

# Early prediction of pathological response in locally advanced rectal cancer based on sequential F-18-FDG PET

Citation for published version (APA):

Hatt, M., van Stiphout, R., le Pogam, A., Lammering, G., Visvikis, D., & Lambin, P. (2013). Early prediction of pathological response in locally advanced rectal cancer based on sequential F-18-FDG PET. *Acta Oncologica*, 52(3), 619-626. <https://doi.org/10.3109/0284186X.2012.702923>

**Document status and date:**

Published: 01/04/2013

**DOI:**

[10.3109/0284186X.2012.702923](https://doi.org/10.3109/0284186X.2012.702923)

**Document Version:**

Publisher's PDF, also known as Version of record

**Document license:**

Taverne

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

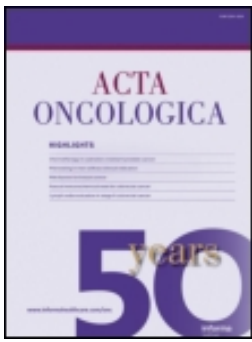
[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.






## Early prediction of pathological response in locally advanced rectal cancer based on sequential $^{18}\text{F}$ -FDG PET

Mathieu Hatt, Ruud van Stiphout, Adrien le Pogam, Guido Lammering, Dimitris Visvikis & Philippe Lambin


To cite this article: Mathieu Hatt, Ruud van Stiphout, Adrien le Pogam, Guido Lammering, Dimitris Visvikis & Philippe Lambin (2013) Early prediction of pathological response in locally advanced rectal cancer based on sequential  $^{18}\text{F}$ -FDG PET, Acta Oncologica, 52:3, 619-626, DOI: [10.3109/0284186X.2012.702923](https://doi.org/10.3109/0284186X.2012.702923)



To link to this article: <https://doi.org/10.3109/0284186X.2012.702923>

 View supplementary material 

 Published online: 08 Aug 2012.

 Submit your article to this journal 

 Article views: 717

 View related articles 

 Citing articles: 7 View citing articles 

ORIGINAL ARTICLE

## Early prediction of pathological response in locally advanced rectal cancer based on sequential $^{18}\text{F}$ -FDG PET

MATHIEU HATT<sup>1,2</sup>, RUUD VAN STIPHOUT<sup>1</sup>, ADRIEN LE POGAM<sup>2</sup>, GUIDO LAMMERING<sup>1</sup>, DIMITRIS VISVIKIS<sup>2</sup> & PHILIPPE LAMBIN<sup>1</sup>

<sup>1</sup>Department of Radiation Oncology (MAASTRO), GROW Research Institute, Maastricht, The Netherlands and <sup>2</sup>INSERM, UMR 1101 LaTIM, Brest, France

### Abstract

**Background.** The objectives of this study were to investigate the predictive value of sequential  $^{18}\text{F}$ -FDG PET scans for pathological tumor response grade (TRG) after preoperative chemoradiotherapy (PCRT) in locally advanced rectal cancer (LARC) and the impact of partial volume effects correction (PVC). **Methods.** Twenty-eight LARC patients were included. Responders and non-responders status were determined in histopathology. PET indices [SUV max and mean, volume and total lesion glycolysis (TLG)] at baseline and their evolution after one and two weeks of PCRT were extracted by delineation of the PET images, with or without PVC. Their predictive value was investigated using Mann-Whitney-U tests and ROC analysis. **Results.** Within baseline parameters, only  $\text{SUV}_{\text{mean}}$  was correlated with response. No evolution after one week was predictive of the response, whereas after two weeks all the parameters except volume were, the best prediction being obtained with TLG (AUC 0.79, sensitivity 63%, specificity 92%). PVC had no significant impact on these results. **Conclusion.** Several PET indices at baseline and their evolution after two weeks of PCRT are good predictors of response in LARC, with or without PVC, whereas results after one week are suboptimal. Best predictor was TLG reduction after two weeks, although baseline  $\text{SUV}_{\text{mean}}$  had smaller but similar predictive power.

Preoperative radiochemotherapy (PRCT) is now considered a standard treatment for patients diagnosed with locally advanced rectal cancer (LARC). A significant tumor downsizing and downstaging, as well as a reduction of the risk for local recurrence and longer survival have been observed as a result of PRCT [1–3]. Within this context,  $^{18}\text{F}$ -FDG PET imaging has been demonstrated as an interesting predictive tool [4]. Indeed, correlations between the pathological tumor response after PRCT and the standard uptake value (SUV) decrease within the tumor have been demonstrated in several studies [5–9]. In most of these studies, the SUV evolution between the baseline positron emission scan (PET) scan and the one acquired before surgery was correlated with the tumor regression grade (TRG) [5–9]. However, allowing for earlier prediction of the response is of higher interest for the clinical practice, since it might enable modifications of the treatment protocol [10]. A few studies have investigated the

early prediction of the pathological tumor response based on  $^{18}\text{F}$ -FDG PET imaging during PRCT [8,9]. Cascini et al. showed that changes in the metabolic activity of the tumor, measured as early as 15 days after the start of PRCT, were predictive for response [8]. More recently, Janssen et al. have examined three different time points of  $^{18}\text{F}$ -FDG PET imaging during PRCT, in order to define the optimal time for early prediction [9]. In this study  $^{18}\text{F}$ -FDG PET scans at baseline, as well as at 8 and 15 days during PRCT were carried out and it was found that the best predictive factor of TRG was the  $\text{SUV}_{\text{max}}$  response index (RI, defined as the percent evolution relative to the pretreatment value) at 15 days. All these studies have considered SUV measurements (max and/or mean) only. It has been demonstrated recently in several studies and various malignancies that other  $^{18}\text{F}$ -FDG PET derived parameters more fully characterizing tumors on a functional level can have statistically higher predictive value than SUV

