



UvA-DARE (Digital Academic Repository)

Optimising the definition of the target volume for lung cancer radiotherapy

Mercieca, S.

Publication date

2020

Document Version

Other version

License

Other

[Link to publication](#)

Citation for published version (APA):

Mercieca, S. (2020). *Optimising the definition of the target volume for lung cancer radiotherapy*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



CHAPTER 3

Comparison of SUVmax and SUVpeak based segmentation to determine primary lung tumour volume on FDG PET-CT correlated with pathology data.

Susan Mercieca,
José S.A. Belderbos,
Judith van Loon,
Kenneth Gilhuijs,
Peter Julyan,
Marcel van Herk.

Radiotherapy and Oncology, 2018; 129(2):227-233.

ABSTRACT

Purpose

The aim of the study was to compare simple SUVmax and SUVpeak based segmentation methods for calculating the lung tumour volume, compared to a pathology ground truth.

Methods

Thirty patients diagnosed with early stage Non-Small Cell lung cancer (NSCLC) underwent surgical resection in the Netherlands between 2006 and 2008. FDG PET-CT scans for these patients were acquired within a median of 20 days before surgery. The tumour volume for each percentage SUVmax and SUVpeak threshold, with and without background correction, were calculated for each patient. The percentage threshold that provided the tumour volume that corresponded best with the pathology volume was considered to be the optimal threshold. The optimal thresholds were plotted as a function of tumour volume using a power-law function and cross-validated using the leave-one-out technique.

Results

The mean optimal percentage threshold was 50% +/-10% and 62% +/-15% for the SUVmax and SUVpeak without background correction respectively and 47% +/-10% and 60 +/-15% for the SUVmax and SUVpeak with background correction respectively. The optimal threshold curves could be fitted well with power-law function. After cross-validation, the correlation between the effective tumour diameter in pathology and autosegmentation was 0.900 and 0.905 for the SUVmax and SUVpeak without background correction respectively and 0.913 and 0.908 for the SUVmax and SUVpeak with background correction respectively.

Conclusion

No benefit was shown on clinical data for the SUVpeak based segmentation method over a SUVmax based one. Both methods can be used to determine the tumour volumes in resected NSCLC tumours.

3.1 INTRODUCTION

An accurate estimation of the lung tumour volume from [18F]-fluorodeoxyglucose positron emission computed tomography (FDG PET-CT) is essential for diagnosis [1], radiotherapy planning [2,3] and therapy response assessment [4,5]. The manual definition of lung tumours on FDG PET-CT images is time consuming and prone to observer variation leading to an inaccurate estimation of the tumour volume [6,7]. To overcome this problem, various automatic and semi-automatic segmentation algorithms have been proposed using amongst others, source-to-background ratios, standardised uptake values (SUV), pipeline segmentation and full kinetic analysis[6,8]. These methods vary significantly in complexity, computational resources used and their ability to account for tumour heterogeneity [8,9]. The European Association of Nuclear Medicine Research (EARL) highlights the need for standardisation of tumour volume measurements so as to facilitate the comparison of treatment outcomes across multiple institutions [10].

The most common segmentation method used to define the tumour volume is based on simple thresholding of the maximum SUV whereby the tumour border image-guided at a percentage threshold of the maximum intensity voxel within a predefined region of interest [9]. Shepherd et al. [9], reviewed 30 different segmentation algorithms used in 13 different institutions. The findings of this study indicate that manual segmentation still provides the most accurate delineation. However, simple threshold segmentation algorithms performed well in patient studies when compared to more complex segmentation algorithms such as pipeline methods even though the latter algorithms account better for tumour heterogeneity. Thresholding methods have the additional advantage of being simpler to use and more widely available. Nevertheless, simple thresholding methods have a number of limitations. First, there is no consensus on the optimal threshold to use for segmentation with proposed values ranging from 15-60% with 42% being the most commonly used threshold [10,11]. Van Loon et al. [14] compared the equivalent tumour volume diameter with pathology volume and found a strong correlation when contouring on FDG PET-CT using the 42% threshold of SUV_{max} with pathology. Some studies argue that a fixed threshold may not be suitable to define all tumour volumes [15,16]. As the SUV_{max} is measured in a single pixel, it may be affected by noise in the image leading to an inaccurate segmentation [15]. Furthermore, it also ignores the effect of lesions size and background FDG uptake [10]. Although the CT could be used to guide the definition of the GTV, it is important to note that the CT tends to overestimate the lung tumour volume, particularly in regions with atelectasis [14].

The SUVpeak has been proposed as an alternative evaluation method and is determined by averaging pixel values over 1.0 cc spherical kernel centred at the high uptake part of the tumour [9]. Since this method does not rely on a single pixel, it is less likely to be affected by noise and scan time, especially for small tumours. It was therefore found to be a more reliable and reproducible method for tumour segmentation in both phantom [16] and manual delineation on CT or FDG PET-CT images as gold standard [13,17-19]. Furthermore, in some studies, the SUVpeak was found to correlate better with clinical outcomes [20–23]. However, to our knowledge, no studies have been published validating the accuracy of using the SUVpeak as a basis to segment lung tumour volumes with pathology data as ground truth.

