

Deriving global quantitative tumor response parameters from ^{18}F -FDG PET-CT scans in patients with non-Hodgkin's lymphoma

Frederic Sampedro^a, Anna Domenech^c, Sergio Escalera^{b,d} and Ignasi Carrió^c

Objectives The aim of the study was to address the need for quantifying the global cancer time evolution magnitude from a pair of time-consecutive positron emission tomography-computed tomography (PET-CT) scans. In particular, we focus on the computation of indicators using image-processing techniques that seek to model non-Hodgkin's lymphoma (NHL) progression or response severity.

Materials and methods A total of 89 pairs of time-consecutive PET-CT scans from NHL patients were stored in a nuclear medicine station for subsequent analysis. These were classified by a consensus of nuclear medicine physicians into progressions, partial responses, mixed responses, complete responses, and relapses. The cases of each group were ordered by magnitude following visual analysis. Thereafter, a set of quantitative indicators designed to model the cancer evolution magnitude within each group were computed using semiautomatic and automatic image-processing techniques. Performance evaluation of the proposed indicators was measured by a correlation analysis with the expert-based visual analysis.

Results The set of proposed indicators achieved Pearson's correlation results in each group with respect to the expert-based visual analysis: 80.2% in progressions, 77.1% in partial response, 68.3% in mixed response, 88.5% in complete response, and 100% in relapse. In the progression and mixed response groups, the proposed indicators

outperformed the common indicators used in clinical practice [changes in metabolic tumor volume, mean, maximum, peak standardized uptake value (SUV_{mean} , SUV_{max} , SUV_{peak}), and total lesion glycolysis] by more than 40%.

Conclusion Computing global indicators of NHL response using PET-CT imaging techniques offers a strong correlation with the associated expert-based visual analysis, motivating the future incorporation of such quantitative and highly observer-independent indicators in oncological decision making or treatment response evaluation scenarios. *Nucl Med Commun* 00:000–000 Copyright © 2015 Wolters Kluwer Health, Inc. All rights reserved.

Nuclear Medicine Communications 2015, 00:000–000

Keywords: image analysis, non-Hodgkin's lymphoma, PET-computed tomography, tumor response

^aDepartment of Radiology, Faculty of Medicine, ^bComputer Vision Center, Building O, Autonomous University of Barcelona, ^cDepartment of Nuclear Medicine, Hospital de Sant Pau and ^dDepartment of Applied Mathematics, Faculty of Mathematics, University of Barcelona, Barcelona, Spain

Correspondence to Frederic Sampedro, MSc, Faculty of Medicine, Autonomous University of Barcelona, 08193 Barcelona, Spain
Tel: +34 699 805 231; e-mail: fredsampedro@gmail.com

Received 5 September 2014 Revised 19 October 2014
Accepted 26 November 2014

Introduction

Fluorine-18 fluorodeoxyglucose (^{18}F -FDG) positron emission tomography-computed tomography (PET-CT) has become a standard imaging method for the time monitoring of treatment response in a variety of tumors [1–3]. From a pair of time-consecutive whole-body PET-CT scans nuclear medicine physicians assess a patient's cancer progression or response condition following a trained visual and semiquantitative analysis of both images. Thereafter, generally, a categorical and qualitative diagnosis is provided, such as 'good response', 'slight progression', or 'strong relapse'. Although this type of information is generally enough in the clinical routine, it lacks observer independence and does not provide a continuous response scale to accurately compare between cases.

In this work we address the need and computation of observer-independent global quantitative tumor response indicators from a pair of time-consecutive PET-CT scans.

This complementary information to the physician's visual analysis would prove especially useful in comprehensive oncological treatment response evaluation and comparison scenarios, as well as in the context of studying possible cancer evolution differences related to particular clinical profiles.

This issue has been partially addressed in the literature in the form of relating time changes in local tumor metabolic activity or volume with surgical outcome parameters [4–8]. Although this methodology is well suited to recognize the value of quantifying PET-CT images, it does not provide a sound framework for designing and evaluating the proposed global response indicators due to several reasons.

