HAL
open science

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, 2013, 54 (3), pp.341-9. 10.2967/jnumed.112.108837 . hal-00748923

HAL Id: hal-00748923

https://hal.science/hal-00748923v1

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 Hatt[1], David Groheux[2,3], Antoine Martineau[2], Marc Espié[4], Elif Hindié[5], Sylvie Giacchetti[4], Anne de Roquancourt[6], Dimitris Visvikis[1], Catherine Cheze-Le Rest[1,7].

[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 ^{18}F-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 (Δ) 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 ΔTLG (96% sensitivity, 92% specificity, 94% accuracy) with significantly higher AUC (0.91 vs. 0.82, p=0.01) than ΔSUV_{max} (63% sensitivity, 92% specificity, 77% accuracy). Sub-group analysis confirmed a significantly higher accuracy for ΔTLG than ΔSUV 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, ^{18}F-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_{max}, or SUV_{peak} averaging the SUV_{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_{mean}) (*12*). One recent study compared SUV_{max} and TLG derived using manual delineation and threshold, showing that SUV_{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_{max}, SUV_{mean}, SUV_{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 (SUV_{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 ≤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 Xl 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 4×4×4mm^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_{max}, SUV_{peak} and SUV_{mean} as well as the MATV and the TLG were then automatically calculated from the tumor delineations. SUV_{peak} was defined as the mean of voxels intensities in a 1.7 cm^3 spheric ROI (27 voxels) centered on SUV_{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^3). SUV_{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_{mean}. The percentage evolution of each parameter between baseline and after the second NAC cycle (Δparam, %) was calculated as :

$$\Delta param = (param_PET_2 - param_PET_1) / param_PET_1 \times 100.$$

Statistical Analysis

Statistical analyses were performed using MedCalc™ (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_1 and PET_2, as well as their evolution $\Delta(PET_1, PET_2)$ 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_{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≥0.07) 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 ΔSUV_{max} (-34%±32%) and ΔTLG (-59%±34%) respectively.

According to Mann-Whitney-U tests, the variation (Δ) of all parameters was statistically different between responders and non-responders, especially for ΔMATV and ΔTLG (*p*<0.0001), as well as ΔSUV_{max} (p=0.0001) (table 2).

According to ROC analysis, the best prediction was achieved using the ΔTLG, with a sensitivity of 96% and specificity of 92% resulting in an accuracy of 94% (table 3). The ΔMATV was also an accurate predictive factor (sensitivity 93%, specificity 88%, accuracy 91%). On the other hand, ΔSUV measurements led to significantly lower AUCs (0.68 for ΔSUV_{mean} to 0.82 for ΔSUV_{max} *vs.* 0.91 and 0.92 for ΔMATV and

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_{max}, ΔSUV_{peak}, ΔSUV_{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_{mean}, ΔSUV_{peak} and ΔSUV_{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_{max}, ΔSUV_{peak}, and ΔSUV_{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 Δ(%) 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 (Δ). Our results are in contradiction with the ones of the only other study to date comparing the performance of ΔSUV_{max} and ΔTLG, that reported a higher predictive value for ΔSUV_{max} (*13*). In this study, ΔSUV_{mean}, ΔSUV_{peak} or ΔMATV predictive values were not considered. In addition, ΔTLG 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 ΔSUV_{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_{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 (Δ).

Pathological responders were associated with a significantly higher decrease of the ^{18}F-FDG PET derived indices considered. Among these, the ΔTLG and ΔMATV were the best predictive factors compared to ΔSUV measurements, with significantly higher accuracy (91% and 94% for MATV and TLG vs. 71%, 76% and 77% for ΔSUV_{mean}, ΔSUV_{peak} and ΔSUV_{max} respectively). With respect to the recent meta-analysis that reported an overall 84% sensitivity and 66% specificity based on ΔSUV_{max} (*8*), our ΔSUV_{max} results demonstrated lower sensitivity (63%) but higher specificity (92%), whereas ΔTLG 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 (Δ) 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 ΔSUV, ΔMATV and ΔTLG 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 ΔMATV and ΔTLG over ΔSUV 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 ΔSUV measurements.

Another interesting result of our study was that ΔSUV_{mean} had significantly lower predictive value than ΔSUV_{peak} and ΔSUV_{max} (AUC of 0.68 *vs.* 0.79 and 0.82, $p<0.02$). Considering the use of ΔSUV_{peak} relative to ΔSUV_{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 ΔSUV_{max} leading to slightly better results than ΔSUV_{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 ΔSUV_{mean} is its higher dependency on partial volume effects compared to SUV_{max} and SUV_{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}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 <3cm 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 (Δ) 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 SUV_{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_{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_{max} (figures 3). The statistical difference between AUCs of ΔSUV_{max} 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_{max} (77%). The advantage of TLG over SUV_{max} 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 ^{18}F-FDG PET for early prediction of response to NAC.

REFERENCES

1. Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol.* 1998;16:2672-2685.

2. Groheux D, Giacchetti S, Espie M, Rubello D, Moretti JL, Hindie E. Early monitoring of response to neoadjuvant chemotherapy in breast cancer with 18F-FDG PET/CT: defining a clinical aim. *Eur J Nucl Med Mol Imaging*. 2011;38:419-425.

3. Bhargava R, Beriwal S, Dabbs DJ, et al. Immunohistochemical surrogate markers of breast cancer molecular classes predicts response to neoadjuvant chemotherapy: a single institutional experience with 359 cases. *Cancer*. 2010;116:1431-1439.

4. von Minckwitz G, Untch M, Blohmer JU, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol.* 2012;30:1796-1804.

5. Groheux D, Hindie E, Giacchetti S, et al. Triple-negative breast cancer: early assessment with 18F-FDG PET/CT during neoadjuvant chemotherapy identifies patients who are unlikely to achieve a pathologic complete response and are at a high risk of early relapse. *J Nucl Med.* 2012;53:249-254.