[10–17]. These include metabolically active tumor volume (MATV, defined as the tumor volume as it can be seen and delineated on a PET image) [11] and total lesion glycolysis (TLG, multiplying MATV and its associated mean SUV) [12]. In addition, some of these studies demonstrated that response could be predicted by extracting these parameters from the pretreatment scan only, therefore potentially eliminating or reducing the need for sequential scans during treatment. Such results have been presented within the context of locally advanced esophageal cancer (LAEC) [11,13], non-Hodgkin lymphoma [14], pleural mesothelioma [15] and cervix and head and neck cancers [16], whereas conflicting results have been recently obtained in rectal cancer [17,18]. Finally, most of these studies have considered the evolution of PET derived parameters without partial volume effects (PVE) correction (PVC). This may lead to biased results especially if MATVs change size and/or shape during treatment. The current study was therefore conducted retrospectively on the cohort previously imaged [9] with the following objectives: 1) determine the predictive value of baseline  $^{18}\text{F}$ -FDG PET derived parameters; 2) investigate the evolution of these parameters during treatment and their associated predictive value; and 3) investigate the impact of PVC on these results.

## Material and methods

### Patients

This study consisted of a retrospective analysis of a cohort of 28 patients (see Table I) diagnosed with non-metastasized LARC previously recruited in a prospective imaging study [9] approved by the medical ethics committee and for which all patients gave written informed consent before entering the study. Although all tumors may not strictly speaking be considered as locally advanced according to staging, none of the patients could be resected immediately because of the high risks of differentiation, bulkiness, and size and location of the tumor. They were therefore all treated with the same PCRT protocol as recommended for LARC patients. All patients underwent PRCT (28 fractions of 1.8 Gy, 5 fractions a week, and concomitant capecitabine, 825 mg/m<sup>2</sup>, twice a day), followed by surgery (total mesorectal excision). Patients underwent  $^{18}\text{F}$ -FDG PET/CT scans at baseline, and on Days 8 and 15 of treatment. From here onwards, these scans will be denoted as PET<sub>i</sub>, with i from 1 to 3.

### $^{18}\text{F}$ -FDG PET/CT acquisitions

The protocol was designed to ensure robust SUV measurements across all three time points. Patients

Table I. Patients' characteristics.

Parameter	Number of patients (%)
Gender	
Male	18 (64)
Female	10 (36)
Age	
Range	44–81
Median	67
TNM Stage	
T1	0 (0)
T2	1 (4)
T3	24 (85)
T4	3 (11)
N0	4 (14)
N1	8 (29)
N2	16 (57)
M0	28 (100)
M1	0 (0)
TRG status	
1	6 (21)
2	6 (21)
3	9 (33)
4	6 (21)
5	1 (4)

received an intravenous injection of FDG after a minimal fasting period of 6 h, with the activity normalized for the weight of the patient as follows:

$$(\text{weight [kg]} \times 4 + 20) \text{ [MBq]}$$

After an uptake period of exactly 60 min, all acquisitions were carried out on a Siemens Biograph 40 TruePoint scanner (Siemens Medical, Erlangen, Germany) with a spatial resolution of approximately 6 mm at full-width-at-half-maximum. Listmode data of the abdominal region were acquired in three-dimensional (3D) mode with scatter, decay, and computed tomography (CT)-based attenuation corrections, and 5 min per bed position. Images were reconstructed using Fourier rebinning and 2D OSEM (four iterations, eight subsets) with voxel size 4.1 × 4.1 × 3 mm<sup>3</sup> without post-filtering.

### Pathological tumor response grade (TRG)

The ground-truth of tumor response to therapy was determined in histology since all patients underwent total mesorectal excision. As proposed by Mandard et al. [19], TRG for each tumor was determined by an experienced pathologist blinded to the imaging data as follows: TRG 1, complete tumor response; TRG 2, residual cancer cells scattered through fibrosis; TRG 3, an increased number of residual cancer cells, with predominant fibrosis; TRG 4, residual cancer outgrowing fibrosis; and TRG 5, no regressive changes within the tumor. Tumors were subsequently grouped into responders (TRG 1–2) and non-responders (TRG 3–5).

*Partial volume effects correction*

All PET images were corrected for PVE using an iterative deconvolution method previously validated on simulated and clinical datasets [20]. This approach iteratively estimates the corrected voxels values through Lucy-Richardson deconvolution [21,22] with prior knowledge (within  $\pm 1$  mm, as it has been shown that a 1 mm error in the PSF led to a negligible impact on measured SUVs [23]) of the scanner's Point Spread Function (PSF), assumed to be spatially invariant in the field of view. In this study the tumors were all in the exact same body region and this assumption has therefore no significant impact on the applied correction on a patient-by-patient comparison basis. Wavelet-based denoising was incorporated using Bayeshrink filtering [24], applied to the residual within each iteration of the deconvolution process. This allows using a sufficient number of iterations to correct for PVE without significant noise addition. This methodology is voxel-based and therefore does not assume homogeneous regional radiotracer distributions for the tumor and/or surrounding background.