First, changes in cancer spread are not taken into consideration, which, as derived from Sampedro *et al.* [9] and described later in this paper, play a key role in measuring the cancer progression or response magnitude. Second, it

is important to note, in the general case, the lack of a well-defined gold standard indicator to compare the proposed global response indicators. In particular, non-Hodgkin's lymphoma (NHL) response or progression magnitude is not well described by any clinical continuous parameter. Even if the international prognostic index is considered the current prognostication system for NHL, prognostic heterogeneity is suggested to exist among the patients within the same international prognostic index risk group [9–11]. In such a scenario, the performance of the proposed indicators can be addressed by using the information resulting from an expert-based ordering by magnitude of the cases where the indicator performance is to be measured, as shown in [3].

Thus, in this work, we start from an ordered set of NHL response/progression cases based on its magnitude (derived by the visual analysis of a consensus of experts focusing on time changes in tumor volume, aggressiveness, and spread). Then, from its associated pair of PET-CT scans, we propose and compute a set of global response/progression indicators by quantifying time changes in the segmented metabolic tumor volumes (also provided by nuclear medicine physicians). Indicator performance is addressed by a correlation analysis with the initial expert-based ordering. Aiming to maximize observer independence in the indicator computations, the possibility of using completely automatic PET tumor volume segmentation techniques is also addressed.

Materials and methods

A set of 178 whole-body FDG-PET/CT scans corresponding to NHL lymphoma patients were acquired from the Phillips Nuclear Medicine workstation at Hospital de Sant Pau (Barcelona, Spain) following all international PET/CT imaging acquisition protocols [12]. From its digital imaging and communications in medicine (DICOM) files, two coregistered three-dimensional volumes were obtained for each scan: a PET volume, in standardized uptake value (SUV) [13] units, and a CT volume (in Hounsfield units) [14]. They corresponded to 89 pairs of time-consecutive scans of the patients. The time elapsed between scans varied depending on the clinical management of each patient, with a median of 3.2 months and an interquartile range of 2 months.

Classification of each cancer evolution condition was carried out by a consensus of three independent nuclear medicine physicians into progression (31), partial response (28), mixed response (nine), complete response (13), or relapse (eight). The classification criteria were based on changes in tumor volume, aggressiveness (represented by its metabolic activity through its SUV), and spread, as illustrated in Fig. 1. Figures 2 and 3 show examples of real cases of each cancer evolution condition.

Note that the cases illustrated in Fig. 1 represent the canonical cancer evolution conditions; that is, in practice,

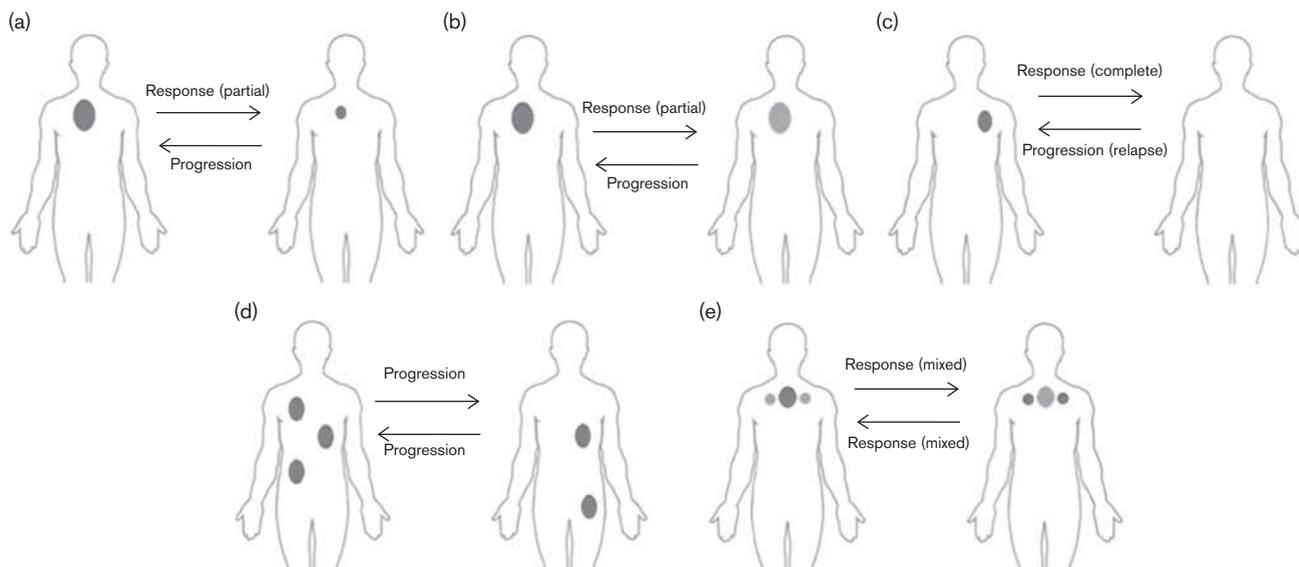
real cases may be combinations of those cases. For instance, a progression case can be presented both with an increase in tumor volume (or uptake) and with the appearance of new lesions, a response case with both a decrease in volume and uptake, or a mixed response case with both increases and decreases in volume and uptake of the persisting tumor lesions.