6. Rousseau C, Devillers A, Sagan C, et al. Monitoring of early response to neoadjuvant chemotherapy in stage II and III breast cancer by [18F]fluorodeoxyglucose positron emission tomography. *J Clin Oncol.* 2006;24:5366-5372.

7. Schwarz-Dose J, Untch M, Tiling R, et al. Monitoring primary systemic therapy of large and locally advanced breast cancer by using sequential positron emission tomography imaging with [18F]fluorodeoxyglucose. *J Clin Oncol.* 2009;27:535-541.

8. Wang Y, Zhang C, Liu J, Huang G. Is 18F-FDG PET accurate to predict neoadjuvant therapy response in breast cancer? A meta-analysis. *Breast Cancer Res Treat.* 2012;131:357-369.

9. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50 Suppl 1:122S-150S.

10. Cazaentre T, Morschhauser F, Vermandel 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.

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;38:1595-1606.

12. Larson SM, Erdi Y, Akhurst T, 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;2:159-171.

13. Tateishi U, Miyake M, Nagaoka T, et al. Neoadjuvant chemotherapy in breast cancer: prediction of pathologic response with PET/CT and dynamic contrast-enhanced MR imaging--prospective assessment. *Radiology.* 2012;263:53-63.

14. Groheux D, Moretti J, Espié M, et al. FDG PET/CT for early prediction of the response of neoadjuvant chemotherapy for breast cancer after two courses of EC (epirubicin + cyclophosphamide). *J Nucl Med.* 2010;51 (Supplement 2):183.

15. Sataloff DM, Mason BA, Prestipino AJ, Seinige UL, Lieber CP, Baloch Z. Pathologic response to induction chemotherapy in locally advanced carcinoma of the breast: a determinant of outcome. *J Am Coll Surg*. 1995;180:297-306.

16. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol.* 2010;17:1471-1474.

17. Hatt M, Cheze le Rest C, Descourt P, et al. Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications. *Int J Radiat Oncol Biol Phys.* 2010;77:301-308.

18. 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;38:663-672.

19. Hatt M, Cheze-Le Rest C, Aboagye EO, et al. Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements. *J Nucl Med.* 2010;51:1368-1376.

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;36:1064-1075.

21. Sheskin D. *Handbook of parametric and nonparametric statistical procedures 5th ed.*: Boca Raton: Chapman & Hall /CRC; 2011.

22. Straver ME, Rutgers EJ, Rodenhuis S, et al. The relevance of breast cancer subtypes in the outcome of neoadjuvant chemotherapy. *Ann Surg Oncol.* 2010;17:2411-2418.

23. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D. Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. *J Nucl Med.* 2011;52:1690-1697.

24. Hatt M, Visvikis D, Le Rest CC. Autocontouring versus manual contouring [Letter to the editor]. *J Nucl Med.* 2011;52:658; author reply 658-659.

25. McDermott GM, Welch A, Staff RT, et al. Monitoring primary breast cancer throughout chemotherapy using FDG-PET. *Breast Cancer Res Treat.* 2007;102:75-84.

26. Groheux D, Giacchetti S, Moretti JL, et al. Correlation of high 18F-FDG uptake to clinical, pathological and biological prognostic factors in breast cancer. *Eur J Nucl Med Mol Imaging.* 2011;38:426-435.

27. Lips EH, Mulder L, de Ronde JJ, et al. Neoadjuvant chemotherapy in ER+ HER2- breast cancer: response prediction based on immunohistochemical and molecular characteristics. *Breast Cancer Res Treat.* 2012;131:827-836.

28. de Ronde JJ, Hannemann J, Halfwerk H, et al. Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response. *Breast Cancer Res Treat.* 2010;119:119-126.

29. Hoetjes NJ, van Velden FH, Hoekstra OS, et al. Partial volume correction strategies for quantitative FDG PET in oncology. *Eur J Nucl Med Mol Imaging.* 2010;37:1679-1687.

30. Tseng J, Dunnwald LK, Schubert EK, et al. 18F-FDG kinetics in locally advanced breast cancer: correlation with tumor blood flow and changes in response to neoadjuvant chemotherapy. *J Nucl Med.* 2004;45:1829-1837.

31. Hatt M, van Stiphout R, le Pogam A, Lammering G, Visvikis D, Lambin P. Early prediction of pathological response in locally advanced rectal cancer based on sequential (18)F-FDG PET. *Acta Oncol.* 2012.

32. 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. *Journal of Nuclear Medicine.* 2012;53:in press.

33. Geets X, Lee JA, Bol A, Lonneux M, Gregoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Im.* 2007;34:1427-1438.

34. Belhassen S, Zaidi H. A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. *Med Phys.* 2010;37:1309-1324.

Figure captions

Figure 1: Illustration of PET_1 (left) and PET_2 (right) for (A) a responder (ΔSUV_{max} -46%, ΔTLG -89%) and (B) a non-responder (both ΔSUV_{max} and ΔTLG -43%). The green contours represent the FLAB delineation.

Figure 2: Box-and-whisker distributions plots of (A) SUV_{max}, SUV_{peak}, SUV_{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 ΔSUV_{max}, ΔSUV_{peak}, ΔSUV_{mean}, ΔMATV and ΔTLG. 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 ΔSUV_{max}, ΔSUV_{peak}, ΔSUV_{mean}, ΔMATV and ΔTLG, (A) without or (B) with partial volume effects correction.

Figure 5: ROC curves related to the prediction of pCR, for ΔSUV_{max}, ΔSUV_{peak}, ΔSUV_{mean}, ΔMATV and ΔTLG, (A) without or (B) with partial volume effects correction.

