Simulated FDG-PET studies for the assessment of SUV quantification methods

J. Silva-Rodríguez a,b, P. Aguiar a,c,e, I. Domínguez-Prado d, P. Fierro d, Á. Ruibal a,c,d,e

a Grupo de Imaxe Molecular e Oncoloxía, Instituto de Investigación Sanitarias (IDIS), Travesía da Choupana S/N 15706, Santiago de Compostela, Galicia, Spain
b L2A2-USC, Facultade de Física, Universidade de Santiago de Compostela, Praza do Obradoiro, s/n, 15782, Santiago de Compostela, A Coruña, Galicia, Spain
c In Vivo Molecular Imaging Group (IMIG), Facultade de Medicina, Universidade de Santiago de Compostela, Praza do Obradoiro, s/n, 15782, Santiago de Compostela, A Coruña, Galicia, Spain
d Servicio de Medicina Nuclear, Complexo Hospitalario Universitario de Santiago de Compostela, Travesía da Choupana S/N 15706, Santiago de Compostela, Galicia, Spain
e Fundación Tejerina, Calle de José Ábasco 40, 28003 Madrid, Spain

A R T I C L E   I N F O

Article history:
Received 19 May 2014
Accepted 9 July 2014
Available online 6 August 2014

Keywords:
PET
SUV
Quantification
Monte Carlo simulation

A B S T R A C T

Aim: To study in detail the accuracy and repeatability of three commonly used methods for SUV estimation in solitary pulmonary nodules.

Material and methods: We have designed a realistic framework based on simulated FDG-PET acquisitions from an anthropomorphic activity model that included solitary pulmonary nodules (different sizes) of well-known SUV. This framework enables us to compare the SUV values obtained from the reconstructed PET images with the real SUV values. Three commonly used methods (SUV max, SUV mean and SUV 50 ) were used to estimate the tumor activity.

Results: Our results showed the tumor activity was overestimated using SUV max and clearly underestimated using SUV mean. Instead, the quantification of SUV 50 showed great agreement with the simulated tumor activity and only slight subestimation was found for very small lesions. On the other hand, SUV mean showed better performance than SUV 50 in terms of repeatability, providing variabilities below 5% for all tumor sizes and for injected doses as low as 111 MBq.

Conclusions: Our findings showed that SUV 50 provided better performance for estimating accurately tumor SUV values in pulmonary nodules, but SUV mean showed better results in terms of repeatability.

Estudios simulados de PET FDG para la evaluación de diferentes métodos de cuantificación de SUV

R E S U M E N

Objetivo: estudiar en detalle la precisión y la repetitividad de tres métodos de uso común en la estimación del SUV de nódulos pulmonares solitarios.

Material y métodos: hemos diseñado una metodología de trabajo basada en la simulación de adquisiciones de estudios PET-FDG a partir de modelos antropomórficos de actividad, que incluyen nódulos pulmonares de diferente tamaño y valor de SUV conocido. Esta metodología nos permite comparar el SUV estimado a partir de la imagen PET con el SUV teórico. La actividad del tumor fue estimada mediante tres métodos conocidos: SUV max, SUV mean y SUV 50.

Resultados: nuestros resultados muestran que, por un lado SUV max sobreestima la actividad en el tumor, mientras SUV 50 subestima el valor de actividad de manera muy significativa. En cambio, la cuantificación de SUV 50 mostró un buen acuerdo con los valores de SUV teóricos o simulados, y únicamente mostró una ligera subestimación para lesiones muy pequeñas. Por otro lado, SUV mean mostró un mejor comportamiento que SUV 50 en términos de repetitividad, proporcionando variabilidades por debajo del 5% para todos los tamaños de lesión y para dosis inyectadas tan bajas como 111 MBq.

Conclusiones: nuestros hallazgos mostraron que SUV 50 proporciona el mejor comportamiento para estimar la actividad en nódulos pulmonares, pero SUV mean mostró mejores resultados en términos de repetitividad.

