Physica Medica 32 (2016) 651–656

Contents lists available at ScienceDirect

Physica Medica

journal homepage: http://www.physicamedica.com

Original paper

The ^{68}Ge phantom-based FDG-PET site qualification program for clinical trials adopted by FIL (Italian Foundation on Lymphoma)

Stephane Chauvie [a,*], Fabrizio Bergesio [a], Federica Fioroni [b], Marco Brambilla [c], Alberto Biggi [a], Annibale Versari [b], Luca Guerra [d], Giovanni Storto [e], Pellegrino Musto [e], Stefano Luminari [f], Maria G. Cabras [g], Monica Balzarotti [h], Luigi Rigacci [i], Maurizio Martelli [j], Umberto Vitolo [k], Massimo Federico [f], Andrea Gallamini [l]

[a] Santa Croce e Carle Hospital, Cuneo, Italy
[b] Arcispedale Santa Maria Nuova-IRCCS, Reggio Emilia, Italy
[c] University Hospital Maggiore della Carità, Novara, Italy
[d] San Gerardo Hospital, Monza, Italy
[e] IRCCS-CROB, Referral Cancer Center of Basilicata, Rionero in Vulture (Pz), Italy
[f] Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy
[g] Businco Hospital, Cagliari, Italy
[h] Humanitas Cancer Centre, Rozzano, Italy
[i] AOU Careggi Hematology Department, Firenze, Italy
[j] University "Sapienza" Roma, Department of Cellular Biotechnologies and Hematology, Italy
[k] Città della Salute Hospital, Torino, Italy
[l] A. Lacassagne Cancer Center, Nice, France

ARTICLE INFO

Article history:
Received 22 September 2015
Received in Revised form 10 March 2016
Accepted 13 April 2016
Available online 28 April 2016

Keywords:
Quantitative imaging
Positron emission tomography
Clinical trials
Lymphoma

ABSTRACT

Purpose: The quantitative assessment of Positron Emission Tomography (PET) scans using standardized uptake value and derived parameters proved to be superior to traditional qualitative assessment in several retrospective or mono-centric prospective reports. Since different scanners give different quantitative readings, a program for clinical trial qualification (CTQ) is mandatory to guarantee a reliable and reproducible use of quantitative PET in prospective multi-centre clinical trials and in every-day clinical life.
Methods: We set up, under the auspices of Italian Foundation on Lymphoma (FIL), a CTQ program consisting of the PET/CT scan acquisition and analysis of ^{18}F and ^{68}Ge NEMA/IEC image quality phantoms for the reduction of inter-scanner variability. Variability was estimated on background activity concentration (BAC) and sphere to background ratio (SBR).
Results: The use of a ^{68}Ge phantom allowed reducing the inter-scanner variability among different scanners from 74.0% to 20.5% in BAC and from 63.3% to 17.4% in SBR compared to using the ^{18}F phantom. The CTQ criteria were fulfilled at first round in 100% and 28% of PET scanners with ^{68}Ge and ^{18}F respectively.
Conclusions: The ^{68}Ge phantom proved a reliable tool for PET scanner qualification, able to significantly reduce the potential sources of error while increasing the reproducibility of PET derived quantitative parameter measurement.
© 2016 Associazione Italiana di Fisica Medica. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Positron Emission Tomography (PET), combined with Computed Tomography (PET/CT), measures the changes in concentrations of tracer uptake in diseased tissue [1]. PET/CT-based evaluation of response to cancer treatment has proved a reliable outcome predictor in several tumours [2–4]. Quantitative metrics for PET/CT scan (Q-PET) interpretation by Standardized Uptake Value (SUV) have been recently shown to improve the prognostic role of PET both at baseline and during treatment [3,5], but only few prospective clinical trials are underway using these metrics for PET/CT scan interpretation[6]. Besides more sophisticated approaches [7] and taking in account its intrinsic limitations [8], SUV is being universally used as a measure of tumour viability by itself or mixed up with more complex indexes [9]. However, typical SUV variability of 40–90% in SUV measurements were

* Corresponding author at: Medical Physics Unit, Santa Croce e Carle Hospital, via Coppino 26 12100, Italy.
E-mail address: chauvie.s@ospedale.cuneo.it (S. Chauvie).

http://dx.doi.org/10.1016/j.ejmp.2016.04.004
1120-1797/© 2016 Associazione Italiana di Fisica Medica. Published by Elsevier Ltd. All rights reserved.

observed [10,11], and only major changes in SUV in response to treatment (⩾30–40%) could be detected [12].

Despite the publication of guidelines for tumour PET/CT imaging [13,14], the lack of standardization [15,16] has hampered in the past the use of SUV as a biomarker in clinical trials. Now, thanks to a better knowledge of the factors affecting SUV measurements [17], guidelines for patient scanning and PET/CT image acquisition [18] are recommended to improve data quality and reproducibility [8,18].

Nonetheless, an inter-scanner variability of SUV measurement up to 100% in a non-harmonized environment [19] and of 25% in a multicenter clinical trial [20], just as consequence of the intrinsic variability of the instrument, is still observed. Hence, a thorough cross-calibration of PET/CT scanners and ancillary instrumentation is the first condition to achieve an accuracy in tracer uptake measurement below 10% [20–27].

