Observational Study

Medicine®

OPEN

Comparison of different automated lesion delineation methods for metabolic tumor volume of ^{18}F-FDG PET/CT in patients with stage I lung adenocarcinoma

Xiao-Yi Wang, MD[a,b], Yan-Feng Zhao, MD[b], Ying Liu, MD[a], Yi-kun Yang, MD[c], Zheng Zhu, MD[b], Ning Wu, MD[a,b,*]

Abstract

The aim of the study was to investigate the suitable segmentation method in small, low uptake and heterogeneous nodules of stage I lung adenocarcinoma.

133 stage I lung adenocarcinoma patients with ^{18}F-FDG PET/CT scans were enrolled in this retrospective study. All lesions were divided into different groups according to nodule density, nodule size, and the maximum standard uptake value (SUV_{max}) level. Four different PET segmentation methods were performed, including percentage threshold of SUV_{max} (T42% and T42% × RC), gradient-based threshold (adaptive iterative algorithm, AT-AIA), and background-related threshold (adaptive thresholding at 40% SUV_{max}, AT40%) approaches. The MTVs were evaluated and compared with CT volume (CTV). Percentage volume error (%VE) compared to CTV was calculated and the correlations between MTVs and CTV were analyzed.

AT-AIA had the highest accuracy in large, high uptake, and solid nodules (72.5%, 72.4%, and 65.6%, respectively). AT40% had the highest accuracy in small, low uptake and nonsolid nodules (56.6%, 56.1%, and 62.6%, respectively). In part-solid nodules, the accuracy of AT-AIA (60.0%) and AT40% (56.7%) were higher than that of T42% and T42% × RC. The MTV of AT-AIA was in excellent correlation with the CTV in solid nodules ($R=0.831$, $P<.001$) and in high uptake nodules ($R=0.830$, $P<.001$). The MTV of AT40% was in good correlation with the CTV in nonsolid nodules ($R=0.686$, $P=.003$) and in part-solid nodules ($R=0.731$, $P<.001$).

AT40% showed best performance in small, low uptake, nonsolid and part-solid lesions. AT-AIA was suitable for large, high uptake, and solid lesions.

Abbreviations: %VE = percentage volume error, ^{18}F-FDG = 2-deoxy-2-[18F]fluoro-D-glucose, AIA = adaptive iterative algorithm, CTV = computed tomography volume, Lung VCAR = lung volume computerized assisted reporting, MTV = metabolic tumor volume, NSCLC = nonsmall cell lung cancer, PET/CT = positron emission tomography/computed tomography, PETVCAR = PET volume computerized assisted reporting, PVE = partial volume effect, ROI = region of interest, SUV_{max} = maximum standard uptake value, TLG = total lesion glycolysis, VOI = volume of interest.

Keywords: adaptive thresholding, delineation method, lung adenocarcinoma, metabolic tumor volume, PET

1. Introduction

Positron emission tomography/ computed tomography with 2-deoxy-2-[^{18}F]fluoro-D-glucose (^{18}F-FDG PET/CT) shows its usefulness in tumor staging and follow-up. Recently, several researches have already proved that maximum standard uptake value (SUV_{max}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG) had prognostic and predictive values in nonsmall cell lung cancer (NSCLC) patients.[1–5] Higher SUV_{max} and adenocarcinoma histology were associated with shorter disease-free survival (DFS).[6] High SUV_{max} and high MTV of the

Editor: Peng Luo.

Authorship: X-YW and NW conceived of the project. X-YW analyzed the data and wrote the paper. Y-FZ, YL, Y-kY, ZZ, and NW provided expert guidance, data, or analysis tools. All authors reviewed the manuscript.

Ethical approval: For the type of retrospective study formal consent is not required. The informed consent was waived because of the retrospective nature of this study.

Funding: This study was funded by the National High Technology Research and Development Program of China (863 Program, Grant No. 2014AA020602) and the Major Project of Beijing Municipal Science and Technology Commission of the People's Republic of China (D141100000214006).

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

[a] PET/CT Center, [b] Department of Diagnostic Radiology, [c] Department of Thoracic Surgery, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

[*] Correspondence: Ning Wu, National Cancer Center/ Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China (e-mail: cjr.wuning@vip.163.com).

Medicine (2017) 96:51(e9365)

Received: 31 August 2017 / Received in final form: 13 November 2017 / Accepted: 27 November 2017

http://dx.doi.org/10.1097/MD.0000000000009365

primary tumor are independent prognostic factors of shorter DFS in early stage of NSCLC without lymph node metastasis.[7,8] In addition, SUV_{max}, MTV, and TLG have prognostic role on NSCLC patients treated with stereotactic body radiation therapy; only MTV and TLG have a predictive value for DFS when tumors are larger than 3 cm.[9,10]

The lesions in those previous studies were generally with solid nodule type as well as high FDG uptake.[5–8] How about the predictive ability in ground-glass opacity nodule (GGN) with low uptake lesions? Goudarzi et al[11] reported that pure bronchioloalveolar carcinoma (BAC) exhibits smaller size, lower uptake, and lower tumor density than invasive adenocacinoma, and many BACs have low SUVs (<2.0). Khalaf et al[12] reported that although the SUV_{max} cutoff value of 2.5 is a useful tool in the evaluation of large pulmonary nodules (>1.0 cm), it has no or minimal value in the evaluation of small pulmonary nodules (≤1.0 cm). Although generally, small GGN with low uptake are going to fall well into the favorable prognostic category, these patients need long-term follow-up exams.

Various automated methods are currently used to segment regions of interest in PET/CT scans, including fixed SUV threshold (e.g., SUV2.5), percentage threshold of SUV_{max} (e.g., T42%), gradient-based threshold (adaptive iterative algorithm, AT-AIA), and background-related threshold (AT40%) approaches. However, up to now, it is still challenging to define MTV accurately for heterogeneous and low uptake lung nodules and prone to inter- and intraobserver variability. It is known that a single threshold SUV method is not universally applicable to all clinical scenarios,[13,14] especially in low FDG uptake lesions. The fixed threshold method was not used in our study since it ignores the background. Currently, the percentage threshold (T42%) method is widely used in lung cancer, which is based on homogenous phantom study with high contrast, so as to more applicable to the solid nodules larger than 20 mm and with high FDG uptake.[7,15] Percentage threshold can be performed rapidly and consistently, with less inter-observer variability. The adaptive iterative delineation method (AT-AIA) is more advanced and complex, which uses an iterative algorithm to find a threshold value that separates the tumor from the background tissue by weighting SUV_{max} and SUV_{mean} within the bounding box. AT-AIA was usually used on solid nodules larger than 20 mm with high FDG uptake.[16,17] But there is not enough clinical evidence that it is suitable for small and low uptake lesions. The AT-AIA method tends to find the largest gradient at the border of the lesions, but in low uptake lesions, the gradient is relatively low. So, this method may not be a good choice in such a clinical scenario. The AT40% method, considering the metabolic contrast between lesion and background uptake and the location of the lesion, may improve the segment accuracy in small and low uptake lesions, although with the manual background region of interest procedure, which may introduce more inter-observer variability than the other methods.[18] Firouzian et al[18] found that these automated lesion delineation methods have high variation in small lesions. The aim of this paper is to investigate the suitable segmentation method in small, low uptake and heterogeneous nodules of stage I lung adenocarcinoma.