*Investigated parameters and analysis*

All parameters were extracted from the original ( $\text{PET}_{1-3}$ ) and PVE corrected ( $\text{PET}_{1-3}^{\text{PVC}}$ ) images, denoted from here onwards  $\text{param}_{1-3}$  and  $\text{param}_{1-3}^{\text{PVC}}$  respectively. For each patient, the tumor was identified on each  $\text{PET}_i$  images by a nuclear physician with more than 10 years experience and subsequently semi-automatically isolated from the bladder in a 3D region of interest (ROI) using an in-house software. This ROI, containing only the MATV and its surrounding background, was automatically transferred to the corresponding corrected  $\text{PET}_i^{\text{PVC}}$  image. MATV were subsequently delineated on both uncorrected and corrected images using an implementation of the Fuzzy Locally Adaptive Bayesian (FLAB) automatic algorithm [25] in the same software. The FLAB approach allows automatic tumor delineation by computing a probability of belonging to a given "class", e.g. tumor or background, for each voxel within the 3D ROI. This probability is calculated by taking into account the voxel's intensity with respect to the statistical distributions (characterized by their mean and variance) of the voxels in the various regions of the image, as well as its spatial correlation with neighboring voxels in 3D. This approach has been validated on simulated and clinical datasets for accuracy, robustness and reproducibility, on both homogeneous and heterogeneous MATVs [25–27].

$\text{SUV}_{\text{max}}$ , and  $\text{SUV}_{\text{mean}}$  as well as the MATV and the TLG were automatically calculated from these

delineations. MATV was defined as the sum of all voxels contained in the FLAB delineated volumes multiplied by the volume of a voxel ( $50.43 \text{ mm}^3$ ). TLG was determined by multiplying the MATV and its associated  $\text{SUV}_{\text{mean}}$ . Response indices (RIs) corresponding to one ( $\text{RI}_2$ ) and two ( $\text{RI}_3$ ) weeks were calculated as the percentage evolution with respect to the baseline value ( $\text{PET}_1$ ) as follows:  $\text{RI}_n = (\text{param}_n - \text{param}_1) / \text{param}_1 \times 100$  for  $n=2$  and  $3$ . Similarly,  $\text{RI}_{2-3}^{\text{PVC}}$  were calculated using the parameters extracted from  $\text{PET}_{1-3}^{\text{PVC}}$  images.

Figure 1 illustrates for a non-responder (Figure 1A) and a responder (Figure 1B) the baseline scan and the scan after two weeks, with the delineation of the tumor on both scans.

*Statistical analysis*

Statistical analyses were performed using Medcalc<sup>TM</sup> (MedCalc Software, Belgium). All quantitative values were expressed as means  $\pm$  standard deviations (SD) and ranges (minimum–maximum). Binary response status based on TRG classification (1–2 vs. 3–5) was correlated with baseline values ( $\text{param}_1$ ) as well as  $\text{RI}_{2-3}$  for early sequential scans ( $\text{PET}_{2-3}$ ) using a Mann-Whitney U-test. For the parameters that were found to be significantly correlated with response, the predictive performance regarding the identification of non-responders was evaluated using ROC analysis. Area under the curve (AUC), along with the best compromise between sensitivity and specificity were reported, as well as results associated with a specificity of 100%, corresponding to the clinical goal of identifying non-responders without erroneously identifying any of the responders. For  $\text{RI}_{2-3}$ , only cut-off values above +30% or below –30% were considered. This constraint was applied because the reproducibility (or test-retest assessment) of the PET derived parameters (both SUV and volume-based measurements) under investigation here has been previously determined to be at such upper and lower limits [27]. Thus characterization of response (or disease progression) based on evolution of PET derived indices must take into account such reproducibility limits. All tests were two-sided and p-values below 0.05 were considered statistically significant.

**Results**

Statistics for baseline and evolution of all derived parameters (with or without PVC) are given in Table II. Results of Mann-Whitney U-tests and subsequent ROCs analyses are provided in Supplementary Table I online at <http://www.informahealthcare.com/doi/abs/10.3109/0284186X.2012.702923> and Table III, respectively.

