Comparison Between 18F-FDG PET Image-Derived Indices for Early Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer.
Mathieu Hatt, David Groheux, Antoine Martineau, Marc Espié, Elif Hindié, Sylvie Giacchetti, Anne De Roquancourt, Dimitris Visvikis, Catherine Cheze Le Rest

To cite this version:
Mathieu Hatt, David Groheux, Antoine Martineau, Marc Espié, Elif Hindié, et al.. Comparison Between 18F-FDG PET Image-Derived Indices for Early Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer.. Journal of Nuclear Medicine, Society of Nuclear Medicine, 2013, 54 (3), pp.341-9. <10.2967/jnumed.112.108837>. <hal-00748923>

HAL Id: hal-00748923
https://hal.archives-ouvertes.fr/hal-00748923
Submitted on 6 Nov 2012
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Comparison between 18F-FDG PET image derived indices for early prediction of response to neoadjuvant chemotherapy in breast cancer

Mathieu Hatt¹, David Groheux²,³, Antoine Martineau², Marc Espié⁴, Elif Hindié⁵, Sylvie Giacchetti⁴, Anne de Roquancourt⁶, Dimitris Visvikis¹, Catherine Cheze-Le Rest¹,⁷.

¹ INSERM, UMR 1101 LaTIM, Brest, France.
² Department of Nuclear Medicine, Saint-Louis Hospital, Paris, France.
³ B2T, Doctoral School, IUH, University of Paris VII, France.
⁴ Department of Medical Oncology, Breast Diseases Unit, Saint-Louis Hospital, Paris, France.
⁵ Department of Nuclear Medicine, Haut-Lévêque Hospital, University of Bordeaux, France.
⁶ Department of Pathology, Saint-Louis Hospital, Paris, France.
⁷ Department of Nuclear Medicine, CHU Milétrie, Poitiers, France.

Corresponding author:
Mathieu HATT
INSERM UMR 1101 LaTIM
CHRU Morvan, 5 avenue Foch
29609 Brest, France

short running foot line: sequential ¹⁸F-FDG PET for breast cancer
ABSTRACT

The goal of this study was to determine the best predictive factor among image-derived parameters extracted from sequential $^{18}$F-FDG Positron Emission Tomography (PET) scans for early tumor response prediction after two cycles of neoadjuvant chemotherapy (NAC) in breast cancer.

Methods: 51 breast cancer patients were included. Responders and non-responders status were determined in histopathology according to the tumor and nodes Sataloff scale. PET indices ($SUV_{max}$ and $SUV_{mean}$, metabolically active tumor volume and total lesion glycolysis (TLG)), at baseline and their evolution ($\Delta$) after two cycles of NAC were extracted from the PET images. Their predictive value was investigated using Mann-Whitney-U tests and ROC analysis. Sub-group analysis was also carried out by considering ER-positive/HER2-negative, triple-negative, and HER2-positive tumors separately. The impact of partial volume correction (PVC) was also investigated using an iterative deconvolution algorithm.

Results: There were 24 pathological non-responders and 27 responders. None of the baseline PET parameters was correlated with response. After two NAC cycles, the reduction of each parameter was significantly associated with response, the best prediction of response being obtained with $\Delta$TLG (96% sensitivity, 92% specificity, 94% accuracy) with significantly higher AUC (0.91 vs. 0.82, $p=0.01$) than $\Delta$SUV$_{max}$ (63% sensitivity, 92% specificity, 77% accuracy). Sub-group analysis confirmed a significantly higher accuracy for $\Delta$TLG than $\Delta$SUV for ER-positive/HER-negative, but not for triple-negative and HER2-positive tumors. PVC had no impact on the predictive value of any of the PET image derived parameters despite significant changes of their absolute values.
Conclusion: Our results suggest that the reduction after two NAC cycles of the metabolically active volume of primary tumor measurements such as ΔTLG predicts histopathological tumor response with higher accuracy than ΔSUV measurements, especially for ER-positive/HER2-negative breast cancer. These results should be confirmed in a larger group of patients as they may potentially increase the clinical value and efficiency of $^{18}$F-FDG PET for early prediction of response to NAC.

Keywords: Breast cancer, Neoadjuvant chemotherapy, $^{18}$F-FDG, tumor delineation, pathological response
Introduction

Preoperative neoadjuvant chemotherapy (NAC) has been used as standard treatment for inflammatory and non-operable locally advanced breast carcinoma patients and it is now increasingly being used for patients with operable but large breast tumors. This strategy allows patients to undergo breast-conserving surgery (BCS) and provides information regarding the efficacy of chemotherapy (1). Early response prediction after one or two cycles of NAC might enable the selection of alternative treatment strategies (2). Breast carcinoma is composite and immunohistochemistry allows defining three main subgroups with different therapeutic response and different outcome (triple negative, HER2+ and luminal tumors). Pathological Complete Response (pCR) is associated with a better outcome in the HER2-overexpressed and triple negative breast cancer patients. On the other hand, recent studies showed than in luminal tumors, especially for luminal A, the impact of pCR on patient’s survival remains less established (3, 4). Thus an intermediate response with tumor shrinkage allowing BCS might be considered as a reasonable clinical objective for this group.

