The Incremental Value of Subjective and Quantitative Assessment of ^{18}F-FDG PET for the Prediction of Pathologic Complete Response to Preoperative Chemoradiotherapy in Esophageal Cancer

Peter S.N. van Rossum[1,2], David V. Fried[3,4], Lifei Zhang[3], Wayne L. Hofstetter[5], Marco van Vulpen[2], Gert J. Meijer[2], Laurence E. Court[3], and Steven H. Lin[1]

[1]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas; [2]Department of Radiation Oncology, University Medical Center Utrecht, Utrecht, The Netherlands; [3]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas; [4]The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas; and [5]Department of Thoracic and Cardiovascular Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas

A reliable prediction of a pathologic complete response (pathCR) to chemoradiotherapy before surgery for esophageal cancer would enable investigators to study the feasibility and outcome of an organ-preserving strategy after chemoradiotherapy. So far no clinical parameters or diagnostic studies are able to accurately predict which patients will achieve a pathCR. The aim of this study was to determine whether subjective and quantitative assessment of baseline and postchemoradiation ^{18}F-FDG PET can improve the accuracy of predicting pathCR to preoperative chemoradiotherapy in esophageal cancer beyond clinical predictors. **Methods:** This retrospective study was approved by the institutional review board, and the need for written informed consent was waived. Clinical parameters along with subjective and quantitative parameters from baseline and postchemoradiation ^{18}F-FDG PET were derived from 217 esophageal adenocarcinoma patients who underwent chemoradiotherapy followed by surgery. The associations between these parameters and pathCR were studied in univariable and multivariable logistic regression analysis. Four prediction models were constructed and internally validated using bootstrapping to study the incremental predictive values of subjective assessment of ^{18}F-FDG PET, conventional quantitative metabolic features, and comprehensive ^{18}F-FDG PET texture/geometry features, respectively. The clinical benefit of ^{18}F-FDG PET was determined using decision-curve analysis. **Results:** A pathCR was found in 59 (27%) patients. A clinical prediction model (corrected c-index, 0.67) was improved by adding ^{18}F-FDG PET–based subjective assessment of response (corrected c-index, 0.72). This latter model was slightly improved by the addition of 1 conventional quantitative metabolic feature only (i.e., postchemoradiation total lesion glycolysis; corrected c-index, 0.73), and even more by subsequently adding 4 comprehensive ^{18}F-FDG PET texture/geometry features (corrected c-index, 0.77). However, at a decision threshold of 0.9 or higher, representing a clinically relevant predictive value for pathCR at which one may be willing to omit surgery, there was no clear incremental value. **Conclusion:** Subjective and quantitative assessment of ^{18}F-FDG PET provides statistical incremental value for predicting pathCR after preoperative chemoradiotherapy in esophageal cancer. However, the discriminatory improvement beyond clinical predictors does not translate into a clinically relevant benefit that could change decision making.

Key Words: ^{18}F-FDG PET; positron emission tomography; texture features; esophageal cancer; treatment response

Preoperative chemoradiotherapy followed by surgery is increasingly applied as standard treatment with curative intent for patients with resectable nonmetastatic esophageal cancer (1). A pathologic complete response (pathCR) to chemoradiotherapy is observed in approximately 25%–30% of esophageal cancer patients (1–4). Many studies have reported that pathCR is associated with favorable overall survival rates (2–4). Whether or not surgery can be safely omitted in patients who achieve a pathCR is an important focus of research, but determining pathCR is difficult without performing surgery (5). A reliable prediction of pathCR before surgery would enable investigators to study the feasibility and outcome of an organ-preserving strategy after chemoradiotherapy that includes omission of surgery and close clinical follow-up. Unfortunately, so far no clinical parameters or diagnostic studies are able to accurately predict which patients will achieve a pathCR (6–9). Therefore, seeking more powerful predictors for pathCR is of high relevance to modern personalized cancer care for which the aim is to tailor treatment to the individual patient.

^{18}F-FDG PET is a well-established imaging modality for the initial staging of esophageal cancer and restaging after preoperative chemoradiotherapy for the detection of distant (interval) metastases (10–12). The value of ^{18}F-FDG PET to predict response to preoperative chemoradiotherapy in patients with esophageal cancer has been studied extensively with wide varying methodologies

and conflicting results (6,10,13). Some investigators examined the value of a (subjective) determination of clinical response by experienced nuclear medicine physicians based on ^{18}F-FDG PET scanning after chemoradiotherapy (11,12,14). Most other ^{18}F-FDG PET studies evaluated the value of quantitative metabolic parameters such as the SUV_{mean} or SUV_{max} within a tumor for assessment of response (5,6,12,13,15). A few studies evaluated more advanced ^{18}F-FDG PET–based metabolic parameters such as metabolic tumor volume (MTV) and total lesion glycolysis (TLG) for response prediction (16–18). Unfortunately, all these methods have not shown sufficient capability of ^{18}F-FDG PET to predict response and guide clinical decision making.

More recently, ^{18}F-FDG PET image texture analysis has been proposed to characterize heterogeneity of intratumoral ^{18}F-FDG uptake, which may provide a useful representation of underlying biologic tumor characteristics (19). Texture analysis refers to a variety of computational methods that describe the relationships between the intensity of voxels and their position within an image (20). Although ^{18}F-FDG PET texture features are not routinely used in clinical evaluation of ^{18}F-FDG PET images, there is increasing recognition that a potential complementary role may exist for prediction of treatment response and prognosis in several cancers (20). The aim of this diagnostic study was to develop and internally validate multivariable prediction models to determine the incremental value of baseline and postchemoradiation ^{18}F-FDG PET scanning for predicting pathCR after chemoradiotherapy in esophageal cancer beyond clinical predictive factors. In a large cohort, the added predictive values of subjective assessment of ^{18}F-FDG PET scans, conventional quantitative metabolic parameters, and comprehensive quantitative ^{18}F-FDG PET texture/geometry features for pathCR were evaluated, respectively.

MATERIALS AND METHODS

This retrospective study was approved by the Institutional Review Board of the MD Anderson Cancer Center (MDACC), and the need for written informed consent was waived. The study was conducted in accordance with the Health Insurance Portability and Accountability Act, the checklist from the STAndards for the Reporting of Diagnostic accuracy studies statement (http://www.stard-statement.org) (21), and the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement (http://www.tripod-statement.org) (22).

