

FDG PET/CT for rectal carcinoma radiotherapy treatment planning: comparison of functional volume delineation algorithms and clinical challenges

Nadia Withofs,^{1a} Claire Bernard,¹ Catherine van der Rest,¹
Philippe Martinive,² Mathieu Hatt,³ Sebastien Jodogne,⁴
Dimitris Visvikis,³ John A. Lee,⁵ Philippe A. Coucke,²
Roland Hustinx¹

*Division of Nuclear Medicine and Oncological Imaging, Department of Medical Physics,¹ CHU of Liege, Liege, Belgium; Division of Radiation Oncology, Department of Medical Physics,² CHU of Liege, Liege, Belgium; Laboratoire de Traitement de l'Information Médicale (LaTIM) UMR 1101,³ Institut national de la santé et de la recherche médicale (INSERM), Brest, France; Department of Medical Physics,⁴ CHU of Liege, Liege, Belgium; Center of Molecular Imaging, Radiotherapy, and Oncology (MIRO),⁵ Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium
nwithofs@chu.ulg.ac.be*

Received 3 September, 2013; accepted 25 April, 2014

PET/CT imaging could improve delineation of rectal carcinoma gross tumor volume (GTV) and reduce interobserver variability. The objective of this work was to compare various functional volume delineation algorithms. We enrolled 31 consecutive patients with locally advanced rectal carcinoma. The FDG PET/CT and the high dose CT (CT_{RT}) were performed in the radiation treatment position. For each patient, the anatomical GTV_{RT} was delineated based on the CT_{RT} and compared to six different functional/metabolic GTV_{PET} derived from two automatic segmentation approaches (FLAB and a gradient-based method); a relative threshold (45% of the SUV_{max}) and an absolute threshold (SUV > 2.5), using two different commercially available software (Philips EBW4 and Segami OASIS). The spatial sizes and shapes of all volumes were compared using the conformity index (CI). All the delineated metabolic tumor volumes (MTVs) were significantly different. The MTVs were as follows (mean ± SD): GTV_{RT} (40.6 ± 31.28 ml); FLAB (21.36 ± 16.34 ml); the gradient-based method (18.97 ± 16.83 ml); OASIS 45% (15.89 ± 12.68 ml); Philips 45% (14.52 ± 10.91 ml); OASIS 2.5 (41.62 ± 33.26 ml); Philips 2.5 (40 ± 31.27 ml). CI between these various volumes ranged from 0.40 to 0.90. The mean CI between the different MTVs and the GTV_{CT} was < 0.4. Finally, the DICOM transfer of MTVs led to additional volume variations. In conclusion, we observed large and statistically significant variations in tumor volume delineation according to the segmentation algorithms and the software products. The manipulation of PET/CT images and MTVs, such as the DICOM transfer to the Radiation Oncology Department, induced additional volume variations.

PACS number: 87.55.D-

Key words: FDG, PET, rectal cancer, segmentation, metabolic volume

^a Corresponding author: Nadia Withofs, Service de Médecine nucléaire, CHU de Liège, B35, Domaine universitaire du Sart Tilman, 4000 Liège, Belgium; phone: +32 4 3667199; fax: +32 4 3668257; email: nwithofs@chu.ulg.ac.be

I. INTRODUCTION

Radiation oncologists increasingly use ^{18}F -fluorodeoxyglucose (FDG) positron emission tomography (PET) during the essential step of gross tumor volume (GTV) delineation. Before incorporating PET in the treatment planning process in a clinical setting, there are a number of factors affecting image quantification and subsequent functional volume segmentation that need to be identified. The first determinants are related to tumor biology (e.g., heterogeneous glycolytic activity within the tumor) and location (e.g., in a region with high surrounding background activity like bladder).^(1,2) Other factors are related to the physics principles of PET/CT acquisition and reconstruction protocols, which have all been identified as potential major factors.^(3,4)

Various (semi-) automatic segmentation methods have been developed. To be clinically implemented, the ideal segmentation algorithm should be accurate, robust, and reproducible, as well as user-friendly. The simplest algorithms are semiautomatic threshold-based methods relying on the standardized uptake value (SUV) of each tumor voxel. These methods are heavily influenced by the intensity of tumor uptake so that various thresholds are applied, depending on individual situation, and regularly fail when the tumor is small or heterogeneous, or when the surrounding background activity is high.^(2,5) More advanced threshold-based segmentation methods take into account the background activity and the signal to background ratio to define the optimal threshold.^(6,7) However, it has been also demonstrated that these adaptive threshold methods may fail to delineate small contrast or highly heterogeneous functional uptakes. Lastly, more advanced automatic algorithms based on image segmentation paradigms, such as gradient-based methods⁽⁸⁾ and the fuzzy locally adaptive Bayesian algorithm (FLAB),⁽⁹⁾ have been proposed. In principle, the advantage of these mostly automatic algorithms is their robustness and reproducibility, taking into account varying imaging conditions such as lesion size, heterogeneity, or tumor uptake intensity in contrast with the surrounding background activity, as well as variable noise characteristics in the reconstructed PET images.^(1,10,11,12)

The primary objective of this work was to compare metabolic GTV delineation, based on FDG PET/CT images and derived from two different automatic segmentation approaches and from threshold-based algorithms, using two commercially available software products. A secondary objective was to investigate volume variations related to the DICOM RT structure transfer between software products from a nuclear medicine to a radiation oncology department. This study focused on patients with rectal carcinoma. The FDG PET/CT is recommended for staging rectal carcinoma that usually displays high FDG uptake and provides a better interobserver agreement for rectal tumor delineation.^(13,14,15,16)

II. MATERIALS AND METHODS

A. Patients

We enrolled 31 consecutive patients with locally advanced rectal carcinoma (LARC) for which a FDG PET/CT was performed before chemoradiotherapy. The locoregional spread of the tumor was assessed using magnetic resonance imaging (MRI) and transrectal ultrasonography. The American Joint Committee on Cancer includes eight different T stages and nine N stages.⁽¹⁷⁾ We thus used a simplified approach and only distinguished the N0 from the N+ stage. The clinical stage was T2 N+ (N = 3); T3 N0 (N = 2); T3 N+ (N = 23), or T4 N+ (N = 3). Mean age of patients was 65 ± 12 yr. The retrospective analysis of the data was approved by the Ethics Committee of our Institution.