Within this context, ¹⁸F-FDG PET imaging has been demonstrated as a potent predictive tool (5-8). Indeed, correlations between the pathological tumor response after completion of NAC and the decrease of tumor standardized uptake values (SUV) after one or two courses of chemotherapy have been demonstrated in several studies. A recent meta-analysis (8) conducted on 19 studies including 920 patients for the early prediction of primary tumor response to NAC reported a pooled sensitivity of 84% and specificity of 66% in identifying responders. The authors emphasized that the low pooled specificity (66%) still calls for caution.
Most studies have considered SUV measurements only; mostly SUV$_{\text{max}}$, or SUV$_{\text{peak}}$ averaging the SUV$_{\text{max}}$ voxel with its neighboring voxels (9). On the other hand, it has been demonstrated in several recent studies and for various malignancies that other $^{18}$F-FDG PET image derived parameters can have statistically significant higher predictive value than SUV in determining tumor response (10, 11). These parameters, which allow for a more comprehensive tumor functional level evaluation, include metabolically active tumor volume (MATV) (11) and total lesion glycolysis (TLG), defined as the product of MATV and its associated mean SUV (SUV$_{\text{mean}}$) (12). One recent study compared SUV$_{\text{max}}$ and TLG derived using manual delineation and threshold, showing that SUV$_{\text{max}}$ had better predictive value than TLG in identifying pCR (13).

To the best of our knowledge, no study has investigated the predictive value, regarding response to NAC for breast cancer, for all of the previously described $^{18}$F-FDG PET image derived parameters (SUV$_{\text{max}}$, SUV$_{\text{mean}}$, SUV$_{\text{peak}}$, TLG, MATV) within the same study, at both the baseline and during treatment time points.

The current study was therefore conducted with the objective of determining the predictive value of several $^{18}$F-FDG PET derived parameters both at baseline and after two NAC cycles.

**MATERIALS AND METHODS**

**Patient Population**

The current study consists of a retrospective analysis of a prospective cohort of 55 consecutive patients diagnosed with breast cancer included in a previous clinical trial (14), performed in agreement with guidelines of the institutional ethical committee with patients' informed consent. PET/CT image datasets of 3 patients could not be
retrieved from the database due to a corrupt archive file, and one was excluded from the analysis due to insufficient initial uptake ($\text{SUV}_{\text{max}}=1.5$) in the primary tumor. Therefore 51 patients were included in the retrospective analysis presented in this work (table 1).

All patients underwent NAC with four cycles of epirubicin + cyclophosphamide, followed by four cycles of docetaxel (+Herceptin in case of HER2+++). Patients underwent $^{18}$F-FDG PET/CT scans at baseline and after the second cycle of NAC. From here onwards, these scans will be denoted as PET$_1$ and PET$_2$ respectively. At completion of chemotherapy, all patients underwent surgery (mastectomy or lumpectomy).

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

All $^{18}$F-FDG PET/CT scans were performed at Saint Louis Hospital in Paris between July 2007 and May 2009. A rigorous imaging protocol was designed to ensure robust SUV measurements across both time points. Blood glucose level had to be $\leq 7$ mmol/L. For both acquisitions, patients received an intravenous injection (in the arm opposite to the breast tumor using a venous line) of $^{18}$F-FDG (5 MBq/kg) after a fasting period of 6 hours. Following an uptake period of 60 minutes, all acquisitions were carried out from mid-thigh level to the base of the skull with the arms raised, on a Philips Gemini XI PET/CT that combines a germanium oxyorthosilicate–based PET scanner and a 16-slice Brilliance CT scanner. CT data were acquired first (120 kV, 100 mAs, no contrast enhancement). PET emission list mode data were acquired in 3D mode, with 2 min per bed position, and reconstructed using a 3D row-action maximum-likelihood (RAMLA) algorithm with voxel size of $4 \times 4 \times 4 \text{mm}^3$. The
attenuation corrected images were normalized for injected dose and body weight and
converted into standardized uptake values (SUVs) defined as:
tracer concentration [kBq/mL] / (injected activity [kBq] / patient body weight [g])

**Tumor Histology and Immunohistochemistry Analysis**

Tumor type was determined on the core needle biopsy performed before neoadjuvant chemotherapy. Immunohistochemical tests were performed on formalin-fixed, paraffin embedded tissues, using specific antibodies and an automated immunostainer (XT Immunostainer; Ventana). Tumors were considered to overexpress HER2 (HER2++) if more than 30% of invasive tumor cells showed definite membrane staining resulting in a so-called fishnet appearance; control by FISH (fluorescence in situ hybridization) or SISH (silver enhanced in situ hybridization) was done for ambiguous cases. Tumors were considered estrogen receptor (ER)–negative or progesterone receptor (PR)–negative if there was less than 10% staining.