Study Population

All consecutive patients with a biopsy-proven resectable adenocarcinoma of the esophagus or gastroesophageal junction and no distant metastases who underwent preoperative chemoradiotherapy followed by surgery at MDACC from January 2006 to October 2013 were extracted from a prospective database. Patients were included only if results from baseline ^{18}F-FDG PET, postchemoradiation ^{18}F-FDG PET, and postchemoradiation endoscopic biopsy were available. Exclusion criteria were nonavailability of a baseline ^{18}F-FDG PET scan acquired at MDACC, non–[18]F-FDG avidity of the tumor at baseline, Siewert type 3 gastroesophageal junction tumors, and an esophageal stent in situ at the time of ^{18}F-FDG PET scanning. In addition, patients with a time interval between completion of chemoradiotherapy and surgery of less than 5 wk or more than 14 wk—indicating urgent and salvage resections, respectively—were excluded. A detailed description of the treatment regimen and ^{18}F-FDG PET image acquisition is provided in Supplemental Appendix A (supplemental materials are available at http://jnm.snmjournals.org).

Histopathologic Assessment

The degree of histopathologic tumor regression in the resected specimen was assigned to 1 of 4 tumor regression grades (TRGs) by experienced pathologists: complete absence of residual cancer (i.e., pathCR), 1%–10% residual carcinoma, 11%–50% residual carcinoma, and greater than 50% residual carcinoma (2). PathCR (i.e., TRG 1) as opposed to any grade of residual carcinoma (i.e., TRGs 2–4) was considered the reference standard. Another commonly made distinction between TRGs 1 and 2 (i.e., 0%–10% residual carcinoma) and TRGs 3 and 4 (i.e., ≥11% residual carcinoma) was not made, because this distinction is rather arbitrary in terms of interpathologist reproducibility and overall survival (2,4,23), and the potential clinical consequences of such a distinction are unclear.

Quantitative Image Analysis

Only the primary esophageal tumors were considered for imaging analysis, as texture analysis cannot be reliably performed on small lesions (such as nodal metastases) because of the small number of voxels (24,25). The primary tumor volume was defined as the volume of interest and delineated using a semiautomatic gradient-based delineation method (26) from commercially available software (MIM Software Inc.), followed by manual editing by 1 interpreter. A rationale for using this method is provided in Supplemental Appendix B. The interpreter was masked to other clinical information and to the reference standard. Contours were extracted, and quantitative ^{18}F-FDG PET analysis was performed using the Imaging Biomarker Explorer software package (27) built in-house with commercial software (Matlab, version 8.4; The MathWorks Inc.).

Test–Retest Analysis

The authors recognized that not all ^{18}F-FDG PET features were sufficiently robust (28,29) and therefore aimed to include only features that had sufficient robustness in the prediction modeling. To fulfill this goal, ^{18}F-FDG PET scans taken outside MDACC (with different types of scanners and scan protocols) were used and compared with diagnostic ^{18}F-FDG PET scans taken within MDACC. A subgroup of 7 patients who underwent this form of double baseline scanning, with an average separation between the 2 scans of 31 d (range, 11–42 d), was identified. A single observer contoured the tumor volumes for the test–retest scans on separate occasions. In this way, both robustness and intraobserver contour variability were incorporated in test–retest analysis. For each image-derived parameter, the intraclass correlation coefficient (ICC) was calculated, using an absolute agreement definition in a 2-way mixed-effects model (30), to quantify the relatedness of the 2 scans per parameter.

Statistical Analysis

Statistical analysis was performed using SPSS 20.0 (IBM Corp.) and R 3.1.2 open-source software (http://www.R-project.org). A P value of less than 0.05 was considered statistically significant.

Model Development. A detailed description on the studied parameters, standardized preselection of variables for multivariable analysis, multivariable model development, and model performance tests and validation is provided in Supplemental Appendix C. The association between clinical parameters and pathCR was studied univariably. A χ^2 test or Fisher exact test (in the case of small cell count) was used for categoric variables, and a Student t test or Mann–Whitney U test was used for parametric or nonparametric continuous clinical variables. The predictive value of each ^{18}F-FDG PET–based parameter for pathCR was quantified using univariable logistic regression analysis providing odds ratios with 95% confidence intervals. Various features were logarithmically transformed to improve on the assumption of linearity on the logit scale.

Because many potential predictors were studied in univariable analysis, a standardized preselection of variables for multivariable analysis

TABLE 1
Clinical Baseline and Treatment-Related Characteristics

Characteristic	PathCR (n = 59)	No pathCR (n = 158)	P
Male sex	54 (91.5)	148 (93.7)	0.558
Age (y)[†]	58.8 ± 12.3	60.1 ± 9.9	0.440
Body mass index (kg/m^2)[†]	29.5 ± 5.3	29.8 ± 5.2	0.632
Hypertension	29 (49.2)	90 (57.0)	0.304
Cardiac comorbidity	14 (23.7)	24 (15.2)	0.141
Diabetes mellitus	12 (20.3)	31 (19.6)	0.906
Chronic obstructive pulmonary disease	4 (6.8)	8 (5.1)	0.739
Smoking	13 (22.0)	38 (24.1)	0.755
Karnofsky performance status[†]	86.4 ± 6.9	85.5 ± 6.6	0.362
Tumor location			0.324
Middle third of esophagus	2 (3.4)	1 (0.6)	
Distal third of esophagus	52 (88.1)	143 (90.5)	
Gastroesophageal junction	5 (8.5)	14 (8.9)	
EUS-based tumor length (cm)[†]	5.0 ± 2.4	5.9 ± 2.7	0.034*
Histologic differentiation grade			0.055
Moderate	34 (57.6)	68 (43.0)	
Poor	25 (42.4)	90 (57.0)	
Signet ring cell adenocarcinoma	5 (8.5)	30 (19.0)	0.061
Clinical T stage			0.006*
cT2	14 (23.7)	15 (9.5)	
cT3	45 (76.3)	143 (90.5)	
Clinical N stage			0.450
cN0	18 (30.5)	58 (36.7)	
cN+	39 (66.1)	98 (62.0)	
Missing	2 (3.4)	2 (1.3)	
Induction chemotherapy	28 (47.5)	50 (31.6)	0.031*
Total radiation dose (Gy)			0.466
45.0	4 (6.8)	6 (3.8)	
50.4	55 (93.2)	152 (96.2)	
Radiation treatment modality			0.405
3-dimensional conformal radiation therapy	1 (1.7)	1 (0.6)	
Intensity-modulated radiotherapy	38 (64.4)	111 (70.3)	
Proton therapy	20 (33.9)	46 (29.1)	
Chemotherapy regimen			0.940
Oxaliplatin/5-fluorouracil	25 (42.4)	64 (40.5)	
Docetaxel/5-fluorouracil	25 (42.4)	67 (42.4)	
Other	9 (15.3)	27 (17.1)	
Postchemoradiation endoscopic biopsy			0.023*
No residual cancer	55 (93.2)	126 (79.7)	
Residual cancer	4 (6.8)	32 (20.3)	
Days from completion chemoradiotherapy to surgery[†]	61.5 ± 20.4	58.3 ± 19.3	0.285
Year of patient accrual			0.072
2006–2007	10 (16.9)	47 (29.7)	
2008–2010	26 (44.1)	63 (39.9)	
2011–2013	23 (39.0)	48 (30.4)	