Several programs for the cross-calibration of PET scanners have been carried out in the recent years to reduce the variability among PET/CT scanners using ^{18}F phantoms: the European Association of Nuclear Medicine (EANM) accreditation program for site of excellence carried out by EARL Ltd (Wien, Austria) [21,22], the UK PET Clinical Trial Network (CTN) [23], the American College of Radiology Imaging Network (ACRIN) [24], the CTN of Society of Nuclear Medicine and Molecular Imaging (SNMMI) [25,26] and the JSCT NHL10 trial [27] in Japan.

The idea to measure the instrumentation factors affecting variance and bias of quantifying tracer uptake using a long half-life isotope phantom has been recently addressed in several publications. Fahey et al. [19] used the standard American College or Radiology (ACR) phantom with the four smaller cylinders filled with a ^{68}Ge epoxy matrix in a ^{18}F-filled background and achieve a coefficient of variation among PET/CT scanners of 9.9–11.3%. Doot et al. [28] used a modified NEMA/IEC Image Quality (IQ) phantom, without lung insert, in which both the six spheres and the background were filled with a ^{68}Ge epoxy matrix and achieved a coefficient of variation among PET/CT scanners of 2.5–9.8%.

Moving from a previous experience in a PET/CT imaging-based multicenter clinical trial [29] conducted on behalf of FIL (Italian Lymphoma Foundation), since 2010 onward we designed and adopted a framework for the PET/CT-based clinical trials, based on three main assumptions. The first is a central PET/CT scan review by a panel of Nuclear Medicine experts to reduce the image interpretation variability [29–32], the second is the use of a standard protocol, shared among PET sites for patient preparation and PET/CT acquisition according to European Association of Nuclear Medicine (EANM) guidelines [14] and, the third, a program for Clinical Trial Qualification (CTQ) of PET/CT scanners. We report here the results of a study conducted on behalf of FIL aimed to compare the results of an innovative method for PET sites CTQ with a ^{68}Ge pre-filled phantom with that obtained with traditional ^{18}F phantom prepared in the PET/CT sites.

2. Material and methods

CTQ was performed first with a ^{18}F phantom on 83 scanners and with a ^{68}Ge pre-filled phantom on 17 scanners.

2.1. Core lab activities with ^{18}F phantoms

All the PET/CT sites participating in PET/CT-based clinical trials conducted on behalf of FIL had to undergo the CTQ coordinated by a central imaging core lab in collaboration with the Italian Associations of Nuclear Medicine (AIMN) and Medical Physics (AIFM). The central core lab was located at the medical physics department of Santa Croce Hospital in Cuneo, where all the CTQ procedures were reviewed. The CTQ required the scanning of two phantoms:

1) Uniformity phantom: a difference between calculated and measured Background Activity Concentration (BAC) lower than 10% was required to assure the correct calibration of the PET/CT scanner in the uniform area of the cylindrical phantom;
2) NEMA/IEC Image Quality (IQ) phantom: a smooth and regular recovery coefficient (RC) curve within the limit presented in the EANM Guidelines [14] was required to assure a good image quality on the six spheres filled with a nominal SBR of 4.

Local personnel scanned the phantoms with the protocol for acquisition and reconstruction used for routine patient imaging. All DICOM images were uploaded to the WIDEN® (Dixit, Torino, Italy) core lab WEB portal [29].

2.2. Core lab activities with ^{68}Ge phantoms

The NEMA/IEC IQ phantom (manufactured by Data Spectrum, Durham, NC) without the "lung" cylinder (see Fig. 1) was filled with ^{68}Ge in an epoxy matrix (Ecklert & Ziegler, Valencia, California). The activity concentrations of the radioactive epoxy added inside each sphere and in the phantom were measured with a radionuclide calibrator tested against a NIST traceable source and were respectively 40.67 kBq/ml and 10.58 kBq/ml at reference time with an uncertainty of ±3%. Nominal Sphere to Background Ratio (SBR) was 3.84. Total activity in the phantom was 108.4 MBq. The ^{68}Ge phantom was imaged with a high resolution CT to demonstrate the absence of air gaps, as previously described [28]. It was then shipped via an authorized courier (Campoverde, Milano, Italy) to the PET/CT sites.

2.3. Measurements and data analysis

BAC, measured in kBq/ml, was defined as the average on six 37 mm diameter circular Regions of Interest (ROI) placed in the uniform area of uniformity and IQ phantoms, far away, at least 2 cm, from the spheres and the phantom's edge. SBR was defined as the ratio between maximum activity concentration of the larger sphere and the BAC. RC was calculated as the ratio between measured maximum and actual activity concentration in each sphere. RC curves were obtained plotting the single RC values as a function of the sphere diameter. The inter-scanner variability (ISV) was defined as the 95% confidence interval (CI) of BAC and SBR. Student's *t*-test was used to compare paired samples. The measurements performed at different time points were scaled to the activity at the reference time, accounting for ^{68}Ge decays. ^{18}F and ^{68}Ge have a half-life of approximately 110 min and 271 days.