B. FDG PET/CT

All 31 FDG PET/CT studies were performed on a Gemini TF PET/CT system (Philips Medical Systems, Cleveland, OH). Images were acquired 60 min after injection of 4 MBq/kg of FDG

(220–440 MBq). All patients were positioned in the radiation treatment position on a flat pallet with the aid of a dedicated RT laser system placed in front of the PET/CT gantry. The low-dose CT (5 mm slice thickness; tube voltage: 120 kV and tube current–time product: 50 to 80 mAs, depending on the patient’s weight) was followed by the PET emission scan with a time per bed position (pbp) depending on the patient’s body mass index (BMI ≤ 25 : 1 min pbp; BMI ≥ 26 and ≤ 32 : 1 min 30 sec pbp; BMI ≥ 33 : 2 min pbp). Data were reconstructed using time of flight (TOF) information, as well as correction for decay, scatter, random, and attenuation (CT data were used for attenuation correction). The reconstructed CT matrix was 512×512 (voxel size $1.17 \times 1.17 \times 5$ mm³) and the PET matrix size was 144×144 (voxel size: $4 \times 4 \times 4$ mm³).

C. CT simulation (CT_{RT})

All CT_{RT} were performed in the Department of Radiation Oncology on a Big Bore system (Philips Medical Systems, Cleveland, OH) using standard clinical parameters (tube voltage: 120 kV and tube current–time product: 250 mAs). For 21 of the 31 patients, the matrix size of the CT images was 1024×1024 (voxel size $0.59 \times 0.59 \times 3$ mm³). For 10 of the 31 patients, the matrix size of the CT images was 512×512 (voxel size $1.17 \times 1.17 \times 3$ mm³). The GTV (GTV_{RT}) was delineated on the CT_{RT} images by a radiation oncologist with a 12-year experience in digestive oncology. He had also access to the FDG PET/CT and MR images. The median (range) delay between the FDG PET/CT and the CT_{RT} was five days (1–12 days).

D. Image coregistration

For each patient, the low dose CT data and the CT_{RT} data were automatically coregistered using the automatic rigid body registration tool called “local correlation” in the Extended Brilliance Workspace EBW-NM 1.5.1 (EBW4, Philips Medical Systems). The same matrix transformation was applied to the FDG PET data. As most commercially available viewers resample the PET images into the CT, we applied a similar methodology for the coregistration process (e.g., all PET images were resampled onto the CT_{RT} space leading to the same voxel size as the CT_{RT}).

E. Segmentation methods

The coregistered FDG PET/CT data were used to delineate the metabolic tumor volumes (MTVs) using two software products: OASIS V1.8.3 (Segami Corporation, Columbia, MD) and EBW (Philips). In each case, the MTVs were delineated using two threshold-based methods. One was based on a relative threshold, including all tumor voxels for which the activity reached 45% of the maximal SUV within the tumor. The other was based on an absolute threshold set at 2.5 SUV and including all tumor voxels for which the SUV was equal or superior to 2.5.^(18,19) From here onwards, the volumes delineated using the OASIS software will be referred to as OA2.5 and OA45%, and the one delineated with the Philips EBW4 as PH2.5 and PH45%.

In addition, two segmentation algorithms were applied to automatically delineate the MTV: FLAB, taking into account image noise and limited spatial resolution through statistical and fuzzy modeling,⁽⁹⁾ and a gradient-based method (GBM) relying on the watershed transform and cluster analysis for segmenting preprocessed PET images that are first denoised using an edge-preserving filter and deblurred using iterative deconvolution.⁽⁸⁾ Whatever the segmentation algorithm, all manipulations were performed by a nuclear medicine physician with 10 years of experience in oncological PET imaging, who identified the tumors and made sure no physiological uptake was included in the volumes delineated by the segmentation algorithms.

F. Data analysis

The matrix sizes of the CT_{RT} (1024×1024) and the low-dose CT (512×512) of the PET/CT were different for 21 patients out of 31. For these 21 patients, we resampled the images of the PET and the low-dose CT before applying any segmentation algorithm in order to match to the CT_{RT} matrix size of 1024×1024 . As a preliminary step in this subpopulation, we applied the different segmentation algorithms on both resampled and non-resampled PET images for

each 21 patients. We then compared the MTVs delineated according to whether the matrix had been resampled or not. We also compared the SUV_{max} and SUV_{mean} extracted from the volumes obtained with or without resampling.

We considered the GTV that was used for treating the patient (GTV_{RT}) as the reference.

In order to visualize the spatial distribution of all delineated MTVs in a single PET image, we exported all MTVs and the GTV_{RT} in the PMOD software (version 3.207; PMOD Technologies Ltd., Zurich, Switzerland). To estimate if the spatial distribution of the different MTVs were concordant, the conformity index (CI) was calculated between each MTV two by two, and between each MTV and the GTV_{RT} . The CI was calculated by dividing the intersection volume by the conjunction volume.⁽²⁰⁾

Lastly, we transferred MTVs from OASIS and Philips Brilliance software products to Pinnacle³ software (Philips Medical Systems) in the Radiation Oncology Department. Each volume was transferred to Pinnacle and PMOD as a DICOM RT structure set via a network connection. We compared the transferred volumes to the initial volumes displayed in OASIS or Philips in order to estimate if the DICOM transfer to Pinnacle or PMOD had an impact on the delineated volume.

G. Statistical analyses

In the 21/31 patients with a CT_{RT} matrix size of 1024×1024 , the Student's *t*-test and Wilcoxon signed-rank test were performed to estimate if the resampling of the PET images had an impact on the delineated MTVs and SUV measurements. In the next step, we compared the MTVs delineated in all the 31 subjects using the ANOVA-2 or Friedman tests. When the ANOVA-2 test was significant, a post-hoc test (Scheffé test) was performed to compare volumes derived from the different segmentation algorithms two by two. Results were considered to be significant at the 5% level ($p < 0.05$). Calculations were done using SAS version 9.2 (SAS Institute, Cary, NC).