Three specific tumor subgroups as described previously were considered, namely triple-negative, ER-positive/HER2-negative (luminal) and HER2-positive.

**Pathological Tumor Response**

Histopathological response was assessed on surgical specimens at completion of NAC. Response was graded according to the Sataloff scale in primary tumor (T) and nodes (N) (15), that was used in the frequently cited study by Rousseau et al (6):

TA: total or nearly-total therapeutic effect, TB: >50% therapeutic effect but less than total or nearly-total effect, TC: <50% therapeutic effect but visible effect, TD: no therapeutic effect, NA: evidence of therapeutic effect and no residual disease, NB: no
node metastases or therapeutic effect, NC: evidence of a therapeutic effect but metastasis still present, ND: metastasis still present and viable, no therapeutic effect. Complete and partial responders [T(A-B) with N(A-B-C)] were considered as histopathological responders, whereas patients with no response or progression [T(C-D), N(D)] were considered as non-responders.

The pCR rate, defined as absence of invasive cancer cells in the primary tumor and in lymph nodes (16) was also evaluated and the predictive power of PET parameters to predict pCR was measured.

**Investigated Parameters and Analysis**

All PET image-derived parameters were extracted from the PET$_1$ and PET$_2$ images. For each patient, the primary tumor was identified on the PET image by a nuclear medicine physician with more than 10 years experience and subsequently semi-automatically isolated in a 3D region of interest (ROI) containing the tumor and its surrounding background. Tumors were subsequently automatically delineated using the Fuzzy Locally Adaptive Bayesian (FLAB) algorithm (17) applied to the previously defined ROI. 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 previously validated on simulated and clinical datasets for accuracy, robustness and reproducibility, on both homogeneous and heterogeneous MATVs (17-19).
Potential impact of partial volume effects (PVE) was also investigated by correcting the image before analysis using a state-of-the-art iterative deconvolution previously validated for oncology PET imaging applications (20). The analysis was subsequently carried out on both the original and PVE corrected (PVC) images.

SUV\textsubscript{max}, SUV\textsubscript{peak} and SUV\textsubscript{mean} as well as the MATV and the TLG were then automatically calculated from the tumor delineations. SUV\textsubscript{peak} was defined as the mean of voxels intensities in a 1.7 cm\textsuperscript{3} spheric ROI (27 voxels) centered on SUV\textsubscript{max} (9). MATV was defined as the sum of all voxels contained in the FLAB delineated volumes multiplied by the volume of a voxel (64 mm\textsuperscript{3}). SUV\textsubscript{mean} was defined as the mean of voxel intensities in the MATV delineated by FLAB. Subsequently, TLG was determined by multiplying the MATV and its associated SUV\textsubscript{mean}. The percentage evolution of each parameter between baseline and after the second NAC cycle (\(\Delta\text{param, }%\)) was calculated as:

\[
\Delta\text{param} = \frac{(\text{param}_{\text{PET}_2} - \text{param}_{\text{PET}_1})}{\text{param}_{\text{PET}_1}} \times 100.
\]

**Statistical Analysis**

Statistical analyses were performed using MedCalc\textsuperscript{TM} (MedCalc Software, Belgium). All parameters’ distributions were expressed as median (or mean depending on normality) ± standard deviations (SD) and range (minimum, maximum). Normality was tested using D’Agostino-Pearson test (21). For each parameter, its correlation with patient response was carried out by testing the statistical difference between responders’ and non-responders’ distributions using a Mann-Whitney U test. For each parameter, their absolute values at PET\textsubscript{1} and PET\textsubscript{2}, as well as their evolution \(\Delta(\text{PET}_1, \text{PET}_2)\) were investigated. The predictive performance regarding the identification of responders was evaluated using receiver operating characteristic
(ROC) analysis. Area under the curve (AUC), sensitivity, specificity, and accuracy were reported. Absolute and associated predictive parameters values extracted from the PVC PET images were compared to original ones. Sub-group analysis was also carried out by considering triple-negative, HER2-positive, and ER-positive/HER2-negative tumors separately. All tests were two-sided and p values ≤0.05 were considered statistically significant.

**RESULTS**

ER, PR and HER2 receptors status are provided in table 1. 13 patients had triple negative tumors, 12 had HER2-positive tumors, and 26 had ER-positive/HER2-negative tumors, which is representative (22). Considering the pathological tumor response established at surgery (after completion of chemotherapy), there were 27 pathological responders (53%) and 24 non-responders (47%). pCR was found in 7 patients (13.7%), among which, 3 had triple negative, and 4 had HER2-overexpressing tumors, and 0 were ER-positive/HER2-negative. Figure 1 illustrates for a pathological non-responder and a responder the two scans, including the delineation of the primary tumor.