3. Results

3.1. Core lab activities with ^{18}F phantoms

Seventy-four sites equipped with 83 PET/CT scanners participated in the ^{18}F phantom CTQ program. Sixty-three out of 83 (76%) scanners fulfilled the CTQ requirements, 14 (17%) did not because of a lack of phantoms or trained personnel, while CTQ is still ongoing on 6 (7%) scanners. For qualified scanners the CTQ was reached at the first round in 28% of the cases, while in 18%, 17% and 13%, two, three or more than three iterations, were required, respectively. Iteration was defined as a dialogue/discussion between a PET/CT site and the core lab requiring new measurements or the re-acquisition of the phantom. The iterations

were due to several reasons, the more frequent were: the NEMA/IEC IQ and uniformity phantoms were not available at the PET/CT site, the personnel performing the acquisition was not available or not experienced in performing PET/CT quality control requiring training from corelab to accomplish the task, the image data were erroneously uploaded and/or the data were missing or incorrect, the activity used to prepare the phantom was missing or incorrect, the scanners need to be re-calibrated, and the standard source for dose calibrators checking was absent.

For the PET/CT scanners fulfilling the CTQ the difference (mean ± standard deviation) between measured and expected BAC in the uniformity phantom was −1.1 ± 4.1% (CI 95%: −9.1% +6.9%, ISV = 16.1%). The difference between measured and expected BAC in the IQ phantom was 5.6 ± 20.9% (CI 95%: −35.4% + 46.6%, ISV = 81.9%). The difference between measured versus expected SBR in the IQ phantom was 6.2 ± 12.0% (CI 95%: −17.3% to +29.7%, ISV = 47.0). In Fig. 3 the curves for average RC of all the PET/CT scanners are shown.

3.2. Core lab activities with ^{68}Ge phantom

Both phantoms underwent 1-h acquisition for ten times on the same scanner at the core lab to ensure that they provide similar quantitative results under controlled conditions of acquisition and reconstruction. The preparation time for ^{18}F phantom was 108 ± 23 min. No preparation is needed for the ^{68}Ge phantom. Differences in average BAC and SBR in the two phantoms were 0.5 ± 0.8% (p = 0.66) and 3.7 ± 0.8% (p = 0.49). No statistically significant differences were appreciable in RC curves as well as seen in Fig. 2. Variability in 10 different acquisitions was 24.6% and 3.3% for BAC and was 30.3% and 9.0% for SBR, for ^{18}F and ^{68}Ge phantoms respectively.

The ^{68}Ge phantom was then circulated across 17 scanners. Differences between measured and expected BAC and SBR were −2.8 ± 5.0% (CI 95%: −13.4% +7.8%; ISV = 21.2%) and 5.0 ± 3.9% (CI 95%: −3.3% +13.3%; ISV = 16.5%), respectively. Three out of 17 scanners were re-calibrated after iteration with core lab because difference between expected and measured BAC was higher than what previewed by the CTQ, which is 10%. The box and whiskers plot of the percentage difference between expected and measured values of the BAC and SBR obtained with IQ ^{18}F and ^{68}Ge phantoms is shown in Fig. 4.

In 11 PET/CT scanners it was possible to compare the results obtained with the ^{18}F and ^{68}Ge IQ phantoms: the differences between expected and measured BAC were significantly reduced when using the ^{68}Ge phantom: −2.6 ± 4.6% (CI 95%: −12.8% +7.6%; ISV = 20.5%) in comparison to the same differences obtained with the ^{18}F phantom −7.8 ± 16.6% (CI 95%: −44.829.2%; ISV = 74.0%) (p = 0.29). Similarly, the differences between expected and measured SBR were significantly reduced when using the ^{68}Ge phantom 5.7 ± 3.9% (CI 95%: −3.0% + 14.4%; ISV = 17.4%) in comparison to the same differences obtained with the ^{18}F phantom: 13.0 ± 14.2% (CI 95%: −18.6% + 44.6%; ISV = 63.3) (p = 0.52). The differences between expected and measured BAC in the uniform phantom used for CTQ were −0.68 ± 3.07% (CI 95%: −7.5% + 6.2%; ISV = 13.7%) for the ^{18}F phantom. The box and whiskers plot of the percentage difference between expected and measured values of the BAC and SBR obtained with IQ ^{18}F and ^{68}Ge phantoms in these 11 PET/CT scanners is shown in Fig. 5.

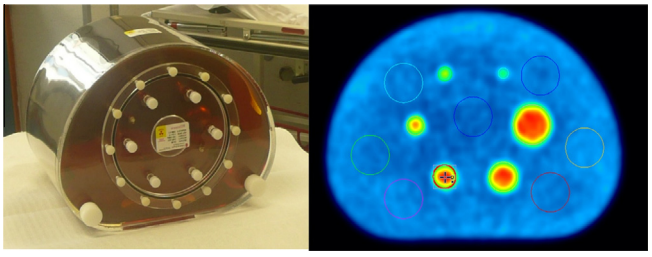

Figure 1. The ^{68}Ge phantom used in this investigation (left) and a PET transaxial image.

4. Discussion

Q-PET is increasingly used in Oncology: with a number of peer-reviewed publication peaking 28.000 in 2012 [33]. Q-PET, based on SUV measurement, allows the proposal of new SUV-derived metrics for tumour burden assessment, not only in lymphoma [34], but also in a number of solid tumours [35], such as lung cancer [36], oesophageal cancer [37], head and neck squamous carcinoma [38], breast cancer [39] and rectal cancer [40]. The proposed indexes for quantitative measurement of viable tumour bulk (Metabolic Tumour Volume, MTV) and active glycolytic tissue (Total Lesion Glycolysis, TLG) prompted clinicians with reproducible and reliable tools for tumour prognostication at baseline in Hodgkin lymphoma [41] and in diffuse large B-cell lymphoma [42], albeit with conflicting results [43]. Quantitative evaluation proved superior to qualitative criteria for early tumour response assessment in diffuse large B-cell lymphoma [44]. These data, however, are of limited value since they were generated in single-centre prospective or multicentre retrospective trials. Not surprisingly, published results of prospective multicentre clinical trials based on Q-PET are still lacking [34], as CTQ procedures are complex and time-consuming. Several factors are known to be responsible for the high variation in Q-PET among PET/CT sites [22], and only a thorough and reproducible CTQ could reduce and quantify this systematic bias [14].