III. RESULTS

A. Impact of PET images resampling on metabolic tumor volumes and SUVs

The resampling of PET images significantly decreased the volumes delineated in the Philips EBW4 workspace (Figs. 1(a) and (b)). The mean⁽²¹⁾ relative variation of the delineated MTV was 26.9% (28.5%) for the PH45% and 7.39% (20.90%) for the PH2.5. The SUV_{max} and SUV_{mean} extracted from the delineated volumes using the Philips EBW4 workspace were significantly higher on resampled PET images than the SUV recorded on non-resampled PET images, using both threshold-based (PH45%; PH2.5) methods (Figs. 1(c) and (d)). On the other hand, the MTVs delineated in OASIS (OA45% and OA2.5) and with the FLAB method were not statistically modified by the resampling.

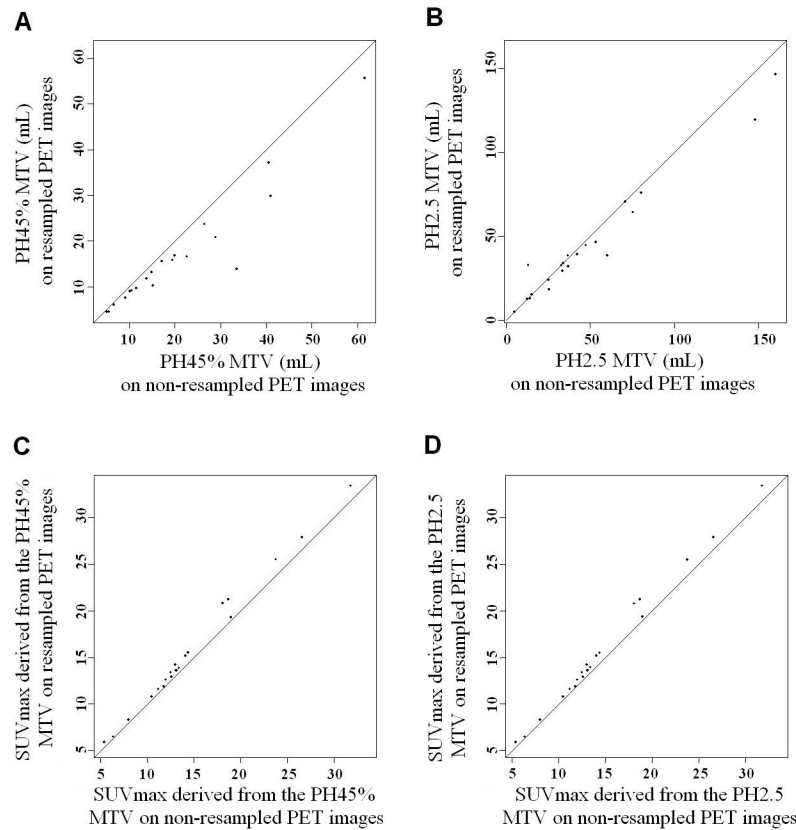


FIG. 1. The scatter-plots (a) and (b) show the metabolic tumor volumes (MTV) without resampling PET images (x-axis) plotted against the MTV (mL) delineated on resampled PET images (y-axis) using the Philips EBW4 software and a 45% SUV_{max} threshold (a) or the SUV2.5 threshold (b). The scatter-plots (c) and (d) show the SUV_{max} extracted from the volumes delineated without resampling PET images (x-axis) plotted against the SUV_{max} extracted from the volumes delineated on resampled PET images (y-axis) using the Philips EBW4 software and a 45% SUV_{max} threshold (c) or the SUV2.5 threshold (d). The resampling of PET images significantly modified the volume delineated in the Philips EBW4 workspace and the extracted SUV_{max} whatever the threshold-based method used (PH45% or PH2.5).

B. Impact of the segmentation algorithm on the volumes and SUVs

The delineated volumes obtained in the 31 patients are presented in Fig. 2. Taken together, the MTVs were significantly different between all approaches ($p < 0.0001$). All MTVs were significantly smaller than the GTV_{CT} , except those defined with the SUV2.5 threshold-based methods (PH2.5 and OA2.5). There were no significant differences between the 45% SUV_{max} threshold (PH45% and OA45%) and the automatic algorithms (gradient-based and FLAB methods), between PH2.5 and OA2.5 or between PH45% and OA45%. Finally, the gradient-based and FLAB volumes were not significantly different.

The MTVs delineated using the same threshold (SUV2.5 or 45% SUV_{max}) in two distinct software products (Philips EBW4 or OASIS) were the most consistent; the intraclass coefficients and the 95% confidence interval lower limits are presented in Table 1.

We also found a significant difference in the spatial distribution of the different MTVs, even when the segmentation algorithm was identical but applied by two different software products (Philips EBW4 or OASIS). The conformity indices (CI) are presented in Table 2. The largest mean CI (≥ 0.87) was observed between the volumes delineated using an identical threshold-based method (45% SUV_{max} or SUV2.5) applied in two different software products (OASIS and Philips EBW4). The CI between all other methods with one another and with the GTV_{CT} were

≤ 0.70 . The values of SUV_{mean} extracted from the different MTVs were significantly different, but the SUV_{max} were identical whatever the chosen algorithms (data not shown).

