

Metrics to Evaluate PET Response to Therapy Based on 3D Analysis

Michael L. Goris¹, Hongyun J. Zhu²

¹Stanford University, Stanford, CA, USA

²University of Arizona Banner University, Tucson, AZ, USA

Email: mlgoris@stanford.edu

How to cite this paper: Goris, M.L. and Zhu, H.J. (2020) Metrics to Evaluate PET Response to Therapy Based on 3D Analysis. *Open Journal of Medical Imaging*, 10, 133-142.

<https://doi.org/10.4236/ojmi.2020.103013>

Received: July 25, 2020

Accepted: August 25, 2020

Published: August 28, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The disadvantage of visualizing tomography by slices is that an important attribute of the object, its volume, is not easily perceived or measured. In oncology this creates a problem, which is addressed here: if early detection and response to treatment are an important prognostic element, then volume is important. The literature has proposed surrogates to volume derived from measures on slices, but geometrically they are not well founded. Actual volume analysis is not complex, and the proposed method applies equally well to organs as to tumors. Volume based measures are more sensitive than individual SUV values, of which the commonly most used is the maximum Standardized Uptake Value (SUV_m). If the tumor volume is defined, it can be replaced by the total tumor SUV (SUV_t). If the metric for change is the ratio after/(before + after), in the patient population analyzed here, the SUV_m metric averages 0.132 for response and 0.662 for progression, the total SUV_t range is 0.069 to 0.734. In contrast to SUV_t , SUV_m is based on a weak sampling method since it is based on the value of a single voxel of more than 10 million.

Keywords

Metrics, PET Response, 3D

1. Introduction

The image of an object (in object space) is the mapping of some of the object's attribute in image space. This mapping is perfect, only if the mapping is one to one and if the relative positioning (neighbor to neighbor relative position) or coherence is maintained. The first condition would require perfect spatial resolution. Short of that, the mapping is one to many (and by extension many to one). In projection images (a mapping from a 3D object space into a 2D image

space) the mapping is not only one to many, but also structurally many to one.

Modern tomographic images computed tomography, Single Photon Computed Tomography and Positron Emission Tomography (CT, SPECT, and PET) map 3D to 3D. Image analysis is not as much affected by overlapping structures when based on image slices. However, the historic evolution of CT and SPECT, and later of PET/CT has started with a technical bias; DICOM was developed to transfer two-dimensional images (e.g. X-rays) and the data structure of CT and SPECT remained a stack of two-dimensional images. The paradox is that an image of a three-dimensional object was reduced to a stack of two-dimensional slices. This reduction, in many cases, is reinforced by the fact that the image is not isomorphic in the third dimension (the distance between slices is not the same as the distance between adjacent points within slices), and that most analytic tools (regions of interest, relative quantification) assume that the data are two-dimensional [1].

The effect on oncology is not good; early detection is based on the assumption that the growth of tumor enlarges the tumor mass and makes treatment less effective [2].

In FDG/PETCT (¹⁸F-fluoroDeoxyGlucose/Positron Emission Tomography with Computed Tomography) as applied in lymphoma, historically the evaluation of the disease stage and response to treatment has been complex. At first, disease was considered present if focal or diffuse FDG uptake above background in a location was not explained by anatomy or physiology [3] [4]. There was no explicit cutoff or a definition of changes. Later a five point scales, still qualitative, and for individual lesions a maximum Standard Uptake Value (SUV_m) was generally used [5] [6].

Eventually the Maximum Standard Uptake Value (SUV_m), and changes in SUV_m (DSUV_m) emerged to eliminate interobserver variation and standardise response to treatment [7] and a predictor of the evaluation of the response and the ultimate response to a full treatment [8].

Volume was introduced as a strange two-dimensional surrogate by the Lugano classification [6]. In general this quantification remains based on scores, rather than the classical description of response or failure: complete response (CR), partial response (PR), stable diseases (SD) progressive disease (PD) and recurrent disease.