The EANM program for site of excellence requires that the local personnel of the PET/CT sites scans the ^{18}F uniform and IQ phantoms and send the images to the EARL ltd core lab, which verifies the calibration accuracy and analyses the RC curve results. A ± 10% difference in the BAC and RC curves standing between the minimum and maximum RC [14] was reached.

The UK program [23], ran among 15 PET/CT sites, requires that the same medical physicist from the corelab scans a ^{18}F IQ phantom, and verifies the calibration accuracy and analyses the RC curve. ±10% difference in BAC and ±0.25 SUV variation in RC was reached.

The results obtained in these CTQ procedures are equivalent to ours: the difference between expected and measured BAC was lower than ±10% (range −6.9 to +9.1%) and ISV was 16.1%. But when BAC was measured in the NEMA IQ phantom, which was not used for qualification, we observed a much higher variability, with ISV = 81.9%. RC curves were between the minimum and maximum RC limits [14] as seen in Fig. 3. The difference between expected and measured SBR was 6.2 ± 12.0%. The relatively high variability of SBR, ISV = 47.0, was accounted by the high probability of error in the filling procedures and the objective difficulty in phantom preparation. In particular the highest variability was found in small-sized spheres.

The low PET/CT site compliance of the ^{18}F phantom-based CTQ program encouraged us to adopt the ^{68}Ge phantom approach, so as to simplify the process and further reduce the inter-scanner variability. The idea to measure the instrumentation factors using a long half-life isotope phantom has been recently addressed in several publications [19,28]. The variability of 10% is the minimum achievable with the standard approach, while at least 5% should be a requirement for using PET/CT in a quantitative way [45], as we are planning in the FIL clinical trials. An optimal inter-scanner

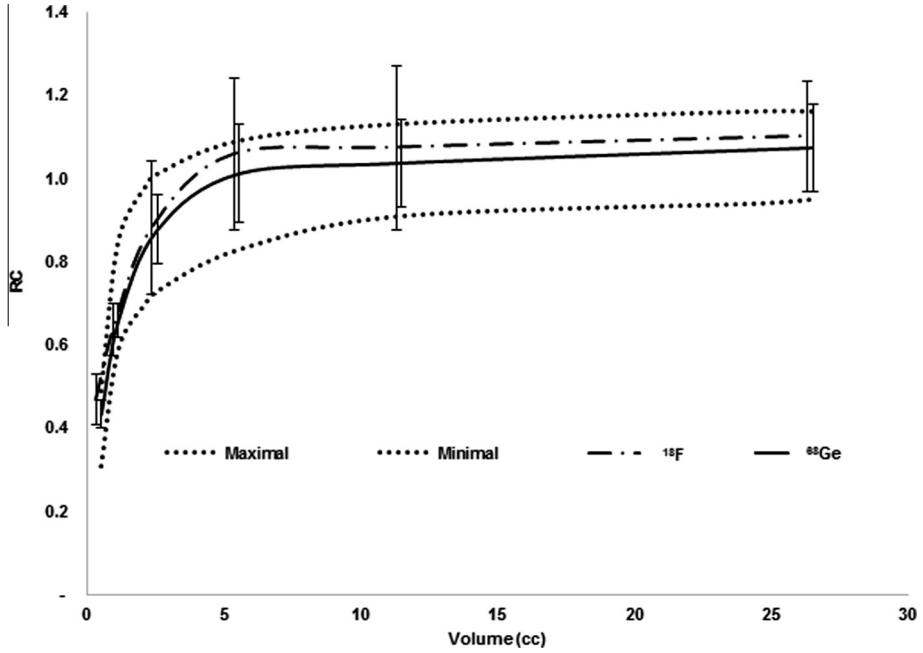

Figure 2. RC curves of ^{18}F and ^{68}Ge IQ phantom. Error bars are one standard deviation on 10 measurements. Minimal and maximal recovery curve (RC) are taken from EANM guidelines [14].