2. Materials and methods

2.1. Subjects

A total of 133 patients with stage I adenocarcinoma who performed ^{18}F-FDG PET/CT scans prior to surgery in our hospital from June 2005 to June 2012 were enrolled in this retrospective study. The informed consent was waived because of the retrospective nature of this study. This was agreed by the local ethics committee and approval from the ethics committee was granted. There were 65 males and 68 females with age ranged 35 to 84 years (mean 60 years). The locations of the lung nodules were as follow: 33 lesions in the left upper lobe, 24 lesions in the left lower lobe, 47 lesions in the right upper lobe, 7 lesions in the right middle lobe, and 22 lesions in the right lower lobe.

All lesions were viewed and judged by 3 experienced radiologists, with 14 years, 17 years, and 30 years working experience respectively. The measurements were performed by 2 senior radiologists, and the time separation between each measurement was 4 weeks. The radiologists were blinded to each other's definition of the lesions. The disagreement was decided by discussion. According to nodule density, the lesions were divided into 3 types: nonsolid, part-solid, and solid. Nonsolid nodule which means pure GGN, defined as an area of hazy increased attenuation that does not obscure underlying bronchial structures or vascular margins on high-resolution computed tomography (HRCT).[19] Part-solid nodule which means mixed GGN, defined as mixed nonsolid and solid components.[19–21] Since the long diameter of 2 cm was the cut-off value of T1a with T1b lung cancer, according to nodule size, the lesions were divided into 2 groups: small lesions (long diameter≤20 mm) and large lesions (long diameter>20 mm). Although the SUV_{max} threshold of 2.5 is generally chosen to maximize sensitivity of malignancy detection, the FDG uptake in early stage lung adenocarcinoma was lower than other lung cancer.[22,23] Therefore, in this study, we considered $SUV < 2.0$ as low uptake lesion, $SUV > 2.0$ as high uptake lesion. According to FDG uptakes, the lesions were divided into 2 groups: low uptake lesions ($SUV_{max} \leq 2$) and high uptake lesions ($SUV_{max} > 2$).

2.2. ^{18}F- FDG PET/CT study

^{18}F-FDG PET/CT was performed using an integrated PET/CT (Discovery ST, GE Healthcare). All patients in this study were scanned on the same PET/CT machine. Patients' blood glucose was between 120 and 200 mg/dL before undergoing PET/CT examination. Patients received 3.70 to 4.44 MBq/kg of ^{18}F-FDG intravenously, followed by a whole body PET/CT scan 60 to 70 minutes later. The PET images were obtained with 3 min acquisition per bed position, with slice thickness of 3.27 mm. Scan from skull vertex to upper-thigh resulted in an acquisition time of 18 to 21 minutes. All PET images were reconstructed using an iterative algorithm (ordered-subset expectation maximization, OSEM) with CT-based attenuation correction. Spiral CT was performed with a tube voltage of 120 kV, tube current of 150 mA, 3.75 mm slice thickness and 3.75 mm interval, at 0.8 s per rotation. The attenuation correction scan was performed from vertex to upper thighs and no contrast was used for this examination. Breathing-hold chest CT without contrast was performed then, with a tube voltage of 120 kV, tube current of 205 mA, slice thickness of 5 mm and 1.25 mm, with 5 mm and 0.8 mm interval respectively, at 0.8 seconds per rotation.

2.3. Automated PET delineation methods

Four different automated PET delineation methods were evaluated and compared (Table 1). All segmentation algorithms were implemented on the same software platform in AW 4.6 workstation (Advantage Workstation, GE Healthcare) to optimize workflow and minimize reproducibility drawbacks.

Table 1

Currently used automated segmental methods.

Automated segmental methods	Example
1. Percentage threshold of SUV_{max}	T42%
2. Percentage threshold of SUV_{max} with recovery coefficients	T42% × RC
3. Gradient-based threshold	Adaptive iterative algorithm (AT-AIA)
4. Background-related threshold	Adaptive thresholding at 40% SUV_{max} (AT40%)
5. Fixed SUV threshold*	SUV2.5

AT-AIA = adaptive iterative algorithm, RC = recovery coefficients, SUV_{max} = maximum standard uptake value.
* Fixed SUV threshold method was not used in this study.

The first method was thresholding at 42% SUV_{max} (T42%): delineates all voxels with SUVs above or equal to 42% of the maximum SUV inside the selected volume of interest (VOI).

The second method was thresholding at 42% SUV_{max} with recovery coefficients (T42% × RC). This was only done on the lesions whose diameter was less than 30 mm (lesion>30 mm, RC=1).

As reported earlier,[24,25] the partial volume effect (PVE) is a physical limitation resulting from the poor spatial resolution of PET systems (4–5 mm) which strongly affects the accuracy of the estimation of radioactivity concentration within structures less than 2 or 3 times of the PET spatial resolution. Among all PVE correction methods, more common ones are based on multiplicative numerical factors, recovering the local radioactivity concentration within any small structure which uptakes ^{18}F-FDG.[26,27] Recovery Coefficients (RC) was obtained as a function of PET resulting from the threshold-isocontour technique.

$$\%\,RC = 100 \times [(C_{hot}/C_{bg-1})/(a_{hot}/a_{bg-1})] \quad (1)$$

$$a_{hot}/a_{bg} = [(C_{hot}/C_{bg-1})/RC] + 1 \quad (2)$$

where C_{hot} and C_{bg} are the average counts measured in the hot sphere ROIs and the average counts in all background ROIs, respectively, whereas a_{hot}/a_{bg} is the ratio of the true radioactivity concentration in the hot sphere and in the background. Gallivanone et al[26] have reported the method for PVE correction of oncological lesions in clinical studies, based on RC and on PET measurements of lesion to background ratio and of lesion metabolic volume. The validation of the PVE correction method resulted to be accurate (>89%) in clinical realistic conditions for lesion diameter > 1 cm, recovering >76% of radioactivity for lesion diameter < 1 cm. Results from patient studies showed that the proposed PVE correction method is suitable and feasible and has an impact on a clinical environment.[26,28] In this study, the RC was derived from PET experimental measurements of small radioactive objects in a priori known object-to-background radioactivity concentration ratio. It came from the work of our previous colleagues[28] (Table 2).