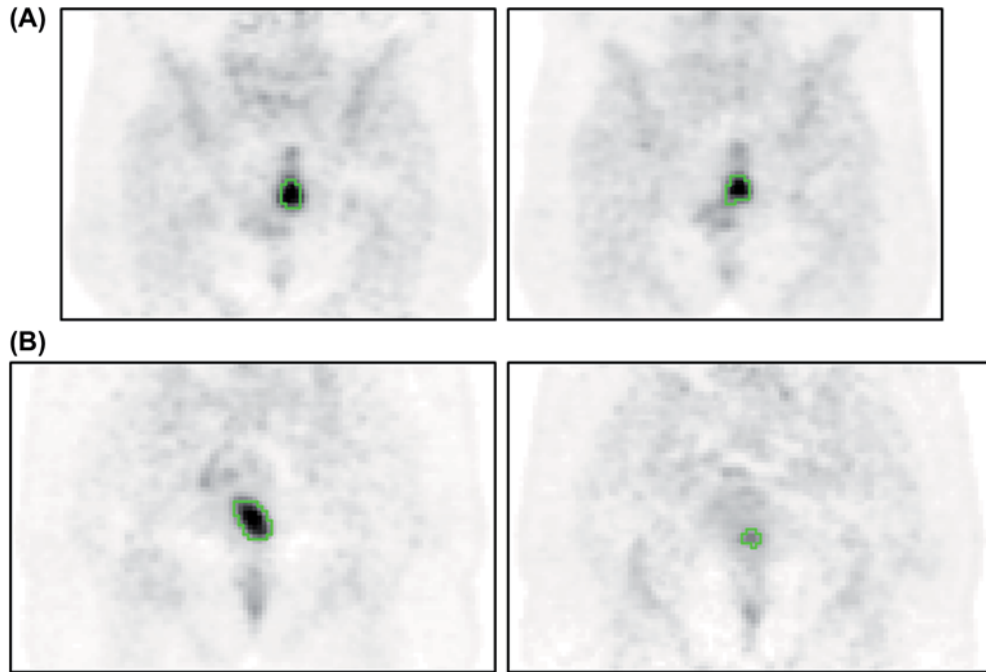


Figure 1. Illustration of the baseline scan (on the left) and the scan at 2 weeks (PET3, on the right) for (A) a responder and (B) a non-responder. The green contour is the FLAB delineation.

#### Pathologic tumor response

Six tumors were characterized by complete pathologic response (TRG 1) and six had residual cancer cells (TRG 2). Nine, six and one tumors were classified as TRG 3, 4 and 5, respectively (Table I). There were therefore 12 responders (43%) and 16 non-responders (57%).

#### Predictive value of baseline parameters

*Without PVC.* Pretreatment SUV measurements were all normally distributed, with  $SUV_{max1}$  and  $SUV_{mean1}$  values for the cohort of  $14.9 \pm 5.2$  and  $8.0 \pm 2.8$ , respectively. On the other hand,  $MATV_1$  and  $TLG_1$  were not normally distributed. Median  $MATV_1$  and  $TLG_1$  values were  $23 \text{ cm}^3$  and  $181 \text{ g}$ , respectively. Most volumes were within the range of  $10\text{--}40 \text{ cm}^3$ , with five tumors above  $100 \text{ cm}^3$ .

Responding tumors were characterized by higher baseline SUV ( $SUV_{max1} 15.6 \pm 3.7$  and  $SUV_{mean} 9.0 \pm 2.3$ ) than non-responding ones ( $SUV_{max1} 12.5 \pm 5.8$  and  $SUV_{mean} 6.4 \pm 2.9$ ). The difference was not significant for  $SUV_{max}$  ( $p = 0.06$ ) contrary to  $SUV_{mean}$  ( $p = 0.02$ ), with an associated AUC of 0.75 (Table III) leading to perfect specificity (100%) but an associated sensitivity of 50% for a cut-off value of 6.1.

No correlation between response and  $MATV_1$  or  $TLG_1$  was found ( $p > 0.1$ ), distributions of these parameters among responders and non-responders being largely overlapped (Table II).

*With PVC.* Pretreatment  $SUV_{max}$  and  $SUV_{mean}$  increased by  $+57 \pm 23\%$  and  $+27 \pm 10\%$  after correction, to  $21.2 \pm 9.7$  and  $10.2 \pm 4.0$ , respectively. On the other hand,  $MATV_1^{PVC}$  were systematically

Table II. Baseline values and evolution (RIs) of each PET derived parameter with respect to baseline values.

Parameter	Baseline value median $\pm$ SD (mean, min, max)	RI <sub>2</sub> (%) median $\pm$ SD (mean, min, max)	RI <sub>3</sub> (%) median $\pm$ SD (mean, min, max)
$SUV_{max}^{PVC}$	$13.9 \pm 5.2$ (14.9, 6.6, 26.7)	$-18 \pm 39$ (-12, -63, +155)	$-32 \pm 25$ (-31, -73, +47)
$SUV_{max}$	$21.2 \pm 9.7$ (23.8, 10.2, 49.3)	$-20 \pm 41$ (-15, -59, +165)	$-43 \pm 27$ (-38, -75, +47)
$SUV_{mean}^{PVC}$	$7.8 \pm 2.8$ (8.0, 3.7, 13.9)	$-20 \pm 36$ (-11, -59, +138)	$-29 \pm 25$ (-28, -66, +54)
$SUV_{mean}$	$10.2 \pm 4.0$ (10.3, 4.5, 20.8)	$-22 \pm 37$ (-12, -56, +153)	$-28 \pm 24$ (-29, -67, +42)
$MATV$	$23 \pm 79$ (49, 2, 397)	$-18 \pm 22$ (-15, -51, +51)	$-36 \pm 29$ (-32, -80, +42)
$MATV^{PVC}$	$19 \pm 76$ (45, 2, 378)	$-15 \pm 22$ (-15, -51, +39)	$-32 \pm 28$ (-32, -78, +37)
$TLG$	$181 \pm 473$ (345, 10, 2235)	$-35 \pm 45$ (-22, -75, +137)	$-57 \pm 37$ (-48, -86, +68)
$TLG^{PVC}$	$196 \pm 518$ (379, 13, 2385)	$-34 \pm 41$ (-24, -75, +130)	$-59 \pm 33$ (-50, -88, +51)

Table III. Predictive value of each parameter baseline values and RIs using ROCs analysis. Only parameters for which a significant correlation was found [see appendix, to be found online, at <http://www.informahealthcare.com/doi/abs/10.3109/0284186X.2012.702923>] are included.