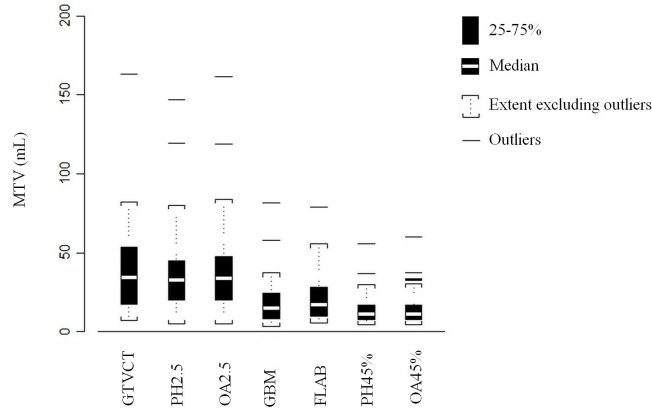


FIG. 2. Box-Whisker plot of tumor metabolic tumor volumes (MTV) delineated using threshold-based methods (45% SUV_{max} or $SUV_{2.5}$) using two different software products (OASIS, Philips EBW4) and automatic segmentation algorithms using the GBM validated by Geets et al.⁽⁸⁾ or a FLAB segmentation approach.

TABLE 1. Intraclass correlation coefficients (95% confidence interval lower limit) for all methods and two by two.

	<i>All Methods</i>	<i>GTV_{RT}</i>	<i>FLAB</i>	<i>OA45%</i>	<i>OA2.5</i>	<i>PH45%</i>	<i>PH2.5</i>	<i>GBM</i>
Overall	0.55 (0.38)							
GTV _{RT}			0.50 (0.09)	0.31 (-0.02)	0.85 (0.74)	0.26 (-0.04)	0.84 (0.72)	0.54 (0.02)
FLAB				0.86 (0.51)	0.49 (0.09)	0.80 (0.31)	0.52 (0.10)	0.91 (0.83)
OA45%					0.33 (-0.02)	0.95 (0.90) ^a	0.36 (-0.01)	0.80 (0.66)
OA2.5						0.28 (-0.03)	0.99 (0.98) ^a	0.53 (0.02)
PH45%							0.31 (-0.03)	0.73 (0.53)
PH2.5								0.56 (0.03)

^a The metabolic tumor volumes delineated using the same threshold (45% SUV_{max} or $SUV_{2.5}$) in two distinct software products (Philips EBW4 or OASIS) were the most consistent.

TABLE 2. Characteristics of the conformity indices between each metabolic tumor volume two by two.

Pair	N	Mean	SD	SE	Min	Max	Student's t-test & Wicoxon test (H0: Mean = 1)
GTV _{RT} vs. FLAB	31	0.31	0.17	0.03	0.0	0.7	<0.0001
GTV _{RT} vs. OA45%	31	0.24	0.16	0.03	0.0	0.6	<0.0001
GTV _{RT} vs. OA2.5	31	0.36	0.21	0.04	0.0	0.7	<0.0001
GTV _{RT} vs. PH45%	31	0.24	0.14	0.03	0.0	0.6	<0.0001
GTV _{RT} vs. PH2.5	31	0.38	0.21	0.04	0.0	0.7	<0.0001
GTV _{RT} vs. GBM	31	0.25	0.15	0.03	0.0	0.6	<0.0001
FLAB vs. OA45%	31	0.69	0.13	0.02	0.4	0.9	<0.0001
FLAB vs. OA2.5	31	0.53	0.17	0.03	0.3	1.0	<0.0001
FLAB vs. PH45%	31	0.69	0.12	0.02	0.5	0.9	<0.0001
FLAB vs. PH2.5	31	0.53	0.15	0.03	0.3	1.0	<0.0001
FLAB vs. GBM	31	0.68	0.11	0.02	0.4	0.8	<0.0001
OA45% vs. OA2.5	31	0.40	0.16	0.03	0.2	0.9	<0.0001
OA45% vs. PH45%	31	0.87	0.10	0.02	0.5	1.0	<0.0001 ^a
OA45% vs. GBM	31	0.70	0.12	0.02	0.4	0.8	<0.0001
OA2.5 vs. PH2.5	31	0.90	0.05	0.01	0.7	1.0	<0.0001 ^a
OA2.5 vs. GBM	31	0.42	0.13	0.02	0.2	0.7	<0.0001
PH45% vs. PH2.5	31	0.40	0.15	0.03	0.2	0.9	<0.0001
PH45% vs. GBM	31	0.70	0.14	0.03	0.4	1.0	<0.0001
PH2.5 vs. GBM	31	0.42	0.12	0.02	0.2	0.7	<0.0001

^a The spatial distribution of the metabolic tumor volumes was significantly different even if the segmentation algorithm was identical (45% SUV_{max} or $SUV_{2.5}$), but applied by two different software products (Philips EBW4 or OASIS). SD = standard deviation; SE = standard error; Min = minimum; Max = maximum.

C. Impact of the DICOM transfer on the volumes

The transfer to PMOD significantly reduced the GTV_{CT} , the volumes delineated with OASIS software (OA2.5 and OA45%) and the gradient-based volumes. It did not modify the volumes delineated with the Philips software (PH2.5 and PH45%) or FLAB. The volume changes are presented in Fig. 3. The Bland-Altman plots show that the greater the MTV, the greater the difference in volume after the transfer in PMOD.

The transfer of the MTVs to Pinnacle significantly changed the MTVs, whichever commercial algorithm was used (Fig. 4). The mean difference was -0.28 ± 0.28 mL ($p = 0.0002$) and -0.25 ± 0.22 mL ($p < 0.0001$) for the volumes delineated in OASIS with 45% and 2.5 thresholds, respectively. Conversely, the transfer from Philips EBW4 to Pinnacle systematically led to an expansion of the volume. The mean difference was 0.21 ± 0.13 mL ($p < 0.0001$) and 0.33 ± 0.16 mL ($p < 0.0001$) for the volumes delineated in Philips EBW4 with 45% and 2.5 thresholds, respectively.