*Significant difference between pathCR group and pathologic noncomplete response group ($P < 0.05$).
[†]Expressed as mean ± SD.
Data are numbers, with percentages in parentheses.

was performed, taking into account test–retest robustness and multicollinearity between parameters (Supplemental Appendix C). Four multivariable logistic regression models were constructed using stepwise backward elimination to study the value of clinical parameters (model 1), the added value of subjective assessment of ^{18}F-FDG PET (model 2), and the subsequent added value of conventional and comprehensive quantitative ^{18}F-FDG PET parameters (models 3 and 4, respectively), for the prediction of pathCR.

Model Performance and Validation. Model discrimination and calibration results were evaluated for all 4 models using c-indices and visual inspection of model calibration plots, respectively. Internal validation using the bootstrap method with 1,000 repetitions was performed to provide insight into potential overfitting and optimism in model performance. Bootstrapping allowed for calculation of bias-corrected c-indices of the 4 models and provided shrinkage factors that were used to adjust the estimated regression coefficients in the final 4 models for overfitting and miscalibration (31). A sensitivity analysis was performed on a subcohort of tumors with an initial MTV above 10 mL, because a recent study suggested that texture analysis may only provide valuable complementary information in tumors larger than 10 mL (24). In addition, the influence of including cases with 2-dimensional-acquisition ^{18}F-FDG PET scans (besides 3-dimensional-acquisition ^{18}F-FDG PET scans) in the model development set was determined by sensitivity analysis.

Clinical Benefit. Because traditional accuracy metrics—such as the c-index—have limited value for clinical decision making in the individual patient, the incremental clinical value of ^{18}F-FDG PET analysis was determined using decision-curve analysis (32). In this net benefit assessment, the adverse effect of falsely predicting pathCR in non-pathCR patients (e.g., false-positives) is incorporated. In fact, a personal decision threshold can be chosen according to the willingness of risking a false-positive result. In the context of predicting pathCR in esophageal cancer, a false-positive finding could result in omission of surgery in a patient with residual disease, which is potentially hazardous

(14,33), and therefore the decision threshold should be high (e.g., ~0.9). To evaluate the incremental value of a model including ^{18}F-FDG PET information, the net benefit (demonstrated on the y-axis) of the model should be higher than the model without that information. Divergence of a model's decision curve from another model's decision curve indicates difference in net benefit, but it depends on the personal decision threshold whether this is of clinical relevance. The decision curves of the 4 models used in the present study are presented together with the net benefit of making the same decision for all patients (i.e., performing surgery in all patients or omitting surgery in all patients). Additional explanation on the decision-curve analysis is provided in Supplemental Appendix C.

RESULTS

From 324 patients with an esophageal adenocarcinoma that met the prespecified inclusion criteria, 107 patients were excluded because of nonavailability of a baseline ^{18}F-FDG PET scan acquired at MDACC ($n = 48$), non–^{18}F-FDG avidity at baseline ($n = 18$), a Siewert type 3 gastroesophageal junction tumor ($n = 15$), an esophageal stent in situ at the time of ^{18}F-FDG PET scanning ($n = 1$), or a time interval between completion of chemoradiotherapy and surgery of less than 5 wk ($n = 1$) or more than 14 wk ($n = 24$). A pathCR was found in 59 (27%) of the 217 eligible patients. Patients in the pathCR group had a mean age of 59 y, and 92% ($n = 54$) of them were men, whereas patients with residual carcinoma had a mean age of 60 y and 94% of them were men (Table 1).

Univariable Analysis

Smaller endoscopic ultrasound (EUS)–based tumor length and less advanced depth of tumor infiltration (i.e., clinical stage T2 vs. T3) were significantly associated with a higher chance of pathCR ($P = 0.034$ and 0.006, respectively). The subgroup of 78 patients

TABLE 2
Univariable Analysis of Subjective and Conventional Quantitative Assessment of ^{18}F-FDG PET for Predicting pathCR

Characteristic	n	Odds ratio	95% confidence interval	P
			Univariable analysis	
Subjective assessment ^{18}F-FDG PET				0.001*
Clinical complete response	60	1.0 (ref)		
No clinical complete response	157	0.30	0.15–0.59	
Baseline SUV$_{max}$ (log)	217	0.63	0.37–1.07	0.087
Baseline SUV$_{mean}$ (log)	217	0.60	0.32–1.13	0.112
Baseline MTV (log)	217	0.85	0.57–1.26	0.408
Baseline TLG (log)	217	0.82	0.61–1.10	0.181
Postchemoradiation SUV$_{max}$ (log)	217	0.32	0.13–0.80	0.015*
Postchemoradiation SUV$_{mean}$ (log)	217	0.64	0.22–1.89	0.420
Postchemoradiation MTV (log)	217	0.34	0.21–0.53	<0.001*
Postchemoradiation TLG (log)	217	0.41	0.28–0.61	<0.001*
ΔSUV$_{max}$ (%)	217	1.00	0.99–1.01	0.701
ΔSUV$_{mean}$ (%)	217	1.01	1.00–1.02	0.146
ΔMTV(%)	217	1.00	0.99–1.00	0.142
ΔTLG (%)	217	1.00	0.99–1.00	0.301

*Significant difference between pathCR group and pathologic noncomplete response group ($P < 0.05$).
ref = reference; Δ = relative change between baseline and postchemoradiation ^{18}F-FDG PET scans.

(36%) who underwent induction chemotherapy before trimodality therapy showed a significantly higher pathCR rate than the group without induction chemotherapy (36% vs. 22%, respectively; $P = 0.031$). The postchemoradiation endoscopic biopsy result was significantly associated with pathCR ($P = 0.023$), although the predictive value of an endoscopic biopsy–based complete response for pathCR was only 30% (55/181).