Introduction

FDG-PET is a non-invasive imaging technique that visualizes the distribution of fluorodeoxyglucose (FDG) in the whole body providing functional and molecular information on tissues. It is
routinely used for staging and treatment response evaluation in oncology.\textsuperscript{1–3} Although it has been traditionally used by visual inspection, the current potential of FDG-PET relies on its capability to provide quantitative information, mainly using the standardized uptake value (SUV). This parameter enables us to objective tumor characterization, reliable differential diagnosis and earlier evaluation and monitoring of treatment response.\textsuperscript{4,5} Nevertheless, SUV has been much discussed and nowadays it is not still widely accepted.\textsuperscript{6,7} This is due to that the accuracy of the SUV is not well documented, in particular the accuracy of changes during the treatment. SUV can be affected by many different uncertainty sources, those accounting for physiological changes between patients or between the same patient, and those attributable to technical issues related to acquisition, reconstruction and quantification. In the past few years, multiple descriptions of the variability sources of the SUV have been carried out to understand some of these factors, thus enabling a more accurate interpretation of SUV values. Regarding the physiological variations, high blood glucose levels, patient motion, high uptake in brown fat have been related to lower SUV values, while increasing uptake period has been related to higher SUV values.\textsuperscript{8–11} Regarding technical issues, image noise and ROI delineation method significantly biased SUV and low spatial resolution or insufficient iterations can lead to lower SUV.\textsuperscript{12–18} On the basis of the latter results, protocols and guidelines for the standardization of FDG-PET studies in multicenter trials have been already proposed.\textsuperscript{19,20} Most of these studies were carried out by using anthropomorphic physical phantoms. These phantoms are ideal for investigating the impact of changing technical factors on the output images, enabling us for applications where patients cannot serve or should not serve. The limitations of the physical phantoms are reduced flexibility for changing shapes and volumes of the internal structures, high cost and cumbersome to use.

An alternative is the use of digital phantoms, so that simulated PET studies are generated from the projection of phantoms by using analytical or Monte Carlo simulation techniques. Thus, respiratory-induced errors in tumor SUV were recently evaluated by using the XCAT anthropomorphic digital phantom.\textsuperscript{21,22} Multiple PET studies were simulated from analytical projections of the phantom. The use of Monte Carlo simulation methods instead of analytical projectors for simulating PET studies is more realistic. Recently, our group conducted a work for studying the impact of extravasated doses on SUV quantification.\textsuperscript{23} We generated PET studies from Monte Carlo simulations of the XCAT phantom, adding multiple complex volumes to the XCAT phantom in order to mimic different extravasated doses.

In this work, we aimed at studying in detail the accuracy and repeatability of three commonly used SUV quantification methods for different tumor sizes and injected doses. For this purpose, we have designed a realistic framework based on simulated FDG-PET acquisitions from an anthropomorphic activity model that included different solitary pulmonary nodules of well-known SUVs. This framework enabled us to compare the SUV values obtained from the reconstructed PET images with the real SUV values. Our work represents the first study aimed at the evaluation of the accuracy of the SUV quantification methods in pulmonary nodules has been carried out using this methodology.

Methods

Our workflow was based on a realistic framework of multiple simulated FDG-PET studies (Fig. 1). We generated whole-body FDG-PET studies of patients with solitary pulmonary nodules of well-known SUVs. This framework enabled us to compare the SUV values obtained from the reconstructed PET images with the real SUV value (simulated value), thus making it possible to evaluate the accuracy and repeatability of SUV quantification under different conditions.