Therefore, the aim of the study was to evaluate the suitability of using SUVpeak in the definition of the tumour volume using pathology data as a gold standard. The objectives of the study were to identify:

- The impact of using background correction on the accuracy of the lung tumour volume segmentation;
- Whether the tumour volume has an impact on the optimal threshold;
- An optimal SUVpeak and SUVmax threshold that can be used to define the lung tumour volume;
- Whether the tumour volume defined on FDG PET-CT using SUVpeak provides a better correlation with the pathology lung tumour volume when compared to SUVmax.

3.2 MATERIALS AND METHODS

3.2.1 Patient Data

Between 2006 and 2008, thirty patients with non-small cell lung cancer (NSCLC) underwent surgical resection in the Netherlands in one of the following hospitals; Netherlands Cancer Institute, The Onze Lieve Vrouwe Gasthuis, or The Maastricht University Medical Centre. The patient and tumour characteristics are available in Table 3.1. Each patient had one lesion, and therefore, a total of thirty lesions were resected.

Table 3.1. Patient and tumour characteristics.

ID	m/f	Age	Tumour location	Days PET-CT and Surgery	Histology	Path Vol (cc)	Path Stage	CT Vol (cc)	CT eqdiam (cm)
1	m	78	LUL	5	SCC	5.02	1B	14.47	3.10
2	m	60	RUL	35	Mixed	5.60	1A	10.41	2.82
3	m	42	RUL	18	Mixed	11.60	1A	8.16	2.58
4	m	N/A	RLL	4	Adeno	1.45	1A	2.51	1.76
5	f	68	RUL	1	N/S	4.77	1A	7.84	2.44
6	f	65	LUL	21	Adeno	4.00	1A	4.86	2.19
7	f	45	RUL	34	Adeno	29.00	1B	49.7	4.53
8	m	54	LUL	28	Large cell	27.25	2A	49.26	4.66
9	m	56	RUL	1	Adeno	3.84	1A	10.95	2.73
10	f	65	RUL	1	Adeno	12.02	1B	21.62	3.54
11	f	59	LUL	56	Adeno	13.00	1B	17.81	3.29
12	f	62	RLL	13	Adeno	1.10	1A	3.86	1.96
13	m	61	LUL	3	Adeno	1.30	1A	5.42	2.30
14	m	77	LUL	3	SCC	0.63	1A	1.87	1.62
15	f	46	RUL	1	Adeno	0.72	1A	1.31	1.34
16	m	47	LUL	41	Large cell	3.09	1A	4.72	2.28
17	m	42	LUL	18	Large cell	28.82	2B	71.88	5.32
18	f	56	RUL	49	Adeno	2.75	1A	8.52	2.01
19	f	66	RUL	20	Adeno	12.30	1B	24.26	3.73
20	m	46	RUL	2	Mixed	3.98	1B	26.67	3.36
21	f	72	LUL	37	Adeno	12.59	1B	14.26	3.17
22	m	80	LLL	52	SCC	42.33	3A	233.83	7.77
23	f	60	LLL	42	Adeno	14.98	1B	30.63	3.95
24	f	42	RUL	35	Adeno	81.89	2B	180.5	6.99
25	m	78	RUL	3	Adeno	1.72	1A	3.82	2.04
26	m	66	LLL	1	Adeno	75.19	3A	172.67	7.08
27	f	64	RML	27	Large cell	24.38	2B	76.16	5.44
28	m	76	RLL	28	Adeno	17.48	2A	32.32	4.11
29	m	75	LUL	57	Large cell	7.50	1A	9.99	2.66
30	m	56	LLL	48	SCC	3.51	1A	6.11	2.24

Abbreviations: Male (m), Female (f), Pathology (Path), Volume (Vol), Squamous Cell Carcinoma (SCC), Left Upper Lobe (LUL), Left Lower Lobe (LLL), Right Upper Lobe (RUL), Right Middle Lobe (RML), Right Lower Lobe (RLL).

3.2.2 Pathology data analysis

Pathology samples were macroscopically and microscopically analysed and tumour volumes calculated as described by van Loon et al. [14]. The lung lobe was first inflated with 10% formaldehyde and then sectioned into 0.5cm thick slices containing the gross tumour. A photograph was then acquired of each separate slice, and the tumour was indexed on each glass slide by two experienced pathologists. The CT scan was used to correct for deformations in the tumour following surgical resection as described by Stroom et al. [24]. To obtain the volume on each slide, the indexed tumour area was calculated on each slide and multiplied by the slice thickness. The gross tumour volumes on each slide were then combined to obtain the total volume.

3.2.3 Acquisition of imaging data

FDG PET-CT scans were acquired between days 1 to 57 (median 20 days) before undergoing surgical resection. The time between the procedures was kept as short as possible so as to reduce this risk of tumour growth, as shown in Table 3.1. FDG was administered to the patients as an intravenous bolus. All patients fasted for 6 hours prior to the scan. The CT slice thickness ranged from 1 to 5 mm, depending on whether the scan was made in the diagnostic process or for study purposes. A free-breathing PET scan was acquired using a voxel size of 5 mm, and a low-dose CT was used to correct for tissue attenuation. The PET scanners were calibrated between participating institutes using a dedicated NEMA phantom as per EARL accreditation standards [25]. The GTV on CT was outlined manually and the tumour volume calculated.

