SHORT COMMUNICATIONS

Inconsistency of Lesion Quantitative Assessment in 2D SUV and 3D SUV Quantification Techniques for [^{18}F]-FDG PET/CT: A Phantom Study

Muhammad Hafiz Hanafi[1], Noramaliza Mohd Noor[2,3], Muhammad Hishar Hassan[1]

[1] Centre for Diagnostic Nuclear Imaging, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2] Department of Imaging, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[3] Department of Radiology, Hospital of Teaching Universiti Putra Malaysia, Persiaran Mardi- UPM, 43400 Serdang, Selangor, Malaysia

ABSTRACT

This study was performed to assess the inconsistency of lesion quantification in standardised uptake value (SUV) [^{18}F]-FDG between Ellipse (2-Dimensional) and Ellipsoid (3-Dimensional) quantification techniques by using PET/CT image quality phantom. Reconstructed images of PET/CT ACR phantom was used to assess the quantification of SUV (SUV_{max}, SUV_{avg} and SUV_{min}) on selected regions of interest. Statistical analysis of paired t-test was performed to compare the lesion quantification in SUV [^{18}F]-FDG between 2D and 3D techniques. The quantification techniques were consistently similar of assessment between 2D SUV_{max} and 3D SUV_{max} at 12mm of ROI lesion with [(0.00 ± 0.02), t(29)=-0.48, p>0.05]. However, the rest of quantification techniques of 2D SUV_{max}, 3D SUV_{max}, 2D SUV_{avg}, 3D SUV_{avg}, 2D SUV_{min} and 3D SUV_{min}, results shown significant inconsistency since the p<0.05. This phantom study has proven that there were inconsistency of lesion quantitative assessment in 2D SUV and 3D SUV quantification techniques for [^{18}F]-FDG PET/CT.

Keywords: Quantification, ^{18}F-FDG, SUV_{max}, SUV_{avg}, SUV_{min}, PET/CT

Corresponding Author:
Muhammad Hafiz Hanafi, MSc
Email: muhammadhafizhanafi90@gmail.com
Tel: +603-9769 2332

INTRODUCTION

Positron emission tomography/computed tomography (PET/CT) for cancer imaging using fluorine-18 fluorodeoxyglucose [^{18}F]-FDG is widely used in oncology cases (1). [^{18}F]-FDG acts as a useful molecular marker for many cancer cases such as follicular lymphoma (2), breast cancer (3), pharynx (4) and others. Thus, the role of PET/CT using [^{18}F]-FDG is getting significant in oncology cases especially in Malaysia (5). Nevertheless, the use of PET/CT in Malaysia is mandatory to comply the quality control (QC) which is verified by an approved medical physicist before subjecting PET/CT examination to the patient as clinical procedure (6). QC process of PET/CT used activated phantom to verify the quantification accuracy of the scanner system. Quantification of technical procedure (using phantom) or clinical procedure (patient) depends on quantification value called standardised uptake value (SUV). Quantification of SUV on phantom can be used as reference of SUV standard because phantom has fixed vials compare to human changed metabolism.

SUV is a method for measuring the amount rate of [^{18}F]-FDG accumulation in tumour tissue. PET/CT scanners are built to measure in vivo radioactivity concentration (Bq/mL). Quantification of SUV acts as an important parameter of semi-quantitative measurement of glucose metabolism to evaluate uptake value through PET/CT images (7–9). SUV is obtained automatically on most modern-day PET/CT scanners and measures normalized radioactivity concentration as follows:

$$SUV = \frac{\text{activity concentration in tissue}}{\text{injected activity/ body size}}$$

SUV_{avg} and SUV_{max} of a target lesion or a region of interest are the two most common ways of reporting SUV (10). However, PET/CT SUV quantification depends on the technical errors such as incorrect units, varying times from injection to scanning, dose infiltration, and image processing such as iterations or filters (7,9–11). Researchers mutually concurred the significance of quantification variation factors are related to the amount of injected [^{18}F]-FDG and the patient size (7). Due to

inconsistent and non-optimised image analysis, a group of European researchers has reported the technical error as a key factor to variation in the measured SUV (9, 11). In such cases, SUV harmonisation would be needed to minimise inter-scanner issue. Current studies reported by fellow researchers served as an on-going effort towards standardisation of FDG-PET imaging to allow comparison variability in quantitative SUV (1, 12). Therefore, this study was performed to investigate the inconsistency quantification analysis using 2D and 3D quantification techniques using phantom accredited by American College of Radiology.

