SUPPLEMENTARY MATERIAL

Supplement to: Borggreve AS et al. Preoperative prediction of pathologic response to neoadjuvant chemoradiotherapy in patients with esophageal cancer using ^{18}F-FDG PET/CT and DW-MRI: a prospective multicenter study.

CONTENTS

Methods

The results of first 20 of the 26 included patients from MDACC have been published previously.[1]

Exclusion criteria

Exclusion criteria included an age of <18 years, previous treatment with thoracic surgery or thoracic radiotherapy, and contraindications for ^{18}F-FDG PET/CT or MRI. The diagnostic work-up consisted of an endoscopy with biopsy for diagnosis, as well as EUS and integrated ^{18}F-FDG PET/CT. In addition, patients who were initially included in the study but eventually did not undergo surgery were excluded from further analyses.

Survival

Survival data was completed for all but 1 patient (n=68) at least up to 18 months after date of surgery (median [IQR] time to censoring: 37 months [33 – 49 months]). Data on disease recurrence and disease free survival (DFS) was completed for all but 5 patients (n=63). Disease recurrence was defined as local or distant recurrence, either based on imaging or histopathological assessment.

^{18}F-FDG PET/CT scan parameters

The timing of the ^{18}F-FDG PET/CT scan in the second or third week of treatment was based on previous studies which demonstrated a potentially superior accuracy at this time point as compared to pre-treatment and post-treatment scanning only.[2,3] The ^{18}F-FDG PET examinations were performed on dedicated PET/CT systems. Patients were instructed to fast for at least 6 hours before ^{18}F-FDG PET and a glucose level within the normal range (80-120 mg/dl) was confirmed. Before ^{18}F-FDG PET, a CT scan without contrast agent was acquired for attenuation correction purposes. ^{18}F-FDG PET scans were acquired 60-90 minutes after administration of ^{18}F-FDG with a dose ranging between 190-370 MBq, in three-dimensional (3D) acquisition mode at 2-5 minutes per bed position.

DW-MRI scan parameters

The DW-MRI examinations were either performed on a 1.5T (UMC Utrecht and NKI-AVL; Achieva, Philips Medical Systems, Best, The Netherlands) or on a 3.0T scanner (MDACC; Discovery MR750, GE Healthcare, Milwaukee, Wisconsin, USA). Transverse diffusion-weighted images were obtained with free breathing and using 3 different b-values (b = 0, 200 and 800 s/mm^2).

Image analysis – tumor delineations

^{18}F-FDG PET/CT imaging analysis, including primary tumor delineation and calculation of metabolic and volumetric parameters, was performed using commercially available software (MIM Software, Cleveland, Ohio, USA). The primary tumor volume was defined as the volume of interest (VOI) and contoured using a semi-automatic gradient-based delineation method – which has been validated in a multi-observer study reporting superior accuracy, consistency and robustness compared with manual and threshold methods[4] – followed by manual editing by two readers.

DW-MRI analysis was performed using an imaging analysis software package (ImageI).[5] The primary tumor – excluding the lumen – was delineated on the DW-MRI scans with a b-value of 200 s/mm^2 using semi-automatic contouring, allowing for manual editing by one reader.[5] Contouring of the tumor was performed conservatively to avoid the edges of the tumor boundaries, as ADC values in the periphery of the tumor may be unreliable due to motion or other image distortions. The DW-MRI scans with b-values of 0, 200, and 800 s/mm^2 were fitted with a mono-exponential model to generate quantitative ADC maps for each slice.[1,6,7]

Statistical analysis

Clinical characteristics that potentially predict response were pre-specified based on previous literature (i.e., clinical T status, histologic subtype, neoadjuvant chemoradiotherapy

regimen and time interval from nCRT to surgery) and were compared between patients with pCR (TRG 1) and non-pCR patients (TRG 2-4), and between good responders (TRG 1-2, GR) and poor responders (TRG 3-4, non-GR) based on the χ^2 or Fisher's exact test for categorical variables and the Mann-Whitney U test for continuous variables.