![Fig. 1. Workflow was based on a realistic framework of multiple simulated FDG-PET studies.](image)

Anthropomorphic digital phantom

We used an anthropomorphic activity and attenuation model based on the XCAT phantom\textsuperscript{22} for an average Caucasian man (175 cm and 77.5 kg). Three spherical lesions were added to the lung in order to mimic solitary pulmonary nodules with SUV of 4.3 (called simulated SUV value) and diameters of 31, 21 and 9 mm. FDG activity concentrations for normal tissues were obtained from the literature.\textsuperscript{24}

Simulated FDG-PET acquisitions

Simulated FDG-PET acquisitions of the previous phantom were generated using the SimSET Monte Carlo package,\textsuperscript{24} which includes simulation of all the relevant processes for energies of interest in nuclear medicine (bellow 1 MeV). Scanner geometry was based on a previous model of SimSET for the General Electric (GE) Advance Nxi PET scanner.\textsuperscript{25} Multiple simulated FDG-PET acquisitions were

![Fig. 2. Coronal views of simulated FDG-PET studies of patients with solitary pulmonary nodules for different injected FDG doses ranging from 9 mCi to 1 mCi.](image)
generated by using different injected doses (9, 6, 3, 2 and 1 mCi) and ten repetitions for each dose for statistical purposes.

**Tomographic reconstruction**

Reconstruction of the simulated FDG-PET acquisitions was performed by using the iterative ordered subset expectation maximization algorithm implemented in the STIR library.

**Quantification and evaluation**

Three commonly used methods of ROI definition (maximum pixel value, manual placement and threshold based) were used to calculate SUV (SUV\textsubscript{max}, SUV\textsubscript{mean} and SUV\textsubscript{50}, respectively). They were obtained from the different noise realizations, injected doses and tumor sizes. Note that for manual ROI delineation (SUV\textsubscript{med}) the attenuation map was used as anatomical reference for ROI delineation. The obtained SUV values were averaged over noise realizations and then compared to the simulated SUV values, for different tumor sizes and injected doses. The repeatability of the SUV measurements was evaluated from ten noise realizations as the standard deviation (STD).

**Results**

Fig. 2 shows coronal views of simulated FDG-PET studies of patients with solitary pulmonary nodules for different injected FDG doses. Two lesions can be: one in left lung corresponding to the

![Graphs showing SUV values for different lesion sizes and injected doses]
Simulated lesion of 21 mm and other lesion in right lung corresponding to the simulated lesion of 31 mm.

Fig. 3 shows the averaged SUV (SUV$_{\text{max}}$, SUV$_{\text{mean}}$, and SUV$_{50}$) for all noise realizations, for different tumor sizes and injected doses. The simulated SUV value is represented by the blue line for comparison. In relation to the quantification of SUV$_{\text{max}}$, it can be derived that SUV$_{\text{max}}$ values were overestimated with respect to the simulated values for lesion sizes of 31 mm and 21 mm, but not for 9 mm. Moreover, the overestimation of SUV$_{\text{max}}$ increased when injected dose was reduced. Instead, the quantification of SUV$_{\text{mean}}$ showed a subestimation with respect to the simulated SUV values for all lesion sizes, without significant changes with the injected dose. Finally, the quantification of SUV$_{50}$ showed great agreement with the simulated SUV values for lesion sizes of 31 mm and 21 mm, and only a slight subestimation should be noted for lesion size of 9 mm. Again, no significant changes were found with the injected dose.

Fig. 4 shows the repeatability of the SUV measurements in terms of STD. As expected, the uncertainty of SUV increased when the injected dose decreased. Furthermore, it can be derived that SUV$_{\text{max}}$ provided much larger uncertainty than SUV$_{\text{mean}}$ and SUV$_{50}$.

**Discussion**

Our workflow provided realistic simulated FDG-PET studies of patients with solitary pulmonary nodules, with different lesion sizes and well-known SUVs. This enabled us to evaluate the accuracy and variability of three commonly used methods of SUV
calculation (SUV\textsubscript{max}, SUV\textsubscript{mean} and SUV\textsubscript{50}) under different conditions. Obviously, it would not be feasible to conduct a similar study with patients, since the real value of SUV would not be known. Below plausible explanations for the results are given.