Univariable analysis of subjective assessment and conventional quantitative features of ^{18}F-FDG PET for predicting pathCR are presented in Table 2. Similar to endoscopic biopsy, subjective assessment of the postchemoradiation ^{18}F-FDG PET scan was significantly associated with pathCR ($P < 0.001$), but only 27 of 60 patients (45%) with a clinical complete response on ^{18}F-FDG PET had a true pathCR. There was a trend toward a lower baseline SUV_{max} in pathologic complete responders ($P = 0.087$). Some of the conventional quantitative features on the postchemoradiation ^{18}F-FDG PET scan were significantly related to a higher

chance of pathCR including lower SUV_{max} ($P = 0.015$), lower MTV ($P < 0.001$), and lower TLG ($P < 0.001$).

Test–retest relatedness of the conventional ^{18}F-FDG PET features was good (ICCs of SUV_{max}, SUV_{mean}, MTV, and TLG were 0.86, 0.87, 0.99, and 0.96, respectively). For each comprehensive ^{18}F-FDG PET feature, the ICCs resulting from test–retest analysis along with univariable analyses for predicting pathCR are presented in Supplemental Tables 1 and 2. In general, test–retest relatedness was excellent for geometry features (median ICC, 0.92); good for first-, second-, and regional higher-order texture features (median ICC, 0.86, 0.83, and 0.85, respectively); and poor for local higher-order texture features (median ICC, 0.69).

Multivariable Analysis

Table 3 shows the finalized multivariable analysis of the 4 prediction models. In the clinical model (model 1), 4 variables remained associated with a higher chance of pathCR (EUS-based

TABLE 3
Finalized Prediction Models for pathCR Using Multivariable Logistic Regression Analysis with Stepwise Backward Elimination

Characteristic	Model 1 Odds ratio (95% confidence interval)	P	Model 2 Odds ratio (95% confidence interval)	P	Model 3 Odds ratio (95% confidence interval)	P	Model 4 Odds ratio (95% confidence interval)	P
EUS-based tumor length (log)	0.48 (0.24–0.95)	0.034*	0.50 (0.24–1.01)	0.054	0.55 (0.57–1.14)	0.107	0.46 (0.19–1.11)	0.085
Clinical T stage		0.077		0.046*		0.185		0.239
cT2	1.0 (ref)		1.0 (ref)		1.0 (ref)		1.0 (ref)	
cT3	0.45 (0.19–1.09)		0.39 (0.15–0.98)		0.53 (0.20–1.36)		0.54 (0.19–1.51)	
Induction chemotherapy		0.008*		0.012*		0.022*		0.008*
No	1.0 (ref)		1.0 (ref)		1.0 (ref)		1.0 (ref)	
Yes	2.44 (1.26–4.74)		2.40 (1.21–4.77)		2.26 (1.12–4.54)		2.80 (1.31–5.98)	
Postchemoradiotherapy endoscopic biopsy		0.035*		0.047*		0.057		0.073
No residual cancer	1.0 (ref)		1.0 (ref)		1.0 (ref)		1.0 (ref)	
Residual cancer	0.30 (0.10–0.92)		0.32 (0.10–0.98)		0.32 (0.10–1.04)		0.31 (0.08–1.12)	
Subjective assessment ^{18}F-FDG PET	Not entered	—		0.001*		0.043*		0.113
Clinical complete response			1.0 (ref)		1.0 (ref)		1.0 (ref)	
No clinical complete response			0.30 (0.15–0.59)		0.45 (0.21–0.98)		0.52 (0.23–1.17)	
Postchemoradiotherapy TLG (log)	Not entered	—	Not entered	—	0.57 (0.37–0.88)	0.011*	0.76 (0.41–1.39)	0.370
Baseline cluster shade (log)	Not entered	—	Not entered	—	Not entered	—	0.19 (0.03–1.03)	0.054
Δrun percentage	Not entered	—	Not entered	—	Not entered	—	1.07 (1.02–1.11)	0.004*
ΔICM entropy	Not entered	—	Not entered	—	Not entered	—	0.97 (0.94–0.99)	0.044*
Postchemoradiotherapy roundness (log)	Not entered	—	Not entered	—	Not entered	—	0.10 (0.03–0.42)	0.001*

Entered variables that were eliminated based on redundancy were year of patient accrual, histologic differentiation grade, and signet ring cell adenocarcinoma (model 1); baseline SUV_{max} and ΔMTV (model 3); and baseline maximum probability (log), Δbusyness, Δcumulative histogram, postchemoradiation skewness, and postchemoradiation long-run high-intensity emphasis (log) (model 4).

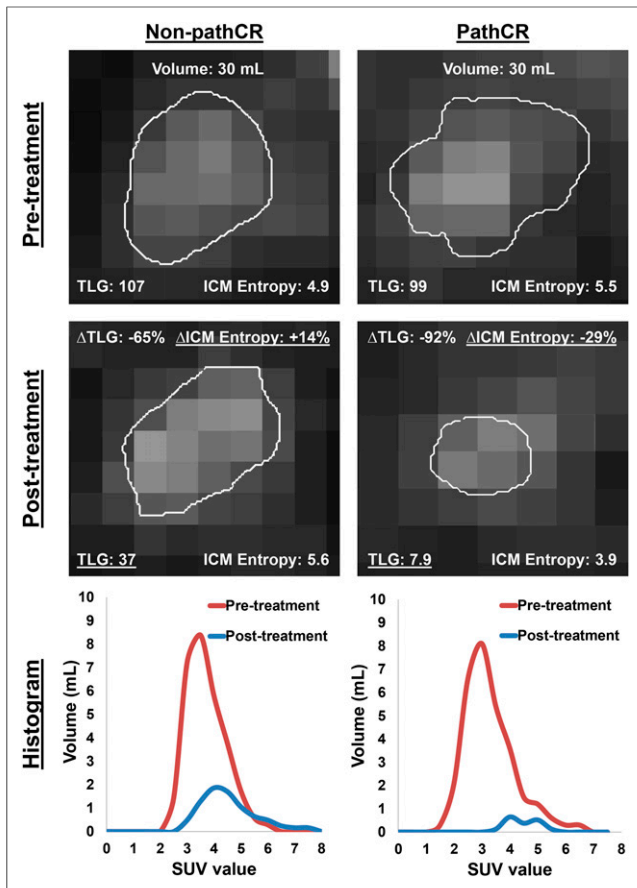ref = reference; ICM = intensity cooccurrence matrix.