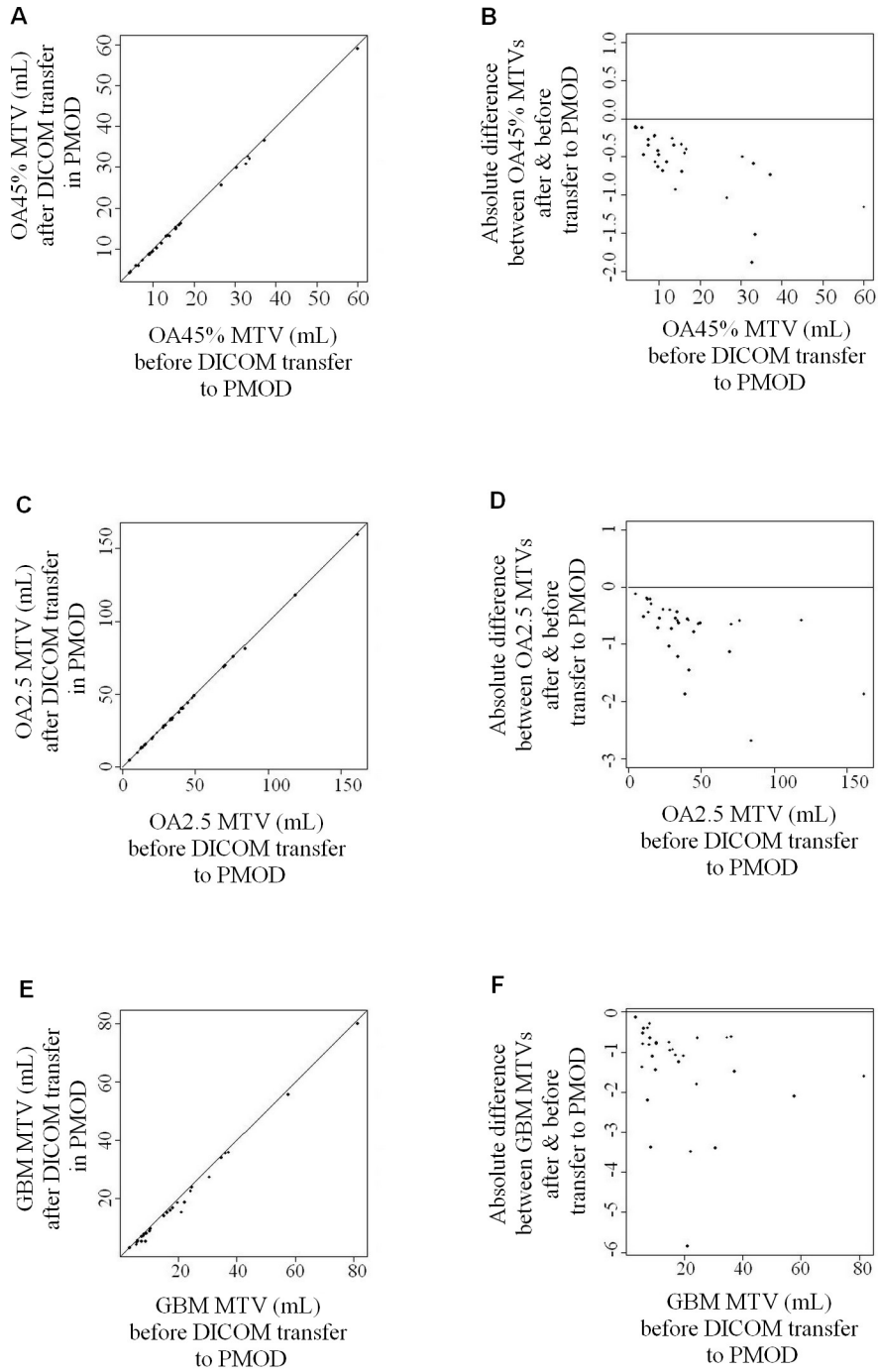


FIG. 3. The scatter plots ((a), (c), (e)) show the metabolic tumor volumes (MTV) before the transfer to PMOD (x-axis) plotted against the MTV (mL) after transfer in PMOD (y-axis) for OASIS ((a): OA45%; (c): OA2.5) and for the gradient-based method (e). The Bland-Altman bias plots ((b), (d), (f)) show the difference between the volumes before the transfer to PMOD (x-axis) against the absolute difference between the volume after the transfer in PMOD and the volume before transfer (y-axis). The importation of the MTVs in PMOD significantly reduced the volumes delineated with OASIS software (OA2.5 & OA45%), and particularly the volumes delineated using the GBM for which the mean difference was superior to 1 mL (mean difference: $-1.38 \text{ mL} \pm \text{SD: } 1.21 \text{ mL}$).