Table 1 - Patients characteristics

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

* 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 oncoprotein (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

Parameter*	PET_1	*p* †	PET_2	*p* †	$\Delta(PET_1, PET_2)$ (%)	*p* †
SUV_{max}	6.8 (2.3, 27.5)	0.8	3.9 (1.4, 30.7)	**0.04**	-34 (-90, +104)	**0.0001**
SUV_{peak}	4.8 (1.9, 22.0)	0.8	2.9 (1.0, 24.5)	0.07	-39 (-90, +73)	**0.0004**
SUV_{mean}	3.7 (1.3, 18.7)	0.7	2.5 (1.0, 17.9)	0.2	-36 (-87, +69)	**0.026**
MATV (cm^3)	14 (2, 227)	0.1	7 (1, 154)	0.09	-45 (-89, +74)	**<0.0001**
TLG	51 (9, 668)	0.4	16 (1, 758)	**0.05**	-59 (-98, +51)	**<0.0001**

* 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

Parameter		AUC [95% CI]	Cut-off value	Sensitivity [95% CI] (%)	Specificity [95% CI] (%)	Accuracy [95% CI] (%)
SUV_{max}	PET_1	0.52 [0.37-0.66]	>11.6	30 [14-50]	83 [63-95]	55 [38-68]
	PET_2	0.67 [0.52-0.79]	<=3.7	67 [46-84]	75 [53-90]	71 [49-87]
	Δ (%)	0.82 [0.68-0.91]	<=-48	63 [42-81]	92 [73-99]	77 [57-90]
SUV_{peak}	PET_1	0.52 [0.37-0.66]	>11.7	22 [9-42]	88 [68-97]	53 [34-69]
	PET_2	0.65 [0.50-0.78]	<=2.8	63 [42-81]	71 [49-87]	67 [46-85]
	Δ (%)	0.79 [0.65-0.89]	<=-42	70 [50-86]	83 [63-95]	76 [54-91]

SUV_{mean}	PET_1	0.53 [0.38-0.67]	<=2.2	19 [6-38]	92 [73-99]	53 [34-69]
	PET_2	0.62 [0.47-0.75]	<=2	59 [39-78]	75 [53-90]	67 [46-85]
	Δ (%)	0.68 [0.54-0.81]	<=-30	74 [54-89]	67 [45-84]	71 [49-88]
MATV (cm^3)	PET_1	0.63 [0.48-0.76]	>7	85 [66-96]	42 [22-63]	65 [48-79]
	PET_2	0.64 [0.49-0.77]	<=11.3	89 [71-98]	42 [22-63]	67 [45-84]
	Δ (%)	0.92 [0.82-0.98]	<=-42	93 [76-99]	88 [68-97]	91 [72-98]
TLG	PET_1	0.57 [0.43-0.71]	>16	93 [76-99]	29 [13-51]	63 [42-81]
	PET_2	0.66 [0.51-0.79]	<=23.3	78 [58-91]	63 [41-82]	71 [49-88]
	Δ (%)	0.91 [0.79-0.97]	<=-56	96 [81-100]	92 [73-99]	94 [79-99]