MATERIALS AND METHODS

PET phantom Flangeless Esser, USA was used in this investigation as image quality phantom. The phantom vials (25mm, 16mm, 12mm and 8mm) were filled and activated with 0.35 millicuries (mCi) [^{18}F]-FDG. Next, the phantom container was filled and activated with 0.83 mCi [^{18}F]-FDG to serve as background radioactivity. PET/CT scanner (Biograph 64, Siemens, USA) was used to scan the phantom. The phantom was carefully aligned so that it was parallel to the patient's table (Fig. 1). For the acquisition, the whole body protocol was used with the same settings that has been used for routine clinical procedures at this institution. The phantom was scanned using these acquisition parameters; time per bed position = 2 minutes, number of bed position = 1, matrix size = 256 pixels, zoom factor = 1, reconstruction type = True X, iterations = 3, subsets = 21, processing filter = Gaussian, FWHM = 4 mm and slice thickness = 5 mm. PET/CT images then, were imported to Leonardo SIEMENS workstation. DICOM images were processed.

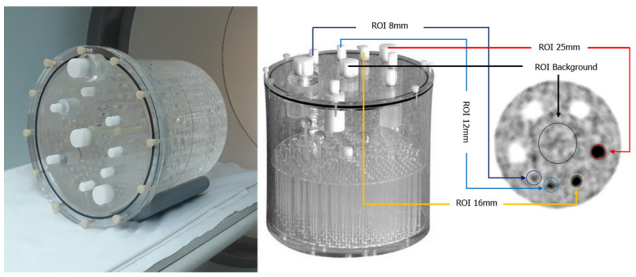

Figure 1: ACR image quality phantom was scanned using PET/CT Siemens Biograph 64 scanner. The phantom was activated with [^{18}F]-FDG at selected region-of-interest (ROI) for quantification assessment of 25 mm, 16 mm, 12 mm, 8 mm and background.

The DICOM headers provided most parameters needed to process images and to compute SUV. Leonardo SIEMENS; SYNGO syngo MMWP VE31A, syngo VE32B WinNT 5.2, Service Pack 2, COEM VE10D 64Bit was used for the quantification of [^{18}F]-FDG PET/CT images. The standard ROIs were placed on the central slice through the vials to obtain SUV measurements. A qualified medical physicist has conducted a quantitative analysis of the SUV supported by a radiographer. A reader for multiple readings. The quantification of SUV were performed with [^{18}F]-FDG using Ellipse (2-dimensional) and Ellipsoid (3-dimensional) quantification techniques.

Statistical analysis was conducted using IBM Statistical Package for Social Science (SPSS) Statistic for Windows, latest version 2020 (IBM Corp., Armonk, N.Y.,USA). Paired t-test was performed to compare the consistency lesion quantification in SUV ^{18}F-FDG between 2D and 3D techniques.

RESULTS AND DISCUSSION

Comparison of quantitative assessment between SUV_{max}, SUV_{avg} and SUV_{min} in phantom lesion

Data was collected from the sample of 30 repeated SUV measurements of phantom lesion which were 2D SUV_{max}, 3D SUV_{max}, 2D SUV_{avg}, 3D SUV_{avg}, 2D SUV_{min} and 3D SUV_{min} as illustrated in Fig. 2. The result of SUV measurements for 2D and 3D were presented in Fig. 3. We demonstrated the result into three separate line graphs to show the differences of quantitative assessment between 2D SUV and 3D SUV techniques (SUV_{max}, SUV_{avg} and SUV_{min}).

We described the result of SUV_{max} as presented in Fig. 3a. There was similarity reading and pattern between 2D SUV_{max} and 3D SUV_{max} as shown at 12mm of ROI lesion. Both of SUV_{max} mean was 3.11 ± 0.01. However, there was an inconsistent pattern between 2D SUV_{max} (2.88 ± 0.01) and 3D SUV_{max} (3.74 ± 0.01) at 16mm of ROI lesion. Additionally at 25mm of ROI lesion, 2D SUV_{max} and 3D SUV_{max} were 3.07 ± 0.01 and 4.00 ± 0.01 respectively. This result highlights the difference of quantitative assessment value between 2D SUV_{max} and 3D SUV_{max} techniques if ROI >12 mm.