**Predictive Value of Parameters’ Absolute Values**

According to the D’Agostino-Pearson test, considering the entire patient cohort (n=51), none of the parameters’ absolute values at baseline and after the second cycle or associated variation (Δ) were normally distributed (table 2). Similarly, the distribution of responders and non-responders groups were not normally distributed (figure 2), except for the variation (Δ) values (figure 3). According to Mann-Whitney-U tests, no significant correlation (p>0.1) was observed between the baseline values and histopathological response (table 2), distributions of
the parameters among responders and non-responders being largely overlapped (figure 2A), leading to low AUC of 0.52-0.63 (table 3).

Primary tumors of non-responders were characterized by higher PET$_2$ SUV$_{\text{max}}$ (5.3±6.2 vs. 2.8±2.9, $p=0.04$) and TLG (30±160 vs. 14±13, $p=0.05$). However, no correlation ($p\geq0.07$) was found for all of the other parameters considered (table 2).

For all parameters, the distributions were largely overlapped (figure 2A-C), as the results of ROC analysis demonstrate, with low AUC of 0.62-0.67 (table 3). For SUV, sensitivity (57-69%) and specificity (71-75%) were limited. On the other hand, for volume-based measurements, sensitivity was higher (89% and 78% for MATV and TLG respectively) but specificity was lower (42% and 63% for MATV and TLG respectively).

**Evolution of Parameters and Associated Predictive Value**

There was a global trend of decreasing MATV and associated tumors’ uptake after the first two NAC cycles, although for some patients PET image derived parameters were also found to be increasing (table 2). The lowest and largest decrease were observed for $\Delta$SUV$_{\text{max}}$ (-34%±32%) and $\Delta$TLG (-59%±34%) respectively.

According to Mann-Whitney-U tests, the variation ($\Delta$) of all parameters was statistically different between responders and non-responders, especially for $\Delta$MATV and $\Delta$TLG ($p<0.0001$), as well as $\Delta$SUV$_{\text{max}}$ ($p=0.0001$) (table 2).

According to ROC analysis, the best prediction was achieved using the $\Delta$TLG, with a sensitivity of 96% and specificity of 92% resulting in an accuracy of 94% (table 3). The $\Delta$MATV was also an accurate predictive factor (sensitivity 93%, specificity 88%, accuracy 91%). On the other hand, $\Delta$SUV measurements led to significantly lower AUCs (0.68 for $\Delta$SUV$_{\text{mean}}$ to 0.82 for $\Delta$SUV$_{\text{max}}$ vs. 0.91 and 0.92 for $\Delta$MATV and $\Delta$TLG respectively).
ΔTLG, p≤0.01) (figure 4A), with significantly lower sensitivity (63-74%), specificity (67-92%) and resulting accuracy (71-77%) (table 3). The least overlapped distributions were those associated with ΔMATV and ΔTLG (figure 3). The optimal cut-off values maximizing sensitivity and specificity were -48%, -42%, -30%, -42% and -56% for ΔSUV$_{\text{max}}$, ΔSUV$_{\text{peak}}$, ΔSUV$_{\text{mean}}$, ΔMATV and ΔTLG respectively (table 3, figure 3).

**Tumor Subgroup Analysis**

When considering ER-positive/HER2-negative patients (n=26), all AUCs were slightly higher but with the same hierarchy as for the entire cohort: ΔTLG and ΔMATV resulted in AUC of 0.96 and 0.98 respectively, whereas ΔSUV measurements led to AUC of 0.69, 0.84 and 0.88 for ΔSUV$_{\text{mean}}$, ΔSUV$_{\text{peak}}$ and ΔSUV$_{\text{max}}$ respectively.

The ROC analysis performed on the 12 HER2-positive and 13 triple-negative patients resulted in non-statistically significant different AUCs for volume-based and SUVs ($p>0.05$) in both cases.

**Prediction of pCR**

The pCR predictive value of all parameters was reduced with respect of the prediction of partial response. The hierarchy was however similar with AUCs of 0.76, 0.79, 0.67, 0.63 and 0.59 for ΔTLG, ΔMATV, ΔSUV$_{\text{max}}$, ΔSUV$_{\text{peak}}$, and ΔSUV$_{\text{mean}}$ respectively (figure 5A).

**Impact of Partial Volume Effects Correction**

The correction of partial volume effects had no significant impact on the resulting variation (Δ) of neither SUV nor volume-based parameters, despite the significant
impact on their PET$_1$ and PET$_2$ absolute values. There were therefore no statistically
significant differences between the AUCs of ROC curves generated using $\Delta$(%) calculated using the original and PVC PET image derived parameters (figure 4B, 5B).

