routine clinical practice for all esophageal cancer and should be interpreted cautiously. Six of the 19 patients (31.6%) with an SUV$_{max}$ between 2.9 and 3.8 were given discordant classifications by the reviewers. For tumor uptake within the relatively challenging SUV$_{max}$ range, we recommend that multiple reviewers be required for reaching a comprehensive consensus. It is necessary to include prospective trials that evaluate FDG PET response based on the Qual$_{4PS}$ criterion to predict outcomes of esophageal cancer and that is embedded in a randomized treatment algorithm.

Conclusions

The proposed Qual$_{4PS}$ interpretation criterion of FDG PET as therapy assessment for ESCC has excellent interobserver agreement and provides good predictive value for survival outcome. It is comparable to, and even better than quantitative criteria with different cut-offs. It can provide important information about which patients will benefit from an esophagectomy after CRT.

Supporting information

S1 Fig. The Kaplan-Meier plots of overall survival for subgrouping patients without stage IV disease according to therapeutic management: dCRT versus trimodality and post-CRT FDG PET results based on (a) Qual$_{4PS}$ and (b) PET-CR.
(TIF)

Author Contributions

Conceptualization: Yung-Cheng Huang, Nan-Tsing Chiu.

Data curation: Yung-Cheng Huang, Hung-I Lu.

References

1. Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. Lancet. 2013; 381 (9864):400–12. https://doi.org/10.1016/S0140-6736(12)60643-6 PMID: 23374478.

2. Mariette C, Piessen G, Triboulet JP. Therapeutic strategies in oesophageal carcinoma: role of surgery and other modalities. Lancet Oncol. 2007; 8(6):545–53. https://doi.org/10.1016/S1470-2045(07)70172-9 PMID: 17540306.

3. Westerterp M, van Westreenen HL, Reitsma JB, Hoekstra OS, Stoker J, Fockens P, et al. Esophageal cancer: CT, endoscopic US, and FDG PET for assessment of response to neoadjuvant therapy—systematic review. Radiology. 2005; 236(3):841–51. https://doi.org/10.1148/radiol.2363041042 PMID: 16118165.

4. Monjazeb AM, Riedlinger G, Aklilu M, Geisinger KR, Mishra G, Isom S, et al. Outcomes of patients with esophageal cancer staged with [(1)(8)F]fluorodeoxyglucose positron emission tomography (FDG-PET): can postchemoradiotherapy FDG-PET predict the utility of resection? J Clin Oncol. 2010; 28 (31):4714–21. https://doi.org/10.1200/JCO.2010.30.7702 PMID: 20876421.

5. Schollaert P, Crott R, Bertrand C, D'Hondt L, Borght TV, Krug B. A systematic review of the predictive value of (18)FDG-PET in esophageal and esophagogastric junction cancer after neoadjuvant chemoradiation on the survival outcome stratification. J Gastrointest Surg. 2014; 18(5):894–905. https://doi.org/10.1007/s11605-014-2488-2 PMID: 24638928.

6. Jeong Y, Kim JH, Kim SB, Yoon DH, Park SI, Kim YH, et al. Role of surgical resection in complete responders on FDG-PET after chemoradiotherapy for locally advanced esophageal squamous cell carcinoma. J Surg Oncol. 2014; 109(5):472–7. https://doi.org/10.1002/jso.23514 PMID: 24301552.

7. Higuchi I, Yasuda T, Yano M, Doki Y, Miyata H, Tatsumi M, et al. Lack of fludeoxyglucose F 18 uptake in posttreatment positron emission tomography as a significant predictor of survival after subsequent surgery in multimodality treatment for patients with locally advanced esophageal squamous cell carcinoma. J Thorac Cardiovasc Surg. 2008; 136(1):205–12. https://doi.org/10.1016/j.jtcvs.2008.02.016 PMID: 18603077.

8. Patnana SV, Murthy SB, Xiao L, Rohren E, Hofstetter WL, Swisher SG, et al. Critical role of surgery in patients with gastroesophageal carcinoma with a poor prognosis after chemoradiation as defined by positron emission tomography. Cancer. 2010; 116(19):4487–94. https://doi.org/10.1002/cncr.25431 PMID: 20629031

9. Ilson DH, Minsky BD, Ku GY, Rusch V, Rizk N, Shah M, et al. Phase 2 trial of induction and concurrent chemoradiotherapy with weekly irinotecan and cisplatin followed by surgery for esophageal cancer. Cancer. 2012; 118(11):2820–7. https://doi.org/10.1002/cncr.26591 PMID: 21990000.

10. Makino T, Miyata H, Yamasaki M, Fujiwara Y, Takiguchi S, Nakajima K, et al. Utility of response evaluation to neo-adjuvant chemotherapy by (18)F-fluorodeoxyglucose-positron emission tomography in locally advanced esophageal squamous cell carcinoma. Surgery. 2010; 148(5):908–18. https://doi.org/10.1016/j.surg.2010.02.016 PMID: 20378140.

11. Abdelsalam M, Bazarbashi S, Abouzied M, Amin T, Soudy H, Rahal M, et al. Whole body 18F-FDG PET predicts progression free and overall survival in squamous cell carcinoma of the esophagus: results of a prospective trial. Hematol Oncol Stem Cell Ther. 2010; 3(4):179–84. PMID: 21150237.

