Baseline (18)F-FDG PET image-derived parameters for therapy response prediction in oesophageal cancer.

Mathieu Hatt, Dimitris Visvikis, Olivier Pradier, Catherine Cheze-Le Rest

Baseline 18F-FDG PET image derived parameters for therapy response prediction in oesophageal cancer

Mathieu Hatt[1], Dimitris Visvikis[1], Olivier Pradier[1,2], Catherine Cheze-le Rest[1]

[1]INSERM, U650 LaTIM, [2]Department of Radiotherapy, CHU Morvan, Brest, France

Running title: [18]F-FDG PET indices for therapy response

Keywords: oesophageal cancer, response to therapy, PET scan, tumour volume, total lesion glycolysis

Corresponding author:

Mathieu HATT
LaTIM, INSERM U650
CHU MORVAN
5 avenue Foch
29609 Brest
France
Tel.:+33298018111

Wordcount: 5206

ABSTRACT

Background: The objectives of this study were to investigate the predictive value of tumour measurements on 18F-FDG PET pretreatment scan regarding therapy response in oesophageal cancer and to evaluate the impact of tumour delineation strategies.

Methods: 50 patients with oesophageal cancer treated with concomitant radio-chemotherapy between 2004 and 2008 were retrospectively considered and classified as complete, partial or non responders (including stable and progressive disease) according to RECIST. The classification of partial and complete responders was confirmed by biopsy. Tumours were delineated on the 18F-FDG pretreatment scan using an adaptive threshold and the automatic Fuzzy Locally Adaptive Bayesian (FLAB) methodologies. Several parameters were then extracted: maximum and peak SUV, tumour longitudinal length (TL) and volume (TV), mean SUV, and Total Lesion Glycolysis (TLG=TV×mean SUV). The correlation between each parameter and response was investigated using Kruskal-Wallis tests and receiver operating characteristic methodology was used to assess performance of the parameters to differentiate patients.

Results: Whereas commonly-used parameters such as SUV measurements were not significant predictive factors of the response, parameters related to tumour functional spatial extent (TL, TV, TLG) allowed significant differentiation of all three groups of patients, independently of the delineation strategy, and could identify complete and non responders with sensitivity above 75% and specificity above 85%. A systematic although not statistically significant trend was observed regarding the hierarchy of the delineation methodologies and the considered parameters, with

slightly higher predictive value obtained with FLAB over adaptive thresholding, and TLG over TV and TL.

Conclusions: TLG is a promising predictive factor of concomitant radio-chemotherapy response with statistically higher predictive value than SUV measurements in advanced oesophageal cancer.

1. Introduction

Oesophageal cancer is the third most common malignancy of the digestive tract and a leading cause of cancer mortality worldwide with an estimated 5-year survival of 15% [1]. Despite the progress made to better understand this disease, its incidence is steadily increasing and there is a growing concern regarding its effective management [2]. The best chance for cure remains surgical resection. However, many patients have already an advanced disease (locally advanced oesophageal carcinoma: LAEC) at diagnosis and may benefit in terms of survival from neoadjuvant therapy prior to surgery [3]. The maximum benefit is for those patients who achieve a complete pathological response with no residual cancer cells in the primary tumour or lymph nodes [4]. A complete response occurs only in 15-30% of cases and is associated with an increased overall survival [5]. On the other hand, patients who do not respond to therapy may be unnecessarily affected by toxicity of an inefficient therapy [6]. Therefore, the development of a diagnostic test offering non invasive response to therapy prediction early in the course of treatment is of a great interest, allowing potential personalization of patient management as for un-operable tumours, chemotherapy and/or radiation therapy remains the only option. Such an assessment becomes more critical when one considers new targeted drugs that could be tested with higher efficiency if applied early [7]. For oesophageal cancer several histological markers such as the tumour suppressor factor gene p53, the proliferative marker Ki67, and the epidermal growth factor receptor, have been evaluated for the prediction of the therapeutic response prior to neoadjuvant therapy. None of these markers or a combination of them can currently predict response with sufficient accuracy [8-9]. Positron Emission Tomography (PET) imaging with 2-(18F)fluoro-2-deoxy-D-glucose (^{18}F-FDG) allows the visualization of the enhanced glucose