FIGURE 1. Examples of representative transverse slices of ^{18}F-FDG PET scans before and after chemoradiotherapy in patient with no pathCR (i.e., non-pathCR) and in patient with a pathCR. These patients initially had comparable tumor volume, TLG, and local tumor texture (as expressed by intensity cooccurrence matrix [ICM] entropy metric). However, in the complete responder, ΔICM entropy metric decreased and posttreatment TLG was markedly lower (underlined), which represent 2 of the important predictors in models 3 and 4. Histograms illustrate 3-dimensional ^{18}F-FDG uptake within volumes of interest.

tumor length, clinical T stage, induction chemotherapy, and postchemoradiation endoscopic biopsy). Adding the subjective assessment of response on postchemoradiation ^{18}F-FDG PET to the clinical model significantly improved the model (model 2). Although 8

conventional quantitative ^{18}F-FDG PET features showed a P value of 0.25 or less in univariable analysis, only 4 (baseline SUV_{max}, ΔMTV, postchemoradiation SUV_{max}, and posttreatment TLG) were preselected for multivariable analysis because of several strong interparameter correlations (i.e., multicollinearity). After stepwise elimination, only posttreatment TLG appeared to provide significant model improvement (model 3).

Many geometry and first- and second-order texture features from the postchemoradiation ^{18}F-FDG PET appeared significantly associated with pathCR in univariable analysis (Supplemental Table 1). However, most of these features were highly correlated with postchemoradiation TLG (i.e., did not provide additional information but would increase the risk of multicollinearity) and were not preselected for multivariable analysis. Subsequently, 9 comprehensive ^{18}F-FDG PET features that met the preselection criteria were added to model 3 to construct model 4. Four of these features remained significantly important to the model after stepwise elimination (i.e., baseline cluster shade, Δrun percentage, ΔICM entropy, and postchemoradiation roundness). Examples of ^{18}F-FDG PET scans before and after chemoradiotherapy in patients with and without a pathCR are provided in Figure 1.

Model Performance

Measures of the performance of each model are presented in Table 4. Discrimination was moderate to good for all 4 models, with c-indices ranging from 0.71 for the clinical model to 0.75, 0.77, and 0.82 for the more complex models including subjective assessment of ^{18}F-FDG PET and conventional and comprehensive quantitative ^{18}F-FDG PET features, respectively (Fig. 2). After internal validation, the corrected c-indices appeared slightly lower (0.67, 0.72, 0.73, and 0.77 for models 1, 2, 3, and 4, respectively), indicating a limited degree of (a combination of) overfitting and bias due to the predictor selection process.

Visual inspection of the consecutive model calibration plots showed a good overall fit (Fig. 3). In general, the models tended to slightly overestimate the incidence of pathCR and underestimate the incidence of residual cancer as a result of minimal overfitting (Fig. 3). In Supplemental Table 3, the logistic regression formulas of the 4 models are presented in which the minimal overfitting and miscalibration were considered by applying a shrinkage factor that was obtained by bootstrapping.

Sensitivity analysis demonstrated no relevant change in discrimination after excluding 45 cases with initial tumor volumes below 10 mL (corrected c-indices, 0.67, 0.70, 0.72, and 0.78 for

TABLE 4
Estimates of Model Performance for 4 Prediction Models

		Discrimination	
Model no	Model type	Apparent c-index	Corrected c-index*
1	Clinical model	0.71 (0.64–0.79)	0.67 (0.60–0.75)
2	+ Subjective assessment ^{18}F-FDG PET	0.75 (0.68–0.82)	0.72 (0.65–0.79)
3	+ Conventional ^{18}F-FDG PET features	0.77 (0.70–0.84)	0.73 (0.66–0.80)
4	+ Comprehensive ^{18}F-FDG PET features	0.82 (0.75–0.88)	0.77 (0.70–0.83)

*Correction after internal validation for both optimism and bias from predictor selection process.
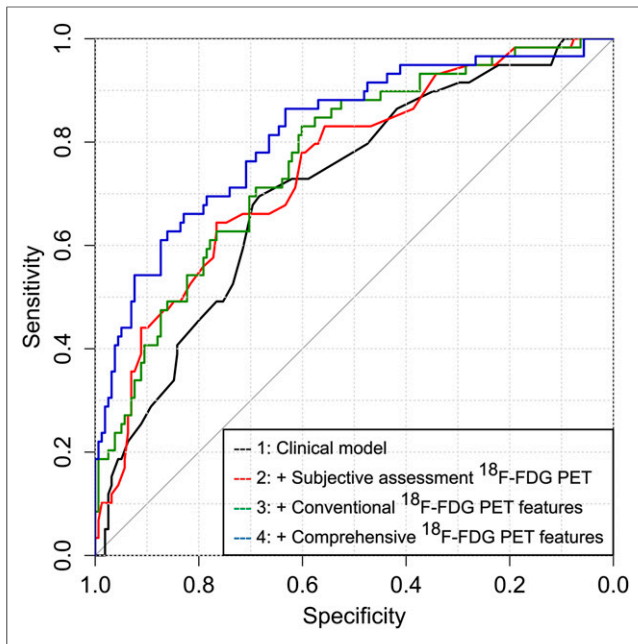Data in parentheses are 95% confidence intervals.

FIGURE 2. Receiver-operating-characteristic curve analysis of the 4 models indicating their ability to discriminate between pathCR and non-pathCR patients.

models 1, 2, 3, and 4, respectively) or after excluding 65 cases with 2-dimensional-acquisition ^{18}F-FDG PET scans (corrected c-indices, 0.69, 0.69, 0.71, and 0.74 for models 1, 2, 3, and 4, respectively).

Clinical Benefit

The decision curves of the models for the prediction of pathCR are displayed in Figure 4. At decision thresholds ranging from approximately 0.3 to approximately 0.7, some incremental value of ^{18}F-FDG PET–based analyses (i.e., models 2–4) beyond clinical predictors (i.e., model 1) was suggested by the divergence between the decision curves. However, at a decision threshold of 0.9 or higher—representing a clinically relevant predictive value for pathCR of 90% or more at which one may be willing to omit surgery—there was no clear incremental value of the prediction models (net benefit, 0% for models 1–3, and 1.8% for model 4).

DISCUSSION

This study demonstrates incremental value of baseline and postchemoradiation ^{18}F-FDG PET scanning for predicting pathCR after preoperative chemoradiotherapy in esophageal cancer beyond clinical predictive factors in terms of discrimination. A clinical prediction model (corrected c-index, 0.67) was improved by adding the subjective assessment of response by an experienced nuclear medicine physician on a postchemoradiation ^{18}F-FDG PET scan (corrected c-index, 0.72) and even more by subsequently adding comprehensive quantitative ^{18}F-FDG PET feature analysis (corrected c-index, 0.77). In terms of clinical benefit, however, both the clinical and the more complex ^{18}F-FDG PET–based multivariable models were not able to provide reliable predictive values for pathCR of approximately 90% or more, which was supported by the lack of net benefit (0%–1.8%) in decision-curve analysis at this clinically relevant decision threshold. In our opinion, these findings suggest that adding simple or complex

^{18}F-FDG PET analyses to clinical parameters for the prediction of pathCR does not aid the clinician to such a reliable extent that one may be willing to omit surgery in predicted complete responders.