Then, the cases of each group are ordered by its magnitude according to the following visual criteria. For progression cases, relative increases in volume or aggressiveness in the existing tumor lesions are considered less severe than the appearance of new tumor lesions in adjacent or distant anatomical locations, respectively. However, all these variables interact, in the sense that strong volume increases of existing lesions may be considered more severe than the sole appearance of small adjacent new lesions. The ordering of partial responses is analogous, but considering the relative volume, aggressiveness, and spread it decreases (emphasizing the global tumor size reduction). From an imaging point of view, the ordering of relapses and complete responses is analogous to that of progressions and partial responses without any tumor presence in one of the scans. Mixed responses are ordered considering the overall balance of tumor volume and uptake increases and decreases of the existing tumor lesions. Figure 3 shows an ordering example of a subset of the progressions and partial responses considered in this study.

The main goal of this work was to analyze the best global quantitative indicators that model each of the cancer evolution groups so as to obtain a continuous analog of the visual qualitative assessment. The performance of each proposed indicator will be addressed by comparing (using the Pearson correlation coefficient) the ordering provided by the medical experts with the order obtained by the indicator of the same cases.

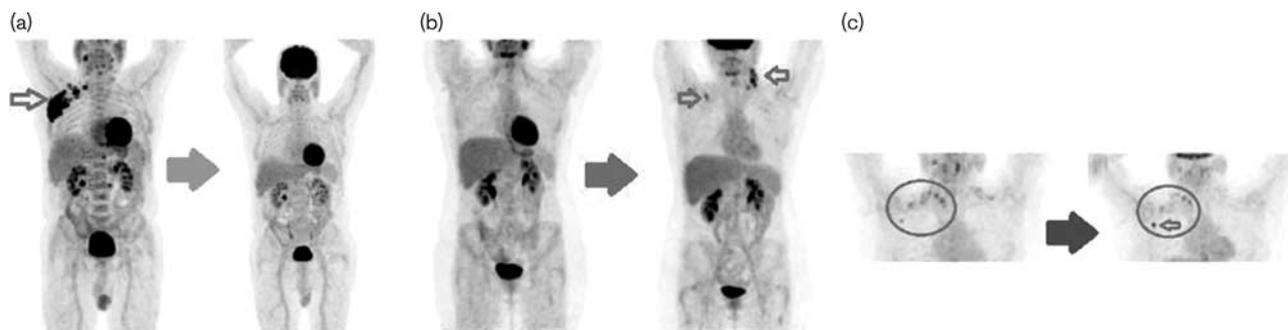
On the design of such global indicators, the ones more commonly used in clinical practice are first considered. Conceptually, in the presence of more than a single tumor lesion or highly heterogeneous tumor tissue (e.g. the presence of necrotic tissue in any of the scans), global changes in SUV_{mean} , SUV_{max} , or SUV_{peak} [4,13,15,16] will not appropriately model the strength of the progression or response condition, as they are unable to model volume increases or the appearance of new tumor lesions. In contrast, global changes in whole-body metabolic tumor volume (WBMTV) or total lesion glycolysis (TLG) [15] offer a better overall description of the magnitude of the cancer evolution. However, they still suffer from conceptual limitations: consider the cases modeled in Fig. 1d and in particular the top right progression case in Fig. 3. In such a case, these indicators may not even be valuable, as the global WBMTV has in fact decreased in time while the case is considered a strong cancer progression.

Fig. 1



Illustrations of the typical cancer evolution scenarios in nuclear medicine. When a single tumor lesion is present, its change in volume or intensity in time defines its progression or response condition (a–c). In the multilesion case, the spread of the cancer into new anatomical locations (regardless of the volume change) is associated with a progression scenario (d), whereas the intensity increase in any of the lesions is clinically associated with a mixed response scenario (e). Intensity of the tumor lesion is represented by its grayscale intensity.