metabolism in viable oesophageal cancer cells and may be of interest within this context. ^{18}F-FDG PET is already well established for staging of oesophageal cancer with a better sensitivity and specificity than the combined use of CT and endoscopic ultrasonography (EUS) to detect distant metastases [10]. PET has also been shown to be promising in assessing response to therapy [11]. Several studies have shown that the reduction of the tumour's metabolic activity as measured by the standard uptake value (SUV) from the baseline to the end of therapy uptake is predictive of a better outcome with however a large variability in the sensitivity and specificity [12]. In addition, a correlation between clinical outcome and a metabolic response observed as early as within the first 2 weeks of treatment has been demonstrated [13]. These findings suggest that tumour activity concentration differences measured on serial ^{18}F-FDG PET scans could possibly be used to individualize treatment. However, it could be more cost-effective and beneficial to the patient to be able to predict therapy response from a single baseline PET scan acquired before the initiation of the treatment. The current study was therefore carried out to investigate the potential value of baseline ^{18}F-FDG PET image derived parameters for the prediction of response to combined radio-chemotherapy in oesophageal cancer. A secondary objective was to investigate the potential influence of the method used to delineate the tumour on the prediction results.

2. Material and methods

2.1 Patients

50 consecutive patients with newly diagnosed oesophageal cancer treated with exclusive concomitant radio-chemotherapy between 2004 and 2008 were included in this study. As part of the routine procedure for the initial staging in oesophageal cancer, each patient was referred for an ^{18}F-FDG PET study before treatment. It

included three courses of 5-fluorouracil/cisplatin and a median radiation dose of 60Gy given in 180cGy daily fractions delivered once daily, 5 days a week for 6-7 weeks. The characteristics of the patients are given in table 1. Most of them (45 out of 50) were male, aged 65±9 years at the time of diagnosis. 74% of the tumours, most of which were squamous cell carcinoma (72%), originated from the middle and lower oesophagus. Response to therapy was evaluated 1 month after the completion of the concomitant radio-chemotherapy using conventional thoraco-abdominal CT and endoscopy. Patients were classified as non responders (NR) including stable and progressive disease, partial responders (PR) or complete responders (CR). Response evaluation was based on CT evolution between pre-treatment and post-treatment scans using RECIST (Response Evaluation Criteria in Solid Tumours) [14]. Patients also underwent fibroscopy in case of partial or complete response. Complete response was confirmed by the absence of visible disease in the high endoscopy and no viable tumor on biopsy. Partial CT response was confirmed by macroscopic residual (>10% viable) on biopsy. No discordance was observed between pathological, when available, and CT evaluation.

The current analysis was carried out after an approval by the institutional ethics review board.

2.2 18F-FDG PET acquisitions

All ^{18}F-FDG PET studies were carried out prior to the initiation of treatment. Patients were instructed to fast for at least 6h before the ^{18}F-FDG administration (5MBq/kg). Static emission images were acquired from head to thigh (2min per bed position) beginning 60min after injection on a Philips GEMINI PET/CT system (Philips Medical Systems, Cleveland, OH USA). Images were reconstructed using the RAMLA 3D

algorithm and CT based attenuation correction. Optimized reconstruction parameters were used for the RAMLA 3D based on the standard optimized clinical protocol (2 iterations, relaxation parameter of 0.05, 5mm 3D Gaussian post-filtering, 4x4x4mm^3 voxels grid sampling). The PET images were corrected for attenuation using CT based attenuation correction.