For SUV_{avg}, the graph shown there was close to homogeneous pattern between 2D SUV_{avg} and 3D

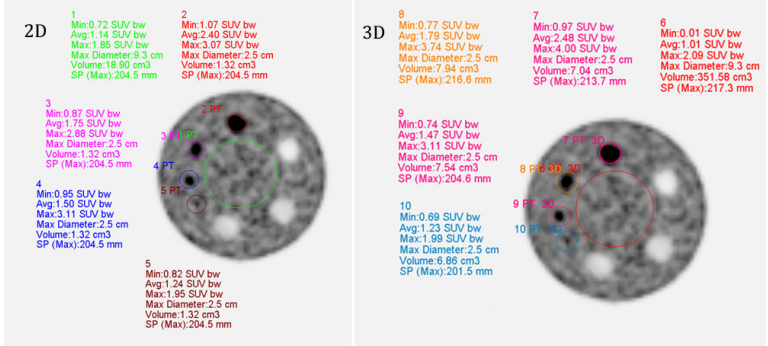

Figure 2: Transversal images of phantom. The quantitative assessment using Ellipse (2D) quantification technique and the quantitative assessment using Ellipsoid (3D) quantification technique.

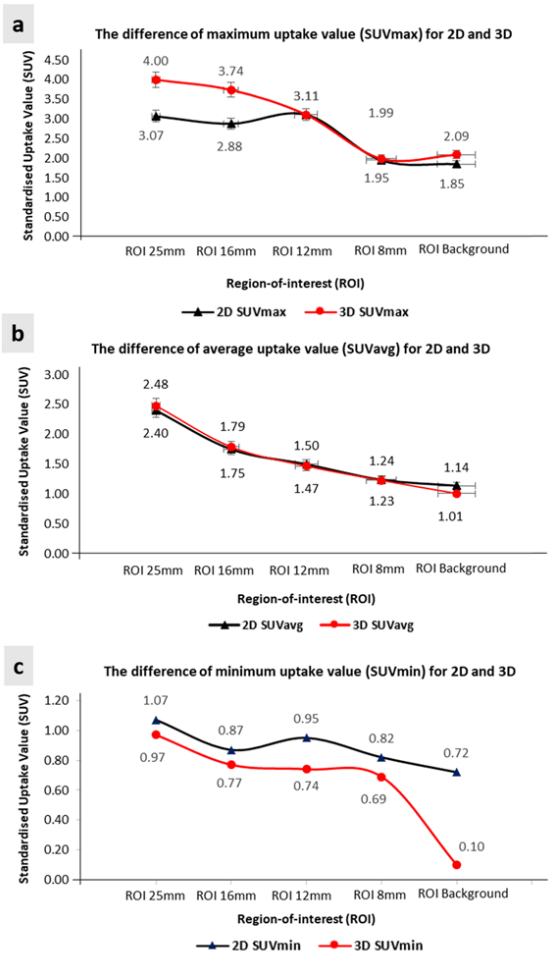

Figure 3: The differences of uptake value for 2D and 3D quantification techniques which are (a) SUV_{max}, (b) SUV_{avg} and (c) SUV_{min}

SUV_{avg} quantitative assessment as demonstrated in Fig. 3b. At 25mm of ROI lesion, 2D SUV_{avg} and 3D SUV_{avg} were 2.41 ± 0.02 and 2.48 ± 0.01 respectively while at background of ROI lesion, 2D SUV_{avg} and 3D SUV_{avg} were 1.14 ± 0.01 and 1.01 ± 0.02 respectively.

A non-identical graph pattern of quantitative assessment between 2D SUV_{min} and 3D SUV_{min} was also discovered. The highest difference shown at background of ROI lesion with the mean of 2D SUV_{min} and SUV_{min} were 0.72 ± 0.01 and 0.10 ± 0.01 respectively as presented in Fig. 3c.

Based on the pattern of displayed in the illustrated graphs, 2D SUV_{avg} and 3D SUV_{avg} was found to be the more favourable quantification techniques as they are the most stable of all ROIs compared to quantitative assessments of SUV_{min} and SUV_{max}. However, the statistical test should be done beforehand to determine whether SUV_{avg} is indeed the best choice of all options.