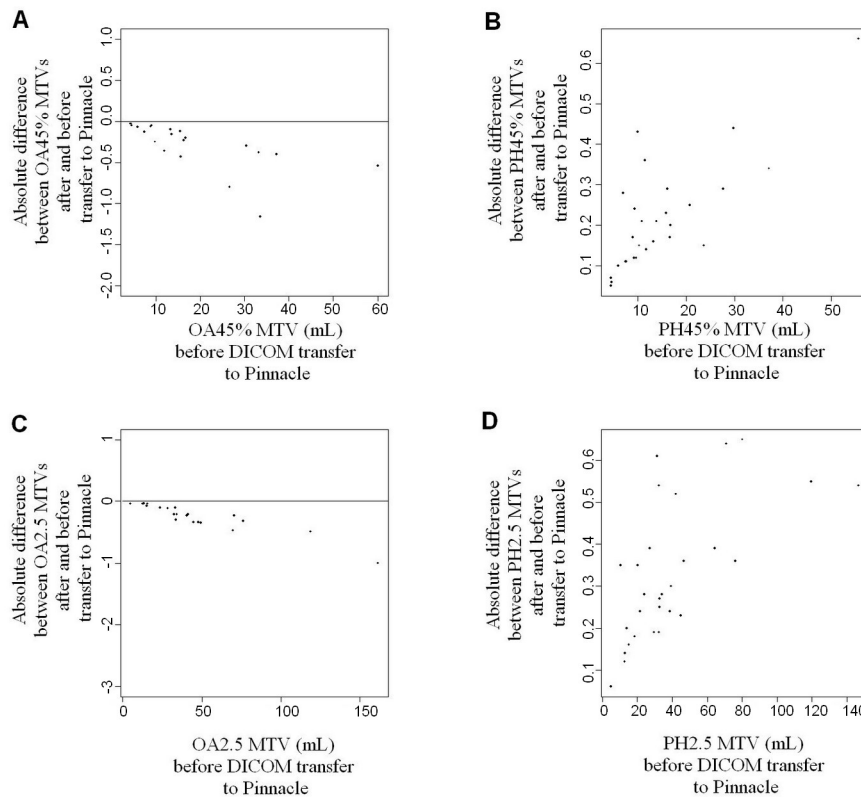


FIG. 4. These Bland-Altman bias plots show the difference between the metabolic tumor volumes (MTV) before the transfer to Pinnacle (x-axis) against the absolute difference between the MTV (mL) after the transfer in Pinnacle and the volume before transfer (y-axis). The transfer from OASIS ((a): OA45%; (c): OA2.5) to Pinnacle led to a reduction of the MTV, systematically. While the transfer from Philips EBW4 ((b): PH45%; (d): PH2.5) to Pinnacle systematically led to an expansion of the MTV.

IV. DISCUSSION

The delineation of the MTV in oncology using PET is increasingly investigated in the literature. At diagnosis, MTV (along with the SUV derived from the MTV) may have a prognostic value; at the time of treatment planning, the MTV may be an additional tool for the delineation of the GTV and, finally during treatment, the MTV may be one of the parameters used for the estimation of the therapeutic response.^(15,16,22,23,24)

Various segmentation algorithms are available, but none has been fully validated and widely implemented in the clinics. This is in part due to the absence of a gold standard, which ideally should be the measurements of corresponding pathological samples.⁽²⁵⁾ Even these are imperfect, as tumor shrinkage may occur after formalin-fixation, for example.^(26,27)

Obviously the segmentation algorithm has a major impact on the metabolic delineated volumes. Considering the absence of a recognized gold standard for assessing the tumor volume *in vivo*, our aim was to compare the MTVs to the clinical reference (i.e., the volume used for treating the patients, defined by a single radiation oncologist). Clearly, the GTV_{CT} cannot be considered representative of the underlying truth, but it is the one used to treat the patients and the only reference at hand. The intraclass correlation coefficient between the MTVs and the GTV_{CT} ranged from 0.26 to 0.85. Furthermore, the conformity index between the volume that was used for treating the patients (GTV_{CT}) and the various MTVs ranged from 0.24 to 0.38,

indicating major spatial variations in addition to changes in volumes. Quite surprisingly, the MTVs derived from the same threshold-based method (e.g., $SUV \geq 45\%$ of the SUV_{max}) actually varied, depending on the commercial software that was used (i.e., Philips EBW4 or Segami OASIS). Such discrepancies are, in fact, easily explained by differences in implementation. For example, the default configuration of the Philips EBW4 does not draw the volume considering a percentage above the maximum pixel value but, in fact, above the average five pixels in the neighborhood of the maximum pixel value. We also observed that the MTV boundaries follow the tumor bordering voxels in Philips EBW4, while the contours in OASIS do not follow the edges of the voxels.

Quality assurance of each institutional PET system, patient preparation, PET images acquisition, and reconstruction parameters are known major factors affecting image quantification and subsequent segmentation.^(3,4) Our work identified additional factors affecting the downstream MTV based on PET images. We found that the modification of the matrix of the CT and the subsequent resampling of the PET images resulted in a variation of the MTV from 0% to 142%; the increase of the extracted SUV_{max} observed on resampled PET images on the Philips EBW4 is certainly related to the selection of five points (voxels or pixels) to average for the maximum SUV.

One would expect the volumes to remain rigorously identical after export as DICOM/RT structures, yet we observed that some changes may occur in the process. Further, the algorithms are affected to variable extents, as the change in volume was statistically significant for OASIS and the GBM when transferred to PMOD, but not for the other methods. In OASIS, the volume boundaries crossed the frontline voxels, but the transfer to PMOD led to a reinterpretation of the boundaries following a specific algorithm implemented in PMOD and leading to subsequent volume changes. If the algorithm implemented follows the bordering voxels for one software and does not for the other one, then the transferred MTV will change. The larger the voxel size and/or the larger the MTV, the larger the volume modifications will be. Following the same principle, the DICOM transfer of a MTV from a commercially available PET workstation (Philips EBW4 or OASIS) to a commercially available treatment planning workstation (Pinnacle) led to significant changes in the MTVs (Fig. 5).

Both stage at diagnosis and tumor regression grade of the total mesorectal excision specimen independently determine patient survival.⁽²⁸⁾ New therapy systems are, therefore, developed to deliver higher doses in a highly conformal manner to irregular target volumes. Considering the current treatment scheme of rectal cancer, the metabolic GTV, does not impact in a dramatic way the clinical target volume (CTV) that includes the GTV with 1 cm isotropic margin, the mesorectal subsite, and the pelvic lymph nodes at high risk of microscopic involvement.⁽²⁹⁾ An additional 8 mm margin is then applied to the CTV to delineate the planning target volume (PTV) in order to minimize organ motion and setup errors. Some of the volume changes observed in this study, particularly those associated with the DICOM transfer, may be accounted for, especially in other tumor types for which the target volume delineation is more complex, such as lung cancer or head and neck carcinomas. These considerations also become of primary importance when highly conformal dose delivery systems are used, in particular intensity-modulated radiation therapy (IMRT) and dose painting. More generally, it should be fully integrated as part of the quality assurance processing that is mandatory for any activity of radiation oncology.⁽³⁰⁾

Our work presents several limitations. First, it lacks a true gold standard for defining the actual tumor volume, even though, as previously discussed, such gold standard does not exist. Phantom studies have been performed and showed differences in volumes depending on the algorithms, but the results cannot directly be translated into the clinical setting, as the complexity and the heterogeneity of clinical PET images are usually much greater than the available phantoms.^(8,9,31) In addition, organs and tissue are not completely immobile in the human body, and movements may occur during or between studies and lead to geographical misses (Fig. 6). In our work, the CI between the MTVs and the GTV_{RT} was ≤ 0.38 and was 0 for four patients. Secondly, the point spread function (PSF) of our PET/CT system was only available for the $4 \times$