The relative changes of the ^{18}F-FDG PET/CT and DW-MRI parameters were also compared between these patient groups using the Mann-Whitney U test to validate findings of previous pilot-studies.[1,7–12] Benjamini-Hochberg corrections were applied to adjust for multiple comparisons and therewith minimizing the false discovery rate.[13] The ability of single modality ^{18}F-FDG PET/CT and DW-MRI parameters to discriminate between different pathologic response groups was quantified using the area under the receiver operating characteristic (ROC) curve (c-statistic).

Secondly, the complementary value of ^{18}F-FDG PET/CT and DW-MRI parameters was assessed using multivariable penalized Ridge regression models to reduce model overfitting in a situation with few events per variable.[14] The applied overall penalty (λ_{min}) represents the minimum mean cross-validated error, which was obtained using 10-fold cross validation. Optimism-corrected c-statistics of the Ridge regression models were obtained by bootstrap resampling.[15] The original dataset was a 1,000 times resampled with replacement to obtain a dataset of the same size. All models were then fitted to the 1,000 bootstrap samples. Each fitted model was then applied both to the resampled dataset from which it was generated, and to the original dataset. The optimism corrected c-statistic was calculated as the original c-statistic minus the optimism, which was calculated as the difference between the c-statistic on the original dataset and the resampled dataset. Also, bootstrapping allowed for reconstruction of 95% confidence intervals (CI) around the c-statistic. The global fit of the models was compared using Akaike Information Criterion (AIC), with the lowest AIC representing the best fit.[16,17] Model calibration of the models was evaluated by visual inspection of the model calibration plots.[17] Variables to be entered in the penalized regression models with pCR or GR as outcome were histopathological tumor type – which

an important factor to impact pathologic response to nCRT based on previous literature – and the ^{18}F-FDG PET/CT and DW-MRI parameter with the highest c-statistic in univariable analyses.

Separate analyses were performed on pCR (TRG 1) versus non-pCR (TRG 2-4), and GR (TRG 1-2) versus non-GR (TRG 3-4). The first analysis aimed at aiding clinical decision-making regarding omission of surgery in anticipated complete responders, whereas the second analysis was deemed as more relevant for potentially modifying or discontinuing nCRT early during treatment. For the latter purpose, only ^{18}F-FDG PET/CT and DW-MRI parameters *during* nCRT were taken into account, as knowledge of response after completion of nCRT would be irrelevant for modification of nCRT.

In order to evaluate whether the best performing model for pCR prediction based on the aforementioned analyses correlate with overall survival (OS) and disease free survival (DFS), multivariable Cox regression analyses were performed with estimation of hazard ratio's (HR) and 95% CI for the included predictors.

All statistical analyses were performed using R 3.1.2 open-source software ('pROC', 'glmnet', 'boot' and 'rms' packages, http://www.R-project.org). A p-value of <0.05 was considered statistically significant.

References

1. Fang P, Musall BC, Son JB, et al. Multimodal Imaging of Pathologic Response to Chemoradiation in Esophageal Cancer. *Int J Radiat Oncol Biol Phys*. 2018;102:996–1001.

2. Chen Y-M, Pan X, Tong L-J, et al. Can 18F-fluorodeoxyglucose positron emission tomography predict responses to neoadjuvant therapy in oesophageal cancer patients? A meta-analysis. *Nucl Med Commun*. 2011;32:1005–10.

3. 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;24:4692–8.

4. 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;82:1164–71.

5. Musall B. Quantitative DWI as an Early Imaging Biomarker of the Response to Chemoradiation in Esophageal Cancer Available from: https://digitalcommons.library.tmc.edu/utgsbs_dissertations/805. 2017. Accessed September 4, 2018.

6. Padhani AR, Liu G, Koh DM, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. *Neoplasia*. 2009;11:102–25.

7. van Rossum PSN, van Lier ALHMW, 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.

8. van Rossum PSN, Fried D V, Zhang L, et al. The Incremental Value of Subjective and Quantitative Assessment of 18F-FDG PET for the Prediction of Pathologic Complete Response to Preoperative Chemoradiotherapy in Esophageal Cancer. *J Nucl Med*. 2016;57:691–700.

9. Li Q-W, Qiu B, Wang B, et al. Prediction of pathologic responders to neoadjuvant chemoradiotherapy by diffusion-weighted magnetic resonance imaging in locally advanced esophageal squamous cell carcinoma: a prospective study. *Dis Esophagus*. 2018;31(2).