2.3 PET image analysis

All considered parameters were extracted from the baseline PET images only. For each patient, the primary tumour was identified on the baseline pre-treatment PET images by a nuclear physician. Three different SUV measurements and three parameters related to the tumour functional dimensions, namely the tumor volume (TV), tumour longitudinal length (TL) and total lesion glycolysis (TLG) [15] were extracted for each primary lesion. SUV measurements considered were SUV_{max}, SUV_{peak} defined as the mean of SUV_{max} and its 26 neighbors (roughly similar to a 1cm ROI), and mean SUV within the delineated tumour (SUV_{mean}). Whereas SUV_{max} and SUV_{peak} are clearly independent of the tumour delineation strategy used, TL, TV, SUV_{mean} and the derived TLG values might depend on the delineation process. To study the impact of this step, we considered two different approaches; namely the automatic Fuzzy Locally Adaptive Bayesian (FLAB) algorithm [16-17] and an adaptive threshold algorithm [18] optimized for the GEMINI PET/CT scanner. Although the first approach is fully automatic, adaptive thresholding requires a manually defined background region of interest (ROI). Therefore two experienced nuclear medicine physicians were considered in the background ROI definition, leading to two series of results denoted as T_{A1} and T_{A2}. TL was determined in longitudinal direction by multiplying the number of slices in the delineated tumour

volume by the PET image slice thickness (4mm). TV was defined as the sum of all voxels contained in the delineated volumes multiplied by the image voxel's volume ($64mm^3$). Finally, TLG was determined by multiplying the SUV_{mean} and associated TV.

2.4 Statistical analysis

The relation between response to therapy and each parameter distribution was studied using the Kruskal-Wallis test [19] as recommended for small, not normally distributed samples.

Receiver operating characteristic (ROC) methodology [20] was used to assess the performance of each parameter to differentiate patients. Two classification tasks were considered: differentiating CR patients from PR and NR, or NR patients from CR and PR. Evaluation was performed in terms of the area under the curve (AUC) as well as specificity and sensitivity.

The significance of the following factors was tested: age, gender, T, N, and M classifications, AJCC (American Joint Committee on Cancer) stage, histology types, SUV_{max}, SUV_{peak}, TL, TV, SUV_{mean}, and TLG. All tests were two-sided and p values <0.05 were considered statistically significant.

3. Results

The range of values for the different image derived indices as well as the mean and standard deviation for the patient population considered are given in table II. All primary lesions were detected by ^{18}F-FDG PET exhibiting a rather high uptake with a SUV_{max} of 9.7±3.9. SUV_{peak} and SUV_{mean} measurements were comparatively lower (8.0±3.3 and 5.8±2.4 respectively).

Correlation between image derived indices and between methodologies

TV and TL measurements were moderately correlated (r=0.77, 0.68 and 0.60 for FLAB, T_{A1} and T_{A2} respectively, p<0.0001). On the other hand, no significant correlation was found between TV and any of the SUV measurements (r<0.2, p>0.1), irrespective of the delineation approach used. High correlations were observed between the TV (r>0.89), TL (r>0.90) or TLG (r>0.93) measurements obtained with the two delineation strategies (p<0.0001). Even higher correlation coefficients (r>0.97, p<0.0001) were observed for the SUV_{mean} measurements derived using the two different tumour segmentation approaches (FLAB and adaptive thresholding). Despite these correlations, certain large differences were observed for few patients between the delineation results of the two segmentation algorithms considered, examples of which are illustrated in figure 1.

Response to therapy analysis

Out of the 50 patients included in the study 25 were classified as PR, while there were 12 CR and 13 NR. Results concerning the predictive value of all considered parameters are summarized in tables III and IV containing the results of the Kruskal-Wallis tests and that of the ROC analysis (considering the AUC, specificity and sensitivity regarding the classification tasks) respectively.

Age, gender, or T, N, M classifications did not allow significant prediction of the response to treatment. The AJCC stage was not significantly (p>0.05) associated with the type of response, despite the fact that all NR were at least stage IIB and could be statistically differentiated from both PR and CR (p<0.05). However, AJCC

stage could not differentiate PR from CR (p>0.05). Finally, there was no statistical correlation between histology type and response *(p=0.3)*.