3.2.4 Data Analysis

All FDG PET-CT images were loaded on in-house developed image processing software known as Worldmatch [26], and a region of interest around the tumour was outlined on all images. A spherical region of interest with a 2.5cm diameter was also defined manually in the healthy lung to calculate the mean SUV in the background. The primary tumour volumes at each percentage SUV_{max} and SUV_{peak} threshold with and without background correction were generated from the FDG PET-CT of each patient. Background correction was performed by subtracting half of the background SUV from the image. The parameters used for the SUV_{peak} determination were a resolution of 0.5mm, with a spherical uniform cross-section filter of 25x25x25 pixels, corresponding to a kernel volume of 1.09cc. The PET-CT data were resampled prior to SUV_{peak} calculation using tri-linearly interpolation while making sure that the original pixel values were included in the resampled data. This was done in order to generate a smoothing kernel close to 1cc.

The percentage SUVmax and SUVpeak thresholds that provided the same segmented tumour volume as the pathology volume was considered to be the optimal threshold. The data were fitted using the power-law function model available on the Statistical Package for the Social Sciences software (SPSS) version 20. Following visual analysis, the resulting curve from the power-law function fitted the shape of the data well. However, to further confirm the goodness of fit, the correlation of determination R^2 and the residual error using the least square criterion were calculated. The least-square criterion is defined as the summation of the mean residual errors squared. The lower the least square criterion, the better is the fit of the model [27].

In order to assess the ability of the power-law function to make volume prediction, the leave-one-out-cross-validation method was used [28]. This technique involves determining the volume for each case using a power-law function that was fitted on all clinical data excluding the corresponding case. The iterative algorithm starts with a 50% threshold, segments the tumour to estimate its volume, looks up the threshold from the power-law function and repeats the latter steps until converged.

The equivalent volume diameter (eqdiam) was then calculated for the pathology and the newly obtained segmented volumes (V) using the formula:

$$\text{Eqdiam} = 2 \left(\frac{3V}{4\pi} \right)^{1/3}.$$

The eqdiam on pathology and the eqdiam using the four segmentation methods were plotted, and the R^2 was calculated. The mean percentage difference between the eqdiam using these methods and the pathology was also calculated. Based on the Shapiro-Wilk test of normality, the Wilcoxon signed-rank test and the paired sample t-test were used accordingly to determine whether there was any statistical difference between the mean eqdiam on pathology and mean eqdiam obtained using the four segmentation methods (i.e. SUVmax and SUVpeak with and without background correction) and also between the means of all segmentation methods.

Finally, Bland-Altman plots were used to evaluate the agreement between the pathology eqdiam and the segmented eqdiam so as to identify any systemic bias or outliers [29]. The lesions were classified according to their location as central and peripheral. In the Bland-Altman plot, the differences in the eqdiam between the segmentation methods and pathology were plotted against the averages of the two eqdiams. The limits of agreement were defined as the mean difference plus and minus 1.96 times the standard deviation of the differences are identified. Any value exceeding the limits of agreement was considered an outlier. A linear regression analysis was also performed so as to identify any systemic bias in the data.

For all statistical tests, a p-value below 0.05 was considered to be statistically significant.

3.2.5 Ethical considerations

The study was approved by the central medical ethics board of the Netherlands Cancer Institute. All patients gave written informed consent to participate in this study and all patient data were anonymised.

3.3 RESULTS

A total of thirty patients diagnosed with NSCLC were analysed. The majority of these patients (77%) were diagnosed with pathologic stage 1 disease, while 17% were diagnosed with stage 2 and 7% had stage 3A disease (Table 3.1).

The mean optimal percentage thresholds were 50% +/-10% and 62% +/-15 for the SUVmax and SUVpeak respectively. The addition of background correction lowered

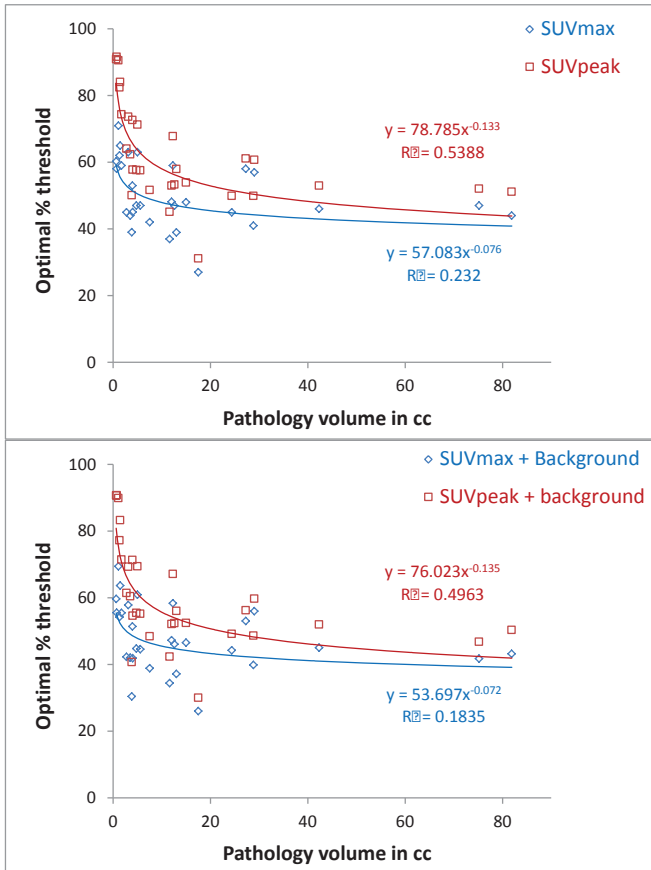


Fig. 3.1. Correlation between optimal percentage threshold and pathology volume (cc) for SUVmax and SUVpeak without and with background correction.