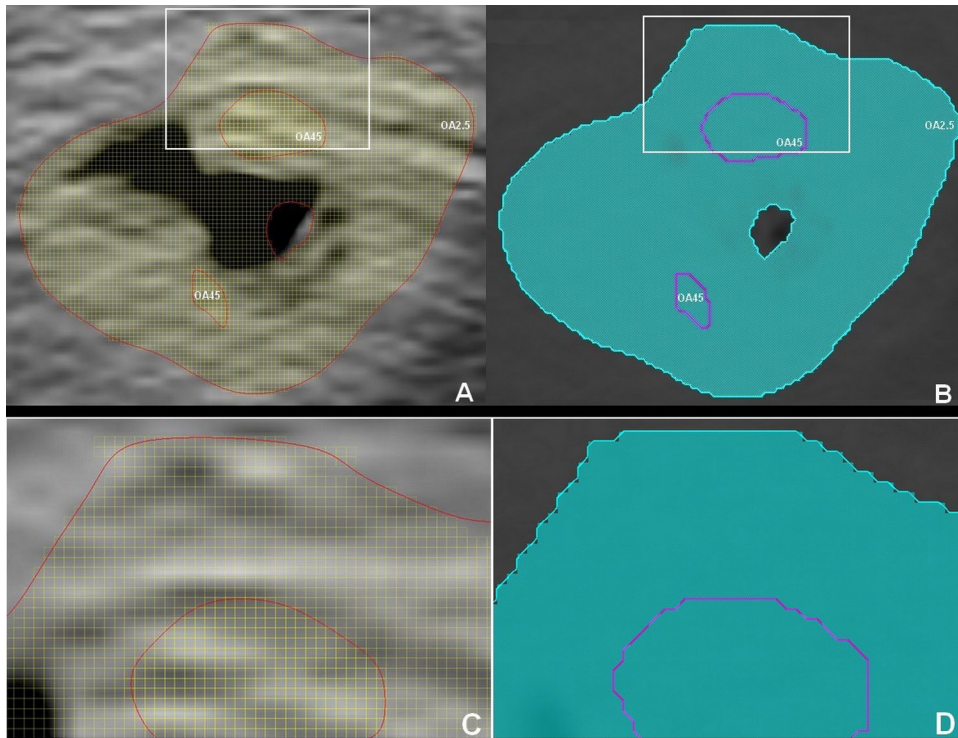


FIG. 5. The transverse slice (a) shows the metabolic tumor volumes (MTVs) contours delineated in Segami OASIS software, using the relative threshold (OA45%) and an absolute threshold (OA2.5). The transverse slice (b) shows the MTVs transferred from Segami OASIS in Pinnacle. Magnified images (c) and (d) extracted from slices (a) and (b), respectively, show the slight difference between the contours delineated and displayed in OASIS (c) and contours displayed in Pinnacle (d) after the DICOM transfer.

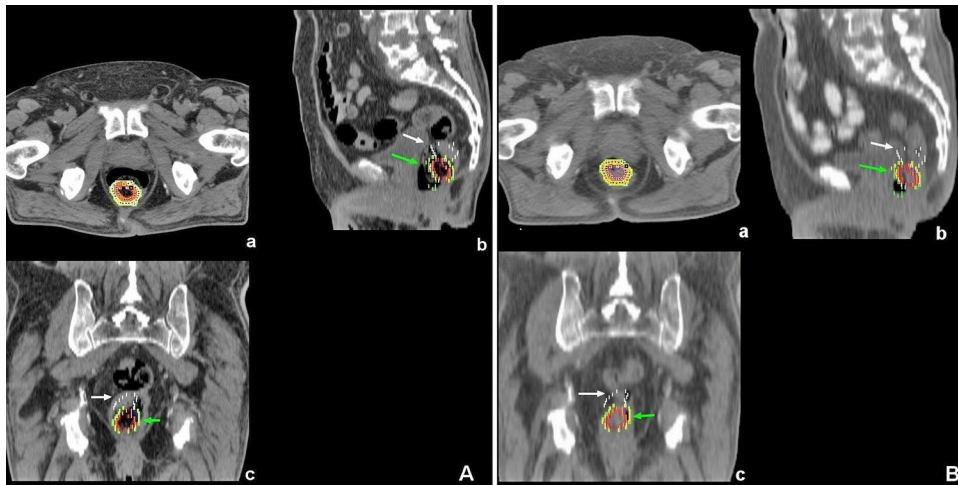


FIG. 6. The figure shows an example of bowel and tumor displacement between the CT_{RT} (left panels) and the CT of the FDG PET/CT (right panels); (a): transverse slice; (b): sagittal slice, and (c): coronal slice). Metabolic tumor volumes (FLAB = purple, OA45% = red; PH45% = orange; OA2.5 = yellow; PH2.5 = green) are indicated by the green arrow and the GTV_{RT} (white contour) by the white arrow.

$4 \times 4 \text{ mm}^3$ voxel size, so that we were not able to perform the segmentation using the GBM on resampled PET images. The GBM has the advantage of being optimized for the individual PET/CT devices, but this requires measuring the PSF of the system. On the other hand, the FLAB algorithm requires less user intervention and is easier to implement, and appears to be highly robust without the need for calibration, at least as long as the parameters of the PET image are within a normal range usually encountered in clinical practice. The MTVs delineated with FLAB were not significantly affected either by the PET images resampling or by the transfer between the software products used in this study. Although this is clearly beyond the scope of this article, segmentation algorithms, such as FLAB or the GBM, should be preferred to methods based upon relative or fixed SUV thresholds, as they are more robust across a wide variety of intensity and heterogeneity of uptake.