The third method was adaptive iterative volume delineation by PET Volume Computerized Assisted Reporting software (PETVCAR, GE Healthcare). The PET and CT co-registration was first assessed once the images were loaded into the PETVCAR software. The primary lung cancer PET gray scale and PET/CT fused images were then reviewed in the axial, sagittal, and coronal planes. A boundary box was placed over the image, which was to auto-contour and segment the region of interest, reviewed and adjusted to ensure this 3-dimensional cube contained all the ^{18}F-FDG PET positive area and excluded the negative normal tissue. This process was repeated until each ^{18}F-FDG PET/CT positive region has been selected and optimized. The lesion metabolic volume was then automatically segmented using an adaptive iterative algorithm (AT-AIA) in PETVCAR which separated the target volume from the background tissue by weighting the SUV_{max} and the SUV_{mean} within the target volume with a weighting factor, represented as a Boolean variable. This weighting factor was automatically set at 0.5[16] (Fig. 1).

Table 2

Recovery coefficients (RC) of discovery ST PET/CT.

Lesion diameter	10 mm	13 mm	17 mm	22 mm
RC	47%	56%	67%	71%
42% × RC	20%	24%	28%	30%

PET/CT = positron emission tomography/computed tomography, RC = recovery coefficients.

And the fourth method was adaptive thresholding at 40% SUV_{max} (AT40%),[18] which adapts the threshold value inside the selected VOI relative to mean background (BG) SUV, calculating T value as thresholding:

$$T = 0.4(SUV_{max} - BG) + BG \quad (3)$$

This delineation method required information of background uptake. The background region needs to be defined by the user which might introduce some variations in the results. The background of lung is heterogeneous; mean background SUV has discrepancy at different regions (apex, central, and peripheral region). The user needs to copy the ROI and select the same location at contralateral lung.

2.4. Computed tomography volume

When lesion density is different from the density of the surrounding tissues, a computed tomography study in the region

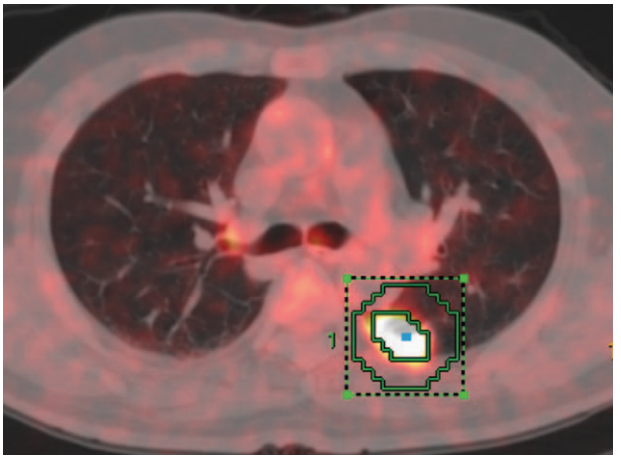

Figure 1. The metabolic tumor volume (MTV) was segmented using an adaptive iterative algorithm (AT-AIA) in PETVCAR. AT-AIA = adaptive iterative algorithm, MTV = metabolic tumor volume, PETVCAR = PET volume computerized assisted reporting.

of interest can provide lesion anatomical volume (computed tomography volume, CTV). CTV was measured through Lung Volume Computerized Assisted Reporting software (Lung VCAR, GE Healthcare) on the 1.25 mm slice thickness images. Lung VCAR is an image analysis software package for Advantage Workstation systems that uses GE's Volume Viewer software. The analysis mode was used which offers a combination of 2D reformatted views with correlated volume rendering views. In this mode, the software zooms on the volume of interest, automatically calculates the volume of the suspicious spot, and displays the calculated volume on the views. Also, depending on the protocol chosen (nodule consistency and circumscribed situation), it displays the consistency of the detected nodules. The actual volume was measured using an automatic nodule sizing algorithm. Upon entering the analysis mode, the software automatically performs the following operations: step 1- definition of a VOI around the nodule; step 2- determination of nodule consistency (solid, part-solid or nonsolid); step 3- determination of nodule circumscribed situation (well circumscribed, vascularized or juxta-pleural). Then the software automatically computes the segmentation: type 1- if the nodule is well circumscribed, the system calculates its volume and displays it on the views; type 2- if the nodule is vascularized, the system proceeds to an automatic vascular tree extraction, followed by a vessel cut, before calculating and displaying its volume on the views; type 3- if the nodule is juxta-pleural, the system separates it from the pleural wall, before calculating and displaying its volume on the views (Fig. 2).

Reproducibility evaluation was achieved for the implementation by repeating the delineation procedure several times in each patient. The inter-observer variability on the delineation process was less prone to happen because the boundary box was auto-contour and segments the region of interest. The user only need to review and adjust to ensure the 3D-box contained all the FDG positive area and excluded the negative normal tissue. Then the lesion metabolic volume was automatically segmented from the different algorithms. The Intraclass Correlation Coefficient (ICC) was used to estimate the reliability between observers when using the fourth delineation method (AT40%).

The follow up of these patients for progression-free survival (PFS) was performed to further validate whether the delineation method classification reasonable.

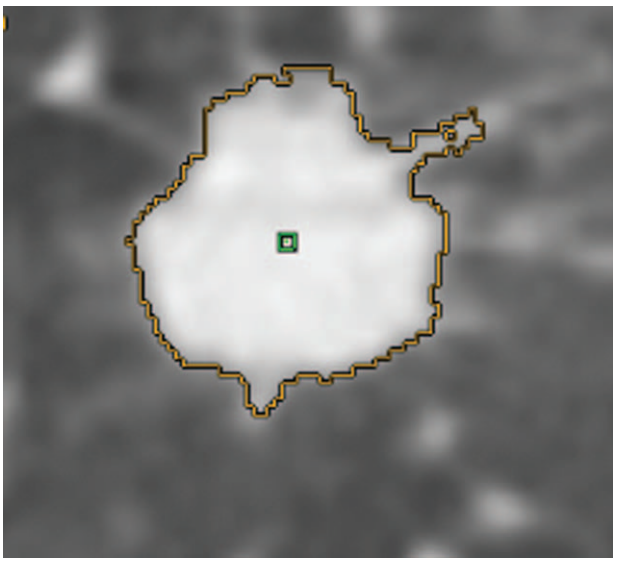

Figure 2. The lesion computed tomography volume (CTV) was measured through Lung VCAR software. CTV = computed tomography volume, VCAR = Lung volume computerized assisted reporting.