DISCUSSION

The current study is the first investigation into the predictive value of $^{18}$F-FDG PET derived parameters on a breast cancer cohort, including baseline three different SUV measurements, MATV and TLG and their evolution ($\Delta$). Our results are in contradiction with the ones of the only other study to date comparing the performance of $\Delta$SUV$_{\text{max}}$ and $\Delta$TLG, that reported a higher predictive value for $\Delta$SUV$_{\text{max}}$ (13). In this study, $\Delta$SUV$_{\text{mean}}$, $\Delta$SUV$_{\text{peak}}$ or $\Delta$MATV predictive values were not considered. In addition, $\Delta$TLG values were derived from tumor volumes delineated through manual contouring using a fixed threshold. Such a delineation approach has been previously shown to be inaccurate for PET imaging (23, 24). In our study a previously validated robust and reproducible MATV segmentation algorithm was used. Another difference is that results in the previously reported study were reported for the prediction of pCR only, corresponding to 17% of the 142 patients, with a resulting cut-off $\Delta$SUV$_{\text{max}}$ value of 83-88%, which led to high specificity (96%) but low sensitivity (67%). In our study, we considered as pathological responders patients with complete or partial response in the primary breast cancer and lymph nodes. According to this criterion we found that MATV and TLG were more predictive than SUV. We refined a subsequent analysis considering pCR only for responders and we observed also than volume parameters were superior (AUC 0.76-0.79) to SUV (0.59-0.67).
In contrast to results in several malignancies such as locally advanced esophageal cancer (11) or non-Hodgkin lymphoma (10), none of the absolute baseline values (with or without PVC) was significantly associated with response. The absolute values of some parameters (SUV$_{\text{max}}$ and TLG) after the second cycle (PET$_2$) were significantly correlated with response but had limited predictive value. The most powerful predictive factors were the evolution between the baseline and the second scan ($\Delta$).

Pathological responders were associated with a significantly higher decrease of the $^{18}$F-FDG PET derived indices considered. Among these, the $\Delta$TLG and $\Delta$MATV were the best predictive factors compared to $\Delta$SUV measurements, with significantly higher accuracy (91% and 94% for MATV and TLG vs. 71%, 76% and 77% for $\Delta$SUV$_{\text{mean}}$, $\Delta$SUV$_{\text{peak}}$ and $\Delta$SUV$_{\text{max}}$ respectively). With respect to the recent meta-analysis that reported an overall 84% sensitivity and 66% specificity based on $\Delta$SUV$_{\text{max}}$ (8), our $\Delta$SUV$_{\text{max}}$ results demonstrated lower sensitivity (63%) but higher specificity (92%), whereas $\Delta$TLG resulted in both higher sensitivity (96%) and specificity (92%). The meta-analysis also demonstrated large differences in sensitivities and specificities among the various considered studies due to combinations of different patient populations, response criteria (pCR and/or partial response) and cut-off values. Our study is in line with previous findings that the reduction ($\Delta$) of PET indices after two cycles are good predictors of response to NAC in breast cancer (8). We further demonstrated the added value of considering more complete tumor characterization in $^{18}$F-FDG PET images through the delineation of the MATV over SUV measurements only. On the one hand, these results were mostly unchanged when performing the analysis on the subgroup of ER+/HER2-
patients. On the other hand, the respective predictive values of $\Delta$SUV, $\Delta$MATV and $\Delta$TLG parameters were similar for the triple-negative and HER2-positive patients. However, since this subgroup analysis included a small number of patients (13 triple-negative and 12 HER2-positive), it needs to be further validated in a larger patient population. Future validation studies on larger groups of patients will be focused on ER-positive/HER2-negative patients since it is for this sub-group that the added value of $\Delta$MATV and $\Delta$TLG over $\Delta$SUV seems to be the most significant. This can be explained by the fact that ER-positive/HER2-negative tumors are known to exhibit lower initial FDG uptake and have therefore lower decrease margins (7, 25, 26). These tumors are also known to only partially respond to NAC, requiring finer characterization of their response (27, 28). Therefore, the addition of functional tumor volume-based metrics logically provides improved predictive accuracy for these tumors. For triple-negative and HER2-positive tumors that exhibit higher FDG uptake and associated decrease margins, as well as better pCR rates, the addition of volume-based metrics might not translate into significantly improved predictive value over $\Delta$SUV measurements.

Another interesting result of our study was that $\Delta$SUV$_{\text{mean}}$ had significantly lower predictive value than $\Delta$SUV$_{\text{peak}}$ and $\Delta$SUV$_{\text{max}}$ (AUC of 0.68 vs. 0.79 and 0.82, $p<0.02$). Considering the use of $\Delta$SUV$_{\text{peak}}$ relative to $\Delta$SUV$_{\text{max}}$, as proposed by the PERCIST recommendations (9), in this study there were no statistically significant differences between the two (AUC 0.82 vs. 0.79, $p=0.2$), despite $\Delta$SUV$_{\text{max}}$ leading to slightly better results than $\Delta$SUV$_{\text{peak}}$. Both these parameters describe the highest activity region of the tumor, whereas the mean value is a better representative measurement of the entire tumor. A potential explanation for the lower predictive
value of $\Delta \text{SUV}_{\text{mean}}$ is its higher dependency on partial volume effects compared to $\text{SUV}_{\text{max}}$ and $\text{SUV}_{\text{peak}}$. It has however been recently shown in a group of 15 breast cancer patients that PVC had limited impact on the SUV reduction (29), without reporting however on the actual impact of this reduction on the overall response predictive value. Another study investigated the impact of PVC on the prediction of response to NAC for breast cancer using dynamic $^{18}\text{F}$-FDG PET. They found that PVC eliminated significant differences in % FDG uptake measurement changes for non-responders (NR) versus partial responders (PR), but not for pCR versus other response categories (30). In the aforementioned work PVC was performed using simplistic recovery coefficients, applied only to tumors with size $<3\text{cm}$ based on MR imaging. In our study, PVC was systematically applied to all images using a voxelwise iterative image deconvolution algorithm previously validated for PET images (20). In this work, the use of PVC did not change either the predictive value of the parameters or their evolution ($\Delta$) or the statistical differences between groups of response, despite a significant change of their absolute values. Finally, PVC did not improve the predictive value of $\text{SUV}_{\text{mean}}$ or the hierarchy of performance between the three different SUV measurements, for neither the partial or pCR response criteria. This lack of PVC impact is consistent with our previous findings (31, 32).