V. CONCLUSIONS

Distinct segmentation algorithms led to significantly different MTVs. Moreover, it was found that the MTVs are affected by the software used for implementing the segmentation algorithm, and that the manipulation of PET/CT images and MTVs, such as the DICOM transfer, may also induce additional volume variations. Furthermore, not all commercially available software products and segmentation algorithms are equally affected by these issues. It is, therefore, recommended to test these aspects before integrating the PET-based volumes in the routine clinical radiation therapy planning, and to standardize all image processing and transfer procedures.

ACKNOWLEDGMENTS

We thank Laurence Seidel and Professor A. Albert for providing statistical data. Part of this work was presented as a poster at the Annual Meeting of the Society of Nuclear Medicine, San Antonio, Texas, June 4–8, 2011.

REFERENCES

1. 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(11):1690–97.
2. Brambilla M, Matheoud R, Secco C, Loi G, Krenqli M, Inglese E. Threshold segmentation for PET target volume delineation in radiation treatment planning: the role of target-to-background ratio and target size. *Med Phys*. 2008;35(4):1207–13.
3. Cheebsumon P, Yaqub M, van Velden FH, Hoekstra OS, Lammertsma AA, Boellaard R. Impact of [(18)F]FDG PET imaging parameters on automatic tumour delineation: need for improved tumour delineation methodology. *Eur J Nucl Med Mol Imaging*. 2011;38(12):2136–44.
4. Matheoud R, Della Monica P, Loi G, et al. Influence of reconstruction settings on the performance of adaptive thresholding algorithms for FDG-PET image segmentation in radiotherapy planning. *J Appl Clin Med Phys*. 2011;12(2):3363.
5. Biehl KJ, Kong FM, Dehdashti F, et al. 18F-FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J Nucl Med*. 2006;47(11):1808–12.
6. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med*. 2005;46(8):1342–48.
7. Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Gregoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol*. 2003;69(3):247–50.
8. 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 Imaging*. 2007;34(9):1427–38.
9. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D. A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET. *IEEE Trans Med Imaging*. 2009;28(6):881–93.

10. 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(4):663–72.
11. Werner-Wasik M, Nelson AD, Choi W, et al. What is the best way to contour lung tumors on PET scans? Multiobserver validation of a gradient-based method using a NSCLC digital PET phantom. *Int J Radiat Oncol Biol Phys*. 2012;82(3):1164–71.
12. Wanet M, Lee JA, Weynand B, et al. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. *Radiother Oncol*. 2011;98(1):117–25.
13. Fletcher JW, Djulbegovic B, Soares HP, et al. Recommendations on the use of 18F-FDG PET in oncology. *J Nucl Med*. 2008;49(3):480–508.
14. Anderson C, Koshy M, Staley C, et al. PET-CT fusion in radiation management of patients with anorectal tumors. *Int J Radiat Oncol Biol Phys*. 2007;69(1):155–62.
15. Bassi MC, Turri L, Sacchetti G, et al. FDG-PET/CT imaging for staging and target volume delineation in preoperative conformal radiotherapy of rectal cancer. *Int J Radiat Oncol Biol Phys*. 2008;70(5):1423–26.
16. Buijssen J, van den Bogaard J, van der Weide H, et al. FDG-PET-CT reduces the interobserver variability in rectal tumor delineation. *Radiother Oncol*. 2012;102(3):371–76.
17. Edge SB BD, Compton CC, Fritz AG, Greene FL, Trotti A. *AJCC Cancer Staging Manual*, 7th edition. New York, NY: Springer; 2010.
18. Day E, Betler J, Parda D, et al. A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients. *Med Phys*. 2009;36(10):4349–58.
19. Jingu K, Ariga H, Kaneta T, et al. Focal dose escalation using FDG-PET-guided intensity-modulated radiation therapy boost for postoperative local recurrent rectal cancer: a planning study with comparison of DVH and NTCP. *BMC Cancer*. 2010;10:127.
20. Niyazi M, Geisler J, Siefert A, et al. FET-PET for malignant glioma treatment planning. *Radiother Oncol*. 2011;99(1):44–48.
21. Van den Broeck A, Vankelecom H, Van Eijsden R, Govaere O, Topal B. Molecular markers associated with outcome and metastasis in human pancreatic cancer. *J Exp Clin Cancer Res*. 2012;31:68.
22. Muralidharan V, Kwok M, Lee ST, Lau L, Scott AM, Christophi C. Prognostic ability of 18F-FDG PET/CT in the assessment of colorectal liver metastases. *J Nucl Med*. 2012;53(9):1345–51.
23. Krengli M, Cannillo B, Turri L, et al. Target volume delineation for preoperative radiotherapy of rectal cancer: inter-observer variability and potential impact of FDG-PET/CT imaging. *Technol Cancer Res Treat*. 2010;9(4):393–98.
24. 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 18F-FDG PET. *Acta Oncol*. 2013;52(3):619–26.
25. Buijssen J, van den Bogaard J, Janssen MH, et al. FDG-PET provides the best correlation with the tumor specimen compared to MRI and CT in rectal cancer. *Radiother Oncol*. 2011;98(2):270–76.
26. Siedschlag C, van Loon J, van Baardwijk A, et al. Analysis of the relative deformation of lung lobes before and after surgery in patients with NSCLC. *Phys Med Biol*. 2009;54(18):5483–92.
27. Summers RM. Polyp size measurement at CT colonography: what do we know and what do we need to know? *Radiology*. 2010;255(3):707–20.
28. Ryan R, Gibbons D, Hyland JM, et al. Pathological response following long-course neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *Histopathology*. 2005;47(2):141–46.
29. Roels S, Duthoy W, Haustermans K, et al. Definition and delineation of the clinical target volume for rectal cancer. *Int J Radiat Oncol Biol Phys*. 2006;65(4):1129–42.
30. Haustermans K, Debucquoy A, Lambrecht M. The ESTRO Breur Lecture 2010: toward a tailored patient approach in rectal cancer. *Radiother Oncol*. 2011;100(1):15–21.
31. Prieto E, Lecumberri P, Pagola M, et al. Twelve automated thresholding methods for segmentation of PET images: a phantom study. *Phys Med Biol*. 2012;57(12):3963–80.