Parameter*		AUC	Cut-off value	Sensitivity (%)	Specificity (%)	Sensitivity for 100% specificity (%)	Cut-off value
SUV <sub>max</sub>	RI <sub>3</sub> (%)	0.77	-43	88	58	31	-16
SUV <sub>max</sub> <sup>PVC</sup>	RI <sub>3</sub> (%)	0.73	-33	50	83	13	-5
SUV <sub>mean</sub>	PET <sub>1</sub>	0.75	6.1	50	100	50	6.1
	RI <sub>2</sub> (%)	0.73	-31	88	33	50	-10
	RI <sub>3</sub> (%)	0.79	-34	81	67	31	-11
SUV <sub>mean</sub> <sup>PVC</sup>	PET <sub>1</sub>	0.76	7.3	44	100	44	7.3
	RI <sub>2</sub> (%)	0.77	-33	81	25	50	-8
	RI <sub>3</sub> (%)	0.75	-42	88	58	19	-8
TLG	RI <sub>2</sub> (%)	0.75	-34	56	67	56	-23
	RI <sub>3</sub> (%)	0.79	-53	63	92	38	-37
TLG <sup>PVC</sup>	RI <sub>2</sub> (%)	0.70	-34	56	50	44	-18
	RI <sub>3</sub> (%)	0.78	-54	63	92	31	-38

smaller with however a small difference of only  $-12 \pm 6\%$  (range  $-3\%$  to  $-24\%$ ). Resulting from a combination of smaller MATV and higher SUV<sub>mean</sub><sup>PVC</sup> TLG<sub>1</sub><sup>PVC</sup> increased by  $+11 \pm 7\%$  (range  $-1\%$  to  $+31\%$ ). The increase of SUV<sub>mean</sub> was strongly correlated with MATV ( $r > 0.7$ ,  $p < 0.0001$ ) whereas increase of SUV<sub>max</sub> was not ( $r < 0.5$ ). Despite these absolute values increase, the differences between responding and non-responding tumors were not significantly altered by PVC. SUV<sub>max1</sub><sup>PVC</sup> still did not significantly separate responding from non-responding tumors ( $p = 0.07$ ) whereas SUV<sub>mean1</sub><sup>PVC</sup> allowed similar differentiation ( $p = 0.02$ ), and no correlation with response was found for MATV<sub>1</sub><sup>PVC</sup> and TLG<sub>1</sub><sup>PVC</sup> ( $p > 0.2$ ).

Regarding ROC analysis, AUCs and associated optimal sensitivities/specificities with PVC were mostly similar as without correction. For SUV<sub>mean1</sub><sup>PVC</sup>, AUC increased from 0.75 to 0.76, although it led to a reduced sensitivity (44% instead of 50%) with the same specificity of 100% (Figure 2A).

*Evolution of parameters during PRCT and associated predictive value*

*Without PVC.* There was a global trend to decreasing MATV and associated uptake of the tumors during PRCT across the cohort of patients, with further reduction at 15 days (Table II). However, these PET parameters were also found to be increasing for some patients (Table II), especially after eight days. The lowest decreases at eight days were observed for MATV and SUV<sub>max</sub> (mean  $-12\%$  and  $-15\%$ , median  $-18\%$  for both) whereas the largest decrease was observed for TLG with a mean of  $-22\%$  (median  $-35\%$ ). Decrease was systematically larger at 15 days, with for instance  $-48\%$  mean TLG decrease (median  $-57\%$ ).

On the one hand, none of the RI<sub>2</sub>s were statistically different between responders and non-responders, except for SUV<sub>mean</sub> ( $-6 \pm 44\%$  vs.  $-25 \pm 12\%$ ,  $p = 0.04$ ) and TLG ( $-18 \pm 53\%$  vs.  $-37 \pm 15\%$ ,  $p = 0.03$ ). On the other hand, RI<sub>3</sub>s of

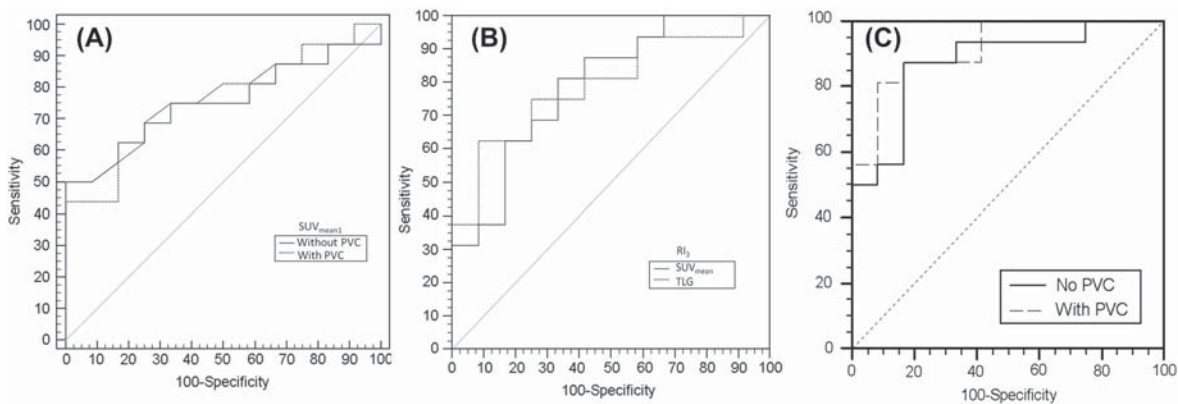


Figure 2. ROCs examples for prediction of non-responders (n = 16), using (A) SUV<sub>mean</sub> with or without PVC, (B) TLG and SUV<sub>mean</sub> RI<sub>3</sub>, and (C) logistic regression combining baseline MATV and SUV<sub>mean</sub> values and their evolution after 8 and 15 days, with and without PVC.

all the parameters except MATV were statistically different between responders and non-responders, especially for TLG and  $SUV_{mean}$  ( $p = 0.009$ ).