Inconsistency of Quantitative Assessment between 2D and 3D quantification techniques in parameter of SUV_{max}, SUV_{avg} and SUV_{min}

A paired t-test was run on the sample of 30 repeated SUV measurements (SUV_{max}, SUV_{avg} and SUV_{min}) of phantom lesion to determine whether there was a statistically significant difference of lesion quantification in SUV [^{18}F]-FDG between 2D and 3D techniques. The statistic test was run on the Pair 1 to Pair 15 as illustrated in Table I.

Table I: Result of statistical paired t-test analysis of quantitative assessment in 2D and 3D SUV quantification techniques using parameter of SUV_{avg}, SUV_{max} and SUV_{min}

SUV Assessment Parameter	Pair	Region of Interest	2D Mean ± SD	3D Mean ± SD	p-value (paired t-test)
SUVavg	Pair 1	25 mm	2.41 ± 0.02	2.48 ± 0.01	p<0.05 (0.00)
	Pair 2	16 mm	1.75 ± 0.02	1.79 ± 0.01	p<0.05 (0.00)
	Pair 3	12 mm	1.50 ± 0.01	1.47 ± 0.01	p<0.05 (0.00)
	Pair 4	8 mm	1.24 ± 0.01	1.23 ± 0.01	p<0.05 (0.02)
	Pair 5	Background	1.14 ± 0.01	1.01 ± 0.02	p<0.05 (0.00)
SUVmax	Pair 6	25 mm	3.07 ± 0.01	4.00 ± 0.01	p<0.05 (0.00)
	Pair 7	16 mm	2.88 ± 0.01	3.74 ± 0.01	p<0.05 (0.00)
	Pair 8	12 mm	3.11 ± 0.01	3.11 ± 0.01	p>0.05 (0.63)
	Pair 9	8 mm	1.95 ± 0.01	1.99 ± 0.02	p<0.05 (0.00)
	Pair 10	Background	1.85 ± 0.01	2.09 ± 0.01	p<0.05 (0.00)
SUVmin	Pair 11	25 mm	1.07 ± 0.01	0.97 ± 0.01	p<0.05 (0.00)
	Pair 12	16 mm	0.87 ± 0.01	0.77 ± 0.01	p<0.05 (0.00)
	Pair 13	12 mm	0.95 ± 0.01	0.74 ± 0.01	p<0.05 (0.00)
	Pair 14	8 mm	0.82 ± 0.02	0.69 ± 0.02	p<0.05 (0.00)
	Pair 15	Background	0.72 ± 0.01	0.10 ± 0.01	p<0.05 (0.00)

For SUV_{avg}, quantitative assessment between 2D SUV_{avg} and 3D SUV_{avg} was consistently different for Pair 1 to Pair 5; result Pair 1 [(-0.08 ± 0.02), t(29)=-20.68, p<0.05]; Pair 2 [(-0.04 ± 0.02), t(29)=-10.37, p<0.05]; Pair 3 [(0.03 ± 0.02), t(29)=-10.02, p<0.05]; Pair 4[(0.01 ± 0.02), t(29)=2.51, p<0.05]; Pair 5 [(0.13 ± 0.02), t(29)=43.46, p<0.05].

Additionally for SUV_{max}, quantitative assessment between 2D SUV_{max} and 3D SUV_{max} was inconsistent for Pair 6 [(-0.94 ± 0.02), t(29)=-285.36, p<0.05], Pair 7 [(-0.86 ± 0.02), t(29)=-219.41, p<0.05], Pair 9 [(-0.04 ± 0.02), t(29)=-9.05, p<0.05] and Pair 10 [(-0.24 ± 0.02), t(29)=--62.42, p<0.05]. The only exception of consistency occurred for Pair 8. Pair 8 shown a definitive constant result between 2D SUV_{max} and 3D SUV_{max} at 12mm of ROI lesion with only [(0.00 ± 0.02), t(29)=-0.48, p>0.05].

Next for SUV_{min}, quantitative assessment between 2D SUV_{min} and 3D SUV_{min} was totally difference for all pairs in this group since the p-values were significant <0.05. Pair 11 [0.10 ± 0.02), t(29)=32.57, p<0.05], Pair 12 [(0.10 ± 0.02), t(29)=29.21, p<0.05], Pair 13 [(0.20 ± 0.02), t(29)=63.01, p<0.05], Pair 14 [(0.13 ± 0.02), t(29)=34.18, p<0.05] and Pair 15 [(0.63 ± 0.02), t(29)=197.73, p<0.05].