2.5. Validation and statistics

Percentage volume error (%VE) was calculated using CTV as reference:

$$\%VE = (Vol_{MTV} - Vol_{CTV})/Vol_{CTV} \times 100 \quad (4)$$

The Vol_{MTV} was the volume of delineated lesions in PET images and Vol_{CTV} was the volume of the delineated lesions in CT images. The discrepancies between the imaging modalities of CT and PET in tumor volume delineation had been reported in previous studies. A difference less than 30% between CTV and MTV was considered clinical acceptable.[29–34] In this study, a difference between ± 50% was considered acceptable, because the lesions in our study were smaller, lower FDG uptake and more heterogeneous than other researches. The accuracy of each method was defined by the percentage of cases which fell within this range. The %VE more than 50% meant overestimated, less than –50% meant underestimated.

The results were evaluated by standard methods including combined t test and Chi-square test. The descriptive data are expressed as the means ± standard deviations. T test was used to analyze the continuous variables, and the chi-square test to compare the categorical variables between groups. The correlation of various MTVs from different segmentation algorithms with CTV was analyzed and evaluated by Pearson correlation. The ordinal data correlation test was performed using the Spearman test. The correlation coefficient (R value) = 0.21–0.40 for the poor consistency, R value = 0.41–0.60 for the moderate consistency, R value = 0.61–0.80 for the good consistency, R value = 0.81–1.00 for the excellent consistency.[35]

The intraclass correlation coefficient (ICC) was used to estimate the reliability between observers when using the fourth delineation method (AT40%). ICC < 0.40 is for the poor reliability, ICC > 0.75 for the good reliability. Progression-free survival (PFS) was compared by employing the Kaplan–Meier method and Cox proportional-hazard model. $P < .05$ were assumed to indicate significant differences. The data were analyzed by SPSS 13.0 software (Chicago).

3. Results

According to the classification of lung nodule types analyzed by 3 experienced radiologists, there were 16 nonsolid nodules, 30 part-solid nodules, and 87 solid nodules in all 133 lesions. The SUV_{max}, diameter, and CTV of 3 lung nodule types were shown in Table 3. There was statistical significance of SUV_{max} between solid and part-solid nodule, solid and nonsolid nodule, part-solid and nonsolid nodule ($t = 4.706$, $P < .001$; $t = 4.539$, $P < .001$; $t = 3.269$, $P = .002$, respectively). But the diameter and CTV had no statistical significance among the 3 types (all $P > .05$). The nodule types had good consistency with $SUV_{max,}$ ($R = 0.680$, $P < .001$) but not with diameter and CTV (Spearman test).

The MTV, %VE, and SUV_{mean} of 4 automated PET delineation methods were shown in Table 4. The comparisons of VE% (t, P value) of different groups were shown in Supplemental Digital Content Tables 1, 2, 3, and 4, http://links.lww.com/MD/C34.

Table 3

SUV_{max}, diameter (mm), and CTV (mm^3) of 3 lung nodule types.

Nodule types	Mean	Std. Deviation	Minimum	Maximum
Nonsolid (n = 16)				
SUV_{max}	0.81	0.33	0.38	1.76
Diameter	20.37	7.79	10.00	31.00
CTV	3804	3283	500	12701
Part-solid (n = 30)				
SUV_{max}	1.65	0.99	0.33	4.81
Diameter	21.93	7.50	10.30	37.00
CTV	6457	4753	430	19131
Solid (n = 87)				
SUV_{max}	4.43	3.17	0.57	21.69
Diameter	23.41	7.50	10.00	44.40
CTV	8053	7296	366	33493

CTV = computed tomography volume, SUV_{max} = maximum standard uptake value.

There was statistical significance between most of each 2 methods of %VE. Figure 3A and B showed the variation of %VE of each delineation method in the 3 types of nodules. The variation of VE % in nonsolid nodule is much larger than that in part-solid nodule and solid nodule. The mean variation of %VE of AT40% is the smallest in the 4 methods. Figure 3C showed that T42% is good at solid nodule, but unstable in part-solid and nonsolid nodule. AT40% and AT-AIA were more stable than the other 2 methods.

According to the criteria that a difference less than ±50% was considered accurate in this study, the segmental accuracy of 4 methods in different nodule type, size and FDG uptake groups were shown in Figure 4. Figure 4 demonstrated that the AT40% method was superior for small, low uptake, nonsolid lesions; the AT-AIA method was superior for large, high uptake, solid lesions. The underestimated and overestimated percentages of 4 methods with variation are shown in Table 5, which showed that AT40% underestimated the lesions in all 3 nodule types compared with the other 3 methods, whereas T42% and T42% × RC usually overestimated the nonsolid and part-solid lesions. However, the comparison of accuracy percentage of 4 methods in different groups had no statistical significance (all $P > .05$).

The MTV of AT-AIA had excellent consistency with CTV in solid nodules ($R = 0.831$, $P < .001$) and also in high uptake nodules ($R = 0.830$, $P < .001$). The MTV of AT40% was in good correlation with the CTV in nonsolid nodules ($R = 0.686$, $P = .003$) and in part-solid nodules ($R = 0.731$, $P < .001$). The

Table 4

MTV (mm^3), %VE, and SUV_{mean} of each automated PET delineation methods.