Figure 2 shows a graphical comparison of the Kruskal-Wallis results considering the predictive value of the different SUV parameters considered. Initial SUV_{max} (fig 2.A) was not predictive of response to therapy (p=0.29) although CR tended to have smaller SUV_{max} (8.1±4.1) than PR and NR (10.2±3.7 and 10.2±3.9 respectively). Similarly, SUV_{peak} (fig 2.B) was not predictive of response to therapy with a mean value of 6.5±3.5 in CR, whereas both PR and NR were characterized by similar higher SUV_{peak} values (8.5±3.1 and 8.4±3.3 respectively) (p=0.14). None of the SUV_{mean} measurements, whatever delineation strategy was used, could significantly predict response to therapy (p>0.19).

On the contrary, all parameters related to tumour spatial extent (TL, TV and TLG) measurements allowed significant (p<0.002) differentiation of the three response groups, irrespective of the segmentation methodology (see figure 3.A-C). For instance, TV as measured by FLAB was 20±25, 32±24 and 72±40 cm^3 for CR, PR and NR patients respectively. The parameter that allowed the best differentiation between the three patient groups was TLG measured by FLAB (K-W test p<0.0001, see figure 3.C), with a TLG of 74±75g, 179±143g and 385±226g for CR, PR and NR patients respectively. Figure 4 shows examples of one CR, one PR and one NR patient with corresponding TLG values.

The ROC analysis results confirmed the limited predictive value of most SUV measurements for the accurate classification of either CR vs. PR and NR, or NR vs. PR and CR (AUC<0.70 and <0.56 respectively). Differences between ROC analysis associated with SUV measurements and those associated with TL, TV and TLG was significant (p<0.05) for both tasks (see examples in figure 5). Better predictive

performances were obtained with TL, TV and TLG measurements with significantly higher AUC (from 0.74 to 0.86) for both tasks (p<0.05). For instance, using FLAB a TLG <58g allowed identifying complete responders with a sensitivity of 75% and a specificity of 92%, and a TLG >196g identified NRs with a sensitivity of 76% and a specificity of 85%. However in terms of predictive performance no significant differences were obtained between TL, TV and TLG measurements for both tasks. In terms of an observed trend, better results were obtained for TLG over TV and TL whatever tumour delineation approach was used (tables III and IV). In addition there was a systematic although not statistically significant trend of better performance for those parameters when obtained with FLAB compared to the use of the adaptive threshold, as demonstrated by higher AUC and smaller confidence intervals, as well as higher sensitivity and specificity for both classification tasks (table IV).

The analysis with respect to histology type (adenocarcinoma vs. squamous cell carcinoma) led to similar results with what was observed when considering the entire population. Within the same context no statistically significant differences were observed between the two patient groups in the hierarchy of parameters and results derived using the different functional tumour volume delineation methods.

The predictive value of TLG, combining TV and SUV_{mean} into one single parameter, was higher than the one of tumour volume, despite the non-significant value of SUV_{mean} alone. Considering together TV and SUV_{mean}, one is able to differentiate different treatment response patient groups (see figure 6). On the one hand, TLG increased the differentiation between CR and NR, as all NR had either a TV above 50cm^3 (8/13) or a SUV_{mean} above 5 (8/13), while 10 out of 12 CR had either a small TV (<15cm^3) (9/12) or SUV_{mean} (<5) (7/12), and half of them (6/12) had both. On the other hand, PR had either a higher SUV_{mean} than CR for volumes below

25cm^3 (6.5±2.7 vs. 4.5±2.4), or lower SUV_{mean} than NR for TV of 25-50cm^3 (5.8±1.8 vs. 7.1±0.9). Therefore the use of TLG increased the differentiation between PR and CR, as well as between PR and NR for volumes below 15cm^3 and between 25 and 50cm^3 respectively.