Figure 1

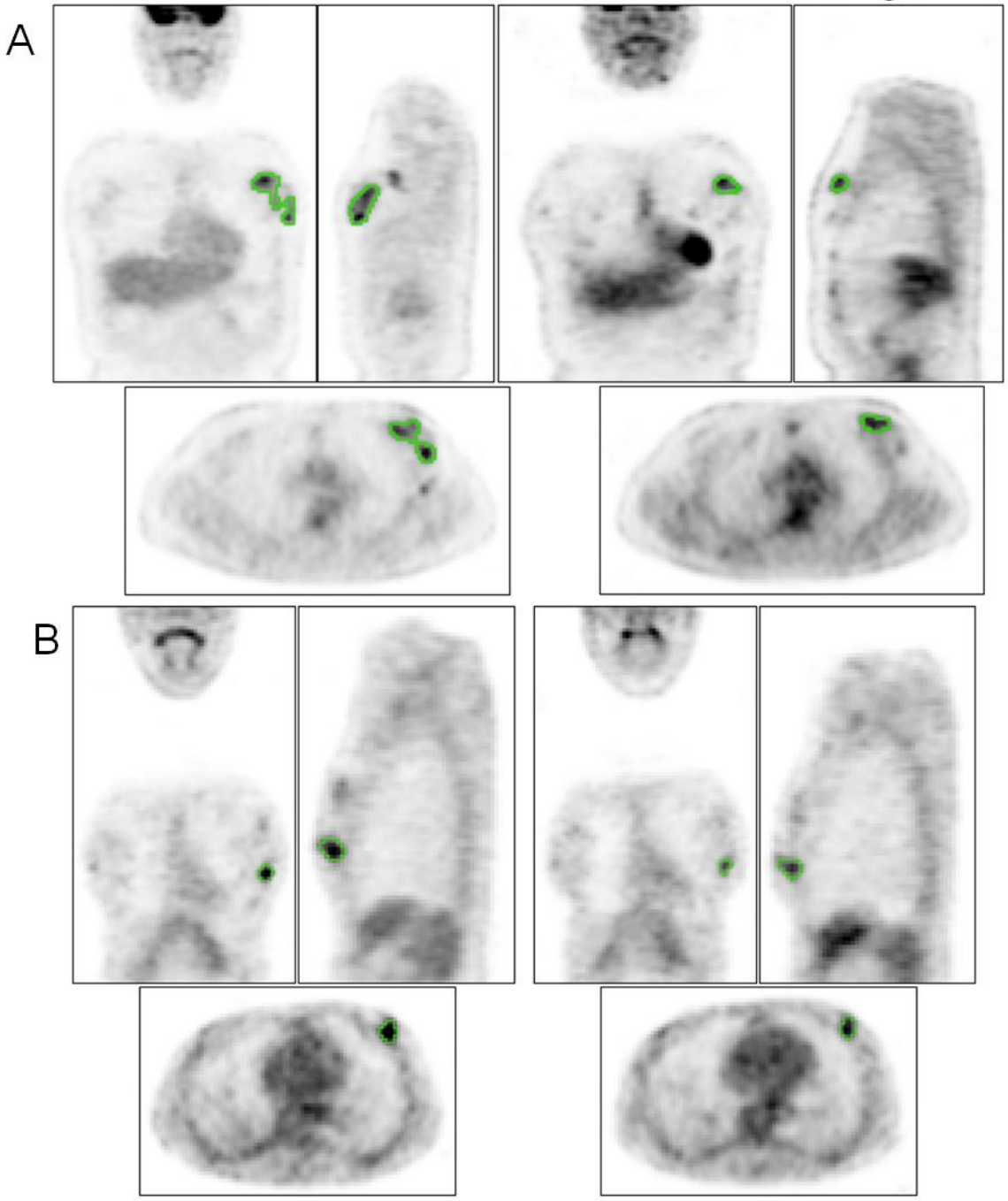