The recent recognition that medical images may contain more useful information than can be perceived with the naked eye led to the field of radiomics, in which comprehensive features can be extracted by computational postprocessing algorithms (*20*). Evidence is slowly emerging that these comprehensive texture and geometry features may yield additional predictive and prognostic information in several solid tumors (*20*). Although a single feature is not directly linked to a specific biologic process, investigators assume that a combination of image-derived textural (and geometry) features may be closely related to underlying biologic processes such as vascularization, cell density, tumor aggressiveness, or hypoxia (*20,34*), which in turn may be related to the degree of chemoradiation resistance. In accordance with what one could reasonably expect, a tumor exhibiting a heterogeneous—compared with a homogeneous—^{18}F-FDG distribution at baseline (i.e., higher cluster shade) was less likely to reach pathCR in our analysis. Similarly, a larger change toward a more homogeneous ^{18}F-FDG uptake and a greater loss in the amount of information (or unpredictability) as a result of chemoradiotherapy—as reflected by Δrun percentage and ΔICM entropy, respectively—resulted in a higher chance of pathCR. After chemoradiotherapy, more roundly shaped ^{18}F-FDG uptake with higher TLG was more likely to represent residual cancer.

Several investigators have studied the predictive value of baseline or postchemoradiation ^{18}F-FDG uptake for treatment response in esophageal cancer. Large recent studies ($n = 88–284$) reported that the predictive value of a clinical complete response based on post-chemoradiation ^{18}F-FDG PET (either or not combined with endoscopic biopsy) for pathCR was low (range, 20%–64%) (*11,12,14*), a finding that was confirmed in the current study (45%). In other
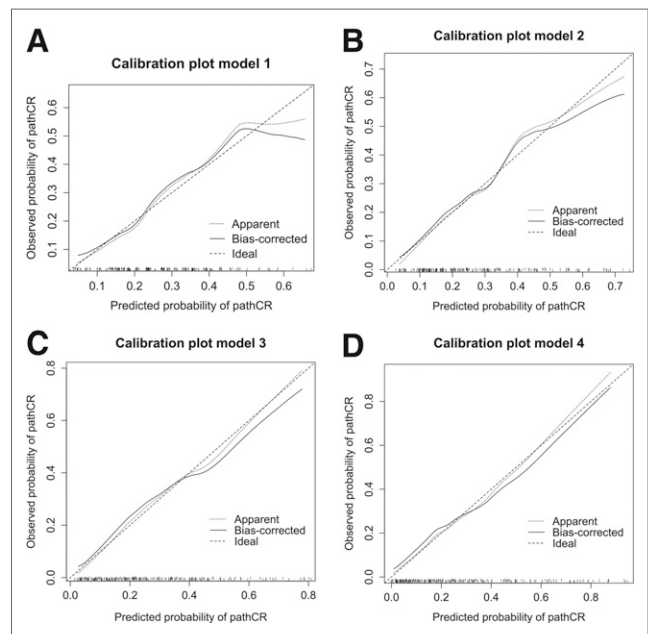

FIGURE 3. (A–D) Calibration plots of the 4 models demonstrating agreement between predicted probability of pathCR by model and observed incidence.
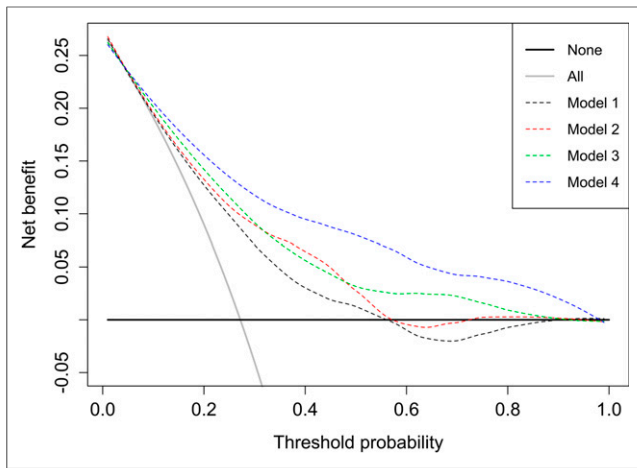
FIGURE 4. Decision curves graphically representing net benefit (*y*-axis) for the 4 models at a range of decision thresholds (i.e., minimum probabilities of pathCR at which one would be willing to change clinical decision making; *x*-axis). The black and gray solid lines represent making same decision in all patients (i.e., omitting surgery in none or in all of the patients, respectively).

studies, pathologic response was variably found to be associated with image-derived quantitative parameters such as baseline SUV_{max} (*35*) and MTV (*17*), or postchemoradiation SUV_{max} (*5*) and SUV_{mean} (*36*), or the relative change in SUV_{max}, MTV, and TLG (*16,37*). In the current study, a similar trend—although nonsignificant—was observed for baseline SUV_{max} ($P = 0.087$), whereas postchemoradiation SUV_{max}, MTV, and TLG were significantly associated with pathCR ($P = 0.015$, $P < 0.001$, and $P < 0.001$, respectively). Importantly, however, in former studies as well as the current study, these subjective and quantitative ^{18}F-FDG PET parameters did not allow differentiation of pathCR from nonpathCR with high accuracy, a distinction that could be useful for clinical decision making. This emphasizes the difference between a significant association of a parameter with pathologic response and a

true predictive value in terms of discrimination or incremental value beyond other predictors. To this regard, we encourage investigators to study new potential diagnostic biomarkers not only for their univariable association with pathologic response, but also for their predictive value and incremental value using a multivariable approach.

In esophageal cancer, 4 recent pilot studies from 2 research groups observed that ^{18}F-FDG PET texture features seemed more informative than conventional metabolic parameters for the prediction of response to chemoradiotherapy (Table 5) (*29,34,38,39*). However, these studies were generally limited by small sample size ($n = 20$ to $n = 50$), heterogeneity of included tumor types and stages, lack of internal validation, and high likeliness of model overfitting and overoptimism of reported results. Also, the used reference standard was suboptimally defined according to RECIST (*29,34*)—which is known to correlate poorly with pathologic response (*8*)—or pathCR was mixed together as 1 group with microscopic residual disease (*38,39*) while these groups are considered different entities associated with different survival rates (*2,4*). In addition, the studies differed from the current study regarding the used delineation method. All these substantial differences with the current study compromise proper comparison.