4. Discussion

Assessment of response to therapy early during treatment plays an important role in patient management as well as in drug development and new criteria including PET have been suggested for this task [21-22]. However, being able to predict response to therapy before the initiation of the treatment would be even more powerful for patient management. In this context, either patient or tumour characteristics could be considered. In our study we focused on functional imaging and different image derived parameters related to tumour uptake using PET. The results of our study demonstrate that tumour volume based parameters derived from baseline FDG PET images in oesophageal cancer are good predictors of response to therapy, with high TL, TV and TLG being associated with poor response to combined radio-chemotherapy. On the contrary, more commonly used parameters such as tumour SUVs were not predictors of response to therapy considering only the baseline FDG PET images. These results further demonstrate the value of tumour volume based PET image derived parameters, since we have previously demonstrated a superior prognostic value of baseline functional TL, TV and TLG over SUV measurements for overall survival on a similar group of oesophageal cancer patients [23].

FDG PET has been previously used for the prediction of response to therapy or prognosis in a variety of malignancies [24]. Considering the predictive value of baseline FDG uptake for therapy response in oesophageal cancer, only few data

showing conflicting results are available [12]. Levine et al. and Rizk et al. reported a high initial SUV_{max} being associated with good response [25-26], whereas Makino et al. and Kato et al. found the opposite [27-28]. These conflicting results can be potentially attributed to differences in patient populations, tumour histology types, as well as treatment, but could also suggest that SUV measurements are unreliable in this context. Although similarly to the results of Kato et al. and Makino et al., our results suggest that lower values of SUV_{max} are associated with a complete response, this trend was not significant. In addition, SUV_{mean} or SUV_{peak}, considered more robust to potential noise bias associated with SUV_{max}, were also not significant predictors of response to therapy in our study.

One of the demonstrated independent predictors of long term survival in oesophageal cancer is longitudinal tumour extension established by pathological examination [29]. It has been previously demonstrated that TL measured on CT images leads to a weak correlation with the pathological TL, associated with a large overestimation [30]. Some authors proposed the estimation of metabolic TL as a surrogate of pathological TL using various thresholds of ^{18}F-FDG PET uptake [31] however conflicting results concerning the predictive value of metabolic TL for response to neoadjuvant radio-chemotherapy have been observed [32-33]. One may argue that TL does not reflect the entire volume of the tumour and could therefore be only considered as a limited surrogate measure of tumour functional spatial extent. This assumption is partly supported by our data, in which only a moderate correlation (r between 0.6 and 0.77) was found between TV and TL, suggesting that TV may bring additional information compared to TL in assessing overall tumour burden. In our study both TV and TL were found to be significant predictive factors of response

to therapy, irrespective of the functional volume delineation strategy, with only a small and non significant improvement of the predictive value of TV over TL.

TV and TLG measured on PET are 3D measurements incorporating metabolically active tumour volume not available from CT data [34]. It has already been demonstrated that a decrease of the TV and TLG can predict response to therapy [35-36]. These studies however have explored differences in indices derived from serial PET images. The value of such indices obtained on the baseline scan only within the context of therapy response prediction in oesophageal cancer has not previously been explored. Because these parameters reflect metabolic information in the entire tumour, they may be more accurate for tumour characterization than a single voxel measure and this may explain why TV and TLG were good predictors of therapy response as demonstrated in our study. Our results are consistent with recent studies in pleural mesothelioma and lymphoma patients that have demonstrated the potential of such indices extracted from baseline 18F-FDG PET scan to predict response to therapy [37-38].

Despite a great potential value, such indices have been only of limited use to date, which can be explained by the limited accuracy, robustness and reproducibility of the available tumour delineation tools [39-40]. In oesophageal cancer only the prognostic value of TV has been studied [23, 41], while there is limited data on the value of TLG [23]. In our study TLG allowed identifying complete responders and non responders with moderate sensitivity (75% and 76% respectively) and high specificity (92% and 85% respectively). Prospective studies with a larger patient population using a predictive model built upon our results should now be carried out to demonstrate the ability of the parameters to discriminate responders from non responders on a patient by patient basis.