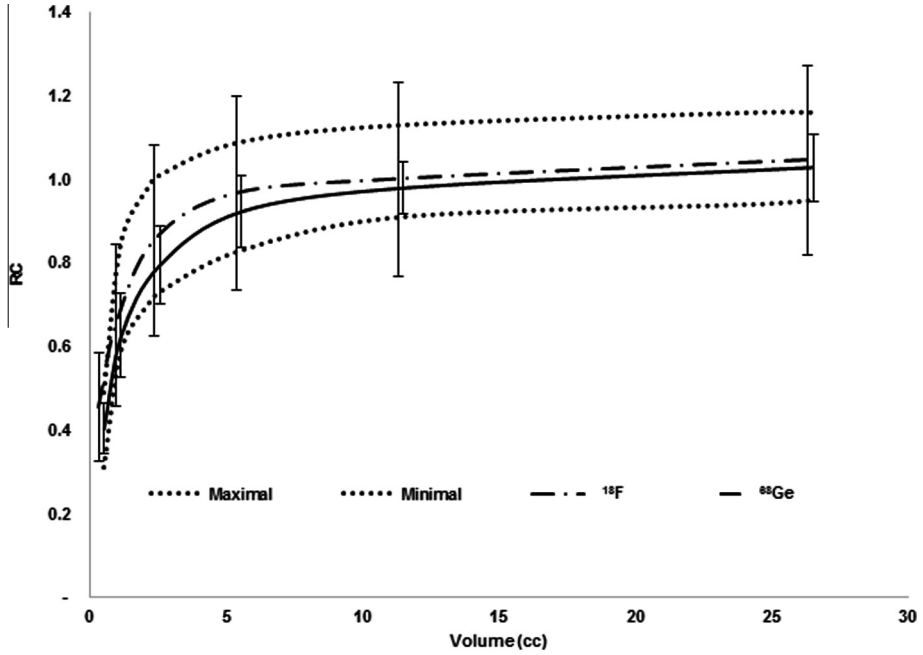

Figure 3. RC curves on all PET scanners that underwent CTQ with ^{18}F and ^{68}Ge IQ phantom. Error bars are one standard deviation on all PET/CT scanners. Minimal and maximal RC are taken from EANM guidelines [14].

variability of 3% has been already demonstrated comparing two PET/CT scanners [46] with a NIST traceable source and an uncertainty as low as 1.1% was also reported [47,48] using a new calibration methodology.

Besides the high variability, several PET sites (17% of scanners) declared themselves unable to accomplish the CTQ, and up to 40% of the scanners required repeated iterations with the core lab while only 28% of them were qualified in a single round. The more critical aspects were the time needed for the CTQ procedure and the absence of dedicated and trained personnel to perform the tests. Some sites lacked the dedicated uniformity or IQ phantoms, some could not cover the cost of the procedure for ^{18}F phantom preparation, and some lacked time to dedicate to quality control of PET/CT scanners. These problems were resolved with the ^{68}Ge phantom. It was shipped to the sites as a sealed source. Unpacking and positioning the phantoms required about 5 min as compared to the nearly 2 h needed for ^{18}F phantom preparation. Moreover, the radiation exposure for the personnel was definitely lower and the risk of contamination during phantom manipulation was eliminated. The acquisition was performed exactly in the same way and with the same time needed for a patient. At the end of CTQ the ^{68}Ge phantom was packed and shipped to the next PET/CT site. The

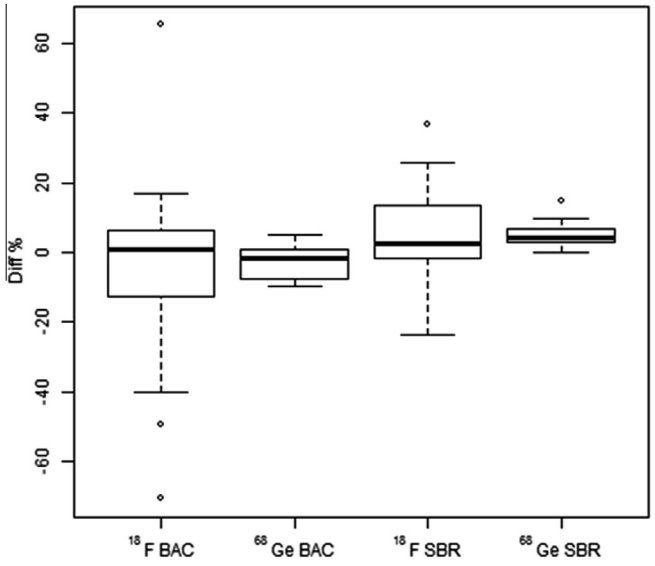

Figure 4. box and whiskers plot of the percentage difference between expected and measured BAC, in ^{18}F (a) and ^{68}Ge (b), and SBR, in ^{18}F (c) and ^{68}Ge (d), measured in IQ phantom for all PET/CT scanners.

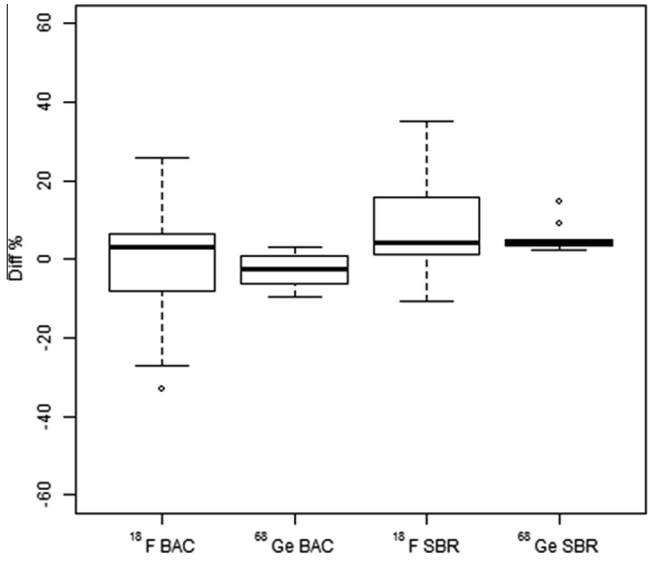

Figure 5. box and whiskers plot of the percentage difference between expected and measured BAC, in ^{18}F (a) and ^{68}Ge (b), and SBR, in ^{18}F (c) and ^{68}Ge (d), measured in IQ phantom for the PET/CT scanners that acquired images of both phantoms.

image analysis was performed remotely in the core lab with known data and no additional information was required from the PET site. Finally CTQ was fulfilled at first round in all cases. Notably, 6 PET/CT sites that were unable to comply with the ^{18}F phantom CTQ were qualified with the ^{68}Ge phantom.