These findings suggested SUV_{max} at 12mm of ROI lesion will provide a consistent quantitative assessment between 2D SUV_{max} and 3D SUV_{max} as simulated

quantification technique for [^{18}F]-FDG PET/CT.

We established our study on PET/CT acquisitions and reconstruction of an Esser Flangeless Deluxe PET Phantom™ body phantom, which is a standard in PET image quality quantitative assessments. The phantom was prepared with attentively by following PET Phantom Testing that has been approved by the ACR Committee on Nuclear Medicine Accreditation. The phantom was widely used in lesion detectability of [^{18}F]-FDG PET/CT as well. The Leonardo workstation succeeded in determining, viewing and processing the phantom images to run hot lesion quantitative analysis with lowest noise errors and possible minimum artefacts.

Current phantom study has discovered there were significant difference and inconsistency of lesion quantitative assessment for [^{18}F]-FDG PET/CT in 2D SUV and 3D SUV quantification techniques for parameter of SUV_{max}, SUV_{avg} and SUV_{min}. Although positive pattern graph was obtained at quantitative assessment of 2D SUV_{avg} and 3D SUV_{avg} displayed in Figure 3(c), statistical analysis shown 2D SUV_{avg} and 3D SUV_{avg} failed to show consistency between them with the exception of Pair 8. Pair 8 with 2D SUV_{max} and 3D SUV_{max} at 25mm of ROI lesion shown significant similarity with 2.41 ± 0.02 and 2.48 ± 0.01 respectively at [(0.00 ± 0.02), t(29)=-0.48, p>0.05]. There rest of pairs have shown that there was significant inconsistency in lesion quantitative assessment. Quantitative assessment of 2D SUV_{max} and 3D SUV_{max} at 25mm of ROI lesion were significantly similar either in graph or statistical analysis.

Nevertheless, comparing the current results to those of previous studies is unavoidable. Previous researches had suggested SUV_{max} as a routinely used metric to report lesion aggressiveness due to its reproducibility. Moreover, SUV_{max} is proportional to SUV_{avg}. Hence, accurately represents the entire lesion (13). In prior clinical study, SUV_{max} may reflect tumour metabolism with more certainty compared to SUV_{avg} (14). We do acknowledge and pay homage to previous studies that shared the investigation of qualitative and quantitative comparison using phantom and various radiopharmaceuticals (13).

Quantification of SUV generally used in previous studies were SUV_{max}, SUV_{avg}, SUVmean and SUVpeak (12). However, contrary to the findings of quantification of SUV, we did not find any lesion quantitative assessment using SUV_{min} either in previous phantom or clinical studies eventhough SUV_{min} has shown inconsistency of quantitative assessment between 2D SUV_{min} and 3D SUV_{min} quantification technique in this study. We does believe however, we have obtained concrete results with this simple method.

The present study has some limitations. We acknowledged data presented was just a small data collections of SUV quantitative assessment (n=30). However, combination of all data collection comes out to approximately n=450 (3 different SUV x 5 different ROI x 30 quantitative assessment). Furthermore, this study focuses on single reader with multiple readings. We cannot fulfil the inter-reader agreement measurement of image quality score as was required for the Cohen's kappa score. Moreover, the procedure of inter-reader agreement has been studied to minimize the variation of SUV among the readers (15). Perhaps, the agreement between the readers would be an exciting endeavour for our next study.

Another limitation we faced was null comparison between current phantom study and clinical study. In clinical study, we can validate the homogenous phantom with heterogeneous organs. We recommended the improvements to the phantom design to make it more heterogeneous to mimic the actual clinical conditions. For example, this is the case for development of 3D organ phantom which is homogenous with real human organs (16). Through this suggestion of improvement of future study, we could suggest all PET/CT cases reporters to harmonise and standardise the quantification parameter by justifying 2D SUV or 3D SUV quantification techniques. This will secure the intra reports of patient's tumour assessment in each PET/CT institution.

Overall, further investigation is needed to study the SUV [^{18}F]-FDG comparison and harmonisation quantitative assessments. We believe further investigation will lead to better consistency and standardised lesion quantification of [^{18}F]-FDG PET/CT case. In the future, this could open opportunities for new pathway of medical physics and molecular imaging field studies in Malaysia.