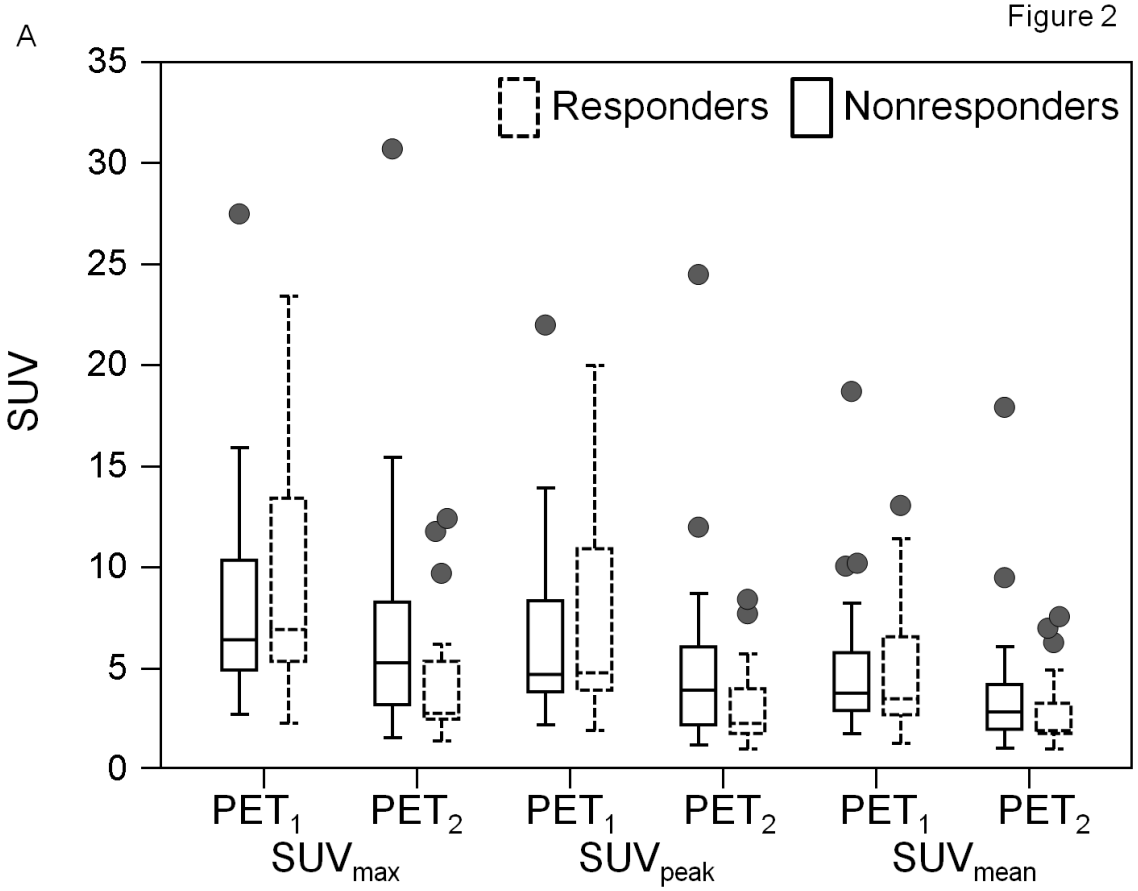
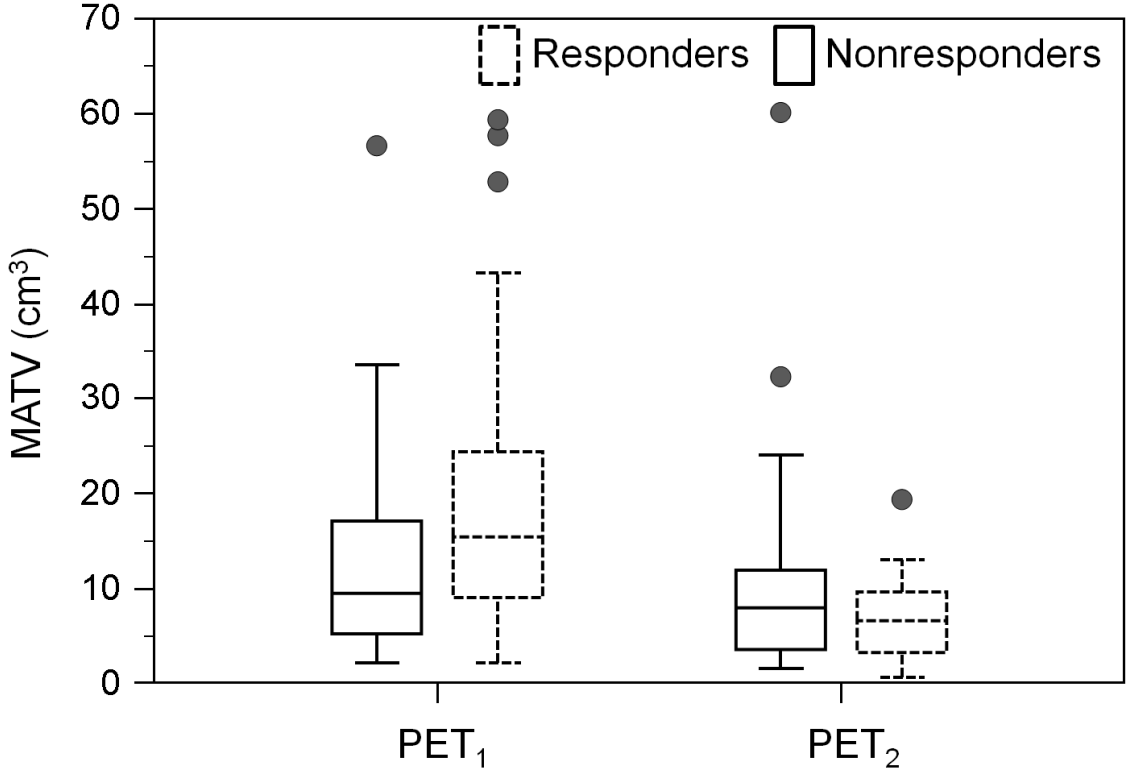