Using the ^{68}Ge phantom we observed a variability in BAC (ISV = 21.2%) 1/4 lower than in ^{18}F phantom (ISV = 81.9%). The variability in SBR (ISV = 16.5%) was 1/3 of what was found with ^{18}F phantoms (ISV = 47.2%). The drastic drop in variability was only due to phantom preparation. Indeed, the phantom variability is high even if multiple ^{18}F phantoms are prepared in a single institution by the same experienced personnel with a variability of 24.6%. Initially, at the start of the CTQ program, a re-calibration of the PET/CT scanners was required when BAC variability was higher than 10%. The experience with the ^{68}Ge phantom prompted us to require a re-calibration in case of an observed BAC variability higher than 3–5%. In this study we did not ship a radionuclide calibrator source along with ^{68}Ge phantom differently from Bouchet et al. [49], who, by combining a radionuclide activity calibration check and a ^{18}F-filled uniform phantom, demonstrated an inter-scanner variability of the 11 PET/CT scanners lower than 10%. Therefore, within our experimental framework it was not possible to separate the bias coming from an inaccurate radionuclide calibrator with respect to the bias due to the whole calibration process. Noteworthy, the overall bias of PET/CT scanner does not cancel out when considering radionuclide activity calibrator and PET/CT scanner separately. Indeed, Doot et al.[50] showed that bias in radionuclide calibrator measurements ranging from −50% to 9% and in BAC ranging frm −27% to 13% lead a corresponding error in SUV measurements from −20 to 47%. A CTQ program is on-going for a clinical trial for follicular lymphoma of the Swiss Oncological Society (SAKK) in which a standard source will be shipped along with the ^{68}Ge phantom. The large reduction in inter-scanner variability was confirmed also when comparing only the 11 PET/CT scanners that performed the CTQ both with the traditional ^{18}F and with the ^{68}Ge IQ phantoms. The ISV decreased from 74.0% to 20.5% in BAC and from 63.3% to 17.4% in SBR.

The results of this study must be interpreted in the light of one limitation: most accreditation schemes require annual measurements to give an idea of intra-scanner variability over time. This was not done in the present study. Only short-time inter-scanner variability was evaluated with 10 repeated measurements on the same scanner in a day. Variability was below 3%. Measurements were also carried out 10 times over a month in the first PET/CT site and variability was still below 3%.

5. Conclusions

In conclusion, our work proved that a ^{68}Ge phantom is a realistic and valid alternative to ^{18}F phantom for image quality assessment in a multicentre clinical trial environment. All the metrics used for image quality assessment, such as BAC, SBR and RC, could be easily measured with the ^{68}Ge phantom with low noise scan. Moreover, using the ^{68}Ge phantom, a much lower radiation exposure is expected for the personnel. The ^{68}Ge phantom simplifies dramatically the procedure of inter-scanner calibration and reduces impressively the inter-scanner variability permitting to achieve a higher accuracy for future Q-PET-based clinical trials.

Funding

This work was supported by the Italian Ministry of Health with Grant RF-CROB-2008-1146343.

Disclosure

SC is co-founder of the spin-off of University of Torino and Istituto Nazionale di Fisica Nucleare, Dixit Ltd. All remaining authors have declared no conflict of interest.

References

[1] Jerusalem G, Hustinx R, Beguin Y, Fillet G. The value of positron emission tomography (PET) imaging in disease staging and therapy assessment. Ann Oncol 2002;13(Suppl 4):227–34.
[2] 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 (Suppl 1):122S–50S. http://dx.doi.org/10.2967/jnumed.108.057307.
[3] Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. J Nucl Med 2009;50(Suppl 1):21S–30S. http://dx.doi.org/10.2967/jnumed.108.057190.
[4] Weber WA. Assessing tumor response to therapy. J Nucl Med 2009;50(Suppl 1):1S–10S. http://dx.doi.org/10.2967/jnumed.108.057174.

[5] Juweid ME, Cheson BD. Positron-emission tomography and assessment of cancer therapy. N Engl J Med 2006;354:496–507. http://dx.doi.org/10.1056/NEJMra050276.

[6] Duehrsen U, Hüttmann A, Müller S, Hertenstein B, Kotzerke J, Mesters R, et al. Positron Emission Tomography (PET) Guided Therapy of Aggressive Lymphomas – a Randomized Controlled Trial Comparing Different Treatment Approaches Based on Interim PET Results (PETAL Trial). Blood 2014;124:391.

[7] Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. Eur J Cancer 1999;35:1773–82. http://dx.doi.org/10.1016/S0959-8049(99)00229-4.

[8] Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med 2009;50(Suppl 1):11S–20S. http://dx.doi.org/10.2967/jnumed.108.057182.

[9] El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit 2009;42:1162–71. http://dx.doi.org/10.1016/j.patcog.2008.08.011.

[10] Kumar V, Nath K, Berman CG, Kim J, Tanvetyanon T, Chiappori AA, et al. Variance of SUVs for FDG-PET/CT is greater in clinical practice than under ideal study settings. Clin Nucl Med 2013. http://dx.doi.org/10.1097/RLU.0b013e318279ffdf.