In this paper we attempt to introduce different metrics. The SUV_m as the maximum SUV over all tumors, the total volume of all detected tumors (expressed in liters) and the sum of all the SUV values in all tumors (SUV_t). A purely mathematical derivation is the average SUV_v, which is in general totally defined by SUV_t and liter

In addition the change is expressed as an index which is the ratio of the new metric divided by the sum of the old metric plus the new metric. The new metric precisely maps into the classical descriptions of response (**Table 1**).

All the quantification is based on a volumetric search, after the elimination of normal organs.

2. Material and Methods

The study includes 17 consecutive lymphoma patients undergoing 48 18F-FDG PET/CT scans in 24 pairs, composed of one before and after treatment scan, or subsequent studies during surveillance. The median time between 2 studies is 79 days, ranging from 21 to 240 days.

Targets are either organs or tumor. The targets are delineated by the operator in 3 orthogonal maximum intensity projections (**Figure 1**). The 3 orthogonal delineations are retro projected as a mask in the image volume. The intersection of the three retro projections becomes a mask effectively separating the target from all surrounding high SUV structures (**Figure 1(A)**). The target is then refined by thresholding until the expected volume and shape are recovered (**Figure 1(B)**). This thresholding is guided by visual clues. For the tumor there is one objective clue: the threshold has to be above the lean body maximum. If the target is a normal organ or structure, the target is erased. The first targets are locations or organs with normally higher SUV's. Discrete solid tumors then are targeted individually (**Figure 2**). When that is done, the collective of targeted tumors can be analyzed for the four metrics: maximum, total, and average SUV (SUV_m , SUV_t , SUV_a) and total volume (liter), and subsequently erased.

After this only marrow activity and unspecified normal and fatty tissue remain. Marrow is handled in toto as an individual tumor. The metrics are evaluated for change by an index (**Table 1**):

$$\frac{\text{After} \times 100}{\text{Before} + \text{After}}$$

The results of this equation are bound between 0 and 100. **Table 1** illustrates the range of this index and the significance. The use of the indexes derived from the metrics rather than the metrics themselves, has the advantage that they all have the same scale.

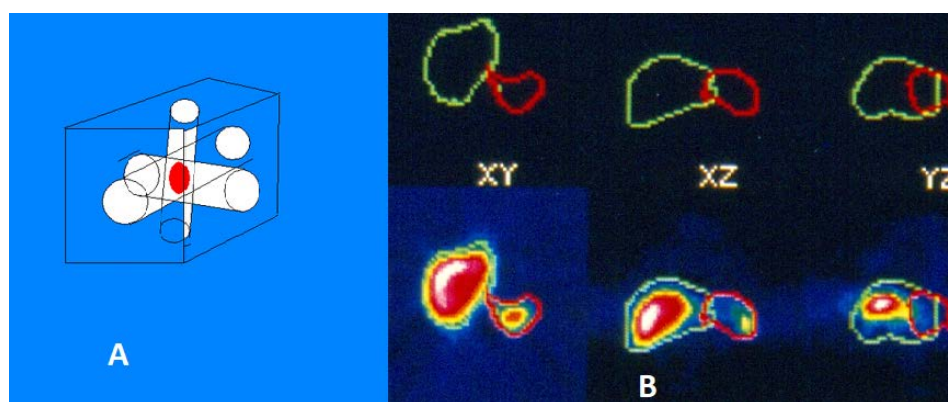


Figure 1. Delineating the organ's or tumor's volumes. Defining organ or tumor volumes. Panel A illustrates the method: 1) The organs or tumors are isolated by circumscribing the organ in three orthogonal MIP's (maximum intensity projection) images. 2) The region of delineation is retro-projected in the image volume and the intersection of the three retro-projections defines a volume containing all but only the targeted organ or tumor. Panel B illustrates: 1) That in some projections the organs overlap (here liver and spleen), but not in all three. In a last step the organic volume is refined by thresholding on the basis of the SUV.