Fig. 2



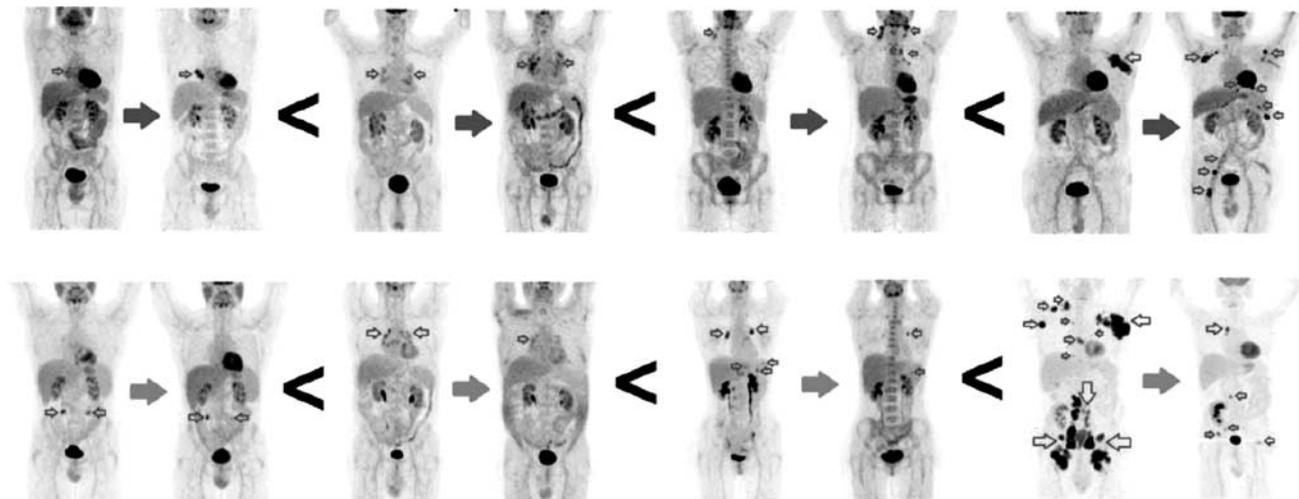
Clinical examples of complete response (a), relapse (b), and mixed response (c). The thick arrows represent the direction of time. Each PET scan is visualized using its maximum intensity projection.

Therefore, we propose a new set of indicators that seek to model more accurately the global progression or response magnitude. With an eye to future technological advances, we focus only on quantitative indicators that can be computed from the PET three-dimensional tumor segmentation masks of both time-consecutive scans. These, in the future, may be obtained accurately in an automatic manner using recent advances in machine learning-based segmentation techniques [17], thus obtaining full observer independence in the whole process. Nevertheless, as current automatic segmentation methods do not achieve the required accuracy to compute reliable indicators in this scenario [17], we focus on the use of expert-guided semiautomatic tumor segmentation masks that, although

introducing a slight observer dependence and a highly time-consuming step, provide accurate and reliable estimators of the underlying phenomena. Figure 4 illustrates this reasoning.

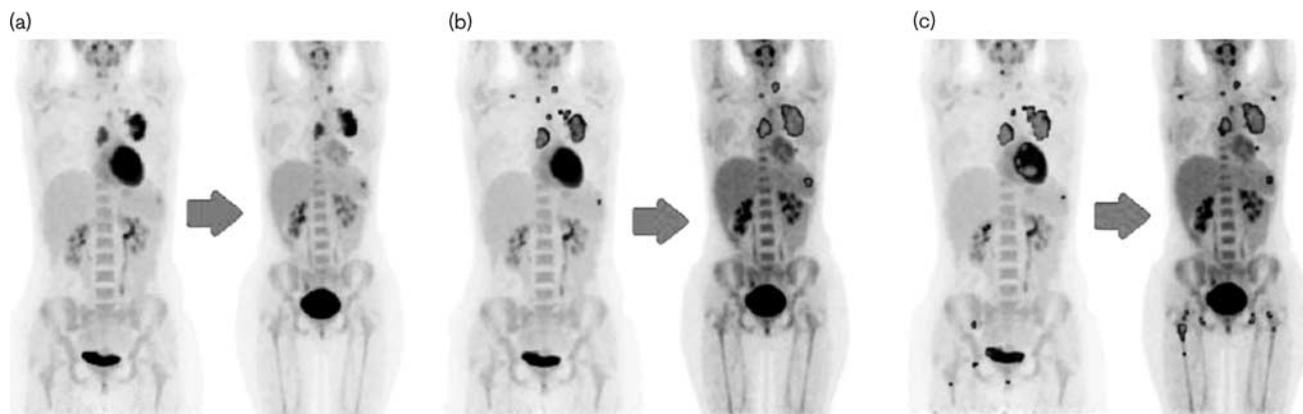
The key clinical variable that the new set of indicators need to model are changes in cancer spread, which are not modeled by the common indicators described before. A first piece of information in this respect is the change in the number of tumor-related lesions. These can be modeled computationally by the change in the number of connected components (ΔNCC) between the pair of PET tumor segmentation masks, as from an image-processing point of view a connected component [18]