In our study, TNM stage and AJCC classification were not good predictors of therapy response. This could be explained by our suboptimal staging procedure. Since we considered only patients referred for exclusive radio-chemotherapy, no patient underwent surgery, and therefore no pathological data was available. Staging was routinely performed using endoscopic ultrasonography and CT which are known to have limited staging performances [10].

Our present study has limitations. Firstly, we considered a group of only 50 patients with predominantly squamous cell carcinomas since it is the most common histological type of oesophageal cancer in European countries. An analysis based on the tumour histology type considering our patient population did not reveal statistically significant differences, although due to the small number of patients with adenocarcinomas, these results would obviously need to be confirmed. Secondly, our study was inherently limited by its retrospective design and as such some selection bias might be present. However, the treatment regime was homogeneous throughout the recruited patients since all were treated in a single institution. In addition, within this patient population no particular selection criteria were applied. Thirdly, the impact of partial volume effects in the measured SUVs was not assessed in this study. The lack of partial volume correction might have played a role in the reduced predictive value of some of the SUV measurements, although it is unlikely because of the large tumour volumes considered in this work ($40\pm35 cm^3$). Lastly, we did consider only primary tumours since the measurements used are simpler to perform in routine clinical practice compared to measurement of overall tumour burden including primary and metastatic lesions. However, given the respective size of metastatic lesions and primary tumours, adding metastatic lesions to the overall TLG would not significantly alter the resulting values and associated conclusions.

5. Conclusion

Our results demonstrated that ^{18}F-FDG baseline image derived parameters related to the metabolic tumour spatial extent (TL, TV and TLG) are good predictors of response to therapy in oesophageal cancer with sensitivity above 75% and specificity above 85%. Commonly used SUV measurements (max, peak, mean) on pre-treatment FDG PET image did not allow statistically significant differentiation of the different response patient groups.

Acknowledgments

This work was partly funded by ANR (French National Research Agency) under the contract ANR-08-ETEC-005-01.

Conflict of interest statement:

The authors declare that they have no conflict of interest

Figures captions

Figure 1: Illustration of differences in tumour delineation depending on the methodology for two patients.

Figure 2: Distributions of NR, PR and CR patients and associated Kruskal-Wallis tests for SUV based image derived indices: (A) SUV_{max} and (B) SUV_{peak}.

Figure 3: Distributions of NR, PR and CR patients and associated Kruskal-Wallis tests for tumour volume related image derived indices: (A) TL (T_{A2}), (B) TV (T_{A1}), and (C) TLG (FLAB).

Figure 4: 18F-FDG PET axial, coronal and sagital images of a (A) complete responder with 20g TLG, (B) partial responder with 100g TLG and (C) non-responder with 750g TLG.

Figure 5: Examples of ROC curves obtained for classification tasks of differentiating (A) CR from NR&PR or (B) NR from PR&CR. Comparison of ROC curves for SUV measurements (SUV_{max} in red, SUV_{peak} in orange and SUV_{mean} in yellow) and TL, TV and TLG measured with FLAB (in light blue, blue and dark blue respectively).

Figure 6: Distribution of CR, PR and NR patients according to their SUV_{mean} and TV as measured by FLAB.

Table captions

Table I: Patient demographic and clinical characteristics

Table II: Image derived parameters definition and associated summary statistics.

Table III: Kruskal-Wallis test results for each parameter considering the ability to differentiate ($p<0.05$) each pair of response group.

Table IV: ROC analysis results with area under the curve (AUC) and associated 95% confidence intervals (CI), specificity (sp) and sensitivity (se) for each parameter regarding the two classification tasks.