Figure 2


A
Responders
Nonresponders
SUV
35
30
25
20
15
10
5
0
PET1
PET2
PET1
PET2
PET1
PET2
SUVmax
SUVpeak
SUVmean

B

Figure 2


Responders
Nonresponders
MATV (cm3)
70
60
50
40
30
20
10
0
PET1
PET2

C

Figure 2

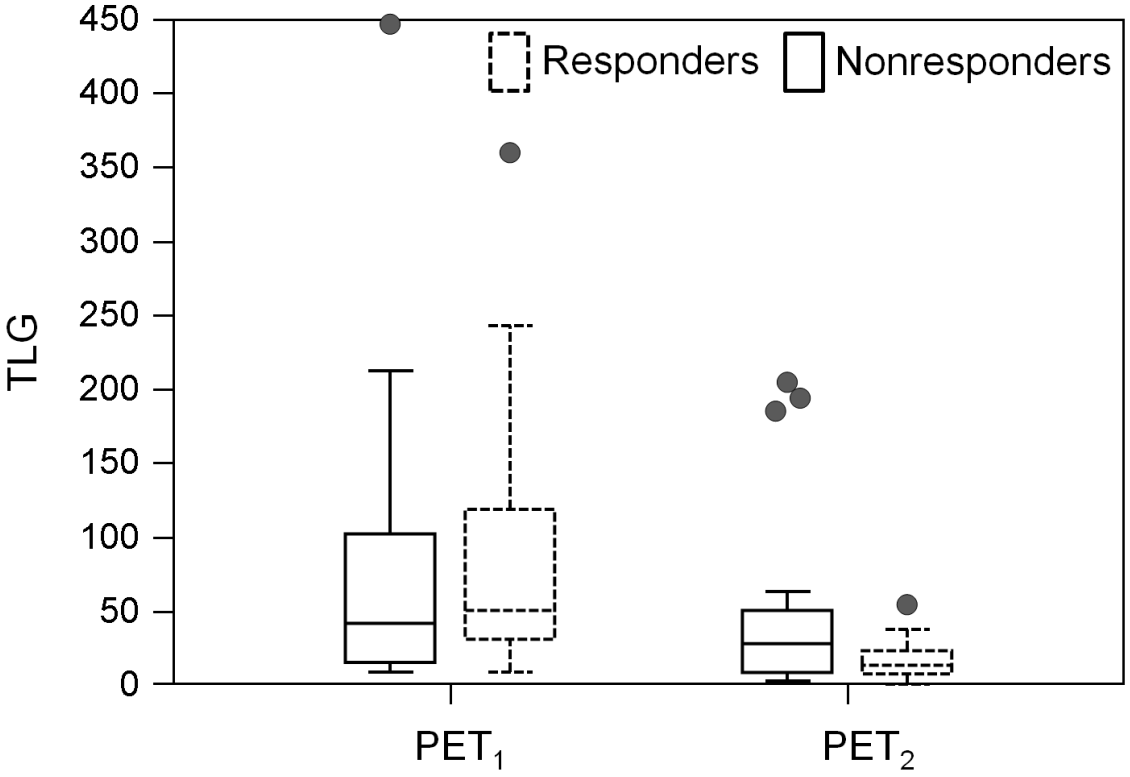
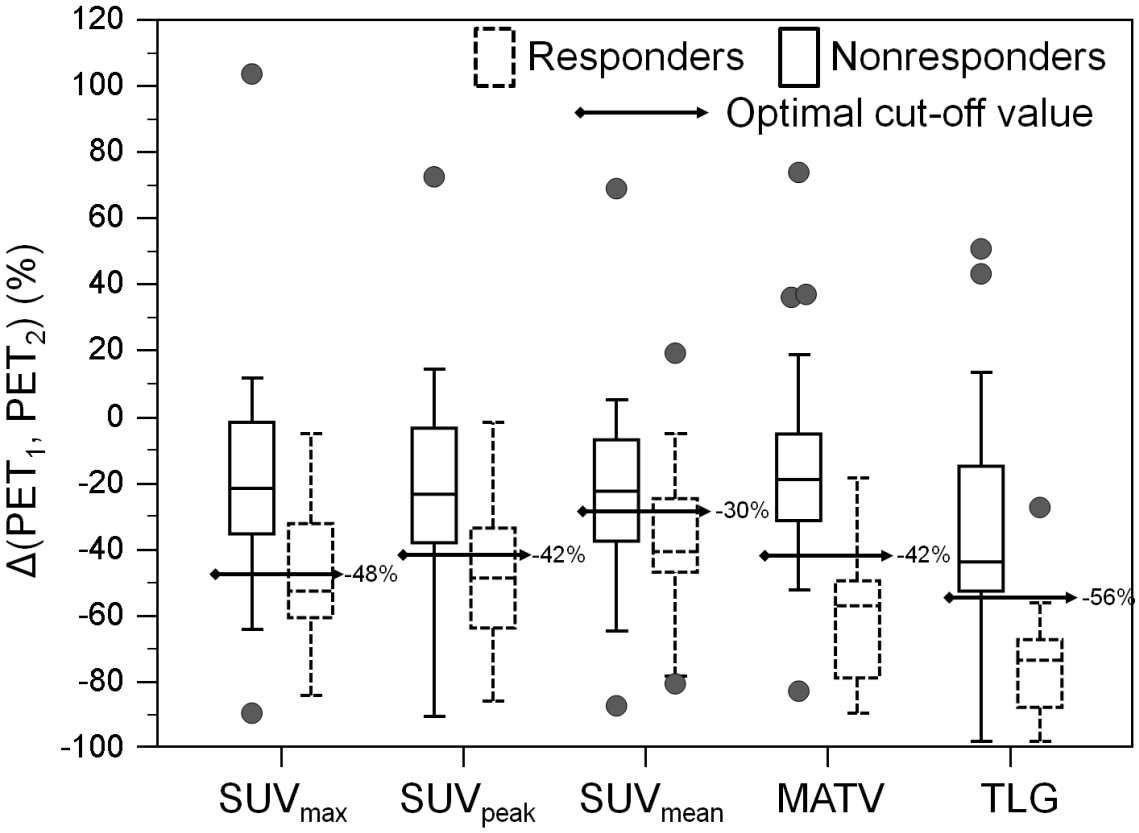
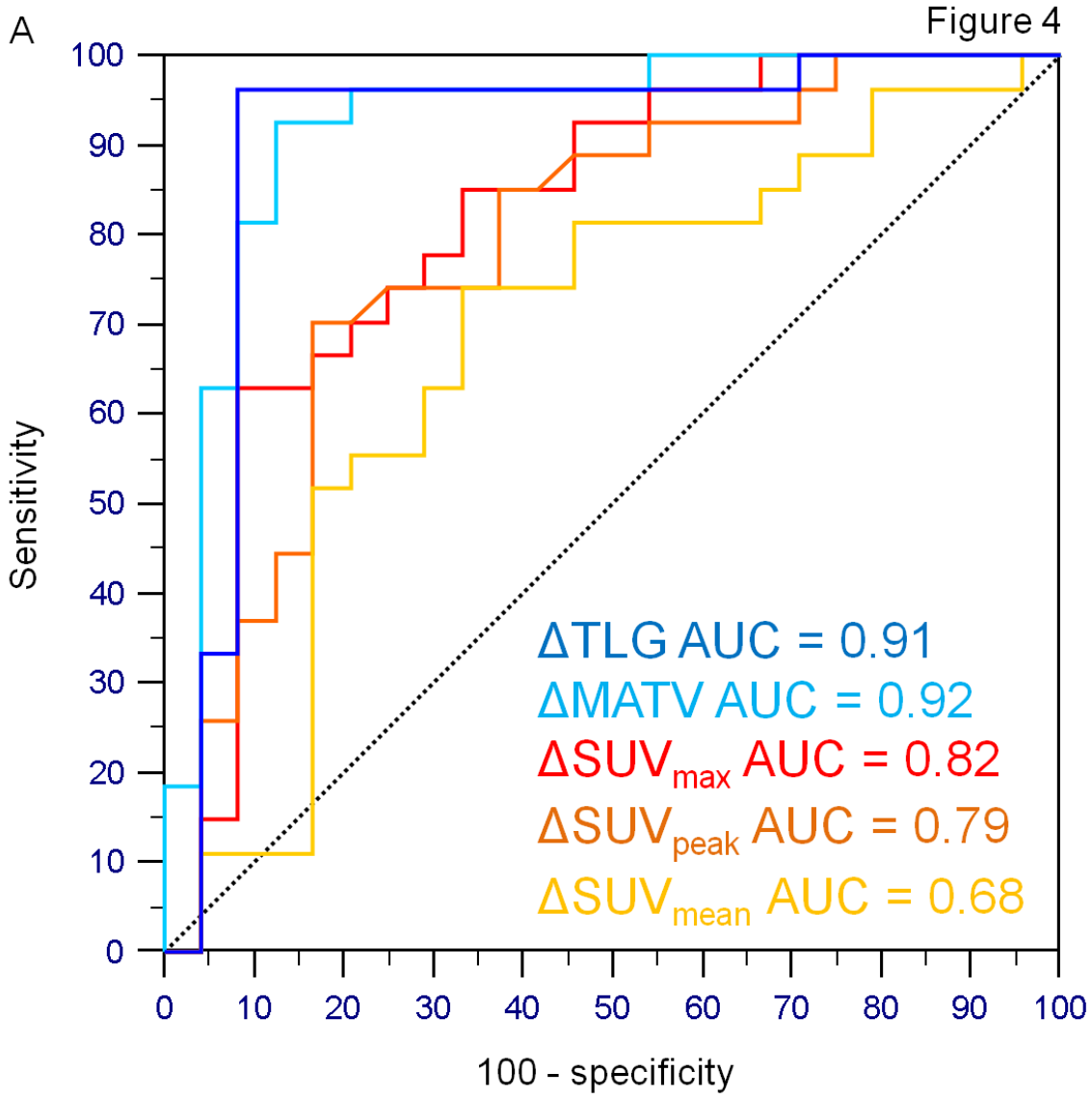
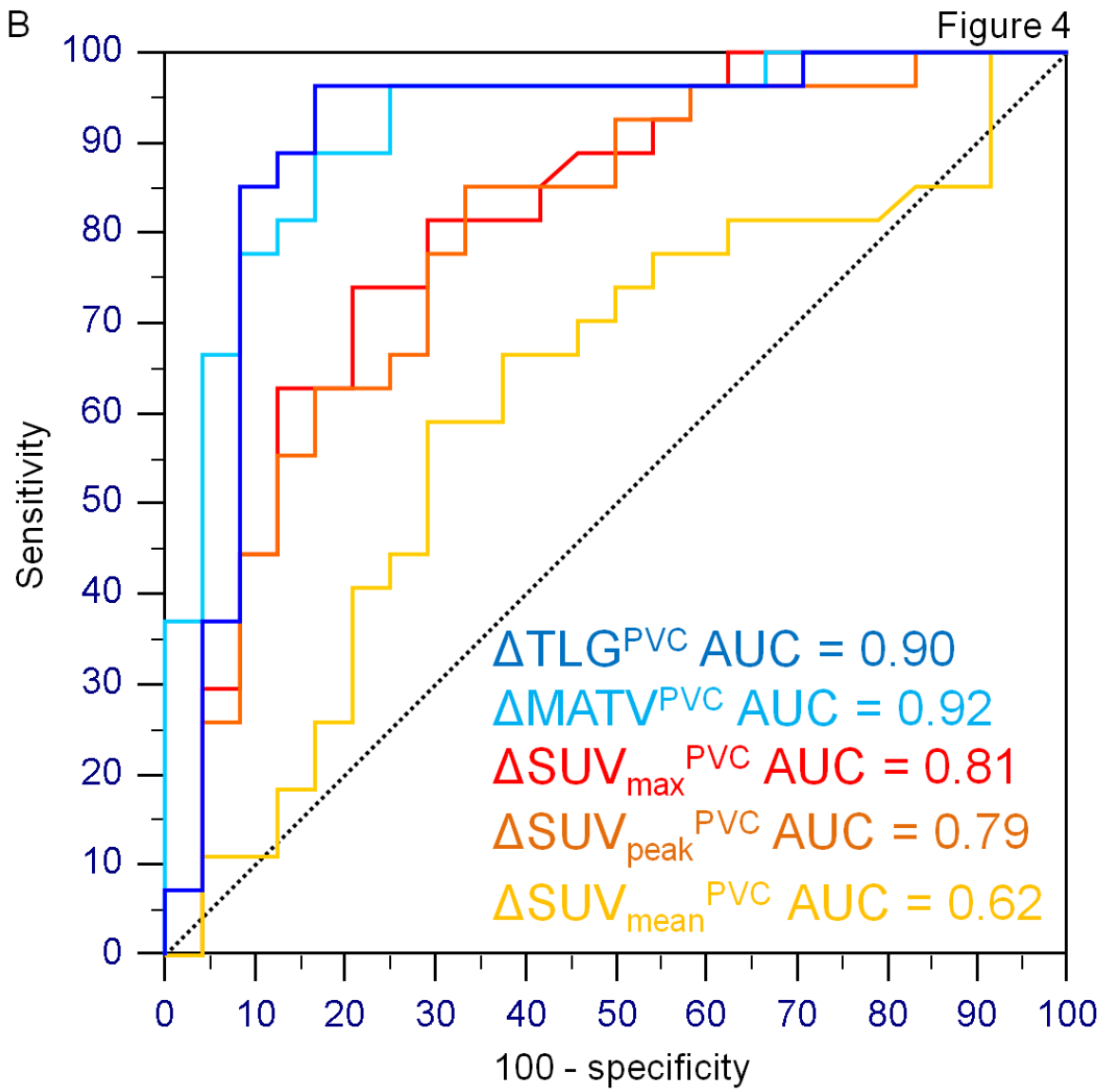