According to ROCs analysis,  $RI_2$ s of neither  $SUV_{mean}$  nor TLG (the only two parameters for which there was a statistical difference at eight days between responders and non-responders) allowed satisfactory prediction of TRG. With an AUC of 0.73, the optimal cut-off value above 30% change ( $-31\%$ ) for  $SUV_{mean}$  was associated with a sensitivity of 88% but a specificity of only 33%.  $RI_2$  of TLG led to an AUC of 0.75 and much lower sensitivity (56%) but higher specificity (67%) with a cut-off value of  $-34\%$ . At eight days, 100% specificity could be achieved only using too low cut-off values ( $-10\%$  only for  $SUV_{mean}$ , and  $-23\%$  for TLG) and at the cost of reduced sensitivity for  $SUV_{mean}$  (50%).

Predictive performance improved at 15 days with higher sensitivities/specificities for all parameters. For  $SUV_{max}$ , a  $-43\%$   $RI_3$  cut-off value was associated with 88% sensitivity and 58% specificity, whereas a  $-34\%$   $RI_3$  for  $SUV_{mean}$  led to 81% sensitivity and 67% specificity. Reduction of TLG using a  $RI_3$  cut-off value of  $-53\%$  led to lower sensitivity (63%) but higher specificity (92%) (Figure 2B). At 15 days, maximizing specificity was associated with low  $RI_3$  cut-off values ( $-11$  and  $-16\%$  for  $SUV_{mean}$  and  $SUV_{max}$ ), except for TLG for which the cut-off value was  $-37\%$ , with however a sensitivity of only 38%. Sensitivities associated with 100% specificity were also between 30% and 40% for the other parameters.

*With PVC.* Evolution of the parameters after 8 and 15 days was not significantly altered by PVC, with very similar values of RIs for all parameters (see Table II). PVC had therefore no impact on the statistical difference between RIs of responders and non-responders. After correction, the  $RI_2$ s still did not allow statistical differentiation between responders and non-responders for most parameters, except for TLG and  $SUV_{mean}$  which was already the case without PVC. Similarly, all the  $RI_3$ s except MATV were still able to statistically differentiate the non-responders from the responders after PVC, but the differentiation was significantly neither improved nor reduced.

Similarly, ROCs results were not significantly altered by PVC, although AUCs varied by up to  $\pm 0.05$ . Sensitivities and specificities corresponding to optimal cut-off values on the other hand were in a few cases significantly modified. For example, the sensitivity associated with  $RI_3$  of  $SUV_{max}$  was decreased after PVC from 88 to 50%, with an increase of specificity from 58 to 83%. Similar

changes were observed for the other parameters at both 8 and 15 days, without nonetheless significantly reduce or increase the predictive performance of the parameters.

## Discussion

Neither the predictive value of baseline PET images nor the one of PET features more fully describing tumors (MATV, TLG) has been extensively determined yet within the context of PCRT early prediction in rectal cancer. Melton and colleagues found in 21 LARC patients that the reduction of PET-based parameters (MATV, TLG and SUV) between baseline and four to six weeks after treatment was correlated with pathological response [17], but they did not investigate the value of baseline parameters. More recently, Chennupati and colleagues did not find a correlation between pathological response and  $SUV_{max}$  or MATV measurements at baseline or reduction after treatment in a cohort of 35 LARC patients, but they did not investigate the value of parameters reduction during the treatment [18]. Janssen and colleagues demonstrated that the reduction of  $SUV_{max}$  after two weeks was a good predictor of response, but they did not investigate other PET parameters beyond  $SUV_{mean}$  or their baseline predictive value [9]. None of these studies investigated the impact of PVC. To the best of our knowledge, the current study is the first investigation on a homogeneous LARC cohort including the impact of PVC, four PET derived parameters ( $SUV_{max}$ ,  $SUV_{mean}$ , MATV, TLG), and three (baseline and after one and two weeks)  $^{18}F$ -FDG PET scans during PCRT.

On the one hand, contrary to results in LAEC [10], neither the baseline MATV nor the derived TLG were significantly associated with response. On the other hand, some of the baseline parameters were of predictive value, such as higher  $SUV_{mean}$  being associated with responding tumors ( $p = 0.02$ ), and allowed prediction of the non-responders with a specificity of 100% but a limited sensitivity of 50%. A similar trend was observed with  $SUV_{max}$ , although without reaching statistical significance ( $p = 0.06$ ).