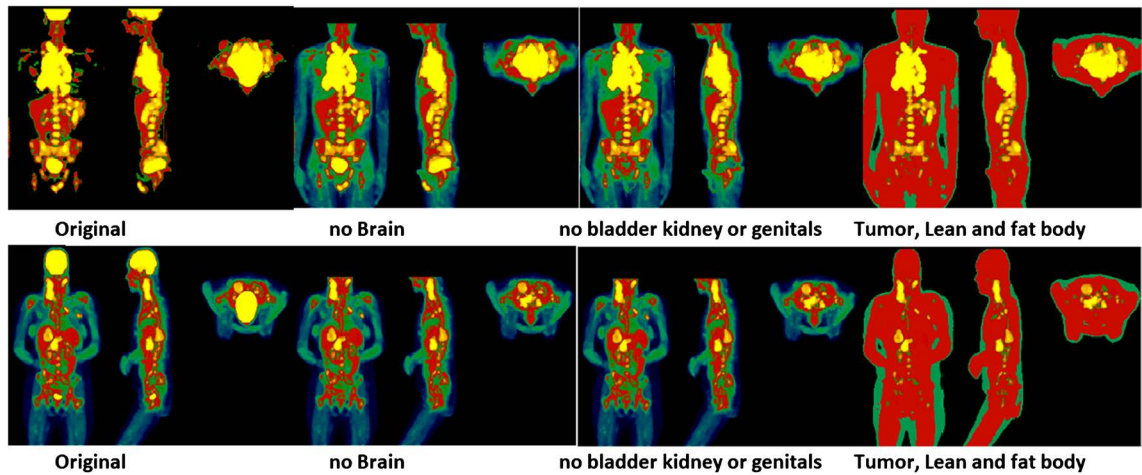


Figure 2. Eliminating normal organs, and separating tumors. Defining FAT and Lean Body. The major step is to eliminate the organs and sites with normally higher SUV's, the second to analyze the organs as they are separated. The 3 metrics are SUV_m, SUV_t and Liter. SUV_m is defined as the maximum of all tumor SUV_m's. The others are additive.

Table 1. The table illustrates the range of value of the index for the range of outcomes or evolution. Since the index' denominator is the sum of the metrics before and after treatment, unless the metrics are used for non-existing abnormalities before and after, the denominator cannot be zero. In the case of a response to treatment, the index is smaller than 50%. Progression ranges from >50% but does not go to 100%. A value of 100% defines a recurrence.

Index (%)	Significance
0.00	complete response
Between 0 and 50	Partial response
0.50	Stable disease
Between 50 and 100	Progression
100	Recurrence

3. Results

The first observation is that the bone marrow response, complete (CR) or partial PR) or evolution, progression (PD) or recurrence (RC) does not follow the evolution of solid tumors [6] [7] [9]. The difference is illustrated in **Table 2**, where globally the index between indexes (solid tumor versus marrow) is significantly different, but not the indexes derived from metrics, within solid tumors or marrow. However, in **Table 3**, the nagging observation that in some cases, frequently in this population, the marrow demonstrates near progression values while solid tumor demonstrates responses with indexes derived from SUV_m, SUV_t and Liter. This is illustrated in **Figure 3**.

The indexes yield not identical information, even within solid tumors, but the results generally correlate. Within marrow, the correlation is weaker for the indexes derived from SUV_m in relation to those derived from SUV_t and Liter (**Table 4**).

Table 2. The analysis of variance is performed as two factors with replication. One factor is marrow versus solid tumor, the other are the indexes derived from the different metrics. Only the former is significant, however, the analysis is global.

ANOVA: The metrics differ not so much within lesions as between lesions					
Source of Variation	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>
Marrow vs. Non-marrow	4964.19	1	4964.19	6.20	0.014
SUV _m vs. SUV _t vs. Liter	38.82	2	19.41	0.02	0.976
Interaction	131.29	2	65.65	0.08	0.921

Table 3. Marrow and No Marrow (solid tumors) congruence and differences. The analysis shown in this table does show that a complete or response in solid tumors is not associated with a corresponding response in the bone marrow. This is true for all three derived indexes (SUV_m, SUV_t, and Liter). The p value is based on a paired two-tailed t-test. For progression disease the discrepancies are opposed, with lower values for marrow than for solid tumors (Legend: CR = complete response, PR = partial response, SD = stable disease, PD = Progressive disease).