Firstly, we will discuss the results obtained from the most commonly used method for tumor quantification, SUV\textsubscript{max}. The overestimation associated with SUV\textsubscript{max} can be explained by the statistical noise level. Obviously, this overestimation increased at low injected doses, as the noise also increases. Particularly interesting is the case of the smallest lesion, where the obtained SUV\textsubscript{max} values were similar to the simulated SUV values.

Regarding to the tumor quantification based on SUV\textsubscript{mean} it provided significant subestimations with respect to the simulated values, for all lesion sizes and without significant changes with the injected dose. This can be explained by the partial volume effect, due to which part of the tumor activity is outside the delineated tumor ROI. As expected, this effect was more important for the smallest lesion.

The tumor quantification based on SUV\textsubscript{50} provided the results that best matched the simulated values. Excellent agreement was found for lesion sizes of 31 mm and 21 mm, and only a slight subestimation was found for the smallest lesion. It can be explained because the tumor quantification based on SUV\textsubscript{50} is not overly affected either by the partial volume effect or statistical noise.

Our findings showed that accurate estimations of SUV are very complicated, especially for small lesions. Both tumor quantifications based on SUV\textsubscript{max} and SUV\textsubscript{mean} represented poor approaches for estimating accurate SUV values compared to the simulated values and only SUV\textsubscript{50} can be considered as an acceptable method for estimating SUV when lesions are large enough. Nevertheless, in most cases, SUV estimations matching the simulated SUV values are not required in clinical routine, where SUV estimations are compared with estimated values from other FDG-PET studies, commonly the same patient. Therefore, the essential issue is the repeatability of the SUV estimation, defined as the statistical variability of multiple SUV estimations of the same tumor and the same patient. The variability of SUV\textsubscript{max} and SUV\textsubscript{50} estimations was ranged between 2% and 8% for injected doses greater than 3 mCi and it increased significantly for doses below this value. Instead, the variability of SUV\textsubscript{mean} was below 5% for all lesion sizes and injected doses greater than 3 mCi. In terms of repeatability, the tumor quantification based on SUV\textsubscript{mean} showed a better performance than SUV\textsubscript{50}, in particular for small lesions, where the ROI used for SUV\textsubscript{50} estimation has a very small number of pixels, making SUV\textsubscript{50} very similar to SUV\textsubscript{max}.

Overall, SUV\textsubscript{50} showed better performance for estimating accurately tumor SUV values. These results were in agreement with recently reported data that showed SUV\textsubscript{50} provides accurate quantifications.\textsuperscript{12} Furthermore, it has to be mentioned that SUV\textsubscript{50} can be automatically segmented from PET images, while an anatomical reference is required for ROI delineation when using SUV\textsubscript{mean}. On the other hand, SUV\textsubscript{mean} showed better results in terms of repeatability. This is because ROI delineation for SUV\textsubscript{50} depends on the value of the single-pixel maximum. Regarding to this, it should be mentioned that ROI for SUV\textsubscript{mean} was accurately segmented from the attenuation map, and therefore inaccuracies related to manual delineation were not included. Another remarkable observation is that a reduction of dose in the range from 9 to 3 mCi would not affect significantly the results of quantification when SUV\textsubscript{mean} and SUV\textsubscript{50} estimations were used. This result is particularly interesting because it was demonstrated for a very old scanner with low sensitivity. Further studies in modern scanners could demonstrate that the dose can be even lower. Here it has to be mentioned that the use of low doses can be feasible for quantification purposes but it could dramatically reduce the detectability of new lesions, because the obtained images had low statistics.

Conclusions

We have carried out an evaluation study of the accuracy and repeatability of SUV estimations obtained from FDG-PET studies in patients with solitary pulmonary nodules. Our findings showed that SUV\textsubscript{50} showed better performance for estimating accurately tumor SUV values, but SUV\textsubscript{mean} showed better results in terms of repeatability.

Financial support

This work was supported by grants 10CSA918001PR, PI11/01806 and P. Aguiar was awarded a public fellowship from Xunta de Galicia POS-A/2013/001.

Conflict of interest

No potential conflicts of interest to this article have been reported.

References