The delineation of breast cancer MATVs is challenging as they often exhibit heterogeneous uptake distributions and complex shapes and/or low tumor-to-background ratios, particularly in the mid-treatment scan. Therefore the use of a robust algorithm is recommended. Within this context, the results presented here using FLAB may be replicated using alternative approaches, such as gradient-based (33) or improved fuzzy C-means (34) methods. Using less robust approaches might
lead to significant differences in the results concerning the value of volume derived parameters \((11, 13)\). In addition, this effect may be more important when considering the temporal evolution of such parameters, as measurements have to be considered with respect to the known physiological reproducibility ranges. Difference values have to be larger than ±30% to characterize response, because of the upper and lower physiological reproducibility limits associated with PET derived measurements \((19)\). Considering MATV and TLG values, these reproducibility limits were obtained using the FLAB method, whereas less robust approaches (fixed and adaptive threshold) lead to larger values \((±50-90\%)\) \((19)\). The FLAB method also allows for more repeatable measurements, with low inter- and intra-observer variability \((18, 19)\). Finally, the imaging protocol in this study was specifically designed to ensure robust and reproducible tumor characterization between both time points. For most patients, measured evolutions were outside the reproducibility limits. Optimal cut-off values were also larger; -48% for \(SUV_{\text{max}}\), which is in accordance to previous findings \((8)\), -42% and -56% for MATV and TLG respectively.

One potential limitation of this study is its retrospective nature and the resulting potential bias, and the lack of reliable outcome information due to the short delay between surgery and last follow-up. Another limitation is the relatively small number of patients due to the previous prospective study that was limited to 55 patients \((14)\), 4 of which could not be exploited in the present study. This small sample is especially restrictive regarding the subgroup analyses of triple negative and HER2-positive patients. The findings of this study need therefore to be validated in larger patient cohorts. Another limitation is that although nodal status was considered to define pathological response, only primary tumors were characterized
in the images to simplify the analysis. Finally, some of the patients were characterized by ΔTLG and ΔMATV decreases that were close to the optimal cut-off values but sufficient to avoid misclassification, whereas it was the opposite for SUV<sub>max</sub> (figures 3). The statistical difference between AUCs of ΔSUV<sub>max</sub> and ΔTLG might therefore be lower in larger prospective studies. However, it has to be emphasized that values of ΔMATV and ΔTLG present a higher spread than ΔSUV measurements as demonstrated in figure 3, showing their higher discriminative power demonstrated by Mann-Whitney-U tests (table 2).

CONCLUSION

The reduction of metabolically active primary tumor volumes and associated activity measurements such as Total Lesion Glycolysis after two neo-adjuvant chemotherapy cycles predicts histopathological tumor response with statistically significant higher accuracy (94%) than SUV<sub>max</sub> (77%). The advantage of TLG over SUV<sub>max</sub> was particularly evident for the 26 patients in the ER-positive/HER2-negative sub-group and therefore we will focus on confirming these results in a larger group of ER-positive/HER2-negative patients as they may potentially increase the clinical value and efficiency of <sup>18</sup>F-FDG PET for early prediction of response to NAC.
REFERENCES


Figure captions

Figure 1: Illustration of PET$_1$ (left) and PET$_2$ (right) for (A) a responder ($\Delta$SUV$_{\text{max}}$ - 46%, $\Delta$TLG -89%) and (B) a non-responder (both $\Delta$SUV$_{\text{max}}$ and $\Delta$TLG -43%). The green contours represent the FLAB delineation.

Figure 2: Box-and-whisker distributions plots of (A) SUV$_{\text{max}}$, SUV$_{\text{peak}}$, SUV$_{\text{mean}}$, (B) MATV and (C) TLG absolute values at PET$_1$ and PET$_2$. One patient with extremely high MATV and TLG values does not appear for readability purposes. The central box represents values from the 25 to 75 percentile. The middle line represents the median. A line extends from minimum to maximum, excluding "outliers" which are displayed as separate points.