	SUV _m			SUV _t			Liter		
	Marrow	No Marrow	P	Marrow	No Marrow	p	Marrow	No-Marrow	P
CR	41.21	0.00	0.0002	45.18	0.00	0.007	45.487	0.000	0.006
PR	37.76	35.71	0.7453	40.54	19.18	0.081	41.214	21.200	0.098
SD	48.14	50.00	0.2283	35.61	50.00	0.267	36.540	50.000	0.284
PD	57.75	76.38	0.3433	69.68	85.18	0.673	68.813	85.520	0.680

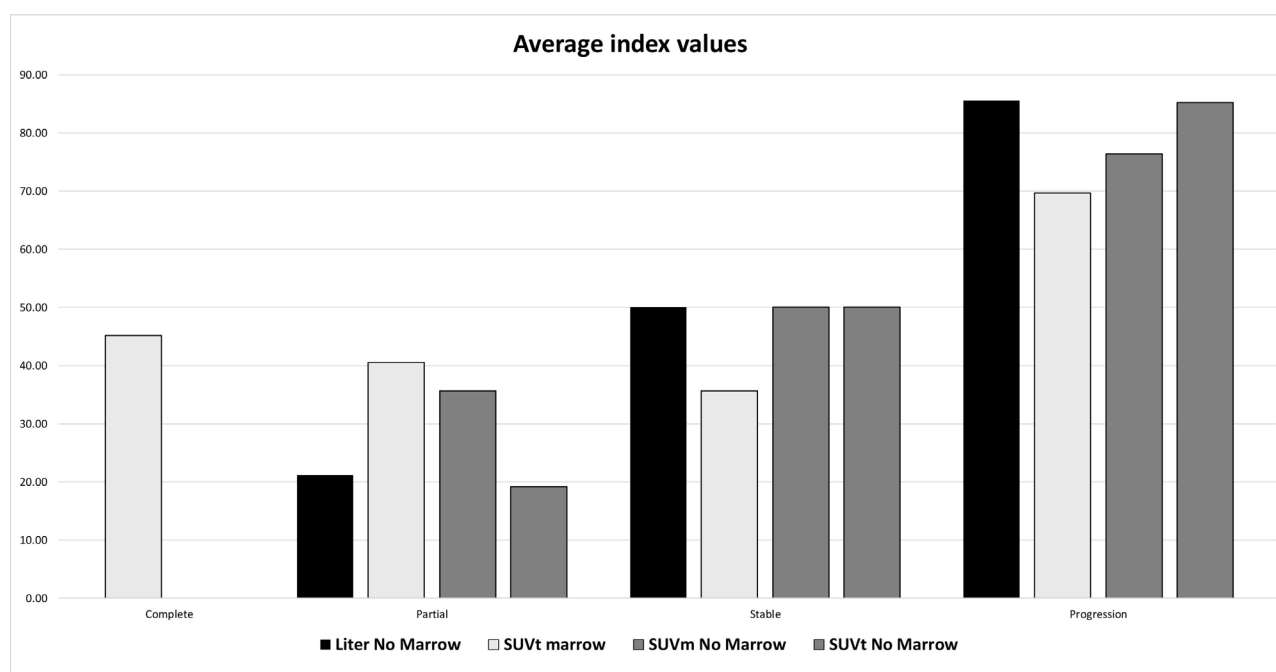


Figure 3. Marrow response differentiation. This figure illustrates that when marrow response seems to contradict solid tumor response, for responses (complete or partial), the marrow index derived from SUV_t tends to be larger, but for stable disease and progression, smaller. Note the agreement between all solid tumor indexes for a complete response.