The reported influences of clinical parameters—including pretreatment EUS-based tumor length and T stage and posttreatment endoscopic biopsy—on the probability of pathCR in our study were in accordance with existing literature (*5,40,41*). The finding that induction chemotherapy before chemoradiotherapy was associated with a significantly higher rate of pathCR in the current series was in line with a recent phase II randomized trial at MDACC that did report a difference in pathCR rate in favor of induction chemotherapy (26% vs. 13%), but this difference was not statistically significant in that trial ($P = 0.094$) (*42*). The difference in significance and nonsignificance of the influence of induction chemotherapy on pathCR between the current study and the randomized trial cannot be fully explained. Potential explanations are possible selection bias through selecting operated patients in the current study only or a lack of statistical power (i.e., sample size) in the randomized trial. However, in the current study the significant difference between

TABLE 5
Overview of Studies on ^{18}F-FDG PET Texture Analysis in Esophageal Cancer for Treatment Response Assessment

Study, year	*n*	Tumor type	Tumor stages	Timing of ^{18}F-FDG PET	Outcome associated with tumor texture
Tixier et al., 2011 (*34*)*	41	Adenocarcinoma, squamous cell carcinoma	I–IV	Prechemoradiotherapy	Clinical response (according to RECIST)
Hatt et al. 2013 (*29*)*	50	Adenocarcinoma, squamous cell carcinoma	I–IV	Prechemoradiotherapy	Clinical response (according to RECIST)
Tan et al., 2013 (*38*)†	20	Adenocarcinoma, squamous cell carcinoma	II–III	Pre- and postchemoradiotherapy	Pathologic response (TRG 1 + 2 vs. 3 + 4)
Zhang et al., 2014 (*39*)†	20	Adenocarcinoma, squamous cell carcinoma	II–III	Pre- and postchemoradiotherapy	Pathologic response (TRG 1 + 2 vs. 3 + 4)
Current study	217	Adenocarcinoma	II–III	Pre- and postchemoradiotherapy	pathCR (TRG 1 vs. 2–4)

*Significant overlap of study populations.
†Complete overlap of study populations.
TRG according to Chirieac et al. (*2*).

the 2 groups was considered not negligible, and induction chemotherapy was therefore entered into the multivariable prediction modeling process.

Several limitations apply to this study. First, slight model overfitting occurred despite the large sample size. Therefore, we reported optimism-corrected model performance estimates resulting from internal validation in addition to apparent estimates. In addition, external validation of the prediction models would be necessary to determine the residual overestimation of generalizability caused by the issue of multiple testing that was inherent to the explorative character of this study. Second, pathologic response in the primary tumor alone was examined, whereas a pathCR of the primary tumor may not completely exclude residual lymph node involvement (*17*), and potentially useful information provided by ^{18}F-FDG PET on lymph node involvement and response could have been missed. Particularly in patients with large lymphadenopathies, texture analysis could potentially provide valuable complementary information (*24,25*). Third, it should be acknowledged that the MTV threshold of 10 mL used for sensitivity analysis may not have been perfectly chosen, because it was based on a previous study that applied slightly different calculation methods for the texture features (*24*). Fourth, the MTV on the postchemoradiation scan was small in many cases, which could have compromised the reliability of the texture features analysis of these scans. Finally, inherently to the retrospective character of this study, the authors could not determine whether more modern reconstruction protocols using smaller voxel sizes, smaller postreconstruction filtering methods, and isotropic rather than anisotropic voxels (*43*) could potentially provide incremental information on tumor heterogeneity. However, our analysis was strengthened by using a prospectively maintained database, including clinical parameters that are practical, restricting analysis to robust features only, studying incremental rather than isolated predictive values, performing internal validation of the developed models, assessing added value in terms of clinical benefit, and providing novel findings.

No comparison with diffusion-weighted MRI (DW-MRI) was performed in this study. However, in a recent prospective explorative study our group found that the treatment-induced change in the median tumor apparent diffusion coefficient during the first 2–3 wk of preoperative chemoradiotherapy as determined on DW-MRI seemed highly predictive of pathCR (*44*). In future research, a multimodality imaging approach—rather than using ^{18}F-FDG PET alone—might prove to provide sufficient incremental predictive value for pathCR beyond clinical predictors to safely guide treatment decision making. Therefore, in a recently initiated multicenter study (ClinicalTrials.gov, NCT02125448) we aim to evaluate the potential complementary value of DW-MRI in addition to ^{18}F-FDG PET for predicting pathCR by performing both DW-MRI and ^{18}F-FDG PET before, during, and after preoperative chemoradiotherapy in patients with esophageal cancer.

CONCLUSION

This study demonstrates that subjective and quantitative assessment of baseline and postchemoradiation ^{18}F-FDG PET provides statistical—but currently not clinically relevant—incremental value for predicting pathCR after preoperative chemoradiotherapy in esophageal cancer. The statistical discriminatory improvement beyond clinical predictors did not translate into a clinically relevant benefit that could change clinical decision making in terms of safely omitting surgery in predicted complete responders. This particular clinical dilemma demands a high predictive accuracy before clinical decision making can be influenced, warranting improvement, development, and validation of current and new imaging—or biomarkers.