the optimal threshold to 47% +/-10% and 60% +/-15% for the SUVmax and SUVpeak, respectively. The optimal threshold could be fitted well with a power-law function (Fig.3.1) with a higher correlation of determination for the optimal percentage threshold for the SUVpeak ($y= 78.785x^{-0.133}$ $R^2=0.539$ $p<0.001$) as opposed to the SUVmax ($y=57.038x^{-0.076}$ $R^2=0.232$ $p=0.185$). When adding background correction, the correlation was 0.184 ($y=53.697x^{-0.072}$ $p=0.018$) and 0.496 ($y=76.023x^{-0.135}$ $p<0.001$) for the SUVmax and SUVpeak respectively (Fig.3.1).

Cross validation on the power-law function gave a correlation of 0.900 and 0.905 was obtained for the SUVmax and SUVpeak, respectively (Fig.3.2). Use of background correction improved the correlation slightly ($R^2=0.913$ for SUVmax and $R^2=0.908$ for SUV).

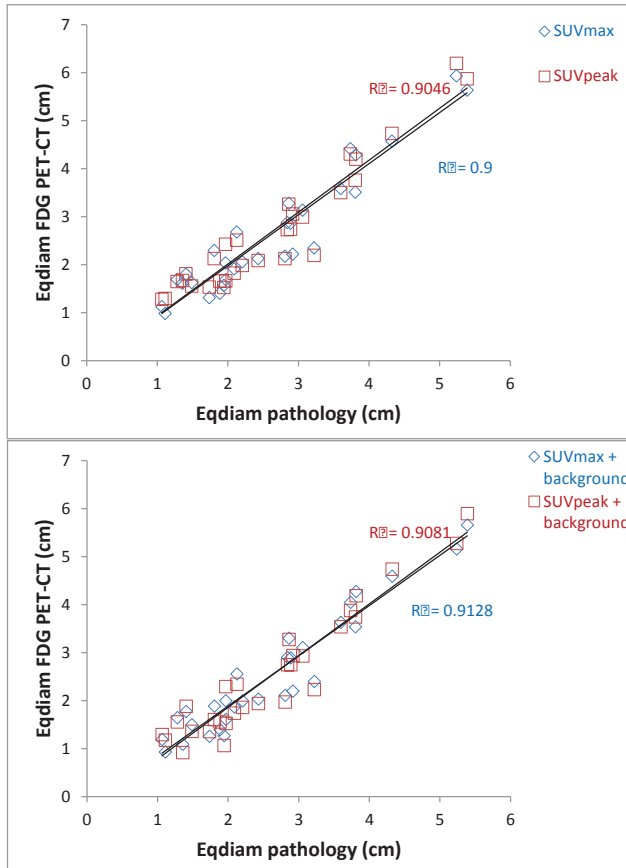


Fig. 3.2. Correlation between the pathology eqdiam and the segmented tumour volume based on the threshold fits of Fig.3.1 for the SUVpeak and SUVmax (without and with background correction).

There was no statistically significant difference between the mean eqdiam on pathology and all the different segmentation methods (Table 2). However, when comparing the mean eqdiam between the segmentation methods, the addition of background correction lead to statistically significantly smaller mean equivalent diameter when compared with the segmentation methods without background correction (Table 3.2).

Table 3.2. The table shows the mean and range difference between the four segmented eqdiam and the pathology eqdiam of the lung tumour. The probability of having systematic bias is also displayed together with the linear regression results from the Bland-Altman plot.

	Mean Eqdiam (cm)	Mean difference from pathology (cm)	Mean % difference from pathology	Range (cm)	p-Value	Slope
SUVmax no background	2.63	+0.01	1.19%	0.82 to -0.80	0.067	0.116
SUVmax + background	2.53	-0.09	-2.73%	0.65 to -0.82	0.110	0.094
SUVpeak no background	2.67	+0.06	2.75%	0.87 to -0.75	0.035	0.131
SUVpeak + background	2.52	-0.09	-3.26%	0.69 to -0.88	0.042	0.125

The mean percentage difference from pathology was -0.62% and -3.06% for the SUVmax and SUVpeak, respectively. When background correction was added, the mean percentage difference was +3.94% for the SUVmax and SUVpeak, respectively. When background correction was added, the mean percentage difference was +3.94% for the SUVmax and +4.38% for the SUVpeak.