Table 4. Correlation between the indexes derived from different metrics, within solid tumors and marrow, and between. Within solid tumors the 3 metrics correlate (Pearson's R), but between solid tumors and marrow there is no correlation between the metrics. Within marrow, there is high correlation between the indexes, from SUV_t and liter (0.993). The correlation is strong, but weak for correlation between indexes derived from SUV_m and from the 2 other metrics (0.643 and 0.619). A two-tailed t-test, with paired observations, is just above significance detecting the difference between the indexes derived from SUV_m , SUV_t and Liter respectively (For those table if Pearson's R > 0.5, p < 0.05).

Metrics	Solid Tumor			Marrow		
	SUV_mNM	SUV_tNM	LiterNM	SUV_mM	SUV_tM	LiterM
SUV_mNM	1					
SUV_tNM	0.943	1				
LiterNM	0.942	0.998	1			
SUV_mM	0.372	0.400	0.413	1		
SUV_tM	0.157	0.055	0.064	0.643	1	
LiterM	0.174	0.066	0.076	0.619	0.993	1

Table 5. The measures are SUV values (not indexes) used for the thresholding level, for the delineation of the fat and lean body. The thresholding is based on a visual evaluation of the expected shape by the operator. But, the limits for whole bodies tend to be stable. Comparing the limits of the first study (A, generally pre-treatment) with the second (B, post treatment) shows no significant differences (The value of P is derived from a paired two tailed t-test).

	Fat	body	Lean	body
	A	B	A	B
Mean	0.17	0.63	0.17	0.59
Median	0.15	0.15	0.51	0.53
SD	0.08	0.11	0.49	0.29
Min	0.03	0.08	0.31	0.38
Max	0.45	0.56	2.88	1.88
Range	0.42	0.48	2.57	1.50
T-test	P =	0.647	P =	0.731

The thresholding limits used visually to delineate organs, tumor, fatty and lean bodies are expressed in the original metrics (SUV). **Table 5** compares the first and second selection for the fat and lean body delineation in each pair (this includes the body devoid of organs, tumors and marrow), show in both cases significant correlation between Liter, SUV_t but not limit or SUV_a .

4. Discussion

This paper addresses three old problems. First, the use of SUV_m as an attempt to classification is somewhat deficient. With exploration in 2D slices, how does one

know that the search for the maximum has been successful? The search is handicapped by being performed mainly on transverse slices, in lesions whose limits have not been set.

In addition, $DSUV_m$ assumes that the locations match before and after. In our case, the definition is not that SUV_m before and after lie in the same voxel, but that they originate in the same set of tumors, which was searched over the total delineated volume.

Second, change cannot be well defined by (Before-after/before) because it does not match the response concepts of oncology. In addition, for a fixed (Before-after), the value of the ratio is very sensitive to the value of “before”. If progression is included, the range of that ratio, potentially, goes from 100 to $-\infty$. The index proposed here does range from 0 to 100, around 50% for no change or stable diseases.

Third, the surrogate for volume from the Lugano classification lacks a link to Euclidean geometry since the summation of two-dimensional shape of a certain thickness does not necessarily result in a three-dimensional object of definable volume. To that extend, it does not allow deriving a volume change. In addition, there is no reason to believe that the longest axis is totally in a single slice.

The exception of the marrow metrics compared to the solid tumors does not resolve, or rather reaffirms the question of the need for a bone marrow biopsy [6] [9] [10]. A total response in solid tumors for the indexes of SUV_m , SUV_t and Liter, should however mitigate the urgency and the fact that responses in solid tumors are associated with higher marrow indexes, but progression and recurrence not. The manual determination of the thresholds (limits) is a weakness [11] [12] but is not exclusively based on the capriciousness of the operator since shapes of organs (Table 6 and Table 7) are known. However, the distribution of

Table 6. The correlations between fat body Liter (A versus B) is 0.95, for the SUV_t 0.87 (all organs and tumors excluded) and between the selected limits 0.64 ($p < 0.01$ for all 3). The analysis is based on the indexes for SUV_t , Liter and SUV_a and the actual SUV for limit. Significant correlations are shown on a gray background and bold. The volume (Liter) is negatively, but slightly correlated with the limit both for A and B.