References

1. Parkin, D.M., et al., *Global cancer statistics, 2002.* CA Cancer J Clin, 2005. **55**(2): p. 74-108.
2. Hayat, M.J., et al., *Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program.* Oncologist, 2007. **12**(1): p. 20-37.
3. Gebski, V., et al., *Survival benefits from neoadjuvant chemoradiotherapy or chemotherapy in oesophageal carcinoma: a meta-analysis.* Lancet Oncol, 2007. **8**(3): p. 226-34.
4. Kelsen, D.P., et al., *Long-term results of RTOG trial 8911 (USA Intergroup 113): a random assignment trial comparison of chemotherapy followed by surgery compared with surgery alone for esophageal cancer.* J Clin Oncol, 2007. **25**(24): p. 3719-25.
5. Chirieac, L.R., et al., *Posttherapy pathologic stage predicts survival in patients with esophageal carcinoma receiving preoperative chemoradiation.* Cancer, 2005. **103**(7): p. 1347-55.
6. Stahl, M., et al., *Clinical response to induction chemotherapy predicts local control and long-term survival in multimodal treatment of patients with locally advanced esophageal cancer.* J Cancer Res Clin Oncol, 2005. **131**(1): p. 67-72.
7. Dragovich, T. and C. Campen, *Anti-EGFR-Targeted Therapy for Esophageal and Gastric Cancers: An Evolving Concept.* J Oncol, 2009. **2009**: p. 804108.
8. Makino, T., et al., *p53 Mutation status predicts pathological response to chemoradiotherapy in locally advanced esophageal cancer.* Ann Surg Oncol, 2010. **17**(3): p. 804-11.
9. Lee, J.M., et al., *Polymorphism in Epidermal Growth Factor Receptor Intron 1 Predicts Prognosis of Patients with Esophageal Cancer after Chemoradiation and Surgery.* Ann Surg Oncol, 2011.
10. van Westreenen, H.L., et al., *Systematic review of the staging performance of 18F-fluorodeoxyglucose positron emission tomography in esophageal cancer.* J Clin Oncol, 2004. **22**(18): p. 3805-12.
11. Krause, B.J., et al., *18F-FDG PET and 18F-FDG PET/CT for assessing response to therapy in esophageal cancer.* J Nucl Med, 2009. **50 Suppl 1**: p. 89S-96S.
12. Kwee, R.M., *Prediction of tumor response to neoadjuvant therapy in patients with esophageal cancer with use of 18F FDG PET: a systematic review.* Radiology, 2010. **254**(3): p. 707-17.
13. Ott, K., et al., *Metabolic imaging predicts response, survival, and recurrence in adenocarcinomas of the esophagogastric junction.* J Clin Oncol, 2006. **24**(29): p. 4692-8.
14. Therasse, P., et al., *New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada.* J Natl Cancer Inst, 2000. **92**(3): p. 205-16.
15. Larson, S.M., 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**(3): p. 159-171.
16. Hatt, 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**(1): p. 301-8.
17. Hatt, M., et al., *A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET.* IEEE Trans Med Imaging, 2009. **28**(6): p. 881-93.
18. Schaefer, A., et al., *A contrast-oriented algorithm for FDG-PET-based delineation of tumour volumes for the radiotherapy of lung cancer: derivation from phantom measurements and validation in patient data.* Eur J Nucl Med Mol Imaging, 2008. **35**(11): p. 1989-99.
19. Kruskal, W. and W. Wallis, *Use of ranks in one-criterion variance analysis.* Journal of the American Statistical Association, 1952. **47**(260): p. 583-621.
20. Metz, C.E., *Basic principles of ROC analysis.* Semin Nucl Med, 1978. **8**(4): p. 283-98.
21. Wahl, R.L., et al., *From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors.* J Nucl Med, 2009. **50 Suppl 1**: p. 122S-50S.
22. Hofman, M.S. and R.J. Hicks, *Restaging: should we percist without pattern recognition?* J Nucl Med, 2010. **51**(12): p. 1830-2.
23. Hatt, M., et al., *Prognostic value of 18F-FDG PET image-based parameters in esophageal cancer: impact of tumor delineation methodology.* European Journal of Nuclear Medicine and Molecular Imaging, 2011.
24. Lucignani, G. and S.M. Larson, *Doctor, what does my future hold? The prognostic value of FDG-PET in solid tumours.* Eur J Nucl Med Mol Imaging, 2010. **37**(5): p. 1032-8.
25. Levine, E.A., et al., *Predictive value of 18-fluoro-deoxy-glucose-positron emission tomography (18F-FDG-PET) in the identification of responders to chemoradiation therapy for the treatment of locally advanced esophageal cancer.* Ann Surg, 2006. **243**(4): p. 472-8.
26. Rizk, N.P., et al., *Predictive value of initial PET-SUVmax in patients with locally advanced esophageal and gastroesophageal junction adenocarcinoma.* J Thorac Oncol, 2009. **4**(7): p. 875-9.