	T42%	T42% × RC	AT-AIA	AT40%	CTV
Total lesions (n = 133)					
MTV	6042 ± 4621	8838 ± 5209	7385 ± 5903	4841 ± 4329	7181 ± 6558
%VE	40% ± 140%	117% ± 191%	42% ± 107%	-6% ± 85%	
SUV_{mean}	2.07 ± 1.93	1.80 ± 1.84	1.84 ± 1.66	2.10 ± 1.84	
Nodule types					
Nonsolid (n = 16)					
MTV	8193 ± 5994	9884 ± 6685	5885 ± 5490	3041 ± 2475	3804 ± 3283
%VE	160% ± 142%	222% ± 183%	74% ± 90%	-12% ± 48%	
SUV_{mean}	0.49 ± 0.18	0.47 ± 0.17	0.53 ± 0.18	0.61 ± 0.19	
Part-solid (n = 30)					
MTV	7054 ± 5603	9946 ± 5297	7446 ± 6353	4393 ± 3619	6457 ± 4753
%VE	67% ± 173%	145% ± 212%	38% ± 97%	-16% ± 67%	
SUV_{mean}	0.98 ± 0.62	0.82 ± 0.44	0.91 ± 0.47	1.08 ± 0.64	
Solid (n = 87)					
MTV	5298 ± 3761	8263 ± 4840	7639 ± 5842	5326 ± 4729	8053 ± 7296
%VE	8% ± 111%	89% ± 179%	37% ± 114%	-2% ± 95%	
SUV_{mean}	2.74 ± 2.07	2.39 ± 2.02	2.41 ± 1.79	2.72 ± 1.97	
Lesion size					
Small (n = 53)					
MTV	3649 ± 2643	6827 ± 4536	3751 ± 2645	2473 ± 2339	2573 ± 1824
%VE	108% ± 179%	240% ± 224%	91% ± 141%	26% ± 116%	
SUV_{mean}	1.32 ± 1.12	1.03 ± 0.82	1.22 ± 0.91	1.38 ± 0.99	
Large (n = 80)					
MTV	7628 ± 4969	10170 ± 5223	9792 ± 6237	6409 ± 4632	10234 ± 6765
%VE	-5% ± 79%	37% ± 108%	9% ± 58%	-28% ± 45%	
SUV_{mean}	2.57 ± 2.19	2.32 ± 2.12	2.26 ± 1.90	2.57 ± 2.11	
Lesion uptake					
Low (n = 57)					
MTV	7186 ± 5413	9771 ± 5706	6506 ± 5550	3977 ± 3296	4403 ± 4041
%VE	134% ± 168%	230% ± 216%	93% ± 137%	25% ± 110%	
SUV_{mean}	0.71 ± 0.28	0.63 ± 0.24	0.73 ± 0.27	0.82 ± 0.29	
High (n = 76)					
MTV	5185 ± 3736	8138 ± 4721	8043 ± 6108	5488 ± 4886	9265 ± 7280
%VE	-31% ± 38%	34% ± 113%	3% ± 52%	-30% ± 48%	
SUV_{mean}	3.09 ± 2.02	2.68 ± 2.02	2.68 ± 1.78	3.05 ± 1.93	

%VE = percentage volume error, CTV = computed tomography volume, high uptake lesion = $SUV_{max} > 2$, large lesion = d > 20 mm, low uptake lesion = $SUV_{max} \leq 2$, MTV = metabolic tumor volume, RC = recovery coefficients, small lesion = d ≤ 20 mm, SUV_{mean} = mean standard uptake value.

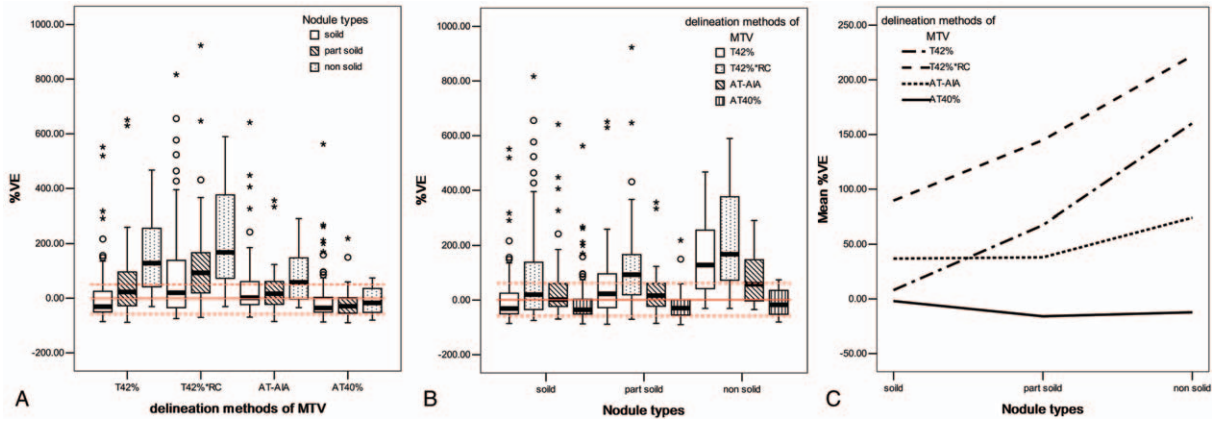

Figure 3. %VE of 4 delineation methods of MTV (A), %VE of 3 nodule types (B), mean %VE of 3 nodule types (C). (A) For each method a group of 3 boxplots are presented, each belonging to lung nodule types. Each boxplot represents the distribution (mean and quartiles) of validation results. (B) For each nodule type, a group of 4 boxplots are presented, each belonging to delineation methods. (C) Mean %VE values are presented with respect to nodule types for 4 methods. %VE = percentage volume error, MTV = metabolic tumor volume.

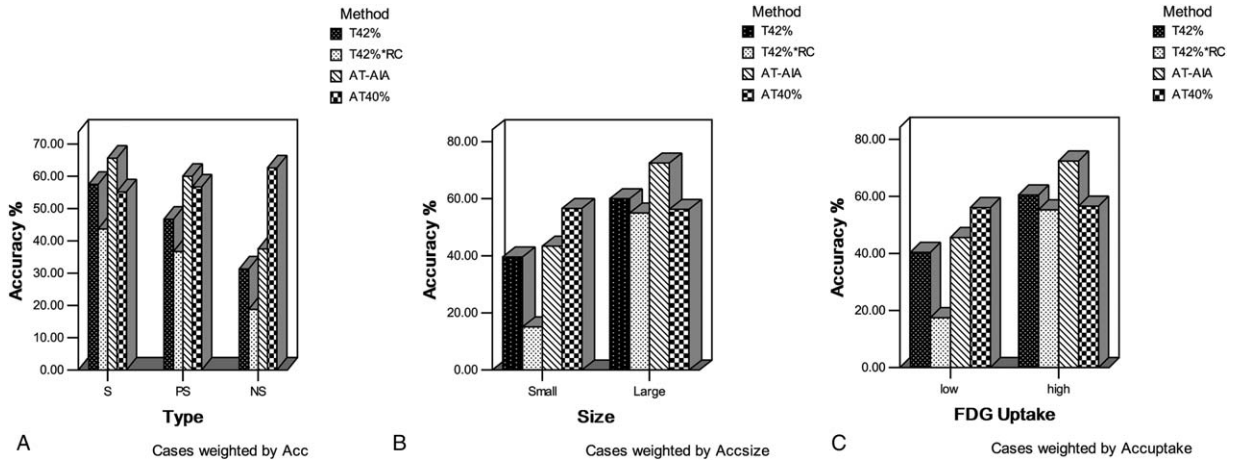

Figure 4. The segmental accuracy of 4 methods in different nodule type (A), size (B), and FDG uptake groups (C). The AT40% method was superior for small, low uptake, nonsolid lesions; the AT-AIA method was superior for large, high uptake, solid lesions. AT-AIA = adaptive iterative algorithm.

R values of AT-AIA and AT40% were higher than that of T42% in most groups. The *R* value of T42% × RC was not as good as those of the other methods (Supplemental Digital Content Table 5, http://links.lww.com/MD/C34).