	A Liter	A SUV_t	A SUV_a	A limit	B Liter	B SUV_t	B SUV_a	B limit
A Liter	1							
A SUV_t	0.790	1						
A SUV_a	-0.178	0.431	1					
A limit	0.429	0.737	0.544	1				
B Liter	0.953	0.716	-0.219	0.282	1			
B SUV_t	0.855	0.870	0.119	0.494	0.831	1		
B SUV_a	-0.365	0.040	0.627	0.228	-0.489	-0.037	1	
B limit	0.179	0.435	0.479	0.640	-0.009	0.172	0.514	1

Table 7. For the lean body, the correlations are less pronounced. A value of 0.6 is associated with a p value <0.01. The analysis is based on the indexes for SUV_t, Liter and SUV_a and the actual SUV for limit. The volume (Liter) is negatively correlated to the limit, but not SUV_t; This is expected as the SSUV values decrease along the edges, while the number of voxels at edge levels represent more volume.

	A Liter	A SUV _t	A SUV _a	A limit	B Liter	B SUV _t	B SUV _a	B limit
A Liter	1							
A SUV _t	0.808	1						
A SUV _a	-0.479	0.059	1					
A limit	-0.021	0.349	0.520	1				
B Liter	0.919	0.781	-0.393	-0.116	1			
B SUV _t	0.804	0.871	-0.159	0.075	0.845	1		
B SUV _a	-0.435	-0.308	0.359	0.087	-0.448	-0.279	1	
B limit	0.080	0.444	0.465	0.273	0.253	0.359	-0.039	1

the PET tracer, expressed in Standardized Uptake Values, assumes that only three variables matter: The dose injected, the physical half-life of the tracer and injected volume (estimated by weight). It avoids the fact that the renal system excretes and that there is competition between organs and tumors. What automated systems will have to do is to find another criterion than SUV to define organ limits, or base SUV values on the total activity remaining in the body.

Unlike identifying and even diagnosing a lung tumor in a clear pulmonary background, automatic systems have not been very good at identifying organs versus tumors in PET, and even less at identifying the type of tumor [13]. The same advantage as for lung tumors applies to brain tumors [14], but some approaches are not only sophisticated but also elegant [15].

The method used here would suggest two steps for the automated systems: eliminate normal organs first, then suspect and hunt for tumors.

5. Conclusions

The method used here would suggest two steps: eliminate normal organs first, then suspect and hunt for tumors. It does not propose a thresholding algorithm to define the limit and volume of tumors. The method proposes, except for the two steps mentioned above, a process essentially based on the fact that the patients, organs, tumors, and the images are three-dimensional, and should be analyzed as such.

The advantage of SUV_t and Liter over SUV_m is not demonstrated, except that: volume has an important prognostic value, and a total SUV value is more sturdy than a single voxel SUV.