Overall, response was associated with higher decrease of the  $^{18}F$ -FDG PET derived indices within the first two weeks of PCRT, especially  $SUV_{mean}$  and TLG ( $p < 0.01$ ), although this decrease was a more efficient predictive factor after two weeks than one, in line with previous findings [9]. First, none of the parameters'  $RI_2$ s except for  $SUV_{mean}$  and TLG were significantly correlated with TRG. Second, an important constraint was that cut-off values of  $RI_{2-3}$ s had to be larger than  $\pm 30\%$ , because of the upper and lower reproducibility limits of such



PET derived measurements [27]. However, measured  $\text{RI}_{2-3}$ s were often within this range, especially after one week, although it was still the case even after two weeks. This may constitute one of the major limitations of early PET based response prediction, and the reproducibility of PET acquisitions clearly need to be improved in that regard. On the one hand, according to ROCs analysis, the reduction of  $\text{SUV}_{\text{mean}}$  after two weeks was associated with an AUC of 0.79, a 100% specificity and 31% sensitivity associated with a  $\text{RI}_2$  cut-off value of  $-11\%$ , which is too low with regard to the reproducibility limits, as it is the case for all other parameters. The only exception was TLG: a  $-37\%$   $\text{RI}_3$  cut-off value led to 100% specificity and 38% sensitivity, whereas a  $-53\%$  cut-off value led to the best compromise of 63% sensitivity and 92% specificity.

Usually, higher SUVs are associated with more aggressive tumors, and therefore may be potentially more resistant. However according to our results, responding tumors were found to have higher initial SUVs. This can be associated with the fact that they were also the ones exhibiting the highest uptake decrease during treatment. In addition, the tumors that were classified as non-responders based on the  $\text{RI}_3$  of TLG (specificity 100%, sensitivity 38%) are almost the same as those classified as non-responders based on their baseline mean SUVs (specificity 100%, sensitivity 50%). Therefore waiting for two weeks did not significantly ( $p > 0.05$ ) improve prediction of response in this context when maximizing specificity, despite increased overall AUCs, which is a new result with respect to previous findings that only investigated the predictive value of RIs and not absolute baseline values [9].

The classification could be further improved using logistic regression of several parameters such as baseline MATV and  $\text{SUV}_{\text{mean}}$  values and their evolution at 8 and 15 days, resulting in an AUC of 0.88, with a sensitivity and specificity of 88% and 83% respectively, whereas a specificity of 100% was associated with 50% sensitivity (Figure 2C). Similar results were obtained with PVC (AUC 0.92, 92% specificity and 81% sensitivity, 56% sensitivity associated with 100% specificity). However, such weighted model combining more than one parameter on 28 patients only is likely to be over fitted to the data and would require validation on a larger cohort.

Regarding PVC, despite significant impact on the absolute values (except for MATV), it had limited impact on the predictive value of the parameters. Similar results regarding the impact of PVC on baseline predictive value were obtained for baseline prediction of chemoradiotherapy response in LAEC [28]. This is the first study however investigating the PVC impact on the predictive value

of sequential scans. This lack of impact may be explained by the fact that although the tumor volumes significantly shrunk during the first two weeks of PCRT ( $-36 \pm 29\%$  after two weeks), their volume were still large (mean  $36 \text{ cm}^3$ ). Since PVE are significant for volumes below  $10\text{--}15 \text{ cm}^3$ , PVC impact should be investigated for the follow-up of tumors exhibiting larger changes and/or smaller volumes.

Some limits of this study have to be emphasized. It is first limited by its retrospective nature and the small number of patients, which led to group the five different classifications of TRG into responders and non-responders. Investigation of other response classification such as complete regression (TRG1) vs. non-response (TRG 2–5) or all five TRG statuses should be conducted in larger prospective studies in order to validate our preliminary results. Finally, the TRG classification, although it constitutes the best measurement of tumor response available, was determined by one pathologist only.

## Conclusion

Our results in this preliminary study suggest that early prediction of tumor response to PRCT in rectal cancer using sequential  $^{18}\text{F}$ -FDG PET scans is challenging due to limited reproducibility of PET imaging, especially after one week that was found to be suboptimal in both robustness due to small evolutions of the parameters and accuracy with no correlation with response. The best compromise between accuracy and robustness of prediction was obtained by considering the reduction of total lesion glycolysis after two weeks of PCRT, with a sensitivity of 63% and specificity of 92%, although the improvement over baseline absolute mean SUV prediction (specificity 100%, sensitivity 50%) was not significant. Finally, partial volume effects correction had no impact on the predictive value of neither the baseline absolute values nor their evolution during treatment. The results of this study require validation in a larger cohort allowing consideration of less restrictive response measures.

**Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## References

- [1] Martin ST, Heneghan HM, Winter DC. Systematic review and meta-analysis of outcomes following pathological complete response to neoadjuvant chemoradiotherapy for rectal cancer. *Br J Surg* 2012;99:918–28.
- [2] Kusters M, Valentini V, Calvo FA, Krempien R, Nieuwenhuijzen GA, Martijn H, et al. Results of European