10. Tan S, Kligerman S, Chen W, et al. Spatial-Temporal [18F]FDG-PET Features for Predicting Pathologic Response of Esophageal Cancer to Neoadjuvant Chemoradiation Therapy. *Int J Radiat Oncol*. 2013;85:1375–1382.

11. Beukinga RJ, Hulshoff JB, Mul VEM, et al. Prediction of Response to Neoadjuvant Chemotherapy and Radiation Therapy with Baseline and Restaging 18 F-FDG PET Imaging Biomarkers in Patients with Esophageal Cancer. *Radiology*. 2018;287:983–992.

12. van Heijl M, Omloo JM, van Berge Henegouwen MI, et al. Fluorodeoxyglucose Positron Emission Tomography for Evaluating Early Response During Neoadjuvant Chemoradiotherapy in Patients With Potentially Curable Esophageal Cancer. *Ann Surg*. 2011;253:56–63.

13. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc*. 1995;57:289–300.

14. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ*. 2015;351:h3868.

15. Steyerberg EW, Harrell FE, Borsboom GJJM, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774–781.

16. Akaike H. Likelihood of a model and information criteria. *J Econom*. 1981;16:3–14.

17. Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. *Circulation*. 2007;115:928–935.

Supplementary Table 1. Tumor regression grades (TRG) per histologic subtype and per treatment regimen.

Treatment regimen	Histologic subtypes	TRG 1	TRG 2	TRG 3	TRG 4	Total
Full cohort	Adenocarcinoma	10 (17.5%)	22 (38.6%)	19 (33.3%)	6 (10.5%)	57 (100%)
	Squamous cell carcinoma	7 (63.6%)	3 (27.3%)	1 (9.1%)	0 (0%)	11 (100%)
	Undifferentiated large cell carcinoma	1 (100%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
Carboplatin/paclitaxel + 41.4 Gy (UMC Utrecht + NKI-AVL)	Adenocarcinoma	4 (11.8%)	12 (35.3%)	12 (35.3%)	6 (17.6%)	34 (100%)
	Squamous cell carcinoma	6 (75%)	2 (25%)	0 (0%)	0 (0 %)	8 (100%)
	Undifferentiated large cell carcinoma	1 (100%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
5-fluorouracil-based + 50.4 Gy (MDACC)	Adenocarcinoma	6 (26.1%)	10 (43.5%)	7 (30.4%)	0 (0%)	23 (100%)
	Squamous cell carcinoma	1 (33.3%)	1 (33.3%)	1 (33.3%)	0 (0%)	3 (100%)

Data are numbers of patients, with row-based percentages in parentheses.