[11] Vanderhoek M, Perlman SB, Jeraj R. Impact of different standardized uptake value measures on PET-based quantification of treatment response. J Nucl Med 2013;54:1188–94. http://dx.doi.org/10.2967/jnumed.112.113332.

[12] Harrison RL, Elston BF, Doot RK, Lewellen TK, Mankoff DA, Kinahan PE. A Virtual Clinical Trial of FDG-PET imaging of breast cancer: effect of variability on response assessment. Transl Oncol 2014;7:138–46. http://dx.doi.org/10.1593/tlo.13847.

[13] Delbeke D, Coleman RE, Guiberteau MJ, Brown ML, Royal HD, Siegel BA, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. J Nucl Med Off Publ Soc Nucl Med 2006;47:885–95.

[14] 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;37:181–200. http://dx.doi.org/10.1007/s00259-009-1297-4.

[15] Graham MM, Badawi RD, Wahl RL. Variations in PET/CT methodology for oncologic imaging at U.S. academic medical centers: an imaging response assessment team survey. J Nucl Med 2011;52:311–7. http://dx.doi.org/10.2967/jnumed.109.074104.

[16] Beyer T, Czernin J, Freudenberg LS. Variations in clinical PET/CT operations: results of an international survey of active PET/CT users. J Nucl Med 2011;52:303–10. http://dx.doi.org/10.2967/jnumed.110.079624.

[17] Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. AJR Am J Roentgenol 2010;195:310–20. http://dx.doi.org/10.2214/AJR.10.4923.

[18] Boellaard R, Oyen WJG, Hoekstra CJ, Hoekstra OS, Visser EP, Willemsen AT, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. Eur J Nucl Med Mol Imaging 2008;35:2320–33. http://dx.doi.org/10.1007/s00259-008-0874-2.

[19] Fahey FHF, Kinahan PEP, Doot RRK, Kocak M, Thurston H, Poussaint TY. Variability in PET quantitation within a multicenter consortium. Med Phys 2010;37:3660–6. http://dx.doi.org/10.1118/1.3455705.

[20] Geworski L, Knoop BO, de Wit M, Ivancević V, Bares R, Munz DL. Multicenter comparison of calibration and cross calibration of PET scanners. J Nucl Med 2002;43:635–9.

[21] Boellaard R, Hristova I, Ettinger S, Sera T, Stroobants S, Chiti A. EARL FDG-PET/CT accreditation program: feasibility, overview and results of first 55 successfully accredited sites. J Nucl Med 2013;54(Suppl. 2):2052.

[22] Zijlstra JM, Boellaard R, Hoekstra OS. Interim positron emission tomography scan in multi-center studies: optimization of visual and quantitative assessments 2009.

[23] Barrington SF, Mackewn JE, Schleyer P, Marsden PK, Mikhaeel NG, Qian W, et al. Establishment of a UK-wide network to facilitate the acquisition of quality assured FDG-PET data for clinical trials in lymphoma. Ann Oncol 2011;22:739–45. http://dx.doi.org/10.1093/annonc/mdq428.

[24] Scheuermann JS, Saffer JR, Karp JS, Levering AM, Siegel BA. Qualification of PET scanners for use in multicenter cancer clinical trials: the American College of Radiology Imaging Network experience. J Nucl Med 2009;50:1187–93. http://dx.doi.org/10.2967/jnumed.108.057455.

[25] Christian P. Use of a precision fillable clinical simulator phantom for PET/CT scanner validation in multi-center clinical trials: the SNM Clinical Trials Network (CTN) Program. J Nucl Med 2012;53(suppl):437. 2012.

[26] Sunderland JJ, Christian PE. Quantitative PET/CT scanner performance chracterization based upon the SNMMI Clinical Trial networ oncology clinical simulator phantom. J Nucl Med 2015;56:145–52.

[27] Daisaki H, Tateishi U, Terauchi T, Tatsumi M, Suzuki K, Shimada N, et al. Standardization of image quality across multiple centers by optimization of acquisition and reconstruction parameters with interim FDG-PET/CT for evaluating diffuse large B cell lymphoma. Ann Nucl Med 2012. http://dx.doi.org/10.1007/s12149-012-0676-2.

[28] Doot RK, Scheuermann JS, Christian PE, Karp JS, Kinahan PE. Instrumentation factors affecting variance and bias of quantifying tracer uptake with PET/CT. Med Phys 2010;37:6035. http://dx.doi.org/10.1118/1.3499298.

[29] Chauvie S, Biggi A, Stancu A. WIDEN: a tool for medical image management in multicenter clinical trials. Clin Trials 2014:1–7.

[30] Barrington S, Qian W, Somer E, Franceschetto A, Bagni B, Brun E, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. Eur J Nucl Med Mol Imaging 2010;37:1824–33.

[31] Gallamini A, Barrington SF, Biggi A, Chauvie S, Kostakoglu L, Gregianin M, et al. The predictive role of interim Positron Emission Tomography on Hodgkin lymphoma treatment outcome is confirmed using the 5-point scale interpretation criteria. Haematologica 2014. http://dx.doi.org/10.3324/haematol.2013.103218.

[32] Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, et al. International validation study for interim PET in ABVD-treated, advanced-stage Hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. J Nucl Med 2013;54:683–90. http://dx.doi.org/10.2967/jnumed.112.110890.