- pooled analysis of IORT-containing multimodality treatment for locally advanced rectal cancer: Adjuvant chemotherapy prevents local recurrence rather than distant metastases. *Ann Oncol* 2010;21:1279–84.
- [3] Maas M, Nelemans PJ, Valentini V, Das P, Rodel C, Kuo LJ, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: A pooled analysis of individual patient data. *Lancet Oncol* 2010;11:835–44.
  - [4] Zhang C, Tong J, Sun X, Liu J, Wang Y, Huang G. (18)F-FDG-PET evaluation of treatment response to neo-adjuvant therapy in patients with locally advanced rectal cancer: A meta-analysis. *Int J Cancer Epub* 2012 Mar 24.
  - [5] Capirci C, Rampin L, Erba PA, Galeotti F, Crepaldi G, Banti E, et al. Sequential FDG-PET/CT reliably predicts response of locally advanced rectal cancer to neo-adjuvant chemo-radiation therapy. *Eur J Nucl Med Mol Imaging* 2007;34:1583–93.
  - [6] Vliegen RF, Beets-Tan RG, Vanhauten B, Driessen A, Oellers M, Kessels AG, et al. Can an FDG-PET/CT predict tumor clearance of the mesorectal fascia after preoperative chemoradiation of locally advanced rectal cancer? *Strahlenther Onkol* 2008;184:457–64.
  - [7] Amthauer H, Denecke T, Rau B, Hildebrandt B, Hunerbein M, Ruf J, et al. Response prediction by FDG-PET after neoadjuvant radiochemotherapy and combined regional hyperthermia of rectal cancer: Correlation with endorectal ultrasound and histopathology. *Eur J Nucl Med Mol Imaging* 2004;31:811–9.
  - [8] Cascini GL, Avallone A, Delrio P, Guida C, Tatangelo F, Marone P, et al. 18F-FDG PET is an early predictor of pathologic tumor response to preoperative radiochemotherapy in locally advanced rectal cancer. *J Nucl Med* 2006;47:1241–8.
  - [9] Janssen MH, Ollers MC, Riedl RG, van den Bogaard J, Buijsen J, van Stiphout RG, et al. Accurate prediction of pathological rectal tumor response after two weeks of preoperative radiochemotherapy using (18)F-fluorodeoxyglucose-positron emission tomography-computed tomography imaging. *Int J Radiat Oncol Biol Phys* 2010;77:392–9.
  - [10] Haustermans K, Debucquoy A, Lambrecht M. The ESTRO Breur Lecture 2010: Toward a tailored patient approach in rectal cancer. *Radiother Oncol* 2011;100:15–21.
  - [11] 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;38:1595–606.
  - [12] Larson SM, Erdi Y, Akhurst T, Mazumdar M, Macapinlac HA, Finn RD, et al. Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging. The visual response score and the change in total lesion glycolysis. *Clin Positron Imaging* 1999;2:159–71.
  - [13] Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med* 2011;52:369–78.
  - [14] Cazaentre T, Morschhauser F, Vermandel M, Betrouni N, Prangere T, Steinling M, et al. Pre-therapy 18F-FDG PET quantitative parameters help in predicting the response to radioimmunotherapy in non-Hodgkin lymphoma. *Eur J Nucl Med Mol Imaging* 2010;37:494–504.
  - [15] Lee HY, Hyun SH, Lee KS, Kim BT, Kim J, Shim YM, et al. Volume-based parameter of (18)F-FDG PET/CT in malignant pleural mesothelioma: Prediction of therapeutic response and prognostic implications. *Ann Surg Oncol* 2010;17:2787–94.
  - [16] El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit* 2009;42:1162–71.
  - [17] Melton GB, Lavelly WC, Jacene HA, Schulick RD, Choti MA, Wahl RL, et al. Efficacy of preoperative combined 18-fluorodeoxyglucose positron emission tomography and computed tomography for assessing primary rectal cancer response to neoadjuvant therapy. *J Gastrointest Surg* 2007;11:961–9; discussion 969.
  - [18] Chennupati SK, Quon A, Kamaya A, Pai RK, La T, Krakow TE, et al. Positron emission tomography for predicting pathologic response after neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Am J Clin Oncol Epub* 2011 Mar 17.
  - [19] Mandard AM, Dalibard F, Mandard JC, Marnay J, Henry-Amar M, Petiot JF, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer* 1994;73:2680–6.
  - [20] Boussion N, Cheze Le Rest C, Hatt M, Visvikis D. Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging. *Eur J Nucl Med Mol Imaging* 2009;36:1064–75.
  - [21] Lucy LB. An iterative technique for the rectification of observed distributions. *Astronom J* 1974;79:745.
  - [22] Richardson WH. Bayesian-based iterative method of image restoration. *J Optical Soc Am* 1972;62:55.
  - [23] Teo BK, Seo Y, Bacharach SL, Carrasquillo JA, Libutti SK, Shukla H, et al. Partial-volume correction in PET: Validation of an iterative postreconstruction method with phantom and patient data. *J Nucl Med* 2007;48:802–10.
  - [24] Chang SG, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans Image Process* 2000;9:1532–46.
  - [25] Hatt M, Cheze le Rest C, Descourt P, Dekker A, De Ruysscher D, Oellers M, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys* 2010;77:301–8.
  - [26] Hatt M, Cheze Le Rest C, Albarghach N, Pradier O, Visvikis D. PET functional volume delineation: A robustness and repeatability study. *Eur J Nucl Med Mol Imaging* 2011;38:663–72.
  - [27] Hatt M, Cheze-Le Rest C, Aboagye EO, Kenny LM, Rosso L, Turkheimer FE, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements. *J Nucl Med* 2010;51:1368–76.
  - [28] Hatt M, Le Pogam A, Visvikis D, Pradier O, Cheze le Rest C. Impact of partial volume effects correction on the predictive and prognostic value of baseline 18F-FDG PET images in esophageal cancer. *J Nucl Med* 2012;53:12–20.

**Supplementary material available online**

Supplementary Appendix