The Bland-Altman plots (Fig.3.3) showed that all data points for all segmentation methods were within the limits of agreement, tumours being located close to the diaphragm. All segmentation (i.e. $\pm 1.96 \times \text{SD}$ of the average) except for case 28 and 26 with both methods seemed to overestimate the eqdiam for tumours larger than 3.5 cm. A statistically significant slightly positive correlation between the mean segmented eqdiam and tumour diameter was noted for the SUVpeak with background correction ($p = 0.042$) and without background correction ($p = 0.035$) but not for the SUVmax as shown in Table 3.3

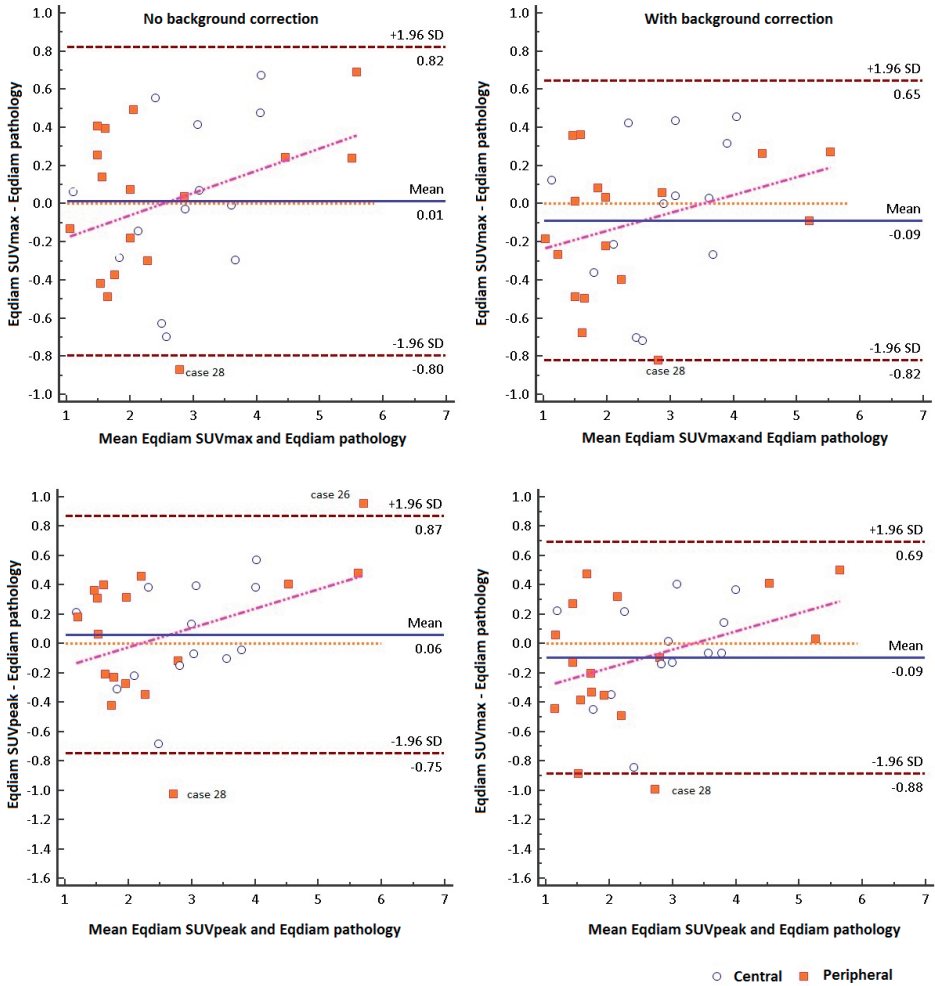


Fig.3.3. Bland-Altman plots of the primary lung tumour eqdiam (cm) obtained using SUVmax and SUVpeak with and without background versus those derived from pathology. Dashed lines represent limits of agreement ($\pm 1.96 \times SD$ of the mean). The pink lines represent the linear regression correlation, showing a slight underestimation for larger tumour volumes. The orange squares represent peripheral tumours, while blue circles represent the central tumours.

Table 3.3. Mean eqdiam (cm) diameter difference between the four segmentation method and pathology using the Wilcoxon sign rank test and paired sample t-test.

Eqdiam	Paired Differences		Paired sample t-test p-Value	Wilcoxon test p-Value
	Mean	SD		
Pathology* – SUVmax no background	-0.01	0.41	0.893	0.853
Pathology* – SUVmax with background	0.09	0.37	0.556	0.217
Pathology* – SUVpeak no background	-0.06	0.41	0.464	0.414
Pathology* – SUVpeak with background	0.09	0.40	0.131	0.209
SUVmax no background – SUVmax + background	0.10	0.19	0.007	0.003
SUVmax no background – SUVpeak no background	-0.05	0.22	0.240	0.600
SUVmax no background – SUVpeak + background	0.11	0.32	0.070	0.086
SUVmax + background – SUVpeak no background	-0.15	0.29	0.008	0.011
SUVmax + background – SUVpeak + background	0.01	0.21	0.837	0.544
SUVpeak no background – SUVmax no background	0.05	0.22	0.240	0.600
SUVpeak no background – SUVpeak + background	0.16	0.24	0.001	<0.001

* Based on the Shapiro-Wilk test, the eqdiam of the pathology data was normally distributed ($p=0.055$) but the eqdiams of all other segmentation methods were not. Therefore both the results of the parametric paired sample t-test and non-parametric equivalent the Wilcoxon signed-rank test are provided.

3.4 DISCUSSION

Segmentation based on SUVmax provides a simple, pragmatic approach to define the tumour volume on FDG PET-CT in routine clinical practice. However, SUVmax measurements are based on a single pixel that can be affected by noise and may, therefore, lead to an inaccurate tumour volume. In view of this, the SUVpeak has been proposed as an alternative since it is less affected by noise and therefore more likely expected to give an accurate estimate of the lung tumour activity and therewith its volume. The SUVpeak has been found to be a suitable alternative to the SUVmax in numerous studies using phantoms, or manual delineations as a gold standard [18,30]. To our knowledge, no studies have been published correlating the segmented volumes using SUVpeak with pathology for lung cancer. Although, studies have been published correlating the segmented tumour volume using SUVmax with pathology [14,24]. However, in these studies, the lung lobe was not always inflated following surgical resection as was performed in our study, potentially leading to an underestimation of the tumour volume [14]. Furthermore, in our study, the tumour volume was measured

in three dimensions and not estimated using the pathology tumour diameter, leading to a more accurate measure.