Figure 3: Box-and-whisker distributions plots of $\Delta$SUV$_{\text{max}}$, $\Delta$SUV$_{\text{peak}}$, $\Delta$SUV$_{\text{mean}}$, $\Delta$MATV and $\Delta$TLG. Optimal cut-off values providing best accuracy in predicting response are displayed. The central box represents values from the 25 to 75 percentile. The middle line represents the median. A line extends from minimum to maximum, excluding "outliers" which are displayed as separate points.

Figure 4: ROC curves related to the prediction of responders, for $\Delta$SUV$_{\text{max}}$, $\Delta$SUV$_{\text{peak}}$, $\Delta$SUV$_{\text{mean}}$, $\Delta$MATV and $\Delta$TLG, (A) without or (B) with partial volume effects correction.

Figure 5: ROC curves related to the prediction of pCR, for $\Delta$SUV$_{\text{max}}$, $\Delta$SUV$_{\text{peak}}$, $\Delta$SUV$_{\text{mean}}$, $\Delta$MATV and $\Delta$TLG, (A) without or (B) with partial volume effects correction.
## Table 1 - Patients characteristics

<table>
<thead>
<tr>
<th></th>
<th>Number of Patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AJCC clinical stage</strong></td>
<td></td>
</tr>
<tr>
<td>IIA</td>
<td>11 (21.5)</td>
</tr>
<tr>
<td>IIB</td>
<td>13 (25.5)</td>
</tr>
<tr>
<td>IIIA</td>
<td>12 (23.5)</td>
</tr>
<tr>
<td>IIIB</td>
<td>14 (27.5)</td>
</tr>
<tr>
<td>IIIC</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Tumor Type</strong></td>
<td></td>
</tr>
<tr>
<td>Invasive ductal, no special type</td>
<td>45 (88)</td>
</tr>
<tr>
<td>Metaplastic</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Lobular</td>
<td>3 (6)</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
</tr>
<tr>
<td>grade-1</td>
<td>4 (8)</td>
</tr>
<tr>
<td>grade-2</td>
<td>28 (55)</td>
</tr>
<tr>
<td>grade-3</td>
<td>17 (33)</td>
</tr>
<tr>
<td>Grade unknown</td>
<td>2 (4)</td>
</tr>
<tr>
<td><strong>Estrogen receptor status</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>30 (59)</td>
</tr>
<tr>
<td>Negative</td>
<td>21 (41)</td>
</tr>
<tr>
<td><strong>HER2 status†</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>12 (23.5)</td>
</tr>
<tr>
<td>Negative</td>
<td>39 (76.5)</td>
</tr>
<tr>
<td><strong>Triple negative Status</strong></td>
<td></td>
</tr>
<tr>
<td>Triple negative</td>
<td>13 (25.5)</td>
</tr>
<tr>
<td>Not Triple negative</td>
<td>38 (74.5)</td>
</tr>
<tr>
<td><strong>Surgery</strong></td>
<td></td>
</tr>
<tr>
<td>Breast-conserving surgery</td>
<td>25 (49)</td>
</tr>
<tr>
<td>Mastectomy</td>
<td>26 (51)</td>
</tr>
<tr>
<td><strong>Pathological Response</strong></td>
<td></td>
</tr>
<tr>
<td>Responders</td>
<td>27 (53)</td>
</tr>
<tr>
<td>Non-responder</td>
<td>24 (47)</td>
</tr>
</tbody>
</table>

* AJCC version 7 (16) according to clinical examination and conventional imaging findings. ** Tumors were considered positive for ER or for PR if more than 10% of cells showed staining by IHC. † Tumors were considered to overexpress c-erbB-2 oncprotein (HER2-positive) if more than 30% of invasive tumor cells showed definite membrane staining resulting in a so-called fishnet appearance.
Table 2 – PET parameters values and correlation with response

<table>
<thead>
<tr>
<th>Parameter*</th>
<th>PET₁</th>
<th>p†</th>
<th>PET₂</th>
<th>p†</th>
<th>Δ(PET₁, PET₂) (%)</th>
<th>p†</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUV&lt;sub&gt;max&lt;/sub&gt;</td>
<td>6.8 (2.3, 27.5)</td>
<td>0.8</td>
<td>3.9 (1.4, 30.7)</td>
<td>0.04</td>
<td>-34 (-90, +104)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SUV&lt;sub&gt;peak&lt;/sub&gt;</td>
<td>4.8 (1.9, 22.0)</td>
<td>0.8</td>
<td>2.9 (1.0, 24.5)</td>
<td>0.07</td>
<td>-39 (-90, +73)</td>
<td>&lt;0.0004</td>
</tr>
<tr>
<td>SUV&lt;sub&gt;mean&lt;/sub&gt;</td>
<td>3.7 (1.3, 18.7)</td>
<td>0.7</td>
<td>2.5 (1.0, 17.9)</td>
<td>0.2</td>
<td>-36 (-87, +69)</td>
<td>0.026</td>
</tr>
<tr>
<td>MATV (cm&lt;sup&gt;3&lt;/sup&gt;)</td>
<td>14 (2, 227)</td>
<td>0.1</td>
<td>7 (1, 154)</td>
<td>0.09</td>
<td>-45 (-89, +74)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TLG</td>
<td>51 (9, 668)</td>
<td>0.4</td>
<td>16 (1, 758)</td>
<td>0.05</td>
<td>-59 (-98, +51)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* median (min, max)
† p value of Mann-Whitney-U test (responders vs. non-responders). **Bold** indicates significant value.