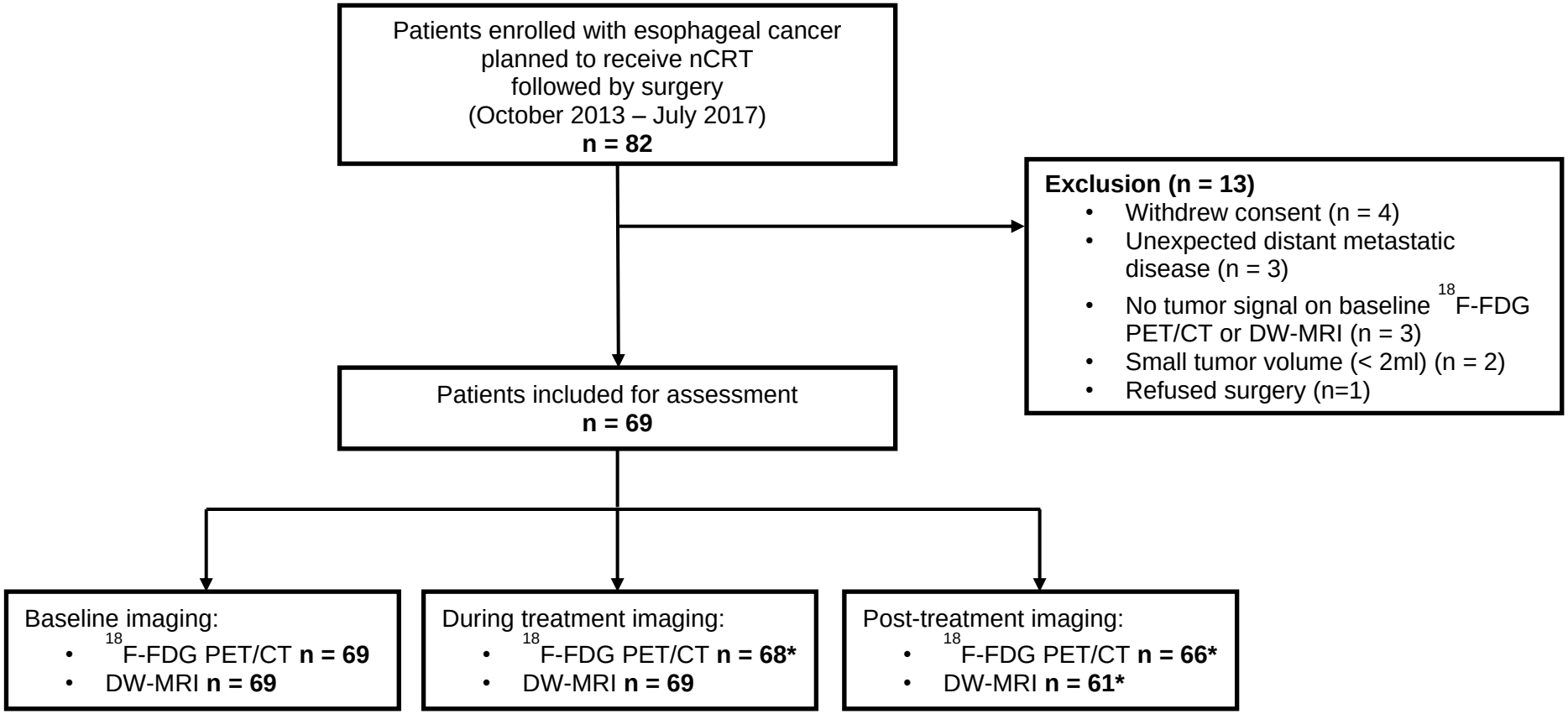
Supplementary Table 2. Intercepts and regression coefficients for Ridge regression analyses on the complementary value of ^{18}F-FDG PET/CT and DW-MRI parameters with pathologic complete response (TRG 1) and good response (TRG 1-2) as outcome variables.

pCR (TRG 1)			GR (TRG 1 +2)		
Intercept and predictors	**β**	**AIC**	**Intercept and predictors**	**β**	**AIC**
Model 1: Histology					
Intercept	-1.55	71.36	Intercept	0.25	88.85
Squamous cell carcinoma †	2.10		Squamous cell carcinoma †	2.06	
Model 2: DW-MRI parameter and histology					
Intercept	-2.20	64.73	Intercept	-0.71	85.83
ΔADC_{during} (%)	0.04		ΔADC_{during} (%)	0.08	
Squamous cell carcinoma †	1.52		Squamous cell carcinoma †	1.08	
Model 3: ^{18}F-FDG PET/CT parameter and histology					
Intercept	-2.51	61.32	Intercept	0.12	71.64
$\Delta SUV_{mean,post}$ (%)	-0.02		$\Delta SUV_{max,during}$ (%)	-0.01	
Squamous cell carcinoma †	1.36		Squamous cell carcinoma †	1.14	
Model 4: DW-MRI and ^{18}F-FDG PET/CT parameters and histology					
Intercept	-2.68	61.13	Intercept	-0.69	72.31
ΔADC_{during} (%)	0.03		ΔADC_{during} (%)	0.06	
$\Delta SUV_{mean,post}$ (%)	-0.02		$\Delta SUV_{max,during}$ (%)	-0.01	
Squamous cell carcinoma †	1.16		Squamous cell carcinoma †	1.14	

ADC apparent diffusion coefficient; *AIC* Akaike Information Criterion; *GR* good response (TRG 1-2); *pCR* pathologic complete response; *SUV* standardized uptake value

† Adenocarcinoma was used as reference category.

Supplementary Figure 1. Study profile.



* Not available due to patients' refusal or an unexpectedly antedated surgical resection.

Supplementary Figure 2. Calibration plots of the penalized regression models as described in Table 3 for pathologic complete response (pCR) prediction (A-D) and good response (GR) prediction (E-H).