Fig. 3



Ordering of a subset of progression (top) and partial response (bottom) cases of this study.

Fig. 4



Cancer progression example (a) and its associated expert-guided (b) and automatic (c) three-dimensional tumor segmentation masks. Note how the presence of errors in automatic segmentation masks may lead to conclude about a false global decrease in WBTMV (c) in an actual volume increase scenario (b). WBTMV, whole-body metabolic tumor volume.

in the tumor segmentation mask can be associated with a single tumor-related lesion [19]. Although clearly a high ΔNCC in magnitude will be likely associated with the strength of the cancer evolution, this parameter suffers from noisy behavior due to possible segmentation inaccuracies [19] and does not quantify the actual volume of the new tumor lesions. Furthermore, it will not recognize the cancer progression scenario illustrated in Fig. 1d, in which, even though a decrease in NCC (i.e. number of tumor lesions) is observed, an underlying progression condition could be present.

To overcome those limitations, another clinically relevant parameter is computed, denoted as V_N . V_N is designed to quantify the amount of new tumor volume that appeared in

the second scan with respect to the first. V_N does not include volume increases of the existing lesions; that is, it only adds up the volume of tumor lesions that appeared in new anatomical locations, thus quantifying the spread strength. Note that this indicator will effectively recognize and quantify the progression cases illustrated in Fig. 1d. The computation of V_N from the pair of time-consecutive PET tumor segmentation masks is nontrivial and described in [20]. In short, both PET scans are realigned and new tumor lesions are detected and quantified from the subtraction of the realigned segmentation masks.

Also, as has been mentioned, the appearance of tumor lesions in new organs or distant anatomical locations is considered to worsen the cancer progression condition.

To model this effect, we introduce the number of significantly new tumor lesions (nSNTL) parameter and approximate it computationally in the following manner. As the set of new tumor lesions of the tumor segmentation mask from the second scan is obtained during the computation of V_N , the remaining task is to identify and count which of those lesions (i.e. connected components in the mask) can be classified as belonging to a new organ or being sufficiently distant from the lesions of the first scan. For that, one of these two conditions must hold: either the mean Hounsfield unit value of a given lesion is significantly different ($P < 0.05$) from that of all the lesions in the first scan, or it is significantly distant ($>1\%$ of the patient's body surface area [21]) from them.

Finally, an indicator that aids in quantifying the magnitude of mixed responses is presented, denoted as A_N . A_N is designed to quantify the amount of tumor volume that increased its activity by more than 20% in the second scan relative to the amount of tumor volume in the first scan. Again, the computation of A_N from the pair of tumor segmentations masks is nondirect and described [20].

Results

Table 1 shows the performance results, in this context, of the common indicators used in clinical practice. Strong correlations are only observed in partial response, relapse, and complete response cases.

Table 2 shows the performance results obtained by combining them with the proposed alternative indicators described in the previous section. Substantial performance increases are shown in progression and mixed response cases. No strong correlation was observed in any scenario if using automatic segmentation procedures.

Finally, although the indicators presented in Table 2 are the ones that obtained the best performance results on our particular data set, very similar results were obtained when using Δ TLG instead of Δ WBMTV (79.6% correlation in progression cases, 76.7% in partial responses, and 90.5% in relapses). Also, the complete response cases were also modeled accurately by Δ WBMTV and Δ TLG (both showing 83.5% correlation).