Table 3 - Predictive value according to ROC analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>AUC [95% CI]</th>
<th>Cut-off value</th>
<th>Sensitivity [95% CI] (%)</th>
<th>Specificity [95% CI] (%)</th>
<th>Accuracy [95% CI] (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUV&lt;sub&gt;max&lt;/sub&gt; PET₁</td>
<td>0.52 [0.37-0.66]</td>
<td>&gt;11.6</td>
<td>30 [14-50]</td>
<td>83 [63-95]</td>
<td>55 [38-68]</td>
</tr>
<tr>
<td>PET₂</td>
<td>0.67 [0.52-0.79]</td>
<td>&lt;=3.7</td>
<td>67 [46-84]</td>
<td>75 [53-90]</td>
<td>71 [49-87]</td>
</tr>
<tr>
<td>Δ (%)</td>
<td>0.82 [0.68-0.91]</td>
<td>&lt;=-48</td>
<td>63 [42-81]</td>
<td>92 [73-99]</td>
<td>77 [57-90]</td>
</tr>
<tr>
<td>SUV&lt;sub&gt;peak&lt;/sub&gt; PET₁</td>
<td>0.52 [0.37-0.66]</td>
<td>&gt;11.7</td>
<td>22 [9-42]</td>
<td>88 [68-97]</td>
<td>53 [34-69]</td>
</tr>
<tr>
<td>PET₂</td>
<td>0.65 [0.50-0.78]</td>
<td>&lt;=2.8</td>
<td>63 [42-81]</td>
<td>71 [49-87]</td>
<td>67 [46-85]</td>
</tr>
<tr>
<td>Δ (%)</td>
<td>0.79 [0.65-0.89]</td>
<td>&lt;=-42</td>
<td>70 [50-86]</td>
<td>83 [63-95]</td>
<td>76 [54-91]</td>
</tr>
<tr>
<td></td>
<td>PET$_1$</td>
<td>PET$_2$</td>
<td>PET$_2$</td>
<td>PET$_2$</td>
<td>PET$_2$</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>SUV$_{\text{mean}}$</td>
<td>0.53 [0.38-0.67]</td>
<td>0.62 [0.47-0.75]</td>
<td>0.68 [0.54-0.81]</td>
<td>0.63 [0.48-0.76]</td>
<td>0.64 [0.49-0.77]</td>
</tr>
<tr>
<td>MATV (cm$^3$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET$_1$</td>
<td>0.57 [0.43-0.71]</td>
<td>0.64 [0.49-0.77]</td>
<td>0.92 [0.82-0.98]</td>
<td>0.63 [0.48-0.76]</td>
<td>0.64 [0.49-0.77]</td>
</tr>
<tr>
<td>TLG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET$_1$</td>
<td>0.57 [0.43-0.71]</td>
<td>0.66 [0.51-0.79]</td>
<td>0.91 [0.79-0.97]</td>
<td>0.57 [0.43-0.71]</td>
<td>0.66 [0.51-0.79]</td>
</tr>
</tbody>
</table>
Figure 2

A

35
30
25
20
15
10
5
0

PET₁
PET₂
PET₁
PET₂
PET₁
PET₂

SUV
SUV\text{max}
SUV\text{peak}
SUV\text{mean}

Respondees
Nonresponders
Figure 2

TLG

PET_1  PET_2

Responders  Nonresponders
Figure 4

ΔTLG AUC = 0.91
ΔMATV AUC = 0.92
ΔSUV\text{\textsubscript{max}} AUC = 0.82
ΔSUV\text{\textsubscript{peak}} AUC = 0.79
ΔSUV\text{\textsubscript{mean}} AUC = 0.68
Figure 4

- $\Delta TLG^{pvc}$ AUC = 0.90
- $\Delta MATV^{pvc}$ AUC = 0.92
- $\Delta SUV_{max}^{pvc}$ AUC = 0.81
- $\Delta SUV_{peak}^{pvc}$ AUC = 0.79
- $\Delta SUV_{mean}^{pvc}$ AUC = 0.62
Figure 5

ΔTLG AUC = 0.76
ΔMATV AUC = 0.79
ΔSUV_{max} AUC = 0.67
ΔSUV_{peak} AUC = 0.63
ΔSUV_{mean} AUC = 0.52
Figure 5

- $\Delta TLG^{PVC}$ AUC = 0.77
- $\Delta MATV^{PVC}$ AUC = 0.76
- $\Delta SUV_{\text{max}}^{PVC}$ AUC = 0.64
- $\Delta SUV_{\text{peak}}^{PVC}$ AUC = 0.63
- $\Delta SUV_{\text{mean}}^{PVC}$ AUC = 0.53