The ICC was 0.933 between observers when using the fourth delineation method (AT40%), which means good reliability.

In the survival analysis, we used adaptive iterative algorithm (AT-AIA) in solid lesions, adaptive thresholding (AT40%) in nonsolid and part-solid lesions. In univariate analysis, MTV was significantly associated with PFS ($P = .04$); patients with high MTV were associated with poor prognosis. In multivariate analysis, only MTV was independent prognostic factors among 5 PET/CT metabolic parameters with a *P* value of .031 (RR, 1.118; 95% CI, 1.010–1.237).

The flowchart of summarizing the methods' performance based on different parameters of 133 lung nodules data is shown

Table 5

The underestimated and overestimated percentage of 4 methods with variation.

	T42%		T42% × RC		PETVCAR		AT40%	
Nodule types	**Under%**	**Over%**	**Under%**	**Over%**	**Under%**	**Over%**	**Under%**	**Over%**
Solid	26.4	16.1	11.5	44.8	6.9	27.6	28.7	16.1
Part-solid	13.3	40.0	3.3	60.0	6.7	33.3	33.3	10.0
Nonsolid	0	68.8	0	81.3	0	62.5	25.0	12.5

The %VE more than 50% means overestimated, less than –50% means underestimated.

Over = overestimated, PETVCAR = PET volume computerized assisted reporting, RC = recovery coefficients, Under = underestimated.

in Figure 5. This diagram provides an overview of the relative performance of the best performing methods for different situations. Depending on the types of data and application, clinicians can use this flowchart to aid their selection of the most appropriate method.

4. Discussion

In general, early stage lung adenocarcimomas are of lower FDG uptake compared to other histological subtypes of NSCLC, especially the nonsolid and some subtype of part-solid adenocarcimomas.[11,12,22,23,36–39] Smaller nodules especially are more likely to have partial volume effects. The investigation of the segmentation method on small heterogeneous lung nodules with low FDG uptake is still rare. Currently, there is no universally accepted segmentation method for such lesion yet. The aim of this paper is to investigate the suitable segmentation method for small, low uptake and heterogeneous lung cancer lesions.

During the comparison of the MTVs, the "true" volume of the lesions needs to be determined. However, there is no appropriate reference for the evaluation of volumes. Although a number of recent papers use macroscopic specimen obtained from histology as reference,[40,41] there is still problematic since the irregular contraction can occur during tissue fixation, and the criterion of contraction rate is quite different. In Schaefer et al's[42] research, they used pathology as the ground truth or CT as a ground truth surrogate, and recommended consensus contours from multiple PET segmentations as a new reference. Nestle et al[17] calculated "expanded" CT volumes according to the smallest margins recommended for motion correction as the standard (the expansion was 0.15 cm lateral, 0.2 cm anteroposterior, and 0.3 cm craniocaudal), and she thought the expanded CTV appeared to be closest to the true PET volumes. Caldwell et al[43] had also reported that the volumes of chest tumors as measured by PET would be equal or larger than the volumes measured by CT. Previous literatures reported that a difference less than 30% between CTV and MTV was considered clinical acceptable.[29–34] In our retrospective study, the previous recorded and described of each specimen slice to estimate the pathological volume was unavailable. Considering all above reasons, we compared MTVs with CTV and evaluated the accuracy using ±50% as criteria, since the lesions of stage I lung adenocarcinoma in this study were smaller, with lower FDG uptake and more heterogeneity than other researches.

The nonsolid and part-solid lung nodules, in our study, were usually with low FDG uptake ($R = 0.68$). It means with the increase of nodule density from GGN to solid nodule, the FDG uptake increased. Our study showed that the delineated MTV were overestimated in most of the cases using T42% in nonsolid and part-solid lung nodules. Moreover, the threshold of T42% × RC is too low, so as to involve more false positive background uptakes in nonsolid and part-solid nodules. AT40% seems, therefore, the best segmentation method in low FDG uptake nodules. It adapts the threshold value according to the mean background SUV. The uptake in normal lung tissue is heterogeneous which might introduce some variations. AT40% is the only method in this study considering the metabolic contrast between target lesion and background uptake information, and considering the location of the lesion at different region of the lung which may improve the accuracy. It appears to be more stable against the heterogeneity of tumor uptake and the broad variation of SUV_{max} values than the other methods in this study.[17] Therefore AT40% should be the optimal choice in nonsolid and part-solid nodules with low uptake. Moreover, in the reproducibility evaluation, there was good reliability between observers when using AT40% method. Furthermore, in the survival analysis, using AT40% method was potentially validated reasonable.

Both AT-AIA and T42% showed good performance in large, high uptake solid lesions in our study. However, AT-AIA seemed

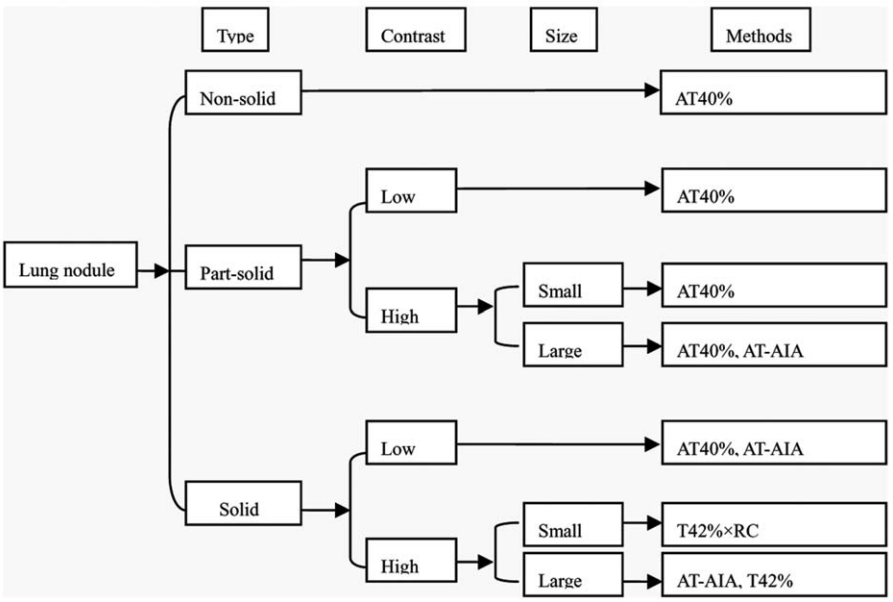

Figure 5. Flow chart summarizing methods' performance in stage I lung adenocarcinoma. The options for each parameter are presented in rectangles and the best performing method this type of data is presented alongside these rectangles. The methods are ordered according to the corresponding %VE ± SD, accuracy, and the correlation coefficient. %VE = percentage volume error, SD = standard deviation.

the best method. It showed the highest correlation value with CTV, and the highest accuracy in this study.