Responders
Nonresponders
TLG
450
400
350
300
250
200
150
100
50
0
PET1
PET2

Figure 3


Responders
Nonresponders
Optimal cut-off value
Δ(PET1, PET2) (%)
120
100
80
60
40
20
0
-20
-40
-60
-80
-100
-48%
-42%
-30%
-42%
-56%
SUVmax
SUVpeak
SUVmean
MATV
TLG


A
Figure 4
Sensitivity
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
100 - specificity
ΔTLG AUC = 0.91
ΔMATV AUC = 0.92
ΔSUVmax AUC = 0.82
ΔSUVpeak AUC = 0.79
ΔSUVmean AUC = 0.68


B
Figure 4
Sensitivity
100 - specificity
0
10
20
30
40
50
60
70
80
90
100
ΔTLG^PVC AUC = 0.90
ΔMATV^PVC AUC = 0.92
ΔSUV_max^PVC AUC = 0.81
ΔSUV_peak^PVC AUC = 0.79
ΔSUV_mean^PVC AUC = 0.62

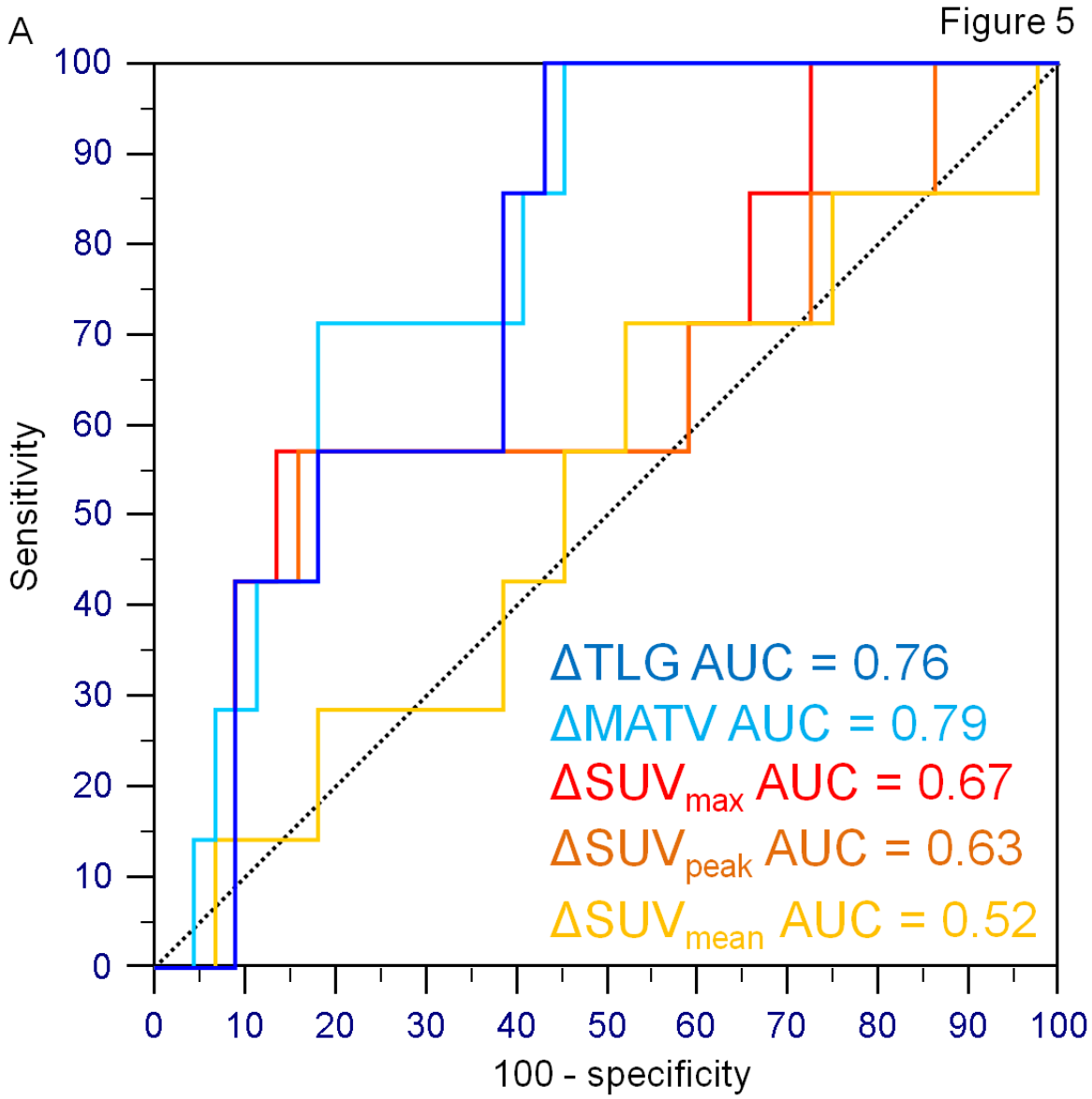
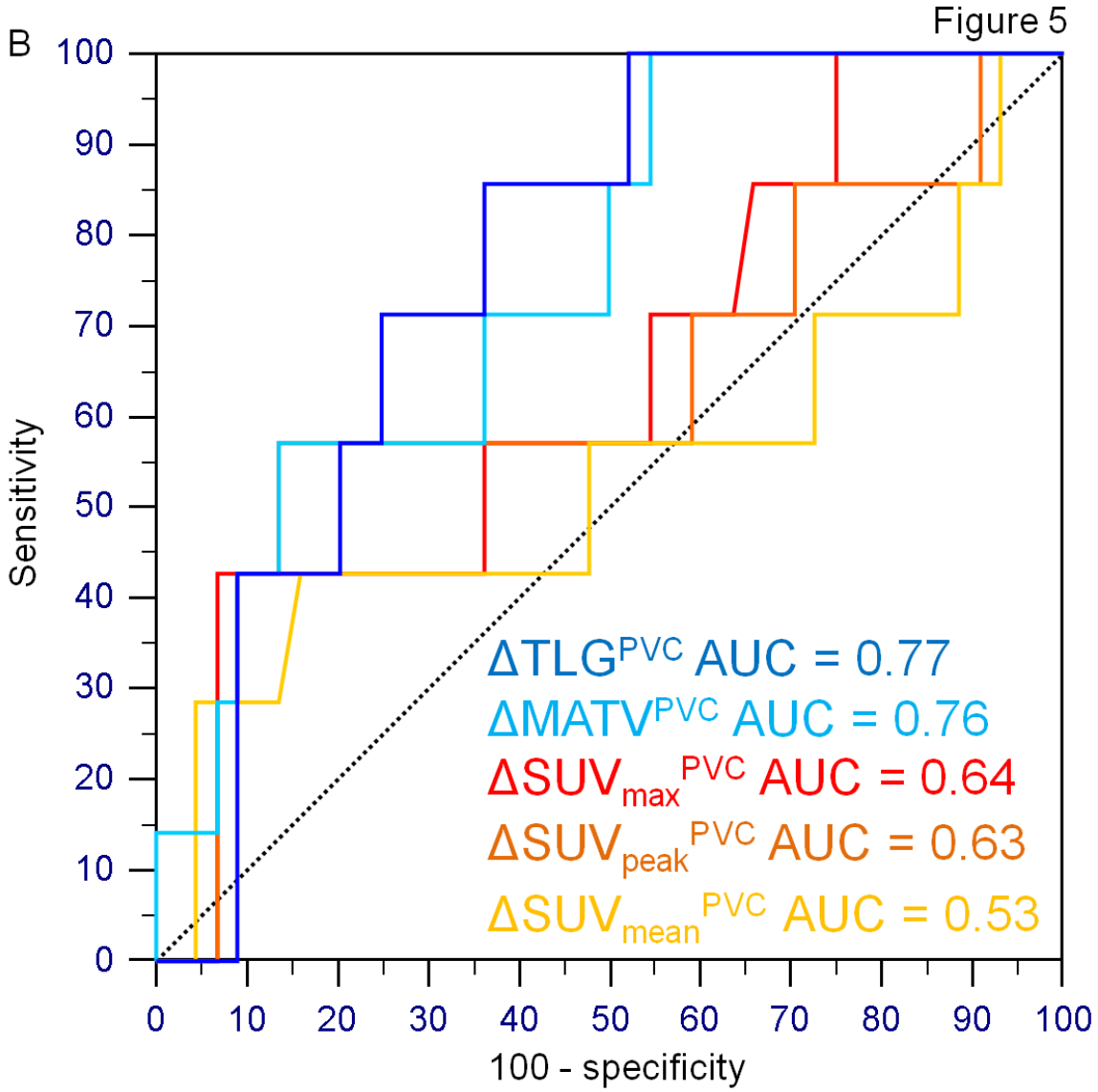

A
Figure 5
Sensitivity
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
100 - specificity
ΔTLG AUC = 0.76
ΔMATV AUC = 0.79
ΔSUVmax AUC = 0.67
ΔSUVpeak AUC = 0.63
ΔSUVmean AUC = 0.52

Figure 5


B
Sensitivity
100 - specificity
0
10
20
30
40
50
60
70
80
90
100
ΔTLG^PVC AUC = 0.77
ΔMATV^PVC AUC = 0.76
ΔSUV_max^PVC AUC = 0.64
ΔSUV_peak^PVC AUC = 0.63
ΔSUV_mean^PVC AUC = 0.53