27. Makino, T., et al., *Utility of response evaluation to neo-adjuvant chemotherapy by (18)F-fluorodeoxyglucose-positron emission tomography in locally advanced esophageal squamous cell carcinoma.* Surgery, 2010. **148**(5): p. 908-18.
28. Kato, H., et al., *Prediction of response to definitive chemoradiotherapy in esophageal cancer using positron emission tomography.* Anticancer Res, 2007. **27**(4C): p. 2627-33.
29. Yendamuri, S., et al., *Esophageal tumor length is independently associated with long-term survival.* Cancer, 2009. **115**(3): p. 508-16.
30. Sillah, K., et al., *Computed tomography overestimation of esophageal tumor length: Implications for radiotherapy planning.* World J Gastrointest Oncol, 2010. **2**(4): p. 197-204.
31. Zhong, X., et al., *Using 18F-fluorodeoxyglucose positron emission tomography to estimate the length of gross tumor in patients with squamous cell carcinoma of the esophagus.* Int J Radiat Oncol Biol Phys, 2009. **73**(1): p. 136-41.
32. Mamede, M., et al., *FDG-PET/CT tumor segmentation-derived indices of metabolic activity to assess response to neoadjuvant therapy and progression-free survival in esophageal cancer: correlation with histopathology results.* Am J Clin Oncol, 2007. **30**(4): p. 377-88.
33. Roedl, J.B., et al., *Assessment of treatment response and recurrence in esophageal carcinoma based on tumor length and standardized uptake value on positron emission tomography-computed tomography.* Ann Thorac Surg, 2008. **86**(4): p. 1131-8.
34. Hong, T.S., et al., *Impact of manual and automated interpretation of fused PET/CT data on esophageal target definitions in radiation planning.* Int J Radiat Oncol Biol Phys, 2008. **72**(5): p. 1612-8.
35. Arslan, N., et al., *Evaluation of response to neoadjuvant therapy by quantitative 2-deoxy-2-[18F]fluoro-D-glucose with positron emission tomography in patients with esophageal cancer.* Mol Imaging Biol, 2002. **4**(4): p. 301-10.
36. Roedl, J.B., et al., *Adenocarcinomas of the esophagus: response to chemoradiotherapy is associated with decrease of metabolic tumor volume as measured on PET-CT. Comparison to histopathologic and clinical response evaluation.* Radiother Oncol, 2008. **89**(3): p. 278-86.
37. Lee, H.Y., 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**(10): p. 2787-94.
38. Cazaentre, T., 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**(3): p. 494-504.
39. Hatt, M., et al., *PET functional volume delineation: a robustness and repeatability study.* Eur J Nucl Med Mol Imaging, 2011.
40. Hatt, M., et al., *Reproducibility of 18F-FDG and 3'-deoxy-3'-18F-fluorothymidine PET tumor volume measurements.* J Nucl Med, 2010. **51**(9): p. 1368-76.
41. Hyun, S.H., et al., *Prognostic value of metabolic tumor volume measured by 18F-fluorodeoxyglucose positron emission tomography in patients with esophageal carcinoma.* Ann Surg Oncol, 2010. **17**(1): p. 115-22.