Table 1 Pearson's correlation results of the common indicators used in clinical practice with respect to the expert-based visual ordering

Correlation (%)	Δ WBMTV	Δ SUV _{mean}	Δ SUV _{max}	Δ SUV _{peak}	Δ TLG
Progression	32.3	5.7	4.2	13.8	30.7
Partial response	76.9	48.1	59.6	43.7	73.8
Mixed response	8.3	20.0	13.3	26.7	25.0
Relapse	100	54.8	90.5	78.6	90.5
Complete response	88.5	44.0	64.8	80.8	83.5

max, maximum; SUV, standardized uptake value; TLG, total lesion glycolysis; WBMTV, whole-body metabolic tumor volume.

Δ Relative change.

Discussion

Several noteworthy conclusions can be drawn from our results. First, note that the conceptual limitations of this set of indicators described in the previous section are empirically observed in our data set. Second, the formulas of the indicators that best model the cases in the data set are a highly consistent mathematical representation of the physician global visual analysis criteria. Third, a substantial performance increase is shown in the progression and mixed response scenarios with respect to the indicators in Table 1, which demonstrates the relevance of the new set of proposed indicators in modeling real NHL evolution cases. Also, the stable results observed in the rest of the scenarios are also coherent, as the change in the overall tumor size and extension (modeled by Δ WBMTV or including the tumor activity information using TLG) is clearly the most important visual criterion in those cases. Fourth, the most difficult (e.g. the one with more discrepancies in the consensus of physicians performing the visual analysis) NHL evolution scenario to order by magnitude was the mixed response, which also showed the worse indicator correlation results. Finally, fifth, as mentioned in the Materials and methods section, current completely automatic tumor segmentation techniques are not capable of offering reliable parameters in this clinical context.

We also considered the possibility of including the time elapsed between scans as another factor in the indicator formulas, as clearly the same cancer progression or response could be considered 'stronger' if it was produced in a shorter period of time. However, we consider that the evaluation of this parameter, in conjunction with other clinical variables such as the specific treatment design of each patient, should be carried out at the oncological management level and not included in the nuclear medicine PET/CT diagnostic quantification framework. Similarly, we only considered mathematical combinations of the proposed indicators that had a sensible clinical basis, and left as future work a possible in-depth analysis on fitting parametrical statistical models to the proposed combined indicator formulas to study the possible asymmetric weight distribution of each indicator.

Finally, we consider that the incorporation of this type of quantitative parameters in nuclear medicine diagnostic frameworks could increase its overall potential. However, a large amount of future work remains. On one hand, expert-guided semiautomatic segmentation of whole-body PET scans is a highly time-consuming task and therefore is typically unfeasible in the clinical routine. In this work we showed that current completely automatic segmentation techniques are unable to provide reliable indicators in this diagnostic context, motivating the initiation of further research in this area. In contrast, the incorporation of this type of indicators at the oncological management level would require a previous in-depth

Table 2 Pearson's correlation results for the indicators that obtained the best performance results on the data set

	Indicator	Correlation (%)	Correlation (%): automatic segmentation
Progression	$\Delta\text{WBMTV} \times V_N \times \text{nSNTL}$	80.2	18.1
Partial response	$\Delta\text{WBMTV} \times (1 + \Delta\text{NCC})$	77.1	32.2
Mixed response	$A_N/\Delta\text{WBMTV}$	68.3	28.3
Relapse	ΔWBMTV	100	45.2
Complete response	ΔNCC	88.5	41.2

Note that V_N and nSNTL are only included in the product if they have a value greater than zero. Results are also reported using completely automatic segmentation techniques.

NCC, number of connected components; nSNTL, number of significantly new tumor lesions; WBMTV, whole-body metabolic tumor volume.

analysis of its exact role as well as its possible limitations in the clinical context, including its performance evaluation within alternative gold standard frameworks.

Conclusion

Addressing the need for obtaining a global continuous and observer-independent representation of the cancer evolution magnitude from a pair of whole-body PET-CT scans, in this work we proposed a set of global indicators of NHL response computed through imaging techniques that offered strong correlation results with the associated expert-based visual analysis.

Acknowledgements

The work of Frederic Sampedro is supported by the Spanish government FPU (Formación del Profesorado Universitario) doctoral grant and Spanish Project TIN2013-43478-P.

Conflicts of interest

There are no conflicts of interest.