T42% is used widespread clinically. This method is based on the homogeneous phantom study with high contrast (8:1). It is usually applicable to tumors whose diameters are larger than 20 mm and with high uptake.[17] It is known that due to the physical principles and the physical limitation resulting from the poor spatial resolution of PET systems, T42% is not applicable to low uptake small lesion.[44] Messa et al[29] and Bradley et al[31] reported the discrepancies between the imaging modalities of PET and CT in tumor volume delineation. When correlated with CTV, PET either underestimated or overestimated the volume due to a number of factors especially partial volume effect.[45] In this study, T42% and T42% × RC overestimated the MTV of nonsolid nodules in almost all of the cases. In these cases, the partial volume effect affects the accurate estimation of FDG uptake strongly. Another reason might be that delineation volumes include the noise of background or nontumor tissues since the contrast between lesion and background is too small to detect.

Firouzian et al[18] reported that lesion size and contrast had impact on the relative performance of the delineation methods. In this study, we considered that lesion type was another important impact factor in addition to lesion size and contrast. The lesion type had good correlation with SUV_{max} in this study. Analyzing nodule type is more straightforward than measuring the SUV_{max} and diameter. Therefore, the radiologist should firstly consider about lung nodule type before selecting delineation method. The survival analysis of these patients was potentially validated that the delineation method classification according to the nodule type and FDG uptakes is reasonable.

The limitations of this study are as follows. The first limitation is the lack of correlation with pathological specimens, so the true representation of the tumor volume is not known. But the correlation of imaging with pathological specimens is problematic because of the contraction that can occur during tissue fixation. Xu et al[46] reported that although the change ratio of the sample dimensions before and after fixation was considered, manual measurement errors could not be avoided, and they failed to overlap the volumes from PET imaging and histopathology because of the lack of reliable markers in pathologic sections. Instead of comparing with pathological specimens, we evaluate the agreement between MTV and CTV using ± 50% as criteria. However, it is also not absolute certainty which needs to be carried out on more future studies. Second, the insufficient numbers of nonsolid and part-solid nodules and the heterogeneous distribution in 3 nodule types leaded to the statistical significance could not be demonstrated. The AT40% method was not strikingly better overall, yet is was superior for small, low uptake, nonsolid lesions. However, the number of these lesions was relatively small. For the future research, it would be helpful to include more nodules, especially nonsolid and part-solid ones, to improve statistical validity. Third, we did not investigate the impact caused by the different proportion of solid components within the part-solid nodules.

5. Conclusions

Lesion type, nodular size, and FDG uptake had big impact on the relative performance of the delineation methods. AT40% showed best performance in small, low uptake, nonsolid and part-solid lesions. AT-AIA was suitable for relatively large, high uptake, solid lesions.

Acknowledgments

A huge number of people helped to investigate and analyze the patients in the past few years. The authors wish to express our gratitude to all of them. Among them are Dr. Rong Zheng, Dr. Wen-jie Zhang, Dr. Ying Liang, Dr. Xiao-meng Li, Dr. Lv Lv, Dr. Jian-hua Geng, technician Shao-meng Du, and all staffs of PET-CT center.

References

[1] Liao S, Penney BC, Zhang H, et al. Prognostic value of the quantitative metabolic volumetric measurement on 18F-FDG PET/CT in Stage IV nonsurgical small-cell lung cancer. Acad Radiol 2012;19:69–77.
[2] Iwano S, Kishimoto M, Ito S, et al. Prediction of pathologic prognostic factors in patients with lung adenocarcinomas: comparison of thin-section computed tomography and positron emission tomography/computed tomography. Cancer Imaging 2014;14:3.
[3] Morimoto D, Takashima S, Sakashita N, et al. Differentiation of lung neoplasms with lepidic growth and good prognosis from those with poor prognosis using computer-aided 3D volumetric CT analysis and FDG-PET. Acta Radiol 2014;55:563–9.
[4] Park B, Kim HK, Choi YS, et al. Prediction of pathologic grade and prognosis in mucoepidermoid carcinoma of the lung using (1)(8)F-FDG PET/CT. Korean J Radiol 2015;16:929–35.
[5] Park SY, Cho A, Yu WS, et al. Prognostic value of total lesion glycolysis by 18F-FDG PET/CT in surgically resected stage IA non-small cell lung cancer. J Nucl Med 2015;56:45–9.
[6] Yoo Ie R, Chung SK, Park HL, et al. Prognostic value of SUVmax and metabolic tumor volume on 18F-FDG PET/CT in early stage non-small cell lung cancer patients without LN metastasis. Biomed Mater Eng 2014;24:3091–103.
[7] Domachevsky L, Groshar D, Galili R, et al. Survival prognostic value of morphological and metabolic variables in patients with stage I and II non-small cell lung cancer. Eur Radiol 2015;25:3361–7.
[8] Tsai YM, Huang TW, Hsu HH, et al. Prognostic significance of the number of removed lymph nodes at lobectomy in patients with positron emission tomography-computed tomography-negative N2 non-small cell lung cancer. Onkologie 2013;36:492–6.
[9] Satoh Y, Onishi H, Nambu A, et al. Volume-based parameters measured by using FDG PET/CT in patients with stage I NSCLC treated with stereotactic body radiation therapy: prognostic value. Radiology 2014;270:275–81.
[10] Vu CC, Matthews R, Kim B, et al. Prognostic value of metabolic tumor volume and total lesion glycolysis from (1)(8)F-FDG PET/CT in patients undergoing stereotactic body radiation therapy for stage I non-small-cell lung cancer. Nucl Med Commun 2013;34:959–63.
[11] Goudarzi B, Jacene HA, Wahl RL. Diagnosis and differentiation of bronchioloalveolar carcinoma from adenocarcinoma with bronchioloalveolar components with metabolic and anatomic characteristics using PET/CT. J Nucl Med 2008;49:1585–92.
[12] Khalaf M, Abdel-Nabi H, Baker J, et al. Relation between nodule size and 18F-FDG-PET SUV for malignant and benign pulmonary nodules. J Hematol Oncol 2008;1:13.
[13] Hyun SH, Ahn HK, Kim H, et al. Volume-based assessment by (18)F-FDG PET/CT predicts survival in patients with stage III non-small-cell lung cancer. Eur J Nucl Med Mol Imaging 2014;41:50–8.
[14] Lin Y, Lin WY, Kao CH, et al. Prognostic value of preoperative metabolic tumor volumes on PET-CT in predicting disease-free survival of patients with stage I non-small cell lung cancer. Anticancer Res 2012;32:5087–91.
[15] Chung HW, Lee KY, Kim HJ, et al. FDG PET/CT metabolic tumor volume and total lesion glycolysis predict prognosis in patients with advanced lung adenocarcinoma. J Cancer Res Clin Oncol 2014;140:89–98.
[16] Moule RN, Kayani I, Prior T, et al. Adaptive 18fluoro-2-deoxyglucose positron emission tomography/computed tomography-based target volume delineation in radiotherapy planning of head and neck cancer. Clin Oncol (R Coll Radiol) 2011;23:364–71.
[17] 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;46:1342–8.
[18] Firouzian A, Kelly MD, Declerck JM. Insight on automated lesion delineation methods for PET data. EJNMMI Res 2014;4:69.