Similarly to previously published studies [15,16], our results indicate that a single fixed threshold is not suitable to define the tumour volume. A higher percentage threshold is required to define the volume for smaller tumours [33]. The mean optimal percentage threshold in our study was found to be 50% +/-10% and 62% +/-15% for the SUVmax and SUVpeak, respectively. The additional background correction lowered the optimal threshold to 47% +/-10% and 60% +/-15% for the SUVmax and SUVpeak, respectively. In previously published studies, the optimal threshold for the SUVmax ranged from 15% to 60% with 42% being the most commonly used threshold [10,11] while a threshold of 50% is recommended for the SUVpeak [27]. For all four segmentation methods, the threshold as a function of pathology tumour volume could be fitted well with a power-law function with a higher correlation for the optimal percentage threshold for the SUVpeak ($R^2=0.539$) as opposed to the SUVmax ($R^2=0.232$). The power function fit was slightly worse with background correction for both the SUVpeak ($R^2=0.496$) and the SUVmax ($R^2=0.184$). However, R^2 tends to be influenced by outliers and does not represent the goodness of fit of power-law function accurately. Analysis of the residual errors using the least-squares criterion showed that the residual errors were similar for all models.

Leave-one-out cross validation of the fitted power-law function showed a very strong correlation above 0.9 for all segmentation methods and the Wilcoxon test showed no statistically significant difference between the mean segmented volumes. This indicates that all methods can be reliably used to segment lung tumour volumes, and no advantage was seen for the SUVpeak over SUVmax, although using background correction provided a slightly better correlation.

The Bland-Altman plot shows that the SUVmax and the SUVpeak lead to a slight overestimation of the tumour volume, while the addition of background correction causes slight underestimation of the tumour volume (Fig.3.3). When analysing the Bland-Altman plot for systematic bias, a slightly positive correlation between the mean pathology-segmented volume and the mean eqdiam difference between the pathology and the segmentation methods was noted. Almost all tumours with an eqdiam above 3.5cm had a larger diameter on FDG PET when compared with pathology. This bias could have been introduced by the fact that the sample size was small and larger tumours were underrepresented in our model since patients with large tumours tend to be medically inoperable. This could also explain why the optimal threshold in our study was often higher than the 42% SUVmax threshold recommended in some studies [9,14].

Our study has other limitations that have to be acknowledged. Variations in tumour histology, intratumour heterogeneity, shape and location as well as scanning time after injection, can affect the FDG uptake [32–35]. Patient characteristics such as variations in glucose levels, weight and age can also have an impact on the uptake of FDG by the tumour [35]. These variables were not taken into account in our study and may have influenced the accuracy of the result. Considering the number of variables that may influence the final segmented tumour volume, irrespective of the method used the final delineations should always be checked visually and to ensure that it has been correctly segmented as recommended by Konert et al. [36].

Furthermore, respiratory tumour motion has been shown to blur the FDG signal leading to an inaccurate estimate of the SUV and an error in the tumour volume estimation [37]. Peripheral lesions and smaller tumours tend to have larger motion amplitudes and therefore are more likely to be affected by errors in volume estimation [37,38].

In our study, the location the segmented volume error did not seem to be influenced by tumour location (central versus peripheral) as shown in Fig.3.3 All the data points were within the limits of agreement with the exception of cases 28 and 26. Both tumours were peripherally located close to the diaphragm, an area that is more prone to motion. As reported by Liu et al. [37], tumours with high motion tend to have decreased lesion activity, leading to an overestimation of the tumour volume. However, in some cases particularly for tumours close to the diaphragm, the SUV can also be overestimated leading to smaller volumes, probably due to the mismatched attenuation correction as observed in case 26. The addition of background correction seems to slightly reduce this error.

For future studies, we recommend using 4D PET-CT to account for this uncertainty, but such scans were not available in this study. A drawback of the 4D PET-CT is that it significantly increases the scanning time, increasing the chance that the patient moves during the procedure. This long procedure may not be suitable for patients suffering from shortness of breath.

Due to logistic issues, the FDG PET-CT scans had to be acquired in different hospitals, potentially leading to variations in the SUV measurements. However, these scanners were calibrated using a standardised phantom to reduce this uncertainty [25]. Furthermore, the tumour might have grown between the acquisition of the FDG PET-CT and the surgical resection. Everitt et al. [39] have shown that there is a 32% probability that the tumour has to be upstaged within 24 days and that the tumour doubling time is estimated to be 66 days. However, we had a relatively short interval (median 20 days) between the two procedures to minimise this risk. No correction was applied for tumour growth since this is very difficult to accurately estimate for each individual

tumour as a number of factors could influence this such as tumour location, size, motion and histology as stated by Everitt et al. [39].

3.5 CONCLUSION

Segmentation based on SUVmax and SUVpeak with and without background correction provided similarly accurate tumour volumes compared to pathology measurements, indicating that all methods can be used to calculate the primary lung tumour volume as long as the appropriate threshold based on tumour volume is applied.