References

- Ciernik IF, Dizendorf E, Baumert BG, Reiner B, Burger C, Davis JB, *et al.* Radiation treatment planning with an integrated positron emission and computer tomography (PET/CT): a feasibility study. *Int J Radiat Oncol Biol Phys* 2003; **57**:853–863.
- Rosenman J. Incorporating functional imaging information into radiation treatment. *Semin Radiat Oncol* 2001; **11**:83–92.
- Czernin J, Phelps ME. Positron emission tomography scanning: current and future applications. *Annu Rev Med* 2002; **53**:89–112.
- Takeda A, Yokosuka N, Ohashi T, Kunieda E, Fujii H, Aoki Y, *et al.* The maximum standardized uptake value (SUV_{max}) on FDG-PET is a strong predictor of local recurrence for localized non-small-cell lung cancer after stereotactic body radiotherapy (SBRT). *Radiother Oncol* 2011; **101**:291–297.
- Zhang H, Wroblewski K, Liao S, Kampalath R, Penney BC, Zhang Y, Pu Y. Prognostic value of metabolic tumor burden from (18)F-FDG PET in surgical patients with non-small-cell lung cancer. *Acad Radiol*, **20**:32–402013.
- Kim MH, Lee JS, Mok JH, Lee K, Kim KU, Park HK, *et al.* Metabolic burden measured by (18)F-fluorodeoxyglucose positron emission tomography/computed tomography is a prognostic factor in patients with small cell lung cancer. *Cancer Res Treat* 2014; **46**:165–171.
- 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**:1114–1119.
- Obara P, Pu Y. Prognostic value of metabolic tumor burden in lung cancer. *Chin J Cancer Res* 2013; **25**:615–622.
- Sampedro F, Domenech A, Escalera S. Obtaining quantitative global tumoral state indicators based on whole-body PET/CT scans: a breast cancer case study. *Nucl Med Commun* 2014; **35**:362–371.
- Jung SH, Yang DH, Ahn JS, Kim YK, Kim HJ, Lee JJ. Serum lactate dehydrogenase with a systemic inflammation score is useful for predicting response and survival in patients with newly diagnosed diffuse large b-cell lymphoma. *Acta Haematol* 2014; **133**:10–17.
- Hermans J, Krol AD, van Groningen K, Kluijn PM, Kluijn-Nelemans JC, Kramer MH, *et al.* International prognostic index for aggressive non-Hodgkin's lymphoma is valid for all malignancy grades. *Blood* 1995; **86**:1460–1463.
- Park JH, Yoon DH, Kim DY, Kim S, Seo S, Jeong Y, *et al.* The highest prognostic impact of LDH among international prognostic indices (IPIs): an explorative study of five IPI factors among patients with DLBCL in the era of rituximab. *Ann Hematol* 2014; **93**:1755–1764.
- Sampedro F, Escalera S, Domenech A, Carrió I. A computational framework for cancer response assessment based on oncological PET-CT scans. *Comput Biol Med* 2014; **55**:92–99.
- Hoffer M. *CT teaching manual*. Leipzig, Germany: Georg Thieme Verlag; 2000. pp. 12–13.
- 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.
- Chan WK, Mak HK, Huang B, Yeung DW, Kwong DL, Khong PL, *et al.* Nasopharyngeal carcinoma: relationship between 18F-FDG PET-CT maximum standardized uptake value, metabolic tumour volume and total lesion glycolysis and TNM classification. *Nucl Med Commun* 2010; **31**:206–210.
- Achury C, Estorch M, Domènech A, Camacho V, Flotats A, Jaller R, *et al.* Interpretation of thyroid incidentalomas in (18)F-FDG PET/CT studies. *Rev Esp Med Nucl Imagen Mol* 2014; **33**:205–209.
- Biancardi A, Segovia M. Adaptive segmentation of MR axial brain images using connected components. *Lect Notes Comput Sci* 2001; **2059**:295–302.
- Byun BH, Kong CB, Park J, Seo Y, Lim I, Choi CW, *et al.* Initial metabolic tumor volume measured by ^{18}F -FDG PET/CT can predict the outcome of osteosarcoma of the extremities. *J Nucl Med* 2013; **54**:1725–1732.
- Sampedro F, Escalera S, Domenech A, Carrió I. Automatic tumor volume segmentation in whole-body PET/CT scans: a supervised learning approach. *J Med Imaging Health Infor* 2015; **5**:1–10.
- Mosteller RD. Simplified calculation of body-surface area. *N Engl J Med* 1987; **317**:1098.