12. Barrington SF, Mikhaeel NG, Kostakoglu L, Meignan M, Hutchings M, Mueller SP, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the International Conference on Malignant Lymphomas Imaging Working Group. J Clin Oncol. 2014; 32(27):3048–58. https://doi.org/10.1200/JCO.2013.53.5229 PMID: 25113771.

13. Marcus C, Ciarallo A, Tahari AK, Mena E, Koch W, Wahl RL, et al. Head and neck PET/CT: therapy response interpretation criteria (Hopkins Criteria)-interreader reliability, accuracy, and survival outcomes. J Nucl Med. 2014; 55(9):1411–6. https://doi.org/10.2967/jnumed.113.136796 PMID: 24947059.

14. Sheikhbahaei S, Mena E, Marcus C, Wray R, Taghipour M, Subramaniam RM. 18F-FDG PET/CT: Therapy Response Assessment Interpretation (Hopkins Criteria) and Survival Outcomes in Lung Cancer Patients. J Nucl Med. 2016; 57(6):855–60. https://doi.org/10.2967/jnumed.115.165480 PMID: 26837337.

15. Scarsbrook A, Vaidyanathan S, Chowdhury F, Swift S, Cooper R, Patel C. Efficacy of qualitative response assessment interpretation criteria at 18F-FDG PET-CT for predicting outcome in locally advanced cervical carcinoma treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2017; 44(4):581–8. https://doi.org/10.1007/s00259-016-3537-8 PMID: 27738729.

16. Huang YC, Lu HI, Huang SC, Hsu CC, Chiu NT, Wang YM, et al. FDG PET using SUVmax for preoperative T-staging of esophageal squamous cell carcinoma with and without neoadjuvant chemoradiotherapy. BMC Med Imaging. 2017; 17(1):1. https://doi.org/10.1186/s12880-016-0171-7 PMID: 28056868.

17. Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. AJR Am J Roentgenol. 2010; 195(2):310–20. https://doi.org/10.2214/AJR.10.4923 PMID: 20651185.

18. Zhu W, Xing L, Yue J, Sun X, Sun X, Zhao H, et al. Prognostic significance of SUV on PET/CT in patients with localised oesophagogastric junction cancer receiving neoadjuvant chemotherapy/chemoradiation:a systematic review and meta-analysis. Br J Radiol. 2012; 85(1017):e694–701. https://doi.org/10.1259/bjr/29946900 PMID: 22337686.

19. Bedenne L, Michel P, Bouche O, Milan C, Mariette C, Conroy T, et al. Chemoradiation followed by surgery compared with chemoradiation alone in squamous cancer of the esophagus: FFCD 9102. J Clin Oncol. 2007; 25(10):1160–8. https://doi.org/10.1200/JCO.2005.04.7118 PMID: 17401004.

20. Stahl M, Stuschke M, Lehmann N, Meyer HJ, Walz MK, Seeber S, et al. Chemoradiation with and without surgery in patients with locally advanced squamous cell carcinoma of the esophagus. J Clin Oncol. 2005; 23(10):2310–7. https://doi.org/10.1200/JCO.2005.00.034 PMID: 15800321.

21. Dimick JB, Goodney PP, Orringer MB, Birkmeyer JD. Specialty training and mortality after esophageal cancer resection. Ann Thorac Surg. 2005; 80(1):282–6. https://doi.org/10.1016/j.athoracsur.2005.01.044 PMID: 15975382.

22. Connors RC, Reuben BC, Neumayer LA, Bull DA. Comparing outcomes after transthoracic and transhiatal esophagectomy: a 5-year prospective cohort of 17,395 patients. J Am Coll Surg. 2007; 205(6):735–40. https://doi.org/10.1016/j.jamcollsurg.2007.07.001 PMID: 18035255.

23. Chang AC, Ji H, Birkmeyer NJ, Orringer MB, Birkmeyer JD. Outcomes after transhiatal and transthoracic esophagectomy for cancer. Ann Thorac Surg. 2008; 85(2):424–9. https://doi.org/10.1016/j.athoracsur.2007.10.007 PMID: 18222237.

24. Downey RJ, Akhurst T, Ilson D, Ginsberg R, Bains MS, Gonen M, et al. Whole body 18FDG-PET and the response of esophageal cancer to induction therapy: results of a prospective trial. J Clin Oncol. 2003; 21(3):428–32. https://doi.org/10.1200/JCO.2003.04.013 PMID: 12560430.

25. Blackstock AW, Aklilu M, Lovato J, Farmer MR, Mishra G, Melin SA, et al. Pathologic complete response may not represent the optimal surrogate for survival after preoperative therapy for esophageal cancer. Int J Gastrointest Cancer. 2006; 37(1):7–14. https://doi.org/10.1385/IJGC:37:1:7 PMID: 17290076.

26. Murthy SB, Patnana SV, Xiao L, Rohren E, Hofstetter WL, Swisher SG, et al. The standardized uptake value of 18-fluorodeoxyglucose positron emission tomography after chemoradiation and clinical outcome in patients with localized gastroesophageal carcinoma. Oncology. 2010; 78(5–6):316–22. https://doi.org/10.1159/000319938 PMID: 20699623.

27. Swisher SG, Moughan J, Komaki RU, Ajani JA, Wu TT, Hofstetter WL, et al. Final Results of NRG Oncology RTOG 0246: An Organ-Preserving Selective Resection Strategy in Esophageal Cancer Patients Treated with Definitive Chemoradiation. J Thorac Oncol. 2017; 12(2):368–74. https://doi.org/10.1016/j.jtho.2016.10.002 PMID: 27729298.