[33] Zaidi H, Alavi A. Trends in PET quantification: opportunities and challenges. Clin Transl Imaging 2014;2:183–5. http://dx.doi.org/10.1007/s40336-014-0065-z.

[34] Kostakoglu L, Chauvie S. PET-derived metabolic volume metrics in lymphoma. Clin Transl Imaging 2015. http://dx.doi.org/10.1007/s40336-015-0135-x.

[35] Tylski P, Stute S, Grotus N, Doyeux K, Hapdey S, Gardin I, et al. Comparative assessment of methods for estimating tumor volume and standardized uptake value in (18)F-FDG PET. J Nucl Med 2010;51:268–76. http://dx.doi.org/10.2967/jnumed.109.066241.

[36] Obara P, Pu Y. Prognostic value of metabolic tumor burden in lung cancer. Chin J Cancer Res 2013;25:615–22. http://dx.doi.org/10.3978/j.issn.1000-9604.2013.11.10.

[37] Shum W-Y, Ding H-J, Liang J-A, Yen K-Y, Chen S-W, Kao C-H. Use of pretreatment metabolic tumor volumes on PET-CT to predict the survival of patients with squamous cell carcinoma of esophagus treated by curative surgery. Anticancer Res 2012;32:4163–8.

[38] Abgral R, Keromnes N, Robin P, Le Roux P-Y, Bourhis D, Palard X, et al. Prognostic value of volumetric parameters measured by 18F-FDG PET/CT in patients with head and neck squamous cell carcinoma. Eur J Nucl Med Mol Imaging 2014;41:659–67. http://dx.doi.org/10.1007/s00259-013-2618-1.

[39] Kajáry K, Tőkés T, Dank M, Kulka J, Szakáll S, Lengyel Z. Correlation of the value of 18F-FDG uptake, described by SUVmax, SUVavg, metabolic tumour volume and total lesion glycolysis, to clinicopathological prognostic factors and biological subtypes in breast cancer. Nucl Med Commun 2015;36:28–37. http://dx.doi.org/10.1097/MNM.0000000000000217.

[40] Jo HJ, Kim S-J, Lee HY, Kim IJ. Prediction of survival and cancer recurrence using metabolic volumetric parameters measured by 18F-FDG PET/CT in patients with surgically resected rectal cancer. Clin Nucl Med 2014;39:493–7. http://dx.doi.org/10.1097/RLU.0000000000000438.

[41] Kanoun S, Rossi C, Berriolo-Riedinger A, Dygai-Cochet I, Cochet A, Humbert O, et al. Baseline metabolic tumour volume is an independent prognostic factor in Hodgkin lymphoma. Eur J Nucl Med Mol Imaging 2014;41:1735–43. http://dx.doi.org/10.1007/s00259-014-2783-x.

[42] Xie M, Wu K, Liu Y, Jiang Q, Xie Y. Predictive value of F-18 FDG PET/CT quantization parameters in diffuse large B cell lymphoma: a meta-analysis with 702 participants. Med Oncol 2015;32:446. http://dx.doi.org/10.1007/s12032-014-0446-1.

[43] Gallicchio R, Mansueto G, Simeon V, Nardelli A, Guariglia R, Capacchione D, et al. F-18 FDG PET/CT quantization parameters as predictors of outcome in patients with diffuse large B-cell lymphoma. Eur J Haematol 2014;92:382–9. http://dx.doi.org/10.1111/ejh.12268.

[44] Meignan M, Sasanelli M, Casasnovas RO, Luminari S, Fioroni F, Coriani C, et al. Metabolic tumour volumes measured at staging in lymphoma: Methodological evaluation on phantom experiments and patients. Eur J Nucl Med Mol Imaging 2014;41:1113–22. http://dx.doi.org/10.1007/s00259-014-2705-y.

[45] Boellaard R. Methodological aspects of multicenter studies with quantitative PET. Methods Mol Biol 2011;727:335–49.

[46] Lockhart CM, MacDonald LR, Alessio AM, McDougald WA, Doot RK, Kinahan PE. Quantifying and reducing the effect of calibration error on variability of PET/CT standardized uptake value measurements. J Nucl Med 2011;52:218–24. http://dx.doi.org/10.2967/jnumed.110.083865.

[47] Zimmerman BE, Cessna JT. Development of a traceable calibration methodology for solid (68)Ge/(68)Ga sources used as a calibration surrogate for (18)F in radionuclide activity calibrators. J Nucl Med 2010;51:448–53. http://dx.doi.org/10.2967/jnumed.109.070300.

[48] Zimmerman BE, Pibida L, King LE, Bergeron DE, Cessna JT, Mille MM. Development of a calibration methodology for large-volume, solid 68 Ge phantoms for traceable measurements in positron emission tomography. Appl Radiat Isot 2013:1–5. http://dx.doi.org/10.1016/j.apradiso.2013.11.049.

[49] Bouchet F, Geworski L, Knoop BO, Ferrer L, Barriolo-Riedinger A, Millardet C, et al. Calibration test of PET scanners in a multi-centre clinical trial on breast cancer therapy monitoring using 18F-FLT. PLoS ONE 2013;8:e58152. http://dx.doi.org/10.1371/journal.pone.0058152.

[50] Doot RK, Pierce LA, Byrd D, Elston B, Allberg KC, Kinahan PE. Biases in multicenter longitudinal PET standardized uptake value measurements. Transl Oncol 2014;7:48–54.