REFERENCES

- [1] Chao F, Zhang H. PET/CT in the staging of non-small-cell lung cancer. *J Biomed Biotechnol.* 2012;2012:1–8. doi: 10.1155/2012/783739.
- [2] Vojtišek R, Muzčík J, Šlampa P, Buđíková M, Hejsek J, Smolák P, et al. The impact of PET/CT scanning on the size of target volumes, radiation exposure of organs at risk, TCP and NTCP, in the radiotherapy planning of non-small cell lung cancer. *Reports Pract Oncol Radiother.* 2014;19:182–90. doi: 10.1016/j.rpor.2013.09.006.
- [3] Delikgoz Soykut E, Ozsahin EM, Yukselen Guney Y, Aytac Arslan S, Derinalp Or O, Altundag MB, et al. The use of PET/CT in radiotherapy planning: contribution of deformable registration. *Front Oncol.* 2013;3:33. doi: 10.3389/fonc.2013.00033.
- [4] Sheikhabahaei S, Mena E, Yanamadala A, Reddy S, Solnes LB, Wachsmann J, et al. The value of FDG PET/CT in treatment response assessment, follow-up, and surveillance of lung cancer. *Am J Roentgenol.* 2017;208:420–33. doi: 10.2214/AJR.16.16532.
- [5] Min SJ, Jang HJ, Kim JH. Comparison of the RECIST and PERCIST criteria in solid tumors: a pooled analysis and review. *Oncotarget.* 2016;7:27848–54. doi: 10.18632/oncotarget.8425.
- [6] Foster B, Bagci U, Mansoor A, Xu Z, Mollura DJ. A review on segmentation of positron emission tomography images. *Comput Biol Med.* 2014;50:76–96. doi: 10.1016/j.compbiomed.2014.04.014.
- [7] van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, et al. PET-CT–based auto-contouring in non–small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol.* 2007;68:771–8. doi: 10.1016/j.IJROBP.2006.12.067.
- [8] Hatt M, Lee JA, Schmidtlein CR, El Naqa I, Caldwell C, De Bernardi E, et al. Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211. *Med Phys.* 2017;44:e1–e42. doi: 10.1002/mp.12124
- [9] Shepherd T, Teras M, Beichel RR, Boellaard R, Bruynooghe M, Dicken V, et al. Comparative study with new accuracy metrics for target volume contouring in PET image guided radiation therapy. *Med Imaging IEEE Trans.* 2012;31:2006–24. doi: 10.1109/TMI.2012.2202322.
- [10] Lasnon C, Enilorac B, Popotte H, Aide N. Impact of the EARL harmonization program on automatic delineation of metabolic active tumour volumes (MATVs). *EJNMMI Res.* 2017;7:30. doi: 10.1186/s13550-017-0279-y.
- [11] Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015;42:328–54. doi: 10.1007/s00259-014-2961-x.
- [12] Gupta A, Sharma P, Patel CD, Maharjan S, Pandey A, Kumar R, et al. Size-dependent thresholding as an optimal method for tumor volume delineation on positron emission tomography-computed tomography: a Phantom study. *Indian J Nucl Med.* 2011;26:22–6. doi: 10.4103/0972-3919.84598.
- [13] Weber WA. Quantitative analysis of PET studies. *Radiother Oncol.* 2010;96:308–10. doi: 10.1016/j.radonc.2010.07.004.
- [14] van Loon J, Siedschlag C, Stroom J, Blauwgeers H, van Suylen R-J, Kneijens J, et al. Microscopic disease extension in three dimensions for non–small-cell lung cancer: development of a prediction model using pathology-validated positron emission tomography and computed tomography features. *Int J Radiat Oncol.* 2012;82:448–56. doi: 10.1016/j.ijrobp.2010.09.001.
- [15] Hoetjes NJ, Van Velden FHP, Hoekstra OS, Hoekstra CJ, Krak NC, Lammertsma AA, et al. Partial volume correction strategies for quantitative FDG PET in oncology. *Eur J Nucl Med Mol Imaging.* 2010;37:1679–87. doi: 10.1007/s00259-010-1472-7.
- [16] Bradley JD, Biehl KJ, Kong F-M, Dehdashti F, Jin J-Y, Mutic S, et al. 18F-FDG PET definition of gross tumour volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate? *J Nucl Med.* 2006;47:1808–12.
- [17] Vanderhoek M, Perlman SB, Jeraj R. Impact of the definition of peak standardized uptake value on quantification of treatment response. *J Nucl Med.* 2012;53:4–11. doi: 10.2967/jnumed.111.093443.
- [18] Akamatsu G, Ikari Y, Nishida H, Nishio T, Ohnishi A, Maebatake A, et al. Influence of statistical fluctuation on reproducibility and accuracy of SUVmax and SUVpeak: a phantom study. *J Nucl Med Technol.* 2015;43:222–6. doi: 10.2967/jnmt.115.161745.