[19] Austin JH, Muller NL, Friedman PJ, et al. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. Radiology 1996;200:327–31.
[20] Hansell DM, Bankier AA, MacMahon H, et al. Fleischner Society: glossary of terms for thoracic imaging. Radiology 2008;246:697–722.
[21] Wormanns D, Hamer OW. Glossary of terms for thoracic imaging—German version of the Fleischner Society Recommendations. Rofo 2015;187:638–61.
[22] LIM, YUJ, LIUN, ZHANGP, FUZ, YANGG. The relationship between fluorodeoxyglucose uptake and clinical stage in patients with adenocarcinoma or squamous cell carcinoma in the lung. Chin J Clin Oncol 2008;35:485–7.
[23] Zhao SJ, Wu N, Zheng R, et al. Primary tumor SUVmax measured on (18)F-FDG PET-CT correlates with histologic grade and pathologic stage in non-small cell lung cancer. Zhonghua Zhong Liu Za Zhi 2013;35:754–7.
[24] Mawlawi O, Podoloff DA, Kohlmyer S, et al. Performance characteristics of a newly developed PET/CT scanner using NEMA standards in 2D and 3D modes. J Nucl Med 2004;45:1734–42.
[25] Teoh EJ, McGowan DR, Macpherson RE, et al. Phantom and clinical evaluation of the bayesian penalized likelihood reconstruction algorithm q. Clear on an LYSO PET/CT system. J Nucl Med 2015;56:1447–52.
[26] Gallivanone F, Canevari C, Gianolli L, et al. A partial volume effect correction tailored for 18F-FDG-PET oncological studies. Biomed Res Int 2013;2013:780458.
[27] Srinivas SM, Dhurairaj T, Basu S, et al. A recovery coefficient method for partial volume correction of PET images. Ann Nucl Med 2009;23:341–8.
[28] L-j YU, W-k WU, ZHAO, Z-s. Improving diagnostic performance of PET/CT in single lung nodule with partial volume effect correction. Chin J Nucl Med 2007;27:306–8.
[29] Messa C, Ceresoli GL, Rizzo G, et al. Feasibility of [18F]FDG-PET and coregistered CT on clinical target volume definition of advanced non-small cell lung cancer. Q J Nucl Med Mol Imaging 2005;49:259–66.
[30] Deniaud-Alexandre E, Touboul E, Lerouge D, et al. Impact of computed tomography and 18F-deoxyglucose coincidence detection emission tomography image fusion for optimization of conformal radiotherapy in non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2005;63:1432–41.
[31] Bradley J, Thorstad WL, Mutic S, et al. Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2004;59:78–86.
[32] Erdi YE, Rosenzweig K, Erdi AK, et al. Radiotherapy treatment planning for patients with non-small cell lung cancer using positron emission tomography (PET). Radiother Oncol 2002;62:51–60.
[33] Ashamalla H, Rafla S, Parikh K, et al. The contribution of integrated PET/CT to the evolving definition of treatment volumes in radiation treatment planning in lung cancer. Int J Radiat Oncol Biol Phys 2005;63:1016–23.
[34] Brianzoni E, Rossi G, Ancidei S, et al. Radiotherapy planning: PET/CT scanner performances in the definition of gross tumour volume and clinical target volume. Eur J Nucl Med Mol Imaging 2005;32:1392–9.
[35] Ohno Y, Koyama H, Onishi Y, et al. Non-small cell lung cancer: whole-body MR examination for M-stage assessment–utility for whole-body diffusion-weighted imaging compared with integrated FDG PET/CT. Radiology 2008;248:643–54.
[36] Zhou J, Li Y, Zhang Y, et al. Solitary ground-glass opacity nodules of stage IA pulmonary adenocarcinoma: combination of 18F-FDG PET/CT and high-resolution computed tomography features to predict invasive adenocarcinoma. Oncotarget 2017;8:23312–21.
[37] Hattori A, Suzuki K, Matsunaga T, et al. Tumour standardized uptake value on positron emission tomography is a novel predictor of adenocarcinoma in situ for c-Stage IA lung cancer patients with a part-solid nodule on thin-section computed tomography scan. Interact Cardiovasc Thorac Surg 2014;18:329–34.
[38] de Geus-Oei LF, van Krieken JH, Aliredjo RP, et al. Biological correlates of FDG uptake in non-small cell lung cancer. Lung Cancer 2007;55:79–87.
[39] Suzawa N, Ito M, Qiao S, et al. Assessment of factors influencing FDG uptake in non-small cell lung cancer on PET/CT by investigating histological differences in expression of glucose transporters 1 and 3 and tumour size. Lung Cancer 2011;72:191–8.
[40] Zhuang M, Dierckx RA, Zaidi H. Generic and robust method for automatic segmentation of PET images using an active contour model. Med Phys 2016;43:4483–94.
[41] Abdoli M, Dierckx RA, Zaidi H. Contourlet-based active contour model for PET image segmentation. Med Phys 2013;40:082507–12.
[42] Schaefer A, Baillet VM, Dewalle-Vignion AS, et al. Impact of consensus contours from multiple PET segmentation methods on the accuracy of functional volume delineation. Eur J Nucl Med Mol Imaging 2016;43:911–24.
[43] Caldwell CB, Mah K, Skinner M, et al. Can PET provide the 3D extent of tumor motion for individualized internal target volumes? A phantom study of the limitations of CT and the promise of PET. Int J Radiat Oncol Biol Phys 2003;55:1381–93.
[44] Bettinardi V, Presotto L, Rapisarda E, et al. Physical performance of the new hybrid PETCT Discovery-690. Med Phys 2011;38:5394–411.
[45] Hanna GG, Hounsell AR, O'Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. Clin Oncol (R Coll Radiol) 2010;22:515–25.
[46] Xu W, Yu S, Ma Y, et al. Effect of different segmentation algorithms on metabolic tumor volume measured on 18F-FDG PET/CT of cervical primary squamous cell carcinoma. Nucl Med Commun 2017;38:259–65.