Technically, 3D images require 3D analysis.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Goris, M.L., Boudier, S. and Briandet, P.A. (1986) Interrogation and Display of Single Photon Emission Tomography Data as Inherently Volume Data. *American Journal of Physiologic Imaging*, **1**, 168-180.
- [2] Hoppe, R.T., Coleman, C.N., Cox, R.S., Rosenberg, S.A. and Kaplan, H.S. (1982) The Management of Stage I-II Hodgkin's Disease with Irradiation Alone or Combined Modality Therapy: The Stanford Experience. *Blood*, **59**, 455-465. <https://doi.org/10.1182/blood.V59.3.455.455>
- [3] Cheson, B.D., Pfistner, B., Juweid, M.E., *et al.* (2007) Revised Response Criteria for Malignant Lymphoma. *Journal of Clinical Oncology*, **25**, 579-586. <https://doi.org/10.1200/JCO.2006.09.2403>
- [4] Juweid, M.E., Stroobants, S., Hoekstra, O.S., *et al.* (2007) Use of Positron Emission Tomography for Response Assessment of Lymphoma: Consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *Journal of Clinical Oncology*, **25**, 571-578. <https://doi.org/10.1200/JCO.2006.08.2305>
- [5] Barrington, S.F., Mikhail, N.G., Kostakoglu, L., *et al.* (2014) Role of Imaging in the Staging and Response Assessment of Lymphoma: Consensus of the International Conference on Malignant Lymphomas Imaging Working Group. *Journal of Clinical Oncology*, **32**, 3048-3058. <https://doi.org/10.1200/JCO.2013.53.5229>
- [6] Cheson, B.D., Fisher, R.I., Barrington, S.F., *et al.* (2014) Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. *Journal of Clinical Oncology*, **32**, 3059-3068. <https://doi.org/10.1200/JCO.2013.54.8800>
- [7] Lin, C., Itti, E., Haioun, C., *et al.* (2007) Early ¹⁸F-FDG PET for Prediction of Prognosis in Patients with Diffuse Large B-Cell Lymphoma: SUV-Based Assessment versus Visual Analysis. *Journal of Nuclear Medicine*, **48**, 1626-1632. <https://doi.org/10.2967/jnumed.107.042093>
- [8] Zhu, H.J., Halkar, R., Alavi, A. and Goris, M.L. (2013) An Evaluation of the Predictive Value Of Mid-Treatment ¹⁸F-FDG-PET/CT Scans in Pediatric Lymphomas and Undefined Criteria of Abnormality in Quantitative Analysis. *Journal of Nuclear Medicine*, **16**, 169-174.
- [9] Juweid, M.E., Wiseman, G.A., Vose, J.M., *et al.* (2005) Response Assessment of Aggressive Non-Hodgkin's Lymphoma by Integrated International Workshop Criteria and Fluorine-18-Fluorodeoxyglucose Positron Emission Tomography. *Journal of Clinical Oncology*, **23**, 4652-4661. <https://doi.org/10.1200/JCO.2005.01.891>
- [10] Cerci, J.J., Trindade, E., Pracchia, L.F., *et al.* (2010) Cost Effectiveness of Positron Emission Tomography In Patients with Hodgkin's Lymphoma in Unconfirmed Complete Remission or Partial Remission after First-Line Therapy. *Journal of Clinical Oncology*, **28**, 1415-1421. <https://doi.org/10.1200/JCO.2009.25.4367>
- [11] Johansson, J., Alakurtti, K., Joutsa, J., Tohka, J., Ruotsalainen, U. and Rinne, J.O. (2016) Comparison of Manual and Automatic Techniques for Substriatal Segmentation in ¹¹C-Raclopride High-Resolution PET Studies. *Nuclear Medicine Communications*, **37**, 1074-1087. <https://doi.org/10.1097/MNM.0000000000000559>
- [12] Kolinger, G.D., García, D.V., Kramer, G.M., *et al.* (2019) Repeatability of [¹⁸F]FDG

PET/CT Total Metabolic Active Tumour Volume and Total Tumour Burden in NSCLC Patients. *EJNMMI Research*, **9**, Article No. 14.

<https://doi.org/10.1186/s13550-019-0481-1>

- [13] Zhang, Y., Oikonomou, A., Wong, A., Haider, M.A. and Khalvati, F. (2017) Radiomics-Based Prognosis Analysis for Non-Small Cell Lung Cancer. *Scientific Reports*, **7**, Article No. 46349. <https://doi.org/10.1038/srep46349>
- [14] Blanc-Durand, P., Van Der Gucht, A., Schaefer, N., Itti, E. and Prior, J.O. (2018) Automatic Lesion Detection and Segmentation of 18F-FET PET in Gliomas: A Full 3D U-Net Convolutional Neural Network Study. *PLoS ONE*, **13**, e0195798. <https://doi.org/10.1371/journal.pone.0195798>
- [15] Zhong, Z., Kim, Y., Zhou, L., *et al.* (2018) 3D Fully Convolutional Networks for Co-Segmentation of Tumors on PET-CT Images. *Proceeding 2018 IEEE 15th International Symposium on Biomedical Imaging*, Washington DC, 4-7 April 2018, 228-231. <https://doi.org/10.1109/ISBL2018.8363561>