- [19] Lodge MA, Chaudhry MA, Wahl RL. Noise considerations for PET quantification using maximum and peak standardized uptake value. *J Nucl Med.* 2012;53:1041–7. doi: 10.2967/jnumed.111.101733.
- [20] Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50:122S–50S. doi: 10.2967/jnumed.108.057307.
- [21] Sher A, Lacoueille F, Fosse P, Vervueren L, Cahouet-Vannier A, Dabli D, et al. For avid glucose tumors, the SUV peak is the most reliable parameter for [18F] FDG-PET/CT quantification, regardless of acquisition time. *EJNMMI Res.* 2016;6:21. doi: 10.1186/s13550-016-0177-8.
- [22] Aristophanous M, Berbeco RI, Killoran JH, Yap JT, Sher DJ, Allen AM, et al. Clinical Utility of 4D FDG-PET/CT scans in radiation treatment planning. *Int J Radiat Oncol.* 2012;82:e99–e105. doi: 10.1016/j.ijrobp.2010.12.060.
- [23] Laffon E, Burger IA, Lamare F, de Clermont H, Marthan R. SUVpeak performance in lung cancer: comparison to average SUV from the 40 hottest voxels. *J Nucl Med.* 2016;57:85–8. doi: 10.2967/jnumed.115.161968.
- [24] Stroom J, Blaauwgeers H, van Baardwijk A, Boersma L, Lebesque J, Theuws J, et al. Feasibility of pathology-correlated lung imaging for accurate target definition of lung tumors. *Int J Radiat Oncol.* 2007;69:267–75. doi: 10.1016/j.ijrobp.2007.04.065.
- [25] Laurence B, Lavent F, Baete K, Gheysens O, Van De Wiele C, Nicolaij D, et al. PET/CT phantom scans to acquire the EARL accreditation. *J Nucl Med.* 2013;54. 1233 1233.
- [26] van Herk M, De Jaeger K, de Munck J, Hoogeman M, Meinders J, Ploeger L, et al. A delineation system for N modalities — software aspects. *Use Comput. Radiat. Ther.* 2000:73–5. doi: 10.1007/978-3-642-59758-9_27. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [27] IBM Knowledge Center. Curve Estimation 2018. https://www.ibm.com/support/knowledgecenter/en/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/idh_curv.html (accessed 5 June 2018).
- [28] Pecori B, Lastoria S, Caracò C, Celentani M, Tatangelo F, Avallone A, et al. Sequential PET/CT with [18F]-FDG predicts pathological tumour response to preoperative short course radiotherapy with delayed surgery in patients with locally advanced rectal cancer using logistic regression analysis. *PLoS One.* 2017;12. doi: 10.1371/journal.pone.0169462e0169462.
- [29] Medcalc. Bland-Altman plot 2018. <https://www.medcalc.org/manual/blandaltman.php> (accessed 29 May 2018).
- [30] Niyazi M, Landrock S, Elsner A, Manapov F, Hacker M, Belka C, et al. Automated biological target volume delineation for radiotherapy treatment planning using FDG-PET/CT. *Radiat Oncol.* 2013;8:180. doi: 10.1186/1748-717X-8-180.
- [31] Sridhar P, Mercier G, Tan J, Truong MT, Daly B, Subramaniam RM. FDG PET metabolic tumor volume segmentation and pathologic volume of primary human solid tumors. *Am J Roentgenol.* 2014;202:1114–9. doi: 10.2214/AJR.13.11456.
- [32] Cheebsumon P, Boellaard R, de Ruyscher D, van Elmpt W, van Baardwijk A, Yaqub M, et al. Assessment of tumour size in PET/CT lung cancer studies: PET- and CT-based methods compared to pathology. *EJNMMI Res.* 2012;2:56. doi: 10.1186/2191-219X-2-56.
- [33] Towards Devic S. Biological target volumes definition for radiotherapy treatment planning: Quo Vadis PET/CT? *J Nucl Med Radiat Ther.* 2013;4. doi: 10.4172/2155-9619.1000158.
- [34] Frings V, van Velden FHP, Velasquez LM, Hayes W, van de Ven PM, Hoekstra OS, et al. Repeatability of metabolically active tumor volume measurements with FDG PET/CT in advanced gastrointestinal malignancies: a multicenter study. *Radiology.* 2014;273:539–48. doi: 10.1148/radiol.14132807.
- [35] Bai B, Bading J, Conti PS. Tumor quantification in clinical positron emission tomography. *Theranostics.* 2013;3:787–801. doi: 10.7150/thno.5629. thno.5629.
- [36] Konert T, Vogel W, MacManus MP, Nestle U, Belderbos J, Grégoire V, et al. PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014. *Radiother Oncol.* 2015;116:27–34. doi: 10.1016/j.radonc.2015.03.014.
- [37] Liu C, Pierce LA, Alessio AM, Kinahan PE, Kinahan PE. The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging. *Phys Med Biol.* 2009;54:7345–62. doi: 10.1088/0031-9155/54/24/007.
- [38] Chirindel A, Adebahr S, Schuster D, Schimek-Jasch T, Schanne DH, Nemer U, et al. Impact of 4D–18FDG-PET/CT imaging on target volume delineation in SBRT patients with central versus peripheral lung tumors. Multi-reader comparative study. *Radiother Oncol.* 2015;115:335–41. doi: 10.1016/j.radonc.2015.05.019.

- [39] Everitt S, Herschtal A, Callahan J, Plumridge N, Ball D, Kron T, et al. High rates of tumor growth and disease progression detected on serial pretreatment fluorodeoxyglucose-positron emission tomography/computed tomography scans in radical radiotherapy candidates with nonsmall cell lung cancer. *Cancer*, 2010;116:5030–7. doi: 10.1002/cncr.25392.