REFERENCES

1. van Hagen P, Hulshof MC, van Lanschot JJ, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med*. 2012;366:2074–2084.
2. Chirieac LR, Swisher SG, Ajani JA, et al. Posttherapy pathologic stage predicts survival in patients with esophageal carcinoma receiving preoperative chemoradiation. *Cancer*. 2005;103:1347–1355.
3. Berger AC, Farma J, Scott WJ, et al. Complete response to neoadjuvant chemoradiotherapy in esophageal carcinoma is associated with significantly improved survival. *J Clin Oncol*. 2005;23:4330–4337.
4. Donahue JM, Nichols FC, Li Z, et al. Complete pathologic response after neoadjuvant chemoradiotherapy for esophageal cancer is associated with enhanced survival. *Ann Thorac Surg*. 2009;87:392–398.
5. Ajani JA, Correa AM, Hofstetter WL, et al. Clinical parameters model for predicting pathologic complete response following preoperative chemoradiation in patients with esophageal cancer. *Ann Oncol*. 2012;23:2638–2642.
6. Kwee RM. Prediction of tumor response to neoadjuvant therapy in patients with esophageal cancer with use of ^{18}F FDG PET: a systematic review. *Radiology*. 2010;254:707–717.
7. Schneider PM, Metzger R, Schaefer H, et al. Response evaluation by endoscopy, rebiopsy, and endoscopic ultrasound does not accurately predict histopathologic regression after neoadjuvant chemoradiation for esophageal cancer. *Ann Surg*. 2008;248:902–908.
8. Westerterp M, van Westreenen HL, Reitsma JB, et al. Esophageal cancer: CT, endoscopic US, and FDG PET for assessment of response to neoadjuvant therapy: systematic review. *Radiology*. 2005;236:841–851.
9. van Rossum PS, van Hillegersberg R, Lever FM, et al. Imaging strategies in the management of oesophageal cancer: what's the role of MRI? *Eur Radiol*. 2013;23:1753–1765.
10. van Rossum PS, van Lier AL, Lips IM, et al. Imaging of oesophageal cancer with FDG-PET/CT and MRI. *Clin Radiol*. 2015;70:81–95.
11. Bruzzi JF, Swisher SG, Truong MT, et al. Detection of interval distant metastases: clinical utility of integrated CT-PET imaging in patients with esophageal carcinoma after neoadjuvant therapy. *Cancer*. 2007;109:125–134.
12. Elliott JA, O'Farrell NJ, King S, et al. Value of CT-PET after neoadjuvant chemoradiation in the prediction of histological tumour regression, nodal status and survival in oesophageal adenocarcinoma. *Br J Surg*. 2014;101:1702–1711.
13. Chen YM, Pan XF, Tong LJ, Shi YP, Chen T. Can ^{18}F-fluorodeoxyglucose positron emission tomography predict responses to neoadjuvant therapy in oesophageal cancer patients? A meta-analysis. *Nucl Med Commun*. 2011;32:1005–1010.
14. Cheedella NK, Suzuki A, Xiao L, et al. Association between clinical complete response and pathological complete response after preoperative chemoradiation in patients with gastroesophageal cancer: analysis in a large cohort. *Ann Oncol*. 2013;24:1262–1266.
15. Ott K, Weber WA, Lordick F, et al. Metabolic imaging predicts response, survival, and recurrence in adenocarcinomas of the esophagogastric junction. *J Clin Oncol*. 2006;24:4692–4698.
16. Roedl JB, Colen RR, Holalkere NS, Fischman AJ, Choi NC, Blake MA. Adenocarcinomas of the esophagus: response to chemoradiotherapy is associated with decrease of metabolic tumor volume as measured on PET-CT: comparison to histopathologic and clinical response evaluation. *Radiother Oncol*. 2008;89:278–286.
17. Blom RL, Steenbakkers IR, Lammering G, et al. PET/CT-based metabolic tumour volume for response prediction of neoadjuvant chemoradiotherapy in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging*. 2013;40:1500–1506.
18. Hatt M, Visvikis D, Pradier O, Cheze-le Rest C. Baseline ^{18}F-FDG PET image-derived parameters for therapy response prediction in oesophageal cancer. *Eur J Nucl Med Mol Imaging*. 2011;38:1595–1606.

19. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit.* 2009;42: 1162–1171.

20. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in ^{18}F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging.* 2013;40:133–140.

21. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ.* 2003;326: 41–44.

22. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162:W1–W73.

23. Karamitopoulou E, Thies S, Zlobec I, et al. Assessment of tumor regression of esophageal adenocarcinomas after neoadjuvant chemotherapy: comparison of 2 commonly used scoring approaches. *Am J Surg Pathol.* 2014;38:1551–1556.

24. Hatt M, Majdoub M, Vallieres M, et al. ^{18}F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med.* 2015;56:38–44.

25. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med.* 2014;55:37–42.

26. Werner-Wasik M, Nelson AD, Choi W, et al. What is the best way to contour lung tumors on PET scans: multiobserver validation of a gradient-based method using a NSCLC digital PET phantom. *Int J Radiat Oncol Biol Phys.* 2012;82:1164–1171.

27. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys.* 2015;42:1341–1353.

28. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET. *J Nucl Med.* 2012;53:693–700.

29. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. *Eur J Nucl Med Mol Imaging.* 2013;40:1662–1671.

30. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1996;1:30–46.

31. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I—development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* 2012;98:683–690.

32. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26:565–574.

33. Piessen G, Messager M, Mirabel X, et al. Is there a role for surgery for patients with a complete clinical response after chemoradiation for esophageal cancer: an intention-to-treat case-control study. *Ann Surg.* 2013;258:793–799.

34. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011;52:369–378.

35. Rizk NP, Tang L, Adusumilli PS, et al. Predictive value of initial PET-SUV$_{max}$ in patients with locally advanced esophageal and gastroesophageal junction adenocarcinoma. *J Thorac Oncol.* 2009;4:875–879.

36. Swisher SG, Erasmus J, Maish M, et al. 2-Fluoro-2-deoxy-D-glucose positron emission tomography imaging is predictive of pathologic response and survival after preoperative chemoradiation in patients with esophageal carcinoma. *Cancer.* 2004;101:1776–1785.

37. Stiekema J, Vermeulen D, Vegt E, et al. Detecting interval metastases and response assessment using ^{18}F-FDG PET/CT after neoadjuvant chemoradiotherapy for esophageal cancer. *Clin Nucl Med.* 2014;39:862–867.

38. Tan S, Kligerman S, Chen W, et al. Spatial-temporal [^{18}F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. *Int J Radiat Oncol Biol Phys.* 2013;85:1375–1382.

39. Zhang H, Tan S, Chen W, et al. Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal ^{18}F-FDG PET features, clinical parameters, and demographics. *Int J Radiat Oncol Biol Phys.* 2014;88: 195–203.

40. Chao YK, Tseng CK, Wen YW, et al. Using pretreatment tumor depth and length to select esophageal squamous cell carcinoma patients for nonoperative treatment after neoadjuvant chemoradiotherapy. *Ann Surg Oncol.* 2013;20:3000–3008.

41. Huang RW, Chao YK, Wen YW, et al. Predictors of pathological complete response to neoadjuvant chemoradiotherapy for esophageal squamous cell carcinoma. *World J Surg Oncol.* 2014;12:170–176.

42. Ajani JA, Xiao L, Roth JA, et al. A phase II randomized trial of induction chemotherapy versus no induction chemotherapy followed by preoperative chemoradiation in patients with esophageal cancer. *Ann Oncol.* 2013;24: 2844–2849.

43. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol.* 2015;60:5471–5496.

44. van Rossum PS, van Lier AL, van Vulpen M, et al. Diffusion-weighted magnetic resonance imaging for the prediction of pathologic response to neoadjuvant chemoradiotherapy in esophageal cancer. *Radiother Oncol.* 2015;115:163–170.