CONCLUSION

In conclusion, the phantom study has proven that there were inconsistence of lesion quantitative assessment in 2D SUV and 3D SUV quantification techniques for [18F]-FDG PET/CT using parameter of SUV_{max}, SUV_{avg} and SUV_{min}. Future investigations are necessary to further support the discussion we have drawn from this study as this is a desirable step for future work regarding niche of phantom study.

ACKNOWLEDGEMENTS

We would like to thank Centre for Diagnostic Nuclear Imaging, Universiti Putra Malaysia for providing the research facility and the use of the PET/CT scanner.

REFERENCES

1. Koopman D, Jager PL, Slump CH, Knollema S, van Dalen JA. SUV variability in EARL-accredited conventional and digital PET. EJNMMI Res. 2019;9(1):106.
2. Yuasa M, Kaji D, Kageyama K, Taya Y, Takagi

S, Yamamoto H, et al. Clinical Significance of Uptake Value on F18-FDG PET/CT and Histological Grade in 164 Patients with Follicular Lymphoma Including Transformation - a Single Center Retrospective Study. Blood. 2019 Nov 13;134(Supplement_1):1529.
3. Paydary K, Seraj SM, Zadeh MZ, Emamzadehfard S, Shamchi SP, Gholami S, et al. The evolving role of FDG-PET/CT in the diagnosis, staging, and treatment of breast cancer. Mol Imaging Biol. 2019;21(1):1–10.
4. Pietrzak AK, Marszalek A, Kazmierska J, Kunikowska J, Golusinski P, Suchorska WM, et al. Sequential delayed [18F]FDG PET/CT examinations in the pharynx. Sci Rep. 2020;10(1):2910.
5. Hanafi MH, Noor NM, Rana S, Saad FFA. Standardisation Techniques of Independent PET/CT Modalities Utilising PET SUV_{max} as a Potential Conversion Marker. Transylvanian Rev. 2017;1(7).
6. Ministry of Health Malaysia. Revised Technical Quality Control Protocol Handbook : Positron Emission Tomography/ Computed Tomography (PET/CT) Systems. Bahagian Kawalselia Radiasi Perubatan; 2018. 44 p.
7. Brendle C, Kupferschläger J, Nikolaou K, la Fougère C, Gatidis S, Pfannenberg C. Is the standard uptake value (SUV) appropriate for quantification in clinical PET imaging?–Variability induced by different SUV measurements and varying reconstruction methods. Eur J Radiol. 2015;84(1):158–62.
8. Ross S, Maguire S, Gilligan P, O'Connell M, Adams A, McCoubrey B. A phantom based study on the effects of contrast agents and kilo-voltage and on standardised uptake value in PET/CT. Phys Medica Eur J Med Phys. 2019;67:203.
9. 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(2):328–54.
10. Plaxton N, Moncayo V, Barron B, Halkar R. Factors that influence standard uptake values in FDG PET/CT. J Nucl Med. 2014;55(supplement 1):1356.
11. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging. 2010;37(1):181.
12. Shimada N, Akamatsu G, Matsumoto K, Daisaki H, Suzuki K, Oda K, et al. A multi-center phantom study towards harmonization of FDG-PET: variability in maximum and peak SUV in relation to image noise. J Nucl Med. 2020;61(supplement 1):1396.
13. Lantos J, Mittra ES, Levin CS, Iagaru A. Standard OSEM vs. regularized PET image reconstruction: qualitative and quantitative comparison using phantom data and various clinical radiopharmaceuticals. Am J Nucl Med Mol Imaging. 2018;8(2):110.
14. Kajáry K, Tokés T, Dank M, Kulka J, Szak6ll Jr S, Lengyel Z. Correlation of the value of 18F-FDG uptake, described by SUV_{max}, SUV_{avg}, metabolic tumour volume and total lesion glycolysis, to clinicopathological prognostic factors and biological subtypes in breast cancer. Nucl Med Commun. 2015;36(1):28–37.
15. Azmi NHM, Suppiah S, Liong CW, Noor NM, Said SM, Hanafi MH, et al. Reliability of standardized uptake value normalized to lean body mass using the liver as a reference organ, in contrast-enhanced 18F-FDG PET/CT imaging. Radiat Phys Chem. 2018;147:35–9.
16. Filippou V, Tsoumpas C. Recent advances on the development of phantoms using 3D printing for imaging with CT, MRI, PET, SPECT, and ultrasound. Med Phys. 2018;45(9):e740–60.